

Assessing Risk in Female Offenders

Kate Anya Geraghty

A thesis submitted to the University of Birmingham for the degree of
Doctorate in Forensic Psychological Practice (Foren.Psy.D.)

Centre for Forensic and Criminological Psychology

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

B15 2TT

July 2015

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Women comprise a minority of the offending population and their crimes are less likely to inflict serious harm when compared with male offenders. Although men may be the predominant perpetrators of violence this does not outweigh the need for evidence informed practice in the assessment and management of risk in female offenders. The gender-specificity debate has dominated the literature with scholars and policy makers calling for a gender responsive approach to the assessment and management of female offenders. This thesis aims to explore the assessment of risk in female offenders. This is achieved through three pieces of research. Firstly, a systematic review of the validity of risk assessments in predicting recidivism and violence for female offenders is presented. This demonstrates that there was great variability with respect to the accuracy of risk assessment tools in predicting recidivism or violence, although tools were found to be better at predicting recidivism than violence. It also demonstrated that there is a significant gap in the empirical base with respect to assessing risk in female offenders. A critique of the HCR-20^{V3} is subsequently presented to assess its reliability and validity. This demonstrates that the tool meets some of the criteria of a sound psychometric measure and that it has good external validity. Limitations of the HCR-20^{V3} are also discussed. The fourth chapter presents an empirical paper. This evaluated the predictive validity of gender-neutral and gender-responsive risk assessments in predicting inpatient violence in female psychiatric offenders. Results indicated that the gender neutral assessments did not perform significantly better than the traditional gender-specific risk assessments in predicting inpatient violence. The final chapter of the thesis concludes by discussing the overall findings as well as the implications for future research and clinical practice. It is suggested that further the gender-responsive approach to female offenders may be better adopted to the management of female offenders rather than to the assessment of their risk.

Dedication

For my Mam,

For everything you have done and given up for me,

For your faith and belief in my ability,

For always being there,

This is for you; I hope it will make you proud.

Cáit

Acknowledgements

Special thanks go to Dr Jessica Woodhams and Dr Darren Bishopp for their support and guidance throughout this thesis. Particularly, during moments of panic and confusion! Both of your expertise and knowledge was very much appreciated.

To [REDACTED] for allowing me to collect data at their sites. Particular thanks go to Dawn, Helena and Geoff for supporting me to set up the project there. Also special thanks go to all the participants. To the victims of female violence: it can only be hoped that the thesis will help contribute towards improving the assessment and management of violence risk.

To my proofreaders extraordinaire. To Magda, for taking the time to read through the thesis. For the phone calls and skyperes which always left me feeling better about my thesis! To Geri, for your support and guidance of my clinical work and for reading through the thesis. Your words of support and encouragement kept my motivation high. And for your sense of humour which always made me smile during some tough times! Thank you both.

To the Monahan's, for the unwavering support you have shown us throughout. We are indebted to you all. There is a piece of all of you in this. You have continually reminded me of the importance of family.

To my fairy Godmother Teresa. Your laugh, humour and outlook in life always brightened me up. I truly could not have asked for a more supportive Aunt!

To Dad, thank you for continuing to support me. It means more than I could say. For your thoughtful emails and phone calls which were cherished throughout my education. And for your continual belief in my ability. Thank you.

To MaryAnne and Derrick. You have done more than I could have asked for. Without you both I could not have done this course. You have gone above and beyond for me. MaryAnne, for your texts and phone calls which let me know that you were always there. Derrick, you helped me settle here and always looked out for me. For the endless cups of tea and trips to get out of the house! The words of encouragement and motivation from you both reminded me of why I was doing the course. I am eternally grateful.

Finally, to my Mam. You have seen me struggle and provided me with continual support both emotionally and financially. For fussing over me and always letting me know that you were there. I could not have got this far without you. We are extremely lucky to have you. Hopefully you will now be able to live your own life.

Dá fhada an lá tagann an tráthnóna

TABLE OF CONTENTS

	Page Number
Abstract	1
Glossary of terms	6
	8
Chapter One	
Introduction	
Chapter Two	24
A systematic review of risk assessment tools for female offenders	
Chapter Three	39
Critique of a psychometric measure: Examining the psychometric Properties of the Historical, Clinical and Risk Management (HCR-20) Version 3 violence risk assessment scheme	
Chapter Four	59
The predictive validity of a female-specific violence risk assessment tool: examining the HCR-20 FAM with female psychiatric offenders	
Chapter Five	112
Discussion	
References	122
Appendices	160

LIST OF TABLES

	Page Number
Table 1 Summary table of predictive validity instruments and study characteristics	30
Table 2 Quality scores for each study	33
Table 3 Violence risk assessment tools	68
Table 4 Psychiatric Classification, Personality Disorder (PD) and Index Offence	82
Table 5 Overall risk ratings for each risk assessment tool	83
Table 6 Mean HCR-20 ^{V2} and subscale scores for violent and non-violent female offenders	86
Table 7 Mean HCR-20 ^{V3} and subscale scores for violent and non-violent female offenders	87
Table 8 Presence of female specific risk factors in HCR-20 ^{V2} FAM and HCR-20 ^{V3} FAM	89
Table 9 Differences in scores between types of aggression and violence perpetrated	91
Table 10 Predictive validity of HCR-20 ^{V2} , HCR-20 ^{V3} , HCR-20 ^{V3} , HCR-20 ^{V3} FAM	93
Table 11 Predictive validity of HCR-20 subscales and individual risk factors	95
Table 12 Predictive validity of Suicidality for rates of self-harm	101

LIST OF APPENDICES

	Page Number
Appendix I HCR-20 ^{V2} and HCR-20 ^{V2} FAM risk factors	160
Appendix II HCR-20 ^{V2} and HCR-20 ^{V3} risk factors	161
Appendix III RAGEE Checklist	162
Appendix IV ROC curves of predictive accuracy of each risk tool	165

LIST OF FIGURES

	Page Number
Figure 1 Flow chart of search process	29
Figure 2-3 Predictive accuracy of HCR-20 ^{V2}	165
Figures 4-5 Predictive accuracy of HCR-20 ^{V2} FAM	165
Figures 6-7 Predictive accuracy of HCR-20 ^{V3}	166
Figures 8-9 Predictive accuracy of HCR-20 ^{V2} FAM	166

