An Exploration into the Value of Protective Factors in Violence Risk Assessment of Psychiatric Inpatients

Rachel Claire Judges, Bsc., Msc.

Thesis submitted to the University of Nottingham for the degree of

Doctorate in Forensic Psychology (D.Foren.Psy)

May 2016

Abstract

This thesis explores the value of including protective factors in the violence risk assessment and risk management processes of forensic mental health services. More specifically it investigates whether assessment of protective factors improves predictive accuracy of violence risk assessment tools, and discusses the implications for clinical practice. The impact on patient motivation to change is also considered. A critique is presented of the Historical Clinical Risk-20 Version 3 (HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013), one of the most popular and widely used violence risk assessment tools. Despite its popularity and good measurable properties, the HCR-20V3 does not include an assessment of protective factors. A systematic review examined research investigating the predictive accuracy of the three violence risk assessment tools recommended for use in forensic mental health services in the National Health Service: HCR-20V3, the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bourman, & de Vries Robbé, 2012), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004). The SAPROF had superior predictive accuracy of absence of violence compared to the other measures; however, limited reliability and validity evidence was found for its use in English forensic inpatient settings. An empirical research project conducted a prospective validation study of the SAPROF, also reporting on the reliability and validity of the measure across a number of domains, and in relation to the HCR-20V3 and START. The SAPROF demonstrated better absence of violence risk predictive abilities than the HCR-20V3 and the START (presence of violence risk);

combined use of the SAPROF and HCR-20V3 significantly increased predictive accuracy of presence of violence risk. Finally, a single case study explores the impact of collaborative risk assessment and management training on a patient's motivation to engage in treatment and interventions to manage risk. Collaborative risk assessment had a positive impact on motivation; however it was not reliably or clinically significant. This thesis provides positive research evidence for the inclusion of protective factors in the violence risk assessment and management process.

Acknowledgements

I would like to thank my academic supervisor, Professor Vincent Egan for his expertise and guidance throughout my thesis, and my systematic review supervisor, Dr Nigel Hunt, for his direction and support. Thank you to Dr Grant Broad, my co-ordinating supervisor, and to Andy Inett, my placement supervisor, for their encouragement and teachings throughout my placement experiences.

I express my extreme gratitude to Dr Lona Lockerbie and Mr Kevin Halpin, without whom the opportunity for me to train to be a Forensic Psychologist would not have been possible. Special thanks go to Dr Lauren Clark for cultivating my reflective skills, and inspiring my practice. I also extend my thanks to all the staff I have worked during my training.

I would like to thank my family and friends for their unwavering support throughout this process and in particular my husband, Mr James Judges, for his never ending encouragement and belief that I would succeed.

Finally, I would like to express my appreciation to the patients I have worked with. I thank them for sharing their experiences with me.

Table of contents

List of Tables

Chapter Two

Chapter Three

Chapter Four

List of Figures

List of Appendices

Chapter Three

Chapter Five

Chapter One

Overview

1.1. Introduction

Violence risk assessment and risk management is a core component of forensic mental healthcare (Rogers, 2000). To provide effective care clinicians need to be aware of the individual's overall needs, and the level of risk they may pose towards others or to themselves (Department of Health; DoH 2007a). Clinicians have a duty of care to reduce and manage any risks to assist in improving an individual's quality of life and recovery (DoH, 2007a). When violence risk is involved there is an additional duty of care towards other patients, professionals, and society as a whole (DoH, 2007a). There is no widely accepted definition of violence. For the purpose of this thesis the following definition, taken from the Historical Clinical Risk-20 Version 3 (HCR-20V3), was applied: '…actual, attempted, or threatened infliction of bodily harm of another person' (Douglas, Hart, Webster, & Belfrage, 2013, p. 2). This includes both physical and serious psychological harm, and is consistent with the World Health Organisation's definition of interpersonal violence (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). Acts of collective violence committed during war, terrorism, or gang conflict are excluded.

Decisions made about how to manage potential for violence should be based on 'knowledge of the research evidence, knowledge of the individual patient and their social context, knowledge of the patient's own experience, and clinical judgement' (p. 7, DoH, 2007a). This thesis aims to examine the research evidence for the inclusion of protective factors in the violence risk assessment and management process.

1.2. What is violence risk assessment and risk management?

Effective treatment of offenders, including those with mental illness, is directed by the principles of risk, need, and responsivity (RNR model; Andrews, Bonta, & Hoge, 1990). The principles allow interventions to be tailored to the patient's individual risks, needs, and personal circumstances or conditions to most effectively reduce and manage risk (Andrews & Bonta, 2006).

The 'risk principle' states that treatment should reflect the level of risk posed by the individual (Andrews & Bonta, 2006). As such, high risk individuals should receive the most intense treatment, and lower risk individuals should be offered less intense treatment. Risk can be defined as the likelihood, imminence, and severity of the occurrence of a negative event (DoH, 2007a). The level of risk is usually determined by the completion of a risk assessment. This identifies static (unchangeable) and dynamic (changeable) personal characteristics or circumstances which may lead to the negative event occurring (Kraemer, Kazdin, Offord, Kessler, Jenson, & Kupfer, 1997). The definition of risk assessment adopted by Whittington and Logan (2011, p. 295) was applied for this thesis: 'The process of gathering information via personal interviews, psychological/medical testing, review of case records, and contact with collateral informants, for use in making decisions pertaining to an individual's risk and its most appropriate, effective, and proportionate prevention or minimization'.

There are three main approaches to risk assessment: unstructured clinical judgement, actuarial, and structured professional (or clinical)

judgement (SPJ). The unstructured clinical judgement approach relies on human decision making, and there are no guidelines regarding what information the clinician should include or discount (Dolan & Doyle, 2002). Decisions are based on the facts available, the clinician's knowledge of the individual, and their intuition and instincts. This information is combined with what the clinician feels is relevant to predict future violence (Mossman, 2004). Some argue this approach is subjective, and it has been demonstrated to have low inter-rater reliability (Hart, 1998).

In contrast, the actuarial method discounts clinical judgement (Hart, 1998), and instead decisions are made based on statistical methods, and empirically based risk factors related to violent behaviour (Doyle & Dolan, 2002). This methodology increases reliability and objectivity (Heilbrun, Yasuhara, & Shar, 2010). However, Mossman (2004) argues actuarial methods are flawed because they rely on static risk factors, and do not account for dynamic factors which could be related to treatment and supervision, resulting in over-prediction of future violence.

The majority of violence risk assessment tools follow the SPJ approach (Douglas et al., 2013), and the DoH (2007a) recommend utilising a tool which follows this method. The SPJ method is a shared approach using clinical judgement and actuarial methods, combining the strengths of each (Dolan & Doyle, 2002). A final clinical judgement regarding the level of risk for future violence is made following consideration of a standardised, empirically-derived checklist of static and dynamic risk factors, clinical experience and knowledge of the patient,

and the patient's own view of their experience (de Vries Robbe, de Vogel, & Stam, 2012). This enables a risk formulation to be developed, where the clinician outlines what may trigger or heighten the risk, and what can be done to manage the risk.

According to a recent international survey (Singh, Fazel, Gueorguieva, & Buchanan, 2014) the most commonly used SPJ violence risk assessment tools are the Historical Clinical Risk-20 (HCR-20V2; Webster, Douglas, Eaves, & Hart, 1997, and HCR-20V3; Douglas et al., 2013), the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, and Webster, 1997), the Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, and Eaves, 1999), and the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, and Forth, 2006).

Risk management aims to prevent and reduce the likelihood of the negative event occurring, or at least minimise the level of potential harm (Andrews & Bonta, 2006). This operates by working with the individual to suggest specific treatment targets, developing flexible strategies to manage risk factors (DoH, 2007a). The 'need principle' states the most effective interventions and management plans are those which focus on the dynamic risk factors (Andrews & Bonta, 2006). Dynamic risk factors are features of the individual's environment and social situation, for example psychological vulnerabilities, substance use, and attitudes which increase their likelihood of risk. Effective targeting of dynamic risk factors during treatment and risk management can result in a change in the level of risk, and reduce the likelihood of reoffending (Andrews & Bonta, 2006).

The 'responsivity principle' states treatment and interventions should be delivered in a way which is consistent with the individual's strengths, for example their learning style, abilities, and personal circumstances (Andrews & Bonta, 2006). It has been suggested treatment aimed at reducing risk of violence should focus on both reducing and managing risk factors, and building and maintaining protective factors (de Ruiter & Nicholls, 2011). Rapp and Goscha (2006) report risk management is most effective when strengths and protective factors are recognised, and incorporated in management plans. This approach has gained further vigour due to the introduction of positive intervention strategies such as the Good Lives Model (GLM; Ward & Stewart, 2003). The GLM assumes that all individuals want to achieve a good life, and will do so by means which are most likely to help attain the positive outcome. The approach promotes the use of pro-social methods of attaining these positive outcomes rather than previously used anti-social methods.

1.3. The emergence of protective factors

The idea that risk factors need to be managed in order to reduce the likelihood of violence has been a longstanding principle (for example, Lombroso, 1887). More recently it has been proposed protective factors should also be considered (de Vries Robbé, 2014). De Ruiter and Nicholls (2011) state treatment to reduce violent reoffending should not only focus on reducing risk factors, but also on increasing and maintaining protective factors. The personal resilience of individuals can be increased by developing personal strengths, and risk management strategies can

be enhanced by placing more emphasis on external and situational protective factors (de Vries Robbé, de Vogel, Koster, & Bogaerts, 2015).

As yet there is no consensus as to how to define a protective factor, or how they work to reduce risk. Rutter (1985) suggests that protective factors are variables within comparable groups of high risk individuals, some with positive behavioural outcomes and others with negative behavioural outcomes, which differ. Others, for example Stattin, Romelsjo, & Stenbacka (1997), propose factors should only be considered protective if they reduce the problem behaviour when the risk is high, but have no influence when the risk is low. Jessor, Turbin, Costa, Dong, Zhang, and Wang (2003) have suggested protective factors may result in a low probability of violence in general (direct positive effect), and a low probability of violence despite the presence of high risk (mediating effect on the relationship between risk factors and violence). It is also suggeted protective factors can help individuals to endure difficult circumstances without becoming violent (increase resilience), and help those who have previously engaged in violent behaviour to desist (de Vries Robbé, 2014). For the purpose of the thesis the definition of protective factors as described by de Vogel, de Ruiter, Bouman, and de Vries Robbé (2012) was adopted: 'any characteristic of a person, his or her environment or situation which reduces the risk of future violent behaviour' (p. 23). This is because this is the definition used in the Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel et al., 2012) manual.

Research in mental health has identified a wide range of protective factors at the individual, family, and community levels which prevent

unfavourable outcomes from occurring (de Vries Robbé, 2014). According to Lösel (2001) protective factors in children and adolescents can include biological features, temperament characteristics, cognitive competencies, childrearing and attachment, school achievement and bonding, peer groups and social networks, self-related and social cognitions, and neighbourhood and community factors. A review into the development of youth violence found protective effects of various factors at the individual, family, school, peer, and neighbourhood levels (Lösel & Farrington, 2012). However, there has been limited focus on protective factors in violent adults, and what promotes desistance from their further violent offending (for example, Ullrich & Coid, 2011).

1.4. Protective factors in violence risk assessment

The best evaluations of violence risk can only explain a moderate amount of variance (Lösel, 2001). Violence risk assessment tools traditionally focus on the assessment of risk factors, but as Rogers (2000) observes, 'risk only evaluations are inherently inaccurate' (p. 589). He argues that without taking into account protective factors, individuals can be classed at a higher risk of violent behaviour, resulting in longer admissions, and increased financial costs (Rogers, 2000). Risk assessments often assist in the allocation of resources, and if protective factors are discounted an unnecessary emphasis may be placed on forensic patients, at the expense of care and treatment for non-forensic patients (Sullivan, Wells, Mortgenstern, & Leake, 1995). Miller (2006) also states that a focus on risk factors can result in longer detention of forensic patients due to its influence on the attitudes of professionals

working with them, heightening the perception that individuals are more likely to reoffend or are more dangerous than they actually are.

Assessment of protective factors in addition to risk factors in the risk assessment process is fairly new (de Vries Robbe et al., 2012). To date there are few violence risk assessment tools which incorporate the assessment of protective factors. These include the Structured Assessment of Violence Risk in Youth (SAVRY; Borum et al., 2006), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004). The HCR-20 is the one of the best validated, and most widely used violence risk assessment tools (Singh et al., 2014), but it does not include an assessment of protective factors. Clinicians using the HCR-20 are guided to make a final judgement of violence risk taking into account the presence and relevance of well-defined risk factors, and although the strengths of an individual would naturally be incorporated, it is based purely on the their clinical judgement (de Vries Robbé, 2014). In 2007 the SAPROF (de Vogel, de Ruiter, Bourman, & de Vries Robbé, 2012) was developed specifically as an assessment of protective factors, to be used alongside a risk factor focused SPJ violence risk assessment tool such as the HCR-20. The SAPROF aims to balance the risk assessment process by offering structure to clinical judgement, and an empirical foundation in terms of protective factors in the violence risk assessment and management process (de Vries Robbé, 2014).

1.5. The present thesis

This thesis aims to explore the research evidence for the inclusion of protective factors in the violence risk assessment and management process in forensic mental health services, in the National Health Service (NHS). More specifically it investigates whether inclusion of protective factors improves the predictive accuracy of violence risk assessment, and discusses the implications for clinical practice. The impact on patient motivation to change is also considered.

Chapter two provides an overview and critique of the HCR-20 version 3 (HCR-20V3; Douglas et al., 2013), exploring its measurable properties, considering its clinical and research applications, whilst also taking into account its use in forensic inpatient mental health settings. Chapter three systematically reviews the research literature on the predictive accuracy of the violence risk assessment tools recommended for use in forensic mental health services in the NHS; the HCR-20, SAPROF, and START (service specification no. C03/S/a; NHS Commissioning Board, 2013). The HCR-20 follows the traditional route of assessing the presence and relevance of risk factors, the SAPROF assesses the presence of protective factors, and the START evaluates the presence of strengths and vulnerabilities. The START is considered a short-term risk assessment (up to three months; Webster et al., 2004), whereas the HCR-20 and SAPROF are used as long-term risk assessments (between six and 12 months; Douglas et al., 2013, and de Vogel et al., 2012).

Chapter four is a research study exploring the implementation of the SAPROF in a forensic inpatient setting, to allow for a greater

understanding of its effectiveness in the risk assessment and management process. The validity and reliability of the SAPROF as an assessment tool is investigated, and compared to the HCR-20 and START across a number of domains, to evaluate whether the assessment of protective factors improves predictive accuracy of violence risk assessment.

Chapter five presents a case study which explores the impact of psycho-education on a patient's motivation to engage in treatment, and interventions to manage risk. The main objective was to establish whether risk assessment tools which follow a strength-based approach and focus on protective factors, have a positive impact on motivation to change when used in the collaborative risk assessment and risk management process.

Finally, chapter six concludes the thesis by presenting a discussion on the findings in relation to the specific aims of the thesis: 1) to critically evaluate the most widely used violence risk assessment tool, the recently updated HCR-20V3; 2) to explore the predictive accuracy of the HCR-20, SAPROF, and START in terms of absence and presence of violence; 3) to investigate the reliability and validity of the SAPROF when implemented in a NHS forensic inpatient mental health service; 4) to evaluate the impact of protective factors in assessment of violence risk on a patient's motivation to change. The clinical implications of the findings and recommendations for future research are also discussed.

Chapter Two

A Critique of the Historical Clinical Risk - 20 Version 3 Risk Assessment Instrument

2.1. Introduction

Assessing and managing the risk of violence in offenders with mental illness is essential for those working in forensic mental health services (Wilson, Desmarais, Nicholls, & Brink, 2010). Such assessment protects the public (and the individual) from harm, but also allows for resources such as supervision and treatment to be allocated appropriately, and for individuals to be managed in the least restrictive environment (Yates, Prescott, & Ward, 2010). A number of tools have been developed to assist the risk assessment and risk management process. In secure adult mental health services NHS England recommends the use of the Historical Clinical Risk - 20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997, and Douglas, Hart, Webster, & Belfrage, 2013; NHS Commissioning Board, 2013).

The HCR-20 is a comprehensive guide for the structured assessment of violence risk. Its properties have been investigated for research purposes. The HCR-20's intention is to establish the presence and relevance of risk factors, enabling the development of appropriate risk management strategies (Douglas, Shaffer, Blanchard, Guy, Reeves, & Weir, 2014). The aim of this critique is to provide an overview of the HCR-20 version 3 (HCR-20V3; Douglas et al., 2013), exploring its properties, and its clinical and research applications, whilst also taking into account its use in forensic mental health settings.

2.2. Overview of the HCR-20V3

2.2.1. Background

Clinicians working within forensic settings are often under pressure from the expectations of the general public and their professional role to predict and manage risk of future violence (Whittington et al., 2013). They are also required to consider the best interests of the individual being assessed, and develop positive risk management strategies for the institutions and services they oblige (Whittington et al., 2013). In forensic mental health services this means integrating the risk of reoffending with appropriate interventions to manage and reduce risk, otherwise known as 'risk-needs assessment' (Andrews, Bonta, & Hoge, 1990).

There is no 'gold standard' when it comes to choosing the correct risk assessment (NICE Guidelines, 2005). The Department of Health recommends the best practice of utilising a tool which follows the structured clinical (or professional) judgement approach (SPJ; Department of Health, 2007a). The SPJ model combines the strengths of actuarial methods with clinical judgement (Dolan & Doyle, 2002). SPJ involves making a judgement about risk based on an assessment of clearly defined static (historical) and dynamic (changeable) factors derived from research, clinical experience and knowledge of the service user, and the service users own view of their experience (Department of Health, 2007a). A recent international survey (Singh, Fazel, Gueorguieva, & Buchanan, 2014) found the HCR-20 (Webster et al., 1997, and Douglas et al., 2013) was one of the most commonly used SPJ tools.

2.2.2. The instrument

The HCR-20 was first pubished in 1995 (Webster, Eaves, Douglas, & Wintrup, 1995) but was updated in 1997 (version 2 or HCR-20V2; Webster et al., 1997). The instrument was constructed following careful consideration of the empirical literature concerning factors that relate to violence (Douglas et al., 2014). More recently, on the basis of extensive clinical beta testing and empirical evaluation, the HCR-20 has been further updated (version 3, or HCR-20V3; Douglas et al., 2013).

The HCR-20V3 is a comprehensive set of professional guidelines for the assessment and management of violence risk, the latter defined as `…actual, attempted, or threatened infliction of bodily harm of another person' (Douglas et al., 2013, p. 2). The HCR-20V3 assists professionals to estimate a person's likelihood of future violence, and determine the most appropriate treatment and management strategies. Its most common applications are within correctional, forensic, and general or civil psychiatric settings, whether in the institution or in the community, and it is applicable for adults aged 18 and above who may pose a risk for future violence (Douglas et al., 2013). The HCR-20V3 is not intended to act as a stand-alone assessment, and it is recommended additional measures should be employed to investigate any unique patterns of violence (Douglas et al., 2013). The authors recommend the HCR-20V3 should be repeated (between 6 and 12 monthly) to take into account changes in circumstance which are inevitable with the nature of risk (Douglas et al., 2013). To administer the HCR-20V3 the professional should be trained in conducting individual assessments, be familiar with the study of violence, and they should have experience of using SPJ tools.

The HCR-20V3 is not an actuarial tool. The guidance regarding score interpretation are not based on algorithms or cut-off scores (Douglas et al., 2014), and the authors state the HCR-20V3 is not intended to give professionals a definitive prediction of violence (Douglas et al., 2013). The outcome of the assessment is an estimate of the likelihood of future violence, and the test authors recommended this should be presented in terms of low, moderate, or high probability of violence (Douglas et al., 2013). In addition, the levels of probability should be considered in terms of imminence, time frames (short and long-term), and in relation to relevant individualised factors. These considerations can help in developing a risk management plan which reports the type and degree of risk presented by the person, and identifies interventions which may reduce the likelihood of that individual exhibiting violence in the future (Douglas et al., 2013).

The HCR-20V3 allows for evaluation of the presence (and relevance) of 20 key violence risk factors in an individual. These are organized into three areas: historical and clinical factors, and risk management. The complete list is shown in table 2.1. The presence of factors is coded using a three level response format: (Y) the factor is definitely or conclusively present; (P) the factor is possibly or partially present, or the risk factor is present, but the information is weak, contradictory, or inconclusive; and (N) the factor is absent, or the professional perceives no evidence the factor is present (Douglas et al., 2013). To rate historical factors, the professional must conduct an exhaustive review of background documents and ideally interview individuals who know the person being assessed. The rating of the

clinical factors and risk management aspects requires a clinical interview with the person being assessed. However, systematic review of the literature found only 16% of research studies included assessments with multiple information sources (see chapter three). In addition to determining the presence or absence of each risk factor, the professional should judge their relevance with respect to the development of future risk management strategies. Relevance is also coded on a three level scale: (low) the factor is of low relevance to the individual's risk for violence; (moderate) the factor is relevant to some degree; and (high) the factor is present, and its role in causing violence or impairing the effectiveness of risk management strategies is likely to be substantial (Douglas et al., 2013). Finally, there is the option to rate a final risk judgement following completion of the assessment (low, medium, and high). De Vries Robbé and de Vogel (2012) have suggested using a five-point scale instead of a three-point scale for the final risk judgement as they have found it to have higher predictive abilities.

Table 2.1

Factors in the HCR-20V3 violence risk assessment scheme

Historical factors Problems with…	**Clinical factors** Problems with…	**Risk management factors** Problems with…
H1: Violence	C1: Insight	R1: Professional services and plans
H2: Other anti-social behavior	C2: Violent ideation or intent	R2: Living situation
H3: Relationships	C3: Symptoms of major mental illness	R3: Personal support
H4: Employment	C4: Instability	R4: Treatment or supervision response
H5: Substance use	C5: Treatment or supervision response	R5: Stress or coping
H6: Major mental disorder		
H7: Personality disorder		
H8: Traumatic experiences		
H9: Violent attitudes		
H10: Treatment or supervision response		

2.2.3. Manual

The HCR-20V3 manual contains the guidelines defining the evaluation and judged presence of the 20 key violence risk factors, and their relevance to the individual being assessed. It defines each of the risk factors and describes key indicators to guide the professional to make the relevant ratings. It contains information to help professionals construct meaningful formulations of violence risk, future risk scenarios, appropriate risk management plans, and informative communication of risk. The manual also highlights how the tool was developed, and outlines research regarding its characteristics in terms of reliability and validity.

Despite detailed descriptions of the risk factors and provision of key indicators, it can be argued that there is subjectivity in the scoring guidelines (Rufino, Boccaccini, & Guy, 2011). This is important because subjectivity can result in reduced predictive validity (Rufino et al., 2011). An example of subjectivity may be the inclusion of non-violent behaviour in the assessment, such as expression of frustration, when this is not included in the definition of violence (Douglas et al., 2013). In fact, the authors of the tool state that the 'definition of violence…is both imperfect and impractical…' (Douglas et al., 2013, p.3).

2.2.4. Research

The development of the HCR-20V3 was guided by empirical literature relating to factors consistent with violence, and it integrates the experience of professionals within the field (Douglas et al., 2014). Since its first publication in 1995, the HCR-20 has been established as one of

the best validated violence risk assessments through both prospective and retrospective research designs. The literature on the HCR-20V3 will be discussed in this critique as an evaluation of its psychometric properties.

2.3. Psychometric Properties of the HCR-20V3

2.3.1. Reliability

2.3.1.1. Internal reliability

In terms of the HCR-20V3, internal reliability can be defined as the degree to which each risk factor determines future violence (i.e. consistency within the tool). See table 2.3 for synthesis of research conducted to date. The internal reliability of the HCR-20V2 has been well established. For example, using Cronbach's Alpha, Belfrage (1998) reported good internal consistency for the historical factors (.96), clinical factors (.89), risk factors (.85), and the HCR-20V2 as a whole (.95). However, there is limited research into the internal reliability of the HCR-20V3.

Bjørkly, Eidhammer, and Selmer (2014) compared the internal consistency of the HCR-20V3 with the HCR-20V2 using Interclass Correlations (ICC). Classifications of ICC values were used as follows: ICC ≥ .75 = excellent, ICC between .60 and < .75 = good, and ICC between .40 and .60 = moderate (Fleiss, 1986). Moderate to good estimates of internal consistency were found between ratings of the two assessments for the same individual, with ICC values of .85, .57, .81, and .84 for the historical factors, clinical factors, risk factors, and overall assessment respectively. Due to their similarity it could be argued the

evidence of good internal reliability for version 2 could be applied to version 3. However, although both versions reflect common underlying dimensions, there are differences, particularly within the clinical factors. A paired sample t-test also revealed significant differences between the historical and clinical factor ratings on version 2 and version 3 (Bjørkly et al., 2014). It was hypothesised this may have been due to more accurate descriptions of item indicators in the HCR-20V3 (Bjørkly et al., 2014), and ulitimately, the tools include different items and definitions.

2.3.1.2. Inter-rater reliability.

In terms of the HCR-20V3, good inter-rater reliability means two independent professionals would code the presence (and relevance) of each risk factor similarly, resulting in the same final risk judgement, in the same time period. See table 2.3 for synthesis of research conducted to date. This construct can be measured using a variety of statistical tests, including Pearson's correlation and ICCs. Classifications of ICC values were used as follows: ICC ≥ .75 = excellent, ICC between .60 and < .75 = good, and ICC between .40 and .60 = moderate (Fleiss, 1986).

Current research suggests good levels of inter-rater reliability for the HCR-20V3. Doyle, Power, Coid, Kallis, Ullrich, and Shaw (2014) investigated the inter-rater reliability of the HCR-20V3 in a sample of discharged patients (*n* = 387) from medium secure units in England and Wales. A prospective cohort design was utilised with a 12 month follow up period. ICC values were reported at the 6 and 12 month period respectively for the overall assessment (.73; .92), historical factors (.72; .91), clinical factors (.69; .90), and risk factors (.76; .93). Despite the

low ICC for clinical factors at the 6 month period (.69) this remains a good level of inter-rater reliability, and all other results show good or excellent inter-rater reliability.

Some sample sizes exploring the integrity of the HCR-20V3 have been small. Douglas and Belfrage (2014) investigated a sample of 32 forensic patients where three professionals jointly interviewed the patients and then independently completed the HCR-20V3. Using ICCs, they likewise found excellent levels of inter-rater reliability for the overall assessment (.94), historical factors (.94), clinical factors (.86), and the final risk judgement (.81), and good levels for risk factors (.69). In a small sample of 15 detainees in Texas, Smith, Kelley, Rulesh, Sorman, and Edens (2014) reported ICC values for the historical, clinical, and risk factors of .92, .67, .68 respectively, showing good to excellent levels of inter-rater reliability.

Kötter, von Franqué, Bolzmacher, Eucker, Holzinger, and Müller-Isberner (2014) investigated the inter-rater reliability of a draft version of the HCR-20V3. Five raters were asked to rate independently the presence and relevance of the HCR-20V3 factors in 30 case vignettes. Inter-rater reliability of the overall assessment was found to be excellent (ICC = .86). However, when exploring the individual factors the ICCs ranged between .06 (H3: Relationships) and .99 (H5: Substance use and H6: Major mental disorder). The mean ICC's for the historical, clinical, and risk management factors were .65, .66, and .73 respectively. The generalizability of the results to the clinical domain can be disputed, as the assessments were completed using case summary vignettes rather than formal assessments of real clients. Strub and Douglas (2009) also

examined a draft version of the historical factors of the HCR-20V3, based on archival data from 80 psychiatric patients. ICCs were completed for the total score of the historical factors using 12 pairs of ratings randomly chosen from the dataset. Inter-rater reliability was found to be acceptable. However, the assessment information was gathered purely from file review and the authors of the tool recommend conducting an interview as well, as the single source of information could have negatively affected the data obtained as it increases bias (DoH, 2007a).

De Vogel, van den Broek, and de Vries Robbé (2014) investigated a preliminary version of the HCR-20V3 using a retrospective file study, with a sample of 25 discharged patients. Using ICCs, inter-rater reliability was found to be good for the overall assessment (.84) and the final risk judgement (.72).

The research described demonstrates the HCR-20V3 has good to excellent inter-rater reliability as classified by Fleiss (1986). However, only one study had a large sample and another investigated a draft version of the tool.

2.3.2. Validity

2.3.2.1. Face validity

Face validity refers to whether the HCR-20V3 appears (to those using it) to examine what it claims to predict (future violence). All the risk factors included in the HCR-20V3 have a literature associating them with violence, and appear relevant to existing risk assessment and management literature which suggests face validity. The inclusion of past, present, and future risk factors, and the addition of relevance

ratings further supports this. However, some of the items may significantly overlap with one another (for example, H1 and H9), so cannot be seen as proven independent predictors. Due to the absence of specific research investigating the face validity of the HCR-20V3 a conclusion can not be reached in this regard.

2.3.2.2. Concurrent validity

For the HCR-20V3 to satisfy the criteria for concurrent validity, it would need to correlate with other tools which propose to estimate risk of violence. See table 2.3 for synthesis of research conducted to date. There is currently limited research using this strategy, but most studies compare the HCR-20V3 with the HCR-20V2 which has been shown to have good concurrent validity (Douglas et al., 2014).

Strub, Douglas, and Nicholls (2014) compared the HCR-20V2 with the HCR-20V3 in a sample of 106 Canadian civil psychiatric and correctional offenders. A correlation of .91 was reported, and correlations of .89, .76, and .81 were reported for the historical, clinical, and risk factors respectively. Douglas and Belfrage (2014) have reported that six international studies found the HCR-20V3 has concurrent validity with the HCR-20V2, and Bjørkly et al. (2014) reported correlations between both versions for the historical factors (.85), clinical factors (.59), risk factors (.81), and the overall assessment (.84).

Comparing the HCR-20V3 with the HCR-20V2 to evaluate concurrent validity is a flawed process. They are similar measures and you would expect them to correlate. It would be more beneficial to investigate comparisons between the HCR-20V3 and other violence risk

assessment tools. Strub and Douglas (2009) compared a draft version the HCR-20V3 historical factors with the HCR-20V2 and the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 2006). The HCR-20V3 correlated significantly with the HCR-20V2 ($r = .60$, $p < .01$) and the VRAG ($r = .60$; $p < .01$), and it was concluded the HCR-20V3 had concurrent validity. However, only the historical factors were investigated and it is therefore unclear whether the clinical and risk factors, or indeed the overall assessment, correlate with the HCR-20V2 and VRAG.

In conclusion, there is currently a lack of research investigating the concurrent validity of the HCR-20V3. Studies comparing the HCR-20V3 to the HCR-20V2 are not sufficient as they are characteristically the same tool (Douglas et al., 2013); for true concurrent validity to be assumed the HCR-20V3 should be compared to other violence risk assessment tools. The research conducted by Strub and Douglas (2009) is more beneficial in this regard as it compares the HCR-20V3 to the VRAG, however if only investigates a draft versions of the historical factors, not the completed version of the tool which differs.

2.3.2.3. Predictive validity

The predictive validity of the HCR-20V3 relates to its ability to predict future violence. See table 2.3 for synthesis of research conducted to date. The predictive accuracy of risk assessment tools is most commonly evaluated using Receiver Operating Characteristic (ROC) analysis (Mossman, 2013). The ROC yields an Area Under the Curve (AUC) which represents the probability a randomly chosen actually

violent person will score greater than a randomly chosen nonviolent person (Swets, 1988). Table 2.2 shows the classification of effect size (Rice and Harris, 2005) and compares AUC and Cohen's d (1969).

Table 2.2

AUC and Cohen's d effect size comparisons and classifications

AUC	**Cohen's d**	**Classification**
0.556	0.200	Small
0.639	0.500	Moderate
0.714	0.800	Large

Initial research suggests the HCR-20V3 has good predictive validity. In a sample of 387 patients discharged from medium secure hospitals in England and Wales, Doyle et al. (2014) investigated the predictive validity of the HCR-20V3 for violence, defined as sexual assault, assaultative acts or threats with a weapon, and all acts of battery, at 6 and 12 month follow up periods. At the 6 and 12 month period (respectively) the AUC values were found to have a moderate to large effect for the overall assessment (.73; .70), historical factors (.63; .63), clinical factors (.75; .71), and risk factors (.67; .63). Correlations with the frequency of violence at 6 and 12 months (respectively) were also explored and found to be significant, but small effect size, for the overall assessment ($r = .23$; $p < .001$; $r = .23$; $p < .001$), historical factors ($r = .14$; $p < .01$; $r = .14$; $p < .01$), clinical factors ($r = .22$; $p < .001$; $r = .24$; $p < .001$), and risk factors ($r = .18$; $p < .001$; $r = .19$; $p < .001$).

Strub et al. (2014) investigated the HCR-20V3's predictive validity in a sample of 106 civil psychiatric patients and offenders. An average AUC value of .74 was reported, and an AUC value of .76 was reported for

the final risk judgement rating. At the 6 - 8 month follow up period 16%, 36%, and 67% of individuals rated low, moderate, and high risk respectively were violent, and at the 4-6 week follow up period 2%, 16%, and 44% of individuals rated low, moderate, and high risk respectively were violent.

Some researchers have stated that because the primary aim of a risk assessment is to predict future violence, prospective study designs are more appropriate when investigating predictive validity (Caldwell, Bogat, & Davidson, 1988). However, the usefulness of retrospective designs has also been recognised, and are commonly more tractable (Maden, 2001). De Vogel et al. (2014) completed a retrospective file study, with a sample of 86 discharged patients. Significant AUC values with violence were reported 1 year (.77), 2 years (.75), and 3 years (.67) following discharge. AUC values were also reported for the final risk judgement rating for the same time periods (.72, .67, and .64 respectively). They also reported that the HCR-20V2 and HCR-20V3 were comparable in their predictive validity. Blanchard and Douglas (2011) also conducted a retrospective study with a sample of 27 civil psychiatric patients and 16 offenders, using a draft version of the HCR-20V3. The AUC was found to be larger for physical violence (.75) in comparison to any violence (.69). The final risk judgement rating yielded the largest effect for both physical violence (.83) and any violence (.76).

In conclusion, the research to date demonstrates the HCR-20V3 is valid in terms of prediction, with moderate to large effects sizes. However, it should be noted that half the reported studies were conducted by the authors of the HCR-20V3 indicating publication.

Selection bias is also present with much of the research being conducted with mental health participants, neglecting prison samples.

2.3.2.4. Content validity

In terms of the HCR-20V3, content validity relates to whether the assessment considers all aspects of violence. There is currently no available research which discusses the content validity of the HCR-20V3. However, the HCR-20V2 was found to be more strongly related to violence than other measures (Douglas & Belfrage, 2014) and the HCR-20V3 was developed with the aim of retaining the core concepts of version 2. It could therefore be assumed version 3 also retains content validity. Nevertheless changes have been made; items have been reformed, relevance ratings and sub items have been included, and there is more emphasis on formulation and risk management. The coding of H7 (personality disorder) no longer requires the completion of a Psychopathy Checklist (PCL-R; Hare, 2003) which was previously thought to be an important factor in the consideration of future violence risk (Guy, Douglas, & Hendry, 2010).

Table 2.3

Data synthesis

Reference	Country	Setting	Sample size	Inter-rater reliability (ICC)	Concurrent validity (*r*)	Predictive validity (AUC)
Bjørkly et al. (2014)	Norway	Inpatient	20		With HCR-20V2: .85 historical factors .59 clinical factors .81 risk factors .84 total scores	
Blanchard & Douglas (2011)	Canada	Inpatients Offenders	27 16			Overall assessment: .75 physical violence .69 any violence Final risk judgement: .83 physical violence .76 any violence
De Vogel et al. (2014)	Netherlands	Community	25 86	.84 overall assessment .72 final risk judgement		 Overall assessment: .77 1 year .75 2 years .67 3 years Final risk judgement: .72 1 year .67 2 years .64 3 years
Douglas & Belfrage (2014)	Canada	Inpatient	32	.94 overall assessment .94 historical factors .86 clinical factors		

Reference	Country	Setting	Sample size	Inter-rater reliability (ICC)	Concurrent validity (*r*)	Predictive validity (AUC)
Doyle et al. (2014)	UK	Community	387	.69 risk factors .81 final risk judgement 6 months: .73 overall assessment .72 historical factors .69 clinical factors .76 risk factors 12 months: .92 overall assessment .91 historical factors .90 clinical factors .93 risk factors		6 months: .73 overall assessment .63 historical factors .75 clinical factors .67 risk factors 12 months: .70 overall assessment .63 historical factors .71 clinical factors .63 risk factors
Kötter et al. (2014)	Germany	Inpatient	30	.86 overall assessment .65 historical factors .66 clinical factors .73 risk factors		
Smith et al. (2014)	USA	Prison	15	.92 historical factors .67 clinical factors .68 risk factors		
Strub & Douglas (2009)	Canada	Inpatient	80		.60 with HCR-20V2 .60 with VRAG	
Strub et al. (2014)	Canada	Inpatient and prison	106		With HCR-20V2: .91 overall assessment .89 historical factors .76 clinical factors .81 risk factors	.74 overall assessment .76 final risk judgement

2.3.3. Appropriate norms

The HCR-20V3 is not an actuarial measure and as such appropriate norms are not suitable. However, Webster et al. (1997) provided normative data for the prevalence of risk factors in a number of different samples (for example civil, forensic psychiatric patients, and correctional offenders) for the HCR-20V2. Due to the revisions to the risk factors these would not be applicable to the HCR-20V3, and further research in this regard is required. In addition, limitations were evident in the data for version 2, including no appropriate information regarding non-criminals, and the questionable diversity of the standardised sample (primarily North American forensic populations).

2.4. Clinical Utility

The clinical utility of a tool refers to its effectiveness in clinical settings and how easy it is to apply (Smart, 2006). Flaata and Marthe (2013) completed a case study, comparing the HCR-20V2 and the HCR-20V3 assessments of a male maximum security patient. Both were completed by two independent raters and hospital staff, and the outcome of the risk assessment, the management plan, and the clinical utility was discussed. It was found that HCR-20V3 provided a better structure for the assessment process, and that the HCR-20V3 was more precise in regards to the risk management plan. The HCR-20V3 was found to be more time consuming to complete than the HCR-20V2, but this was counterbalanced by the provision of a more comprehensive violence risk assessment. A study exploring the implementation of the HCR-20V3 (de Vogel & de Vries Robbé, 2013), found on average it took an extra 27

minutes to code the HCR-20V3 than the HCR-20V2. Participants reported there were factors on the HCR-20V3 which were easier to code, but that the unfamiliarity of the new factors made it harder. Participant's ratings of usefulness reported that sub items within the risk factors (89%), indicators to aid coding (78%), relevance ratings (75%), risk scenarios (74%), risk formulation (73%), and final risk judgement ratings (67%) were useful in the HCR-20V3. In general the HCR-20V3, in comparison to the HCR-20V2, was more applicable, clear, structured, detailed, and more specific and dynamic.

Wærp (2013) completed a case study with a female forensic psychiatric patient, focusing on changes in repeated assessments of the HCR-20V2 and HCR-20V3. It was reported the sub-items of the HCR-20V3 helped to structure judgement, the presence and relevance ratings improved the development of formulations, and the specification of time frame and priority of case ratings enhanced the structure of the risk management plan. Bjørkly et al. (2014) investigated the clinical utility of the HCR-20V3 in comparison to the HCR-20V2 and found version 3 enabled the development if a more systematic and detailed violence risk assessment.

Therefore it can be concluded that the research to date shows the HCR-20V3 is beginning to prove its clinical utility. However, there is a lack of research in this area (just three studies), and two of those were case studies which are difficult to generalise (Yin, 2003).

2.5. Limitations of the HCR-20V3

In addition to the practical concerns outlined above, the HCR-20V3 has a number of limitations. There is an assumption of its worth on the evidence base of previous versions, and despite reassurances from the tool's authors that this data should be applicable to the HCR-20V3, this view is yet to be supported by substantive research. It should also be recognised that much of the current research has been conducted by one or more of the originating authors which may have resulted in bias.

Of the research described in this critique, only two studies had sample sizes of over 100 (Doyle et al., 2014, and Strub et al., 2014), and the majority of studies were conducted outside the United Kingdom (Canada, Norway, Germany, Netherlands, and Sweden), making the results of the remaining studies hard to generalise to the United Kingdom. It should also be noted that a number of the studies described in this critique were based on draft versions of the HCR-20V3, which differ from the published assessment. Therefore the results may not be applicable to the final published HCR-20V3.

As with much research investigating risk assessment, there is limited information about the use of the HCR-20V3 with women. The manual suggests its use with female populations is possible; however there is currently no published research to support this. The Female Additional Manual (FAM; de Vogel, de Vries Robbé, van Kalmthout, & Place, 2014) has been developed for use alongside the HCR-20V3, and it aims to provide more concrete guidelines for gender-sensitive risk assessment and management for women. The HCR-20V3's lack of

evidence base for use with women highlights the need for additional guidelines.

The final limitation this critique focuses on is the HCR-20V3's omission of assessed protective factors. Protective factors are considered in the HCR-20V3 in so much as they contribute to the formulation and the risk management plan; however there is no guidance in terms of how to formally identify protective factors. There is increasing evidence to suggest assessment of protective factors should be included when assessing the risk of future violent behaviour. It is suggested a risk assessment is unbalanced if both risk factors and protective factors are not taken into account (Rogers, 2000). Rogers (2000) states that focusing purely on risk factors can result in negative outcomes for forensic services, as individuals become classed as being at a higher risk of violent behaviour. This can result in longer admissions, impacting negatively on resources (Miller 2006), and professionals viewing their clients as more dangerous than they actually are (Rogers, 2000).

Ullrich and Coid (2009) observed that treatment to reduce violent reoffending should focus on reducing risk factors, and also increasing and maintaining protective factors. Building on positive factors within treatment has always been integral to clinical practice, but the idea these factors can represent protection in regards to future violent behaviour is fairly new (de Vries Robbé, de Vogel, & Stam, 2012).

Recently, the Structured Assessment of Protective Factors (SAPROF; De Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009) has been developed for use in combination with a reliable and valid violence risk assesment tool, such as the HCR-20, and it aims to balance the risk

assessment of future violent behaviour by assessing the presence or absence of protective factors. The HCR-20V3's neglect of protective factors highlights the need for such an assessment.

2.6. Conclusion

Assessing and managing the risk of violence in offenders with mental illness is essential in forensic mental health services, and the use of the HCR-20 as an assessment tool has been recommended. The aim of this critique was to provide an overview of the HCR-20V3, exploring its measurable properties, and clinical and research applications, whilst also taking into account its use in forensic mental health settings. The HCR-20 is considered the most researched and best empirically guided risk assessment of violence, and it has been widely adopted (Douglas et al., 2014). Version 3 of the instrument was introduced in 2013, and as such the evidence base for its reliability, validity, and clinical utility is still in its infancy. However, if it maintains the core principles of the HCR-20V2, as stated by its authors, it is possible that it will prove itself a similarly reliable and valid assessment. Despite its limitations the research to date is supportive of this, demonstrating high levels of internal and inter-rater reliability, and good levels of concurrent and predictive validity. Its clinical utility has also been supported.

The research examining the HCR-20V3 has a number of limitations, including author bias, sample sizes, and neglect of validation with females. It can also be argued that the omission of an assessment of protective factors within the HCR-20V3 may result in the risk assessment being unbalanced (Rogers, 2000), and the need for a

concurrent additional tool such as the SAPROF is imperative. Future research may wish to consider whether the inclusion of protective factors in risk assessment strategies improves predictive accuracy, and more specifically the impact on predictive accuracy if the SAPROF was used in conjunction with the HCR-20V3.

This critique demonstrates the importance of ensuring an evidence base when selecting a risk assessment tool. Failure to do so could result in negative consequences for the individual being assessed (e.g. longer detention) and the wider population (e.g. financial implications).

Chapter Three

The Inclusion of Protective Factors in Violence Risk Assessment of Psychiatric Inpatients: a systematic review

Abstract

This review systematically examined the research literature on the inclusion of protective factors in violence risk assessment tools. It investigated the predictive accuracy of the violence risk assessment tools recommended for use in forensic mental health services in the NHS (service specification no. C03/S/a; NHS Commissioning Board, 2013), and whether the inclusion of protective factors improved the risk assessment process. Following PRISMA guidelines, research investigating the predictive validity of the Historical Clinical Risk-20 (HCR-20V2; Webster, Douglas, Eaves, & Hart, 1997, and HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013), the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2012), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004) in a population of adult male offenders, with a primary diagnosis of mental illness or personality disorder, a history of violence, and an outcome measure of future violent behaviour or re-offending was included. A total of 11,847 participants were investigated in 44 studies, and the HCR-20 was the most widely evaluated tool. The SAPROF was found to have the highest level of predictive validity with a mean AUC of .74, followed by the HCR-20 with a mean AUC of .72. The START had the lowest level of predictive validity with a mean AUC of .70. It was difficult to establish with any certainty whether the inclusion of protective factors improved the risk assessment process, but the combined use of the HCR-20 with the SAPROF improved predictive accuracy, promoting the inclusion of protective factors.

3.1. Introduction

3.1.1. Background

This review systematically examined the research literature on the inclusion of protective factors in violence risk assessment tools. It investigated the predictive accuracy of the violence risk assessments tools which are recommended for use in forensic mental health services in the National Health Service (NHS), and whether the inclusion of protective factors improves the risk assessment process.

Assessing and managing the risk of violence in individuals with mental illness who have also offended is essential in forensic mental health services (Wilson, Desmarais, Nicholls, & Brink, 2010). Clinicians are often under pressure from expectations of the general public to predict and manage risk, but they are also required to consider the best interests of the individual being assessed, and develop positive risk management strategies (Whittington et al., 2013). In forensic mental health services this means integrating the risk of reoffending with appropriate interventions to manage and reduce this risk, otherwise known as 'risk-needs assessment' (Andrews, Bonta, & Hodge, 1990). This allows for resources such as supervision and treatment to be allocated appropriately, and for individuals to be managed in the least restrictive environment.

A number of violence risk assessment instruments have been developed. In the NHS's forensic mental health services it is recommended that the HCR-20 (HCR-20V2; Webster, Douglas, Eaves, & Hart, 1997, and HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013), the Structured Assessment of Protective Factors (SAPROF; de Vogel, de

Ruiter, Bouman, & de Vries Robbe, 2012), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, and Desmarais, 2004) are all used (service specification no. C03/S/a; NHS Commissioning Board, 2013).

The HCR-20 is a comprehensive set of professional guidelines for the assessment and management of violence risk. The measure assists professionals to evaluate a person's likelihood of future violence, and determine the most appropriate treatment and management strategies. It was first developed in 1995 (Webster, Eaves, Douglas, & Wintrup, 1995), updated in 1997 (HCR-20V2; Webster et al., 1997), and it was further updated in 2013 (HCR-20V3; Douglas et al., 2013).

The SAPROF is a structured assessment guideline developed for use in combination with a reliable and valid violence risk assesment tool, such as the HCR-20. It documents and quantifies the presence or absence of protective factors. The tool also offers guidelines for future treatment and risk management.

The START likewise assesses risk of future violence, but also evaluates other related risks such as self-neglect. In addition to appraising risk, it is designed as a tool to help structure regular clinical assessment in terms of evaluation, monitoring, and treatment planning. The START is considered a short-term risk assessment (up to three months; Webster et al., 2004), whereas the HCR-20 and SAPROF are used as long-term risk assessments (between six and 12 months; Douglas et al., 2013, and de Vogel et al., 2012).

Traditionally violence risk assessment tools have focused on the assessment of risk factors, but there is increasing evidence to suggest

protective factors should be included when assessing the risk of future violent behaviour (for example, Miller, 2006). Of the risk assessment tools suggested for use within the NHS, the SAPROF assesses the presence of protective factors, whereas the START evaluates the presence of both protective factors and risk factors. The HCR-20 follows the traditional route of assessing the presence of risk factors alone.

The predictive accuracy of risk assessment tools is most commonly evaluated using Receiver Operating Characteristic (ROC) analysis (Mossman, 2013). The ROC yields an Area Under the Curve (AUC) which represents the probability that a randomly chosen actually violent person will score greater than a randomly chosen non-violent person (Swets, 1988). Table 3.1 shows the classification of effect size (Rice & Harris, 2005), and compares AUC and Cohen's d (1969).

Table 3.1

AUC and Cohen's d effect size comparisons and classifications

AUC	Cohen's d	Classification
0.556	0.200	Small
0.639	0.500	Moderate
0.714	0.800	Large

Whether using prospective or retrospective research designs, the HCR-20 has been established as one of the best validated violence risk assessments (Singh, Fazel, Gueorguieva, & Buchanan, 2014). Douglas, Shaffer, Blanchard, Guy, Reeves, and Weir (2014) found that across forensic psychiatric samples, the predictive validity of the HCR-20 V1 and V2 AUC values ranged between .42 (Schaap, Lammers, & de Vogel, 2009) and .89 (McDermott, Quanbeck, Busse, Yastro, & Scott, 2008). Research exploring the predictive validity of the HCR-20V3 has found

AUC values ranging between .70 (Doyle et al., 2014) and .75 (de Vogel, van den Broek, & de Vries Robbé, 2014).

The predictive validity of the START has been investigated to a lesser degree and recent research has reported modest AUC values of .63 (Quinn, Miles, & Kinane, 2013), .59 (Troquete et al., 2014), and .65 (Whittington et al., 2014). Research exploring the predictive validity of the SAPROF is even more limited. De Vries Robbé, de Vogel, and de Spa (2011) found the SAPROF to have good predictive validity for absence of violent reconvictions in the short (1 year), medium (3 years), and long term (11 years), with AUC values of .85, .80, and .74 respectively. They also found that creating a new measure where violence risk is balanced by protective factors by subtracting the SAPROF total score from the HCR-20 total score (overall total score of risk and protection index; ORP index), yielded better predictive validity than using the HCR-20 on its own (de Vries Robbé et al., 2011). This serves to highlight the argument that 'risk-only evaluations are inherently inaccurate' (Rogers, 2000, p. 589).

3.1.2. Summary of existing reviews

A scoping search found no published systematic reviews investigating the inclusion of protective factors in violence risk assessment tools, although a number of systematic reviews exploring the use of violence risk assessments using the HCR-20, were found. For example, Singh, Grann, and Fazel (2011) conducted a systematic review and meta-regression comparing violence risk assessment tools. They found a median AUC value of .70 for the HCR-20. Published reviews

including investigation of the SAPROF and START were not identified. Most recently, a systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour was identified (Whittington et al., 2013), investigating which risk assessment tools have the highest level of predictive validity for a violent outcome. There were 16 studies that investigated the predictive validity of the HCR-20, and found a mean AUC value of .69 (Whittington et al., 2013).

Whittington et al. (2013) declare that for a risk assessment tool to be effective a number of elements should be present, including the recognition of protective factors. This highlights the emerging presumed importance of protective factors when completing violence risk assessments but does not identify their impact on predictive accuracy.