Glossary of terms

Term	Definition
Actuarial risk assessments	Static instruments which assess risk based on explicit rules and a small number of static factors
AUC	Area Under the Curve. A statistic used to evaluate the predictive accuracy of a tool. It is a discrimination index assessing the probability that a randomly selected recidivist will score higher than a randomly selected non-recidivist.
Base Rates	Represent the percentage in the population who commit a violent or criminal act
Gender neutral	Characteristics of criminal behaviour that apply equally to men and women
Gender specific	Characteristics that uniquely differentiate females and males. Used interchangeably with 'gender responsive' or 'gender sensitive'.
GPCSL	General Personality and Cognitive Social Learning Model of behaviour: integrated model developed to understand offending behaviour.
HCR-20	The Historical, Clinical and Risk Management Scale. A risk assessment tool assessing risk for future interpersonal violence
FAM	Female Additional Manual. Additional guidelines for assessing risk for interpersonal violence in women
Predictive validity/accuracy	Refers to the ability of an instrument to correctly assess the likelihood of reoffending
Selection Ratios	Cut off scores used to classify individuals into levels of violence (e.g. low, medium, high)
RAGEE Guidelines	Risk Assessment Guidelines for the Evaluation of Efficacy in studies assessing the efficacy of risk assessment tools.
ROC	Receiver Operating Characteristic- used to evaluate the predictive accuracy of a tool. ROC produces a statistic known as an AUC to determine predictive accuracy
RNR	A model of offender rehabilitation that outlines three principles for reducing reoffending: the level of risk, criminogenic need and responsivity principles.
SPJ	Structured Professional Judgement- a type of risk assessment tool that assesses risk based

Assessing risk in female offenders

	on static and dynamic risk factors evidenced to be associated with future offending. Also uses clinical judgement and emphasizes the importance of dynamic risk factors
UCJ	Unstructured clinical judgement. A type of risk assessment tool based on clinical expertise alone

CHAPTER ONE

Introduction

Female Offenders

Presently female offenders comprise 4.48% of the total prison population in the UK (Ministry of Justice [MoJ], 2015). In 2013, 25% of those sentenced in the courts, 25% of those who received convictions and 15% of those under supervision in the community were female (MoJ, 2014a). The discrepancy in male and female offenders is well documented within criminal justice statistics both in the UK and internationally. In the US, 24% of those arrested are female (Snyder & Mulako-Wangota, 2014). Of females serving sentences in 2012 in the UK, 27.4% were for violent offending (41.2% of these were classified as wounding while 24.3% were classified as murder), followed by drugs offences (16.6%) and theft and handling stolen goods (14.6%; MoJ, 2012). Similar patterns of offending have been found for men (MoJ, 2012). Further, violence against the person and theft and handling stolen goods were found to be the offence groups that have the highest number of arrests for men and women (MoJ, 2014b). Additionally, the conviction ratio for males and females was found to be consistently increasing for women over the past decade (MoJ, 2014b).

Female offenders are also considered to be less likely to reoffend with lower rates of recidivism in the US (Deschenes, Owen & Crow, 2006) and UK (MoJ, 2015). However, females incarcerated for property or drugs offences are also considered more likely to recidivate than those with any other offence history (Deschenes et al., 2006). Furthermore, women with property and drug offences are more likely to recidivate sooner than women who have served sentences for violent offences (Deschenes et al., 2006). Although men may be the predominant perpetrators of offending, this does not outweigh the need for evidence informed practice in the assessment and treatment of female offenders. Further, the offence patterns of males and females are similar which reinforces

the need for adequate assessment of risk and treatment needs as well as tailoring interventions to meet the needs of female offenders. There is an increasing recognition of the need for research to include female offenders and to differentiate this from research on male offenders.

The nature of female crime

On the whole women are considered to commit less crime than men and their crime is less likely to be violent (de Vogel, de Vries Robbé, van Kalmhout & Place, 2012, 2014; Monahan et al., 2001). It has been suggested that societal roles regarding the expectations of women has led to differential treatment of women within the criminal justice system, therefore affecting the rates of female crime and severity of sentencing (Lloyd, 1995). Women are also less likely to commit crime in pairs or groups, are less likely to use weapons and their crimes are less likely to result in injury to their victim (Greenfield & Snell, 1999; Kruttschnitt, Gartner & Ferraro, 2002). In the US, where a weapon is used, women were found to be more likely to use a gun when committing a crime alongside a male and more likely to use a knife when on their own (Koons-Wift & Schram, 2003). Furthermore, violence in women is more likely to be relational violence, child abuse or violence towards a relative (Monahan et al., 2001; Nicholls, 2001; Robbins, Monahan & Silver, 2003). It is also considered to be less instrumental and more reactive in nature (Monahan et al., 2011; Nicholls, 2001). However, it is acknowledged that there are females whose violence is similar to male offenders. Here instrumental aggression, hostility, robberies and gang membership are more prevalent (Babcock, Miller & Siard, 2003; Batchelor, 2005; Bottos, 2007). Data from the United States also suggested that the

use of lethal violence was more likely to be perpetrated towards intimate partners/spouses (United States Department of Justice, 1998).

Theories on female offending

Several theories have been put forward in the psychological and criminological literature to explain female criminality. This has included theories attempting to account for the differences between male and female offending. However, there is little consensus among scholars as to the aetiology of female offending (Becker & McCorkel, 2011). Broadly, theories of female criminality can be categorised into gender-neutral and gender-specific theories.

Gender-neutral theories of offending include developmental theories, social learning theories and integrated models which incorporate a number of theories.

Development theories posit that there are two categories of antisocial individuals, namely life-course persistent and adolescent limited (Moffitt, 1993). Life course persistent offenders commit a range of offences, including violence, and offending begins at an early age which persists throughout the lifespan. Antisocial behaviours occur due the interaction between neuropsychological vulnerabilities (temperament, behaviour, cognitive abilities) and the environment (poor family environment, poor socio-economic environment).

Adolescent limited offenders on the other hand, begin and cease offending in adolescence and the theory proposes that they are less likely to commit violent offences. Social learning theory is also offered as a way of explaining antisocial behaviour and is based on the work of Bandura (1977). Here behaviour is learned through observation or conditioning where antisocial behaviours become habitual. This theory was then developed to incorporate a cognitive aspect, appraisal, where the complex interaction of thoughts and feelings on

subsequent behaviours was highlighted (Andrews & Bonta, 2010; Mc Guire, 2004).

Building on this the General Personality and Cognitive Social Learning Model (GPCSL) developed, offering an integrated approach to explain offending behaviour (Andrews and Bonta, 2010). Under this model offending behaviour is viewed as based upon the individual weighing up the rewards and costs that encourage offending behaviour. These are influenced by interpersonal factors such as family, school, work, attitudes, feelings and antisocial beliefs as well as the modelling of criminal behaviour by family and/or antisocial peers. These personal, interpersonal and environmental factors are proposed as operating within a broader context of social, political and cultural influences which act as secondary influences on offending behaviour (Andrews & Bonta, 2010). As such an individual engages in offending behaviours based upon a number of factors including characteristics of the situation (e.g. the victim), the emotional state of the individual (e.g. anger), individual attitudes and personality (antisocial values) and the environment. Offending behaviour thus occurs due to how the individual appraises a situation and their ability to self-regulate. Notably, the GPCSL model of offending highlights that factors including age, ethnicity and gender, are not assumed to directly account for variations in criminal behaviour. These are proposed as having influence through the primary GPSCS factors highlighted above.

In terms of gender-specific theories, early theories of female criminality often used biological determinants to explain criminal behaviour. For instance, it was argued that female criminals had masculine characteristics which predisposed them to crime and were thus more similar to male criminals than non-criminal females. Here women were assumed to lack the 'feminine qualities' that acted as a protective factor from crime (Lombroso & Ferrero, 1895). Biological research in recent years may offer some support here. Dabbs et al. (1988) demonstrated a correlation between high testosterone levels in women, a

hormone typically produced in much higher quantities in males than females, and violent behaviour in prison. Further research also found link found between testosterone, criminal behaviour and aggressively dominant behaviour in prison (Dabbs & Hargrove, 1997).

Nonetheless, the research base appears be correlational and therefore the functional relationship between testosterone and aggression/violence is not yet fully understood.

Further, high levels of testosterone has also been linked with alcoholism (Stalenheim et al., 1998) and substance use is a risk factor for offending which may explain the why violent offenders may have higher levels of testosterone.

Moving away from the biological determinism approach, Pollak (1950), in his seminal book *The Criminality of Women*, offered a more integrated approach to explaining female crime which proposed a gender-neutral and gender-specific approach. He asserted that female crime occurred owing to a combination of biological, psychological and societal factors. He outlined two positions. Firstly, he suggested that women may commit as much crime as men but the frequency of female crime is underestimated. He suggested that this may be due to the victims more likely to be children; the victims do not complain, or if the victims are adult males, their ego prevents them from reporting the crime. The second approach, which could be classified as gender-specific, suggested that women's criminality is different in that it is also deceitful and therefore less easily detected. Further gender-specific theories examine female crime as a societal process. Here the focus was on female crime resulting from emancipation (Bishop, 1931; Thomas, 1923). As a result of emancipation, women were felt to be exposed to greater economic distress and increased opportunities to commit crime (Steffensmeier & Schwartz, 2004). Within this view less emphasis is placed on an individual and the psychological factors for female crime and more on the social construct of female crime (Morash, 2009). However, these theories have been criticised for being too narrow in focus, reliant on a patriarchal view of

offending where women do not conform to prescribed gender roles and do not offer an inclusive approach to account for the differences in criminal activity, of which gender is only one (Belknap & Holsinger, 2006; Kruttschnitt & Carbone-Lopez, 2006).

Some researchers have sought to explain the disparities through outlining differential motivations underpinning male and female offending. For instance, Nicholls, Greaves and Moretti (2008) highlighted that the type of violence perpetrated by women and the function of female offending is unique. Zaplin (2008) suggested the motivations that influence the nature and timing of female offending are different to men. It has been suggested that different methods of socialization may account for the differences in male and female crime. Here boys are encouraged to act more assertively whereas girls are encouraged to bond with others (Bowie, 2007). Adult females have been found to be less likely to describe themselves in terms of their individual characteristics, and more likely to describe themselves in terms of their relationships with others (Cross & Madson, 2007). Some have suggested that lower rates of female offending are attributable to females acquiring social cognitive skills earlier in life which can be attributed to socialization as well as increased interhemispheric communication, fewer frontal lobe deficits and greater verbal ability (Bennett, Farrington & Huesman, 2005). This leads to greater empathic understanding, perspective taking and prosocial reasoning which are emphasized more in the socialization of girls in comparison with boys.

Gender-specificity debate

Feminist theory has driven research into a more gendered approach to the assessment and management of female offenders (Belknap & Holsinger, 2006; Daly, 1992; Reisig, Holtfreter, & Morash, 2006; Simpson, Yahner & Duggan, 2008). This posits the need for a

'gendered' approach to the assessment of, and interventions for, female offenders. Here emphasis was placed on the role of victimization, relationship difficulties, mental health difficulties and poverty as salient factors for female offending. Blanchette (2002) asserted that low self-esteem, victimization in childhood and/or adulthood and self-harm and/or suicide were female specific criminogenic needs. This was supported by research among female prisoners in the UK which emphasised the importance of personal/emotional factors in predicting recidivism among female offenders (Hollin & Palmer, 2005).

Caulfield (2010) suggested that such factors represent distinct criminogenic needs for female offenders and there should be a shift from focusing on gender-neutral factors (i.e. factors within traditional risk assessments) to factors that take account of these (i.e. gender-specific/gender responsive). However, this does not necessarily equate to unique risk factors for female offending. As such while the risk factors for male and female offenders may be similar the expression of these in women may be different. Therefore, factors that are relevant to both men and women (e.g., mental health and/ or relationship instability) should be considered differently for both male and female offenders (van Voorhis, Wright, Salisbury, & Bauman, 2010). This is also reflected in research in the UK which found that data from the Offender Assessment System (OASys) identified that 59% of women, in comparison to 29% of men, had problems with relationships including adverse childhood experiences, poor family relationships and abuse (MoJ, 2009a). Furthermore, trauma is considered a risk factor for offending in risk assessments (e.g. HCR-20^{V3}, Douglas, Hart, Webster & Belfrage, 2013) however, the prevalence of trauma in women appears to be higher. With respect to victimization, 53% of women in prison, in comparison with 27% of men, reported having experienced emotional, physical or sexual abuse as a child (MoJ, 2012). In the UK twice as many women as men reported being a victim of non-sexual partner abuse, and women were seven times as likely to have

experienced sexual assault as men (MoJ, 2014b). As such, while a relationship between traumatic experiences and later offending exists for both men and women, this relationship appears to be stronger for women (Belknap & Holsinger, 2006). In a review of young female offenders, Chesney-Lind, Morash & Stevens (2008) found that female offenders were likely to abuse substances to cope with experienced victimization and engage in criminal behaviours such as prostitution and involvement in drugs in order to survive. This then makes them vulnerable to further victimisation which reinforces the cycle of victimisation and antisocial behaviours. Furthermore, the relationship between victimisation and violent offending can be mediated by factors such as depression; anxiety; self-harm and substance use difficulties (Underwood, Kupersmidt & Coie, 1996).

Bonta, Pang and Wallace-Capretta (1995) conducted a study correlating female criminogenic risk factors with reconviction and found that whilst some of the risk factors were the same for men, the strength of the relationship between the factors for men and women appeared to differ, with, educational levels and criminal associates playing a less important role for female offenders. Study of the Offender Assessment System (OASys) in 2000 found that drug misuse, followed by thinking style, relationships and attitudes were the most common criminogenic risks for women and criminal history, followed by thinking skills, attitudes, drug misuse and antisocial lifestyle associated were more common for men (HM Prison Service, 2013).