3.1.3. Aims and objectives

The aim of this review was to examine systematically the research literature on the inclusion of protective factors in violence risk assessment tools. It investigated the predictive accuracy of violence risk assessment tools used in forensic mental health services in the NHS; the HCR-20, the SAPROF, and the START. It compared the predictive accuracy of these tools to establish whether the inclusion of protective factors improves the risk assessment process.

3.2. Method

Evidence for the predictive accuracy of the HCR-20, the SAPROF, and the START was assessed by conducting a systematic review of

published and unpublished resources. The review followed the PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009).

3.2.1. Identification of studies

3.2.1.1. Search strategy: sources of literature

The sources of literature are outlined in table 3.2. Electronic bibliographic databases and gateways were searched in an attempt to identify all relevant publications. Grey literature and the reference lists of previous reviews and retrieved resources were also searched, and further attempts to identify studies were made by contacting experts in the field. These sources were specified as previous systematic reviews in the area of violence risk assessment found them appropriate (for example, Singh et al., 2011). The search included both published and unpublished resources, and those obtained in a foreign language were noted.

Table 3.2

Data sources

Source	Database	Search completed
Electronic bibliographic databases	EBSCOhost: Cumulative Index to Nursing and Allied Health Literature (CINAHL)	09th March 2015
	OVID: PsycINFO (1806 to March Week 1 2015)	09th March 2015
	OVID: Embase (1980 to 2015 Week 10)	09th March 2015
	OVID: MEDLINE (R) (1946 to March Week 1 2015)	09th March 2015
	ProQuest: Applied Social Sciences Index and Abstracts (ASSIA)	09th March 2015
	Web of Science (Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (A&HCI), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science and Humanities (CPCI-SSH)) (Timespan = 2000-2015)	11th March 2015
Gateways	Cochrane Library	09th March 2015
	Campbell Library	11th March 2015
Grey literature sources	Google Scholar	11th March 2015
Reference lists from identified reviews	A search of reference lists of resources which met the inclusion criteria took place	14th April 2015
Contact with experts	Michiel de Vries Robbe (Van der Hoeven Kliniek, Utrecht)	14th April 2015
	Dr Vivienne De Vogel (Van der Hoeven Kliniek, Utrecht)	14th April 2015
	Dr Kevin Douglas (Simon Fraser University, Canada)	14th April 2015
	Dr Christopher Webster (Simon Fraser University, Canada)	14th April 2015

3.2.1.2. Search strategy: search terms

The search strategy was broad and not restricted to a single type of study. The search was limited to references published from the year 2000 onwards, as the first violence risk assessment including the evaluation of protective factors was developed in 2004 (START; Webster et al., 2004). The following outlines the search terms applied to all

databases, and this was modified to meet the requirements of each database (see appendix 3.A):

(forensic mental health patient/offender)

AND

(HCR-20/SAPROF/START)

AND

(predictive validity/area under curve)

3.2.1.3. Data management

EndNote (bibliographic software) was used during the review process to manage the data.

3.2.2. Study selection (inclusion and exclusion criteria)

After identification of studies using the search strategy described, all duplicate references were identified and removed, and all irrelevant references were excluded by screening the reference titles. The inclusion criteria (see table 3.3) were applied to the remaining resources in two stages. In stage 1 two reviewers (the researcher and a research assistant with experience of completing violence risk assessment tools) applied the inclusion criteria to the titles and abstracts of all remaining references. Full-text papers of any titles and abstracts considered relevant by either reviewer were obtained, and their relevance was again assessed using inclusion criteria (stage 2). See appendix 3.B for the full inclusion and exclusion form used at both stages. References that did

not meet the criteria were excluded and their details were recorded alongside the reasons for exclusion.

Table 3.3

Inclusion criteria

Population	Adult males with a diagnosis of mental illness and/or personality disorder, and a history of violent behaviour
Exposure	HCR-20V2, HCR-20V3, START, SAPROF
Outcome	HCR-20/START/START vulnerabilities/SAPROF: Actual, attempted, threatened harm to others/self SAPROF/START strengths: Absence of actual, attempted, threatened harm to others/self

The violence risk assessment tools selected for inclusion in this review have primarily been validated for use in an adult male sample (de Vogel et al., 2012). For this reason, research based solely on a female or adolescent sample was excluded. Research conducted using samples without a diagnosis of mental illness (MI) or personality disorder (PD) were excluded because it would not reflect the population found within forensic mental health services.

To investigate the predictive validity of the specified violence risk assessment tools two outcomes were identified: i) actual, attempted, threatened harm to others/self; and ii) absence of actual, attempted, threatened harm to others or self. This reflects that the HCR-20V2, HCR-20V3, START, and the SAPROF (used with either the HCR-20V2 or HCR-20V3) assess likelihood of future violence, while the SAPROF used independently assesses absence of future violence. Actual, attempted, threatened harm to others or self was considered to be self-reported or observed aggression or violence towards others or self, violent reoffending, violent reconviction, violent recidivism, or readmission due to violence. Violence to self was included as research conducted by

Abidin et al. (2013) had found the HCR-20, SAPROF, and START to be predictive of self-harm (to varying degrees).

All quantitative study designs were included. Qualitative study designs were not included as they would not have reported a measure of predictive validity. To reduce publication bias, no initial limits were set on language, despite the researcher lacking resources to translate non-English language studies within the time constraints. Details of these studies were recorded (see figure 3.1), but they were excluded from the review, as were any references which were un-obtainable despite contacting the author. Those studies focused on the HCR-20V2 which were included in the systematic review conducted by Whittington et al. (2013) were excluded to avoid double counting.

3.2.3. Quality assessment

The quality of the remaining references was assessed using adapted checklists from the Critical Appraisal Skills Programme (CASP; CASP, 2013; see appendices 3.C and 3.D). Adaptations included specifying the violence risk assessment tools used, a clear definition of the outcome measure, and an appropriate measure of predictive validity (for example, AUC value). The quality assessment was completed by the researcher. To aid consistency of the process, 20% of the references were independently assessed by a second reviewer (a research assistant with experience of completing violence risk assessment tools). An intra-class correlation coefficient (ICC) of .98 (95% confidence interval = .92 – 1.00) was achieved between the two assessors for this data set, which falls in the excellent range (Fleiss, 1986).

3.2.4. Data extraction

Following the quality assessment, data were extracted from all included studies using a data extraction form (see appendix 3.E). Data to be extracted included sample size, risk assessment tool, outcome measure, length of follow-up, base rate, and predictive validity. Due to time constraints it was not possible to contact the authors of relevant resources if there were missing data or additional data required, therefore this information was recorded as 'unknown'. To reduce errors or inconsistences the researcher repeated the data extraction process, and cross-checked both sets of extracted data. Following completion of the data extraction tables the data were reported using the process of narrative synthesis (Boland, Cherry, & Dickson, 2014).

3.3. Results

3.3.1. Quantity of research available

The search identified 5,947 citations. Following removal of 566 duplicates, and 5,148 irrelevant references, 233 references remained for screening using the inclusion and exclusion criteria. During stage 1 screening 137 references were excluded on the basis that their title and abstract did not meet the inclusion criteria. This resulted in 96 references being retained for stage 2 screening. After applying the inclusion criteria to the full-texts of the remaining references (stage 2), 34 references were excluded; 12 did not have the specified population, the exposure was not as required in 15, and 6 were excluded due to the study design. One reference was excluded as it was a dissertation, and

the studies within were already included in the review. The full-texts of three references were excluded because they could not be obtained despite author contact, and six non-English language references were excluded. A further thirteen references were excluded due to being included in a previously completed systematic review (Whittington et al., 2013). Following contact with experts in the field, three additional references were included, and a further one reference was obtained via hand searching of reference lists of previous reviews and retrieved resources. Subsequently 44 references were included. The results of the study selection stage have been reported using a PRISMA diagram (Moher et al., 2009), see figure 3.1. Appendices 3.F, 3.G, 3.H, and 3.I outline reasons for exclusion in regards to each reference.

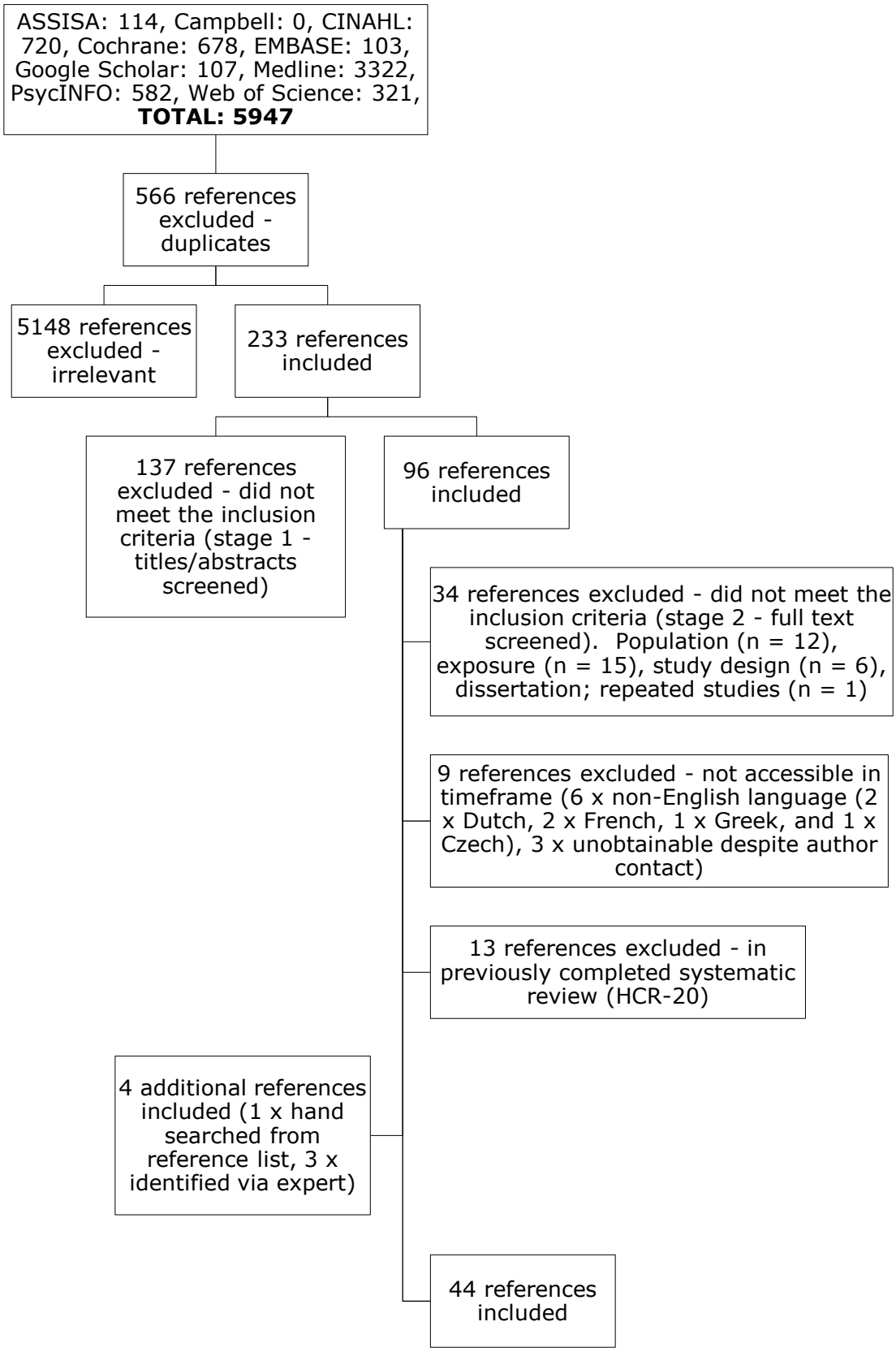


Figure 3.1. Identification of included studies

3.3.2. Quality assessment

All included studies were assessed for quality. Quality scores ranged between 28 and 47 for cohort studies (maximum score achievable was 66), with a mean of 38.5 (SD = 4.9), and between 47 and 50 for case control studies (maximum score achievable was 74) with a mean of 48.5 (SD = 2.1). Appendices 3.J and 3.K detail the results of the quality assessment process.

The majority of the studies were cohort designs (*n* = 42), and 29 of these employed a prospective study design (73%). Two studies were case control designs, and both were prospective studies. The suitability of the follow-up period was guided by the time frames for update recommended by the authors of the tools; for the HCR-20 and SAPROF follow-up period greater than 12 months was considered good, whereas for the START a follow-up period greater than 3 months was considered good. More than half of the studies had a follow-up period which was considered suitable (66%), whereas 9% were determined to have an unsuitable follow-up period (less than 6 months for the HCR-20 and SAPROF, and less than 1 month for the START). Participants lost during the follow-up period were reported in 23% of studies, though 93% of studies did not discuss the implications of lost participants. Table 3.4 outlines further details of study design and follow-up periods.

In terms of selection bias 28 studies (64%) utilised a large number of participants ($n \geq 100$), 20% used a sample recruited across more than one site, and 39% included all participants who were available. This highlights a particular area for improvement in relation to recruiting

larger samples, across different sites, and including all relevant participants.

Measurement bias was considered for the exposure construct (risk assessment tool), and the outcome (the presence or absence of actual, attempted, threatened harm to others/self). The risk assessment tool was clearly stated in all studies, and the assessor rating the risk assessment tool was blind to the outcome in 27% of studies, although it was unclear if this was the case in another 70% of studies. The risk assessment tool manual defines the level of training assessors should receive prior to completion, and this was reflected in 66% of studies. The manual also recommends the assessments should be completed using more than two sources of information (file review, interview with patient/offender, interview with those who know patient/offender). This was reflected in 16% of studies. Many used two sources (34%) or one source (32%). Consensus rating was used in 14% of studies (where the assessment is completed by more than one assessor who agree a rating). Inter-rater reliability was assessed in 50% of studies and found to be excellent in 82% of those studies.

Measurement bias in terms of the outcome was considered in relation to whether it was clearly defined, and whether the outcome measure reflected the defined outcome. This was found to be the case in 89% and 23% of studies respectively. The assessor of outcome was reported to be blind to the exposure in 20% of studies but it was unclear in 70%. Missing information was only discussed in 9% of studies. The quality of future research could be improved by following recommendations of the tool's authors in terms of completing the

assessments, ensuring the assessor is blind to the outcome and exposure (and that it is clearly reported), ensuring the defined outcome is reliable and valid, and making sure missing information is considered and dealt with appropriately.

The results were clearly reported in all studies, as was predictive validity, and the majority of studies reported outcome base rates (91%). Construct validity was considered in 36% of studies, and confounding factors considered in 50% of studies. The results were considered to fit with other available evidence in the majority of studies (82%), and the implications for practice were discussed in 73% of studies. The discussion and consideration of confounding factors would be an area of improvement for the quality of future research studies.

3.3.3. Study characteristics

Table 3.4 outlines the characteristics of each study. Some studies investigated more than one risk assessment tool, and as such the HCR-20V2 was investigated in 31 studies, the HCR-20V3 in three studies, the SAPROF in eight studies, and the START in 15 studies. Additional risk assessment tools were investigated in 18 of the studies as comparators, but the results of these were not discussed within this review. The HCR-20V2 was the most widely evaluated tool, which indicates it has been the most robustly tested. The HCR-20V3, SAPROF, and START are relatively newer tools, and as such have not been the focus of research to the same degree. Half of the studies examining the SAPROF were completed by the authors of the tool or researchers affiliated with them, as such author bias was present. Similarly 42% of research investigating the

predictive validity of the START was completed by the authors (or those affiliated with) of the tool.

Included studies were conducted in 18 different countries. The United Kingdom (England, Northern Ireland, Scotland, and Wales) accounted for the majority of studies (34%). The studies investigating the HCR-20V2 used participants from the broadest range of countries. The research focused on the SAPROF used participants primarily from the Netherlands (50%), while research investigating the START used participants primarily from Canada (47%). This is a reflection of research being completed by the authors of the assessment tool; 50% of research conducted on the SAPROF was completed by the tools authors, and 42% of START research was completed by the tools authors. This highlights author bias in the research.

For all included studies the research settings comprised inpatient (43%), community (46%), or both inpatient and community (11%) contexts. For the HCR-20V2 more than half of the studies were conducted in the community (52%); all research investigating the HCR-20V3 was completed in the community. SAPROF research was primarily conducted in the community (50%), whilst 74% of START research was using inpatient samples.

As mentioned in the quality assessment, the majority of studies were cohorts which employed a prospective approach. Overall 44 studies the follow-up period ranged from 30 days to 11 years. In terms of the HCR-20V2 and HCR-20V3, the follow-up period ranged from 6 months to 11 years, and 6 months to 36 months respectively. The follow-up period

for the SAPROF ranged from 6 months to 11 years, and from 30 days to 12 months for the START.

Table 3.4

Study characteristics

Risk tool	Reference	Additional assessment	Country	Setting	Study design	Follow-up
HCR-20V2	Abidin et al. (2013)		Ireland	Inpatient	Prospective	6 months
	Arbach-Lucioni et al. (2011)	PCL-SV	Spain	Inpatient	Prospective	12 months
	Barber-Rioja et al. (2012)	PCL-SV	USA	Community	Retrospective	12 months
	Chu et al. (2011a)		Australia	Inpatient	Retrospective	6 months
	Coid et al. (2009)	PCL-R VRAG RM2000 OGRS-II	England and Wales	Community	Prospective	*M* 1.97 years
	Coid et al. (2011)	PCL-R VRAG	England and Wales	Community	Prospective	*M* 1.97 years
	Coid et al. (2013)	VRAG OGRS-II	England and Wales	Community	Prospective	*M* 1.97 years
	Coupland (2015)	VRS	Canada	Inpatient and community	Retrospective	*M* 9.7 years
	de Borba Telles et al. (2012)	PCL-R	Brazil	Inpatient	Prospective	12 months
	de Vries Robbé & de Vogel (2011)	PCL-R	Netherlands	Inpatient	Prospective	12 months
	de Vries Robbé et al. (2011)		Netherlands	Community	Retrospective	36 months
	de Vries Robbé et al. (2013)		Netherlands	Community	Retrospective	*M* 11 years
	de Vries Robbe et al. (2014)		Netherlands	Inpatient	Prospective	12 months
	Dernevik et al. (2002)		Sweden	Inpatient	Prospective	12 months
	Desmarais et al. (2012)	PCL-SV	Canada	Inpatient	Retrospective	12 months
	Dolan and Blattner (2010)		England	Community	Retrospective	12 months
	Doyle et al. (2012)	VRAG VRS	England	Community	Prospective	5 months
	Gray et al. (2011a)		England and Wales	Community	Retrospective	24 months
	Ho et al. (2013)		China	Community	Prospective	12 months

Risk tool	Reference	Additional assessment	Country	Setting	Study design	Follow-up
	Langton et al. (2009)	VRS Static-99 RM200 PCL-R	England	Inpatient	Prospective	*M* 570 days
	McDermott et al. (2008)	PLC-R	USA	Inpatient	Prospective	*M* 2.5 years
	Michel et al. (2013)		Canada Finland Germany Sweden	Community	Prospective	24 months
	Neves et al. (2011)	PLC-R	Portugal	Community	Prospective	*M* 13 months
	O'Shea et al. (2014)		England	Inpatient	Prospective	3 months
	Pedersen et al. (2010)		Denmark	Community	Retrospective	*M* 6 years
	Pedersen et al. (2012)		Denmark	Inpatient and community	Prospective	*M* 21 months
	Snowden et al. (2010)	VRAG	England and Wales	Community	Retrospective	24 months
	Troquete et al. (2014)		Netherlands	Community	Prospective	6 months
	Viljoen (2014)		Canada	Community	Prospective	12 months
	Vojt et al. (2013)		Scotland	Inpatient and community	Prospective	*M* 31 months
	Wilson et al. (2013)*		Canada	Inpatient	Prospective	12 months
HCR-20V3	de Vogel et al. (2014)		Netherlands	Community	Retrospective	36 months
	Doyle et al. (2014)		England and Wales	Community	Prospective	12 months
	Strub et al. (2014)	HCR-20V2	Canada	Community	Prospective	6-8 months
SAPROF	Abidin et al. (2013) ^					
	Barnard-Croft (2014)		England and Wales	Community	Prospective	12 months
	Coupland (2015) ^					
	de Vries Robbé & de Vogel (2011) ^					
	de Vries Robbé et al. (2011) ^					
	de Vries Robbé et al. (2013) ^					
	de Vries Robbe et al. (2014) ^					

Risk tool	Reference	Additional assessment	Country	Setting	Study design	Follow-up
	Viljoen (2014) ^	PCL-R				
START	Abidin et al. (2013) ^					
	Braithwaite et al. (2010)		Canada	Inpatient	Prospective	30 days
	Chu et al. (2011a) ^	LSI-R SV PCL-R VRAG				
	Chu et al. (2011b)		Australia	Inpatient	Retrospective	1 month
	Desmarais et al. (2010)		Canada	Inpatient and community	Prospective	12 months
	Desmarais et al. (2012) ^					
	Gray et al. (2011b)		Wales	Inpatient	Prospective	*M* 114 days
	Nicholls et al. (2006)		Portugal	Inpatient and community	Prospective	12 months
	Nonstad et al. (2010)		Canada	Inpatient	Prospective	3 months
	Quinn et al. (2013)		England	Inpatient	Retrospective	6 months
	Troquete, et al. (2014)^					
	Viljoen (2014)					
	Whittington, et al. (2014)		England	Inpatient	Prospective	30 days
	Wilson et al. (2010)*		Canada	Inpatient	Prospective	12 months
	Wilson et al. (2013) *^					

Note. Additional risk assessment tools: LSI-R SV (Level of Service Inventory – Revised Short Version; Andrews & Bonta, 1998), OGRS-II (Offenders Group Reconviction Scale – II; Copas & Marshall, 1998; Taylor, 1999), PCL-R (Psychopathy Checklist – Revised; Hare, 2003), PCL-SV (Psychopathy Checklist – Short Version; Hart, Cox, & Hare, 1995), RM2000 (Risk Matrix 2000; Thornton, 2005; Thornton et al., 2003), Static-99; (Hanson & Thornton, 2000; Harris, Phenix, Hanson, & Thornton, 2003), VRAG (Violence Risk; Quinsey, Harris, Rice, & Cormier, 1998), and VRS (Violence Risk Scale; Wong & Gordon, 2006).
*case control study.
^ data described above under different risk assessment tool.

3.3.3.1. Participant characteristics

The participant characteristics of all included references can be found in tables 3.5 and 3.6. There were a total of 11,847 participants investigated, and all studies were treated as separate pieces of research. It appeared the same sample was used in 3 studies (Coid et al., 2009, 2011, and Coid, Ullrich, & Kallis, 2013), although different risk assessment tools were examined. The same samples were also used in Wilson, Desmarais, Nicholls, and Brink, (2010) and Wilson, Desmarais, Nicholls, Hart, and Brink (2013), again examining different risk assessment tools. The smallest sample size was 30 (Wilson et al., 2010, 2013), and the largest sample size was 1,657 (Coid et al., 2009).

All samples were at least 60% male, and 16 studies used samples which were 100% male. The mean age of the samples ranged from 31 years old to 47 years old, this information was unclear in 1 study. The percentage of participants who had a history of violence ranged from 18% to 100%, though this information was unclear in 23 studies. The percentages of participants with a diagnosis of MI ranged from 5% to 100% (unclear in 15 studies). Finally, the percentages of participants with a diagnosis of PD ranged from 0.7% to 98% (unclear in 13 studies). The main reason for information being unclear was that a history of violent behaviour, MI, and PD were often divided into different categories (for example, murder and arson, or schizophrenia and mood disorders), and the participants frequently fell into more than one category, resulting in it being impossible to separate this data into the categories required by the review.

Table 3.5

Participant characteristics

Risk tool	Reference	Sample size	% male	*M* age	% history of violence	% diagnosis of MI	% diagnosis of PD
HCR-20V2	Abidin et al. (2013)	100	94	40	Unknown	100	Unknown
	Arbach-Lucioni et al. (2011)	78	74	43	Unknown	90	10
	Barber-Rioja, et al. (2012)	131	69	37	Unknown	10	Unknown
	Chu et al. (2011a)	66	80	34	44	85	20
	Coid et al. (2009)[a]	1657	82	31	Unknown	Unknown	73
	Coid et al. (2011) [a]	1353	100	31	Unknown	44	74
	Coid et al. (2013) [a]	1396	100	31	Unknown	Unknown	Unknown
	Coupland (2015)	178	100	38	Unknown	30	76
	de Borba Telles et al. (2012)	68	100	43	Unknown	59	15
	de Vries Robbé & de Vogel (2011)	245	100	Unknown	60	Unknown	Unknown
	de Vries Robbé et al. (2011)	105	100	31	Unknown	19	83
	de Vries Robbé et al. (2013)	188	100	32	100	15	66
	de Vries Robbe et al. (2014)	185	79	41	70	53	89
	Dernevik et al. (2002)	54	89	34	100	69	Unknown
	Desmarais et al. (2012)	120	100	38	80	Unknown	Unknown
	Dolan and Blattner (2010)	72	87	36	64	73	31
	Doyle et al. (2012)	114	62	40	18	55	4
	Gray et al. (2011a)	890	100	38	Unknown	Unknown	18
	Ho et al. (2013)	220	75	44	99	Unknown	5
	Langton et al. (2009)	44	100	34	89	Unknown	Unknown
	McDermott et al. (2008)	238	86	47	86	100	98
	Michel et al. (2013)	248	96	38	57	82	22
	Neves et al. (2011)	258	87	35	73	5	2
	O'Shea et al. (2014)	505	69	40	Unknown	Unknown	Unknown
	Pedersen et al. (2010)	107	100	36	94	Unknown	18

Risk tool	Reference	Sample size	% male	*M* age	% history of violence	% diagnosis of MI	% diagnosis of PD
	Pedersen et al. (2012)	81	100	36	96	Unknown	9
	Snowden et al. (2010)	1016	85	32	Unknown	77	18
	Troquete et al. (2014)	310	94	40	Unknown	93	69
	Viljoen (2014)	102	61	47	22	100	34
	Vojt et al. (2013)	109	100	39	Unknown	93	7
	Wilson et al. (2013) [b]	30	100	37	73	87	Unknown
HCR-20V3	de Vogel et al. (2014)	86	100	32	100	21	62
	Doyle et al. (2014)	387	89	38 (median)	Unknown	Unknown	Unknown
	Strub et al. (2014)	106	69	34	Unknown	82	Unknown
SAPROF	Abidin et al. (2013) ^						
	Barnard-Croft (2014)	409	89	38	Unknown	89	6
	Coupland (2015) ^						
	de Vries Robbé & de Vogel (2011) ^						
	de Vries Robbé et al. (2011) ^						
	de Vries Robbé et al. (2013) ^						
	de Vries Robbe et al. (2014) ^						
	Viljoen (2014) ^						
START	Abidin et al. (2013) ^						
	Braithwaite et al. (2010)	34	79	44	Unknown	Unknown	21
	Chu et al. (2011a) ^						
	Chu et al. (2011b)	50	76	35	Unknown	86	20
	Desmarais et al. (2010)	119	90	39	64	85	Unknown
	Desmarais et al. (2012) ^						
	Gray et al. (2011b)	44	64	40	Unknown	Unknown	14
	Nicholls et al. (2006)	137	89	38	Unknown	100	.7
	Nonstad et al. (2010)	47	83	36	Unknown	100	15
	Quinn et al. (2013)	80	74	38	96	Unknown	13
	Troquete et al. (2014)^						

Risk tool	Reference	Sample size	% male	*M* age	% history of violence	% diagnosis of MI	% diagnosis of PD
	Viljoen (2014) ^						
	Whittington et al. (2014)	50	88	39	Unknown	Unknown	Unknown
	Wilson et al. (2010)[b]	30	100	37	73	87	Unknown
	Wilson et al. (2013) [b]^						

Note. [a] same sample used across studies.
[b] same sample used across studies.
^ data described above under different risk assessment tool.

Table 3.6

Participant characteristics dependent on the risk assessment tool studied

	HCR-20V2	HCR-20V3	SAPROF	START
n studies	31	3	8	15
Total participants	10,628	579	1,412	1,319
% male	84	86	90	85
n studies 100% male	14	0	4	3
M age range	31 – 47	32 - 38	31 - 47	34 - 47
% history of violence range	18 – 100	100	22 – 100	22 - 96
% diagnosis of MI range	5 – 100	21 – 82	19 – 100	85 - 100
% diagnosis of PD range	2 - 98	62	5 - 89	0.7 - 69

3.3.3.2. Outcome characteristics

Table 3.7 details the defined outcome, and the average base rates. The studies included in the review utilised a number of different outcomes to investigate the predictive validity of the risk assessment tools. The most common outcomes, as defined in each study, were violence (13 studies) and aggression (12 studies). Base rates were reported in all but 2 studies, and were often divided into categories dependent on follow-up periods, and sub-levels of the defined outcome. The reported base rates ranged between 11% and 73%. In terms of the HCR-20V2, the reported base rates ranged between 11% and 73%, and for the HCR-20V3 it was between 23% and 34%. The base rates ranged between 11% and 53% for the SAPROF, and between 15% and 65% for the START. The problem of comparing base rates for differing outcomes should be acknowledged. For example, studies investigating the SAPROF used absence of violence as an outcome which is likely to produce a much higher base rate than an outcome of presence of violence, increasing the possibility for true positives regardless of the accuracy of the assessment tool (Conway & Murrie, 2007).

Table 3.7

Outcome characteristics

Risk tool	Resference	Outcome	Average base rate
HCR-20V2	Abidin et al. (2013)	Adverse events	7.1 per 10,000 patient-days at risk
	Arbach-Lucioni et al. (2011)	Aggression	54%
	Barber-Rioja et al. (2012)	Re-incarceration and non-compliance	43%
	Chu et al. (2011a)	Violence	29%
	Coid et al. (2009)	Reconviction	27%
	Coid et al. (2011)	Reconviction	29%
	Coid et al. (2013)	Reconviction	25%
	Coupland (2015)	Institutional and community recidivism	53%
	De Borba Telles et al. (2012)	Violent and/or anti-social behaviour	73%
	de Vries Robbé & de Vogel (2011)	Violence	Unknown
	De Vries Robbé et al. (2011)	Recidivism	19%
	De Vries Robbé et al. (2013)	Recidivism	30%
	De Vries Robbe et al. (2014)	Aggression	11%
	Dernevik et al. (2002)	Violence	57%
	Desmarai, et al. (2012)	Aggression	55%
	Dolan and Blattner (2010)	Failure (readmission or reconviction)	56%
	Doyle et al. (2012)	Violence	25%
	Gray et al. (2011a)	Reconviction	15%
	Ho et al. (2013)	Violence	27%
	Langton et al. (2009)	Aggression	39%
	McDermott et al. (2008)	Physical aggression	25%
	Michel et al. (2013)	Any aggressive behaviour	18%
	Neves et al. (2011)	Recidivism	35%
	O'Shea et al. (2014)	Aggressive and violent incidents	61%
	Pedersen et al. (2010)	Recidivism	65%
	Pedersen et al. (2012)	Aggression and recidivism	43%

Risk tool	Resference	Outcome	Average base rate
	Snowden et al. (2010)	Reconviction	12%
	Troquete et al. (2014)	Violent and criminal behaviour, and START definition	21%
	Viljoen (2014)	Aggression	48%
	Vojt et al. (2013)	Violent incident and reconviction	25%
	Wilson et al. (2013)	Institutional violence	50%
HCR-20V3	De Vogel, et al. (2014)	Violent reconviction	Unknown
	Doyle et al. (2014)	Violence	23%
	Strub et al. (2014)	Violence	34%
SAPROF	Abidin et al. (2013)^		
	Barnard-Croft (2014)	Violence	29%
	Coupland (2015) ^		
	de Vries Robbé & de Vogel (2011) ^		
	De Vries Robbé et al. (2011) ^		
	De Vries Robbé et al. (2013) ^		
	De Vries Robbe et al. (2014) ^		
	Viljoen (2014) ^		
START	Abidin et al. (2013) ^		
	Braithwaite et al. (2010)	Challenging behaviour	48%
	Chu et al. (2011a) ^		
	Chu et al. (2011b)	Aggression	15%
	Desmarais et al. (2010)	Aggression	44%
	Desmarais et al. (2012) ^		
	Gray et al. (2011b)	Aggression	36%
	Nicholls et al. (2006)	Aggression	65%
	Nonstad et al. (2010)	Violent incidents (physical)	35%
	Quinn et al. (2013)	Aversive incidents	165 aversive incidents
	Troquete et al. (2014) ^		
	Viljoen (2014) ^		

Risk tool	Resference	Outcome	Average base rate
	Whittington, et al. (2014)	Aggression	52%
	Wilson et al. (2010)	Institutional violence	50%
	Wilson et al. (2013) ^		

Note. ^ data described above under different risk assessment tool.

3.3.4. Results data synthesis

Inter-rater reliability was primarily investigated using intra-class correlations, where the critical values are: ICC ≥ .75 = excellent, ICC between .60 and < .75 = good, and ICC between .40 and .60 = moderate (Fleiss, 1986). Table 3.8 outlines the inter-rater reliability for each study, and the mean inter-rater reliability data for each risk tool are displayed in table 3.9. Full results can be found in appendix 3.L. Inter-rater reliability ranged between .41 for the overall risk rating on the START (Viljoen, 2014), and .98 for the HCR-20V2 (Coid et al., 2009, 2011, and Coid, Ullrich, & Kallis, 2013). Overall, the HCR-20V3 had the highest inter-rater reliability with a mean of .88 (.83-.92), and the START had the lowest with a mean of .75 (.41-.91), however this is still in the excellent range.

Table 3.8

Results: Inter-rater reliability and AUC values

Risk tool	Reference	Inter-rater reliability	Follow-up	AUC
HCR-20V2	Abidin et al. (2013)		6 months	.88*
	Arbach-Lucioni et al. (2011)		1-4 months	.75*
			5-8 months	.69*
			9-12 months	.77*
	Barber-Rioja et al. (2012)	ICC = .90	12 months	.75*
	Chu et al. (2011a)		1 month	.73*
			3 months	.74*
			6 months	.61*
	Coid et al. (2009)	ICC = .98	*M* 1.97 years	.68*
	Coid et al. (2011)	ICC = .98	*M* 1.97 years	.67
	Coid et al. (2013)	ICC = .98	*M* 1.97 years	.62*
	Coupland (2015)	ICC = .94*	*M* 9.7 years	.65*
	De Borba Telles et al. (2012)		12 months	.78*
	de Vries Robbé & de Vogel (2011)		12 months	.79*
	De Vries Robbé et al. (2011)		12 months	.81
			24 months	.77
			36 months	.68
	De Vries Robbé et al. (2013)	ICC= .74	12 months	.84
			36 months	.73
			Long term	.64
	De Vries Robbe et al. (2014)		12 months	.78*
	Dernevik et al. (2002)		12 months	.71*
	Desmarais et al. (2012)		12 months	.79*
	Dolan and Blattner (2010)		12 months	.86
	Doyle et al. (2012)	ICC= .88*	5 months	.68
	Gray et al. (2011a)	r = .80	24 months	.69*
	Ho et al. (2013)	ICC = .37	6 months	.72*
			12 months	.69*
	Langton et al. (2009)		12 months	.59*
			Full period	.69*
	McDermott et al. (2008)	ICC = .97	6 months	.77*
			Full period	.64*
	Michel et al. (2013)	ICC = .73*	6 months	.67*
			12 months	.71*
			18 months	.71
			24 months	.72*
	Neves et al. (2011)		*M* 12.82 months	.82*
	O'Shea et al. (2014)		3 months	.66*
	Pedersen et al. (2010)	ICC = .90	*M* 6 years	.74*
	Pedersen et al. (2012)		*M* 21 months	.68*
	Snowden et al. (2010)	ICC = .80	24 months	.70*
	Troquete et al. (2014)		3 months	.59*
			6 months	.60*
	Viljoen (2014)	ICC = .80	6 months	.53*
			12 months	.70*
	Vojt et al. (2013)		*M* 31 months	.63*
	Wilson et al. (2013)	ICC = .88	3 months	.86
			6 months	.81
			9 months	.74
			12 months	.85*

Risk tool	Reference	Inter-rater reliability	Follow-up	AUC
HCR-20V3	De Vogel et al. (2014)	ICC = .83	12 months	.77
			24 months	.75
			36 months	.67
	Doyle et al. (2014)	ICC = .92	6 months	.73
			12 months	.70
	Strub et al. (2014)		4-6 weeks	.75*
			6-8 months	.67*
SAPROF	Abidin et al. (2013)	r = .83	6 months	.81*
	Barnard-Croft (2014)	ICC > .90	6 months	.73*
			12 months	.72*
	Coupland (2015)	ICC = .77	*M* 9.7 years	.65*
	de Vries Robbé & de Vogel (2011)		12 months	.79*
	De Vries Robbé et al. (2011)	ICC = .88	12 months	.85
			24 months	.80
			36 months	.74
	De Vries Robbé et al. (2013)	ICC = .79	12 months	.85
			36 months	.75
			Long term	.73
	De Vries Robbe et al. (2014)		12 months	.75*
	Viljoen (2014)	ICC = .75	6 months	.59*
			12 months	.61*
START	Abidin et al. (2013)	r = .77*	6 months	.70*
	Braithwaite et al. (2010)		30 days	.62*
	Chu et al. (2011a)		1 month	.76*
			3 months	.81*
			6 months	.76*
	Chu et al. (2011b)		1 month	.74*
	Desmarais et al. (2010)	ICC = .87	12 months	.74*
	Desmarais et al. (2012)	ICC = .91*	12 months	.79*
	Gray et al. (2011b)		*M* 114 days	.58*
	Nicholls et al. (2006)		12 months	.70*
	Nonstad et al. (2010)		3 months	.77*
	Quinn et al. (2013)		1 month	.63*
			3 months	.58*
			6 months	.56*
	Troquete et al. (2014)	ICC = .57*	3 months	.59*
			6 months	.62*
	Viljoen (2014)	ICC = .41*	6 months	.61*
			12 months	.61*
	Whittington et al. (2014)		30 days	.65*
			M 231 days	.72*
	Wilson et al. (2010)	ICC = .88*	3 months	.72*
			6 months	.81*
			9 months	.71*
			12 months	.77*
	Wilson et al. (2013)	ICC = .88*	3 months	.74*
			6 months	.81*
			9 months	.71*
			12 months	.80*
ORP index	Coupland (2015)		*M* 9.7 years	.65*
	de Vries Robbé & de Vogel (2011)	ICC = .80	12 months	.82*
	De Vries Robbé et al. (2011)		12 months	.85
			24 months	.81

Risk tool	Reference	Inter-rater reliability	Follow-up	AUC
			36 months	.72
	De Vries Robbé et al. (2013)		12 months	.87
			36 months	.76
			Long term	.70
	De Vries Robbe et al. (2014)		12 months	.79*
	Viljoen (2014)		6 months	.57*
			12 months	.61*

Note. * mean ICC/AUC.

The predictive validity of each risk assessment tool was investigated using area under the curve statistics (AUC; see table 3.1). Many studies investigated the predictive validity at a variety of follow-up periods, and differing outcome categories, and some studies also differentiated between MI diagnoses. Table 3.8 details the AUC data for each study. Where multiple outcome or diagnostic categories were used, a mean AUC value is reported. The full results can be found in appendix 3.L. Table 3.9 displays the mean AUC values for each risk tool.

Overall, AUC values ranged between .53 (HCR-20V2; Viljoen, 2014), and .88 (HCR-20V2; Abidin et al., 2013). The SAPROF was found to have the highest level of predictive validity, with a mean AUC value of .74 (range .59 - .85), followed by the HCR-20V2 and HCR-20V3 which both had a mean AUC value of .72 (range .53 - .88, and .76 - .77 respectively), all had a large effect size, though the START had the lowest level of predictive validity with a mean AUC value of .70 (range .56 - .81), which was a moderate effect size. For studies focused on the SAPROF 79% had AUC values producing a large effect size, followed by the HCR-20V3 and the START, with 57% of studies resulting in large effect size AUC values, and 54% of studies focused on the HCR-20V2 had large effect size AUC values. Studies focused on the START had the most

AUC values producing a small effect size (32%), and studies focused on the SAPROF had the least (2%).

In terms of follow-up periods the HCR-20V2, HCR-20V3, SAPROF, and START were all most predictive between 6-12 months (*M* AUC = .74, .73, .74, and .71 respectively).

The SAPROF's ORP index was investigated in 6 studies. Inter-rater reliability was reported in one study and was in the excellent range (.80). The level of predictive validity as reported by the AUC value ranged between .57 and .87, with a mean AUC value of .74. AUC values producing large effect sizes were reported in 70% of studies, the remainder produced 10% and 20% moderate and small effect sizes respectively. In terms of follow-up periods, predictive validity was highest between 6-12 months (*M* AUC = .75).

A meta-analysis of the data generated in this systematic review was considered. Although the homogeneity of the included sample supports the completion of a meta-analysis, the high clinical heterogeneity of the outcome measures contradicts this. The outcome measure was operationalised differently in all studies. For example, the outcome of 'violence' is different from 'any aggressive behaviour' as an outcome which may include acts such as verbal aggression. In addition, the studies investigating the SAPROF, and some of those examining the START, used absence of violence as an outcome rather than presence of violence which would have been used when investigating the HCR-20. Therefore a meta-analysis was not conducted.

Table 3.9

Results: Mean data for each risk tool

Risk tool	No. of studies	Mean quality score (range)	Mean sample size (range)	Mean inter-rater reliability (range)	Mean follow-up, months (range)	Mean AUC (range)
HCR-20V2	31	39 (28-47)	331 (30-1657)	.85 (.37-.98)	16 (1-120)	.72 (.53-.88)
HCR-20V3	3	39 (29-46)	193 (86-387)	.88 (.83-.92)	14 (1-36)	.72 (.67-.77)
SAPROF	8	42 (28-47)	189 (102-409)	.82 (.75-.88)	26 (6-120)	.74 (.59-.85)
START	15	39 (31-50)	88 (30-310)	.75 (.41-.91)	6 (1-12)	.70 (.56-.81)
ORP index	6	41 (28-47)	167 (102-245)	.80	18 (6-120)	.74 (.57-.87)

3.4. Discussion

The search strategy identified 53 references which satisfied the inclusion criteria. Of these, 13 references had been included in a previous systematic review and were subsequently excluded. Consequently 40 references obtained from the search strategy were included in the review. Contacting professionals in the field and hand searching the reference lists of included resources revealed four further references. This allowed confidence in the search strategy, and it can be concluded that all relevant research was included in this systematic review. The HCR-20 is one of the most commonly used SPJ tools, and is reportedly one of the best validated violence risk assessments (Singh et al., 2014); it was therefore no surprise that the majority of references obtained during the search focused on the HCR-20. Similarly, the SAPROF is the newest violence risk assessment, and as such resulted in the fewest research studies.

The aim of this review was to examine systematically the research literature on the inclusion of protective factors in violence risk assessment. It investigated the predictive accuracy of the violence risk assessment tools which are recommended for use in forensic mental health services in the NHS (HCR-20, SAPROF, and START), and compared the predictive accuracy of these tools to establish whether the inclusion of protective factors improves the risk assessment process.

The HCR-20, which assesses risk factors, and the SAPROF, which assesses protective factors, demonstrated good predictive validity with a large effect size, while the START, which assesses both risk and protective factors, demonstrated moderate predictive validity with a

moderate effect size in relation to prediction of future violence (according to the classifications defined by Rice & Harris, 2005). These results suggest that the violence risk assessment tools recommended for use in forensic mental health services in the NHS have predictive validity.

In line with previous systematic reviews, for example Whittington et al. (2013) and Singh et al. (2011), the HCR-20 was found to have good predictive validity. There continues to be limited research investigating the value of protective factors, with fewer research studies focused on the SAPROF and the START. Overall, the SAPROF was found to have greater predictive validity (in predicting absence of violence) than the HCR-20 (in predicting presence of violence), with a mean AUC value of .74, and 79% of studies demonstrating AUC values with a large effect size. This compares to a mean AUC of .72 for the HCR-20V2 and HCR-20V3, with 54% and 57% of studies demonstrating AUC values with a large effect size respectively. While the START was found to have lower predictive validity than the HCR-20 (mean AUC value = .70), 57% of studies demonstrated AUC values with a larger effect size which is similar to that obtained from the HCR-20. These mixed results suggest it is difficult to establish whether the inclusion of protective factors improves the risk assessment process.

The results observed support the research of de Vries Robbé et al. (2011), who found that instruments assessing violence risk which included protective factors had better predictive validity than using a negative risk-based HCR-20 alone. When the HCR-20 and SAPROF were used together to create a total score of risk and protection (the ORP index), the overall predictive validity was higher than when the HCR-20

was used alone (mean AUC value = .74). This is a small increase, and both values fall in the large effect range, but it further evidences that the inclusion of protective factors improves the predictive accuracy of violence risk assessment tools. However, thus far, it is difficult to support the argument that to assess the risk of future violence accurately both risk and protective factors need to be considered (Rogers, 2000), when the predictive accuracy of risk assessments focused purely on risk factors do not differ dramatically from those which include protective factors.

Overall, the methodological quality of the studies included was acceptable. Strengths included the use of prospective study designs in the majority of research studies. Some researchers have stated that because the primary aim of a risk assessment is to predict future violence, prospective study designs are more appropriate when investigating predictive validity (Caldwell, Bogat, & Davidson, 1988). Recommendations made by the risk assessment tools' authors in terms of follow-up periods, and assessors training requirements were followed in the majority of research studies, and inter-rater reliability was excellent.

Limitations and areas for improvement include problems in regard to sample selection, measurement bias, and author bias. The included studies highlight the need to recruit larger samples and those which are more representative of the defined population. Many studies included participants from just one psychiatric unit or excluded some participants, due to lack of capacity to consent for example. The implications of excluded and lost participants were rarely discussed. This is particularly pertinent when investigating violence risk assessment tools, as the

reason for participant loss for example, may be that they were moved due to an increase (or decrease) in their violent behaviour or relapse of their illness, and had this been included in the results of the study it may have had an impact on its outcome. Similarly, few studies discussed missing information in terms of the exposure measure or the outcome criterion. Finally, confounding factors were rarely discussed or considered in the included studies. For example, it could be argued that the reason for reduced predictive validity is because the assessment classified an individual at high risk of future violence, and as such, an increased level of management and supervision was put in place (Andrews & Bonta, 2007), subsequently reducing the opportunity for future violence.

The current systematic review has a number of strengths. The search strategy was comprehensive, and reflected all relevant research. To reduce bias in the study selection stage, two reviewers completed stage 1 of applying the inclusion criteria, and the quality assessment was also completed by two reviewers, yielding an ICC of .98 for inter-rater reliability. Publication bias was reduced by including unpublished resources such as theses.

In terms of limitations, language bias may have been present due to 17% of studies being excluded on the basis of language or due to the full text being unavailable. All studies included in the review were treated as separate studies, despite there being evidence some samples overlapped, and it is possible further studies overlapped but were not identified. This possible double counting may have resulted in an over-estimation of the number of participants included in this review.

Additionally, due to time constraints it was not possible to contact the authors of relevant resources if there were missing data, which may have had implications in terms of the studies included, particularly in relation to the defined population. The data extraction was completed by a single reviewer (the researcher) which may have resulted in errors or inconsistencies. However this was minimised by the reviewer repeating the data extraction process, and cross-checking that both sets of extracted data were the same.

Author bias should be acknowledged. Half of the research investigating the predictive accuracy of the SAPROF was completed by the authors (or those affliated with) of the tool. Similarly 42% of research investigating the predictive validity of the START was completed by the authors (or those affiliated with) of the tool.

Despite the inclusion criteria specifying an outcome of 'violent behaviour', each study operationalised violence outcome differently. For example, using official reconviction data is different from 'any aggression behaviour'. The outcome may result in an under-estimation of violent behaviour, as it does not include that which is not registered or documented. In addition, the outcomes differed for studies investigating the SAPROF (and some examining the START) as they assessed absence rather than presence of violence. This affects the base rate which would have been much higher for absence of violence, increasing the opportunity for true positives independent of the accuracy of the assessment tool (Conway & Murrie, 2007). Future studies may wish to take into consideration these limitations to improve the quality of research in this area.

The findings of this systematic review can be tentatively generalised to adult males with a diagnosis of mental illness or personality disorder, who have a history of violent behaviour, for whom this review was directed. Many of the studies included a small percentage of females, which may have skewed the results. In some studies the data regarding diagnosis was not clear, which may have had a similar effect. In terms of whether the findings can be generalised to forensic mental health services within the NHS, much of the research was conducted in Western countries, and approximately a third of the studies were completed in the United Kingdom (n = 14), supporting generalisation.

3.4.1. Implications for practice

Caution should be used when applying the findings of this review to professional practice. When completing research into the predictive validity of violence risk assessment tools, researchers tend to focus purely on the total scores achieved on the tools; however, these total scores are not used within professional practice. The risk assessment, and final risk judgement is made integrating clinical and structured professional judgement evidence (for example, Douglas et al., 2013); predictive validity based purely on the total scores may not be reflective of the level of risk identified by the clinician.

The findings support the continued use of the HCR-20, the SAPROF, and the START in forensic mental health settings within the NHS. However, it could be questioned why three violence risk assessment instruments are utilised instead of just one. The SAPROF

was found to have the greatest predictive validity in terms of absence of violence, but it cannot be used independently of another validated risk assessment tool. The START had the lowest level of predictive validity, lying below the large effect size classification, indicating it may not be suitable for use in practice, and it could be withdrawn.

Yet the START evaluates a number of different areas of risk, not just violence. The authors of the HCR-20 (Douglas et al., 2013) and SAPROF (de Vogel et al., 2012) recommend they are used as 'long term' risk assessment tools, whilst the START proposes to be a 'short term' risk assessment tool (Webster et al., 2004). Both of which are arguments for the START's continued use. The results of the systematic review found optimum predictive validity found between the 6 and 12 month follow-up period when using the SAPROF and HCR-20V3, but the START displayed better predictive validity between the 6 and 12 month follow-up period, rather than follow-up periods below 6 months. This has implications for the use of the START as a short term risk assessment tool, and future research may wish to explore further its predictive validity at shorter follow-up periods (between 1 and 3 months), rather than focusing on follow-up periods of 6 months and above.

3.4.2. Conclusions

In conclusion, this systematic review examined the research literature on the inclusion of protective factors into violence risk assessment. It investigated the predictive accuracy of the violence risk assessments tools which are recommended for use in forensic mental health services in the NHS, and found the HCR-20 and SAPROF

predictively valid (presence and absence of violence respectively), producing a large effect size (the START a moderate effect size), suggesting they have adequate predictive validity. It was difficult to establish with any certainty whether the inclusion of protective factors improves the risk assessment process, but it appeared the use of the HCR-20 and SAPROF together, creating an overall total score of risk and protection, improved predictive accuracy for future violence, supporting the inclusion of protective factors.