Although the difference in the risk profiles of male and female offenders has yet to be fully explored, policy makers are recognising the importance of adopting a gender-responsive approach to the assessment and intervention for female offenders. This is reflected in the publication of the Corston Report (Home Office, 2007) and the English and Welsh Government Green Paper which asserted that female offenders may have a different profile of risks (Ministry of Justice, 2010). Furthermore, the National Offender

Management Service (NOMS), in 2008, put in place gender-specific standards for women's prisons; which have been reviewed in 2012 and 2014 (MoJ, 2014a). The standards included adapting services to support women who have been abused, raped, experienced domestic violence or been involved in prostitution. Although there have been some developments in the literature on female offending (Blanchette & Brown, 2006; Chesney-Lind & Pasko, 2013), there has been little advancement within risk assessment practice with male-oriented tools being applied to females. While professionals now have access to a range of risk assessments which are designed for specific settings, specific populations and for specific forms of violence (Hart & Logan, 2011), there have been few tools developed and validated for use with female offenders. This could be argued as being a systematic bias within the risk assessment field.

Risk, Needs and Responsivity (RNR)

The RNR model of offender rehabilitation outlines three principles aimed at reducing reoffending: the risk, need and responsivity principles (Andrews & Bonta, 2003). The *risk principle* asserts that the intensity of offender programs should match the offender's risk level with higher risk offenders attending programs of higher intensity. The *need principle* contends that the treatment programs should reduce dynamic risk factors, which are empirically associated with recidivism (for example pro-criminal attitudes). Finally, the *responsivity principle* informs the delivery of intervention programs appropriate for the individual. Risk is viewed as an indicator of clinical need and the level of risk co-varies with the depth of criminogenic needs which subsequently inform intervention (Ward et al., 2007). As such accurately assessing risk is important to ensure adequate management

strategies are put in place to monitor and reduce this risk. However, in practice there are a number of issues with assessing risk accurately.

Assessing risk- premises, principles and practices

Three generations of risk assessment tools are identified (Bonta, 1996). The first generation is referred to as ‘unstructured clinical judgement’ (UCJ) whereby risk was predicated on the expertise of professional experience and knowledge. However, studies have found that this form of prediction is no better than chance in predicting future risk (Hanson & Bussière, 1998), as risk assessment professionals were found to be accurate in only one of every three predictions of violent behavior among psychiatric populations (Monahan, 1986). Furthermore, the process at which a decision was made was unclear and it was found to have low reliability and validity (Monahan & Steadman, 1996). Further, the UCJ approach is not necessarily replicable or testable, a key standard necessary for an evidenced informed approach to practice (Sackett & Rosenberg, 1995). For instance, two clinicians could assess the same individual with a particular risk profile and arrive at different conclusions regarding the level of risk they pose (Heilbrin, Yasuhara & Shah, 2010).

The lack of empirical support for isolated clinical predictions of violent behaviour led to the development of the second generation of risk assessment tools, commonly referred to as actuarial risk assessment (Bonta, 1996). Decisions on risk are made according to fixed and explicit rules based on a small number of factors which have been demonstrated statistically to be associated with the outcome (Meehl, 1954). Monahan (1986) contended that predicting future offending can be enhanced with the use of actuarial risk assessments. They are also cost effective and may be useful for settings

which demand resource efficient tools (Craig & Beech, 2010). However, actuarial measures have been found wanting with respect to predicting violence, demonstrating low levels of accuracy. Hart, Michie & Cooke (2007) conducted an evaluation of the ability of actuarial tools to predict violence. The authors found that the 95% Confidence Intervals (CIs: range of scores where an individual's true score may lie [Warner, 2008]), were so wide for individual risk assessment that actuarial tools could not predict risk with any degree of certainty. As such the authors advised that extreme caution should be taken when using these measures as their predictive accuracy may be too low to support their use when making management decisions about individual offenders (Hart et al., 2007).

Actuarial measures also have little utility in predicting fluctuating violence potential (Skeem, Monahan & Mulvey, 2002). Violence and its future likelihood, is not static, it is reliant on a combination of key risk factors including the individual's adaptiveness, resistance to change and his/her intentions (Scott, 1977, p. 128). This is also known as dynamic risk. When predicting violence among psychiatric patients, dynamic variables related to risk are considered to be crucial in predicting future violence (Ogloff & Daffern, 2006). Interventions guided by actuarial measures which fail to take account of dynamic risk may lead to prescribing unnecessary restrictive interventions across the course of their detention (restraint, seclusion, increased observation) which may be unachievable, unjustified and perhaps unethical (Ogloff & Daffern, 2006). In clinical settings practitioners also need to know how to manage and supervise an offender. Thus a probability estimate is of little clinical value as the items used to derive an individual's risk do not need to be casually related to risk nor emphasise characteristics that may moderate potential treatments of such behaviour (responsivity factors). Therefore, actuarial measures have little utility in supporting agencies to *manage* risk.

These issues led to the development of the third generation of risk assessment tools, Structured Professional Judgment (SPJ; Bonta, 1996). Such tools follow an idiographic approach to risk assessment (Craig & Beech, 2010). The SPJ guidelines reflect conceptual developments within the field of violence risk assessment which highlight the importance of other features of risk such as the nature, severity, imminence of violence (dynamic risk) as well as how to manage this (Ogloff, 2006, p.3). There is an increasing body of evidence which supports the SPJ approach to risk assessment as being reliable and valid (Otto & Douglas, 2010). It is also the recommended approach by the DoH in the UK (2007). The SPJ approach has been criticised for a number of reasons. These include adopting a deficits focused approach to the assessment process thereby neglecting protective factors through focusing on the negative characteristics of an individual, and failing to underpin assessments by theories on the causal nature of antisocial behaviours (de Vogel, 2005).

Singh, Grann and Fazel's (2011) meta-analysis results suggest that tools designed for specific populations have greater predictive potency than tools designed for a broad range of offenders/offence type. Therefore, using tools for specific populations and/or specific types of offences is recommended (de Vogel et al., 2012, 2014). Further, improving risk assessment and management in female offenders also has public health implications. For instance, research has demonstrated an intergenerational transfer of risk of violence between mothers and children, and mothers who have a history of violent offending are more likely to raise aggressive children (Kim, Capaldi, Pears, Kerr & Owen, 2009; Motz, 2001; Serbin et al., 1998).

Importance of risk assessment

Accurately assessing risk is important to determine the most appropriate risk group an offender should be placed into, and subsequently the management and treatment approach suitable to address that risk. Risk assessment is a guiding influence in sentencing practices, release decisions and effective correctional programming in terms of risk levels and treatment needs. Therefore, effective risk assessment is crucial in clinical practice not only for decision making but also in supporting practitioners to identify the most appropriate intervention to reduce risk of reoffending (Craig, Browne & Beech, 2008).

Assessing risk of harm can also inform subsequent management plans. Although risk assessment and risk management are conceptually distinct they can be considered to be related. For instance, violence risk assessment and violence risk management are different steps, phases or aspects of the same general process (Douglas et al., 2013). The goal is to understand what kinds of violence an individual may perpetrate, as well as helping to develop intervention plans to prevent future violence from occurring.

It is also important for the offender being assessed, as risk assessments can influence sentence length, treatment selection, release from custody and supervision within the community (Craig et al., 2008). Further, if an offender is wrongly classified into a risk level this can have important implications for the offender, the staff managing the offender as well as the public. If an offender is wrongly classified as high risk this may put unnecessary restrictions on their freedom. If they are wrongly classified as low risk this may lead to fewer risk management strategies put in place and increased opportunity to offend putting both staff and the public at risk (Johnstone, *n.d.*).

Aims of thesis

The aim of the present thesis is to explore the assessment of risk in female offenders.

Although many scholars have criticised services for not adopting a female-centred approach (Pollack, 2005), others have also suggested that there is a dearth of research to support the adoption of ‘gender-specific’ approaches to risk and treatment (Heilbrun et al., 2008; Zaplin, 2008). This thesis hopes to contribute to the research base on the ongoing debate on the suitability of gender-neutral or gender-specific risk assessments. It has been suggested that the absence of risk assessment tools which have been validated on female offenders has significant implications for the prevention of future female crime (de Vogel & de Vries Robbé, 2013). However, it is not clear if gender-specific risk assessments would be warranted given the confused literature base on whether female offenders have unique and different risk factors in comparison to male offending. As such the thesis aims to:

1. Explore the suitability of male-oriented risk assessment tools for female offenders;
2. Examine the validity of a female specific risk assessment tool for violence in inpatient female offenders;
3. Compare male and female specific risk assessments of violence;
4. Contribute towards resolving the debate on the necessity of adopting a gender-responsive approach to assessing risk in female offenders;

To achieve these aims a systematic review of empirical research on the validity of risk assessment tools for female offenders was undertaken. This is presented in Chapter Two. Given the sparse literature base on female offending, the review examines risk assessment tools for predicting reoffending and future violence. Systematically examining the extent to which current risk assessment tools available adequately assess risk of reoffending within female offenders and the quality of research within this area is explored. Chapter

Three presents an empirical research paper on the predictive validity of a recently published female specific violence risk assessment: the HCR-20 FAM (de Vogel et al, 2012; 2014). It also evaluates Version 2 and Version 3 of the HCR-20 and compares the validity of the each risk assessment in predicting inpatient violence. This was presented at the Division of Forensic Psychology (DFP) Conference in 2015. Chapter Four presents a critical review of the HCR-20^{V3} (Douglas et al., 2013). This is a widely used violence risk assessment tool within correctional and mental health services (Douglas & Reeves, 2010). The review explores the psychometric properties of the HCR-20 through a critical review of the empirical research on the reliability and validity of the HCR-20^{V3}, evaluating its strengths and limitations. Chapter Five concludes the thesis by drawing together the work presented through a discussion of the main findings and considering the implications of the thesis for future research and clinical practice.

CHAPTER TWO

The predictive validity of risk assessment tools for female offenders: a systematic review.

Paper Published:

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

<http://dx.doi.org/10.1016/j.avb.2015.01.002>



The predictive validity of risk assessment tools for female offenders: A systematic review



Kate Anya Geraghty*, Jessica Woodhams

School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

ARTICLE INFO

Article history:

Received 11 January 2014
Received in revised form 15 December 2014
Accepted 6 January 2015
Available online 14 January 2015

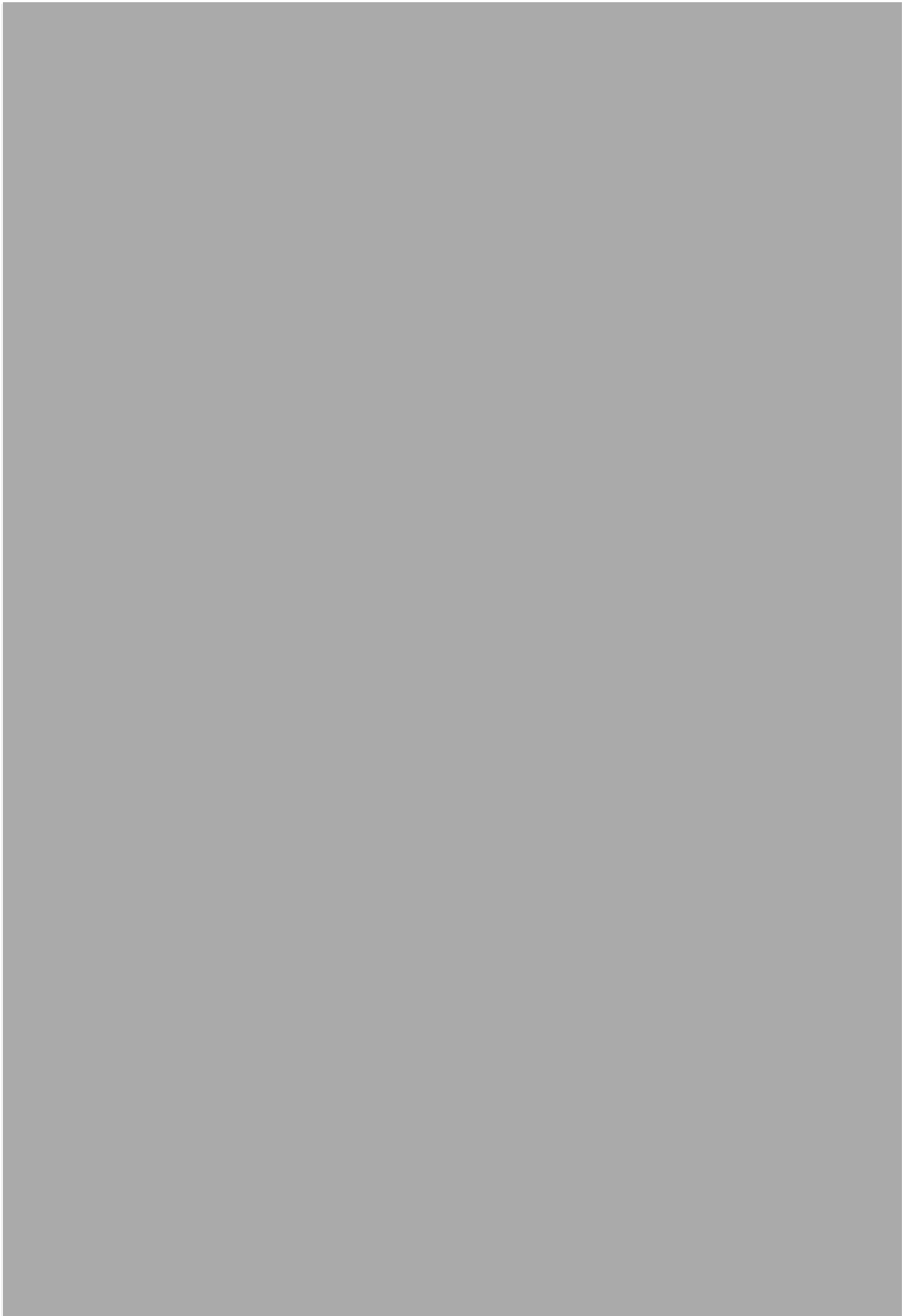
Keywords:

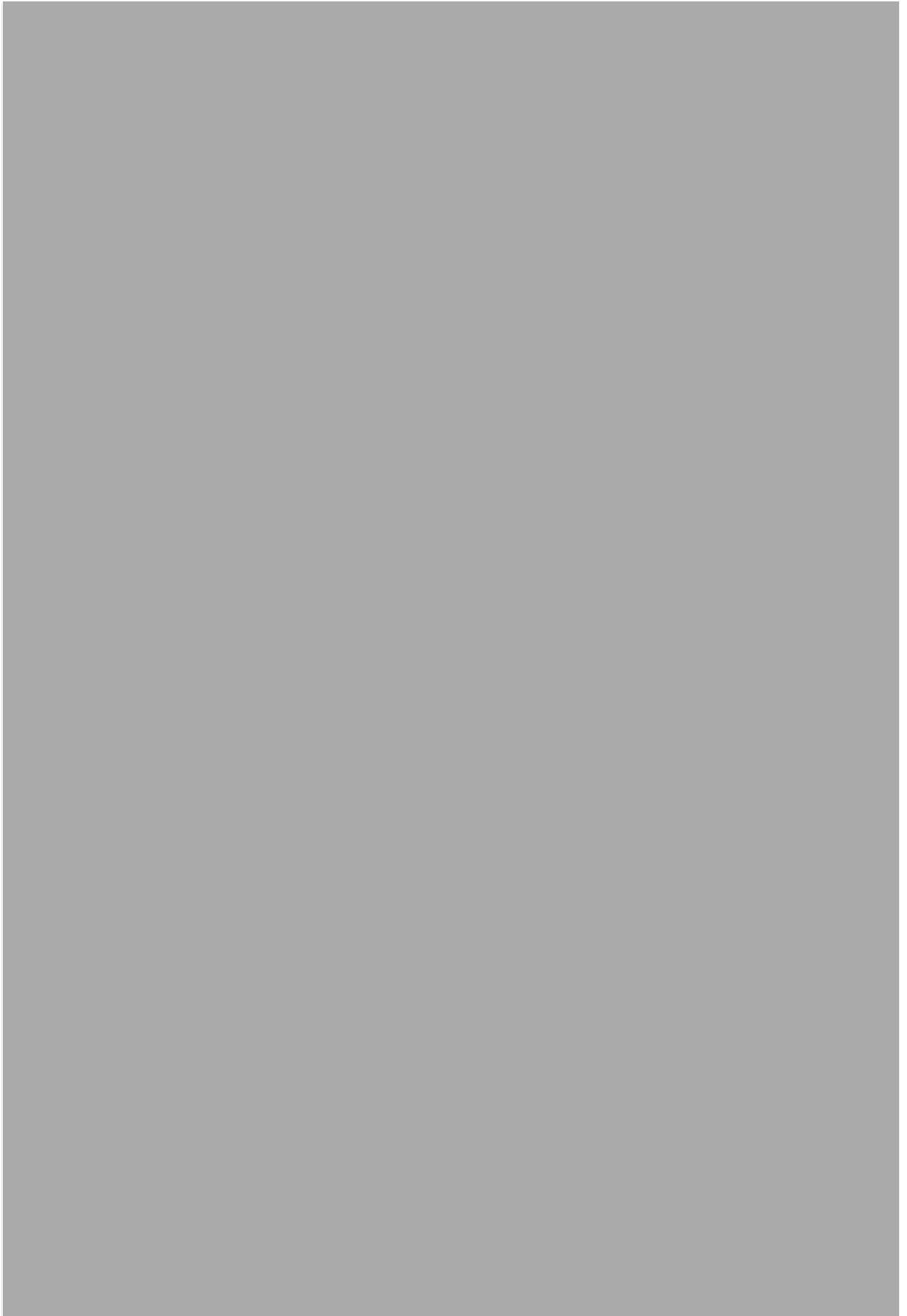
Risk assessment
Female offenders
Predictive validity