The predictive accuracy of violence risk assessments focused purely on risk factors do not differ dramatically from those which include protective factors. This invites questioning of the value of protective factors as inherently worthy. However, the argument remains that in practice neglect of protective factors can result in individuals with mental illness, who have offended, being classed as at a higher risk of violent behaviour than is appropriate (Rogers, 2000). In addition, and thinking about it from the point of view of the patient, treatment to reduce violent reoffending should not focus purely on reducing risk factors which can be demoralising and demotivating (Miller, 2006), but also on increasing and maintaining protective factors which serves for a more positive focus for the individual to build upon (Rogers, 2000).

Chapter Four

The Structured Assessment of Protective Factors for Violence Risk: a prospective validation study of psychiatric inpatients

Abstract

The aim this research study was to explore the value of protective factors in the assessment of violence risk. The use of the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009b) was explored to allow for a greater understanding of the effectiveness of its application in forensic inpatient settings. The validity and reliability of the SAPROF as an assessment tool was investigated across a number of domains to evaluate whether the inclusion of protective factors improves predictive accuracy of violence risk assessment tools. A prospective cohort research design and quantitative analysis were employed. The total follow-up period was six months, and the sample consisted of 108 inpatients in low and medium secure forensic services. Information from the SAPROF, HCR-20 V3, and START assessments was collected, and evidence of absence and presence of violence obtained. The SAPROF was found to have good internal reliability, concurrent, construct, and discriminative validity. The SAPROF demonstrated good predictive accuracy for absence of violence (AUC = .75), and results suggested its predictive abilities were superior to the HCR-20V3 (AUC = .68) in predicting violence, and the START (AUC = .64) in predicting presence and absence of violence, although not significantly better. Combining the use of the SAPROF with the HCR-20V3 significantly increased the predictive accuracy for presence of violence. The SAPROF was found to have incremental validity over the HCR-20V3 and the START, suggesting there is additional value in the consideration of protective factors in the assessment of violence risk.

4.1. Introduction

4.1.1. Background

In forensic mental health services violence risk assessment and management is central to effective care and treatment (Wilson, Desmarais, Nicholls, & Brink, 2010). A number of violence risk assessment tools have been developed to create a structured and consistent approach, but there is no 'gold standard' when it comes to choosing the correct instrument (NICE Guidelines, 2005). Traditionally, tools have focused on the assessment of risk factors, but there is increasing evidence to suggest protective factors should be included (for example, Miller, 2006), and individual's strengths emphasised (Department of Health; DoH, 2007a).

The National Health Service (NHS) recommend the Historical Clinical Risk - 20 (HCR-20 V2; Webster, Douglas, Eaves, & Hart, 1997; HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013), the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2012), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004) are used as violence risk assessment tools (service specification no. C03/S/a; NHS Commissioning Board, 2013). The HCR-20 follows the traditional route of assessing the presence of risk factors, whereas the SAPROF assesses the presence of protective factors, and the START evaluates the presence of both protective factors and risk factors.

This research study focused on the use of the SAPROF, specifically it's predictive abilities for absence of violence, in comparison to the HCR-20V3 and the START (which predict presence of violence). The SAPROF

is a structured assessment guideline developed for use in combination with a reliable and valid violence risk assesment tool, such as the HCR-20. It aims to balance the risk assessment of future violent behaviour by documenting and quantifying the presence or absence of protective factors such as empathy, leisure activities, and social network (de Vogel et al., 2012).

ROC analysis is the most common method used to investigate the predictive accuracy of risk assessment tools (Mossman, 2013). The ROC yields an Area Under the Curve (AUC), and in risk assessment this represents the probability a randomly chosen violent individual will score greater than a randomly chosen non-violent individual on the measure (Swets, 1988). Since the SAPROF aims to identify protective factors against violence risk, research investigating its predictive accuracy tends to use absence of violence as the outcome, which contrasts with traditional investigation of the predictive accuracy of violence risk assessment tools where presence of violence is used as the outcome.

Systematic review of the literature identifed just eight studies exploring the predictive validity of the SAPROF. Half of these studies were completed by the authors of the SAPROF or researchers affiliated with them, as such author bias is present strengthening the justification for further research.

Research was primarily conducted in community settings outside of the United Kingdom (UK), demostrating a need for further research to be conducted in NHS inpatient forensic mental health services. Only one study was completed in England and Wales, but it used a community sample. Using a prospective cohort study, Barnard-Croft (2014)

investigated the ability of the SAPROF to predict non-violent outcomes in participants discharged from a medium secure setting, comparing the instrument across differing diagnostic groups. Strong significant correlations were found between higher scores on the SAPROF and non-violent outcomes. It was able to significantly predict non-violent outcomes for particpants with a primary diagnosis of serious mental illness (AUC = .78 and .69, at 6 and 12 months respectively). However, it was unable to significantly predict non-violent outcomes for participants with a primary diagnosis of personality disorder, or a co-morbid diagnosis of personality disorder.

Other research completed in community settings found AUC values for predictive validity of absence of violence between .59 (Viljoen, 2014) and .85 (de Vries Robbé, de Vogel, & de Spa, 2011, and de Vries Robbé, de Vogel, Wever, Douglas, & Nijman, 2014).

In a Dutch inpatient setting, de Vries Robbé and de Vogel (2011) found good predictive validity for absence of violent (AUC = .77) and sexual (AUC = .81) offences. They found comparable results for the predictive validity of the HCR-20 for presence of violent (AUC = .74) and sexual (AUC = .85) offences. Also in a Dutch inpatient setting, a prospective investigation of the difference in dynamic risk and protective factors between different stages in treatment, and the predictive validity of the SAPROF and HCR-20 for aggressive incidents during clinical treatment, found total scores on the SAPROF increased as treatment progressed and HCR-20 total scores decreased. Good significant predictive validity was found for the SAPROF for absence of aggressive incidents (AUC = .76) and the HCR-20 for presence of aggressive

incidents (AUC = .77; De Vries Robbé, de Vogel, Wever, Douglas, & Nijman, 2014).

Outside of the Netherlands, Abidin, Davoren, Naughton, Gibbons, Nulty, and Kennedy (2013) compared the SAPROF with previously validated risk assessments, and prospectively tested its ability to predict absence of violence or self harm in a secure hospital setting in Ireland. They found total scores on the SAPROF had strong negative correlations with the HCR-20, and predicted absence of violence (AUC = .85) and absence of self harm (AUC = .77).

In Canada, Coupland (2015) investigated the predictive validity of the SAPROF for absence of community and institutional recidivism. Significant predictive validity for absence of conviction of violent offences pre-treatment (AUC = .64), post-treatment (AUC = .65), at the point of release (AUC = .71), and for absence of violent offences which did not result in conviction (AUC = .70, .71, and .75 respectively) were found. The SAPROF was not found to significantly predict absence of institutional recidivism.

De Vries Robbé et al. (2011) investigated the predictive validity of a new measure where risk factors were balanced by protective factors. They developed the 'Overall Total Score of Risk and Protection Index (ORP index) by subtracting the SAPROF total score from the HCR-20 total score. This measure had better predictive validity for violent reconvictions in the community than the HCR-20, and this has been supported by other researchers where the AUC values have ranged between .64 and .87 (for example, de Vries Robbé et al., 2013, and

Viljoen, 2014). This construct has not been investigated in a NHS inpatient forensic mental health service.

Four studies have investigated the additional value of using the SAPROF in addition to the HCR-20, studying the incremental validity of the tools. De Vries Robbé et al. (2013) found significantly improved predictive validity for presence of violence in the three year and long-term follow-up periods, but not the one year follow-up period. Conversely Abidin et al. (2013) found the SAPROF had a significant interactive effect with the dynamic factors on the HCR-20, but not the tool as a whole. It was concluded the SAPROF was not consistently better than the HCR-20 when predicting adverse events, but it has the advantage of covering a wider context of behaviours. Viljoen (2014) found mixed results, reporting a significant improvement, specifically for presence of verbal and sexual aggression, but no improvement in the area of predicting presence of general violence. Similarly, Coupland (2015) found the SAPROF did not add value at the pre- and post-treatment stage, however at the point of release good incremental validity was found. Again, none of these studies were completed in the UK and the majority used community settings.

The research to date suggests the SAPROF has good predictive validity as an absence of violence risk assessment tool, and results consistently report higher levels of predictive validity than the HCR-20, particularly when using the ORP index. However, there are conflicting results in terms of incremental validity, and whether using the SAPROF and the HCR-20 together improves the risk assessment process. There are currently no published data replicating the validity of the SAPROF as

applied to a UK inpatient sample. More specifically, its validity has not been investigated in a NHS inpatient forensic mental health service, where it has been compared to that of the HCR-20V3, and the incremental validity of the conjunctive use of the SAPROF and HCR-20V3 has been explored.

The authors of the SAPROF state it can be used with both male and female individuals (de Vogel et al., 2012). However, only half of published research includes female participants, and only one of those explicitly investigated gender differences in the validity of the SAPROF (Viljoen, 2014). Viljoen (2014) found the SAPROF did not significantly predict any of the outcome variables for female participants, indicating a need for further research to explore the applicability of the SAPROF as a risk assessment tool for use with women.

Similarly previous research has not explicitly investigated the application of the SAPROF as a risk assessment tool to individuals with a diagnosis of learning disability, focusing instead on mental illness and personality disorder (for example, Barnard-Croft, 2014). Presence of intelligence is considered a protective factor in the SAPROF (de Vogel et al., 2012), and some of the items may be more difficult to attain for an individual with a learning disability, for example self-control and work. In 2010/11, only 6.6% of adults with learning disabilities were reported to be in some form of paid employment (Foundation for People with Learning Disabilities). This suggests individuals with a diagnosis of learning disability may be negatively discriminated against by the SAPROF.

Finally, research supporting the applicability of the SAPROF to individuals in differing stage of their treatment, for example level of hospital security, stage of recovery, and length of admission remains in its infancy. Given the aim of inpatient admission to forensic services is to manage and reduce risk (Andrews & Bonta, 2007) it can be assumed that as individuals move through their care pathway the SAPROF assessment should identify an increased number of protective factors. This is supported by research completed by de Vries Robbé, de Vogel, and de Spa (2011) who found presence of dynamic protective factors increased following engagement in treatment.

4.1.2. Aims and objectives

The aim of this research study was to explore the value of protective factors in the assessment of violence risk. The use of the SAPROF was explored to allow for a greater understanding of the effectiveness of its application in forensic inpatient settings. The validity and reliability of the SAPROF as an assessment tool was investigated across a number of domains to evaluate whether the inclusion of protective factors improves predictive accuracy of violence risk assessment.

The objective of this research study was to examine whether the SAPROF, in comparison to the HCR-20V3 and START, is a reliable and valid assessment tool in terms of:

1. Internal reliability: Is the SAPROF consistent within itself?

2. Concurrent validity: Is the SAPROF (assessing absence of violence) related to the:
    a. HCR-20V3 (assessing presence of violence)?
    b. START (assessing presence and absence of violence)?
3. Construct validity: Are patients' total SAPROF scores representative of associated of levels of protection (absence of violence) in one month, three month, and six month follow-up periods?
4. Predictive validity: Do total SAPROF scores prospectively accurately predict absence of violence in a NHS inpatient forensic mental health service, in one month, three month, and six month follow-up periods?
5. Incremental validity. Does the use of the SAPROF with the HCR-20V3 improve predictive accuracy for presence of violence in a NHS inpatient forensic mental health service, in one month, three month, and six month follow-up periods?
6. Discriminative validity: Does the total SAPROF score discriminate between:
    a. Gender?
    b. Patients in different levels of security?
    c. Patients with differing diagnoses?
    d. Patients at different stages of their care pathway?
    e. Patients with regard to their length of stay in forensic inpatient settings?

In line with the research questions, the hypotheses are as follows:

1. The SAPROF will have good internal reliability.
2.
    a. Total SAPROF scores will be significantly negatively correlated with total HCR-20V3 scores.
    b. Total SAPROF scores will be significantly positively correlated with total START strength scores, and total SAPROF scores will be significantly negatively correlated with total START vulnerability scores.
3. There will be a significant positive correlation between higher total SAPROF scores and increased absence of violence (number of violent incident free days).
4. Total SAPROF scores will significantly predict absence of violence (number of violent incident free days).
5. Combining the use of the SAPROF with the HCR-20V3 will significantly increase the predictive accuracy of presence of violence.
6.
    a. There will be no significant difference in total SAPROF scores between genders.
    b. Total SAPROF scores for patients in low secure units will be significantly higher than those in medium secure units.
    c. Total SAPROF scores for patients with mental illness will be significantly higher than those with learning disability.

d. Total SAPROF scores for patients on pre-discharge and rehabilitation wards will be significantly higher than those on acute wards.
e. There will be a significant positive correlation between total SAPROF scores and patients with a longer length of admission.

4.2. Method

4.2.1. Study design

A prospective cohort research design using quantitative analysis was employed. The total follow-up period was six months, with one and three months durations included. Data was collected as part of a service delivery evaluation.

4.2.2. Setting

Four forensic hospitals offering local inpatient provision to service users whose offending behavior and mental health needs require they are detained in secure conditions under the Mental Health Act (MHA; 1983, 2007 amendments) were involved. Each service specialised in the assessment, treatment, and rehabilitation of adults with complex needs. The services were a medium secure unit (MSU) for adult men and women. It had 65 beds over five wards providing different levels of care (acute, sub-acute, rehabilitation, and pre-discharge). Of the 65 beds, 16 were for women. A low secure unit (LSU) for adult men, providing 20 beds, with 15 beds located on a low intensity unit (rehabilitation care), and five beds within a high dependency unit (acute care). A LSU for

adult men with a learning disability, with 20 beds over two wards which offered different levels of care (acute and rehabilitation), and a 12 bed inpatient service (pre-discharge) for male patients with a learning disability.

4.2.3. Participants

All inpatients assessed prospectively during the period 01 May 2014 to 31 May 2015 were included in the sample (*N* = 108). Patients were excluded if their length of stay post assessment was shorter than the follow-up period. For example, patients with a length of stay post-assessment of three months were excluded from the six month follow-up, but included in the one and three month follow-ups. As such, 108 participants were included in the one and three month follow-up and 100 participants were included in the six month follow-up.

4.2.4. Criterion measures

The SAPROF (de Vogel et al., 2012) evaluates the presence of 17 internal, motivation, and external protective factors as three subscales (see table 4.1). These subscales were referred to as SAPROF/I, SAPROF/M, and SAPROF/E respectively. Two of the factors are static. The factors are coded on a three point scale, based on the degree to which the protective factor is present: 'no' (0); 'perhaps' (1); and 'yes' (2). The presence of key items (a protective effect that is already present) and goal items (a protective effect which may occur after intervention) are established. The assessor gives a final judgement which reflects the extent of protection which can be 'low'; 'moderate'; or

'high'. The SAPROF gives an integrative final risk judgement which combines and weighs the risk and protective factors. Key items, goal items, the final judgement, and integrative final risk judgement were not included for the purposes of this research. The total SAPROF score was utilised for the purpose of analysis by summing the presence of the 17 factors (maximum score 34).

Table 4.1

Factors in the SAPROF

Internal factors	Motivational factors	External factors
1. Intelligence (static)	6. Work	13. Social network
2. Secure attachment in childhood (static)	7. Leisure activities	14. Intimate relationship
3. Empathy	8. Financial management	15. Professional care
4. Coping	9. Motivation for treatment	16. Living circumstances
5. Self-control	10. Attitudes towards authority	17. External control
	11. Life goals	
	12. Medication	

The HCR-20V3 (Douglas et al., 2013) is a comprehensive set of professional guidelines for the assessment and management of violence risk. It assists professionals estimate a person's likelihood of future violence, and consider the most appropriate treatment and management strategies. The HCR-20V3 allows for the evaluation of the presence (and relevance) of 20 violence risk factors in the areas of historical, clinical, and risk management over three subscales (see table 4.2). These subscales were referred to as HCR-20/H, HCR-20/C, and HCR-20/R respectively. The historical factors are static, and the clinical and risk management items are dynamic. The presence of factors is coded using a three level response format: 'no' (0); 'possibly' (1); and 'yes' (2). The risk factors relevance with respect to the development of future risk management strategies are judged, and also coded on a three level

scale: ‘low’; ‘moderate’; or ‘high’. There is the option to rate a final risk judgement following completion of the assessment (low, medium, and high). Risk factor relevance and the final risk judgement were not included for the purposes of this research. The total HCR-20 score was utilised for the purpose of analysis by summing the presence of the 20 factors (maximum score 40).

The HCR-20 has been established as one of the best validated violence risk assessments using both prospective and retrospective research designs (Singh, Fazel, Gueorguieva, & Buchanan, 2014). Research exploring the predictive validity of the HCR-20 V3 has found AUC values ranging between .70 (Doyle, Coid, Archer-Power, Dewa, Hunter-Didrichsen, Stevenson, & Shaw, 2014) and .75 (de Vogel, van den Broek, & de Vries Robbé, 2014).

Table 4.2

Factors in the HCR-20V3

Historical items Problems with…	Clinical items Problems with…	Risk management items Problems with…
H1. Violence	C1. Insight	R1. Professional services and plans
H2. Other anti-social behavior	C2. Violent ideation or intent	R2. Living situation
H3. Relationships	C3. Symptoms of major mental illness	R3. Personal support
H4. Employment	C4. Instability	R4. Treatment or supervision response
H5. Substance use	C5. Treatment or supervision response	R5. Stress or coping
H6. Major mental disorder		
H7. Personality disorder		
H8. Traumatic experiences		
H9. Violent attitudes		
H10. Treatment or supervision response		

The START (Webster, Martin, Brink, Nicholls, & Desmarais, 2004) is a further set of guidelines designed to evaluate mental disorder, monitor progress, plan treatment, and begin the process of estimating

future risk to self and others. In addition to assessing risk of violence it aims to inform decision making in terms of self-harm, suicide, unauthorised leave, substance abuse, self-neglect, and victimization. There are 20 dynamic factors (with two optional case specific items) which are considered in terms of strengths (protective factors) and vulnerabilities (risk factors; see table 4.3). This differentiation was referred to as START/S and START/V respectively. The factors are coded on a three point scale: minimally present (0); moderately present (1); or maximally present (2). Key and critical items can be selected, where a key item reflects a prominent strength, and a critical item identifies a factor which needs specific attention in treatment planning and supervision. Consideration of these factors allows for a specific risk estimate of 'low', 'moderate', or 'high' to be made for each area of risk. The key and critical items, and the specific risk estimates were not included for the purposes of this research. Total START/S and START/V scores were utilised for the purpose of analysis by summing the presence of the 20 factors (maximum score 40 respectively).

The predictive validity of the START has been investigated to a lesser degree than the HCR-20, but recent research has reported lower AUC values, for example, .63 (Quinn, Miles, & Kinane, 2013), .59 (Troquete, van den Brink, Beintema, Mulder, van Os, Schoevers, & Wiersma, 2014), and .65 (Whittington, Bjorngaard, Brown, Nathan, Noblett, & Quinn, 2014).

Table 4.3

Factors in the START

Strengths and vulnerabilities	
1. Social skills	11. Social support
2. Relationships	12. Material resources
3. Occupational	13. Attitudes
4. Recreational	14. Medication adherence
5. Self-care	15. Rule adherence
6. Mental state	16. Conduct
7. Emotional state	17. Insight
8. Substance use	18. Plans
9. Impulse control	19. Coping
10. External triggers	20. Treatability

4.2.5. Outcome measures

The outcome measure to investigate the validity of the SAPROF and START/S was absence of violence (AoV), whereas for the HCR-20V3 and START/V it was presence of violence (PoV).

4.2.5.1. AoV

Defined as the absence of '*actual, attempted, or threatened infliction of bodily harm of another person'* (Douglas et al., 2013, p. 2) or self. AoV was measured by determining the total number of violent 'incident free' days for each participant. A day was classed as 'incident free' if there were no recorded violent incidents on that day. The maximum total number of violent 'incident free' days possible post assessment was 30, 91, and 183 at the one, three, and six month follow-up periods respectively. Violence to self was included as research conducted by Abidin et al. (2013) had found the HCR-20, SAPROF, and START to be predictive of self-harm (to varying degrees). A distinction was made between externalised and internalised violence as absence of violent harm-to-others (A/HO) and harm-to-self (A/HS).

4.2.5.2. PoV.

Defined as the presence of '*actual, attempted, or threatened infliction of bodily harm of another person'* (Douglas et al., 2013, p. 2) or self. PoV was measured by determining the total number of violent 'incident' days for each participant. It was classed as an 'incident' if there were recorded violent incidents on that day. For research purposes if more than one violent incident occurred on the same day they were treated as a separate violent 'incident' day. For example, three violent incidents occurring on the same day were the equivalent of three violent 'incident' days (the unit of measure is days, not the incidents themselves). A distinction was made between externalised and internalised violence as presence of violent harm-to-others (A/HO) and harm-to-self (A/HS).

4.2.6. Procedure

All assessments were completed routinely as part of a patient's admission. The SAPROF and the HCR-20V3 were completed on a six monthly basis by qualified psychologists, who conducted an exhaustive review of background documents, discussion with individuals who knew the person being assessed, and completed a clinical interview with the person being assessed. The START was updated every three months by members of the multidisciplinary team (MDT; nursing staff, medical professionals, psychologists, occupational therapists, and social workers) who discussed and agreed ratings. All persons implementing the tools were trained in their use. This information was stored electronically in

the patient's case file, and the scores based on the ratings given were transferred into a database for analysis by the researcher. The SAPROF's ORP index was calculated by subtracting the total SAPROF score from the total HCR-20V3 score.

The follow-up periods began from the date their assessment was completed. Violent incident follow-up data was obtained from information recorded on Incident Reporting Information System (IRIS; the local procedure for recording all incidents) and documented Risk Incidents on RiO (the patient administration system). Any member of staff could complete an IRIS form, and all incidents of violence were included. Any member of staff could record a Risk Incident on RiO, and incidents documented as 'abuse', 'aggressive', 'assault actual – perpetrator', 'assault threat – perpetrator', 'self-harm actual', 'self-harm threat', 'suicide', and 'violent' were included. The researcher collated the data and calculated the total number of violent 'incident free' and 'incident' days.

The researcher transferred all data to a SPSS (statistical package for the social sciences) database for further analysis.

4.2.7. Ethical considerations

The study was approved by local governance procedures, including local NHS Trust and Service Level permissions. All data was stored securely and anonymously to ensure confidentiality.

4.2.8. Statistical methods

Analysis was conducted using SPSS 22.0 for Windows. Cronbach's alpha was used to measure internal reliability. Cronbach's alpha yields a coefficient (α) ranging between 0 (no correlation, therefore no internal consistency) and 1 (perfect correlation, therefore complete internal consistency). Usually a result of .80 and above implies an acceptable level of internal reliability (Bryman, 2012).

Pearson's correlation coefficient was used to investigate concurrent validity. Good concurrent validity would be suggested if significant relationships between the SAPROF and the HCR-20V3 and the START were found. Pearson's correlation coefficient was also used to investigate construct validity. Good construct validity would be suggested if significant relationships between the risk assessment tool and relevant outcome were found. A correlation (*r*) between .1 and .3 indicates a small effect, between .3 and .5 indicates a medium effect, and .5 or greater indicates a large effect (Cohen, 1988).

Sensitivity and specificity, and positive predictive power (PPP) and negative predictive power (NPP) were calculated to examine true positive and true negative predictive abilities of each tool. Table 4.4 outlines the definitions of these. To calculate these values, the optimal cut-off point of the ROC curve was identified (Hart, Webster, & Menzies, 1993) using Clinical Calculator 1 (VassarStats). For the SAPROF and START/S scores below the cut-off were considered to have low levels of protection (high risk), and scores above the cut-off were considered to have high levels of protection (low risk; optimal cut-off point = 17 and 16 for SAPROF and START/S respectively). For the HCR-20V3 and START/V, scores below

the cut-off were considered to have low levels of risk, and scores above the cut–off were considered to have high levels of risk (optimal cut-off point = 35 and 28 for the HCR-20V3 and START/V respectively).

Table 4.4

Definitions of sensitivity, specificity, PPP, and NPP

Test	SAPROF or START/S	HCR-20V3 or START/V
Sensitivity	Percentage of non-violent individuals correctly identified as having high levels of protection.	Percentage of violent individuals correctly identified as having high levels of risk.
Specificity	Percentage of violent individuals correctly identified as having low levels of protection.	Percentage of non-violent individuals correctly identified as having low levels of risk.
PPP	Non-violent individuals correctly identified as 'high protection'.	Violent individuals correctly identified as 'high risk'
NPP	Violent individuals correctly identified as 'low protection'.	Non-violent individuals correctly identified as 'low risk'

ROC analysis was used to investigate predictive validity. When evaluating the AUC values generated by ROC analysis, significant AUC's show the risk assessment tool significantly predicts absence or presence of violent incidents during the specified follow-up period, and the actual AUC value indicates how strong the effect was. Rice and Harris (2005) suggest AUC's above .56 should be considered small effects, AUC's above .64 as medium effects, and AUC's above .71 as large effects. Therefore AUC values of a greater value and significance indicate greater predictive ability. StAR (statistical comparison of ROC curves; Vergara, Norambuena, Ferrada, Slater, & Melo, 2008) was used to allow comparison of significance.

A multiple regression (hierarchical, enter) analysis was conducted to investigate incremental validity of the protective measures. Multiple

regression yields a correlation (R) between multiple independent variables and a dependent variable. The amount of variance the independent variables have is indicated (R^2). The associated F-ratio represents the improvement in prediction resulting from the regression. If this value is greater than 1 improvement is greater than any inaccuracy within the model, and the significance of this is calculated. A coefficient (β) demonstrates the individual contribution of each independent variable and their significance is specified.

An independent-samples t-test was conducted to determine if there were significant differences in protection levels between genders, level of security, and diagnosis. Logistic regression (binominal, enter) was used to determine if SAPROF outcome can be predicted by gender, level of security, and diagnosis. One-way Analysis of Variance (ANOVA) was used to determine if there were significant differences in protection levels between stages of care pathway. Post hoc comparisons were corrected using the Bonferroni test. Due to skewed distribution, Spearman's rank-order correlation coefficient was calculated to determine if there were significant differences between protection levels of patients in regard to their length of admission.

4.3. Results

4.3.1. Participants and descriptive data

The sample comprised 108 participants (93 men, 15 women) with a mean age of 40.21 years (SD = 13.2). The average length of admission at the time of assessment was 908.17 days (SD = 975.77), which equated to 2.69 years. The ethnicity of most participants was

White-British (80%). The typical diagnosis (as defined by information recorded in H6 of the HCR-20V3) was schizophrenia (51%). Slightly more than half of participants were detained in conditions of medium security (56%) and the remainder in low security (45%). The majority were detained under sections 37 and 41 (44%) of the Mental Health Act (1983; 2007). Most participants were in the acute stage of their care pathway (42%), followed by the rehabilitation stage, and the pre-discharge stage. The most common recorded index offence (H1 of the HCR-20V3) was assault (35.2%), followed by arson, then sexual assault. For full details of the participants please see table 4.5.

Table 4.5

Participant data

	N (%)
Gender	Male = 93 (86.1)
	Female = 15 (13.9)
Ethnicity	White-British = 86 (79.6)
	Asian or Asian British-Indian = 3 (2.8)
	Mixed-White and Black Caribbean = 2 (1.9)
	Other = 17 (15.8)
Diagnosis	Schizophrenia or Paranoid Schizophrenia = 55 (50.9)
	Schizo-Affective Disorder = 9 (8.3)
	Depression = 4 (3.7)
	Personality Disorder = 2 (1.9)
	Bi-Polar Disorder = 1 (.9)
	Learning Disability = 16 (14.8)
	Autistic Spectrum Disorder = 3 (2.8)
	Learning disability and Schizophrenia = 7 (6.5)
	Undiagnosed = 6 (5.6)
Section of Mental	S3 = 27 (25)
Health Act	S37 = 15 (13.9)
(1983; 2007)	S37/41 = 48 (44.4)
	S47/49 = 8 (7.4)
	S48/49 = 5 (4.6)
	Other = 5 (5.4)
Level of security	Low secure = 48 (44.4)
	Medium secure = 60 (55.6)
Stage of Care	Acute = 45 (41.7)
Pathway	Rehabilitation = 43 (39.8)
	Pre-Discharge = 20 (18.5)
Index Offence	Murder = 4 (3.7)
	Assault (including wounding, ABH, and GBH) = 38 (35.2)
	Arson = 12 (11.1)
	Aggravated burglary = 4 (3.7)
	Rape = 6 (5.6)
	Sexual assault = 10 (9.3)
	Other sexual offence = 6 (5.6)
	Other = 25 (25.86)
Age (years)	*M* = 40.21, SD = 13.2 (range = 19 - 89)
Length of Admission (days)	*M* = 980.93, SD = 974.8 (range = 59 – 4582)

The SAPROF and START were completed for all 108 patients. The HCR-20V3 was completed for 97 of these patients, the remaining 11 patients having had a Risk of Sexual Violence Protocol (RSVP; Hart, Kropp, Laws, Klaver, Logan, & Watt, 2003) due to their having a sexual index offence. This data was excluded from analyses. Table 4.6 outlines the descriptive data.

Table 4.6

Risk assessment descriptive data and internal reliability

	N	*M*	SD	Internal reliability
SAPROF	108	18.79	5.02	.80
SAPROF/I		4.32	2.40	.70
SAPROF/M		7.55	3.00	.75
SAPROF/E		6.93	.93	.23
ORP index	97	7.35	10.13	
HCR-20V3	97	26.24	5.99	.78
HCR-20/H		14.87	3.04	.60
HCR-20/C		6.01	2.60	.75
HCR-20/R		5.40	2.54	.70
START	108			
START/S		21.09	6.75	.91
START/V		22.90	7.89	.91

4.3.2. Internal reliability

Table 4.6 displays the results. The SAPROF total score was found to be highly reliable (α = .80). Removal of any items, except 'secure attachment in childhood' and 'external control' resulted in a lower Cronbach's alpha. Removal of 'secure attachment in childhood' and 'external control' led to a small improvement in internal reliability (α = .81 respectively), so there was no value in removing these items. The SAPROF/I and SAPROF/M total scores had good levels of reliability (α = .70 and .75 respectively) compared to SAPROF/E total score. Despite the low reliability of SAPROF/E total score removal of related items did not result in a large increase in the overall internal reliability of the SAPROF total score (α = .81). Comparison using the Feldt test (1969) found the SAPROF total score was not significantly more reliable than the HCR-20V3 total score (p = .32), but the START/S and START/V total scores were significantly more reliable than the SAPROF total score (p < .001).

4.3.3. Concurrent validity

Table 4.7 displays the results of the correlation analysis (Pearson's *r*). There was a significant negative correlation between the SAPROF and HCR-20V3 total scores (r = -.64, n = 97, p < .001), and this was a large effect. The scatter graph for this result in figure 4.1 displays the correlation. There was a significant positive correlation between the ORP and HCR-20V3 total score (r = .918, n = 97, p < .001), also a large effect.

There was a significant positive correlation between the SAPROF total score and the START/S total score (r = .52, n = 108, p < .001), a large effect, and there was a significant negative correlation between the SAPROF total score and the START/V total score (r = -.44, n = 108, p < .001), a medium effect. The ORP and START/V total score positively correlated (r = .375, n = 97, p < .001) with a medium effect, and there was a significant negative correlation with START/S total score (r = -.408, n = 97, p < .001), again a medium effect.

These results suggest the SAPROF total score has good concurrent validity, and is related to the HCR-20V3 total score and START total score. However, it is also possible that the similarity of some of the items across the instruments may have confounded these results, for example, self-control (SAPROF), instability (HCR-20V3), and impulse control (START). The relationship of the ORP to the HCR-20V3 total score is stronger than the SAPROF's total score relationship; however it is likely this is due to the contribution of the HCR-20V3's score to its development (HCR-20V3 total score minus SAPROF total score).

Table 4.7

Relationship between SAPROF and other risk assessment tools

	SAPROF	SAPROF/I	SAPROF/M	SAPROF/E	**ORP**
					(*n* = 97)
HCR-20V3 (*n* = 97)	-.64**	-.61**	-.51**	-.25*	.92**
HCR-20/H	-.36**	-.39**	-.27*	-.10	.60**
HCR-20/C	-.56**	-.53**	-.49**	-.09	.73**
HCR-20/R	-.51**	-.43**	-.40**	-.38**	.71**
START (*n* = 108)					**(*n* = 97)**
START/S	.52**	.29*	.57**	.26*	-.41**
START/V	-.44**	-.31*	-.42**	-.21*	.38**

*significant at the p < 0.01 level
**significant at the p < 0.001 level

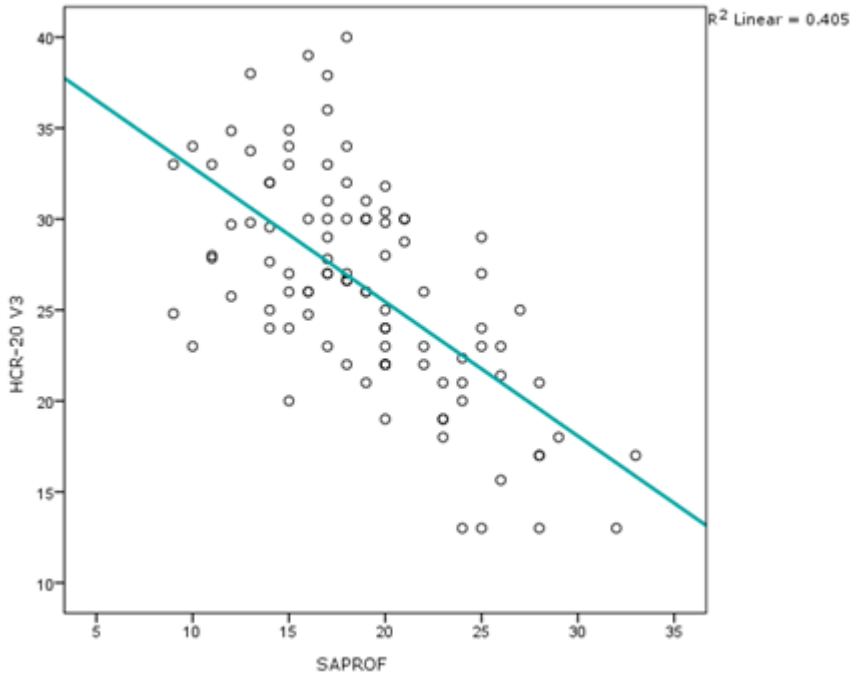


Figure 4.1. Scatter graph to show the correlation between SAPROF and HCR-20V3

4.3.4. Construct validity

Descriptive statistics for AoV and PoV within a 6 month period can be seen in table 4.8. On average, patients were AoV (absence of violence) on 99% of the days, and so PoV (presence of violence) was low frequency. At the one, three, and six month follow-up 83%, 75%, and

62% of participants were AoV respectively. At the one, three, and six month follow-up 86%, 79%, and 68% of participants were A/HO respectively (94%, 92%, and 88% of participants were A/HS respectively). Taking the mode and range into account the scope of AoV varies between patients, with some being free from violent incidents, and others having multiple violent incidents.

Table 4.8

Descriptive statistics of absence or presence of violent incidents over a 6 month period

	M	*SD*	Mode	Range
AoV	181.47	3.84	183	161 – 183
A/HO	181.96	2.68	183	167 - 183
A/HS	182.52	2.30	183	162 - 183
PoV	1.53	3.84	0	0 – 22
P/HO	1.05	2.68	0	0 – 16
P/HS	.48	2.30	0	0 – 21

Note. AoV = Absence of violence, A/HO = Absence of harm-to-others, A/HS = Absence of harm-to-self, PoV = Presence of violence, P/HO = Presence of harm-to-others, P/HS = Presence of harm-to-self.

Good construct validity is suggested if significant relationships between the risk assessment tool and AoV or PoV were found. Table 4.9 displays the results of the Pearson's correlation analysis. The SAPROF total score and ORP were significantly positively correlated with AoV and PoV respectively at all follow-ups, and significantly positively correlated with A/HO and P/HO respectively at the one and six month follow-ups. There was a significant positive correlation between the SAPROF total score and A/HS at the three and six month follow-ups. The ORP did not correlate with P/HS at any follow-up period. These were all small effects. The SAPROF/I total score was significantly correlated with all outcomes at all follow-ups (small and medium effects), except A/HS at the one month

follow-up. The SAPROF/E total score was not significantly correlated with AoV at any of the follow-ups, and SAPROF/M total score was only significantly correlated with AoV at the 6 month follow-up.

The SAPROF total score and ORP held stronger construct validity than the HCR-20V3 total score, which was only correlated with PoV at the six month follow-up. The HCR-20V3 total score correlations with P/HO were comparable to the SAPROF total score and ORP, with significant positive correlations at the one and six month follow-ups. There were no significant correlations between the HCR-20V3 total score and P/HS at any follow-up, which was weaker than the SAPROF total score but comparable to the ORP. The HCR-20/H total score and HCR-20/R total score were not significantly correlated with any of the outcomes at any follow-up. The HCR-20/C total score was significantly correlated with PoV and P/HO at all follow-ups, but not with P/HS at any follow-up. As with the SAPROF total score and ORP, all effects were small.

The SAPROF total score and ORP held stronger construct validity than the START/S total score, which was not correlated with any outcome measures at any follow-up period. There was a small significant positive correlation between the START/V total score and PoV at all follow-ups, which is comparable to the SAPROF total score and ORP. There were no significant correlations between START/V total score and P/HO at any follow-up which is weaker than the SAPROF total score and ORP. There were significant correlations between START/V total score and P/HS at the three and six months follow-up, which is comparable to the SAPROF total score but stronger than the ORP. Again all effects were small.

These results suggest the SAPROF total score (and the ORP) hold good construct validity, and overall the SAPROF total score is slightly more superior to the HCR-20V3 and the START total scores. However, it should be noted that using *r* as a guide, all effect sizes for all risk assessment tools were small, and comparing the high base rate of AoV to the low base rate of PoV may be problematic (Conroy & Murrie, 2007). In addition, the problem of multiple comparisons and the increased chance of false positives should be acknowledged (Abdi, 2007).

Table 4.9

Relationship between risk assessment tools and AoV / PoV

Follow-up	1 month				3 months				6 months			
	n	AoV or PoV	A/HO or P/HO	A/HS or P/HS	*n*	AoV or PoV	A/HO or P/HO	A/HS or P/HS	*n*	AoV or PoV	A/HO or P/HO	A/HS or P/HS
SAPROF	108	.22*	.22*	0.15	108	.20*	0.15	.17*	100	.26**	.21*	.19*
SAPROF/I	108	.26**	.25*	0.13	108	.24**	.25**	.18*	100	.33**	.30**	.20*
SAPROF/M	108	0.15	0.15	0.11	108	0.13	0.08	0.13	100	.17*	0.14	0.13
SAPROF/E	108	0.03	0.03	0.1	108	-0.03	-0.11	0.07	100	0.01	-0.08	0.11
ORP	97	.21*	.21*	0.12	97	.20*	0.67	0.15	90	.24*	.21*	0.15
HCR-20V3	97	0.16	.17*	0.06	97	0.16	0.16	0.1	90	.18*	.18*	0.09
HCR-20/H	97	0.03	0.08	-0.02	97	0.05	0.08	-0.02	90	0.08	0.13	-0.01
HCR-20/C	97	.19*	.26**	0.03	97	.22*	.24*	0.11	90	.22*	.24*	0.08
HCR-20/R	97	0.14	0.05	0.14	97	0.1	0.04	0.14	90	0.12	0.04	0.15
START/S	108	-0.01	-0.04	0.04	108	0.05	0.01	0.07	100	0.08	0.04	0.09
START/V	108	.17*	0.1	0.14	108	.20*	0.13	.19*	100	.22*	0.13	.22*

Note. AoV = Absence of violence, A/HO = Absence of harm-to-others, A/HS = Absence of harm-to-self, PoV = Presence of violence, P/HO = Presence of harm-to-others, P/HS = Presence of harm-to-self.

*significant at the $p < 0.05$ level (one-tailed)

**significant at the $p < 0.01$ level (one-tailed)

4.3.5. Predictive validity

Use of correlation analysis to investigate the predictive accuracy of violence risk assessment tools has been criticised due to its reliance on base rates (prevalence of violence) or selection ratios (proportion of individuals predicted to be violent; for example, Cohen, 1969). Base rates are particularly problematic in this research due to the comparison of a high level base rate of absence of violence (resulting in a higher likelihood of true positives independent of the assessment tool) to a low level base rate of presence of violence. It is common to use sensitivity and specificity, and positive and negative predictive power to explore predictive accuracy (Baldessarini, Finkelstein, & Arana, 1983). Table 4.10 shows how the SAPROF and ORP compare to the HCR-20V3 and the START. At the six month follow-up, the sensitivity, specificity, PPP, and NPP of the SAPROF was 82%, 61%, 78%, and 67% respectively for AoV. However, as can be seen for A/HS and P/HS, these methods have also been criticised for similar reasons as correlation analysis (Rice & Harris, 1995).

Table 4.10

Sensitivity, specificity, PPP, and NPP for the risk assessment tools

6 months	SAPROF	ORP	HCR-20V3	START/S	START/V
AoV / PoV					
Sensitivity	.82	.61	.12	.88	.48
Specificity	.61	.75	.95	.30	.82
PPP	.78	.59	.57	.68	.62
NPP	.67	.77	.65	.59	.73
A/HO / P/HO					
Sensitivity	.79	.14	.11	1.00	.46
Specificity	.61	.98	.94	0	.79
PPP	.82	.80	.43	.69	.50
NPP	.57	.72	.70	-	.77
A/HS / P/HS					
Sensitivity	1	0	.01	1	.46
Specificity	0	1	.99	.10	.79
PPP	.89	-	.50	.90	.50
NPP	-	.89	.90	1	.77

Note. AoV = Absence of violence, A/HO = Absence of harm-to-others, A/HS = Absence of harm-to-self, PoV = Presence of violence, P/HO = Presence of harm-to-others, P/HS = Presence of harm-to-self.

Rice and Harris (1995) suggest ROC analysis is the superior method for evaluating the accuracy of risk assessment tools because the AUC is independent of selection ratios and base rates. Table 4.11 displays the results of the ROC analysis for each risk assessment tool and absence or presence of violent incidents. The ROC curve for each risk assessment tool and relevant outcome at the 6 month follow-up period is shown in figures 4.2 and 4.3. AUC values ranged from .67 to .75 for the SAPROF total score predicting absence of violence, from .56 to .75 for the ORP predicting presence of violence, from .47 to .69 for the HCR-20V3 total score predicting presence of violence, and from .44 to .78 for the START total score predicting both absence and presence of violence.

At the one and six month follow-ups the SAPROF total score had significant large predictive abilities for AoV, A/HO and A/HS. Significant medium predictive abilities at the three month follow-up were found for

AoV and A/HO. Medium predictive ability at the three month follow-up for A/HS was found, but this was not significant. Predictive ability for AoV was most accurate at the 6 month follow-up, and for A/HO and A/HS it was most accurate at the one month follow-up. For AoV and A/HO the SAPROF total score was found to be significantly more accurate than the START/S total score ($p < .001$ and $p = .01$ respectively). There were no other significant differences between the predictive accuracy of the SAPROF total score and HCR-20V3 or START total scores.

The ORP had significant large predictive abilities for PoV at the three and six month follow-ups, and significant medium predictive abilities at the one month follow-up. It also had significant large predictive abilities for P/HO at the one and six month follow-ups, and significant medium predictive abilities at the three month follow-up. Predictive ability for all outcomes was most accurate at the 6 month follow-up. For PoV the ORP was found to be significantly more accurate than the HCR-20V3 total score ($p = .02$) and the START/S total score ($p = .02$). For P/HO the ORP was significantly more accurate than the START/S total score ($p = .03$). There were no other significant differences between the predictive accuracy of the ORP and HCR-20V3 or START total scores.

These results suggest the SAPROF total score and the ORP have good predictive validity, and overall are superior to the HCR-20V3 and START total scores. At the 6 month follow-up for AoV, the SAPROF total score was significantly more predictively accurate than the START/S total score, whereas for PoV, the ORP was significantly more predictively accurate than the HCR-20V3 and START/S total scores.

Table 4.11

ROC analysis for each risk assessment tool and absence or presence of violent incidents

	SAPROF		ORP		HCR-20V3		START/S		START/V	
Follow-up	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
1 month										
AoV / PoV	.74**	.63 - .85	.70**	.57 - .83	.63	.50 - .77	.48	.32 - .64	.62	.48 - .75
A/HO / P/HO	.74**	.63 - .85	.72**	.60 - .84	.67*	.52 - .80	.44	.27 - .62	.57	.42 - .72
A/HS / P/HS	.83*	.70 - .97	.56	.28 - .85	.47	.20 - .74	.65	.38 - .92	.68	.42 - .94
3 months										
AoV / PoV	.70**	.58 - .82	.71**	.59 - .82	.67*	.55 - .78	.50	.37 - .62	.64*	.52 - .76
A/HO / P/HO	.69**	.57 - .82	.70**	.58 - .81	.67*	.55 - .78	.48	.34 - .62	.61	.48 - .74
A/HS / P/HS	.67	.52 - .82	.64	.48 - .81	.59	.41 - .77	.61	.45 - .78	.73*	.58 - .87
6 months										
AoV / PoV	.75***	.65 - .85	.75***	-65 - .86	.68**	.57 - .79	.59	.47 - .70	.68**	.57 - .80
A/HO / P/HO	.72**	.61 - .82	.73***	.62 - .84	.69**	.58 - .80	.55	.43 - .67	.63*	.51 - .75
A/HS / P/HS	.71*	.58 - .85	.64	.48 - .79	.55	.37 - .72	.66	.51 - .80	.78**	.65 - .91

Note. CI = Confidence interval, AoV = Absence of violence, A/HO = Absence of harm-to-others, A/HS = Absence of harm-to-self, PoV = Presence of violence, P/HO = Presence of harm-to-others, P/HS = Presence of harm-to-self.
*significant at the $p < 0.05$ level (two-tailed)
**significant at the $p < 0.01$ level (two-tailed)
***significant at the $p< 0.001$ level (two-tailed)

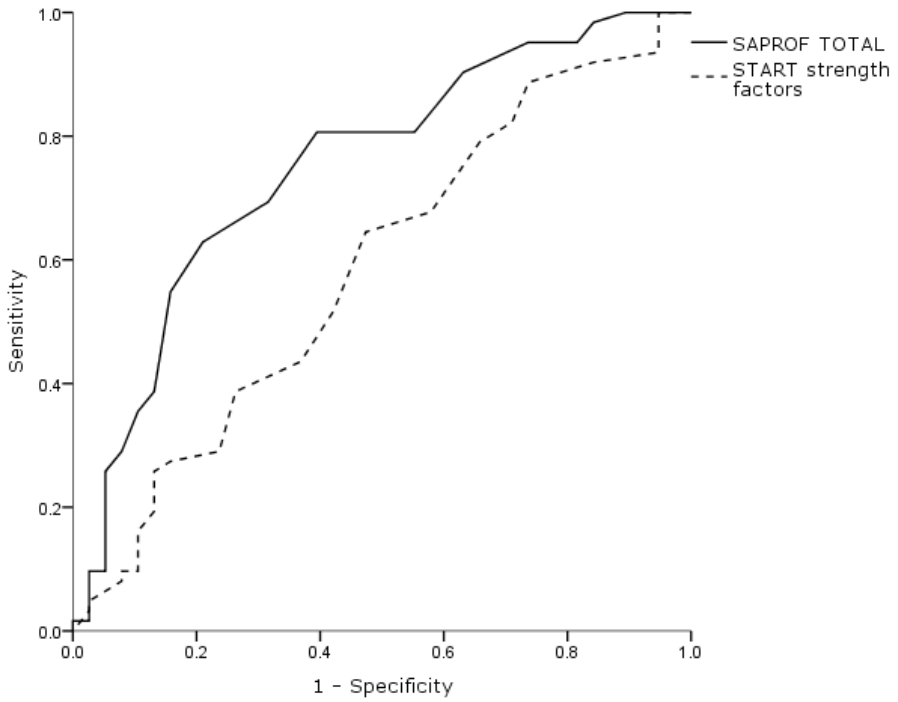


Figure 4.2. ROC curve for risk assessment tools and AoV at 6 month follow-up period

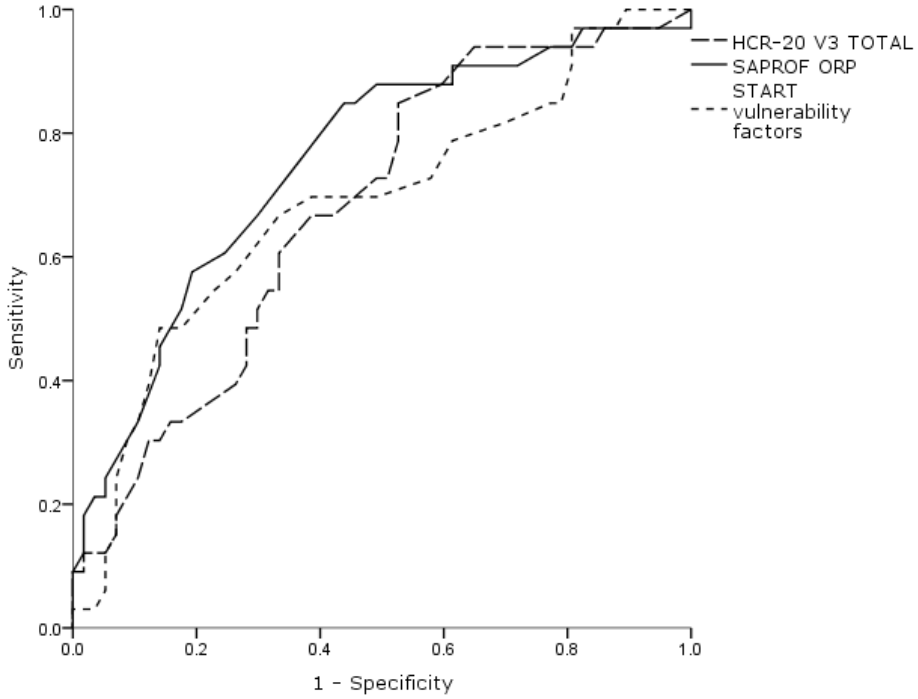


Figure 4.3. ROC curve for risk assessment tools and PoV at 6 month follow-up period

4.3.6. Incremental validity

Table 4.12 shows the results of the multiple regression analysis for the use of the SAPROF in addition to the HCR-20V3 when predicting presence of violence. In the first step of the hierarchical multiple regression the HCR-20V3 total score was entered. For six month follow-up this model was not significant (F (1, 88) = 3.06; p = .08), and explained only 3% of the variance in violent incidents. After entry of the SAPROF total score at step two the total variance explained by the model overall was 7%, and was significant (F (2, 87) = 3.21; $p < .05$). The introduction of the SAPROF total score explained an additional 4% variance in violent incidents, after controlling for the HCR-20V3 total score (*R2 change* = .04; F(1, 87) = 3.28; p = .07). In the final model the SAPROF total score was the better predictor (β = .24, p = .07), compared to the HCR-20V3 total score (β = .03, p = .82) but the effect was trend only. At the one and three month follow-ups there was no significant effect.

These results indicate that using the SAPROF total score in addition to the HCR-20V3 total score significantly increases the prediction of violence at the six month follow-up. Interestingly, neither the HCR-20V3 nor SAPROF total scores were found to be predictive when used independently. There was no impact of incremental validity at the one and three month follow-ups.

Table 4.12

Hierarchical regression of SAPROF and HCR-20V3

Follow-up						ANOVA		Coefficients				95% confidence interval	
		R	R^2	R^2 *change*	*F change*	*df*	*F*	*B*	*SE*	β	*t*	Lower	Upper
1 month													
Model 1	HCR-20V3	.16	.03	.03	2.46	95(1)	2.46						
Model 2	HCR-20V3							.00	.02	.02	.18	-.04	.05
	SAPROF	.23	.05	.03	2.68	94(2)	2.59	.05	.03	.21	1.64	-.10	.10
3 months													
Model 1	HCR-20V3	.16	.03	.03	2.60	95(1)	2.60						
Model 2	HCR-20V3							.02	.05	.05	.36	-.09	.13
	SAPROF	.22	.05	.02	1.98	94(2)	2.30	.09	.06	.18	1.41	-.04	.21
6 months													
Model 1	HCR-20V3	.18	.03	.03	3.06	88(1)	3.10						
Model 2	HCR-20V3							.02	.09	.03	.23	-.16	.20
	SAPROF	.26	.07	.04	3.28	87(2)	3.21*	.19	.10	.24	1.81	-.02	.39

*significant at the $p < 0.05$ level

4.3.7. Discriminative validity

4.3.7.1. Gender

Table 4.13 compares the mean SAPROF scores by gender using an independent subjects t-test. Female patients had higher levels of protection overall, and male patients had higher levels of risk in terms of the ORP. However, there were no significant differences in SAPROF scores between gender.

The results of the logistic regression are shown in table 4.14. A test of the full model against a constant only model was not statistically significant, indicating the SAPROF total score did not reliably distinguish between gender (χ^2 = 1.12, p = .29, df = 1). The model explained 2% (Nagelkerke's *R2*) of the variance in gender, and correctly classified 86.1% of cases (100% for male and 0% for female). The Wald criterion demonstrated the SAPROF total score was not a significant predictor of absence of violence.

To summarise, levels of protection as determined by the SAPROF total score and levels of risk as determined by the ORP do not differ between gender, and the SAPROF total score cannot be predicted by gender. These results suggest the SAPROF total score does not discriminate between genders.

Table 4.13

Independent-samples t-test of SAPROF between genders

	Male (*n* = 93)	Female (*n* = 15)				95% confidence interval	
	M (SD)	*M* (SD)	*t*	*df*	*M* difference	Lower	Upper
SAPROF	18.58 (4.97)	20.07 (5.37)	-1.06	106	-1.49	-4.26	1.28
SAPROF/I	4.27 (2.35)	4.67 (2.77)	-.59	106	-.40	-1.73	.93
SAPROF/M	7.38 (2.99)	8.60 (2.90)	-1.48	106	-1.22	-2.87	.42
SAPROF/E	6.95 (.96)	6.80 (.78)	.56	106	.146	-.37	.66
ORP	8.11 (9.82)	3.20 (11.09)	1.75	95	4.91	-.68	10.50

Table 4.14

Logistic regression for SAPROF and gender (N = 108)

		MI	LD	% correct	Predictor	B	*SE*	Wald's χ^2	*df*	*Exp(B)*	95% confidence interval Lower	Upper
Step 0	MI	93	0	100.0	Constant	-1.83	.28	43.00*	1	.16		
	LD	15	0	0								
	Overall %			86.1								
Step 1	MI	93	0	100.0	SAPROF	.06	.06	1.13	1	1.06	.95	1.18
	LD	15	0	0								
	Overall %			86.1								

*significant at the $p < 0.001$ level

4.3.7.2. Level of security

Table 4.15 outlines the mean SAPROF scores for each level of security, and the results of the t-test. Patients in the medium secure unit (MSU) reported higher levels of protection. Patients in the low secure unit (LSU) had higher levels of risk in terms of the ORP. Patients in the LSU had significantly lower scores on the SAPROF, and patients in the LSU had significantly higher scores for the ORP.

The results of the logistic regression are shown in table 4.15. A test of the full model against a constant only model was statistically significant, indicating the SAPROF total score reliably distinguished between patients in LSU and patients in MSU ($\chi^2 = 14.87$, $p < .001$, $df = 1$). The model explained 17% (Nagelkerke's *R2*) of the variance in level of security, and correctly classified 63.9% of cases (58.3% for LSU and 68.3% for MSU). The Wald criterion demonstrated the SAPROF total score was a significant predictor of absence of violence, and increasing SAPROF scores were associated with an increased likelihood of being in an MSU.

These results suggest the SAPROF total score had good discriminative validity in terms of level of security as it judged patients in conditions of medium security had a higher level of protection than patients in low security. It was able to discriminate between protection levels, and the ORP discriminated between risk levels, of patients in different levels of security. In addition SAPROF total score was predicted by level of security.

Table 4.15

Independent-samples t-test of SAPROF between different levels of security

	LSU (*n* = 48) *M* (SD)	MSU (*n* = 60) *M* (SD)	*t*	*df*	*M* difference	95% confidence interval Lower	Upper
SAPROF	16.79 (3.40)	20.38 (5.55)	-4.14**	100	-3.59	-5.31	-1.87
SAPROF/I	3.08 (1.82)	5.32 (2.35)	-5.41**	106	-2.22	-3.05	-1.41
SAPROF/M	6.83 (2.44)	8.12 (3.29)	-2.31*	105	-1.28	-2.38	-.19
SAPROF/E	6.90 (.59)	6.95 (1.14)	-.32	92	-.54	-.39	.28
	(*n* = 38)	(*n* = 59)					
ORP	12.61 (6.35)	3.97 (10.69)	4.99**	95	8.64	5.20	12.08

*significant at the p < 0.05 level
**significant at the p < 0.001 level

Table 4.16

Logistic regression SAPROF and level of security (N = 108)

		LSU	MSU	% correct	Predictor	*B*	*SE*	Wald's χ^2	*df*	*Exp(B)*	95% confidence interval Lower	Upper
Step 0	LSU	0	48	0	Constant	.22	.19	1.33	1	1.25		
	MSU	0	60	100.0								
	Overall %			55.6								
Step 1	LSU	28	20	58.3	SAPROF	.17	.05	12.16*	1	1.18	1.08	1.30
	MSU	19	41	68.3								
	Overall %			63.9								

*significant at the p < 0.001 level

4.3.7.3. Diagnosis

Table 4.17 outlines the mean SAPROF scores for each diagnosis, and the results of the t-test. Patients with mental illness (MI) had higher levels of protection. Patients with learning disability (LD) had higher levels of risk in terms of the ORP. Patients with MI had significantly higher scores on the SAPROF compared to patients with LD, and patients with MI had significantly lower ORP scores compared to patients with LD.

The results of the logistic regression are shown in table 4.18. A test of the full model against a constant only model was statistically significant, indicating the SAPROF total score reliably distinguished between patients with MI and patients with LD ($\chi^2 = 12.36$, $p < .001$, $df = 1$). The model explained 20% (Nagelkerke's *R2*) of the variance in diagnosis and correctly classified 72.2% of cases (91% for MI and 23.3% for LD). The Wald criterion demonstrated the SAPROF total score was a significant predictor of absence of violence, and increasing SAPROF scores were associated with a decreased likelihood of having a diagnosis of LD.

These results suggest the SAPROF total score showed good discriminative validity in terms of diagnosis as it judged patients with a diagnosis of MI to have higher levels of protection than patients with LD. It was able to discriminate between protection levels, and the ORP discriminated between risk levels, of patients with MI and patients with LD. In addition SAPROF total score was predicted by diagnosis.

Table 4.17

Independent-samples t-test of SAPROF between different diagnoses

	MI (*n* = 78) *M* (SD)	LD (*n* = 30) *M* (SD)	*t*	*df*	*M* difference	95% confidence interval Lower	Upper
SAPROF	19.91 (.59)	15.87 (.57)	4.96*	87	4.04	2.42	5.67
SAPROF/I	5.15 (.25)	2.17 (.21)	9.10*	97	2.99	2.34	3.61
SAPROF/M	7.81 (.36)	6.87 (.45)	1.47	106	.94	-.33	2.21
SAPROF/E	6.95 (.12)	6.87 (.12)	.41	106	.08	-.32	.48
ORP	5.05 (1.17)	15.18 (1.07)	-6.38*	72	-10.13	-13.29	-6.97

*significant at the p < 0.001 level

Table 4.18

Logistic regression for SAPROF and diagnosis (N = 108)

		MI	LD	% correct	Predictor	*B*	*SE*	Wald's χ^2	*df*	*Exp(B)*	95% confidence interval Lower	Upper
Step 0	MI	78	0	100.0	Constant	-.6	.22	19.78*	1	.39		
	LD	30	0	0								
	Overall %			72.2								
Step 1	MI	71	7	91.0	SAPROF	-.20	.06	12.36*	1	.82	.73	.91
	LD	23	7	23.3								
	Overall %			72.2								

*significant at the p < 0.001 level

4.3.7.4. Stage of care pathway

Table 4.19 displays the mean SAPROF scores for the stage of care pathway and the results of the ANOVA. Patients in the acute stage had the lowest levels of protection, and patients in the rehabilitation stage had the highest. Patients in the rehabilitation stage had the lowest levels of risk as determined by the ORP, and patients in the acute stage had the highest.

Table 4.19

One-way between subjects ANOVA of the SAPROF between differing stages of care pathway

	Acute (*n* = 45) *M* (SD)	Rehabilitation (*n* = 43) *M* (SD)	Pre-Discharge (*n* = 20) *M* (SD)	*df*	*f*
SAPROF	17.29 (5.49)	20.21 (4.62)	19.10 (3.89)	2 (105)	3.97*
SAPROF/I	4.16 (2.42)	4.81 (2.51)	3.65 (1.95)	2 (105)	1.83
SAPROF/M	6.40 (3.09)	8.42 (2.65)	8.25 (2.77)	2 (105)	6.22*
SAPROF/E	6.73 (1.16)	6.98 (.67)	7.25 (.79)	2 (105)	2.28
ORP	9.61 (10.09)	4.22 (9.58)	8.38 (10.20)	2 (94)	3.08

*significant at the $p < 0.05$ level

There was a significant effect of the stage of care pathway on the SAPROF ($F(2, 105) = 3.97$, $p < .05$). Post hoc comparisons corrected using the Bonferroni test indicated that the mean score for the acute stage SAPROF total score was significantly lower compared to the rehabilitation stage SAPROF total score (mean difference = -2.92, $p < .05$). However, the pre-discharge stage SAPROF total score did not significantly differ from the acute and rehabilitation stage SAPROF total scores.

There was a significant effect of the stage of care pathway on the SAPROF/M total score ($F(2, 105) = 6.22$, $p < .01$). Post hoc comparisons

corrected using the Bonferroni test indicated the mean score for the acute stage SAPROF/M total score was significantly lower compared to the rehabilitation stage SAPROF/M total score (mean difference = -2.02, $p < .01$). However, the pre-discharge stage SAPROF/M total score did not significantly differ from the rehabilitation and acute stage SAPROF/M's total score.

There was marginal yet significant effect of the stage of care pathway on the ORP ($F(2, 94) = 3.08$, $p = .051$). Post hoc comparisons corrected using the Bonferroni test indicated that the mean score for the acute stage ORP was significantly higher compared to the rehabilitation stage ORP. However, the pre-discharge stage ORP did not significantly differ from the acute and rehabilitation stage ORP's.

These results suggest that the SAPROF total score discriminates between protection levels, and the ORP discriminates between risk levels, of patients at differing stages. Specifically the results suggest the SAPROF total score and ORP can discriminate between acute and rehabilitation stages, although not between pre-discharge and acute/rehabilitation stages.

4.3.7.5. Length of admission

As shown in table 4.4, there was huge variation in the patient's length of admission (m = 980.93 days, SD = 974.8, range = 59 – 4582 days). The Spearman's rank-order correlation coefficient showed there was no correlation between length of admission and the SAPROF total score or ORP. The results show the SAPROF total score cannot

distinguish between patients in regard to their length of admission in forensic inpatient settings alone.

4.4. Discussion

The aim this research study was to explore the value of protective factors in the assessment of violence risk. The use of the SAPROF was examined to allow for a greater understanding of the effectiveness of its application in forensic inpatient settings. The validity and reliability of the SAPROF as an assessment tool was investigated across a number of domains to evaluate whether the inclusion of protective factors improves predictive accuracy of violence risk assessment tools.

The SAPROF total score was found to have good internal reliability supporting hypothesis one. Good internal reliability means each protective factor within the SAPROF contributes to determining levels of protection. Internal reliability of the SAPROF/E total score was very low, suggesting external factors contribute to a lesser degree; however it is likely this is due to three out of five factors being rated similarly due to the homgeniety of the sample in terms of professional care, living circumstances, and external control, rather than a true deficit of the internal reliability for this scale. Removal of these items did not result in a great improvement to the overall internal reliability. This is comparible to research conducted by Abidin et al. (2012) and Coupland (2015) who reported high levels internal reliability (α = .88 and α = .87 - .89 respectively). Therefore use of the SAPROF, with 20 items across three scales, as a reliable assessment tool is supported.

Hypothesis two was supported, as the SAPROF total score was found to have good concurrent validity. This means significant correlations with other risk assessments tools were found, and suggests the SAPROF total score was related to previously validated tools (HCR-20V3 and START). Abidin et al. (2012) also found significant correlations between the SAPROF and the START, and de Vries Robbé et al. (2014) found significant correlations between the SAPROF total score and the HCR-20V3 total score. Vijolen (2014) found similar results with significant correlations between SAPROF total scores and HCR-20 and START/V total scores, however, conversely significant correlations were not found between the SAPROF and START/S total scores. This strengthens the evidence for the use of the SAPROF as a tool which proposes to estimate levels of protection in the violence risk assessment and management process.

The SAPROF total score demonstrated good construct validity supporting hypothesis three. Significant positive correlations between the SAPROF total score and absence of violence were found, suggesting patients' SAPROF assessments are representative of their protective factors. This is supportive of previous research completed (for example, de Vries Robbé et al., 2011, and Barnard-Croft, 2015). Therefore the SAPROF measures what it claims to be measuring, and this study can endorse its use as an assessment of protective factors in the violence risk assessment and management process. The SAPROF total score was found to be superior to the HCR-20V3 and START total scores in terms of construct validity, which further endorses its use.

The SAPROF total score demonstrated good predictive validity supporting hypothesis four. This means SAPROF total scores significantly predicted the outcome, and suggests SAPROF assessments accurately predict absence of violence. The AUC value found for absence of violence at the 6 month follow-up period (.82) was comparable to previous research using the same timescale, for example, Abidin et al. (2013) and Barnard-Croft (2014) reported AUC values of .81 and .73 respectively, although greater than that found by Viljoen (2014) who reported an AUC of .59 at the 6 month follow-up period. It is possible these differences could be a result of the differing samples and settings, and also outcomes operationalised differently.

The SAPROF total score was also found to have superior predictive accuracy for absence of violence than the HCR-20V3 and START total scores had for presence of violence. These findings are supportive of the existing body of research into the predictive validity of the SAPROF for absence of violence (for example, de Vries Robbé et al., 2011), and further support its use in the violence risk assessment and management process. The SAPROF/M and SAPROF/E total scores were not representative of absence of violence at any of the follow-ups, suggesting that internal factors are more important protectors against violence than motivational or external factors. This study is also supportive of previous research which has demonstrated the SAPROF can be utilised to assess the risk of harm to self (Abidin et al., 2013). The HCR-20V3 does not appear to capture the risk of self-harm, and as such the use of the SAPROF as an additional assessment would be supported if this risk were suspected.

Combining the use of the SAPROF total score with the HCR-20V3 total score significantly increased the predictive accuracy for presence of violence, supporting hypothesis five. This means use of the SAPROF in addition the HCR-20V3 improved the accuracy of the risk assessment relative to using the HCR-20V3 as a lone assessment. These results add to the mixed findings of existing research (for example de Vries Robbé et al., 2013 and Abidin et al., 2013), but suggest the SAPROF positively influences risk assessment and can be used in addition to the HCR-20V3 to improve the process.

In terms of discriminative ability the SAPROF total score was unable to discriminate between genders, supporting hypothesis six (a). This suggests males and females achieved similar results on the SAPROF, and as such it appears suitable for use with both genders as recommended by the tools authors (de Vogel et al., 2012). However, it is possible these results were due to the limited number of females in the sample, rather than a true reflection of the SAPROF's inability to distinguish between genders. Rumgay (2004) suggests women may respond differently to protective factors. Further research is required to establish whether males and females do indeed have the same protective factors against violence, much the same as further research is needed into whether risk factors are the same.

The SAPROF total score had good discriminative validity in terms of level of security for the participant. This means it was able to discriminate between protection levels (and the ORP discriminated between risk levels) of patients in different levels of security. However hypothesis six (b) was not supported; it was expected those in the LSU

would have higher levels of protection due to the admission criteria, which would result in those referred to MSU's having higher levels of risk than those referred to LSU's (Andrew & Bonta, 2007). It is possible this result could be due to the support and supervision offered in conditions of medium security which inherently result in higher levels of protection (service specification no. C03/S/a; NHS Commissioning Board, 2013).

The SAPROF total score showed good discriminative validity in terms of diagnosis and supported hypothesis six (c), as it was able to discriminate between protection levels of patients with MI and patients with LD, and found that patients with MI have higher levels of protection. As a result, it is possible patients with a diagnosis of LD need a higher level of support to enable development of more protective factors. It is possible that due to their learning disability, criterion for protection such as intelligence, self-control, and financial management, can never really be met, and therefore may generally have lower levels of protection, highlighting the need for external measures to manage risk or a specialist instrument for this population. Further investigation is required in this area as applicability of the SAPROF as an assessment tool of protectives in individuals with a learning disability has not been conducted previously.

The SAPROF total score showed good discriminative validity in terms of care pathway stages. This means it was able to discriminate between protection levels of patients in the acute and rehabilitation stages of their care pathway. However, it was unable to discriminate between the pre-discharge and rehabilitation or acute stages. Hypothesis six (d) was therefore only partly supported. Surprisingly, patients in the

pre-discharge stage had lower levels of protection than those in the rehabilitation stage. This may be because as patients are approaching discharge thoughts turn towards how they will present in the community, and as such, levels of protection may fall because the support of professionals in all areas reduces. Three out of five factors on SAPROF/E are rated positively present due to detention in hospital and will automatically reduce following discharge.

The SAPROF total score did not show discriminative validity in terms of length of stay. This means the SAPROF was unable to distinguish between patients in regard to their length of admission in forensic inpatient settings. As such hypothesis number six (e) was rejected. As with stage of care pathway it is possible that the protection offered by detention in hospital over-rules protection in other areas.

The discriminative validity results in terms of care pathway and length of admission are in contrast with research conducted by de Vries Robbé et al. (2014), who found SAPROF total scores increased through their care pathway. It is possible differing support systems in the Netherlands where this research was complete may explain the differences in the results.

4.4.1. Study limitations

Although low and medium secure settings were included, these results found may not generalise to other settings. The outcomes may not be applicable to community settings as factors leading to inpatient violence may be different from those resulting in community violence. The sample was relatively homogenous, only had a small percentage of

females (14%), and even fewer diagnosed with a personality disorder (2%), and the results may not be generalizable to these populations. In addition, there were very few ethnic minorities which may not be reflective of the number in forensic mental health services in general.

The study size and follow-up length may have been insufficient; a longer follow-up may have generated a higher base rates. However, in practice it is recommended risk assessments are updated on a regular basis due to the dynamic nature of risk (Douglas et al., 2013), and therefore the 6 month follow-up is reflective of this. Replication in other settings and populations with a variety of follow-up periods may be useful.

Due to the study design, the analysis of inter-rater reliability was not possible, and the researcher was not blind to the assessment information when collecting the follow-up data. Both these factors may have impacted on the results of the research, and future investigations may wish to conduct such work blind to clinical information.

The use of total scores on each of the assessment tools and calculated optimal cut-off scores is a noteworthy limitation of this research. In practice the final judgement ratings of low, medium, and high are utilised by professionals to guide their assessment of level of protection (and risk; Douglas et al., 2013). However, it was not possible to investigate this construct as the setting used for research did not consistently utilise the final judgement ratings. Future research should consider using the final judgement ratings in addition to the total scores.

A further limitation may be the method of gathering outcome data which may have been influenced by bias. It was local policy for all

violent incidents to be reported via IRIS forms and risk incidents recorded on RiO, and we can be reasonably confident that these outcomes are representative of all incidents of violence. However, it is a subjective decision made by professionals as to the threshold criteria which results in an incident being reported. Some incidents may not have been captured by this measure, and so the index behaviour may be an under-representation of the actual phenomenon. In addition, as patients move through their care pathway they are likely to be supervised to a lesser degree, and incidents may go more un-noticed. Conversely, if patients are considered to have low levels of protection and high levels of risk, they are likely to be more closely supervised, with enhanced management plans, which are likely to result in fewer opportunities (and lower motivation) to engage in violent behaviour. This may have had an impact on the base rates of absence and presence of violence reported.

In terms of base rates, due to the use of absence of violence (SAPROF and START) and presence of violence as outcomes (HCR-20V3 and START), the base rates for each outcome were extremely different (99% versus 1% respectively) which may be problematic. Investigation of high base rates is likely to produce higher incidences of the investigated construct (absence of violence) being identified independent of the usefulness of the tool (Conroy & Murrie, 2007).

Finally, self-harm was included in the outcome as a form of violence as research conducted by Abidin et al. (2013) had found the HCR-20, SAPROF, and START to be predictive of self-harm (to varying degrees). However, it is not generally considered violent behaviour and

is not included in the definition of violence in the HCR-20V3 (Douglas et al., 2013).

4.4.2. Implications for practice

In practice, professionals want to know whether the individual they are assessing is at risk of future violence, not whether a group of individuals are at risk, and there is criticism of the use of AUC analyses to investigate the use of violence risk assessment tools on this basis (Cooke & Michie, 2013). There are suggestions that AUC values can be subject to over-optimistic interpretations (Sjöstedt & Grann, 2002), and Szmukler, Everitt, and Leese (2012) state 'even a highly statistically significant AUC is of limited value in clinical practice' (p. 895). In reality the 'scores' on violence risk assessment tools are not employed, nor used to predict future violence. The instruments are used to guide formulation and management plans. The resultant AUC values are irrelevant to practice (Hart & Cooke, 2013). Despite this, research such as the current study is important to add to empirical evidence and recommendations for best clinical practice (Cooke & Michie, 2013). This is one of the first piece of research investigating the reliability and validity of the SAPROF in an English inpatient sample, and enhances the evidence base for its application in this area.

The SAPROF was found to predict the absence of violence in a patients admission, and as such can be used as an evaluation of protective factors, alongside risk factors, in the process of violence risk assessment. This research provides corroborative evidence for the continued use of the SAPROF in forensic mental health settings. It

supports the recommendation that use of the SAPROF enhances the risk assessment process, and improves the validity of a risk factor focused assessment such as the HCR-20V3 (de Vogel et al., 2012). The use of the SAPROF in addition to the HCR-20V3 is supportive of the strengths based approach recommended by the DoH (2007a).

Due to the SAPROF's higher level of predictive accuracy (for absence of violence) than the HCR-20V3 and START, it could be argued the SAPROF could be used instead of these tools. However it was not designed for this purpose, it was developed as a complimentary assessment to be used alongside the HCR-20V3 to allow for a more balanced assessment, and it could be argued that disregard of risk factors would tip the balance in the other direction, and may result in a less comprehensive assessment. Although this is not reflected in the results, which support assessment of only protective factors to evaluate violence risk.

The SAPROF's ORP index also achieved good concurrent, construct, predictive, and discriminative validity. It was found to be significantly more predictive for presence of violence than the HCR-20V3, and is supportive of previous studies using the method (for example, De Vries Robbé et al., 2014). This advocates the use of a measure where violence risk is balanced by protective factors (de Vries Robbé et al., 2011); this is a research concept, and it is difficult to comprehend how this can be applied in practice. The ORP is based on substracting the SAPROF total score from the HCR-20 total score, and as discussed, 'scores' are not used in practice.

4.4.3. Conclusions

The aim of this research study was to explore the value of protective factors in the assessment of violence risk. The validity and reliability of the SAPROF as an assessment tool was investigated across a number of domains to evaluate whether the inclusion of protective factors improves predictive accuracy of violence risk assessment tools. It was found to have good internal reliability, and good concurrent, construct, predictive, incremental, and discriminative validity. The SAPROF was found to have enhanced construct and predictive validity over the HCR-20V3, suggesting there is value considering protective factors when attempting any assessment of violence risk.

Chapter Five

The Importance of a Strengths-Based Approach to Collaborative Risk Assessment and Risk Management: A research case study

Abstract

The aim of this case study was to explore the impact of psycho-education on a patient's motivation to engage in treatment and interventions to manage risk. The main objective was to establish whether risk assessment tools, which follow a strengths-based approach and focus on protective factors, have a positive impact on motivation to change when used in the collaborative risk assessment and management process. Risk management should be positive, collaborative, and strengths-based (Department of Health; DoH, 2007a). To enable recovery and rehabilitation individuals need to develop insight in to the risks related to their offending behaviour and mental illness (Birchwood, Spencer, & McGovern, 2000). In addition, individuals need to develop insight into the need for change, and it is essential for this to happen before effective interventions can take place (Bordin, 1979). Effective collaborative risk assessment and management may encourage insight into offending behaviour and mental illness to develop (Horstead & Cree, 2013). The case reviewed was a patient who took part in a ten week, collaborative risk assessment and risk management, psycho-education group programme. Motivation to change was measured using the University of Rhode Island Change Assessment (URICA; McConnaughy, Prochaska, & Velicer, 1983) after the patient viewed the team's ratings on the Historical Clinical Risk-20 Version 3 (HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013), Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2012), and Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004). The results were compared to establish if

one assessment increased motivation more than the others. Readiness to change increased after viewing the HCR-20V3 and the SAPROF; however reliable and clinically significant change was not indicated.

5.1. Introduction

5.1.1. Background

The purpose of risk assessment and risk management in forensic mental health services is to balance the risk of reoffending with appropriate interventions to manage and reduce risks. This is known as a 'risk-needs assessment' (Andrews, Bonta, & Hoge, 1990). It allows for least restrictive management proportionate to the level of risk and recommendation of the most appropriate intervention and support (Department of Health; DoH, 2007a).

Risk management should be positive, collaborative, and strengths-based (DoH, 2007a). Positive risk management acknowledges 'risk can never be completely eliminated'; and that risk management plans will contain decisions associated with risk (DoH, 2007a; p. 10). It is suggested that individuals should be supported to take reasonable risks, taking into account independence, well-being, and choice (DoH, 2007b). Collaborative risk management between the individual, carer, and professional entails developing a therapeutic relationship based on warmth, empathy, and a sense of trust, where the risk management process is explained at the earliest opportunity (DoH, 2007a). This allows for the development of a collaborative risk management plan in an open and transparent environment, which encourages the individual to have more autonomy over their care and treatment (Ryan & Deci, 2000). In terms of strengths-based risk management, Rapp and Goscha (2006) suggest when an individual's strengths are acknowledged, in addition to their vulnerabilities, and strategies to address problems are built around

their positive skills, the individual feels more able to cope, resulting in more effective risk management.

To enable recovery and rehabilitation individuals need to develop insight into the risks associated with their offending behaviour and mental illness (Birchwood, Spencer, & McGovern, 2000). In addition, individuals need to develop insight into the need for change, and it is essential for this to happen before effective interventions can take place (Bordin, 1979). Effective collaborative risk assessment and risk management may allow for insight into offending behaviour and mental illness to develop (Horstead & Cree, 2013). In contrast, if individuals are excluded from the risk assessment process this may result in mistrust of professionals, rejection or denial of risk issues, and possibly strengthen any cognitive distortions regarding their risk (Horstead & Cree, 2013). This is unlikely to create conditions supportive of change.

Prochaska and DiClemente (1986) developed the Stages of Change model which proposed four levels; pre-contemplation, contemplation, action, and maintenance. This was later updated to include determination (see figure 5.1; Prochaska & Velicer, 1997). It was suggested that if you can identify an individual's stage of change, you can develop interventions which are most appropriate to meeting their needs, and in turn they can move through the stages. It could be argued that collaborative risk assessment and risk management may increase insight into the need to change, and enable movement through the stages of change (Prochaska & Velicer, 1997). If an individual is aware of their risks and protective factors, and the impact they have, it is

possible they may be more motivated to engage in interventions to help manage risk and maintain protection (Bordin, 1979).

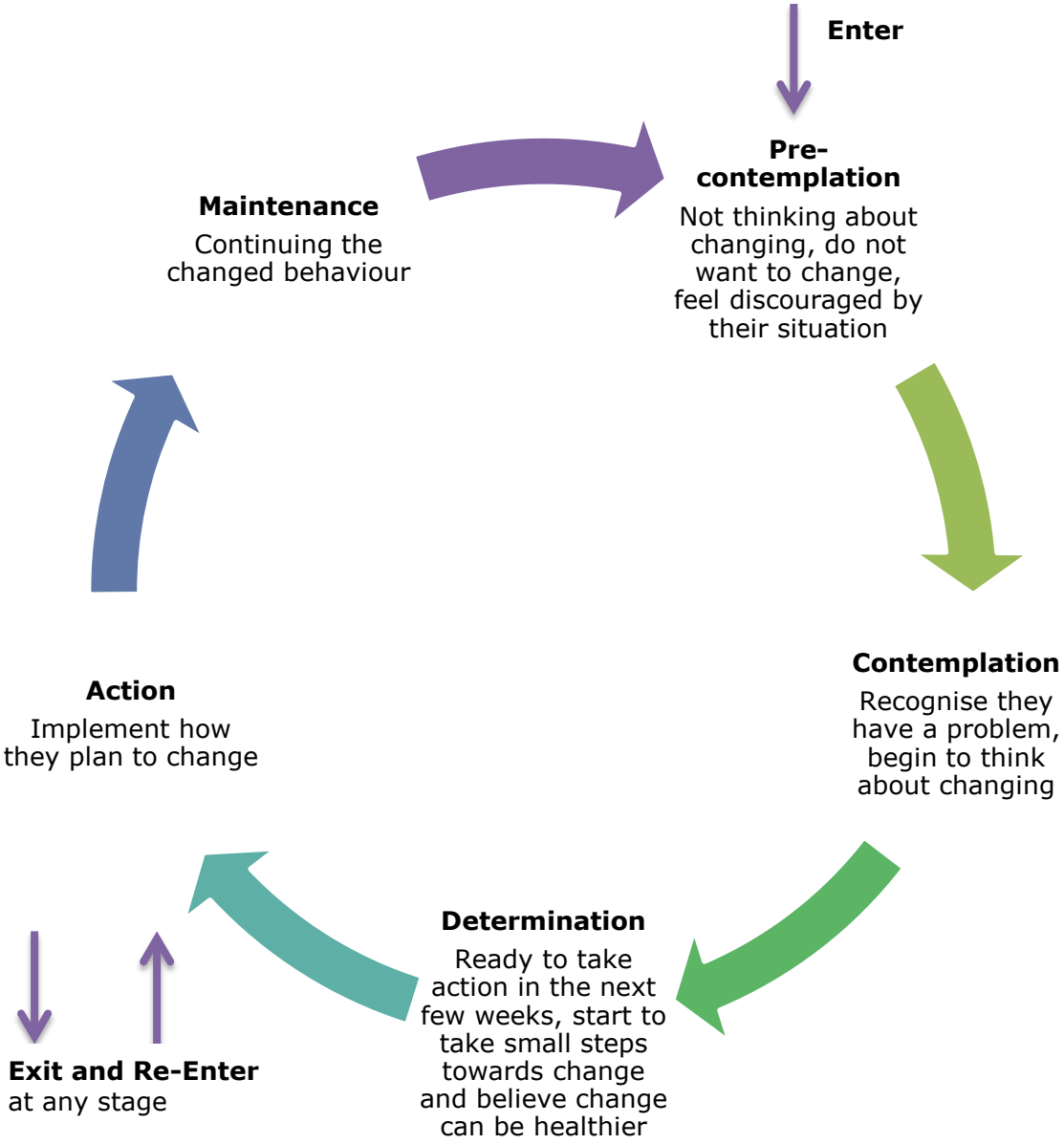


Figure 5.1. Stages of change model (adapted from Prochaska & Velicer, 1997)

To facilitate collaborative risk assessment and risk management, current best practice recommends the process should be explained as soon as possible (DoH, 2007a), and this could take the form of psycho-education. This is deemed to be an important and ethically crucial aspect of treatment for individuals with mental illness (Bauml, Frobose,

Kraemer, Rentrop, & Pitschel-Walz, 2006), and it has been reported that educating patients about their illness can increase coping strategies and awareness of relapse indicators (Aho-Mustonen, Miettinen, Koivisto, Timonen, & Raty, 2008). It inherently follows that the same could be applied to risk assessment and risk management. Psycho-education may need to include the definition of risk and risk management (including the difference between risk and protective factors), the purpose of risk management, and how it relates to treatment planning. There should be a focus on developing strengths, using positive treatment and interventions to further enhance the motivation of the individual and inspire a lifestyle free from offending (Horstead & Cree, 2013).

Psycho-education may also include information on the types of risk assessment used. In conditions of medium security it is recommended the Historical Clinical Risk Management - 20 (HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013) is used to assess potential risk of violence (service specification no. C03/S/a; NHS Commissioning Board, 2013). The HCR-20V3 evaluates the presence and relevance of risk factors, and uses these to formulate risk of violence and develop management plans. Due to its focus on risk factors the HCR-20V3 does not explicitly follow the strengths-based process recommended. Protective factors are considered in so much as they contribute to the formulation and the risk management plan; however there is no guidance in terms of how to identify protective factors.

It has been argued that collaborative risk management based solely on a focus on risk factors may be detrimental to individuals, as it can be demoralising and demotivating (Miller, 2006). To balance the risk

assessment and risk management process, taking into account risk factors and protective factors (Laub & Lauritsen, 1994), the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbe, 2012), and the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, & Desmarais, 2004) are also recommended for use as risk assessment tools in medium secure services (service specification no. C03/S/a; NHS Commissioning Board, 2013). The SAPROF evaluates the presence of protective factors to assess risk of future violence, thus focusing on the individual's strengths. The START assesses risk of future violence, but also evaluates other related issues such as self neglect, and it considers the presence of both strengths (protective factors) and vulnerabilities (risk factors).

There are a number of theories which support an emphasis on strengths and protective factors in the process of risk assessment and risk management. The Theory of Reasoned Action (TRA; Fishbien & Azjen, 2010) claims one of the strongest predictors of behaviour is intention. This is predicted by the attitude an individual holds towards the behaviour (behavioural beliefs), what the individual thinks others believe about the behaviour (normative beliefs), and the extent the individual believes they can control the behaviour (control beliefs). The more positive these beliefs are towards the behaviour, the more likely the individual is to perform it. In terms of reducing likelihood of offending the TRA suggests individuals would have negative attitudes towards offending, good social support which would disapprove of offending, and confidence in their ability not to offend. These elements could be viewed as protective factors or strengths. However, it should be noted that the

TRA assumes rationality and cognition are all that is needed to change behaviour, and this is not always the case (Kippax & Crawford, 1993).

The Good Lives Model (GLM) of offender rehabilitation (Ward and Stewart, 2003) is another theory supporting the use of protective factors in risk assessment and risk management. The GLM approach is based on the assumption that capabilities and strengths in people should be built on in order to reduce their risk of reoffending. It suggests that individuals offend because they are attempting to achieve a valued outcome in their life, but their route to this often involves harmful and antisocial behaviours due to vulnerabilities within the individual and their environment. Intervention and treatment is viewed as something which should add to the individual's strengths, rather than something which removes or manages a problem. This could be achieved through employment, leisure activities, and social integration. Solution-Focused Treatment (SFT; De Jong & Berg, 2008) follows the same tenet where individuals are encouraged to explore and build on positive personal goals. Andrews, Bonta, and Wormith (2011) argue that despite the popularity of the positive, strength—based focus of the GLM, it actually adds very little to the overarching theory of offender rehabilitation, the risk, need, and responsivity model (RNR; Andrews & Bonta, 2006) of offender assessment and treatment. This model suggests that to reduce the risk posed by offenders the treatment should be proportional to the *risk*, focused on the *need* (factors related to the offending), and delivered in a way that is *responsive* to the individual's abilities.

Strength-based risk management can also be enhanced by external and environmental factors. For example, detention or inpatient

treatment not only reduces the opportunity of offend, but offers support from professionals which may be emotional and practical, including motivational work. Control theory (Cochran, Wood, & Arneklev, 1994) suggests external factors can be seen as important socialisation methods which discourage offending behaviour. This is extremely important for individuals who have not yet developed personal internal strengths, or for those who may not be able to.

A review of the literature did not identify any research that has explored the use of risk assessment tools which assess the presence of protective factors (such as the SAPROF), and the impact on an individual's motivation to change (and engage in treatment). However, it could be inferred that if a strengths-based approach improves motivation, a risk assessment following the strengths-based approach would do so also.

Case studies appear to have been neglected in forensic mental health literature, possibly due to misunderstandings regarding their scientific value and rigour (Robinson, 2012). They are useful for investigating experiences at an individual level, allowing for greater understanding, theory development, and analytical generalisation (Yin, 2003). Perhaps most importantly, case studies allow for 'real-life' situations to be scrutinised and for views to be tested as they unfold in practice (Flyvbjerg, 2006). Davies, Howells, and Jones (2007) report it can be difficult to apply the results of larger studies, such as cohort studies or randomised control trials, to individuals in forensic mental health services, due to limited focus and generic conclusions. Case studies are particularly important in terms of formulation (Kreis & Cooke,

2012). Formulation is an important part of risk assessment and risk management, and provides an individualised and comprehensive view of the individual's needs and risks (and strengths), which then guides intervention (Cooke, 2010). This is something which is difficult to explore in large scale research.

5.1.2. Aims and objectives

A collaborative risk assessment and risk management training group programme for patients was introduced in a regional medium secure unit in May 2015. The aim of this case study was to explore the impact of the training on a patient's motivation to engage in treatment, and interventions to manage risk. The main objective was to establish whether risk assessment tools which follow a strengths-based approach and focus on protective factors, in comparison to those which focus on risk factors alone, have a positive impact on motivation to change (in terms of the Stages to Change model, Prochaska and DiClemente, 1986) when used in the collaborative risk assessment and risk management process. It was hypothesised that strengths-based collaborative risk assessment and management would have a positive impact on motivation to change.

5.2. Method

5.2.1. Setting

This was conducted in a medium secure unit (MSU) for adult men and women. The service aims to offer local inpatient provision to service users whose offending behaviour and mental health needs require they

are detained under the Mental Health Act (MHA; 1983, 2007 amendments) in secure conditions. It specializes in the assessment, treatment, and rehabilitation of adults with complex needs. It provides 65 beds, and has five wards which provide different levels of service to meet the patient's needs (acute, sub-acute, rehabilitation, and pre-discharge care). Of the 65 beds, 16 are for women (acute, rehabilitation, and pre-discharge care).

5.2.2. Case introduction

The case reviewed was a patient who took part in a ten week collaborative risk assessment and risk management psycho-education programme. For the purposes of this case study and to maintain anonymity, the patient will be referred to as 'Mr O'. Mr O was a 63 year old British Caucasian male. He was detained under sections 37 and 41 of the Mental Health Act (1983; 2007 amendments). His diagnosis was bipolar affective disorder. The index offence was grievous bodily harm (GBH) with intent against a woman previously unknown to him.

Mr O was born in a small village in the South East of England, and he grew up with his mother and father. He had one brother who was five years older. Mr O reports he had a happy childhood, and there are no reports or evidence of traumatic experiences during this period. His father worked as an agricultural worker before moving on to a horticultural nursery, and his mother was employed as a farm worker.

Mr O reports being bullied whilst at primary school, but this stopped when he went to secondary school. He did well at school, and left at the age of 16 years old, obtaining employment as a trainee

draftsman and designer. He remained in this position for 12 months, but had to leave due to the company going out of business. He worked as a landscape gardener for 14 months, but his employer was unable to keep him on due to the employment of a new worker who could drive, unlike Mr O. Between 1970 and 1976 Mr O worked for various construction companies as a labourer. During this period he gained a number of skills which allowed him to become a self-employed builder. In 1976 he began working in a colliery, where he remained until 1984 when he was made redundant as the coal mine closed. Mr O returned to his career as a self-employed builder until 1995, when he began training to become a registered general nurse. He graduated in 1998 and embarked on a career in general nursing, first working on general medical wards at a local hospital, followed by employment in a hospice where he continued working until 2012. His employment ended following an incident where he was unable to get home from the hospice due to poor weather. He had attempted to contact the director on a number of occasions, as transport had been guaranteed in such circumstances. The director did not like Mr O's persistence, and wrote to him the next day to dismiss him. The dismissal was challenged, and Mr O was able to continue working, although he did not return to the hospice. He once again returned to building work, and continued nursing on an agency basis.

Mr O's father died in 1980 at the age of 65 years old. He died from lung cancer and secondary liver cancer. His brother died two years later of leukaemia. In 1996 Mr O's mother died following an overdose of amitriptyline with whiskey. He reports his mother suffered from depression which he believes started after the death of his father, and

was exacerbated by the death of his brother. Mr O reports he had made a number of attempts to have his mother reviewed by the general practitioner, but they had not been prepared to visit her at home.

Mr O married his wife in 1971, and they remain together. His wife had a daughter from a previous relationship, and they had a son together in 1984 following a period of 12 years during which they had been trying to conceive. Their son was diagnosed with Asperger's Syndrome.

Mr O had a history of depression and anxiety which appeared to related to post traumatic stress disorder (PTSD), following a transient ischaemic attack which was the result of a motorbike accident in 2006. His symptoms were managed in the community, and he had no admissions to psychiatric services. Mr O reported that his symptoms of anxiety appeared to have become progressively worse since November 2013, and affected his ability to concentrate and remember things. He described 'stomach knots' and shaking which would get worse when he tried to do something, for example leave the house. In 2014 it was queried whether he was suffering from bipolar disorder due to the cyclical nature of his illness. Mr O reported suffering from low mood and depressive episodes with periods of elated mood where he would become energetic, over confident, and reckless.

Prior to Mr O's arrest for the index offence there was no recorded history of psychosis. At the time of his offence Mr O has reported experiencing auditory command hallucinations from a male voice that he refers to as 'Mr Fox'. Mr O reported he was taught by a 'Mr Fox' at school, and feels this is the foundation of the voice. He reported feeling frightened to disobey the instructions that the voice gave him. He was

unsure how long he had been hearing the voice for, but did not think it had been present for longer than a few months. Prior to the index offence the voice had instructed him to harm someone. At the time of the intervention it was unclear whether the offence was in response to a delusional system or whether it was an impulsive act.

There was no evidence of violent or anti-social behaviour before the index offence, which occurred in January 2014, when Mr O was convicted of GBH with intent, following an allegation that he had stabbed a 21 year old female shop worker without provocation. Mr O reportedly had no connection with the victim prior to the incident.

Mr O was admitted to the acute ward at the MSU in February 2014 following assessment in prison. During his early he was diagnosed as suffering from a severe depressive episode with psychotic features, and bipolar affective disorder. He continued to refer to 'Mr Fox', and his physical health deteriorated as he stopped eating and had limited fluid intake. Emergency treatment with electroconvulsive therapy (ECT) was implemented due to a high risk of suicide, and deterioration in his physical state. Subsequently Mr O's mental state improved, and during his 18 month admission he progressed through the care pathway. At the time of the current assessment, intervention, and evaluation Mr O resided on the pre-discharge ward.

5.2.3. Measures

The HCR-20V3 (Douglas et al., 2013) is a comprehensive set of professional guidelines for the assessment and management of violence risk. The HCR-20V3 assists professionals to evaluate a person's

likelihood of future violence and determine appropriate treatment and management strategies. The HCR-20V3 allows for the appraisal of the presence (and relevance) of 20 key violence risk factors. These are organized into three areas: historical, clinical, and risk management. The historical factors are static, whereas the clinical and risk management items are dynamic. The complete list is shown in table 4.2. The presence of factors is coded using a three level response format: 'yes' the factor is definitely or conclusively present; 'possible' the factor is possibly or partially present, or the risk factor is present, but the information is weak, contradictory, or inconclusive; 'no' the factor is absent, or the professional perceives no evidence the factor is present. The assessor should judge the relevance of factors with respect to the development of future risk management strategies. Relevance is also coded on a three level scale: 'low' the factor is of low relevance to the individual's risk for violence; 'moderate' the factor is relevant to some degree; 'high' the factor is present and its role in causing violence or impairing the effectiveness of risk management strategies is likely to be substantial. There is the option to rate a final risk judgement following completion of the assessment (low, medium, and high).

The SAPROF (De Vogel et al., 2012) was developed to complement the risk assessment of future violent behaviour in offenders and forensic psychiatric patients. It is a structured assessment guideline designed for use in combination with a reliable and valid violence risk assesment tool, such as the HCR-20V3 (Douglas et al., 2013). It aims to document and quantify the presence or absence of 17 protective factors in three areas: internal, motivation, and external (see table 4.1). All but two of the

criteria (intelligence and secure attachment in childhood) are dynamic. The factors are coded on a three point scale based on the degree to which the protective factor is present. 'No' means the protective factor is not present; 'perhaps' means the factor may be present but there is no conclusive evidence, or the factor is present only to some extent; 'yes' means the protective factor is definitely or clearly present. There is also the option to rate factors in regard to key items (a protective effect that is already present), and goal items (a protective effect which may occur after intervention). After rating the presence of the protective factors the assessor gives a final judgement which reflects the extent of 'protection'. 'Low' indicates there is little or no protection, 'moderate' refers to a moderate degree of protection, and 'high' indicates there is a high degree of protection. The SAPROF gives an integrative final risk judgement which combines and weighs the risk and protective factors.

The START (Webster et al., 2004) is a set of guidelines designed to evaluate mental disorder, monitor progress, plan treatment, and begin the process of estimating future risk to self and others. In addition to assessing risk of violence it aims to inform decision making in terms of self-harm, suicide, unauthorised leave, substance abuse, self-neglect, and victimization. There are 20 dynamic factors (with two additional case specific items) which are considered as strengths (protective factors) and as vulnerabilities (risk factors). These are outlined in table 4.3. The factors are coded on a three point scale; maximally present, moderately present, and minimally present. Key and critical items can also be selected, where a key item reflects a prominent strength and a critical item identifies a factor which needs specific attention in treatment

planning and supervision. Consideration of these areas allows for a specific risk estimate, 'low', 'moderate', or 'high' to be made for each area of risk (violence, self-harm, substance abuse, etc.).

5.2.3.1. Outcome measures

Motivation to change was measured using the University Rhode Island Change Assessment (URICA; McConnaughy, Prochaska, & Velicer, 1983) scale. The URICA consists of 32 questions which are answered on a five point Likert scale ranging from 1 (strong disagreement) to 5 (strong agreement). Responders are asked how closely they agree or disagree with each question. Appendix 5.A shows the questionnaire. There are four subscales which measure the assumed stages of change; pre-contemplation, contemplation, action, and maintenance. Its use is recommended in treatment and research to assess clinical progress and motivational readiness to change. The URICA yields a readiness score which indicates the individual's level of motivation to change in relation to the Stages of Change model. The individual's responses to each of the subscales are summed, and divided by the number of questions on the subscale. These mean scores are then totalled, and the pre-contemplation mean is subtracted from the total to obtain the readiness to change score.

The URICA was designed to assess attitudes toward changing problem behaviours (McConnaughy et al., 1983). It has good internal consistency, with coefficient alphas ranging between .79 and .89 (McConnaughy, DiClemente, Prochaska, & Velicer, 1989). It has also been shown to have good concurrent and predictive validity (DiClemente & Hughes, 1990). Although most commonly used to measure change in

terms of substance abuse, it has been found effective in measuring change of attitudes towards treatment. Dozois, Westra, Collins, Fung, and Garry (2004) administered the URICA to a sample of individuals with panic disorder at initial assessment, and following cognitive behavioural therapy. They found that higher scores on the action stage scale predicted engagement in treatment. However, as case study data can be difficult to generalise to wider populations, it can also be difficult to apply group research to individuals. As such this should be noted when utilising instruments to assess individuals.

5.2.4. Procedure

5.2.4.1. Referral details

Mr O was referred to the Risk and Protection Awareness training by his multidisciplinary team (MDT). The aim was to improve his insight in relation to his risk factors and protective factors to enable him to progress further in his care pathway, and to allow him to develop a collaborative risk assessment and risk management plan with his MDT. To establish the suitability of Mr O's referral his risks and needs were explored following the RNR model.

5.2.4.2. Treatment needs analysis

The assessment of risk allows the needs of the individual to be identified. All admissions are treated the same: during a patient's admission the HCR-20V3, SAPROF, and START are used to assess risk of future violence and other related risks. The HCR-20V3 and the SAPROF are completed on a six monthly basis, and the START is updated every

three months. The SAPROF and HCR-20V3 assessment were completed by a qualified psychologist, who conducted an exhaustive review of background documents, discussion with individuals who knew Mr O, and a clinical interview with Mr O. The START was completed following the consensus model, where members of the MDT (nursing staff, medical professionals, psychologists, occupational therapists, and social workers) discuss and agree the ratings for each factor.

The most recent assessments completed at the time of attendance at the training programme were used for evaluation of suitability for attendance, and also during the course of the intervention. They were reviewed, and if Mr O was found to have difficulties in relation to insight into his offending behaviour and mental illness, as identified by factor 'C1: Problems with insight' on the HCR-20V3, it was deemed an intervention aimed to improve this area, such as the Risk and Protection Awareness training, was needed.

5.2.4.3. General treatment considerations

To ensure the intervention was responsive to Mr O's abilities, his engagement in previous psychological interventions was explored, and any difficulties he experienced were highlighted. His level of cognitive functioning was reviewed to see if he might need any additional support during the intervention, and he was asked if he could foresee any difficulties or problems engaging. His mental health was assessed to ensure this would not limit his engagement, and this included consideration of his ability to cope with the possible difficult emotions associated with viewing his risk assessment, for example, shame in terms

of previous violence, distress in terms of past traumatic experiences, or anger in terms of reaction to the MDT's opinion.