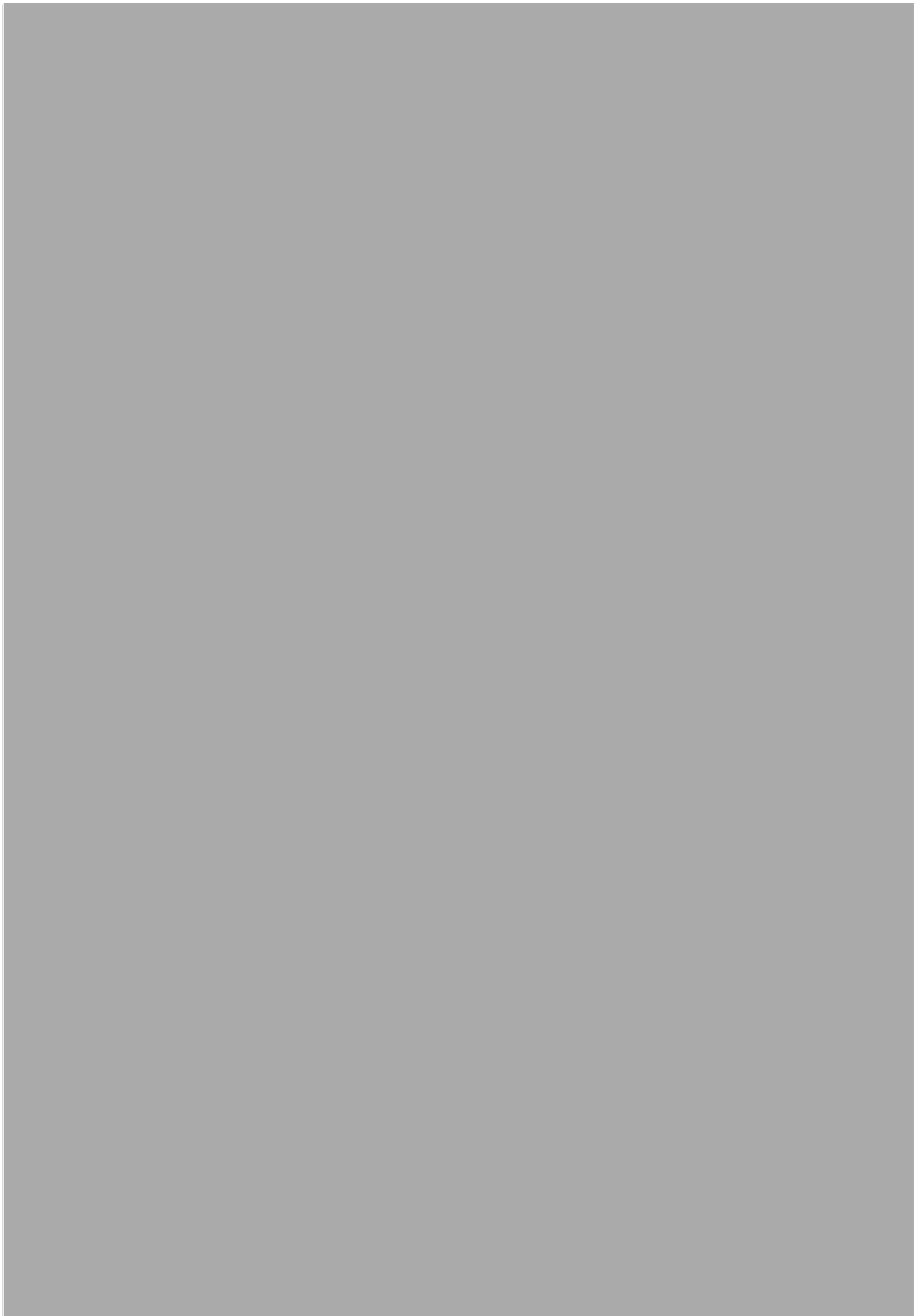
ABSTRACT

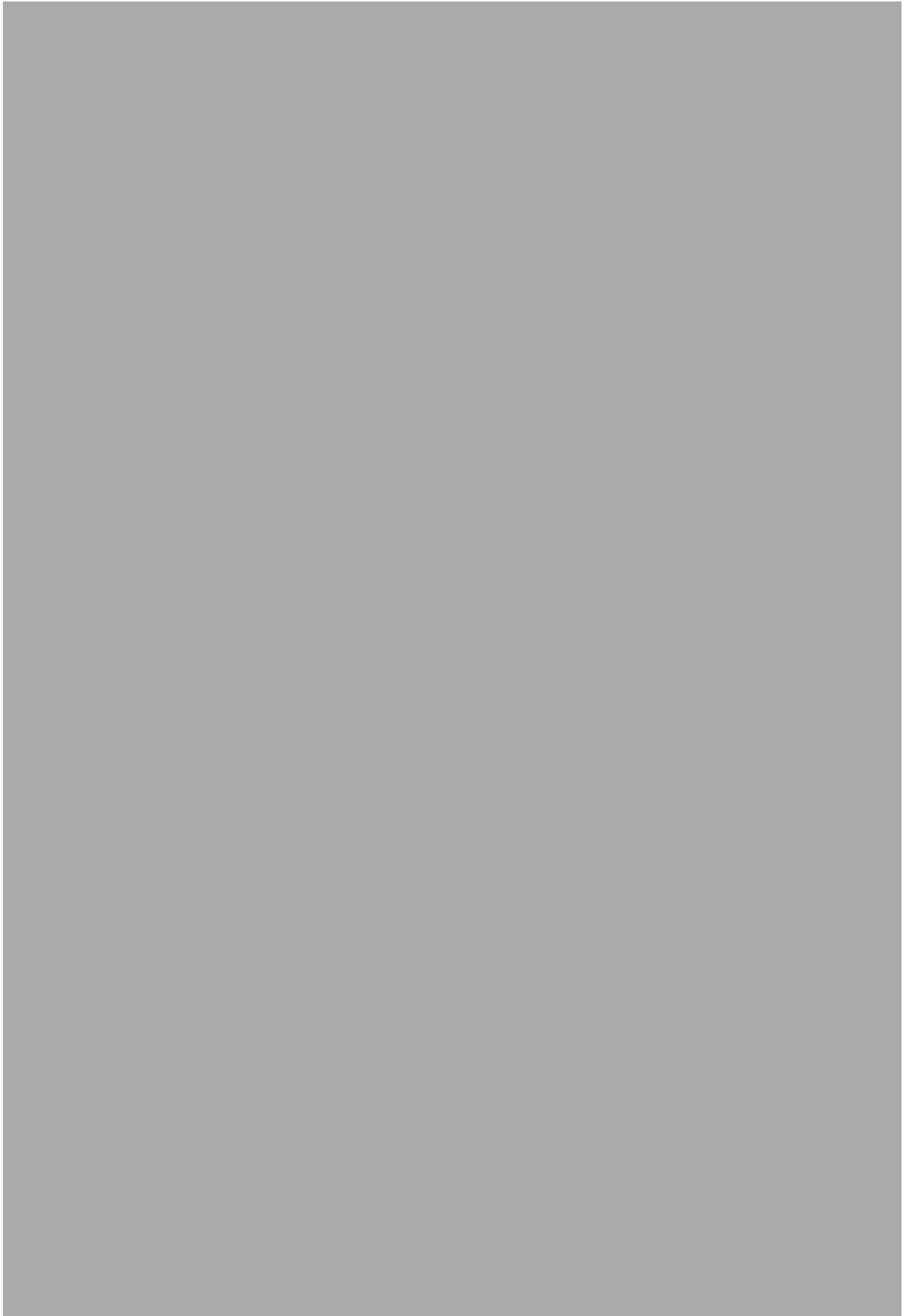
Assessing an offender's risk level is important given the impact of criminal behavior on victims, the consequences for the offender, and for society more generally. A wide range of assessment tools have been developed to assess risk in offenders. However, the validity of such tools for female offenders has been questioned. We present a systematic literature review of studies examining the accuracy with which risk assessment tools can predict violence and recidivism in female offenders. Five databases were searched, reference lists of relevant publications were hand searched, and an online search engine was used to identify studies. Fifteen studies were subject to review which evaluated nine risk assessment instruments (COMPAS, CAT-SR, HCR-20, LSI, PLC-R, OGRS, RISC, RM2000V, VRAG). The quality of these studies was systematically examined using a detailed quality assessment. The review findings indicate that the most effective tool for assessing both violence and recidivism in women was the LSI. There was variability in the quality scores obtained, with studies limited by measurement issues and standards of reporting results. Future research should aim to improve the quality of studies in this area, assess predictive accuracy across *subtypes* of female offenders, and compare correctional and psychiatric samples independently.