5.2.4.4. Assessment

Mr O's suitability to participate in the intervention was further assessed with the use of a qualitative questionnaire which aimed to gather information about his current knowledge and understanding of the risk assessment and risk management process. This was completed by one of the training facilitators two weeks before the training was due to commence. Mr O was considered suitable to attend the training as there were deficits in his knowledge and understanding in this regard. Only patients who had a comprehensive awareness of risk assessment and risk management were excluded. A copy of the pre-training questionnaire can be seen in appendix 5.B. He was asked to complete the URICA questionnaire during the same session to provide a baseline measure of readiness to change.

5.2.4.5. Intervention

The 'Risk and Protection Awareness Training' was a group psycho-education programme which aimed to improve patients' understanding of the risk assessment and risk management process, and how it relates to their treatment plan. It was predicted that following completion participants would be in a position to work collaboratively with their MDT to develop their risk assessment and risk management plans.

There were ten one-hour training sessions in total, and a summary of the content can be seen in table 5.1. Mr O was one of eight patients

who attended the training programme. During the first session the participants were given an overview of what would be involved. The remainder of the session described the risk assessment and risk management process at the MSU. This included information about what risk assessment and risk management is, the differences between risk and protective factors, and the differences between static and dynamic factors. An overview of the risk assessment tools used was given, including the steps involved in completing the assessments: information gathering, rating the presence and relevance of factors, risk formulation, and the development of management strategies. During sessions two to seven patients were given further information about the core risk assessments used at the MSU (HCR-20V3, SAPROF, and START), and details about how they are used to guide treatment during their admission. They were given the opportunity to rate the presence of risk and protective factors for themselves, and they were able to compare these to the ratings given by the MDT. During session eight participants were given information about the process of risk formulation, how it is developed, and how it is used. In sessions nine and ten participants were encouraged to think about management plans, how they link to treatment strategies, and what they would personally put in place to manage their risk factors and maintain and build on their protective factors.

Throughout the training skills required for effective risk management were utilised to assist, build, and refine such abilities, for example brainstorming, problem solving, planning, and goal setting. The information was presented in verbal and visual formats, with the use of

charts to help patients rate the presence of risk and protective factors, and handouts which they could take away with them, to help aid learning and retention. At the end of the training each participant received a 'summary collaborative risk assessment and risk management plan' which they could use to help them work with their MDT to update future risk assessments and risk management plans.

Table 5.1

Risk and Protection Awareness training session outline

Number	Content
1	What is risk assessment and management?
2	HCR-20 overview and rating own HCR-20
3	Continue HCR-20 overview and rating own HCR-20 Comparing own HCR-20 to MDT HCR-20
4	SAPROF overview and rating own SAPROF
5	Continue SAPROF overview and rating own SAPROF Comparing own SAPROF to MDT SAPROF
6	START overview and rating own START
7	Continue START overview and rating own START Comparing own START to MDT START
8	What is formulation?
9	What are management plans?
10	Completing own summary risk assessment and management plan

5.2.5. Evaluation

5.2.5.1. Treatment outcome

Following attendance at the training it was anticipated that patients would be able to work more collaboratively with the MDT to develop their risk assessment, risk management plan, and treatment goals. The patient administration system was reviewed in the month after completion of the training to establish whether this had occurred.

5.2.5.2. Motivation

It was also envisaged that training would increase patients' motivation to engage in their treatment plan and give them more autonomy in relation to their care. In addition to the URICA being completed during pre-training assessment, it was used at different time points during the training. These were after viewing each of the risk assessments completed by their psychologist and MDT (end of sessions three, five, and seven).

5.2.5.3. Ethical considerations

Informed consent was obtained from Mr O for this information to be used for the case study. His capacity to consent was agreed by the MDT prior to approaching him. See appendix 5.C for a copy of the information sheet and consent form.

5.3. Results

5.3.1. Treatment needs analysis

Mr O was one of three group participants to receive a conclusively present rating on C1 of the HCR-20V3, suggesting he had a risk factor in the area of insight in terms of his offending behaviour and mental illness. Mr O's level of insight was not significantly different from the group participants ($z = 1$), but his lack of insight was considered a perpetuating factor for future offending. As such, improving his insight was determined to be a need to reduce the risk of Mr O reoffending in the future. Improving insight may enhance his motivation to engage in any further treatment recommended during his admission (Bordin, 1979). Mr

O was considered an appropriate referral to the Risk and Protection Awareness training.

5.3.2. General treatment considerations

Since his admission to the ward Mr O had engaged in individual psychological therapy and group treatment. He had not had any difficulties in these processes, and it could therefore be assumed that a group training programme was suitable. Mr O had average cognitive functioning abilities, according to the Wechsler Adult Intelligence Scale (WAIS-IV, Wechsler, 2008), with no obvious deterioration relative to his pre-morbid state, and there were no concerns regarding his reading and writing skills. He did not think he required any additional support, and did not think he would have any particular difficulties engaging. Mr O also reported there was nothing he would feel uncomfortable discussing in a group training setting. He said if he did experience any difficulties he would feel confident in asking staff for extra support, as he has done in the past. There were no concerns about Mr O attending the training.

5.3.3. Assessment

In the qualitative questionnaire Mr O reported that he would like to attend the training to gain more of an understanding of why 'I did what I did' (in reference to his index offence). He said he felt enthusiastic about attending, and felt it would be a good opportunity to 'learn about risks'. Mr O appeared to have a basic understanding of risk assessment and risk management, but was not aware of the specific risk assessment tools used during his admission. He reported, and it was corroborated through

review of his notes, that his risk assessment had not been discussed with him, and he had not seen a copy of the report. When asked what he felt might be a risk or protective factor, he gave examples in reference to his own experiences. For example, he reported that his sense of humour and use of language was a risk factor because it led others to misinterpret what he meant. On the other hand, he also thought his communication skills were a protective factor, along with an ability to empathise with others. He was able to identify that feeling agitated (but not being sure of what was causing his agitation) was an example of a relapse indicator, and that when feeling this way he could let his support system know so they could monitor his behaviour.

The qualitative questionnaire indicated that Mr O did not have a comprehensive understanding of risk assessment and the risk management process, and therefore was a suitable candidate for attendance at the Risk and Protection Awareness training.

5.3.3.1. URICA

Mr O achieved a readiness to change score of 9.86 (maximum achievable 15) prior to attending the training. This placed him in the contemplation stage of change. This suggests Mr O recognised that he had a mental illness and had engaged in offending behaviour in the past, and was considering what he needed to be doing to manage this. All results can be seen in figure 5.7.

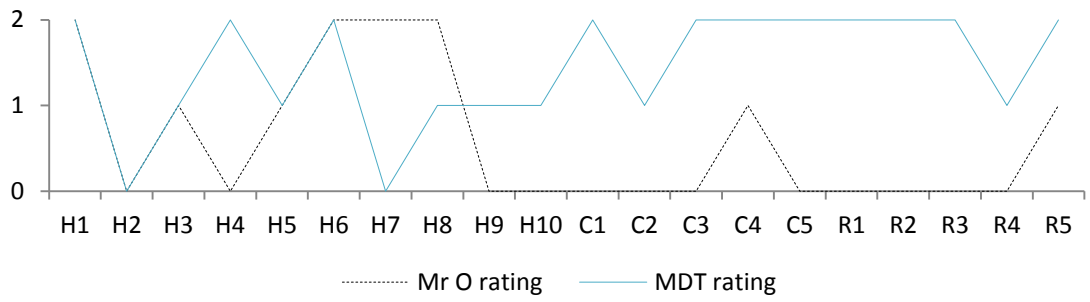
5.3.4. Intervention

Mr O was one of eight patients who attended the 'Risk and Protection Awareness Training'. He attended all 10 of the training sessions and he engaged well throughout. Mr O completed his own ratings of the presence of risk and protective factors on the HCR-20V3, SAPROF, and START, and compared his own opinions to that of his MDT. Figures 5.2, 5.3, and 5.4 show his ratings in comparison to the MDT ratings.

Mr O's ratings on the HCR-20V3 differed considerably from that of the MDT; specifically he appeared to under-estimate his risk factors in comparison to the opinion of the MDT. For example, on the HCR-20V3 he did not think he had definite or possible risk factors in 11 out of 20 areas, where as the MDT felt he had definite or possible risk factors in all areas except 'other anti-social behaviour'. Interestingly, Mr O felt he had risk factors in the areas of 'personality disorder' (H7) and 'traumatic experiences' (H8), which were not reflected in the MDT ratings. He said he had only recently been diagnosed with a personality disorder which may explain the discrepancy between his rating and the MDT rating. In terms of 'traumatic experiences' the MDT felt this was a possible risk factor, whereas Mr O felt it was a definite risk factor.

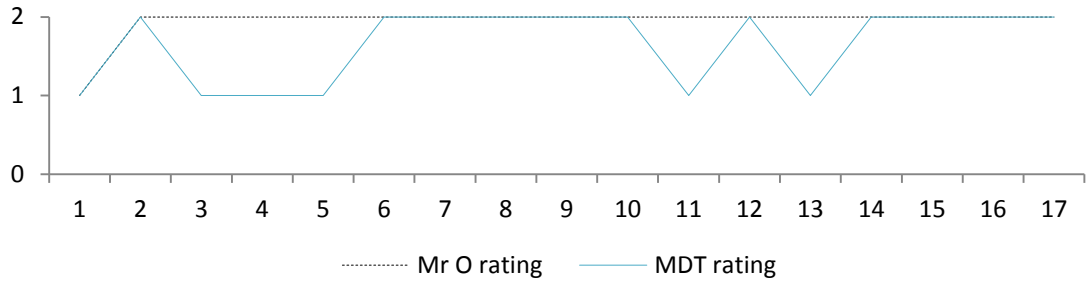
In contrast, Mr O's ratings on the SAPROF were very similar to the MDT ratings. They only differed in five out of 17 areas, and both Mr O and the MDT felt the protective factor was present, but the degree to which it was present varied. Mr O felt he had definite protective factors whereas the MDT felt they were possible protective factors.

Similarly, on the START Mr O's opinions on his strengths (protective factors) were comparable to the MDT ratings, while his opinions in regard to his vulnerabilities (risk factors) were at odds to their ratings. Mr O felt he had maximum or moderate vulnerabilities in just three areas, where as the MDT felt he had maximum or moderate vulnerabilities in all areas except 'substance use' (8). As with the SAPROF, Mr O and the MDT felt strengths were present in all areas, but the degree to which they were present varied.



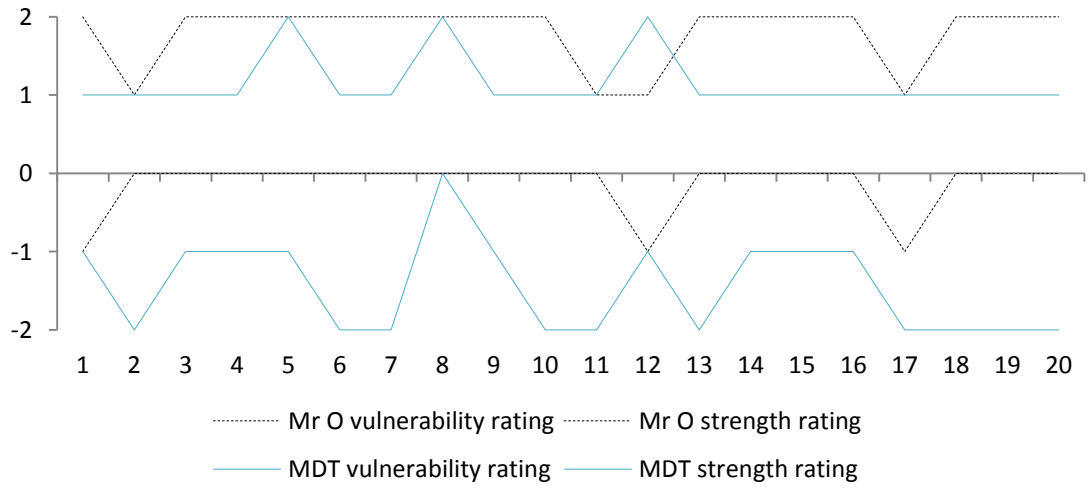
Presence:
2 = (Yes) Information indicates the factor is present
1 = (Possible) Information indicates the factor is possibly or partially present
0 = (No) Information indicates the factor is not present or does not apply

Figure 5.2. HCR-20 assessment



Presence:
2 = (Yes) The protective factor is clearly present
1 = (Perhaps) The protective factor may be present or is present to some extent
0 = (No) The protective factor is clearly absent, or there is no evidence that the protective factor is present

Figure 5.3. SAPROF assessment



Presence:
2 = Maximally present
1 = Moderately present
0 = Minimally present

Figure 5.4. START assessment

Due to the complex nature of risk formulation, participants of the training were given information about how a formulation is developed, and how it may relate to their treatment plans. However, they did not develop their own risk formulations. Mr O was able to use his identification of risk and protective factors to reflect on what a risk formulation related to his offending behaviour may look like. Figure 5.5 outlines the risk formulation using the MDT risk assessments as a basis for development.

There appeared to be no early life experiences such as removal from the family home or childhood abuse which could be considered predisposing factors. Mr O has a diagnosis of bipolar affective disorder, which may have predisposed him to be more likely to engage in risk behaviour. He also experienced a series of bereavements within his family (father, brother, mother, grandson) and his wife's family which may have contributed to this predisposition. Factors which appeared to be catalysts for Mr O's difficulties (precipitating factors) may have included being made redundant from his job as a nurse, a motorcycle accident, and difficulty managing the symptoms of his bipolar affective disorder. A more immediate precipitating factor appeared to be the experience of auditory hallucinations which commanded him to hurt others. Factors which maintained the possibility of risk behaviour (perpetuating factors) may have included post-traumatic stress disorder resulting from his motorcycle accident, which increased his symptoms of anxiety and depression, his diagnosis of bipolar affective disorder, and resultant difficulties coping with stress.

Mr O also had a number of protective factors which may have reduced the likelihood of engaging in risk behaviour. These included an average level of intelligence which would have allowed him to think rationally and act purposefully, a secure attachment in childhood, the ability to empathise with others, abstinence from drugs and alcohol, and the support of his wife.

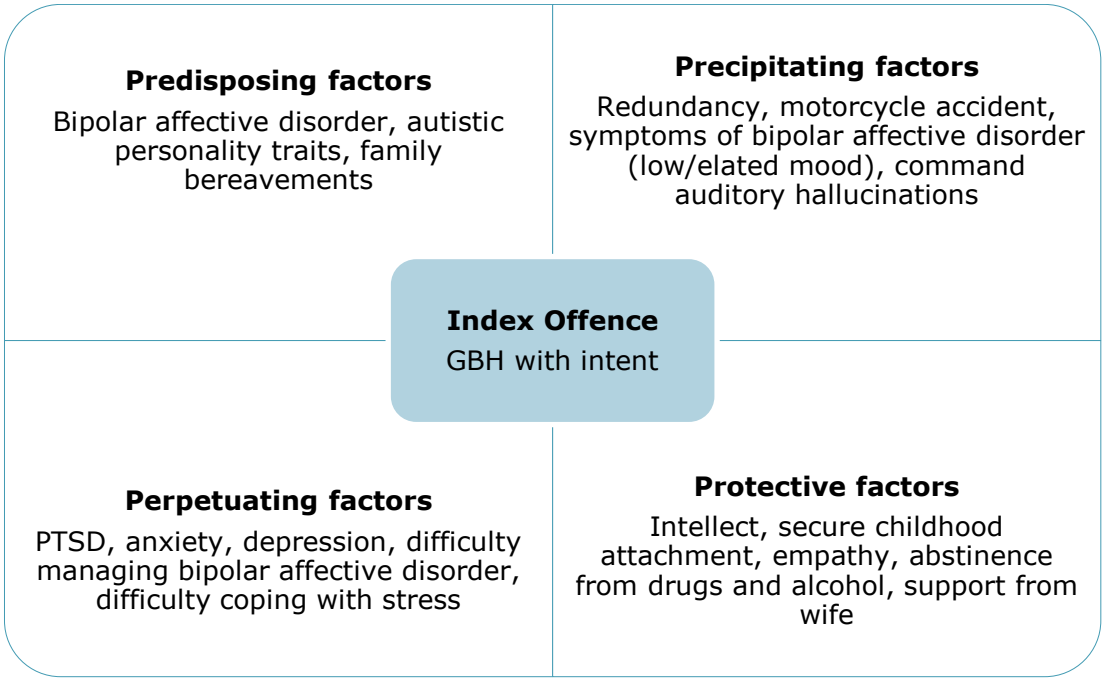


Figure 5.5. Mr O's risk formulation

In line with the risk assessment and risk management procedure Mr O was encouraged to identify which risk factors were relevant to the development of future risk management strategies, and which protective factors were essential for the prevention of future violent behaviour. He was also asked to identify what his relapse indicators were, and what strategies could be put into place to help manage his risk factors, and maintain and build on his protective factors. Finally, he was asked to set

some goals to aid the management of his risk factors and development of protective factors. Figure 5.6 outlines the plan he developed.

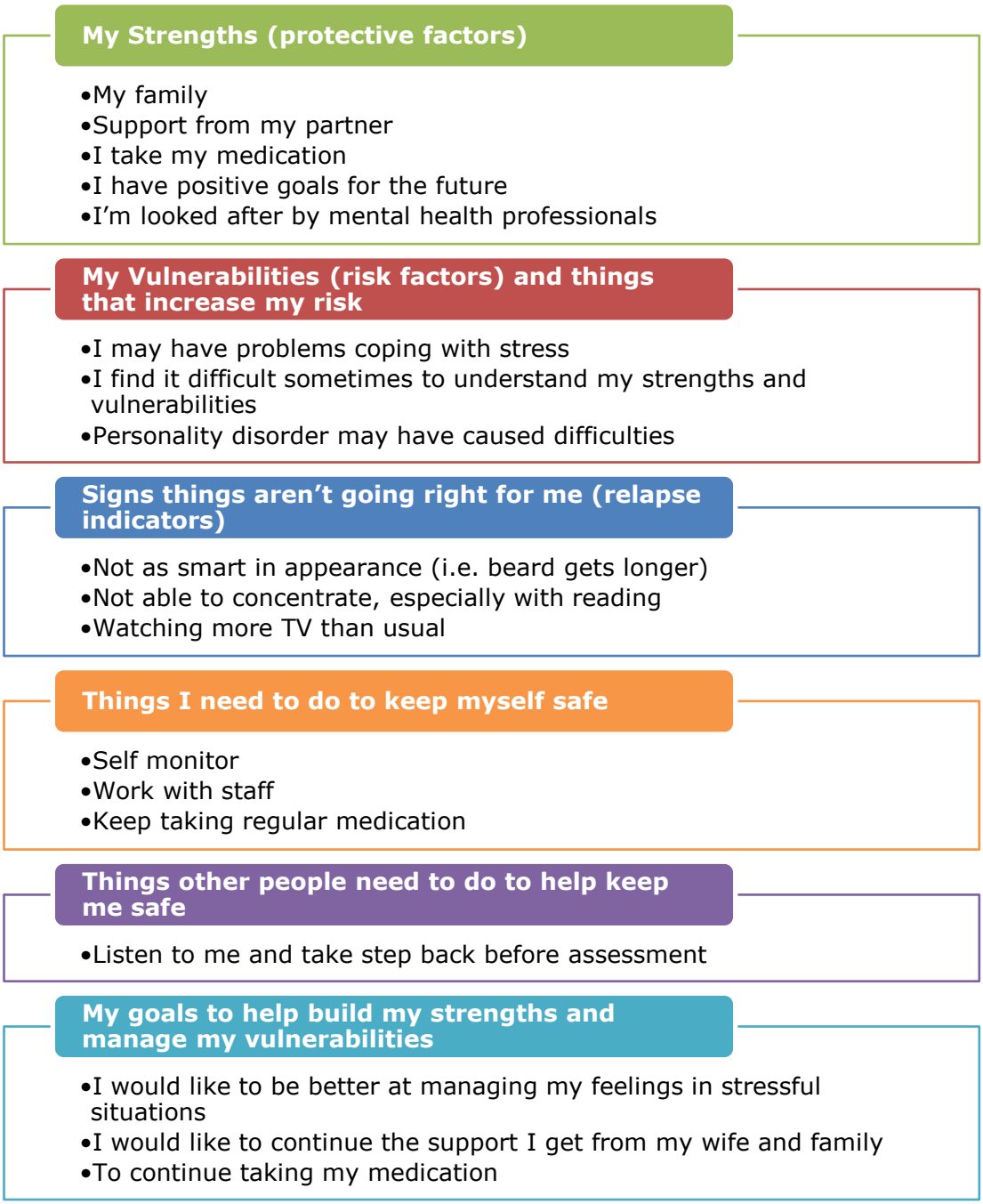


Figure 5.6. Mr O's summary collaborative risk assessment and risk management plan

5.3.5. Evaluation

5.3.5.1. Treatment outcome

Mr O appeared keen to collaborate with his MDT in the development of his risk assessment and risk management plans, evidenced by him approaching his psychologist and requesting to speak to her about his assessment. Review of the patient administration system revealed that following attendance at the Risk and Protection Awareness training Mr O met with his psychologist to collaboratively develop his risk assessment and risk management plans. Attendance at the training programme and subsequent collaborative discussion with his psychologist may have increased Mr O's insight into his risks in terms of his mental illness and offending behaviour, enabling progression through his care pathway, and it may have increased his ability to address his risk of reoffending by developing shared goals for treatment.

5.3.5.2. Motivation

Mr O's motivation to change was measured using the URICA after he viewed the team's ratings on the HCR-20V3, SAPROF, and START. The results are shown in figure 5.7. McConnaughy et al. (1983) state that scores less than 8 suggest the individual is in the pre-contemplation stage, scores between 8 and 11 suggest the contemplation stage, scores between 11 and 14 suggest the individual is in the action stage, and scores greater than 14 suggest the action stage of change. Prior to attending the training Mr O's readiness to change score was 9.86 (z = 1.26), which fell in the contemplation stage. After viewing the MDT's ratings of his risk factors on the HCR-20V3, Mr O's readiness to change

score was 10.00 (z = 1.26), which remained in the contemplation stage. The readiness to change score rose further to 10.86 (z = 1.47) after he viewed the MDT's rating of his protective factors on the SAPROF. This is approaching the action stage. After viewing the MDT's ratings of his strengths and vulnerabilities on the START, Mr O's readiness to change score reduced to 9.29 (z = 1.33). This was his lowest score, although it was higher than the baseline measure, and remained in the contemplation stage. Although higher at all stages, none of Mr O's readiness to change scores were significantly different from the group mean. Investigation of clinically significant change and reliable change (Jacobson & Truax, 1991) found no reliable (RC = 1.78) or clinically significant (criterion A = >11.75) improvement.

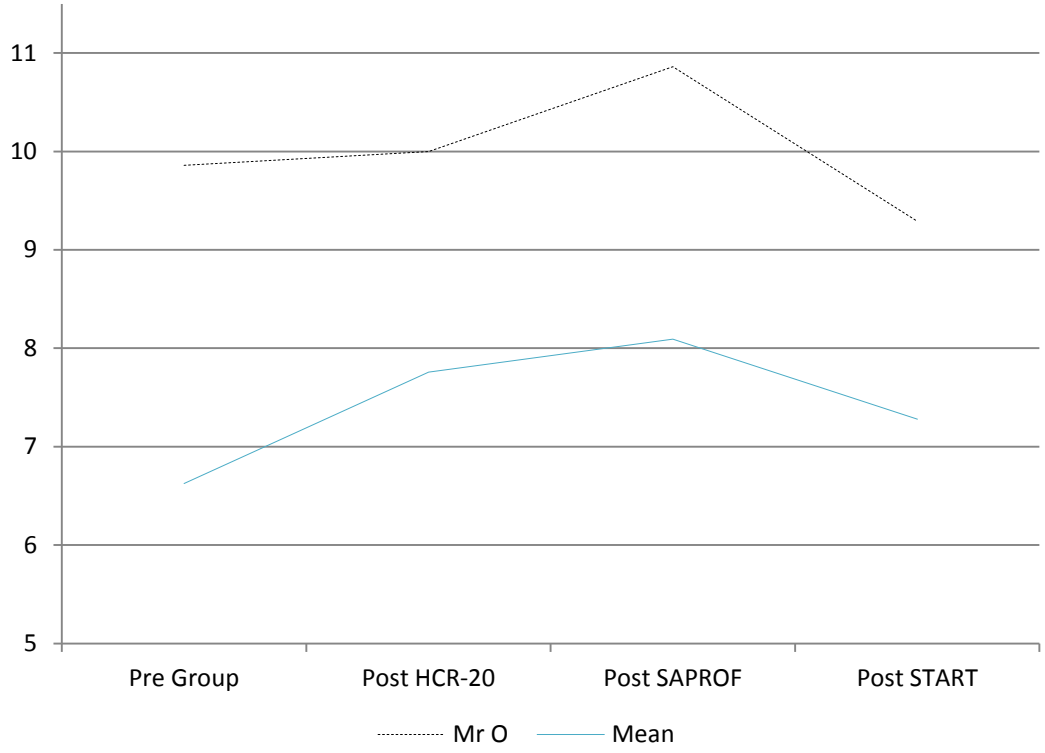


Figure 5.7. Mr O's readiness to change in comparison to the group mean

5.3.6. Further interventions identified in relation to reducing risk of future violence

Taking into account the risk factors and protective factors identified in the risk assessments completed, Mr O may benefit from interventions to help manage his risks and maintain and build on his protective factors. In line with the strengths-based approach, these should focus on developing his strengths, using positive treatment and interventions to further enhance his motivation.

Mr O may benefit from further offence focused work. This may involve attendance at the Offending Behaviour Group which follows a cognitive behavioural therapy (CBT; Beck, 1967) framework which has been found effective in offender treatment (Craig, Dixon, & Gannon, 2013). The GLM is incorporated in the programme which would allow for a strengths focused approach (Ward and Stewart, 2003).

Mr O may benefit from engaging in bereavement counselling to allow him to come to terms with the loss of family members, including his parents and brother. He may also benefit from attaining more coping skills to allow him to cope with the symptoms of his bipolar disorder, anxiety, and depression. This could be completed on an individual or group basis, perhaps following a CBT framework.

Mr O's relationship with his wife was a primary protective factor. To maintain this they may benefit from some family focused psycho-education to enable her to help support him in coping with his mental illness, management of risk factors, and maintaining and building protective factors. As Mr O approaches discharge he may wish to think about what leisure activities he would like to pursue to help him to

structure his time in a positive manner, and how he will maintain a positive relationship with mental health professionals to reduce the likelihood of deterioration in his mental state.

5.4. Discussion

The aim of this case study was to explore the impact of collaborative risk assessment and risk management training on a patient's motivation to engage in treatment and interventions to manage risk. The main objective was to establish whether risk assessment tools which follow a strengths-based approach and focus on protective factors, in comparison to those which focus on risk factors, have a positive impact on motivation to change in terms of the Stages to Change model (Prochaska & DiClemente, 1986), when used in the collaborative risk assessment and risk management process. It was hypothesed that strengths-based collaborative risk assessment would have a positive impact on motivation to change.

Mr O's motivation increased after reviewing the HCR-20V3 and SAPROF assessments. A possible explanation may be that viewing his risk and protective factors increased Mr O's motivation to take action and manage them to enable his progression through the care pathway (Ryan & Deci, 2000). Prior to attending the training, as identified in the assessment interview, he had limited understanding about what his risk and protective factors were. Increasing his awareness may have helped encourage him to think about what he could do to change (Bordin, 1979).

After viewing the SAPROF, Mr O's readiness to change score rose to 10.86. This was the peak level over the course of the training. This

provides some support for a strengths-based approach towards risk assessment and risk management. It suggests that understanding his risk factors may have increased his motivation to change, but understanding his protective factors might have increased his motivation to change further. However, Mr O's motivation decreased to its lowest point after viewing the START assessment. This is surprising because the START contains risk factors (vulnerabilities) which are balanced by protective factors (strengths).

Despite strengths-based collaborative risk assessment having some positive impact on motivation to change, the hypothesis was not supported as the change was not a reliable or clinicially significant. There was no significant difference between strengths-based assessments and those which focused on risk factors.

This case study demonstrated that a strengths-based shared understanding of risk and protection did not have a signiciant impact on motivation to change. This has a number of clinical implications. Research suggests inclusion of the assessment of protective factors, such as use of the SAPROF, and not just using tools which assess purely risk factors, such as the HCR-20V3, balances the risk assessment for professionals (Laub & Lauritsen, 1994), but this case study does not support increased motivation for change in patients. This suggests inclusion of protective factors may not be useful for patients.

The importance of the collaborative process was highlighted (DoH, 2007a). When rating the presence of risk factors on the HCR-20V3, Mr O identified two which the MDT did not consider important. One of these was explained by a recent diagnosis, which had been made following

completion of the risk assessment. The other was in the area of 'traumatic experiences', suggesting that Mr O felt this was more of a risk area than the MDT perceived. The MDT may have acknowledged the issue, but due to clinical judgement may have felt the evidence did not justify a 'yes' rating. A possible implication is that a treatment area was not identified, and as such the risk factor goes un-managed, which could result in an impact on future risk behaviours. Alternatively, the differences may have arisen because Mr O and the MDT have a different level of experience in this sort of rating. The MDT have received professional training, and are proficient in completing risk assessments. On the other hand Mr O's experience is limited, and he has no professional training in the area of risk assessment. During the collaborative process patients and the MDT may continue to disagree in their ratings, but it might also allow them to discuss their opinions to ensure nothing is overlooked, and to develop a shared understanding which they can use to jointly develop the management plan.

Another element which supports the collaborative process, and which is demonstrated by the differences in Mr O's and the MDT ratings, is the impact of a shared understanding on insight (Birchwood et al., 2000). Although insight was not measured or assessed in this case study it could be inferred that sharing the MDT views may have increased Mr O's understanding of his mental illness and offending behaviour. His ratings, compared to those of the MDT's, demonstrated he tended to over-estimate his protective factors and under-estimate his risk factors. This may result in him finding it difficult to understand why he has been referred to attend treatment in those areas. The collaborative process

may increase insight into the reasons for completing recommended interventions. This in turn may increase motivation to engage.

If a patient's motivation increases following collaborative risk assessment using a collaborative approach (albeit not significantly) it could be argued that a strengths-based risk management plan may also increase motivation. Future interventions may benefit from following this approach, for example those which employ the GLM framework. In addition, risk management plans that have been developed collaboratively between the individual and the MDT are more likely to be positively engaged with than those imposed on the individual. At the end of the training Mr O had developed his own risk management plan which reflected a number of needs, including the requirement to improve his skills to be able to manage stressful situations. Since he has recognised this need, he is more likely to be motivated to engage in an intervention to address it.

5.4.1. Limitations

One of the main limitations of this case study is the process of the Risk and Protection Awareness training itself. During the course the participants are given an outline of what the risk or protective factor means, and asked to rate how much they think that factor relates to them. They are then given a copy of the MDT's rating which they can compare to their own. However, the MDT risk assessment shared does not include natural language explanations as to why that classification has been given. This could lead to confusion about why they have been given the rating, and actually decrease motivation rather than increase it.

In an attempt to counter-act this, participants were encouraged to speak to their psychologists between training sessions to receive any clarification they felt they needed, but this would have occurred after their motivation was measured, so although clinically the impact was minimised, on a research level this was not addressed.

In addition, the order in which the risk assessments were introduced in the training programme may have impacted the results, and as such should be considered a confounding factor.

Use of the URICA to measure motivation to change has its limitations because there are mixed results in terms of its usefulness. For example, McMurran, Theodosi, and Sellen (2006) found the URICA did not consistently provide evidence of motivation for therapy and motivation to change. They recommended that better measures of motivation need to be developed. In addition, Blanchard, Morgenstern, Morgan, Labouvie, and Bux, (2003) found that readiness to change did not predict end of treatment outcomes, and they concluded the URICA showed limited clinical utility. Despite the limitations of the URICA there are few other tools available for use or validated for use in this setting (McMurran et al., 2006). In this case study motivation was also evaluated by investigating whether Mr O met with his MDT to collaborate on his risk assessment and risk management plan. However this only occurred after completion of the training, and did not contribute to the measure of motivation following sharing of each assessment within the training. Future explorations of motivation to change in this area may consider using alternative or multiple measures of motivation.

This case study did not measure insight directly, but instead inferred insight of risk areas, protection, and need for treatment increased following sharing of risk assessment information. However, giving information does not necessarily mean the individual will take on board that information and use it. An interview following completion of training may have established whether insight into risk and protective factors was present, and longer term follow-up may have been of benefit to investigate whether engagement in treatment increased following attendance at the training.

The nature of a case study means the results cannot be generalised to the wider population. To achieve this it may be useful to complete a similar investigation but with the use of a larger cohort study. However, this would lose the detail which this case study has provided, for example the details of Mr O's formulation and the development of his individualised risk management plan.

5.4.2. Future directions

Professionals seek to determine if the risk assessments they are using are effective in evaluating the risk of future violence. It would be interesting to investigate whether the contribution of patient's opinions to the risk assessment process increases or decreases the predictive validity of the risk assessment tool used.

5.4.3. Conclusions

To enable risk management to be positive, collaborative, and strengths-based in forensic mental health settings, patients (and their

carers) need to be involved in the process, and assessments and plans need to focus on their strengths and protective factors (DoH, 2007a). This helps to improve insight into mental illness and offending behaviour, along with insight into the need to change, which enables individuals to effectively engage in interventions to help manage risk (Birchwood et al., 2000 and Bordin, 1979). Psycho-education is a useful tool to implement this (Baumi et al., 2006). It should be commenced as early as possible, and continued throughout contact with services to enable the collaborative process to be fully integrated (DoH, 2007a). This encourages individuals to become proficient at assessing and managing their own risk, reducing the likelihood of future offending or relapse of mental illness (Horstead & Cree, 2013). However, this was not supported in this case study where no reliable or clinically significant change was found. A larger cohort study should be considered for future research to allow further exploration of this area.

Chapter Six

Discussion

The aim of this thesis was to explore and enhance the research evidence for the inclusion of protective factors in the violence risk assessment and management process in forensic mental health services, in the National Health Service (NHS). More specifically it investigated whether assessment of protective factors improves predictive accuracy of violence risk assessment, and discusses the implications for clinical practice. The impact on patient motivation to change is also considered.

Chapter two critically evaluated the most widely used violence risk assessment tool, the recently updated Historical Clinical Risk-20 Version 3 (HCR-20V3; Douglas, Hart, Webster, & Belfrage, 2013). The HCR-20V3 assists professionals to estimate a person's likelihood of future violence, and determine the most appropriate treatment and management strategies. An overview of the instrument was provided, and its measurable properties explored, considering its clinical and research applications, whilst also taking into account its use in forensic mental health settings. The HCR-20 is considered the most researched and best empirically guided risk assessment of violence, and it has been widely adopted (Douglas, Shaffer, Blanchard, Guy, Reeves, & Weir 2014). Version 3 of the instrument was introduced in 2013, and as such the evidence base for its reliability, validity and clinical utility is still in its infancy. Despite limited research, the evidence available suggested high levels of internal and inter-rater reliability, and good levels of concurrent and predictive validity. Its clinical utility was also supported.

One of the criticisms of the HCR-20V3 was the omission of assessment of protective factors in addition to risk factors. The manual suggests protective factors should be considered when developing the

formulation and management plan; however there is no guidance in terms of what constitutes a protective factor. Development of the HCR-20V3 was guided by empirical literature relating to factors consistent with violence, of which there is a plethora (Douglas et al., 2014). It is possible that assessment of protective factors alongside risk factors was not included because the empirical basis of what protects an individual from future violence remains in its infancy. As such there is no consensus as to how to define a protective factor, or how they work to reduce risk. As the popularity of including protective factors in the violence risk assessment process increases this is something which will no doubt improve. Chapter one concluded that the neglect of assessment of protective factors alongside risk factors in the HCR-20V3 supported the need of an additional tool such as the Structured Assessment of Protective Factors (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2012).

Chapter three systematically reviewed the research literature on the predictive accuracy of the violence risk assessment tools recommended for use in forensic mental health services, in the NHS: HCR-20, SAPROF, and Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicolls, and Desmarais, 2004). It aimed to establish whether the instruments including assessment of protective factors (the SAPROF and START) had enhanced predictive abilities (for absence of violence). Following PRISMA guidelines, research investigating the predictive validity of the HCR-20, SAPROF, and START in a population of adult male offenders, with a primary diagnosis of mental illness or personality disorder, and an outcome measure of future

violent behaviour or re-offending was included. The results suggested that the assessment tools have useful predictive validity. The SAPROF was found to have the highest level of predictive validity (for absence of violence), followed by the HCR-20 (for presence of violence), and the START had the lowest level of predictive validity (for absence and presence of violence). It was difficult to establish with any certainty whether the inclusion of protective factors improved the risk assessment process, but the combined use of the HCR-20 with the SAPROF, creating a new measure of risk balanced by protection, improved predictive accuracy for presence of violence, promoting the inclusion of protective factors.

Despite the identified research supporting the use of the SAPROF, the systematic review identified only eight empirical studies investigating the tools use, and only one of those was completed in England. In addition, the research conducted using an English population was completed using a community sample. This highlighted the need for further research regarding the use of the SAPROF in forensic mental health services in the NHS.

Chapter four investigated the reliability and validity of the SAPROF when implemented in a NHS forensic mental health service, to allow for a greater understanding of the effectiveness of its application in forensic inpatient settings. The SAPROF was found to have good internal reliability, concurrent, construct, and discriminative validity. The SAPROF demonstrated good predictive accuracy for absence of violence, and the results suggested the tool's predictive abilities were superior to the HCR-20V3, and the START, although not significantly. Combining the use of

the SAPROF with the HCR-20V3 significantly increased the predictive accuracy for presence of violence. The results heightened the argument for assessment of protective factors in the risk assessment process, and supported the use of the SAPROF alongside the HCR-20V3.

Chapter five evaluated the impact of protective factors in assessment of violence risk on a patient's motivation to change. The clinical implications of the findings and recommendations for future research were also discussed. The main objective was to establish whether risk assessment tools, which follow a strengths-based approach and focus on protective factors, had a positive impact on motivation to change when used in the collaborative risk assessment and management process. The results showed that readiness to change was at its highest after the patient viewed his SAPROF assessment, although there was no reliable or significant change in motivation levels. This suggested a focus on protective factors did not have a positive impact on motivation to change superior to that of a focus on risk factors.

6.1. Limitations

This thesis has some important limitations which should be noted. Chapter two provided a critical evaluation of the HCR-20V3 (Douglas et al., 2013), however because the tool is relatively new, research investigating its reliability and validity remains in its infancy. In addition, much of the research was conducted by the tools authors or researchers associated with the authors. Only two studies included in the critique had sample sizes of over 100 (Doyle et al., 2014, and Strub et al., 2014), and the majority of studies were conducted outside the United Kingdom

(Canada, Norway, Germany, Netherlands, and Sweden), making the results of the remaining studies hard to generalise. It should also be noted that a number of the studies described in the critique were based on draft versions of the HCR-20V3, which will have differed from the published assessment. Therefore the research described may not be applicable to the final published HCR-20V3.

Some selection bias may have been present in chapter three which systematically reviewed the research literature on the predictive accuracy of the HCR-20, SAPROF, and START risk assessments. Language bias may have been present with 17% of studies excluded on the basis of language or due to the full text being unavailable. Additionally, due to time constraints it was not possible to contact the authors of relevant resources if there were missing data, which may have had implications in terms of the studies included, particularly in relation to the defined population. Author bias was also an important factor in the review. Half of the research investigating the predictive accuracy of the SAPROF and 42% of research investigating the predictive validity of the START was completed by the authors (or those affliated with) of the tool.

A meta-analysis was not conducted after completion of the systematic review. The main reason for this was differing operationalisations of the outcome of 'violent behaviour'. For example, the outcome of 'violence' is different from 'any aggressive behaviour' which may include acts such as verbal aggression. In addition, the outcomes differed for studies investigating the SAPROF (and some examining the START) as they assessed absence rather than presence of violence.

The comparison of absence of violence to presence of violence has a noteworthy impact on base rate levels, which should be acknowledged as limitation for both the systematic review and the research outlined in chapter four which investigated the reliability and validity of the SAPROF as an assessment tool. The base rate was much higher for absence of violence, and investigation of high base rates is likely to produce higher incidences of the investigated construct (i.e. absence of violence) being identified independent of the usefulness of the tool (Conroy & Murrie, 2007).

In terms of other limitations for the empirical research, although low and medium secure settings were included, the results found may not generalise to other settings. The outcomes may not be applicable to community settings and the sample was relatively homogenous. As with the systematic review, the outcome measure of the empirical research may have been problematic as although it was local policy for all violent incidents to be reported, it was a subjective decision made by professionals as to whether an incident made the threshold for reporting, and only those observed would have been reported.

Self-harm was included in the outcome as a form of violence in both the systematic review and the empirical project because research conducted by Abidin et al. (2013) had found the HCR-20, SAPROF, and START to be predictive of self-harm (to varying degrees). However, it is not generally considered violent behaviour and is not included in the definition of violence in the HCR-20V3 (Douglas et al., 2013). The inclusion of self-harm will have inflated the number of reported violent

incidents, and as such impacted on the base rates of absence and presence of violence.

Overall, it should be remembered that in practice risk assessments are not used to predict whether or not a patient will be violent in the future. The outcome of the risk assessment is recommendations for management plans and treatment during the patient's admission. Whereas in research, risk assessments are reduced to numerical outcomes, and the qualitative information (which is the main benefit of SPJ tools) is lost. The use of total scores on each of the assessment tools and calculated optimal cut-off scores is a noteworthy limitation of the empirical project (and the systematic review), as predictive validity based purely on the total scores (and generated cut-offs) may not be reflective of the level of risk identified by the clinician.

This brings into question the relevance of investigating the predictive accuracy of violence risk assessment tools, and its application to clinical practice. It could be argued it may be more effective to evaluate future violence dependent on the management and treatment plans implemented in response to the risk assessment, rather than the outcome of the risk assessment itself.

Finally, in terms of limitations for the case study outlined in chaper five, which investigated the impact of collaborative risk assessment and management psycho-education training on a patient's motivation to engage in treatment, and interventions to manage risk, the study design itself was limiting. The nature of a case study means the results cannot be generalised to the wider population, insight was not directly measured in relation to collaborative risk assessment and management, and the

order in which the risk assessments were introduced in the training programme may have impacted the results. In addition, the use of the URICA to measure motivation to change has its limitations because there are mixed results in terms of its usefulness (McMurran, Theodosi, & Sellen, 2006).

6.2. Implications for practice

Despite the limitations discussed there are a number of implications for practice. This thesis provides research evidence for the inclusion of protective factors in the violence risk assessment and management process in forensic mental health services in the NHS. The HCR-20, SAPROF, and START were found to hold good predictive validity for absence and presence of violence (dependent on the tool used), and the inclusion of protective factors improved predictive validity of the risk assessment tools, supporting the continued use of these instruments.

The SAPROF's Overall Risk Protection (ORP) index also achieved good concurrent, construct, predictive, and discriminative validity, and it was significantly more accurate than the HCR-20V3 in predicting presence of violence, advocating a measure where violence risk is balanced by protective factors. However, the ORP index is purely a research concept, and is a result of substracting the SAPROF total score from the HCR-20 total score. As already discussed, in practice 'scores' are not used, and as such it is difficult to comprehend how this can be applied in practice. Perhaps though it further supports the proposal of considering protective factors in addition to risk factors when using clinical judgement to evaluate risk of future violence.

It should also be remembered that even the best evaluations of violence risk can only explain a moderate amount of variance (Lösel, 2001). Although all the evaluated risk assessment tools held good predictive validity, the thesis also showed the propensity for the assessments to identify non-violent individuals as violent, and vice versa. This highlights the instruction that risk assessment tools should be used to help guide clinical decision-making, but should not be used to replace this process.

Despite the limitations of applying such research to practice, the findings remain important as they add to empirical evidence and recommendations for best clinical practice (Cooke & Michie, 2013). This thesis found value in the inclusion of protective factors in violence risk assessment, and adds to the empirical basis of what constitutes a protective factor.

With the popularity of strength-based approaches to treatment increasing, such as the Good Lives Model (GLM, Ward & Stewart, 2003), it could be reasoned strength-based risk assessment will become similarly prevalent, and further support the additional use of protection focused assessments. However, criticisms of the GLM approach state that it adds nothing to the overarching theory of offender rehabilitation, the risk, need, and responsivity model (RNR; Andrews, Bonta, and Wormith, 2011), and the same might be said for strength-based risk assessment. Perhaps what is more important though is the perception of staff and patients. If staff who complete risk factor focused assessments see their patients as inherently 'risky' (Rogers, 2000), and patients who view risk factor focused assessments become demotivated (Miller, 2006),

surely it would be better for services (in terms of financial cost and resources), and patients (in terms of wellbeing) to focus on the assessment of positive protective factors.

The importance of collaborative development of risk assessments and management plans is well documented (DoH, 2007a), and the case study goes some way to show how emphasis on protective factors can assist this process. However, the case study did not show that a focus on protective factors was superior to a focus on risk factors. Despite this, if collaborative risk assessment and management improves patient involvement, it could be argued this may enhance patient insight into their offending behaviour, and in turn reduce the likelihood of future offending. With the increased involvement of patients in the development of their risk assessments and management plans, it may be worth considering the accuracy of these collaborative assessments, in comparison to those completed without the input of the patient. It could be predicted that due to the enhanced knowledge of their personal circumstances, the patient's involvement would improve the assessments accuracy, and due to invested interest engagement in management and treatment plans would improve (DoH, 2007a).

The need for an assessment of protective factors in addition to risk factors is highlighted, and in practice it is recommended the most appropriate tool for the identified need should be used (DoH, 2007a). This research would suggest that as the most accurate assessment, the SAPROF should be this instrument. However, the SAPROF was not developed to be an assessment used independently of another valid SPJ tool, and it could be argued the balance would be tipped in the opposite

direction if risk factors were discounted. This is not reflected in the results, which support assessment of specifically protective factors to evaluate violence risk.

6.3. Suggestions for future research

Future research should take into account the limitations discussed. The systematic review identified a number of confounding factors which would be areas of improvement for the quality of future research studies. These included study design, follow-up periods, sample size, implication of lost participants, selection bias, and measurement bias. Systematic reviews investigating similar constructs should consider the possibility of a meta-analysis, and ensure the outcomes selected allow for this to occur.

Research into the reliability and validity of the SAPROF remains in its infancy, and as such the empirical research should be replicated in other settings and populations with a larger sample size, and a variety of follow-up periods. Due to the study design, the analysis of inter-rater reliability was not possible, and the researcher was not blind to the assessment information when collecting the follow-up data. Both these factors may have impacted on the results of the research, and future investigations may wish to conduct such work blind to clinical information. Research may also wish to consider using a standardised tool to explore the outcome measure, to reduce subjective report of violent behaviour. Removal of self-harm as the outcome should be also be considered.

In practice the final judgement ratings of low, medium, and high are utilised by professionals to guide their assessment of level of protection (and risk; Douglas et al., 2013). However, it was not possible to investigate this construct as the setting used for research did not consistently utilise the final judgement ratings. Future research may wish to consider using the final judgement ratings in addition to the total scores.

Finally, as suggested in the case study in chapter five, it would be interesting to investigate whether the contribution of patient's opinions to the risk assessment and management process increases or decreases the predictive validity of the risk assessment tool used.

6.4. Conclusions and recommendations

This thesis aimed to explore and enhance the research evidence for the inclusion of protective factors in the violence risk assessment and management process in forensic mental health services, in the NHS. Value was found in the inclusion of protective factors which improved the predictive accuracy of violence risk assessment. However, strengths-based collaborative risk assessment did not improve patient motivation to a greater extent than traditional risk based methods. It is recommended assessment of protective factors should continue to be considered in the risk assessment process in NHS forensic mental health services.

References

Abdi, H. (2007). The Bonferonni and Šidák Corrections for Multiple Comparisons. In N. Sakind (Ed.). *Encyclopedia of Measurement and Statistics* (page numbers unavailable). Thousand Oaks (CA): Sage.

*Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. G. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: Prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BMC Psychiatry, 13*(197). Retrieved from http://link.springer.com/article/10.1186%2F1471-244X-13-197.

**Abou-Sinna, R., & Luebbers, S. (2012). Validity of assessing people experiencing mental illness who have offended using the Camberwell Assessment of Need-Forensic and Health of the Nation Outcome Scales-Secure. *International Journal of Mental Health Nursing, 21*(5), 462-470. doi: http://dx.doi.org/10.1111/j.1447-0349.2012.00811.x.

Aho-Mustonen, K., Miettinen, R., Koivisto, H., Timonen, T., & Raty, H. (2008). Group psychoeducation for forensic and dangerous non-forensic long-term patients with schizophrenia. A pilot study. *European Journal of Psychiatry, 22*(2), 84-92. doi: 10.4321/S0213-61632008000200004.

Andrews, D. A., & Bonta, J. (1998). *The psychology of criminal conduct* (2nd ed.). Cincinnati, OH: Anderson.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis.

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

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

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

**Babalola, O., Gormez, V., Alwan, N. A., Johnstone, P., & Sampson, S. (2014). Length of hospitalisation for people with severe mental illness. *Cochrane Database of Systematic Reviews,* (1). Retrieved from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000384.pub3/abstract doi:10.1002/14651858.CD000384.pub3.

Baldessarini, R., Finkelstein, S., & Arana, G. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*(5), 569-573. doi: 10.1001/archpsyc.1983.01790050095011.

*Barber-Rioja, V., Dewey, L., Kopelovich, S., & Kucharski, L. (2012). The Utility of the HCR-20 and PCL:SV in the Prediction of Diversion Noncompliance and Reincarceration in Diversion Programs. *Criminal Justice and Behavior, 39*(4), 475-492. doi:

http://dx.doi.org/10.1177/0093854811432609.

* Bardnard-Croft. (2014). *A quantitative research report on the validity of the SAPROF to predict community violence, with a focus on comparisons across different diagnostic groups*. (unpublished doctoral dissertation). University of Manchester, UK.

Bauml, J., Frobose, T., Kraemer, S., Rentrop, M., & Pitschel-Walz, G. (2006). Psychoeducation: A basic psychotherapeutic intervention for patients with schizophrenia and their families. *Schizophrenia Bulletin, 32*(1), S1-S9. doi: 10.1093/schbul/sbl017.

Beck, A. (1969). *Depression – Clinical Experiemental and Theoetical Aspects*. New York: Harper and Row.

Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *Journal of Forensic Psychiatry, 9*(2), 328-338. doi: 10.1080/09585189808402200.

**Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *Journal of Forensic Psychiatry, 11*(1), 167-175. doi: http://dx.doi.org/10.1080/095851800362445.

Birchwood, M., Spencer, E., & McGovern, D. (2000). Schizophrenia: early warning signs. *Advances in Psychiatric Treatment, 6,* 93-101. Retrieved from http://schizophrenia.com/.

Bjørkly, S., Eidhammer, G., & Selmer, L. (2014). *Concurrent validity and clinical utility of the HCR-20V3 compared to the HCR-20: Similar tools but improved method*. Journal of Forensic Nursing, 10(4), 234-242. doi: 10.1097/JFN.0000000000000047.

**Bjorkly, S., Hartvig, P., Roaldset, J. O., & Singh, J. P. (2014). Norwegian developments and perspectives on violence risk assessment. *Criminal Justice and Behavior, 41*(12), 1384-1397. doi: http://dx.doi.org/10.1177/0093854814547949.

Blanchard, A. & Douglas, K. (2011). *APLS Undergraduate Paper Award: The historical-clinical-risk management-Version 3: The inclusion of idiographic relevance ratings in violence risk assessment*. Poster presented at the annual conference of the American-Psychology Law Society, Miami, FL.

Blanchard, K. A., Morgenstern, J. Morgan, T. J., Labouvie, E., Bux, D. A. (2003). Motivational subtypes and continuous measures of readiness for change: concurrent and predictive validity. *Psychology of Addictive Behaviors, 17*(1), 56-65. doi: 10.1037/0893-164X.17.1.56.

**Blum, F. M. (2004). *Psychopathy, psychosis, drug abuse, and reoffense among conditionally released offenders* (unpublished doctoral dissertation). University of Southern California, California.

Boer, D., Hart, S., Kropp, P., & Webster, C. (1997). *Manual for the Sexual Violence Risk-20. Professional guidelines for assessing risk of sexual violence.* Burnaby, BC: Simon Fraser University, Mental Health, Law, and Policy Institute.

Boland, A., Cherry, G., & Dickson, R. Eds. (2014). *Doing a systematic review. A students guide.* London: Sage.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, research and practice*, *16*(3), 252-260. doi: 10.1037/h0085885.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.

*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). *The International Journal of Forensic Mental Health, 9*(4), 271-281. doi: http://dx.doi.org/10.1080/14999013.2010.534378.

Bryman, A. 2012. *Social research methods* (4th ed.). Oxford: Oxford University Press.

**Buchanan, A. (2014). Predicting violent offences by released prisoners. *The British Journal of Psychiatry, 204*(3), 240. doi: http://dx.doi.org/10.1192/bjp.204.3.240.

Caldwell, R., Bogat, G., & Davidson, W. (1988). The assessment of child abuse potential and the prevention of child abuse and neglect: A policy analysis. *American Journal of Community Psychology, 16*(5), 609-624. Retrieved from http://link.springer.com/article/10.1007/BF00930017.

**Campbell, M. A., French, S., & Gendreau, P. (2009). The Prediction of Violence in Adult Offenders. *Criminal Justice and Behavior, 36*(6), 567-590. doi: http://dx.doi.org/10.1177/0093854809333610

CASP (2013). Critical Appraisal Skills Programme: Making sense of evidence. *Critical Appraisal Skills Programme.* Retrieved from www.casp-uk.net/

**Castelletti, L., Rivellini, G., & Stratico, E. (2014). Predictive efficacy of violence risk assessment tools, implications for forensic and

general psychiatry in Italy. A literature review. *Journal of Psychopathology / Giornale di Psicopatologia, 20*(2), 153-162. Retrieved from http://www.gipsicopatol.it/issues/2014/vol20-2/07-Castelletti-abstract.html.

**Cesniene, I. (2010). Violence risk assessment in two forensic samples using the HCR-20 in Lithuania. *European Psychiatry, 25*(1), 2-131. doi: 10.1016/S0924-9338(10)70745-8.

**Chakhssi, F., de Ruiter, C., & Bernstein, D. (2010). Reliability and validity of the Dutch version of the Behavioural Status Index: A nurse-rated forensic assessment tool. *Assessment, 17*(1), 58-69. doi: http://dx.doi.org/10.1177/1073191109338815.

**Chu, C. M., Daffern, M., & Ogloff, J. R. (2013). Predicting aggression in acute inpatient psychiatric setting using BVC, DASA, and HCR-20 Clinical scale. *Journal of Forensic Psychiatry & Psychology, 24*(2), 269-285. doi: http://dx.doi.org/10.1080/14789949.2013.773456.

*Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & Daffern, M. (2011a). The Short- to Medium-Term Predictive Accuracy of Static and Dynamic Risk Assessment Measures in a Secure Forensic Hospital. *Assessment, 20*(2), 230-241. doi: 10.1177/1073191111418298.

*Chu, C. M., Thomas, S. D., Ogloff, J. R., & Daffern, M. (2011b). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) in a secure forensic hospital: Risk factors and strengths. *The International Journal of Forensic Mental Health, 10*(4), 337-345. doi: http://dx.doi.org/10.1080/14999013.2011.629715.

**Claix, A., & Pham, T. H. (2004). Evaluation of the HCR-20 Violence

Risk Assessment Scheme in a Belgian forensic population. *Encephale, 30*(5), 447-453. doi: 10.1016/S0013-7006(04)95459-0.

Cochran, J. K., Wood, P. B., & Arneklev, B. J. (1994). Is the religiosity–delinquency relationship spurious? A test of arousal and social control theories. *Journal of Research in Crime and Delinquency, 31*(1), 92–123. doi: 10.1177/0022427894031001004.

Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.

*Coid, J., Ullrich, S., & Kallis, C. (2013). Predicting future violence among individuals with psychopathy. *The British Journal of Psychiatry, 203*(5), 387-388. doi: http://dx.doi.org/10.1192/bjp.bp.112.118471.

*Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., Rogers, R. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology, 77*(2), 337-348. doi: http://dx.doi.org/10.1037/a0015155.

*Coid, J., Yang, M., Ullrich, S., Zhang, T., Steve, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *The Journal of Forensic Psychiatry & Psychology, 22*(1), 3-21. doi: http://dx.doi.org/10.1080/14789949.2010.495990.

Conway, M. & Murrie, D. (2007). *Forensic Assessment of Violence Risk: A guide for risk assessment and risk management*. New Jersey: John Wiley & Sons.

Cooke, D. (2010). Personality disorder and violence: Understand violence risk: An introduction to the special issue. *Journal of Personality Disorders*, *24*(5), 539–550. doi: 10.1521/pedi.2010.24.5.539.

Cooke, J., & Michie, C. (2013). Violence risk assessment: From prediction to understanding, from what? to why? In C. Logan & E. Johnstone (Eds.), *Managing clinical risk: A guide to effective practice* (pp. 3–26). New York, NY: Routledge.

Copas, J., & Marshall, P. (1998). The Offender Group Reconviction Scale: The statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society*, *47*(1), 159-171. doi: 10.1111/1467-9876.00104.

**Cote, G., Crocker, A. G., Nicholls, T. L., & Seto, M. C. (2012). Risk assessment instruments in clinical practice. *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie, 57*(4), 238-244. Retrieved from http://europepmc.org/abstract/med/22480589.

* Coupland, R. (2015). *An examination of dynamic risk, protective factors and treatment related change in violent offenders*. (unpublished doctoral dissertation). University of Saskatchewan, Saskatoon.

Craig, L, Dixon, L., and Gannon, T. (Eds.) (2013). *What works in offender rehabilitation. An evidenced-based approach to assessment and treatment.* West Sussex: Wiley-Blackwell.

**Crocker, A. G., Braithwaite, E., Laferriere, D., Gagnon, D., Venegas, C., & Jenkins, T. (2011). START changing practice: Implementing a

risk assessment and management tool in a civil psychiatric setting. *The International Journal of Forensic Mental Health, 10*(1), 13-28. doi: http://dx.doi.org/10.1080/14999013.2011.553146.

**Daffern, M. (2007). The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior, 12*(1), 116-130. doi: 10.1016/j.avb.2006.03.005.

**Daffern, M., & Howells, K. (2007). The prediction of imminent aggression and self-harm in personality disordered patients of a high security hospital using the HCR-20 Clinical Scale and the Dynamic Appraisal of Situational Aggression. *The International Journal of Forensic Mental Health, 6*(2), 137-143. doi: http://dx.doi.org/10.1080/14999013.2007.10471258.

**Dahle, K.-P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry, 29*(5), 431-442. doi: http://dx.doi.org/10.1016/j.ijlp.2006.03.001.

Davies, J., Howells, K., & Jones, L. (2007). Evaluating innovative treatment in forensic mental health: A role for single case methodology? *The Journal of Forensic Psychiatry & Psychology*, *18*(3), 353–367. doi: 10.1080/14789940701443173.

**Davoren, M., Abidin, Z., Naughton, L., Gibbons, O., Nulty, A., Wright, B., & Kennedy, H. G. (2013). Prospective study of factors influencing conditional discharge from a forensic hospital: The DUNDRUM-3 programme completion and DUNDRUM-4 recovery

structured professional judgement instruments and risk. *BMC Psychiatry, 13*(1), 185. doi: 10.1186/1471-244X-13-185.

**Davoren, M., O'Dwyer, S., Abidin, Z., Naughton, L., Gibbons, O., Doyle, E., . . . Kennedy, H. G. (2012). Prospective in-patient cohort study of moves between levels of therapeutic security: The DUNDRUM-1 triage security, DUNDRUM-3 programme completion and DUNDRUM-4 recovery scales and the HCR-20. *BMC Psychiatry, 12*(80). doi: 10.1186/1471-244X-12-80.

*de Borba Telles, L. E., Folino, J. O., & Taborda, J. G. V. (2012). Accuracy of the Historical, Clinical and Risk Management Scales (HCR-20) in predicting violence and other offenses in forensic psychiatric patients in Brazil. *International Journal of Law and Psychiatry, 35*(5-6), 427-431. doi: http://dx.doi.org/10.1016/j.ijlp.2012.09.001.

de Jong, P., & Berg, I. K. (2008). *Interviewing for solutions* (3rd Edition). Belmont, CA: Thompson Brooks.

**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. (2004). Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *Journal of Forensic Psychiatry and Psychology, 15*(1), 145-164. doi: 10.1080/14789940410001655916.

**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 & Psychotherapy, 12*(3), 226-240. doi: http://dx.doi.org/10.1002/cpp.452.

**de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law, 12*(3), 321-336. doi: http://dx.doi.org/10.1080/10683160600569029.

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2007). *SAPROF. Richtlijnen voor het beoordelen van beschermende factoren voor gewelddadig gedrag. Dutch version.*Utrecht: Forum Educatief.

**de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2009a). *Structured Assessment of Protective Factors for Violence Risk (SAPROF)*. Utrecht: Forum Educatief.

**de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2009b). *Guidelines for the assessment of protective factors for violence risk.* Paper presented at the English Version. Utrecht, The Netherlands: Forum Educatief.

** de Vogel, d. V., Ruiter, d. C., Bouman, Y., & de Vries Robbé, M. (2010). *SAPROF: Structured Assessment of PROtective Factors for violence risk: Leitlinien für die Erfassung von protektiven Faktoren bei einem Risiko für gewalttätiges Verhalten*. Utrecht: Forum Educatief.

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbe, M. (2012). *Guidelines for the assessment of protective factors for violence risk (2nd ed)*. Utrecht: De Forensische Zorgspecialisten.

**de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of Discharge and Risk of Recidivism Measured by the HCR-20: A Retrospective Study in a Dutch Sample of Treated Forensic Psychiatric Patients. *The International Journal of Forensic Mental Health, 3*(2), 149-165. doi: http://dx.doi.org/10.1080/14999013.2004.10471204.

de Vogel, V. & de Vries Robbé, M. (2013). *Innovation in risk assessment: The value of the HCR-20 V3 in forensic clinical practice in the Netherlands*. Symposium presented at the annual conference of the International Association of Forensic Mental Health Services, Masstricht, Netherlands.

** de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2014). *Female Additional Manual: Additional guidelines for the HCR-20V3 for assessing risk for violence in women*. Utrecht: Van der Hoeven Kliniek.

**de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International Journal of Forensic Mental Health, 10*(3), 171-177. doi: 10.1080/14999013.2011.600230.

* de Vogel, V., van den Broek, E., & de Vries Robbé, M. (2014). The Use of the HCR-20V3 in Dutch Forensic Psychiatric Practice. *International Journal of Forensic Mental Health, 13*(2), 109-121. doi: 10.1080/14999013.2014.906518.

**de Vries Robbé, M. (2014). Protective Factors (Doctoral thesis). Retrieved from

http://www.saprof.com/_files/Phd%20Thesis%20Protective%20Factors.%20Michiel%20de%20Vries%20Robbe.pdf.

**de Vries Robbé, M. Protective Factors: Validation of the SAPROF for Sexual Offenders. Unpublished.

*de Vries Robbé, M., & de Vogel, V. (2011). *Protective factors for violent behaviour: Clinical results with the SAPROF.* Paper presented at the tenth Conference of the International Association of Forensic Mental Health Services, Vancouver, Canada.

**de Vries Robbé, M., & de Vogel, V. (2013). Protective Factors for violence risk. Bringing balance to risk assessment and management. In C. Logan & L. Johnstone (Eds), Managing clinical risk: A guide to effective practice (p. 293-310). London: Routledge.

*de Vries Robbé, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study of the SAPROF. *International Journal of Forensic Mental Health, 10*(3), 178-186. doi: 10.1080/14999013.2011.600232.

*de Vries Robbé, 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 & Psychology, 24*(4), 440-457. doi: http://dx.doi.org/10.1080/14789949.2013.818162.

**de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting

reductions in postdischarge community recidivism. *Law and Human Behavior, 39*(1), 53-61. doi: http://dx.doi.org/10.1037/lhb0000089.

**de Vries Robbé, M., de Vogel, V., Koster, K., & Bogaerts, S. (2015). Assessing protective factors for sexually violent offending with the SAPROF. *Sexual abuse: a journal of research and treatment*, *27*(1), 51-70. doi: 10.1177/1079063214550168.

**de Vries Robbé, M., de Vogel, V., & Stam, J. (2012). Protective factors for violence risk: The value for clinical practice. *Psychology, 3*(12), 1259-1263. doi: 10.4236/psych.2012.312A187.

*de Vries Robbé, M., de Vogel, V., Wever, E., Douglas, K., & Nijman , H. Risk and Protective Factors for Inpatient Aggression. Unpublished. In de Vries Robbé, M. (2014). Protective Factors (Doctoral thesis). Retrieved from http://www.saprof.com/_files/Phd%20Thesis%20Protective%20Factors.%20Michiel%20de%20Vries%20Robbe.pdf.

**de Vries Robbé, M., Mann, R., Maruna, S., & Thornton, D. (2014). An exploration of protective factors supporting desistance from sexual offending. *Sexual abuse: A journal of research and treatment*. doi: 10.1177/1079063214547582.

Department of Health. (2007a). *Best Practice in Managing Risk. Principles and evidence for best practice in the assessment and management of risk to self and others in mental health services.* Retrieved from http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/prod

_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_076512.pdf.

Department of Health. (2007b). *Independence, choice and risk: a guide to best practice in supported decision making*. Retrieved from http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/dh_074773.

*Dernevik, M., Grann, M., & Johansson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the HCR-20. *Psychology, Crime & Law, 8*(1), 93-111. doi: http://dx.doi.org/10.1080/10683160208401811.

**Desmarais, S., 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., Read, J. D., & Brink, J. (2010). Confidence and accuracy in assessments of short-term risks presented by forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology, 21*(1), 1-22. doi: http://dx.doi.org/10.1080/14789940903183932.

*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. doi: http://dx.doi.org/10.1037/a0026668.

**Desmarais, S. L., Sellers, B. G., Viljoen, J. L., Cruise, K. R., Nicholls, T. L., & Dvoskin, J. A. (2012). Pilot implementation and preliminary evaluation of START:AV assessments in secure juvenile correctional facilities. *The International Journal of Forensic Mental Health, 11*(3), 150-164. doi: http://dx.doi.org/10.1080/14999013.2012.737405.

**Desmarais, S. L., Van Dorn, R. A., Telford, R. P., Petrila, J., & Coffey, T. (2012). Characteristics of START assessments completed in mental health jail diversion programs. *Behavioral Sciences & the Law, 30*(4), 448-469. doi: http://dx.doi.org/10.1002/bsl.2022.

DiClemente, C. & Hughes, S. (1990). Stages of change profiles in outpatient alcoholism treatment. *Journal of Substance Abuse, 2*(2), 217–235. doi: 10.1016/S0899-3289%2805%2980057-4.

**Dietiker, J., Dittmann, V., & Graf, M. (2007). Risk assessment of sex offenders in a German-speaking sample. Applicability of PCL-SV, HCR-20+3, and SVR-20. *Der Nervenarzt, 78*(1), 53-61. doi: http://dx.doi.org/10.1007/s00115-006-2110-2.

*Dolan, M., & Blattner, R. (2010). The utility of the Historical Clinical Risk-20 Scale as a predictor of outcomes in decisions to transfer patients from high to lower levels of security-A UK perspective. *BMC Psychiatry, 10*(76), 1-8. doi: 10.1186/1471-244X-10-76.

Dolan, M. & Doyle, M. (2002). Violence risk prediction: Clinical and actuarial measures and the role of the psychopathy checklist. *British Journal of Psychiatry, 177*, 303-311. doi: 10.1192/bjp.177.4.303.

**Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale

second edition (VRS-2) in a British forensic inpatient sample. *Journal of Forensic Psychiatry & Psychology, 18*(3), 381-393. doi: 10.1080/14789940701489390.

**Dolan, M., & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behavior, 30*(6), 469-483. doi: http://dx.doi.org/10.1002/ab.20044.

**Douglas, K. S. (2004). Making structured clinical decisions about violence risk: Reliability and validity of the HCR-20 violence risk assessment scheme. *Dissertation Abstracts International*, 64, 8B.

**Douglas, K. S. (2014a). Introduction to the Special Issue of the HCR-20 Version 3. *International Journal of Forensic Mental Health, 13*(2), 91-92. doi: 10.1080/14999013.2014.913746.

**Douglas, K. S. (2014b). Version 3 of the Historical-Clinical-Risk Management-20 (HCR-20V3): Relevance to Violence Risk Assessment and Management in Forensic Conditional Release Contexts. *Behavioral sciences & the law, 32*(5), 557-576. doi: 10.1002/bsl.2134.

**Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *The International Journal of Forensic Mental Health, 13*(2), 130-139. doi: http://dx.doi.org/10.1080/14999013.2014.908429

Douglas, K., Hart, S., Webster, C., & Belfrage, H. (2013). *HCR-20V3. Accessing risk for violence. User Guide.* 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. doi: 10.1080/14999013.2014.906519.

**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., & Ogloff, J. R. (2003a). Violence by Psychiatric Patients: The Impact of Archival Measurement Source on Violence Base Rates and Risk Assessment Accuracy. *Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie, 48*(11), 734-740. Retrieved from https://ww1.cpa-apc.org/Publications/Archives/CJP/2003/.../douglas.pdf.

**Douglas, K. S., & Ogloff, J. R. P. (2003b). The impact of confidence on the accuracy of structured professional and actuarial violence risk judgments in a sample of forensic psychiatric patients. *Law and Human Behavior, 27*(6), 573-587. doi: 10.1023/B:LAHU.0000004887.50905.f7.

**Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*(10), 1372-1379. doi: 10.1176/appi.ps.54.10.1372.

Douglas, K. S., Shaffer, C., Blanchard, A., Guy, L. S., Reeves, K., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and*

annotated bibliography. Retrieved from https://kdouglas.wordpress.com/hcr-20/.

**Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative Validity Analysis of Multiple Measures of Violence Risk in a Sample of Criminal Offenders. *Criminal Justice and Behavior, 32*(5), 479-510. doi: http://dx.doi.org/10.1177/0093854805278411.

**Dowsett, J. (2005). Measurement of risk by a community forensic mental health team. *Psychiatric Bulletin, 29*(1), 9-12. doi: http://dx.doi.org/10.1192/pb.29.1.9.

*Doyle, M., Carter, S., Shaw, J., & Dolan, M. (2012). Predicting community violence from patients discharged from acute mental health units in England. *Social Psychiatry and Psychiatric Epidemiology, 47*(4), 627-637. doi: http://dx.doi.org/10.1007/s00127-011-0366-8.

** Doyle, M., Coid, J., Archer-Power, L., Dewa, L., Hunter-Didrichsen, A., Stevenson, R., Shaw, J. (2014). Discharges to prison from medium secure psychiatric units in England and Wales. *The British Journal of Psychiatry, 205*(3), 177-182. doi: 10.1192/bjp.bp.113.136622.

** 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. doi: 10.1046/j.1365-2850.2002.00535.x.

**Doyle, M., & Dolan, M. (2006). Predicting Community Violence from Patients Discharged from Mental Health Services. *The British Journal of Psychiatry, 189*, 520-526. doi:

http://dx.doi.org/10.1192/bjp.bp.105.021204.

**Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology, 7*(2), 141-154. doi: http://dx.doi.org/10.1348/135532502760274756.

**Doyle, M., Lewis, G., & Brisbane, M. (2008). Implementing the Short-Term Assessment of Risk and Treatability (START) in a forensic mental health service. *Psychiatric Bulletin, 32*(11), 406-408. doi: 10.1192/pb.bp.108.019794.

**Doyle, M., Power, L. A., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting Post-Discharge Community Violence in England and Wales Using the HCR-20V3. *International Journal of Forensic Mental Health, 13*(2), 140-147. doi: 10.1080/14999013.2014.906517.

Dozois, J. D., Westra, A. H., Collins, A. K., Fung, T. S., & Garry, J. K. (2004). Stages of change in anxiety: psychometric properties of the University of Rhode Island Change Assessment (URICA) scale. *Behaviour Research and Therapy, 42*(6), 711-29. doi: 10.1016/S0005-7967%2803%2900193-1.

**Dunbar, E., Quinones, J., & Crevecoeur, D. A. (2005). Assessment of hate crime offenders: The role of bias intent in examining violence risk. *Journal of Forensic Psychology Practice, 5*(1), 1-19. Retrieved from http://edunbar.bol.ucla.edu/papers/assessmentroleofbiasintentrisk.pdf.

**Edworthy, R., & Khalifa, N. (2014). A clinical database for measuring

outcomes in a low-secure service: A feasibility study. *Journal of Forensic Practice, 16*(1), 58-67. doi: http://dx.doi.org/10.1108/JFP-08-2012-0006.

**Eidhammer, G., Selmer, L. E., & Bjørkly, S. A brief report from a comparison of the HCR-20 and the HCR-V3 concerning internal consistency and clinical utility. Retrieved from http://www.researchgate.net/publication/266160089_A_brief_report_from_a_comparison_of_the_HCR-20_and_the_HCR-V3_concerning_internal_consistency_and_clinical_utility.

**Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *BMJ (Online), 345*(7868). doi: 10.1136/bmj.e4692.

Fishbein, M., & Azjen, I. (2010). *Predicting and changing behaviour: The reasoned action approach*. New York: Psychology Press.

Flaata, A. & Marthe, K. (2013). *A single case study illustration from a high security ward*. Symposium presented at the annual conference of the International Association of Forensic Mental Health Services, Maastricht, Netherlands.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*, 219–245. doi: 10.1177/1077800405284363.

Feldt, L. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*(3), 363-373. doi: 10.1007/BF02289364.

Fleiss, J. (1986). *Design and analysis of clinical experiments*. New York: John Wiley and Sons.

**Fluttert, F., Van Meijel, B., Webster, C., Nijman, H., Bartels, A., & Grypdonck, M. (2008). Risk management by early recognition of warning signs in patients in forensic psychiatric care. *Archives of psychiatric nursing, 22*(4), 208-216. doi: 10.1016/j.apnu.2007.06.012.

**Fujii, D. E., 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. doi: 10.1176/appi.ps.56.6.711.

**Gairing, S., de Tribolet-Hardy, F., Vohs, K., & Habermeyer, E. (2013). Criminological and psychiatric characteristics of inmates of preventive detention and their impact on the treatability. *Der Nervenarzt, 84*(1), 65-71. doi: http://dx.doi.org/10.1007/s00115-011-3434-0.

**Garcia-Mansilla, A., Rosenfeld, B., & Cruise, K. R. (2011). Violence risk assessment and women: predictive accuracy of the HCR-20 in a civil psychiatric sample. *Behavioral Sciences & the Law, 29*(5), 623-633. doi: http://dx.doi.org/10.1002/bsl.1005.

**Gravier, B., & Lustenberger, Y. (2005). Assessment of the risk of violent behaviours: State of the discipline. *Annales Medico-Psychologiques, 163*(8), 668-680. doi: 10.1016/j.amp.2005.07.019.

*Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., . . . Snowden, R. J. (2011b). The Short-Term Assessment of Risk

and Treatability (START): A prospective study of inpatient behavior. *The International Journal of Forensic Mental Health, 10*(4), 305-313. doi: http://dx.doi.org/10.1080/14999013.2011.631692.

**Gray, N. S., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J. (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV, and the HCR-20. *Psychological Assessment, 19*(4), 474-479. doi: http://dx.doi.org/10.1037/1040-3590.19.4.474.

**Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71*(3), 443-451. doi: http://dx.doi.org/10.1037/0022-006X.71.3.443.

**Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative Efficacy of Criminological, Clinical, and Personality Measures of Future Risk of Offending in Mentally Disordered Offenders: A Comparative Study of HCR-20, PCL:SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*(3), 523-530. doi: http://dx.doi.org/10.1037/0022-006X.72.3.523.

**Gray, N. S., Taylor, J., & Snowden, R. J. (2008). Predicting violent reconvictions using the HCR-20. *The British Journal of Psychiatry, 192*(5), 384-387. doi: http://dx.doi.org/10.1192/bjp.bp.107.044065.

*Gray, N. S., Taylor, J., & Snowden, R. J. (2011a). Predicting violence using structured professional judgment in patients with different mental and behavioral disorders. *Psychiatry Research, 187*(1-2), 248-253. doi: http://dx.doi.org/10.1016/j.psychres.2010.10.011.

**Green, D., Belfi, B., Griswold, H., Schreiber, J. M., Prentky, R., & Kunz, M. (2014). Factors Associated with Recommitment of NGRI Acquittees to a Forensic Hospital. *Behavioral Sciences & the Law, 35*(5), 608-626. doi: 10.1002/bsl.2132.

**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 & Psychology, 15*(2), 278-292. doi: http://dx.doi.org/10.1080/1478994032000199095.

**Guy, L. S. (2010). Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey. *Dissertation Abstracts International, 71,* 4B.

**Guy, L. S., Douglas, K. S., & Hendry, M. C. (2010). The role of psychopathic personality disorder in violence risk assessments using the HCR-20. *Journal of Personality Disorders, 24*(5), 551-580. doi: http://dx.doi.org/10.1521/pedi.2010.24.5.551.

**Guy, L. S., Packer, I. K., & Warnken, W. (2012). Assessing risk of violence using structured professional judgment guidelines. *Journal of Forensic Psychology Practice, 12*(3), 270-283. doi: 10.1080/15228932.2012.674471.

**Habermeyer, E., Gairing, S., & Lau, S. (2010). Criminal risk

assessment: Does psychopathology still play a role? *Forensische Psychiatrie, Psychologie, Kriminologie, 4*(4), 258-263. doi: http://dx.doi.org/10.1007/s11757-010-0080-5.

Hanson, R.K, and Thornton, D. (2000). Improving Risk Assessment for Sexual Offenders: A Comparison of Three Actuarial Scales. *Law and Human Behaviour, 24*(1), 119-136. doi: 10.1.1.456.3059.

Hare, R. (2003). *Manual for the Revised Psychopathy Checklist (2nd ed.).* Toronto, ON, Canada: Multi-Health Systems.

Harris, A., Phenix, A., Hanson, R., & Thornton, D. (2003). *Static-99 coding rules revised*. Ottawa, Canada: Public Safety and Emergency Preparedness Canada.

Hart, S. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*(1), 121-137. doi: 10.1111/j.2044-8333.1998.tb00354.x.

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

Hart, S., Webster, C., & Menzies, R. (1993). *A note on portraying the accuracy of violence predictions. Law and human behaviour, 17*(6), 695-700. doi: 10.1007/BF01044690.

**Hartvig, P., Alfarnes, S., Ostberg, B., Skjonberg, M., & Moger, T. A. (2006). Brief checklists for assessing violence risk among patients discharged from acute psychiatric facilities: a preliminary study.

Nordic Journal of Psychiatry, 60(3), 243-248. doi: 10.1080/08039480600780532.

**Hartvig, P., Roaldset, J., Moger, T., Ostberg, B., & Bjorkly, S. (2011). The first step in the validation of a new screen for violence risk in acute psychiatry: The inpatient context. *European Psychiatry, 26*(2), 92-99. doi: http://dx.doi.org/10.1016/j.eurpsy.2010.01.003.

**Heilbrun, K., & Holliday, S. B. (2013). Psychological assessment in forensic contexts *APA handbook of testing and assessment in psychology, Vol 2: Testing and assessment in clinical and counseling psychology* (pp. 271-284). Washington, DC: American Psychological Association; US.

Heilbrun, K., Yasuhara, K., & Shar, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. Otto & K. Douglas (Ed.). *The Oxford Handbook of Clinical Psychology*. New York: Oxford University Press.

**Hill, A., Rettenberger, M., Habermann, N., Berner, W., Eher, R., & Briken, P. (2012). The utility of risk assessment instruments for the prediction of recidivism in sexual homicide perpetrators. *Journal of Interpersonal Violence, 27*(18), 3553-3578. doi: http://dx.doi.org/10.1177/0886260512447570.

**Hilterman, E. L., Philipse, M. W., & de Graaf, N. D. (2011). Assessment of offending during leave: Development of the Leave Risk Assessment in a sample of Dutch forensic psychiatric patients. *The International Journal of Forensic Mental Health, 10*(3), 233-243. doi: http://dx.doi.org/10.1080/14999013.2011.598601.

**Ho, H., Thomson, L., & Darjee, R. (2009). Violence risk assessment: the use of the PCL-SV, HCR-20, and VRAG to predict violence in mentally disordered offenders discharged from a medium secure unit in Scotland. *Journal of Forensic Psychiatry & Psychology, 20*(4), 523-541. doi: 10.1080/14789940802638358.

*Ho, R. M. Y., Lau, J. S. F., Cheung, H. H. K., Lai, T. T. S., Tam, V. F. L., Chan, W. L., . . . Yan, C. K. (2013). Evaluation of a model of violence risk assessment (HCR-20) among adult patients discharged from a gazetted psychiatric hospital in Hong Kong. *Journal of Forensic Psychiatry & Psychology, 24*(4), 479-495. doi: 10.1080/14789949.2013.809467.

**Hodgins, S., Tengstrom, A., Eriksson, A., Osterman, R., Kronstrand, R., Eaves, D., . . . Vartiainen, H. (2007). A Multisite Study of Community Treatment Programs for Mentally Ill Offenders with Major Mental Disorders: Design, Measures, and the Forensic Sample. *Criminal Justice and Behavior, 34*(2), 211-228. doi: http://dx.doi.org/10.1177/0093854806291248.

**Horstead, A., & Cree, A. (2013). Achieving transparency in forensic risk assessment: a multimodal approach. *Advances in Psychiatric Treatment, 19*(5), 351-357. doi: 10.1192/apt.bp.112.010645.

**Hurducas, C. C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of surveys. *The International Journal of Forensic Mental Health, 13*(3), 181-192. doi: http://dx.doi.org/10.1080/14999013.2014.942923.

**Inett, A., Wright, G., Roberts, L., & Sheeran, A. (2014). Predictive validity of the START with intellectually disabled offenders. *Journal*

of Forensic Practice, 16(1), 78-88. doi: 10.1108/JFP-12-2012-0029.

**Izycky, A., Braham, L., Williams, L., & Hogue, T. (2010). Goal attainment scaling: usefulness of a tool to measure risk in violent mentally disordered offenders. *British Journal of Forensic Practice, 12*(2), 14-22. doi: 10.5042/bjfp.2010.0182.

Jacobson, N. & Truax, P. (1991). Clinical Significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59(1), 12-19. doi: 10.1037/0022-006X.59.1.12.

Jessor, R., Turbin, M., Costa, F., Dong, Q., Zhang, H., & Wang, C. (2003). Adolescent problem behaviour in China and the United States: A cross-national study of psychosocial protective factors. *Journal of Research on Adolescene, 13*(3), 329-360. doi: 10.1111/1532-7795.1303004/.

**Jovanovic, A. A., Tosevski, D. L., Ivkovic, M., Damjanovic, A., & Gasic, M. J. (2009). Predicting violence in veterans with posttraumatic stress disorder. *Vojnosanitetski Pregled, 66*(1), 13-21. doi: 10.2298/VSP0901013J.

**Jung, S., Ledi, D., & Daniels, M. K. (2013). Evaluating the concurrent validity of the HCR-20 scales. *Journal of Risk Research, 16*(6), 697-711. doi: http://dx.doi.org/10.1080/13669877.2012.726247.

Kippax, S. & Crawford, J. (1993). Flaws in the Theory of Reasoned Action. In D. Terry, C. Galloios, & M. McCamish (eds.). *The Theory of Reasoned Action: Its application to AIDS-Preventive Behaviour* (pp. 253-269). UK: Pergamon, Press.

**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*(4), 129-132. doi: 10.1192/pb.bp.108.019810.

**Kötter, S., von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20V3 in Germany. *International Journal of Forensic Mental Health, 13*(2), 122-129. doi: 10.1080/14999013.2014.911784.

Kraemer, H., Kazdin, A., Offord, D., Kessler, R., Jenson, P. & Kupfer, D. (1997). Achieves of General Psychiatry, 54(4), 337-343. doi: 10.1001/archpsyc.1997.01830160065009.

Kreis, M. & Cooke, D. (2012). The Manifestation of Psychopathic Traits in Women: An Exploration Using Case Examples. *International Journal of Forensic Mental Health, 11*(4), 267-279. doi: 10.1080/14999013.2012.746755.

**Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*(4), 471-489. doi: http://dx.doi.org/10.1177/009385480102800405.

Kropp, P., Hart, S., Webster, C., & Eaves, D. (1999). *Spousal Assault Risk Assessment guide (SARA)*. Toronto: Multi-Health Systems.

**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., Dahlberg, L., Mercy, J., Zwi, A., & Lozano, R. (Eds.). (2002). *World report on violence and health.* Geneva: World Health Organisation.

**Langton, C. M. (2007). Assessment implications of "What Works" research for dangerous and severe personality disorder (DSPD) service evaluation. *Psychology Crime & Law, 13*(1), 97-111. doi: 10.1080/1068160600869841.

*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 & Law, 15*(7), 635-659. doi: 10.1080/10683160802516224.

Laub, J. & Lauritsen, J. (1994). The Precursors of Chronic Offending Across the Life Course. *Federal Probation, 3*, 51-57. doi: unavailable.

**Lewis, A. H., & Webster, C. D. (2004). General instruments for risk assessment. *Current Opinion in Psychiatry, 17*(5), 401-405. doi: 10.1097/01.yco.0000139978.68060.5a.

**Lindsay, W. R., Hastings, R. P., & Beail, N. (2013). Why Do Some People with Intellectual Disability Engage in Offending Behaviour and What Can We Do About It? Editorial. *Journal of Applied Research in Intellectual Disabilities, 26*(5), 351-356. doi: 10.1111/jar.12042.

**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. *International Journal of Offender Therapy and Comparative Criminology, 52*(1), 90-111. doi: http://dx.doi.org/10.1177/0306624X07308111.

**Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*(2), 409-443. doi: http://dx.doi.org/10.1037/1076-8971.7.2.409.

**Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., & Coid, J. W. (2011). A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology, 27*(4), 547-573. doi: http://dx.doi.org/10.1007/s10940-011-9137-7.

**Logan, C. (2014). The HCR-20 Version 3: A Case Study in Risk Formulation. *International Journal of Forensic Mental Health, 13*(2), 172-180. doi: 10.1080/14999013.2014.906516.

Lösel, F. (2001). Nonviolence: Protective Factors. In N. Smelser & P. Baltes (Eds.). *International Encyclopaedia of the Social and Behavioural Sciences*. Oxford: Pergamon Press.

Lösel, F. & Farrington, D. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine, 43*(1), s8-s23. doi: 10.1016/j.amepre.2012.04.029.

Lowry, R. (2001-2005). Clinical Calculator 1 [computer software]. Available from http://vassarstats.net/clin1.html.

**Lussier, P., Verdun-Jones, S., Deslauriers-Varin, N., Nicholls, T., & Brink, J. (2010). Chronic Violent Patients in an Inpatient Psychiatric Hospital: Prevalence, Description, and Identification.

Criminal Justice and Behavior, 37(1), 5-28. doi: http://dx.doi.org/10.1177/0093854809347738.

**Lv, Y., Han, C.-B., & Wang, X.-P. (2013). The reliability and validity of the Historical, Clinical, Risk management-Chinese version. *Chinese Journal of Clinical Psychology, 21*(6), 984-987. Retrieved from http://caod.oriprobe.com/articles/41010268/The_Reliability_and_Validity_of_the_Historical__Clinical__Risk_Managem.htm.

**Macpherson, G. J., & Kevan, I.-M. (2004). Predictive validity of the HCR-20 Violence Risk Assessment Scheme within a maximum security special hospital. *Issues in Forensic Psychology, 5*, 62-80. Retrieved from http://shop.bps.org.uk/publications/publication-by-series/issues-in-forensic-psychology/issues-in-forensic-psychology-no-5-risk-assessment-and-management.html.

Maden, A. (2001). Practical application of structured risk assessment. *British Journal of Psychiatry, 178*, 479. doi: 10.1192/bjp.178.5.479.

**Mann, B., Matias, E., & Allen, J. (2014). Recovery in forensic services: facing the challenge. *Advances in Psychiatric Treatment, 20*(2), 125-131. doi: 10.1192/apt.bp.113.011403.

**Martinaki, S., Tsopelas, C., Ploumpidis, D., Douzenis, A., Tzavara, H., Skapinakis, P., & Mavreas, V. (2013). Evaluation of dangerousness of Greek mental patients. *Psychiatrike = Psychiatriki, 24*(3), 185-196. Retrieved from http://www.pubfacts.com/detail/24185085/Evaluation-of-dangerousness-of-Greek-mental-patients.

**Matiasko, M. (2010). Possibilities of violent acts risk assessment of

offender with mental disorder. [Czech]. [Moznosti predikcie nasilneho jednania u pachatel'a s dusevnou poruchou.]. *Psychiatrie, 14*(1), 12-17. doi: unavailable.

McConnaughy, E., DiClemente, C., Prochaska, J. & Velicer, W. (1989). Stage of change in psychotherapy: A follow-up report. *Psychotherapy: Theory, Research, Practice, Training, 26*(4), 494-503. doi: 10.1037/h0085468.

McConnaughy, E., Prochaska, J.,& Velicer, w. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, research, and practice, 20*(3), 368-375. doi: 10.1037/h0090198.

**McDermott, B. E., Dualan, I. V., & Scott, C. L. (2011). The predictive ability of the Classification of Violence Risk (COVR) in a forensic psychiatric hospital. *Psychiatric Services, 62*(4), 430-433. doi: http://dx.doi.org/10.1176/appi.ps.62.4.430.

*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 & the Law, 26*(6), 759-777. doi: http://dx.doi.org/10.1002/bsl.842.

**McDougall, C., Pearson, D. A. S., Willoughby, H., & Bowles, R. A. (2013). Evaluation of the ADViSOR project: Cross-situational behaviour monitoring of high-risk offenders in prison and the community. *Legal and Criminological Psychology, 18*(2), 205-228. doi: 10.1111/j.2044-8333.2012.02059.x.

**McKeown, A. (2010). Female offenders: Assessment of risk in forensic

settings. *Aggression and Violent Behavior, 15*(6), 422-429. doi: http://dx.doi.org/10.1016/j.avb.2010.07.004.

McMurran, M., Theodosi, E., & Sellen, J. (2006). Measuring engagement in therapy and motivation to change in adult prisoners: a brief report, *Criminal Behaviour and Mental Heath, 16*(2), 124-129. doi: 10.1002/cbm.625.

**McNiel, D. E., Gregory, A. L., Lam, J. N., Binder, R. L., & Sullivan, G. R. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology, 71*(5), 945-953. doi: http://dx.doi.org/10.1037/0022-006X.71.5.945.

**Megargee, E. I. (2013). Psychological assessment in correctional settings. In J. Graham & J. Naglieri (Eds), *Handbook of psychology* (pp. 394-424). Hoboken, NJ: John Wiley & Sons Inc; US.

*Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Muller-Isberner, R., Hodgins, S. (2013). Using the HCR-20 to predict aggressive behavior among men with schizophrenia living in the community: Accuracy of prediction, general and forensic settings, and dynamic risk factors. *The International Journal of Forensic Mental Health, 12*(1), 1-13. doi: http://dx.doi.org/10.1080/14999013.2012.760182.

Miller, H. (2006). A dynamic assessment of offender risk, needs, and strengths in a sample of general offenders. *Behavioural Sciences and the Law, 24*, 767-782. doi:10.1002/bsl.728.

**Mills, J. F., & Gray, A. L. (2013). Two-tiered violence risk estimates: A validation study of an integrated-actuarial risk assessment instrument. *Psychological Services, 10*(4), 361-371. doi:

http://dx.doi.org/10.1037/a0032608.

**Mills, J. F., Kroner, D. F., & Morgan, R. D. (2011). *Clinician's guide to violence risk assessment*. New York, NY: Guilford Press; US.

**Mills, J. F., Kroner, D. G., & Hemmati, T. (2007). The validity of violence risk estimates: An issue of item performance. *Psychological Services, 4*(1), 1-12. doi: http://dx.doi.org/10.1037/1541-1559.4.1.1.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Empidemiology, 62*(10), 1006-1012. doi: 10.7326/0003-4819-151-4-200908180-00135.

**Mokros, A., Stadtland, C., Osterheider, M., & Nedopil, N. (2010). Assessment of risk for violent recidivism through multivariate Bayesian classification. *Psychology, Public Policy, and Law, 16*(4), 418-450. doi: http://dx.doi.org/10.1037/a0021312.

Moons, P., Boriau, M., & Ferdinande, P. (2008). Self-extubation risk assessment tool: predictive validity in a real-life setting. *Nursing in Critical Care, 13*(6), 310-314. doi: http://dx.doi.org/10.1111/j.1478-5153.2008.00305.x.

**Morrissey, C., Beeley, C., & Milton, J. (2014). Longitudinal HCR-20 scores in a high-secure psychiatric hospital. *Criminal Behaviour and Mental Health, 24*(3), 169-180. doi: 10.1002/cbm.1893.

Mossman, D. (2004). Understanding prediction instruments. In R. Simon & L. Gold (Eds.). The American Psychiatric Publishing

Textbook of Forensic Psychiatry (pp. 501-523). Washington, DC: American Psychiatric Publishing, Inc.

Mossman, D. (2013). Evaluating Risk Assessments Using Receiver Operating Characteristic Analysis: Rationale, Advantages, Insights, and Limitations. *Behavioural Sciences and the Law, 31*(1), 23-39. doi: 10.1002/bsl.2050.

**Mudde, N., Nijman, H., van der Hulst, W., & van den Bout, J. (2011). Predicting aggression during the treatment of forensic psychiatric patients by means of the HCR-20. *Tijdschrift voor Psychiatrie, 53*(10), 705-713. Retrieved from http://www.tijdschriftvoorpsychiatrie.nl/en/issues/446/articles/9212.

**Muller-Isberner, R., Webster, C. D., & Gretenkord, L. (2007). Measuring progress in hospital order treatment: Relationship between levels of security and C and R scores of the HCR-20. *The International Journal of Forensic Mental Health, 6*(2), 113-121. doi: http://dx.doi.org/10.1080/14999013.2007.10471256.

**Murphy, D. (2007). Theory of mind functioning in mentally disordered offenders detained in high security psychiatric care: its relationship to clinical outcome, need and risk. *Criminal Behaviour & Mental Health, 17*(5), 300-311. doi: 10.1002/cbm.664.

**Nanayakkara, V., O'Driscoll, C., & Allnutt, S. (2012). HCR-20 risk profiles of patients referred to forensic mental health services in New South Wales. *Psychiatry, Psychology and Law, 19*(5), 711-719. doi: http://dx.doi.org/10.1080/13218719.2011.615819.

National Institute of Clinical Excellence (2005). *Violence. The short term management of disturbed/violent behaviour in inpatient psychiatric settings and emergency departments*. Clinical Practice Guidelines. Retrieved from http://www.nice.org.uk/nicemedia/pdf/cg025fullguideline.pdf.

**Nedopil, N. (2009). Gender specific and structured approach to violence risk assessment - The Munich Prognosis project (MPP). *European Psychiatry, 24*(1), S148. doi: 10.1016/S0924-9338(09)70381-5.

*Neves, A. C., Goncalves, R. A., & Palma-Oliveira, J. M. (2011). Assessing risk for violent and general recidivism: A study of the HCR-20 and the PCL-R with a non-clinical sample of Portuguese offenders. *The International Journal of Forensic Mental Health, 10*(2), 137-149. doi: http://dx.doi.org/10.1080/14999013.2011.577290.

NHS Commissioning Board. (2013). *NHS Standard Contract for Medium and Low Secure Mental Health Services (adults)*. Retrieved from http://www.england.nhs.uk/wp-content/uploads/2013/06/c03-med-low-sec-mh.pdf.

**Nicholls, T. L. (2004). Violence risk assessments with female NCRMD acquittees: Validity of the HCR-20 and PCL-SV. *Dissertation Abstracts International, 64,* 8B.

*Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., & Martin, M.-L. (2006). The Short-Term Assessment of Risk and Treatability (START): A Prospective Validation Study in a Forensic Psychiatric Sample. *Assessment, 13*(3), 313-327. doi:

http://dx.doi.org/10.1177/1073191106290559.

**Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Research report - Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL : SV, and VSC. *Behavioral Sciences & the Law, 22*(1), 127-158. doi: 10.1002/bsl.579.

*Nicholls, T. L., Petersen, K. L., Brink, J., & Webster, C. (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. doi: 10.1080/14999013.2011.600234.

**Nieberding, R. J., Moore, J., & Dematatis, A. P. (2002). Psychological assessment of forensic psychiatric outpatients. *International Journal of Offender Therapy and Comparative Criminology, 46*(3), 350-363. doi: http://dx.doi.org/10.1177/03024X02046003008.

**Nilsson, T., Wallinius, M., Gustavson, C., Anckarsater, H., & Kerekes, N. (2011). Violent recidivism: A long-time follow-up study of mentally disordered offenders. *PLoS ONE Vol 6(10), Oct 2011, ArtID e25768, 6*(10). doi: http://dx.doi.org/10.1371/journal.pone.0025768.

*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'Shea, L, & Dickens, G. L. (2014). Short-Term Assessment of Risk

and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*(3), 990-1002. doi: http://dx.doi.org/10.1037/a0036794.

**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*(2), 255-270. doi: http://dx.doi.org/10.1016/j.avb.2012.11.016.

*O'Shea, L. E., Picchioni, M. M., Mason, F. L., Sugarman, P. A., & Dickens, G. L. (2014a). 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.

**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*, No Pagination Specified. doi: http://dx.doi.org/10.1016/j.comppsych.2014.07.010.

**O'Shea, L., Picchioni, M., McCarthy, J., Mason, F., & Dickens, G. (2015). Predictive validity of the hcr-20 for inpatient aggression: The effect of intellectual disability on accuracy. *Journal of Intellectual Disability Research*, No Pagination Specified. doi: http://dx.doi.org/10.1111/jir.12184.

**Ogloff, J. R., & Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An Instrument to Assess Risk for Imminent

Aggression in Psychiatric Inpatients. *Behavioral Sciences & the Law, 24*(6), 799-813. doi: http://dx.doi.org/10.1002/bsl.741.

**Olsson, H., Strand, S., Kristiansen, L., Sjoling, 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.

*Pedersen, L., Rasmussen, K., & Elsass, P. (2010). Risk assessment: The value of structured professional judgments. *The International Journal of Forensic Mental Health, 9*(2), 74-81. doi: http://dx.doi.org/10.1080/14999013.2010.499556.

*Pedersen, L., Ramussen, K., & Elsass, P. (2012). HCR-20 violence risk assessments as a guide for treating and managing violence risk in a forensic psychiatric setting. *Psychology, Crime & Law, 18*(8), 733-743. doi: http://dx.doi.org/10.1080/1068316X.2010.548814.

**Penney, S. R., McMaster, R., & Wilkie, T. (2014). Multirater reliability of the historical, clinical, and risk management-20. *Assessment, 21*(1), 15-27. doi: http://dx.doi.org/10.1177/1073191113514107.

**Pham, T., Chevrier, I., Nioche, A., Ducro, C., & Reveillere, C. (2005). Psychopathy, risk assessment and rehabilitation. *Annales Medico-Psychologiques, 163*(10), 878-881. doi: http://dx.doi.org/10.1016/j.amp.2005.10.016.

**Pham, T. H., Ducro, C., Marghem, B., & Reveillere, C. (2005). Prediction of recidivism among prison inmates and forensic patients in Belgium. *Annales Medico-Psychologiques, 163*(10), 842-845. doi: 10.1016/j.amp.2005.09.013.

**Philipse, M. W. G., Koeter, M. W. J., Van Der Staak, C. P. F., & Van Den Brink, W. (2005). Reliability and Discriminant Validity of Dynamic Reoffending Risk Indicators in Forensic Clinical Practice. *Criminal Justice and Behavior, 32*(6), 643-664. doi: http://dx.doi.org/10.1177/0093854805279946.

**Pillay, S. M., Oliver, B., Butler, L., & Kennedy, H. G. (2008). Risk stratification and the care pathway. *Irish Journal of Psychological Medicine, 25*(4), 123-127. Retrieved from http://www.researchgate.net/profile/Harry_G_Kennedy/publication/216054421_Risk_stratification_and_the_care_pathway/links/09e4150106259df3a9000000.pdf.

Pollock, P., Stowell-Smith, M. & Gopfert, M. (Eds.). (2006). *Cognitive Analytic Therapy for Offenders: A New Approach to Forensic Psychotherapy.* London: Routledge.

**Polvi, N. H. (2001). The prediction of violence in pre-trial forensic patients: The relative efficacy of statistical versus clinical predictions of dangerousness. *Dissertation Abstracts International, 61, 7*B.

Prochaska, J. & DiClemente, C. (1986). Toward a comprehensive model of change. *Treating Addictive Behaviours, Applied Clinical Psychology, 13*(1), 3-27. doi: 10.1007/978-1-4613-2191-0_1.

Prochaska, J. & Velicer, W. (1997). The transtheoretical model of health behaviour change. *American Journal of Health Promotion, 12*(1), 38-38. doi: 10.4278/0890-1171-12.1.38.

**Pyott, J. (2005). Use of HCR-20 in routine psychiatric practice. *Psychiatric Bulletin, 29*(9), 352. doi: http://dx.doi.org/10.1192/pb.29.9.352.

*Quinn, R., Miles, H., & Kinane, C. (2013). The validity of the Short-Term Assessment of Risk and Treatability (START) in a UK medium secure forensic mental health service. *The International Journal of Forensic Mental Health, 12*(3), 215-224. doi: http://dx.doi.org/10.1080/14999013.2013.832714.

Quinsey, V., Harris, G., Rice, M., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Quinsey, V., Harris, G., Rice, M. & Cormier, C. (2006). Violence Risk Appraisal Guide (VRAG). In V. Quinsey, G. Harris, M. Rice, & C. Cormier, *Violent Offenders: Apppraising and managing risk (2nd Ed.)*. Washington DC: American Psychological Association.

Rapp, C. & Goscha, R. J., *The Strengths Model: Case Management with People with Psychiatric Disabilities*. New York: Oxford University Press.

**Reimann, B. J., & Nussbaum, D. (2011). Predicting seclusion in a medium secure forensic inpatient setting. *The International Journal of Forensic Mental Health, 10*(2), 150-158. doi: http://dx.doi.org/10.1080/14999013.2011.578299.

**Reynolds, L. M., Jones, J. C., Davies, J. P., Freeth, D., & Heyman, B. (2014). Playing the game: Service users' management of risk status in a UK medium secure forensic mental health service. *Health, Risk & Society, 16*(3), 199-209. doi:

http://dx.doi.org/10.1080/13698575.2014.911820.

Rice, M. & Harris, G. (1995). Violent recidivism: assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*(5), 737-748. doi: 10.1037/0022-006X.63.5.737.

Rice, M. & Harris. G. (2005). Comparing effects sizes in follow up studies: ROC area, Cohen's d, and r. *Law and Human Behaviour, 29*(5), 615-620. doi: 10.1007/s10979-005-6832-7.

**Rice, M. E., & Harris, G. T. (2013). Psychopathy and violent recidivism *Handbook on psychopathy and law* (pp. 231-249). New York, NY: Oxford University Press; US.

**Rizzo, M., & Smith, S. (2012). Risk assessment and risk management among severely mentally ill patients requiring hospitalisation. *Journal of Forensic Psychiatry & Psychology, 23*(4), 497-509. doi: 10.1080/14789949.2012.704638.

**Roaldset, J. O., Hartvig, P., Linaker, O. M., & Bjorkly, S. (2012). A multifaceted model for risk assessment of violent behaviour in acutely admitted psychiatric patients. *Psychiatry Research, 200*(2-3), 773-778. doi: http://dx.doi.org/10.1016/j.psychres.2012.04.038.

**Roberts, A. D. L., & Coid, J. W. (2007). Psychopathy and offending behaviour: Findings from the national survey of prisoners in England and Wales. *Journal of Forensic Psychiatry & Psychology, 18*(1), 23-43. doi: 10.1080/14789940601012698.

Robinson, O. (2012). A war of words. *The Psychologist*, *25*(2), 164–166. doi: 10.1080/14999013.2012.746755.

Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behaviour, 24*(5), 595-605. doi:10.1023/A:1005575113507.

**Rogers, R., & Jackson, R. L. (2005). Sexually Violent Predators: The Risky Enterprise of Risk Assessment. *Journal of the American Academy of Psychiatry and the Law, 33*(4), 523-528. Retrieved from http://www.jaapl.org/content/33/4/523.full.pdf.

**Rossegger, A., Frank, U., Elbert, T., Fries, D., & Endrass, J. (2010). Recidivism following release from prison in Switzerland: Validity of HCR-20. *Schweizer Archiv fur Neurologie und Psychiatrie, 161*(7), 254-259. Retrieved from http://www.sanp.ch/docs/2010/2010-07/2010-07-063.pdf.

**Rufino, K. A., Boccaccini, M. T., & Guy, L. S. (2011). Scoring Subjectivity and Item Performance on Measures Used to Assess Violence Risk: The PCL-R and HCR-20 as Exemplars. *Assessment, 18*(4), 453-463. doi: 10.1177/1073191110378482.

Rumgay, J. (2004). Scripts for safer survival: Pathways out of female crime. *The Howard Journal, 43*(4), 405-419. doi: 10.1111/j.1468-2311.2004.00338.x.

Rutter, M. (1985). Resilience in the face of adversity. *British Journal of Psychiatry, 147*(6), 598-611. doi: 10.1192/bjp.147.6.598.

Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78. doi: 10.1037/0003-066X.55.1.68.

**Schaap, G., Lammers, S., & de Vogel, V. (2009). Risk assessment in

female forensic psychiatric patients: a quasi-prospective study into the validity of the HCR-20 and PCL-R. *Journal of Forensic Psychiatry & Psychology, 20*(3), 354-365. doi: 10.1080/14789940802542873.

**Selenius, H., Hellstrom, A., & Belfrage, H. (2011). Aggression and risk of future violence in forensic psychiatric patients with and without dyslexia. *Dyslexia, 17*(2), 201-206. doi: 10.1002/dys.425.

**Sevilla-Sanchez, D., Espaulella-Panicot, J., de Andres-Lazaro, A. M., Torres-Allezpuz, R., Soldevila-Llagostera, M., & Codina-Jane, C. (2012). Potentially inappropriate medication on admission to a medium-stay unit according to STOPP and START criteria. *Revista Espanola de Geriatria y Gerontologia, 47*(4), 155-157. doi: 10.1016/j.regg.2012.02.013.

**Sinani, L., Kola, V., Cenko, F., Elezi, F., Balaj, K., Saraci, J., . . . Gjolena, G. (2013). Assessment and treatment of patients with violent behaviors in emergency psychiatry. The HCR-20 violence risk assessment scheme. *European Psychiatry, 28*(1), 1. doi: 10.1016/S0924-9338(13)77013-5.

**Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., . . . Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *The International Journal of Forensic Mental Health, 13*(3), 193-206. doi: http://dx.doi.org/10.1080/14999013.2014.922141.

** Singh, J. P., Fazel, S., Gueorguieva, R., & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk

assessment instruments. *British Journal of Psychiatry, 204*(3), 180-187. doi: 10.1192/bjp.bp.113.131938.

**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. doi: http://dx.doi.org/10.1016/j.cpr.2010.11.009.

**Singh, J. P., Serper, M., Reinharth, J., & Fazel, S. (2011). Structured assessment of violence risk in schizophrenia and other psychiatric disorders: A systematic review of the validity, reliability, and item content of 10 available instruments. *Schizophrenia Bulletin, 37*(Suppl 5), 899-912. doi: http://dx.doi.org/10.1093/schbul/sbr093.

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*(2), 179 –183. doi:10.1080/14999013.2002.10471172.

**Skipworth, J. (2005). Rehabilitation in forensic psychiatry: Punishment or treatment? *Journal of Forensic Psychiatry and Psychology, 16*(1), 70-84. doi: 10.1080/14789940512331309830.

Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care, 18*(5), 377-382. doi: 10.1093/intqhc/mzl034.

Smith, S., Kelley, S., Rulesh, A., Sorman, K., & Edens, J. (2014). Adapting the HCR-20V3 for pre-trial settings. *International Journal*

of Forensic Mental Health, 13(2), 160-171. doi: 10.1080/14999013.2014.906520.

**Smith, H., & White, T. (2007). Feasibility of a structured risk assessment tool in general adult psychiatry admissions. *Psychiatric Bulletin, 31*(11), 418-420. doi: http://dx.doi.org/10.1192/pb.bp.106.012518.

**Smith, S. T., Kelley, S. E., Rulseh, A., Sörman, K., & Edens, J. F. (2014). Adapting the HCR-20V3 for Pre-trial Settings. *International Journal of Forensic Mental Health, 13*(2), 160-171. doi: 10.1080/14999013.2014.906520.

*Snowden, R., Gray, N. S., & Taylor, J. (2010). Risk assessment for future violence in individuals from an ethnic minority group. *The International Journal of Forensic Mental Health, 9*(2), 118-123. doi: http://dx.doi.org/10.1080/14999013.2010.501845.

**Snowden, R. J., Gray, N. S., Taylor, J., & MacCulloch, M. J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine, 37*(11), 1539-1549. doi: 10.1017/S0033291707000876.

**Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2006a). Evaluation of risk assessment instruments for sex offenders. *Der Nervenarzt, 77*(5), 587-595. doi: http://dx.doi.org/10.1007/s00115-005-1945-2.

**Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2006b). Evaluation of risk assessment instruments for sex offenders. [German]. [Ruckfallprognosen bei sexualstraftatern - Vergleich der pradiktiven validitat von prognoseinstrumenten.].

Nervenarzt, 77(5), 587-595. doi: 10.1007%2Fs00115-005-1945-2.

**Stadtland, C., & Nedopil, N. (2005). Psychiatric disorders and the prognosis for criminal recidivism. *Nervenarzt, 76*(11), 1402-11 doi: 10.1007/s00115-004-1808-2.

**Stalans, L. J., Yarnold, P. R., Seng, M., Olson, D. E., & Repp, M. (2004). Identifying three types of violent offenders and predicting violent recidivism while on probation: A classification tree analysis. *Law and Human Behavior, 28*(3), 253-271. doi: 10.1023/B:LAHU.0000029138.92866.af.

**Stanfill, M. L., O'Brien, S., & Viglione, D. J., Jr. (2014). Multimethod violence risk assessment *Multimethod clinical assessment* (pp. 379-402). New York, NY: Guilford Press; US.

Stattin, H., Romelsjo, A., & Stenbacka, M. (1997). Personal resources as modifiers of the risk for future criminality. *British Journal of Criminology, 37*(2), 198–223. doi: 10.1093/oxfordjournals.bjc.a014155.

**Strand, S., & Belfrage, H. (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime & Law, 7*(1), 71-79. doi: http://dx.doi.org/10.1080/10683160108401784.

Strub, D. & Douglas, K. (2009). *New version of the HCR-20 violence risk assessment scheme: Evaluation of a draft of the revised Historical scale.* Poster presented at the American-Psychology Law Society conference, San Antonio, Texas.

*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. *The International Journal of Forensic Mental Health, 13*(2), 148-159. doi: http://dx.doi.org/10.1080/14999013.2014.911785.

**Sturup, J., Karlberg, D., Fredriksson, B., Lihoff, T., & Kristiansson, M. (2015). Risk assessments and recidivism among a population-based group of Swedish offenders sentenced to life in prison. *Criminal Behaviour and Mental Health* (online). doi: 10.1002/cbm.1941.

**Sturup, J., Monahan, J., & Kristiansson, M. (2013). Violent behavior and gender of Swedish psychiatric patients: A prospective clinical study. *Psychiatric Services, 64*(7), 688-693. doi: http://dx.doi.org/10.1176/appi.ps.201200266.

**Stübner, S., GROß, G., & NEDOPIL, N. (2006). Inpatient risk management with mentally ill offenders: results of a survey on clinical decision-making about easing restrictions. *Criminal behaviour and mental health, 16*(2), 111-123. doi: 10.1002/cbm.619.

Sullivan, G., Wells, K., Mortgenstern, H. & Leake, B. (1995). Identifying modifiable risk factors for rehospitalisation: A case-control study of seriously mentally ill persons in Mississippi. *American Journal of Psychiatry, 152*(12), 1749-1756.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285-1293. doi:10.1126/science.3287615.

Szmukler, G., Everitt, B., & Leese, M. (2012). Risk assessment and receiver operating characteristic curves. *Psychological Medicine, 42*(5), 895–898. doi: 10.1017/S003329171100208X.

**Tardiff, K., & Hughes, D. M. (2011). Structural and clinical assessment of risk of violence *Handbook of forensic assessment: Psychological and psychiatric perspectives* (pp. 335-359). Hoboken, NJ: John Wiley & Sons Inc; US.

Taylor, R. (1999). *Predicting reconvictions for sexual and violent offences using the revised Offender Group Reconviction Scale. Home Office Research Findings, 104*. London, UK: Home Office.

**Telles, L. E. d. B., Day, V. P., Folino, J. O., & Taborda, J. G. V. (2009). Reliability of the Brazilian version of HCR-20 Assessing Risk for Violence. *Revista Brasileira de Psiquiatria, 31*(3), No Pagination Specified. doi: http://dx.doi.org/10.1590/S1516-44462009005000001.

**Tengstrom, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry, 55*(4), 243-249. doi: http://dx.doi.org/10.1080/080394801681019093.

**Teo, A. R., Holley, S. R., Leary, M., & McNiel, D. E. (2012). The relationship between level of training and accuracy of violence risk assessment. *Psychiatric Services, 63*(11), 1089-1094. doi: http://dx.doi.org/10.1176/appi.ps.201200019.

**Thomson, L., Davidson, M., Brett, C., Steele, J., & Darjee, R. (2008). Risk assessment in forensic patients with schizophrenia: The predictive validity of actuarial scales and symptom severity for offending and violence over 8 - 10 years. *The International Journal of Forensic Mental Health, 7*(2), 173-189. doi: http://dx.doi.org/10.1080/14999013.2008.9914413.

**Tiegreen, J. A. (2010). Comparison of dynamic versus static factors in risk assessment: Standardization of the violence clinical risk indicator. *Dissertation Abstracts International, 70,* 9B.

**Tozdan, S., Briken, P., Yoon, D., & von Franqué, F. (2014). Risk and Protective Factors Among Sexual Offenders: Relapse Prediction and Changes During Treatment. *Psychiatrische Praxis*. doi: 10.1055/s-0034-1387404.

*Troquete, N. A., van den Brink, R. H., Beintema, H., Mulder, T., van Os, T. W., Schoevers, R. A., & Wiersma, D. (2014). Predictive Validity of the Short-Term Assessment of Risk and Treatability for Violent Behavior in Outpatient Forensic Psychiatric Patients. *Psychological Assessment*, No Pagination Specified. doi: http://dx.doi.org/10.1037/a0038270.

Ullrich, S. & Coid, J. (2009). Protective factors for violence among released prisoners – effects over time and interactions with static risk. *Journal of Consulting and Clinical Psychology, 79*(3), 381-390. doi: 10.1037/a0023613.

**Ullrich, S., & Coid, J. (2011). Protective factors for violence among released prisoners—Effects over time and interactions with static risk. *Journal of Consulting and Clinical psychology, 79*(3), 381-390. doi: 10.1037/a0023613.

**Van Den Berg, J. W., & de Vogel, V. (2011). Risk assessment with intellectual disabled forensic psychiatric patients: How useful are risk assessment instruments?. [Dutch]. [Risicotaxatie bij forensisch psychiatrische patienten met een lichte verstandelijke handicap: Hoe bruikbaar zijn risicotaxatie-instrumenten?.]. *Tijdschrift voor*

Psychiatrie, 53(2), 83-93. Retrieved from http://www.tijdschriftvoorpsychiatrie.nl/en/issues/436/articles/8405.

**van den Brink, R. H. S., Hooijschuur, A., van Os, T., Savenije, W., & Wiersma, D. (2010). Routine Violence Risk Assessment in Community Forensic Mental Healthcare. *Behavioral Sciences & the Law, 28*(3), 396-410. doi: 10.1002/bsl.904.

**van den Broek, E., & de Vries Robbé, M. (2008). *The supplemental value of the SAPROF from a treatment perspective: A counterbalance to risk.* Paper presented at the eighth Conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Vergara, Norambuena, Ferrada, Slater, & Melo, 2008. StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics, 9*(265), 1-5. doi: 10.1186/1471-2105-9-265.

*Viljoen, S. (2014). *Examining the utility of strengths and protective factors in violence risk assessment measure*. (unpublished doctoral dissertation). Simon Fraser University, Canada.

**Viljoen, S., Nicholls, T., Greaves, C., Ruiter, C., & Brink, J. (2011). Resilience and successful community reintegration among female forensic psychiatric patients: A preliminary investigation. *Behavioral sciences & the law, 29*(5), 752-770. doi: 10.1002/bsl.1001.

**Vitacco, M. J., Erickson, S. K., Kurus, S., & Apple, B. N. (2012). THE ROLE OF THE VIOLENCE RISK APPRAISAL GUIDE AND HISTORICAL, CLINICAL, RISK-20 IN U.S. COURTS: A Case Law

Survey. *Psychology Public Policy and Law, 18*(3), 361-391. doi: 10.1037/a0025834.

**Vladejic, S. S., Vladejic, S. D., & Popovic, I. N. (2011). Assessing risk of violence: Are only forensic patients aggressive? *European Psychiatry, 26*(1), 798. doi: 10.1016/S0924-9338(11)72503-2.

*Vojt, G., Thomson, L. D., & Marshall, L. A. (2013). The predictive validity of the HCR-20 following clinical implementation: Does it work in practice? *Journal of Forensic Psychiatry & Psychology, 24*(3), 371-385. doi: http://dx.doi.org/10.1080/14789949.2013.800894.

Wærp, J. (2013). *Changes in repeated HCR-20 measurement.* Symposium presented at the annual conference of the International Association of Forensic Mental Health Services, Maastricht, Netherlands.

**Walters, G. D., Kroner, D. G., DeMatteo, D., & Locklair, B. R. (2014). The Impact of Base Rate Utilization and Clinical Experience on the Accuracy of Judgments Made with the HCR-20. *Journal of Forensic Psychology Practice, 14*(4), 288-301. doi: 10.1080/15228932.2014.941726.

Ward, T. & Stewart, C. A. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime, and Law, 9*(3), 125-143. doi: 10.1080/1068316031000112557.

**Warren, J. I., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., & Van Patten, I. (2005). Understanding the risk factors for violence and criminality in women: The concurrent validity of the PCL-R and HCR-20. *International Journal of Law and*

Psychiatry, 28(3), 269-289. doi: http://dx.doi.org/10.1016/j.ijlp.2003.09.012.

**Watt, A., Topping-Morris, B., Rogers, P., Doyle, M., & Mason, T. (2003). Pre-admission nursing assessment in a Welsh Medium Secure Unit (1991-2000): Part 2-comparison of traditional nursing assessment with the HCR-20 risk assessment tool. *International Journal of Nursing Studies, 40*(6), 657-662. doi: 10.1016/S0020-7489(03)00017-8.

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

Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk (Version 1).* Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University, and Forensic Psychiatric Services Commission of British Columbia.

Webster, C., Martin, M., Brink, J., Nicolls, T., & Desmarais, S. (2004). *Short Term Assessment of Risk and Treatability (START). Clinical guide for evaluation and recovery. Version 1.1.* Ontario: St Joseph's Healthcare Hamilton.

**Webster, C. D., Muller-Isberner, R., & Fransson, G. (2002). Violence risk assessment: Using structured clinical guides professionally. *The International Journal of Forensic Mental Health, 1*(2), 185-193. doi: http://dx.doi.org/10.1080/14999013.2002.10471173.

**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 & the law, 24*(6), 747-766. doi: 10.1002/bsl.737.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition.* San Antonio: Pearson.

*Whittington, R., Bjorngaard, 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.

Whittington, R., Hockenhull, J., McGuire, J., Leitner, M., Barr, W., Cherry, M., Flentje, R., Quinn, B., Dundar, Y., & Dickson, R. (2013). A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002-8. *Health Technology Assessment, 17*(50), 1-128. doi: 10.3310/hta17500.

Whittington, R. & Logan, C. (2011). *Self harm and violence: towards best practice in managing risk in mental health services.* Chichester: Wiley Blackwell.

*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). *The International Journal of Forensic Mental Health, 9*(4), 282-293. doi: http://dx.doi.org/10.1080/14999013.2010.534694.

*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. doi: http://dx.doi.org/10.1037/lhb0000025.

Wong, S. C. P. & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: a treatment friendly violence risk assessment tool. *Psychology, Public Policy and Law, 12*(3), 279 -309*. doi:* 10.1037/1076-8971.12.3.279.

**Yao, X., Li, Z., Arthur, D., Hu, L., & Cheng, G. (2012). The application of a violence risk assessment tool among Chinese psychiatric service users: A preliminary study. *Journal of Psychiatric and Mental Health Nursing, 19*(5), 438-445. doi: http://dx.doi.org/10.1111/j.1365-2850.2011.01821.x.

Yin, K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

**Yoon, D., Spehr, A., & Briken, P. (2011). Structured assessment of protective factors: A German pilot study in sex offenders. *Journal of Forensic Psychiatry & Psychology, 22*(6), 834-844. doi: http://dx.doi.org/10.1080/14789949.2011.617831.

* included in systematic review

** excluded from systematic review

Appendices

Appendix 3.A: Search syntax

EBSCOhost: Cumulative Index to Nursing and Allied Health Literature (CINAHL), completed 09th March 2015

S1. "short term assessment of risk and treatability" Boolean/Phrase (3)

S2. "START" Boolean/Phrase (13,440)

S3. "structured assessment of protective factors" Boolean/Phrase (0)

S4. "SAPROF" Boolean/Phrase (0)

S5. (MH "Predictive Validity") OR "predictive validity" OR (MH "Reliability and Validity") OR (MH "Instrument Validation") Boolean/Phrase (33,940)

S6. (MH "Data Analysis, Statistical") OR "statistical validity" Boolean/Phrase (39,319)

S7. "measurement" Boolean/Phrase (86,822)

S8. (MH "Data Analysis, Statistical") OR "statistical analysis" Boolean/Phrase (46,187)

S9. "statistical measurement" Boolean/Phrase (5)

S10. (MH "Validity") OR "test validity" Boolean/Phrase (8.292)

S11. "test reliability" OR (MH "Test-Retest Reliability") Boolean/Phrase (13,098)

S12. "specificity" OR (MH "Sensitivity and Specificity") Boolean/Phrase (58,914)

S13. "sensitivity" Boolean/Phrase (93,914)

S14. "accuracy" Boolean/Phrase (24,892)

S15. (MH "ROC Curve") OR "area under the curve" Boolean/Phrase (15,533)

S16. "area under curve" Boolean/Phrase (265)

S17. "receiver operating characteristic" Boolean/Phrase (3,497)

S18. "AUC" Boolean/Phrase (3,950)

S19. "ROC" Boolean/Phrase (14,452)

S20. "mentally ill offender*" OR (MH "Mentally Ill Offenders") Boolean/Phrase (1,535)

S21. "mentally disordered offender*" Boolean/Phrase (126)

S22. "MDO" Boolean/Phrase (10)

S23. "offender*" Boolean/Phrase (8,491)

S24. "violent offender*" Boolean/Phrase (93)

S25. "patient*" Boolean/Phrase (1,007,666)

S26. "violent patient*" Boolean/Phrase (204)

S27. (MH "Inpatients") OR "inpatient*" Boolean/Phrase (75,115)

S28. "violent inpatient*" Boolean/Phrase (3)

S29. "service user*" Boolean/Phrase (3,599)

S30. "violent service user*" Boolean/Phrase (1)

S31. "client*" Boolean/Phrase (32,688)

S32. "violent client*" Boolean/Phrase (14)

S33. "forensic psychiatric patient*" Boolean/Phrase (43)

S34. (MH "Psychiatric Patients") OR "psychiatric patient*" Boolean/Phrase (9,498)

S35. "violent psychiatric patient*" Boolean/Phrase (3)

S36. (MH "Public Offenders") Boolean/Phrase (2,724)

S37. "criminal*" Boolean/Phrase (4,455)

S38. "prisoner*" OR (MH "Prisoners") Boolean/Phrase (6,403)

S39. S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 Boolean/Phrase (1,066,788)

S40. S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 Boolean/Phrase (279,362)

S41. "hcr-20" Boolean/Phrase (30)

S42. "HCR-20V3" Boolean/Phrase (1)

S43. "hcr-20 version 3" Boolean/Phrase (0)

S44. "historical clinical risk management" Boolean/Phrase (3)

S45. S1 OR S2 OR S3 OR S4 OR S41 OR S42 OR S43 OR S44 Boolean/Phrase (13,470)

S46. S39 AND S40 AND S45 Boolean/Phrase (790)

S47. S39 AND S40 AND S45 Published Date: 20000101-20151231 (720)

OVID: PsycINFO (1806 to March Week 1 2015), completed 09th March 2015

1. (short term assessment of risk and treatability).mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (56)

2. START.mp. (23.351)

3. structured assessment of protective factors.mp. (18)

4. SAPROF.mp. (10)

5. HCR-20.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (255)

6. HCR-20V3.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (4)

7. HCR-20 version 3.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (6)

8. historical clinical risk management.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (159)

9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 (23,693)

10. predictive validity.mp. or exp Statistical Validity/ (19,638)

11. statistical validity.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (14,933)

12. exp Measurement/ or exp "Predictability (Measurement)"/ or exp Statistical Measurement/ or measurement.mp. (329,071)

13. Statistical analysis.mp. or exp Statistical Analysis/ (82,305)

14. statistical measurement.mp. (1,339)

15. test validity.mp. or exp Test Validity/ (57,888)

16. test reliability.mp. or exp Test Reliability/ (40,923)

17. specificity.mp. (26,964)

18. sensitivity.mp. (76,935)

19. accuracy.mp. (53,432)

20. area under curve.mp. (146)

21. area under the curve.mp. (1,609)

22. AUC.mp. (1,457)

23. receiver operating characteristic*.mp. (2,971)

24. ROC.mp. (2,410)

25. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (492,047)

26. exp Mentally Ill Offenders/ or mentally ill offender*.mp. (3,350)

27. mentally disordered offender*.mp. (626)

28. MDO.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (40)

29. offender*.mp. (28,024)

30. violent offender*.mp. (1,041)

31. patient*.mp. (568,894)

32. exp Patient Violence/ or violent patient*.mp. (1,343)

33. service user*.mp. (38,688)

34. violent service user*.mp. (2)

35. client*.mp. (120,404)

36. exp Clients/ or violent client*.mp. (7,815)

37. exp Psychiatric Patients/ or forensic psychiatric patient*.mp. (27,461)

38. psychiatric patient*.mp. (35,473)

39. violent psychiatric patient*.mp. (35)

40. exp Criminals/ (16,647)

41. criminal*.mp. [mp=title, abstract, heading word, table of contents, key concepts, original title, tests & measures] (45,280)

42. exp Prisoners/ or prisoner*.mp. (13,853)

43. exp Hospitalized Patients/ or inpatient*.mp. (46,459)

44. violent inpatient*.mp.(19)

45. 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 (723,441)

46. 9 and 25 and 45 (661)

47. limit 46 to yr="2000 -Current" (582)

OVID: Embase (1980 to 2015 Week 10), completed 09th March 2015

1. (short term assessment of risk and treatability).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 17

2. START.mp. (135,588)

3. structured assessment of protective factors.mp. (3)

4. SAPROF.mp. (5)

5. predictive validity.mp. or exp predictive validity/ (9,061)

6. statistical validity.mp. (303)

7. statistical analysis.mp. or exp statistical analysis/ (1,407,563)

8. statistical measurement.mp. (53)

9. test validity.mp. (450)

10. exp reliability/ or test reliability.mp. (113,431)

11. Specificity.mp. or exp "sensitivity and specificity"/ (615,539)

12. sensitivity.mp. (939,078)

13. exp accuracy/ or accuracy.mp. (517,009)

14. area under curve.mp. or exp area under the curve/ (82,519)

15. area under the curve.mp. (95,994)

16. AUC.mp. (54,625)

17. exp receiver operating characteristic/ or exp roc curve/ or receiver operating characteristic*.mp. (62,995)

18. ROC.mp. (43,880)

19. measurement.mp. or exp measurement/ (1,926,025)

20. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 (4,478,432)

21. exp offender/ or mentally ill offender*.mp. or exp mental patient/ (27,307)

22. mentally disordered offender*.mp. (478)

23. MDO.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (196)

24. offender*.mp. (15,352)

25. violent offender*.mp. (542)

26. patient*.mp. or exp patient/ or exp hospital patient/ (6,685,862)

27. violent patient*.mp. (453)

28. inpatient*.mp. (98,492)

29. violent inpatient*.mp. (10)

30. service user*.mp. (3,708)

31. violent service user*.mp. (0)

32. client*.mp. (50,345)

33. violent client*.mp. (20)

34. exp prisoner/ or forensic psychiatric patient*.mp. (11,733)

35. psychiatric patient*.mp. (13,347)

36. violent psychiatric patient*.mp. (24)

37. criminal*.mp. (27,103)

38. prisoner*.mp. (13,922)

39. 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 (6,767,656)

40. HCR-20.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (160)

41. HCR-20V3.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (0)

42. HCR-20 version 3.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (0)

43. historical clinical risk management.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] (35)

44. 1 or 3 or 4 or 40 or 41 or 42 or 43 (178)

45. 20 and 39 and 44 (106)

46. limit 45 to yr="2000 -Current" (103)

OVID: MEDLINE (R) (1946 to March Week 1 2015), completed 09th March 2015

1. (short term assessment of risk and treatability).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (8)

2. START.mp. (94,375)

3. structured assessment of protective factors.mp. (1)

4. SAPROF.mp. (3)

5. HCR-20.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (83)

6. HCR-20V3.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (0)

7. HCR-20 version 3.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (0)

8. historical clinical risk management.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (17)

9. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 (94,457)

10. exp "Reproducibility of Results"/ or predictive validity.mp. (286,392)

11. exp Data Interpretation, Statistical/ or statistical validity.mp. (49,457)

12. measurement.mp. (417,739)

13. statistical analysis.mp. (59,164)

14. statistical measurement.mp. (34)

15. test reliability.mp. (624)

16. area under curve.mp. or exp Area Under Curve/ (28,980)

17. receiver operating characteristic*.mp. (27,780)

18. ROC.mp.(42,273)

19. test validity.mp. (279)

20. exp "Sensitivity and Specificity"/ (431,374)

21. area under the curve.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (45,260)

22. Specificity.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (821,105)
23. sensitivity.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (780,218)
24. accuracy.mp. (209,310)
25. AUC.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (33,177)
26. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 (2,100,766)
27. exp Prisoners/ or exp Mentally Ill Persons/ or mentally ill offender*.mp. (16,851)
28. mentally disordered offender*.mp. or exp Criminals/ (1,453)
29. MDO.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (142)
30. offender*.mp. (7,378)
31. violent offender*.mp. (352)
32. patient*.mp. (4,717,092)
33. violent patient*.mp. (364)
34. inpatient*.mp. or exp Inpatients/ (6,664)

35. violent inpatient*.mp. (9)

36. service user*.mp. (2,307)

37. violent service user*.mp. (0)

38. client*.mp. (37,852)

39. violent client*.mp. (16)

40. forensic psychiatric patient*.mp. (120)

41. psychiatric patient*.mp. (10,839)

42. violent psychiatric patient*.mp. (17)

43. criminal*.mp. (21,174)

44. prisoner*.mp. (14,587)

45. 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 (4,787,657)

46. 9 and 26 and 45 (4,560)

47. limit 46 to yr="2000 -Current" (3,322)

ProQuest: Applied Social Sciences Index and Abstracts (ASSIA), completed 09th March 2015

(short term assessment of risk treatability OR START OR structured assessment of protective factors OR SAPROF OR HCR-20 OR HCR-20V3 OR HCR-20 version 3 OR historical clinical risk management) AND (predictive validity OR statistical validity OR measurement OR statistical analysis OR statistical measurement OR test validity OR test reliability OR specificity OR sensitivity OR accuracy OR area

under curve OR area under the curve OR AUC OR receiver operating characteristic* OR ROC) AND (mentally ill offender* OR mentally disordered offender* OR MDO OR offender* OR violent offender* OR patient* OR violent patient* OR inpatient* OR violent inpatient* OR service user* OR violent service user* OR client* OR violent client* OR forensic psychiatric patient* OR psychiatric patient* OR violent psychiatric patient* OR criminal* OR prisoner*) (114)

Web of Science (Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (A&HCI), Conference Proceedings Citation Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science and Humanities (CPCI-SSH)) (Timespan = 2000-2015), completed 11th March 2015

1. TS = ("short term assessment of risk and treatability" OR "structured assessment of protective factors" OR SAPROF OR "historical clinical risk management" OR HCR-20 version 3 OR HCR-20 20 V3 OR HCR-20) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (229)

2. TI=(START) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (18,809)

3. #2 OR #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (19,028)

4. TS = (predictive validity OR statistical validity OR measurement OR statistical analysis OR statistical measurement OR test validity OR test reliability OR specificity OR sensitivity OR accuracy OR "area under curve" OR "area under the curve" OR AUC OR "receiver operating characteristic*" OR ROC) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (2,860,663)

5. TS = ("mentally ill offender*" OR "mentally disordered offender*" OR MDO OR offender* OR "violent offender*" OR patient* OR "violent patient*" OR inpatient* OR "violent inpatient*" OR "service user*" OR "violent service user*" OR client* OR "violent client*" OR "forensic psychiatric patient*" OR "psychiatric patient*" OR "violent psychiatric patient*" OR criminal* OR prisoner*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (3,003,071)

6. #5 AND #4 AND #3 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2000-2015* (321)

Cochrane Library, completed on 09th March 2015

1. short term assessment of risk and treatability (3)

2. START (38,231)

3. structured assessment of protective factors (279)

4. SAPROF (0)

5. HCR-20 (3)

6. HCR-20V3 (0)

7. HCR-20 version 3 (1)

8. historical clinical risk management (793)

9. #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 (38,898)

10. predictive validity (1,783)

11. statistical validity (16,565)

12. measurement (39,998)

13. statistical analysis (39,370)

14. statistical measurement (7,825)

15. test validity (12,660)

16. test validity (12,660)

17. specificity (18,236)

18. sensitivity (51,166)

19. accuracy (14,630)

20. area under curve (15,996)

21. area under the curve (15,942)

22. AUC (10,057)

23. receiver operating characteristic (2,184)

24. ROC (2,911)

25. #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 (131,892)

26. mentally ill offender (8)

27. mentally disordered offender (9)

28. MDO (38)

29. offender (216)

30. violent offender (34)

31. patient (163,520)

32. violent patient (202)

33. service user (982)

34. violent service user (34)

35. client (1,639)

36. violent client (42)

37. forensic psychiatric patient (56)

38. psychiatric patient (6,118)

39. violent psychiatric patient (134)

40. criminal (714)

41. prisoner (82)

42. inpatient (6,565)

43. violent inpatient (86)

44. #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #42 (167,531)

45. #9 and #25 and #44 (6,670)

46. #1 or #3 or #4 or #5 or #6 or #7 or #8 (1056)

47. #46 and #25 and #44 (678)

Appendix 3.B: Inclusion and exclusion criteria

	INCLUDE CRITERIA	EXCLUDE CRITERIA
Population	▪ Males ▪ Adults aged 18 years and over ▪ Primary diagnosis of mental illness or personality disorder ▪ History of violent behaviour (convicted or not)	▪ Females only ▪ Children and adolescents aged 17 years and under only ▪ Other primary diagnosis, e.g. learning disability, or no diagnosis only ▪ History of sexually violent behaviour only ▪ No history of violent behaviour
Exposure	▪ Full HCR-20V2 and/or V3 and/or ▪ Full START and/or ▪ Full SAPROF	▪ No violence risk assessment tool ▪ Risk assessment tool not listed in inclusion criteria ▪ Select scales of HCR-20V2/V3/START/ SAPROF only
Outcome	▪ HCR-20/START and/or START vulnerabilities/SAPROF: ▪ Actual, attempted, threatened harm to others/self ▪ Self-reported/observed aggression to others/self ▪ Self-reported/observed violence to others/self ▪ Violent reoffending ▪ Violent reconviction ▪ Violent recidivism ▪ Readmission ▪ SAPROF/START strengths: ▪ Absence of actual, attempted, threatened harm to others/self ▪ Absence of self-reported/observed aggression to others/self ▪ Absence of self-reported/observed violence to others/self ▪ Absence of violent reoffending ▪ Absence of violent reconviction ▪ Absence of violent recidivism ▪ Absence of readmission	▪ HCR-20/START and/or START vulnerabilities/SAPROF: ▪ Self-reported/observed sexual aggression only ▪ Self-reported/observed sexual violence only ▪ Sexual reoffending only ▪ Sexual reconviction only ▪ Sexual recidivism only ▪ SAPROF/START strengths: ▪ Absence of self-reported/observed sexual aggression only ▪ Absence of self-reported/observed sexual violence only ▪ Absence of sexual reoffending only ▪ Absence of sexual reconviction only ▪ Absence of sexual recidivism only

Setting	▪ Any setting	▪ Setting is not considered grounds for exclusion
Study design	▪ All quantitative study designs ▪ Quantitative and qualitative study designs combined	▪ No statistical analysis of predictive validity (AUC) ▪ Literature review or systematic review ▪ Meta-analysis

Appendix 3.C: Quality Assessment Form: Cohort

A group of participants is identified and followed over time to assess specific outcomes. There may or may not be a concurrent control group.

Reference:

Question	Score	Comments
Are the results of the study valid?		
Were the study objectives clear?		
1. Will a cohort design address the objectives? - Prospective (Y) - Retrospective (P)		
Selection bias (Was the cohort recruited in an acceptable way?)		
2. Was the cohort clearly defined?		
3. Was the cohort representative of a defined population? - multi-site (Y) - range of age, ethnicity, diagnosis, violent offence, length of stay (P)		
4. Was everybody included who should have been included? - all inpatients of a ward/unit (Y) - all discharged of a ward/unit (Y) - nothing special about the cohort (Y) - Subjects excluded, but valid reason reported (P)		
Measurement bias (exposure, i.e. risk assessment tool)		
5. Was the risk assessment tool clearly stated?		
6. Was the same risk assessment tool used across the cohort? If not, were the subjects allocated using the same procedure?		
7. Were the assessors trained/experienced enough to be competent in applying the risk assessment tool (as defined in the manual)?		
8. Was the assessor blind to the outcome?		
9. Was the risk assessment tool completed using information gathered from more than one source (as directed in the manual)? - more than 2 sources (Y) - 2 sources (P)		
10. Was consensus rating used?		
11. Was inter-rater reliability assessed?		
12. Was inter-rater reliability for total risk assessment tool: - r = 0.80 or above (Y) - ICC = .75 or above (Y)		
13. Was missing information discussed and dealt with appropriately (pro-rated as directed in the manual)?		
Measurement bias (outcome, i.e. actual, attempted, threatened harm to others/self)		
14. Was the outcome clearly defined?		
15. Was the outcome data source clearly stated?		
16. Was the outcome measure clearly stated, and did it truly reflect the defined outcome? Or, if an objective tool was used to measure the outcome, and was it reliable and valid?		

17. Was the same measure of outcome used across the sample? If not, were the subjects allocated using the same procedure?		
18. Was the assessor blind to the exposure?		
19. Was missing information discussed and dealt with appropriately?		
Follow up period		
20. Was the follow up period long enough? HCR-20/SAPROF: - 12 months or longer (Y) - 6 - 12 months (P) START: - 3 months or longer (Y) - 1-2 months (P)		
21. Were subjects lost during the follow up period discussed, and the reasons why recorded?		
22. Were the implications of lost subjects discussed?		
What are the results?		
23. Were the results clearly reported (AUC values)?		
24. Was predictive validity clearly stated?		
25. Was the base rate of the outcome reported?		
26. If two risk assessment tools are used (e.g. HCR-20 and SAPROF) is incremental validity discussed (effect of combined tools on predictive validity)?		
27. Was construct validity discussed (correlation between risk assessment tool and outcome)?		
28. Was concurrent validity discussed (correlation between risk assessment tool and previously validated risk assessment tools)?		
29. Were confounding factors identified and/or discussed?		
30. Taking into account any bias identified, are the results reliable?		
Will the results help locally?		
31. Can the results be applied to the local population (generalizable to all adult male offenders with a diagnosis of mental illness or personality disorder, and a history of violent behaviour)?		
32. Do the results of this study fit with other available evidence?		
33. Are the implications of this study for practice discussed?		
Quality score	(66)	*Y = 2, P = 1, N = 0, U = unclear*
No. Unclear		

Appendix 3.D: Quality Assessment Form: Case Control

A group of participants with a particular condition are matched for age and other characteristics with a control group of participants who do not have the condition.

Reference:

Question	Score	Comments
Are the results of the study valid?		
Were the study objectives clear?		
1. Will a case control study address the objectives?		
Selection bias (Were the cases and controls recruited in an acceptable way?)		
2. Were cases clearly defined?		
3. Were controls clearly defined?		
4. Were cases and controls randomly selected from the population?		
5. Were cases reliably assessed as such?		
6. Were controls reliably assessed as such?		
7. Were demographics of cases and controls clear?		
8. Were demographics of cases and controls comparable?		
Measurement bias (exposure, i.e. risk assessment tool)		
9. Was the risk assessment tool clearly stated?		
10. Was the same risk assessment tool used across the cases and controls? If not, were the subjects allocated using the same procedure?		
11. Were the assessors trained/experienced enough to be competent in applying the risk assessment tool (as defined in the manual)?		
12. Was the assessor blind to the outcome?		
13. Was the risk assessment tool completed using information gathered from more than one source (as directed in the manual)? - more than 2 sources (Y) - 2 sources (P)		
14. Was consensus rating used?		
15. Was inter-rater reliability assessed?		
16. Was inter-rater reliability for total risk assessment tool: - r = 0.8 or above (Y) - ICC = .75 or above (Y)		
17. Was missing information discussed and dealt with appropriately (pro-rated as directed in the manual)?		
Measurement bias (outcome, i.e. actual, attempted, threatened harm to others/self)		
18. Was the outcome clearly defined?		
19. Was the outcome data source clearly stated?		
20. Was the outcome measure clearly stated, and did it truly reflect the defined outcome? Or, if an objective tool was used to measure the outcome, and was it reliable and valid?		
21. Was the same measure of outcome used across the sample? If not, were the subjects allocated using the same procedure?		
22. Was the assessor blind to the exposure?		

23. Was missing information discussed and dealt with appropriately?		
Follow up period		
24. Was the follow up period long enough? HCR-20/SAPROF: - 12 months or longer (Y) - 6 - 12 months (P) START: - 3 months or longer (Y) - 1-2 months (P)		
25. Were subjects lost during the follow up period discussed, and the reasons why recorded?		
26. Were the implications of lost subjects discussed?		
What are the results?		
27. Were the results clearly reported (AUC values)?		
28. Was predictive validity clearly stated?		
29. Was the base rate of the outcome reported?		
30. If two risk assessment tools are used (e.g. HCR-20 and SAPROF) is incremental validity discussed (effect of combined tools on predictive validity)?		
31. Was construct validity discussed (correlation between risk assessment tool and outcome)?		
32. Was concurrent validity discussed (correlation between risk assessment tool and previously validated risk assessment tools)?		
33. Were confounding factors identified and/or discussed?		
34. Taking into account any bias identified, are the results reliable?		
Will the results help locally?		
35. Can the results be applied to the local population (generalizable to all adult male offenders with a diagnosis of mental illness or personality disorder, and a history of violent behaviour)?		
36. Do the results of this study fit with other available evidence?		
37. Are the implications of this study for practice discussed?		
Quality score	(74)	*Y = 2, P = 1, N = 0, U = unclear*
No. Unclear		

Appendix 3.E: Data extraction form

Reference:
Risk tool
Additional assessment
Country
Setting
Study design
Follow-up
Sample size
% male
M age
% history of violence
% diagnosis of mental illness
% diagnosis of personality disorder
Outcome
Average base rate
Inter-rater reliability
AUC

Appendix 3.F: References excluded - stage 1 (137)

Reference	Reason for exclusion
Abou-Sinna and Luebbers (2012)	Study design
Babalola, Gormez, Alwan, Johnstone, and Sampson (2014)	Exposure and study design
Bjorkly, Hartvig, Roaldset, and Singh (2014)	Exposure and study design
Buchanan (2014)	Study design
Campbell, French, and Gendreau (2009)	Study design
Castellettic, Rivellini, and Stratico (2014)	Study design
Chakhssi, de Ruiter, and Bernstein (2010)	Exposure
Cote, Crocker, Nicholls, and Seto (2012)	Study design
Crocker, Braithwaite, Laferriere, Gagnon, Venegas, and Jenkins (2011)	Study design
Davoren, Abidin, Naughton, Gibbons, Nulty, Wright, and Kennedy (2013)	Study design
Daffern and Howels (2007)	Exposure
Davoren, Abidin, Naughton, Gibbons, Nulty, Wright, and Kennedy (2013)	Study design
De Ruiter and Nicholls (2011)	Study design
De Vogel and de Ruiter (2004)	Study design
De Vogel, de Ruiter, Bouman, and de Vries Robbé (2009a)	Study design
De Vogel, de Ruiter, Bouman, and de Vries Robbé (2009b)	Study design
De Vogel, de Vries Robbé, van Kalmthout, and Place (2014)	Population and exposure
De Vries Robbé (2014)	Population and outcome
De Vries Robbé and de Vogel (2013)	Study design
De Vries Robbé, de Vogel, and Stam (2012)	Study design
De Vries Robbé, Mann, Maruna, and Thornton (2014)	Exposure and study design
Desmarais, Collins, Nicholls, and Brink (2011)	Study design
Desmarais, Sellers, Viljoen, Cruise, Nicholls, and Dvoskin (2012)	Population
Dietiker, Dittmann, and Graf (2007)	Study design
Dolan and Fullam (2007)	Exposure and study design
Douglas (2004)	Outcome
Douglas (2014)	Study design
Douglas (2014)	Study design
Douglas and Belfrage (2014)	Study design
Douglas, Hart, Webster, Belfrage, Guy, and Wilson (2014)	Study design
Douglas and Kropp (2002)	Study design
Douglas and Ogloff (2003)	Study design
Doyle, Coid, Archer-Power, Dewa, Hunter-Didrichsen, Stevenson, Wainwright, Kallis, Ullrich, and Shaw (2014)	Exposure and study design
Doyle and Dolan (2002)	Exposure and study design
Doyle, Lewis, and Brisbane	Study design
Dunbar, Quinones, and Crevecoeur (2005)	Study design
Edworthy and Khalifa (2014)	Exposure and study design
Eidhammer, Selmer, and Bjorkly (2013)	Study design

Reference	Reason for exclusion
Fazel, Singh, Doll, and Grann (2012)	Study design
Fluttert, Van Meijel, Webster, Nijman, Bartels, and Grypdonck (2008)	Exposure and study design
Gairing, de Tribolet-Hardy, Vohs, and Habermeyer (2013)	Study design
Gravier and Lustenberger (2005)	Exposure and study design
Guy (2010)	Study design
Guy, Douglas, and Hendry (2010)	Study design
Guy, Packer, and Warnken (2012)	Study design
Habermeyer, Gairing, and Lau (2010)	Study design
Hartvig, Alfarnes, Ostberg, Skjonberg, and Moger (2006)	Exposure
Hartvig, Roaldset, Moger, Ostberg, and Bjorkly (2011)	Exposure
Heibrun, Holliday, and Brooks (2013)	Exposure and study design
Hilterman, Philipse, and de Graaf (2011)	Exposure
Hodgins, Tengstrom, Eriksson, Osterman, Kronstrand, Eaves, Hart, Webster, Ross, Levin, Levander, Tuninger, Muller-Isberner, Freese, Tiiihonen, Kotilainen, Repo-Tiihonen, Vaananen, Eronen, Vokkolainen, and Vartiainen (2007)	Exposure and study design
Horstead (2013)	Study design
Hurducas, Singh, de Ruiter, Petrila (2014)	Study design
Inett, Wright, Roberts, and Sheeran (2014)	Population
Izycky, Braham, Williams, and Hogue (2010)	Exposure
Jung, Ledi, and Daniels (2013)	Study design
Khiroya, Weaver, Maden (2009)	Study design
Kötter, von Franqué, Bolzmacher, Eucker, Holsinger, and Müller-Isberner (2014)	Study design
Kroppan, Nesset, Nonstad, Pederson, Almvik, and Palmstierna (2011)	Study design
Langton (2007)	Exposure and study design
Lewis and Webster (2004)	Study design
Lindsay, Hastings, and Beail (2013)	Population, exposure, and study design
Litwack (2001)	Study design
Liu, Yang, Ramsey, Li, and Coid (2011)	Study design
Logan (2014)	Study design
Long and Dolley (2012)	Population and exposure
Lussier, Verdun-Jones, Deslauriers-Varin, Nicholls, and Brink (2010)	Study design
Lv, Han, and Wang (2013)	Study design
Mann, Matias, and Allen (2014)	Exposure and study design
McDermott, Dualan, and Scott (2011)	Exposure
McDougall, Pearson, Willoughby, and Bowles (2013)	Exposure and study design
McKeown (2010)	Population
Megargee (2013)	Study design
Mills, Kroner, and Hemmati (2007)	Exposure
Mills, Kroner, and Morgan (2011)	Study design
Mills and Gray (2013)	Exposure
Moons, Boriau, and Ferdinande (2008)	Exposure

Reference	Reason for exclusion
Muller-Isberner, Webster, and Gretenkord (2007)	Study design
Murphy (2007)	Study design
Nanayakkara, O'Driscoll, and Allnutt (2012)	Study design
Nicholls (2004)	Population
Nieberding, Moore, and Dematatis (2002)	Study design
O'Shea and Dickens (2014)	Study design
O'Shea, Mitchell, Picchioni, and Dickens (2013)	Study design
O'Shea, Picchioni, Mason, Sugarman, and Dickens (2014)	Outcome
Ogloff and Daffern (2006)	Exposure
Olsson, Strand, Kristiansen, Sjoling, and Asplund (2013)	Study design
Penney, McMaster, and Wilkie (2014)	Study design
Philipse, Koeter, Van Der Staak, and Van Den Brink (2005)	Exposure and study design
Pillay, Oliver, Butler, and Kennedy (2008)	Study design
Pyott (2005)	Study design
Reimann and Nussbaum (2011)	Outcome
Reynolds, Jones, Davies, Freeth, and Heyman (2014)	Study design
Rice and Harris (2013)	Exposure and study design
Rizzo and Smith (2012)	Study design
Roaldset, Olav, Hartvig, Linaker, and Bjorkly (2012)	Exposure
Roberts and Coid (2007)	Exposure
Rogers and Jackson (2005)	Population, exposure, and study design
Rossegger, Frank, Elbert, Fries, and Endrass (2010)	Population
Rufino, Boccaccini, and Guy (2011)	Study design
Schaap, Lammers, and de Vogel (2009)	Population
Selenius, Hellstrom, and Belfrage (2011)	Study design
Sevilla-Sanchez, Espaulella, de Andres-Lazaro, Torres-Allezpuz, Soldevila-Llagostera, and Codina-Jane (2012)	Exposure
Singh, Desmarais, Hurducas, Arbach-Lucioni, Condemarin, Dean, Doyle, Folino, Godoy-Cervera, Grann, Ho, Large, Nielsen, Pham, Rebocho, Reeves, Rettenberger, de Ruiter, Seewald, and Otto (2014)	Study design
Singh, Fazel, Gueorguieva, and Buchanan (2014)	Study design
Singh, Grann, and Fazel (2011)	Study design
Singh, Serper, Reinharth, and Fazel (2011)	Study design
Skipworth (2005)	Exposure and study design
Smith, Kelley, Rulseh, Sörman, and Edens (2014)	Study design
Smith and White (2007)	Study design
Snowden, Gray, Taylor, and MacCulloch (2007)	Exposure
Stadtland, Hollweg, Kleindienst, Dietl, Reich, and Nedopil (2005)	Population
Stadtland and Nedopil (2005)	Exposure and study design
Stalans, Yarnold, Seng, Olson, and Repp (2004)	Exposure
Stanfill, O'Brien, and Viglione (2014)	Study design
Strand and Belfrage (2001)	Population and study design
Stübner, Groβ, and Nedopil (2006)	Exposure and study design
Sturup, Monahan, and Kristiansson (2013)	Exposure

Reference	Reason for exclusion
Tardiff and Hughes (2011)	Study design
Telles, Day, Folino, and Taborda (2009)	Study design
Tengstrom (2001)	Exposure
Tiegreen (2010)	Exposure
Tozdan (2014)	Population
Ullrich and Coid (2011)	Exposure and study design
Van Den Berg and de Vogel (2011)	Population and study design
Van Den Brink, Hooijschuur, van Os, Savenije, and Wiersma (2010)	Exposure
Van den Broek and de Vries Robbé (2008)	Study design
Viljoen, Nicholls, Greaves, Ruiter, and Brink (2011)	Population
Vitacco, Erickson, Kurus, and Apple (2012)	Study design
Vladejic, Vladejic, and Popovic (2011)	Study design
Vogel, Ruiter, Bouman, and Robbé (2010)	Study design
Walters, Kroner, DeMatteo, and Locklair (2014)	Study design
Warren, south, Burnette, Rogersm Friend, Bale, and Van Patten (2005)	Population
Watt, Topping-Morris, Rogersm Doyle, and Mason (2003)	Study design
Webster, Muller-Isberner, and Fransson (2002)	Study design
Webster, Nicholls, Martin, Desmarais, and Brink	Study design
Yao, Li, Hu, and Cheng (2012)	Exposure

Appendix 3.G: References excluded - stage 2 (34)

Reference	Reason for exclusion
Belfrage, Fransson, and Strand (2000)	Study design
Blum (2004)	Exposure
Cesniene (2010)	Exposure
Chu, Daffern, and Ogloff (2013)	Exposure
Daffern (2007)	Study design
Dahle (2006)	Population
De Vogel, de Vries Robbé, de Ruiter, and Bouman (2011)	Study design
De Vries Robbé (2014)	Dissertation – studies within already included
De Vries Robbé, de Vogel, Douglas, and Nijman (2015)	Population and outcome
De Vries Robbée, de Vogel, Koster, and Bogaerts (2015)	Population and outcome
Desmarais, van Dorn, Telford, Petrila, and Coffey (2012)	Study design
Douglas, Yeomans, and Boer (2005)	Population
Doyle, Dolan, and McGovern (2002)	Exposure
Garcia-Mansilla, Rosenfeld, and Cruise (2011)	Exposure
Gray, Hill, McGleish, Timmons, MacCulloch, and Snowden (2003)	Exposure
Green, Griswold, Schreiber, Prentky, and Kunz (2014)	Exposure
Grevatt, Thomas-Peter, and Hughes (2004)	Exposure
Hill, Rettenberger, Habermann, Berner, Eher, and Briken (2012)	Population
Ho, Thomson, and Darjee (2009)	Exposure
Jovanovic, Tosevski, Ivkovic, Damjanovic, and Gasic (2009)	Population
Kroner and Mills (2001)	Population
Lindsay, Hogue, Taylor, Steptoe, Mooney, O'Brien, Johnston, and Smith (2008)	Population
Mokros, Stadtland, Osterheider, and Nedopil (2010)	Exposure
Morrissey, Beeley, and Milton (2014)	Study design
Nicholls, Ogloff, and Douglas (2004)	Exposure
Nicholls, Petersen, Brink, and Webster (2011)	Study design
Nilsson, Wallinius, Gustavson, Anckarsater, and Kerekes (2011)	Exposure
O'Shea, Picchioni, McCarthy, Mason, and Dickens (2015)	Population
Polvi (2001)	Exposure
Stadtland, Hollweg, Kleindienst, Dietl, Reich, and Nedopil (2006)	Population
Sturup, Karlberg, Fredriksson, Lihoff, and Kristiansson (2015)	Population
Teo, Holley, Leary, and McNiel (2012)	Exposure
Thomson, Davidson, Brett, Steele, and Darjee (2008)	Exposure
Yoon, Spehr, and Briken (2011)	Population

Appendix 3.H: References excluded – not available in timeframe (9)

Reference	Reason for exclusion
Claix and Pham (2004)	Non English language (Dutch)
Martinaki, Tsopelas, Ploupidis, Douzenis, Tzavara, Skapinakis, and Mavreas (2013)	Non English language (Greek)
Matiasko (2010)	Non English Language (Czech)
McNiel, Gregory, Lam, Binder, and Sullivan (2003)	Unobtainable despite author contact
Mudde, Nijman, van der Hulst, and van den Bout (2011)	Non English language (Dutch)
Nedopil (2009)	Unobtainable despite author contact
Pham, Chevrier, Nioche, Ducro, and Reveillere (2005)	Non English language (French)
Pham, Ducros, Marghem, and Reveillee (2005)	Non English language (French)
Sinani, Kola, Cenko, Elezi, Balaj, Saraci, Dervishi, and Gjolena (2013)	Unobtainable and unable to contact authors

Appendix 3.I: References excluded – in previously completed systematic review

De Vogel, de Ruiter, Hildebrand, Bos, and van de Ven (2004)
De Vogel and de Ruiter (2005)
De Vogel and de Ruiter (2006)
Dolan and Khawaja (2004)
Douglas, Ogloff, and Hart (2003)
Douglas and Ogloff (2003)
Dowsett (2005)
Doyle and Dolan (2006)
Fujii, Tokioka, Lichton, and Hishinuma (2005)
Gray, Snowden, MacCulloch, Phillips, Taylor, and MacCulloch (2004)
Gray, Fitzgerald, Taylor, MacCulloch, and Snowden (2007)
Gray, Taylor, and Snowden (2008)
Macpherson and Kevan (2004)

Appendix 3.J: Quality assessment (part 1)

	Are the results of the study valid?																					
	Selection bias				Measurement bias (exposure)									Measurement bias (outcome)						Follow-up period		
Cohort	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Abidin, et al. (2013)	✓	✓	?	✓	✓	✓	U	✓	✓	×	✓	?	×	✓	✓	✓	✓	?	×	?	×	×
Arbach-Lucioni, et al. (2011)	✓	✓	?	✓	✓	✓	✓	U	?	×	×	×	×	✓	✓	?	✓	✓	×	✓	×	×
Barber-Rioja, et al. (2012)	?	✓	?	×	✓	✓	✓	U	×	?	✓	✓	×	✓	✓	?	✓	✓	×	✓	×	×
Barnard-Croft (2014)	✓	✓	✓	✓	✓	✓	U	U	?	×	✓	✓	×	✓	✓	?	✓	U	×	✓	✓	✓
Braithwaite et al. (2010)	✓	✓	?	?	✓	✓	✓	U	×	×	×	×	×	✓	✓	?	✓	✓	×	?	×	×
Chu, et al. (2011)	?	✓	?	?	✓	✓	✓	✓	?	?	×	×	×	?	×	?	✓	U	×	?	×	×
Chu, et al. (2011)	?	✓	?	?	✓	✓	✓	✓	✓	?	×	×	×	✓	✓	?	✓	U	×	?	×	×
Coid, et al. (2009)	✓	✓	?	?	✓	✓	✓	U	?	×	✓	✓	×	?	✓	?	✓	U	×	?	?	?
Coid, et al. (2011)	✓	✓	✓	?	✓	✓	U	U	?	×	✓	✓	×	✓	✓	?	✓	U	×	?	×	×
Coid, et al. (2013)	✓	✓	?	?	✓	✓	✓	U	?	×	✓	✓	×	?	✓	?	✓	U	×	?	?	?
Coupland (2015)	?	✓	?	?	✓	✓	✓	✓	×	×	✓	?	×	✓	✓	?	✓	?	×	×	✓	×
De Borba Telles, et al. (2012)	✓	✓	U	✓	✓	✓	U	U	✓	U	×	×	×	✓	✓	✓	✓	U	×	✓	×	×
De Vogel, et al. (2014)	?	✓	?	✓	✓	✓	U	U	U	×	✓	✓	×	×	×	×	✓	U	×	✓	×	×
de Vries Robbé & de Vogel (2011)	✓	✓	?	U	✓	✓	✓	U	U	✓	✓	✓	×	×	×	×	U	U	×	U	×	×
De Vries Robbé, et al. (2011)	?	✓	?	?	✓	✓	✓	✓	?	×	?	✓	×	✓	✓	✓	✓	U	×	✓	✓	×
De Vries Robbé, et al. (2013)	?	✓	?	✓	✓	✓	✓	✓	×	✓	✓	?	×	✓	✓	?	✓	U	×	✓	×	×
De Vries Robbe, et al. (2014)	✓	✓	?	✓	✓	✓	✓	U	?	✓	×	×	×	✓	✓	?	✓	U	×	?	×	×
Dernevik, et al. (2002)	✓	✓	?	✓	✓	✓	✓	U	U	?	×	×	×	✓	✓	✓	✓	U	×	✓	×	×
Desmarais, et al. (2010)	✓	✓	?	✓	✓	✓	✓	U	×	×	✓	✓	×	✓	✓	✓	✓	✓	×	✓	×	×
Desmarais, et al. (2012)	?	✓	?	?	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓	?	✓	?	×	✓	×	×
Dolan and Blattner (2010)	?	✓	?	?	✓	✓	✓	✓	×	×	×	×	×	✓	×	✓	✓	U	×	U	×	×

Doyle, et al. (2012)	✓	✓		?		✓		✓	✓	?	U	✓	×	✓	✓	×	✓	✓	✓	✓	U	×	×	✓	×
Doyle, et al. (2014)	✓	✓		✓		✓		✓	✓	✓	U	?	✓	✓	✓	×	✓	✓	✓	✓	U	×	✓	✓	×
Gray, et al. (2011)	?	✓		✓		U		✓	✓	?	✓	?	×	✓	✓	×	?	✓	?	✓	U	×	✓	×	×
Gray, et al. (2011)	✓	✓		?		?		✓	✓	✓	U	?	×	×	×	×	✓	✓	?	✓	U	×	✓	✓	×
Ho, et al. (2013)	✓	✓		?		✓		✓	✓	✓	U	×	×	✓	?	×	✓	✓	?	✓	✓	?	✓	?	×
Langton, et al. (2009)	✓	✓		?		?		✓	✓	✓	U	?	×	×	×	✓	✓	✓	?	✓	U	×	✓	✓	×
Mcdermott, et al. (2008)	✓	✓		?		×		✓	✓	✓	U	U	×	?	✓	×	✓	✓	?	✓	U	?	✓	×	×
Michel, et al. (2013)	✓	✓		✓		?		✓	✓	✓	U	✓	×	✓	?	×	✓	✓	?	✓	U	×	✓	✓	×
Neves, et al. (2011)	✓	✓		?		?		✓	✓	✓	U	✓	×	×	×	×	✓	✓	?	✓	U	×	✓	?	×
Nicholls, et al. (2006)	✓	✓		?		✓		✓	✓	✓	✓	?	×	×	×	×	✓	✓	?	✓	✓	×	✓	×	×
Nonstad, et al. (2010)	✓	✓		?		?		✓	✓	?	U	U	✓	×	×	×	✓	✓	?	✓	U	×	×	×	×
O'Shea, et al. (2014)	✓	✓		?		?		✓	✓	?	U	?	×	×	×	×	✓	✓	✓	✓	✓	×	×	×	×
Pedersen, et al. (2010)	?	✓		?		?		✓	✓	U	U	×	×	×	✓	×	✓	✓	?	✓	U	×	✓	×	×
Pedersen, et al. (2012)	✓	✓		?		?		✓	✓	✓	U	U	?	×	×	×	✓	✓	?	✓	U	×	✓	×	×
Quinn, et al. (2013)	?	✓		?		✓		✓	✓	?	✓	×	✓	×	×	×	✓	✓	?	✓	U	×	✓	×	×
Snowden, et al. (2010)	?	✓		✓		?		✓	✓	✓	✓	×	×	✓	✓	×	✓	✓	?	✓	U	×	✓	×	×
Strub, et al. (2014)	✓	✓		✓		✓		✓	✓	✓	U	✓	×	×	×	×	✓	✓	✓	✓	U	×	?	×	×
Troquete, et al. (2014)	✓	?		?		?		✓	✓	✓	U	U	×	✓	?	×	✓	✓	?	✓	?	×	✓	×	×
Viljoen (2014)	✓	✓		?		?		✓	✓	U	?	?	×	✓	?	×	✓	✓	?	✓	U	×	✓	?	×
Vojt, et al. (2013)	✓	✓		?		?		✓	✓	U	U	U	×	×	×	×	✓	✓	?	✓	U	×	✓	✓	×
Whittington, et al. (2014)	✓	✓		?		?		✓	✓	U	U	×	×	×	×	?	✓	✓	?	✓	U	×	✓	✓	×

Case Control	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Wilson, et al. (2010)	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	U	×	×	?	✓	×	✓	✓	?	✓	✓	×	✓	×	×
Wilson, et al. (2013)	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	U	U	×	×	✓	✓	?	✓	✓	?	✓	✓	×	✓	×	×

Appendix 3.K: Quality assessment (part 2)

	What are the results?								Will the results help locally?			Total unclear	Quality score (66)
Cohort	23	24	25	26	27	28	29	30	31	32	33		
Abidin, et al. (2013)	✓	✓	✓	×	✓	✓	?	?	?	✓	✓	1	45
Arbach-Lucioni, et al. (2011)	✓	✓	✓	×	✓	×	?	?	?	✓	?	1	39
Barber-Rioja, et al. (2012)	✓	✓	✓	?	×	×	×	?	?	✓	?	1	38
Barnard-Croft (2014)	✓	✓	×	×	✓	×	?	?	✓	?	?	3	42
Braithwaite et al. (2010)	✓	✓	✓	×	×	×	?	?	?	✓	✓	1	35
Chu, et al. (2011a)	✓	✓	✓	×	×	×	×	?	?	?	✓	1	31
Chu, et al. (2011b)	✓	?	✓	×	×	×	×	?	?	✓	✓	1	35
Coid, et al. (2009)	✓	✓	✓	×	×	×	×	?	✓	✓	✓	2	39
Coid, et al. (2011)	✓	✓	✓	×	×	×	?	?	✓	✓	?	3	37
Coid, et al. (2013)	✓	✓	✓	×	×	×	×	?	✓	✓	✓	2	39
Coupland (2015)	✓	✓	✓	✓	✓	✓	✓	?	?	✓	✓	0	46
De Borba Telles, et al. (2012)	✓	✓	✓	×	×	×	?	?	?	✓	✓	5	35
De Vogel, et al. (2014)	✓	✓	×	×	×	✓	×	?	?	✓	?	4	29
de Vries Robbé and de Vogel (2011)	✓	✓	×	✓	×	×	×	?	?	✓	?	6	28
De Vries Robbé, et al. (2011)	✓	✓	✓	✓	✓	✓	×	?	?	✓	✓	1	47
De Vries Robbé, et al. (2013)	✓	✓	?	✓	✓	✓	?	?	?	✓	✓	1	47
De Vries Robbe, et al. (2014)	✓	✓	×	✓	✓	×	×	?	?	✓	✓	2	38
Dernevik, et al. (2002)	✓	✓	✓	×	✓	×	×	?	✓	✓	?	3	38
Desmarais, et al. (2010)	✓	✓	✓	×	✓	×	?	?	?	✓	✓	1	44
Desmarais, et al. (2012)	✓	✓	✓	✓	?	✓	×	×	?	✓	✓	0	43
Dolan and Blattner (2010)	✓	✓	✓	×	×	×	?	?	?	✓	✓	2	34
Doyle, et al. (2012)	✓	✓	✓	×	✓	×	?	?	✓	✓	✓	2	44

Doyle, et al. (2014)	✓	✓	✓	×	×	×	×	?	✓	✓	✓	2	46
Gray, et al. (2011a)	✓	✓	✓	×	×	×	×	?	✓	✓	?	2	37
Gray, et al. (2011b)	✓	✓	✓	×	✓	×	×	?	?	✓	?	2	37
Ho, et al. (2013)	✓	✓	✓	×	×	×	?	?	?	✓	✓	1	42
Langton, et al. (2009)	✓	✓	✓	×	✓	✓	✓	?	×	?	×	2	40
McDermott, et al. (2008)	✓	×	✓	×	✓	×	?	?	?	U	×	5	33
Michel, et al. (2013)	✓	✓	✓	×	×	×	?	?	✓	✓	✓	2	43
Neves, et al. (2011)	✓	✓	✓	✓	✓	✓	✓	?	?	✓	✓	2	44
Nicholls, et al. (2006)	✓	✓	✓	×	✓	×	×	?	?	✓	✓	0	41
Nonstad, et al. (2010)	✓	✓	✓	×	?	×	×	×	?	✓	✓	3	34
O'Shea, et al. (2014)	✓	✓	✓	×	×	×	×	?	?	✓	✓	1	34
Pedersen, et al. (2010)	✓	✓	✓	×	×	×	?	?	?	✓	✓	2	35
Pedersen, et al. (2012)	✓	✓	✓	×	×	×	?	?	?	?	✓	3	34
Quinn, et al. (2013)	✓	✓	✓	×	×	✓	×	?	?	✓	✓	1	38
Snowden, et al. (2010)	✓	✓	✓	×	×	×	×	?	?	✓	✓	1	39
Strub, et al. (2014)	✓	✓	✓	×	✓	✓	✓	?	✓	?	✓	2	43
Troquete, et al. (2014)	✓	✓	✓	✓	×	×	?	?	?	?	✓	2	38
Viljoen (2014)	✓	✓	✓	✓	×	✓	?	?	?	?	✓	2	41
Vojt, et al. (2013)	✓	✓	✓	×	×	×	×	?	?	✓	✓	4	33
Whittington, et al. (2014)	✓	✓	✓	×	×	×	×	?	?	✓	?	3	32

Case Control	27	28	29	30	31	32	33	34	35	36	37	Total unclear	Quality score (74)
Wilson, et al. (2010)	✓	✓	✓	×	✓	×	?	?	?	✓	✓	1	50
Wilson, et al. (2013)	✓	✓	✓	×	×	×	×	?	?	✓	✓	2	47

Appendix 3.L: Data from studies evaluating the predictive validity of specified violence risk assessment tools in adults, with a diagnosis of mental illness and/or personality disorder, with a history of violence

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
HCR-20V2	Abidin, et al. (2013)		6 months	.87 harm to others	.78	.96
				.88 harm to self	.79	.97
	Arbach-Lucioni, et al. (2011)		1-4 months	.75		
			5-8 months	.69		
			9-12 months	.77		
	Barber-Rioja, et al. (2012)	ICC = .90	12 months	.71 re-incarceration	.61	.79
				.79 non-compliance	.71	.86
	Chu, et al. (2011a)		1 month	.78 interpersonal violence	.66	.91
				.68 verbal threat	.45	.91
				.72 any inpatient aggression	.58	.87
			3 months	.75 interpersonal violence	.60	.90
				.69 verbal threat	.46	.92
				.78 any inpatient aggression	.64	.93
			6 months	.62 interpersonal violence	.40	.84
				.62 verbal threat	.35	.88
				.59 any inpatient aggression	.38	.81
	Coid, et al. (2009)	ICC = .98	*M* 1.97 years	.67 violence	.63	.71
				.69 acquisitive	.66	.72
				.67 any	.64	.70
	Coid, et al. (2011)	ICC = .98	*M* 1.97 years	.67 violence	.63	.71
	Coid, et al. (2013)	ICC = .98	*M* 1.97 years	.68 no DSM-IV Axis I disorder	.62	.75
				.64 DSM-IV Axis I disorder	.60	.69
				.62 schizophrenia	.52	.72
				.63 lifetime depression	.56	.70

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.63 drug dependence	.57	.68
				.60 alcohol disorder	.53	.68
				.70 no Axis II disorder	.60	.79
				.58 Axis II disorder (not ASPD)	.45	.71
				.60 ASPD	.56	.64
				.67 low PCL-R score	.62	.72
				.61 medium PCL-R score	.56	.67
				.44 high PCL-R score	.30	.58
	Coupland (2015)	Pre-treatment: ICC = .93 Post treatment: ICC = .94	*M* 9.7 years	Community recidivism (convictions): Pre-treatment: .64 all violent .65 nonsexual violent .75 any recidivism Post treatment: .72 all violent .72 nonsexual violent .81 any recidivism	 .55 .55 .64 .63 .64 .73	 .74 .75 .86 .80 .81 .90
				Community recidivism (all charges): Pre-treatment: .70 all violent .70 nonsexual violent .75 any recidivism Post treatment: .77 all violent .78 nonsexual violent	 .59 .59 .64 .69 .69	 .80 .81 .87 .86 .86

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.83 any recidivism	.75	.91
				Institutional recidivism: pre-treatment:		
				.55 major	.46	.64
				.49 minor	.40	.57
				.50 violent	.39	.61
				.50 any	.41	.59
				Institutional recidivism: HCR-20 post treatment:		
				.60 major	.51	.69
				.48 minor	.39	.56
				.54 violent	.41	.67
				.51 any	.42	.59
	De Borba Telles, et al. (2012)		12 months	.82 general offending		
				.73 violent offending		
	de Vries Robbé & de Vogel (2011)		12 months	.74 violent		
				.85 sexual		
				.79 total		
	De Vries Robbé, et al. (2011)		12 months	.81		
			24 months	.77		
			36 months	.68		
	De Vries Robbé, et al. (2013)	ICC = .74	12 months	.84	.73	.95
			36 months	.73	.62	.84
			Long term	.64	.56	.73
	De Vries Robbe, et al. (2014)		12 months	.79 total		
				.80 total male		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.75 total female		
				.68 intramural/supervised		
				.85 unsupervised/transmural		
				.73 violent		
				.89 sexual		
				.81 major mental illness		
				.77 personality disorder		
				.76 high psychopathy		
	Dernevik, et al. (2002)		12 months	.68 total incidents	.52	.83
				.68 total incidents excluding self-harm	.54	.83
				.64 weeks in high risk management	.46	.82
				.64 high risk management excluding self-harm	.46	.82
				.78 medium risk management	.62	.95
				.82 medium risk management excluding self-harm	.67	.98
				.71 low risk management	.51	.91
	Desmarais, et al. (2012)		12 months	.80 any aggression	.72	.88
				.80 verbal aggression	.72	.88
				.79 physical aggression – objects	.69	.89
				.75 physical aggression – others	.65	.86
	Dolan and Blattner (2010)		12 months	.86	.78	.94
	Doyle, et al. (2012)	Historical: ICC = .97 Clinical: ICC = .85 Risk: ICC = .83	5 months	.68	.56	.80

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
	Gray, et al. (2011a)	r = .80		Reconviction: .69 all mental disorders .72 schizophrenia .62 personality disorder .63 substance use .80 mental retardation .63 mood disorder Violent reconviction: .73 all mental disorders .74 schizophrenia .62 personality disorder .65 substance use .80 mental retardation .67 mood disorder		
	Ho, et al. (2013)	ICC = .37	6 months	.67 verbal violence .71 violence against others .79 violent conviction .70 any violence		
			12 months	.62 verbal violence .68 violence against others .79 violent conviction .68 any violence		
	Langton, et al. (2009)		12 months	.58 physical aggression .60 damage to property	.39 .38	.78 .82
			Full period	.68 physical aggression .70 damage to property	.52 .53	.84 .87
	McDermott, et al. (2008)	Risk:	6 months	.69 impulsive aggression		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
		ICC = .97		.89 predatory aggression		
				.73 psychotic aggression		
			Full period	.67 impulsive aggression		
				.68 predatory aggression		
				.57 psychotic aggression		
	Michel, et al. (2013)	Historical:	6 months	.74 forensic		
		ICC =.90		.60 general psychiatric		
		Clinical:				
		ICC = .78				
		Risk:				
		ICC = .52				
			12 months	.67 forensic		
				.74 general psychiatric		
			18 months	.70 forensic		
				.72 general psychiatric		
			24 months	.72 forensic		
				.72 general psychiatric		
	Neves, et al. (2011)		*M* 12.82	.84 general recidivism	.78	.90
			months	.81 violent behaviour	.72	.89
				.82 non-violent recidivism	.74	.90

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
	O'Shea, et al. (2014)		3 months	Any aggression:		
				.72 full sample	.66	.78
				.70 male	.62	.77
				.78 female	.68	.86
				.74 schizophrenia	.66	.81
				.71 personality disorder	.64	.79
				.69 schizophrenia and personality disorder	.61	.77
				.67 developmental	.57	.76
				.64 organic	.52	.76
				.64 Caucasian	.55	.72
				.66 non-Caucasian	.52	.78
				Physical aggression:		
				.66 full sample	.59	.72
				.62 male	.54	.70
				.70 female	.58	.81
				.70 schizophrenia	.61	.79
				.66 personality disorder	.58	.74
				.62 schizophrenia and personality disorder	.54	.69
				.56 developmental	.47	.67
				.52 organic	.39	.65
				.55 Caucasian	.44	.66
				.64 non-Caucasian	.48	.79
	Pedersen, et al. (2010)	ICC = .90	*M* 6 years	.73 any crime	.63	.83
				.74 violent crime	.64	.83
	Pedersen, et al. (2012)		*M* 21 years	.70 inpatient aggression	.57	.83
				.66 violent recidivism	.52	.80
	Snowden, et al. (2010)	ICC = .80	24 months	.71 all offenders		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.72 white offenders		
				.66 black offenders		
	Troquete, et al. (2014)		3 months	.59 violent and criminal behaviour	.48	.69
				.58 as defined by START	.47	.69
			6 months	.61 violent and criminal behaviour	.52	.70
				.58 as defined by START	.48	.69
	Viljoen (2014)	ICC = .80	6 months	.55 violence		
				.51 physical aggression		
				.61 verbal aggression		
				.57 sexual aggression		
				.42 serious physical aggression		
				.51 serious verbal aggression		
				.48 serious sexual aggression		
				.56 most serious incident		
				.55 most serious violent incident		
				.56 most serious sexual incident		
			12 months	.60 violence		
				.56 physical aggression		
				.62 verbal aggression		
				.59 sexual aggression		
				.52 serious physical aggression		
				.61 serious verbal aggression		
				.47 serious sexual aggression		
				.58 most serious incident		
				.57 most serious violent incident		
				.63 most serious sexual incident		
	Vojt, et al. (2013)		*M* 31 months	.50 all incidents	.39	.61
				.54 minor incidents	.43	.65

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.86 serious incidents	.76	.96
				.60 any convictions	.33	.87
	Wilson, et al. (2013)	ICC = .88	3 months	.86 physical aggression or sexually inappropriate behaviour	.73	1.00
			6 months	.81 physical aggression or sexually inappropriate behaviour	.65	.97
			9 months	.74 physical aggression or sexually inappropriate behaviour	.51	.98
			12 months	.81 physical aggression or sexually inappropriate behaviour	.57	1.00
				.88 any aggression	.76	1.00
HCR-20V3	De Vogel, et al. (2014)	ICC = .83	12 months	.77 total score .82 summary risk ratings		
			24 months	.75 total score .74 summary risk ratings		
			36 months	.67 total score .71 summary risk ratings		
	Doyle, et al. (2014)	ICC = .92	6 months	.73		
			12 months	.70		
	Strub, et al. (2014)		4-6 weeks	Presence: .78 combined .88 patient .70 offender Relevance: .71 combined .82 patient .63 offender		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
			6-8 months	Presence:		
				.46 combined		
				.70 patient		
				.79 offender		
				Relevance:		
				.68 combined		
				.72 patient		
				.64 offender		
SAPROF	Abidin, et al. (2013)	r = .83	6 months	.85 harm to others	.72	.97
				.77 harm to self	.60	.93
	Barnard-Croft (2014)	ICC > .90	6 months	.78 MI	.72	.84
				.70 PD	.47	.95
				.71 co-morbid	.57	.84
			12 months	.69 MI	.62	.76
				.76 PD	.55	.96
				.71 co-morbid	.56	.81
	Coupland (2015)	Pre-treatment:	*M* 9.7 years	Community recidivism (convictions):		
		ICC = .73		Pre-treatment:		
				.64 all violent	.55	.73
				.65 nonsexual violent	.56	.74
		Post treatment:		.73 any recidivism	.64	.83
		ICC = .79				
				Post treatment:		
				.65 all violent	.57	.74
		At release:		.66 nonsexual violent	.58	.75
		ICC = .80		.72 any recidivism	.63	.81

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				Pre-release:		
				.71 all violent	.63	.80
				.72 nonsexual violent	.64	.80
				.76 any recidivism	.67	.85
				Community recidivism (all charges):		
				Pre-treatment:		
				.70 all violent	.61	.79
				.71 nonsexual violent	.62	.80
				.72 any recidivism	.62	.83
				Post treatment:		
				.71 all violent	.62	.80
				.72 nonsexual violent	.64	.81
				.72 any recidivism	.62	.82
				Pre-release:		
				.75 all violent	.67	.84
				.76 nonsexual violent	.68	.85
				.75 any recidivism	.66	.85
				Institutional recidivism:		
				Pre-treatment:		
				.55 major	.46	.64
				.50 minor	.41	.59
				.50 violent	.38	.62
				.52 any	.43	.60

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				Institutional recidivism:		
				Post treatment:		
				.57 major	.47	.66
				.51 minor	.42	.59
				.49 violent	.37	.61
				.53 any	.44	.61
	de Vries Robbé & de Vogel (2011)		12 months	.77 violent		
				.81 sexual		
				.78 total		
	De Vries Robbé, et al. (2011)	ICC = .88	12 months	.85		
			24 months	.80		
			36 months	.74		
	De Vries Robbé, et al. (2013)	ICC = .79	12 months	.85	.74	.96
			36 months	.75	.65	.85
			Long term	.73	.66	.81
	De Vries Robbe, et al. (2014)		12 months	.75 total		
				.76 total male		
				.71 total female		
				.66 intramural/supervised		
				.78 unsupervised/transmural		
				.72 violent		
				.84 sexual		
				.79 major mental illness		
				.68 personality disorder		
				.76 high psychopathy		
	Viljoen (2014)	ICC = .75	6 months	.59 violence		
				.52 physical aggression		
				.60 verbal aggression		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.61 sexual aggression		
				.50 serious physical aggression		
				.62 serious verbal aggression		
				.68 serious sexual aggression		
				.58 most serious incident		
				.52 most serious violent incident		
				.70 most serious sexual incident		
			12 months	.59 violence		
				.54 physical aggression		
				.63 verbal aggression		
				.60 sexual aggression		
				.55 serious physical aggression		
				.67 serious verbal aggression		
				.58 serious sexual aggression		
				.60 most serious incident		
				.58 most serious violent incident		
				.71 most serious sexual incident		
START	Abidin, et al. (2013)	START/S:	6 months	START/S:		
		r = .69		.78 harm to others	.65	.90
				.64 harm to self	.45	.83
		START/V:				
		r = .85		START/V:		
				.82 harm to others	.71	.94
				.54 harm to self	.46	.85
	Braithwaite et al. (2010)		30 days	START/S:		
				.65 aggression toward others	.56	.74
				.57 self-harm	.38	.76

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				START/V:		
				.66 aggression toward others	.56	.75
				.58 self-harm	.20	.75
	Chu, et al. (2011a)		1 month	.75 interpersonal violence	.60	.91
				.79 verbal threat	.59	.99
				.74 any inpatient aggression	.59	.90
			3 months	.79 interpersonal violence	.63	.95
				.82 verbal threat	.51	1.00
				.83 any inpatient aggression	.68	.98
			6 months	.74 interpersonal violence	.54	.95
				.79 verbal threat	.55	1.00
				.74 any inpatient aggression	.55	.93
	Chu, et al. (2011b)		1 month	START/S:		
				.71 any inpatient aggression	.56	.86
				.75 interpersonal violence	.59	.91
				.64 verbal threat	.44	.84
				START/R:		
				.76 any inpatient aggression	.59	.93
				.78 interpersonal violence	.61	.94
				.77 verbal threat	.53	1.00
	Desmarais, et al. (2010)	ICC = .87	12 months	Full sample:		
				Low/moderate confidence:		
				.83 any aggression		
				.85 verbal aggression		
				.80 physical aggression - objects		
				.77 physical aggression – others		
				.88 self-harm		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				High confidence: .70 any aggression .68 verbal aggression .70 physical aggression – objects .65 physical aggression – others .57 self-harm Inpatient subsample: Low/moderate confidence: .88 any aggression .94 verbal aggression .90 physical aggression – objects .83 physical aggression – others .76 self-harm High confidence: .60 any aggression .61 verbal aggression .56 physical aggression – objects .60 physical aggression – others .63 self-harm		
	Desmarais, et al. (2012)	START/S: ICC = .93 START/V: ICC = .95	12 months	START/S: .76 any aggression .75 verbal aggression .77 physical aggression – objects .80 physical aggression – others	.68 .66 .65 .70	.85 .84 .89 .89

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
		START/R: ICC = .85		START/V:		
				.79 any aggression	.71	.87
				.79 verbal aggression	.70	.87
				.80 physical aggression – objects	.71	.89
				.77 physical aggression – others	.66	.88
				START/R:		
				.80 any aggression	.72	.88
				.78 verbal aggression	.69	.86
				.84 physical aggression – objects	.76	.92
				.85 physical aggression – others	.77	.93
	Gray, et al. (2011b)		*M* 114 days	START/S:		
				.21 violence to others		
				.28 verbal aggression		
				.61 self-harm		
				START/V:		
				.68 violence to others		
				.74 verbal aggression		
				.48 self-harm		
				START/R:		
				.65 violence to others		
				.70 verbal aggression		
				.86 self-harm		
	Nicholls, et al. (2006)		12 months	Full sample:		
				.67 verbal aggression	.61	.73
				.69 physical aggression – objects	.62	.76

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.65 physical aggression – others	.57	.72
				.65 sexually inappropriate	.43	.86
				.66 self-harm	.54	.77
				Inpatient sample:		
				.72 verbal aggression	.58	.86
				.67 physical aggression – objects	.52	.83
				.70 physical aggression – others	.55	.85
				.92 sexually inappropriate	.79	1.05
				.67 self-harm	.50	.84
	Nonstad, et al. (2010)		3 months	.77 START/S	.64	.91
				.77 START/V	.64	.91
	Quinn, et al. (2013)		1 month	.14 START/S		
				.85 START/V		
				.91 START/R		
			3 months	.32 START/S		
				.74 START/V		
				.68 START/R		
			6 months	.43 START/S		
				.67 START/V		
				.58 START/R		
	Troquete, et al. (2014)	START/S:	3 months	HCR-20-START/V:		
		ICC = .49		.59 violent and criminal behaviour	.49	.70
				.58 as defined by START	.47	.69
		START/V:				
		ICC = .64		HCR-20-START/V-START/S:		
				.59 violent and criminal behaviour	.49	.70
		START/R:		.58 as defined by START	.47	.69

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
		ICC = .58				
				HCR-20-START/V-START/S-START/R:		
				.62 violent and criminal behaviour	.52	.72
				.58 as defined by START	.47	.69
			6 months	HCR-20-START/V:		
				.62 violent and criminal behaviour	.52	.71
				.59 as defined by START	.48	.69
				HCR-20-START/V-START/S:		
				.62 violent and criminal behaviour	.52	.72
				.59 as defined by START	.48	.69
				HCR-20-START/V-START/S-START/R:		
				.65 violent and criminal behaviour	.56	.74
				.62 as defined by START	.52	.72
	Viljoen (2014)	START/S:	6 months	START/S:		
		ICC = .44		.63 violence		
				.65 physical aggression		
		START/V:		.61 verbal aggression		
		ICC = .56		.60 sexual aggression		
				.62 serious physical aggression		
		START/R:		.62 serious verbal aggression		
		ICC = .24		.61 serious sexual aggression		
				.66 most serious incident		
				.66 most serious violent incident		
				.57 most serious sexual incident		
				START/V:		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.54 violence .50 physical aggression .59 verbal aggression .60 sexual aggression .51 serious physical aggression .64 serious verbal aggression .53 serious sexual aggression .55 most serious incident .52 most serious violent incident .61 most serious sexual incident		
				START/R: .68 violence .69 physical aggression .68 verbal aggression .64 sexual aggression .61 serious physical aggression .62 serious verbal aggression .72 serious sexual aggression .67 most serious incident .69 most serious violent incident .54 most serious sexual incident		
			12 months	START/S: .60 violence .60 physical aggression .64 verbal aggression .55 sexual aggression .64 serious physical aggression		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.56 serious verbal aggression .51 serious sexual aggression .62 most serious incident .64 most serious violent incident .59 most serious sexual incident START/V: .54 violence .52 physical aggression .61 verbal aggression .59 sexual aggression .54 serious physical aggression .62 serious verbal aggression .48 serious sexual aggression .56 most serious incident .54 most serious violent incident .69 most serious sexual incident START/R: .67 violence .67 physical aggression .63 verbal aggression .66 sexual aggression .72 serious physical aggression .66 serious verbal aggression .71 serious sexual aggression .69 most serious incident .68 most serious violent incident		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.62 most serious sexual incident		
	Whittington, et al. (2014)		30 days	.55 START/S	.47	.64
				.74 START/V	.64	.84
			M 231 days	.75 START/S	.59	.89
				.69 START/V	.52	.85
	Wilson, et al. (2010)	START/S: ICC = .85	3 months	.74 START/S		
				.70 START/V		
		START/V: ICC = .90				
			6 months	.81 START/S		
				.81 START/V		
			9 months	.71 START/S		
				.70 START/V		
			12 months	.80 START/S		
				.73 START/V		
	Wilson, et al. (2013)	START/S: ICC = .85	3 months	START/S: .74 physical aggression or sexually inappropriate behaviour	.52	.97
		START/V: ICC = .90		START/V: .73 physical aggression or sexually inappropriate behaviour	.48	.98
			6 months	START/S: .81 physical aggression or sexually inappropriate behaviour	.64	.99
				START/V:		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.81 physical aggression or sexually inappropriate behaviour	.65	.98
			9 months	START/S:		
				.71 physical aggression or sexually inappropriate behaviour	.49	.93
				START/V:		
				.70 physical aggression or sexually inappropriate behaviour	.44	.96
			12 months	START/S:		
				.80 physical aggression or sexually inappropriate behaviour	.54	1.00
				.84 any aggression	.70	.98
				START/V:		
				.73 physical aggression or sexually inappropriate behaviour	.41	1.00
				.82 any aggression	.67	.98
HCR-20/ SAPROF	Coupland (2015)		*M* 9.7 years	Community recidivism (convictions):		
				Pre-treatment:		
				.66 all violent	.56	.75
				.66 nonsexual violent	.57	.76
				.71 any recidivism	.59	.82
				Post treatment:		
				.65 all violent	.56	.74
				.66 nonsexual violent	.57	.75
				.70 any recidivism	.60	.81

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				Pre-release:		
				.65 all violent	.56	.74
				.66 nonsexual violent	.57	.75
				.71 any recidivism	.61	.82
				Community recidivism (all charges):		
				Pre-treatment:		
				.70 all violent	.60	.80
				.71 nonsexual violent	.61	.81
				.70 any recidivism	.59	.82
				Post treatment:		
				.69 all violent	.60	.79
				.70 nonsexual violent	.61	.79
				.70 any recidivism	.61	.81
				Pre-release:		
				.72 all violent	.63	.82
				.73 nonsexual violent	.64	.82
				.72 any recidivism	.61	.82
				Institutional recidivism:		
				Pre-treatment:		
				.54 major	.45	.63
				.51 minor	.43	.60
				.54 violent	.42	.66
				.51 any	.42	.60

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				Institutional recidivism:		
				Post treatment:		
				.61 major	.52	.71
				.52 minor	.44	.61
				.64 violent	.53	.76
				.54 any	.45	.62
	de Vries Robbé & de Vogel (2011)	ICC = .80	12 months	.81 violent		
				.84 sexual		
				.82 total		
	De Vries Robbé, et al. (2011)		12 months	.85		
			24 months	.81		
			36 months	.72		
	De Vries Robbé, et al. (2013)		12 months	.87	.76	.97
			36 months	.76	.65	.86
			Long term	.70	.62	.78
	De Vries Robbe, et al. (2014)		12 months	.80 total		
				.82 total male		
				.74 total female		
				.70 intramural/supervised		
				.85 unsupervised/transmural		
				.76 violent		
				.88 sexual		
				.82 major mental illness		
				.75 personality disorder		
				.79 high psychopathy		
	Viljoen (2014)		6 months	.58 violence		
				.53 physical aggression		

Risk tool	Research	Inter-rater reliability	Follow-up	Predictive validity (AUC)[c]	Lower 95% CI	Upper 95% CI
				.64 verbal aggression		
				.62 sexual aggression		
				.46 serious physical aggression		
				.58 serious verbal aggression		
				.56 serious sexual aggression		
				.58 most serious incident		
				.55 most serious violent incident		
				.64 most serious sexual incident		
			12 months	.61 violence		
				.58 physical aggression		
				.67 verbal aggression		
				.63 sexual aggression		
				.54 serious physical aggression		
				.66 serious verbal aggression		
				.51 serious sexual aggression		
				.61 most serious incident		
				.60 most serious violent incident		
				.68 most serious sexual incident		

Appendix 5.A. URICA

Name: Date:

Each statement below describes how a person might feel when starting therapy or approaching problems in his/her life. Please indicate the extent to which you tend to agree or disagree with each statement.

In each case, make your choice in terms of how you feel right now, not what you have felt in the past or would like to feel.

For all the statements that refer to your 'problem', answer in terms of problems related to mental health or offending behaviour. The words 'here' and 'this place' refer to the MSU.

Circle the response that best describes how much you agree or disagree with each statement.

		1	2	3	4	5
1	As far as I am concerned, I don't have any problems that need changing.	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
2	I think I might be ready for some self-improvement	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
3	I am doing something about the problems that had been bothering me	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
4	It might be worthwhile to work on my problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
5	It worries me that I might slip back on a problem I have already changed, so I am here to seek help	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
6	I am not the problem one. It doesn't make much sense for me to consider changing	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
7	I am finally doing some work on my problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
8	I have been thinking that I might want to	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

	change something about myself					
9	I have been successful in working on my problem, but I'm not sure I can keep up the effort on my own	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
10	At times my problem is difficult, but I am working on it	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
11	Trying to change is pretty much a waste of time for me because the problem doesn't have anything to do with me	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
12	I'm hoping that I will be able to understand myself better	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
13	I guess I have faults, but there's nothing that I really need to change	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
14	I am really working hard to change	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
15	I have a problem and I really think I should work on it	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
16	I'm not following through with what I had already changed as well as I had hoped, and I want to prevent a relapse of the problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
17	Even though I'm not always successful in changing, I am at least working on my problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
18	I thought once I had resolved the problem I would be free of it, but sometimes I still find myself struggling with it	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
19	I wish I had more	Strongly	Disagree	Undecided	Agree	Strongly

	ideas on how to solve my problems	disagree				Agree
20	I have started working on my problem, but I would like help	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
21	Maybe someone or something will be able to help me	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
22	I may need a boost right now to help me maintain the changes I've already made	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
23	I may be part of the problem, but I don't really think I am	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
24	I hope that someone will have some good advice for me	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
25	Anyone can talk about changing, I'm actually doing something about it	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
26	All this talk about psychology is boring. Why can't people just forget about their problems	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
27	I'm struggling to prevent myself from having a relapse of my problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
28	It is frustrating, but I feel I might have recurrence of a problem I thought I had resolved	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
29	I have worries, but so does the next guy	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
30	I am actively working on my problem	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
31	I would rather cope with my faults than try to change them	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
32	After all I had done to try and change my	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

	problem every now and again it comes back to haunt me					

Appendix 5.B. Pre-training questionnaire

Risk and Protection Awareness Training

You have been put forward by your multidisciplinary team as someone who may be interested in attending this training. This is an opportunity for you to learn more about your risk assessment and risk management plans, and it will help you to work collaboratively with your multidisciplinary team in the development of these.

The aims of this group are to:

- Improve your understanding of the risk assessment and risk management process.
- Improve your understanding of how your risk assessment and risk management plan relates to your treatment plan.
- Help you to work collaboratively with the team in developing your risk assessment, risk management plan, and treatment plan.
- Help you to feel more in control of your treatment.
- Increase your motivation to engage in your treatment plan.

At the end of the group you will have learnt what each of the core risk assessments used are, and how they are used to guide your treatment during your stay. You will have been able to see an overview of your own risk assessments and you will be in a position to have a discussion with your team psychologist about the content of your risk assessment. You will also be able raise discussions with your team if your views differ from theirs.

- 10 x 1 hour sessions, with no break.
- It is important to attend all sessions.
- If you miss a session due to circumstances beyond your control the material you have missed will be covered with you individually.
- There are no individual sessions in this group but if you feel you need extra support please ask.

Pre Group Interview

Do you think you have completed this group, or a group like it before?

Information gathered from this assessment will be used to create a care plan for your attendance at the training. Attendance at this group can help you to achieve goals within all sections of the My Shared Pathway. Do you have any personal goals which you think this training could help you with?

How do you feel about being asked to take part in this training?

Is there anything you think may prevent you from taking part or anything you find particularly difficult?

Is there anything you would not feel comfortable discussing within a group setting?

Do you have any worries about the training?

Can name any of the risk assessments used here?

Has anyone ever discussed with you what is in your risk assessment and risk management report?

Have you seen copies of your risk assessment and risk management reports?

Can you tell me what a risk factor (or vulnerability) is?

Can you tell me what your risk factors are?

Can you tell what a protective factor (or strength) is?

Can you tell me what your protective factors are?

Do you know what the signs are if things aren't going right for you (your relapse indicators)?

Can you tell me what you need to do to keep yourself safe or to reduce your risk of reoffending in the future?

Can you tell me what others need to do to keep you safe or to reduce your risk of reoffending in the future?

It is hoped that by the end of the group you will have an increased understanding about all of these things, and you will also be aware of what your team's views are.

Do you have any questions?

Additional comments…

Appendix 5.C. Information sheet and consent form

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

CASE STUDY INFORMATION SHEET

The value of protective factors

During your admission risk assessments of future violence are completed. Until recently these risk assessments focused purely on risk factors. In 2014, the SAPROF was introduced to assess the presence of protective factors and it is incorporated into your risk assessment with the aim of providing a more 'balanced' assessment.

We would like to explore the impact of the introduction of the SAPROF, focusing on the following questions:

1) Does the presence of protective factors result in an absence of violence?
2) Does a risk assessment based on risk and protective factors increase the motivation of an individual to change their behaviour more than an assessment based purely on risk factors?

We hope that the results of this evaluation will help to further improve the process of risk assessment and risk management.

Who is organising this case study?
Rachel Whitehead (Forensic Psychologist in Training)
Professor Vincent Egan (University of Nottingham)
Dr Grant Broad (Clinical and Forensic Psychologist).

The outcomes of the evaluation will form part of Doctoral project for Rachel Whitehead who is a postgraduate student at the University of Nottingham.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will get this information sheet to keep and you will be asked to sign a consent form. If you decide to take part you can still withdraw at any time without giving a reason. If you decide to not take part or to withdraw at any stage it will not affect your leave, your rights and privileges, or your access to medical care. You should also know that taking part in this research will not increase your rights or privileges in any way.

What will happen to me if I take part?
You will firstly be asked to sign a consent form to show you have agreed to take part. A report will be written which describes your past experiences, focusing particularly on the presence or absence of risk and protective factors at the time of your index offence and how these have changed during your admission. In addition, your engagement in the Risk and Protection Awareness Training will be explored, focusing on whether your motivation to change was affected by the risk assessments that were shared with you.

Basic demographic information about you, such as your age, ethnicity, diagnosis, and how long you have been in hospital for will also be collected from your hospital records. No information collected about you will have your name on it.

What are the possible benefits from taking part?
You will receive no direct benefit from taking part in the case study but your participation may mean that we can improve services for residents in units like this one. Taking part will help us gain a further understanding about the importance of considering protective factors in our risk assessments.

Will my taking part in the case study be kept confidential?
We will make a record in your notes that we have seen you for the purpose of this case study. However, any information collected will not be communicated to your clinical team or anyone else outside of the research team.

Should you disclose either the intention to harm yourself, harm another individual, attempt to escape, disclose a previously unknown offence, or act in any way that may result in a breach of security, it would be the duty of the researcher to inform your clinical team of such information so that they may take appropriate action.

- Any information removed from the hospital will have your name removed and will be stored in a secure location at the University.
- Your consent form (which you signed) will be kept in a locked cabinet separate to any other information you have provided.
- Your identity will not be recorded as part of your data, and will not be revealed in any publication that may result from this case study. Data will be collected with only a participation number to identify it.
- The data collected in this study will be used only for the purpose described in this form, and will be available only to the research team.
- Data gathered from this study will be maintained as long as required by regulations, which is up to 5 years following the publication of empirical articles or communications describing the results of the study.

What if I have a concern about the case study?
If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated in the course of this case study, please contact Dr. Lona Lockerbie (Lead for Psychological Practice and Quality) in writing, providing a detailed description of your concern.

What will happen to the results of the case study?
Rachel Whitehead will write up the results of the case study for a Doctorate degree research project in Forensic Psychology. You will get the opportunity to read this if you wish. The results will also be used to make revisions and improvements to the risk assessment process. It will not be possible for anyone to tell that you took part in the case study. However, we will keep the data, without identifying information for up to 5 years after publication.

Thank you for taking the time to hear about this case study. It has important implications and so I hope that you will consider taking part in it.

Rachel Whitehead
Forensic Psychologist in Training

CONSENT FORM
(Final version 1.0: 20.05.15)

Title of Case Study: The value of protective factors

Name of Researcher: Rachel Whitehead

Name of Participant: Mr O

1. I confirm that I have read and understand the information sheet version number 1 dated 20.05.2015 for the above case study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my care, treatment, and legal rights being affected. ☐

3. I understand that relevant sections of my treatment and psychiatric notes, and the data collected in the case study, may be looked at by authorised individuals from the University of Nottingham, the researcher's supervisors, and regulatory authorities where it is relevant to my taking part in the study. I give permission for these individuals to have access to these records and to collect, store, analyse, and publish anonymous information obtained from my participation in this case study. I understand that my personal details will be kept confidential. ☐

4. I understand and agree that data from my engagement with the Risk and Protection Awareness Training will be used to write this case study, along with information about the presence or absence of risk and protective factors at the time of my index offence and how these have changed during admission. ☐

5. I am aware my responsible clinician knows that I am taking part in this case study; however they will not be informed of the results without my consent (unless you disclose anything as described in the information sheet). ☐

6. I agree to take part in the above study. ☐

_____

Name of Participant Date Signature

_____

Name of Researcher Date Signature

3 copies: 1 for participant, 1 for the project notes and 1 for the care and treatment notes of participant