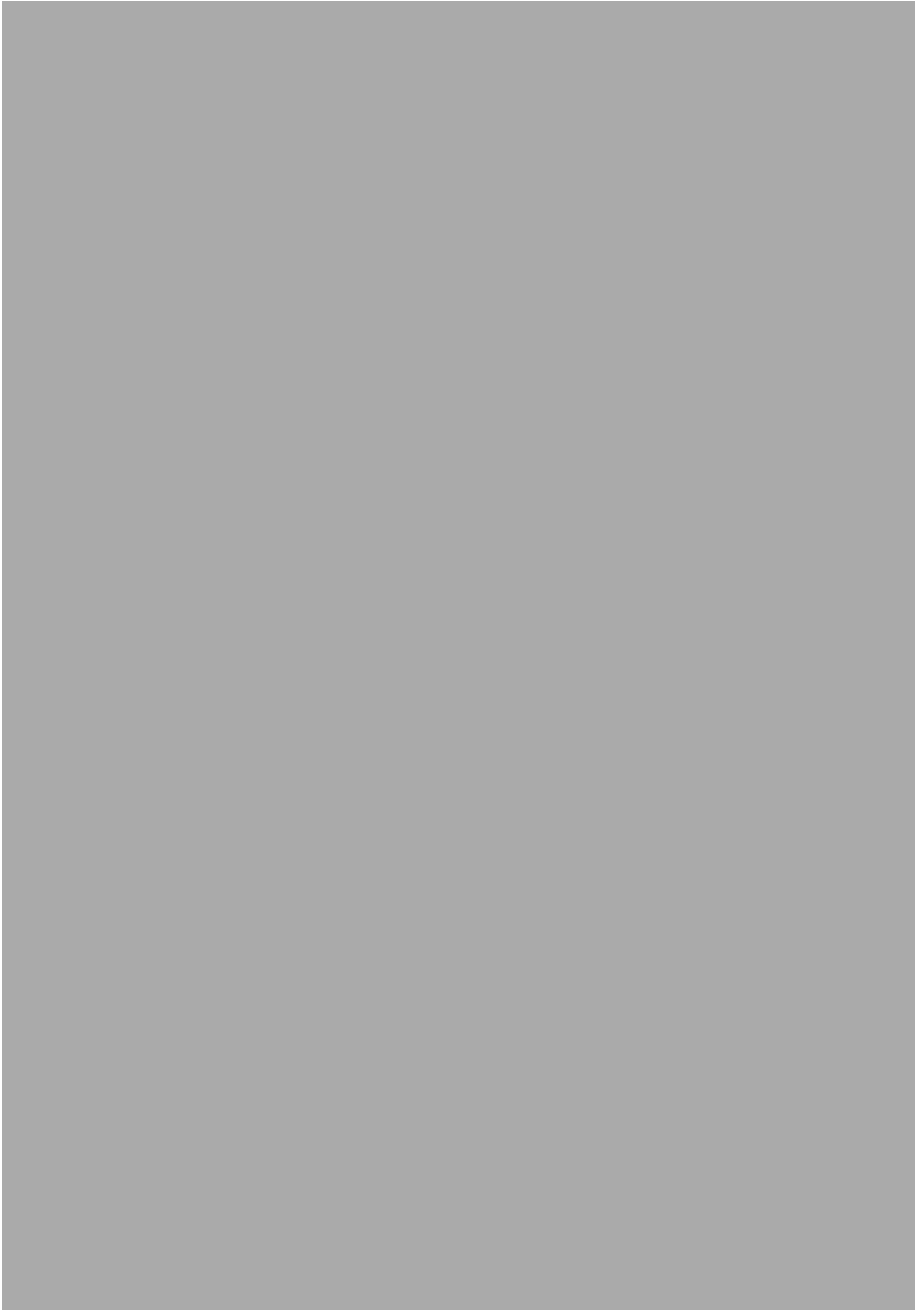
© 2015 Elsevier Ltd. All rights reserved.

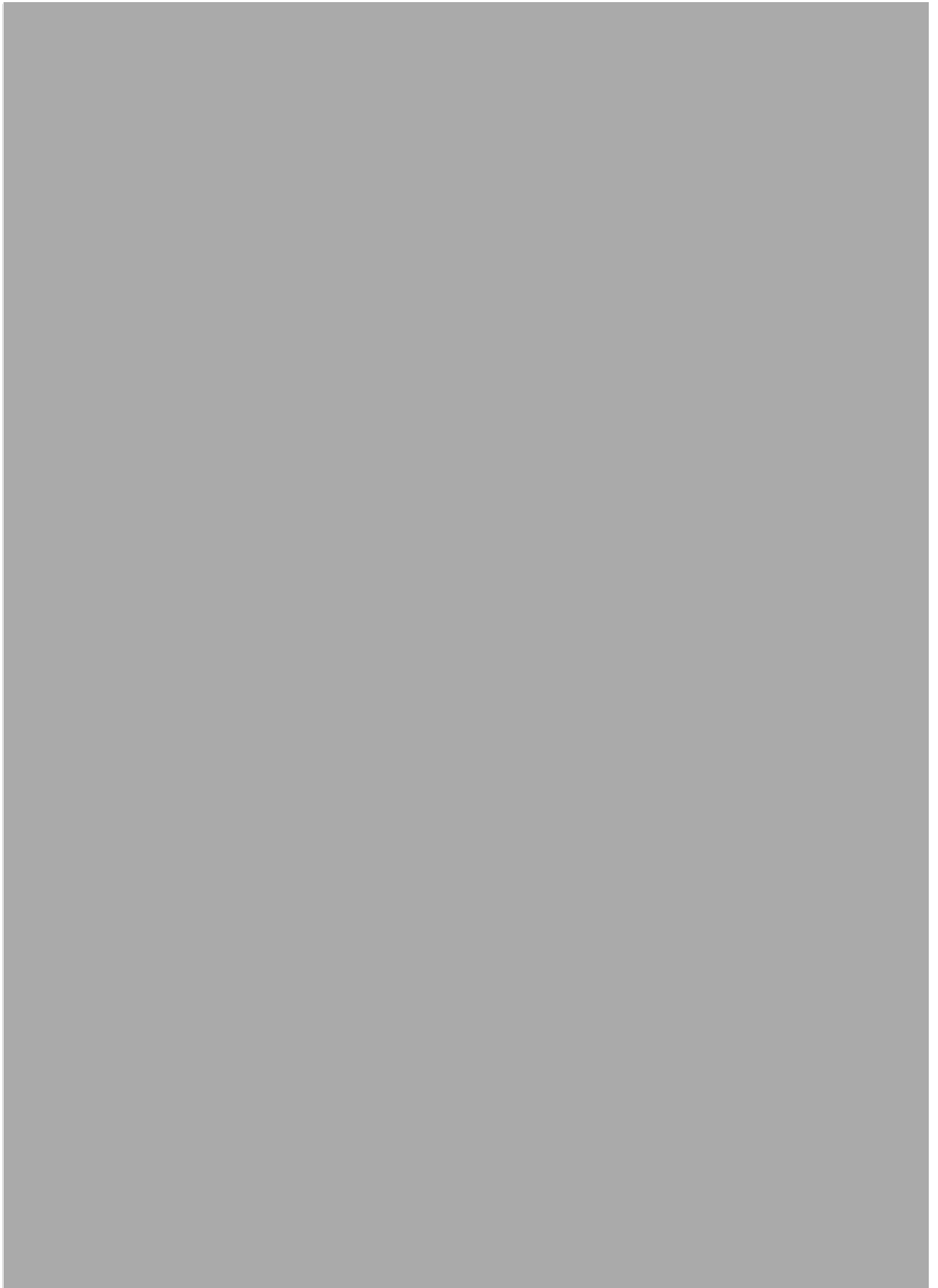


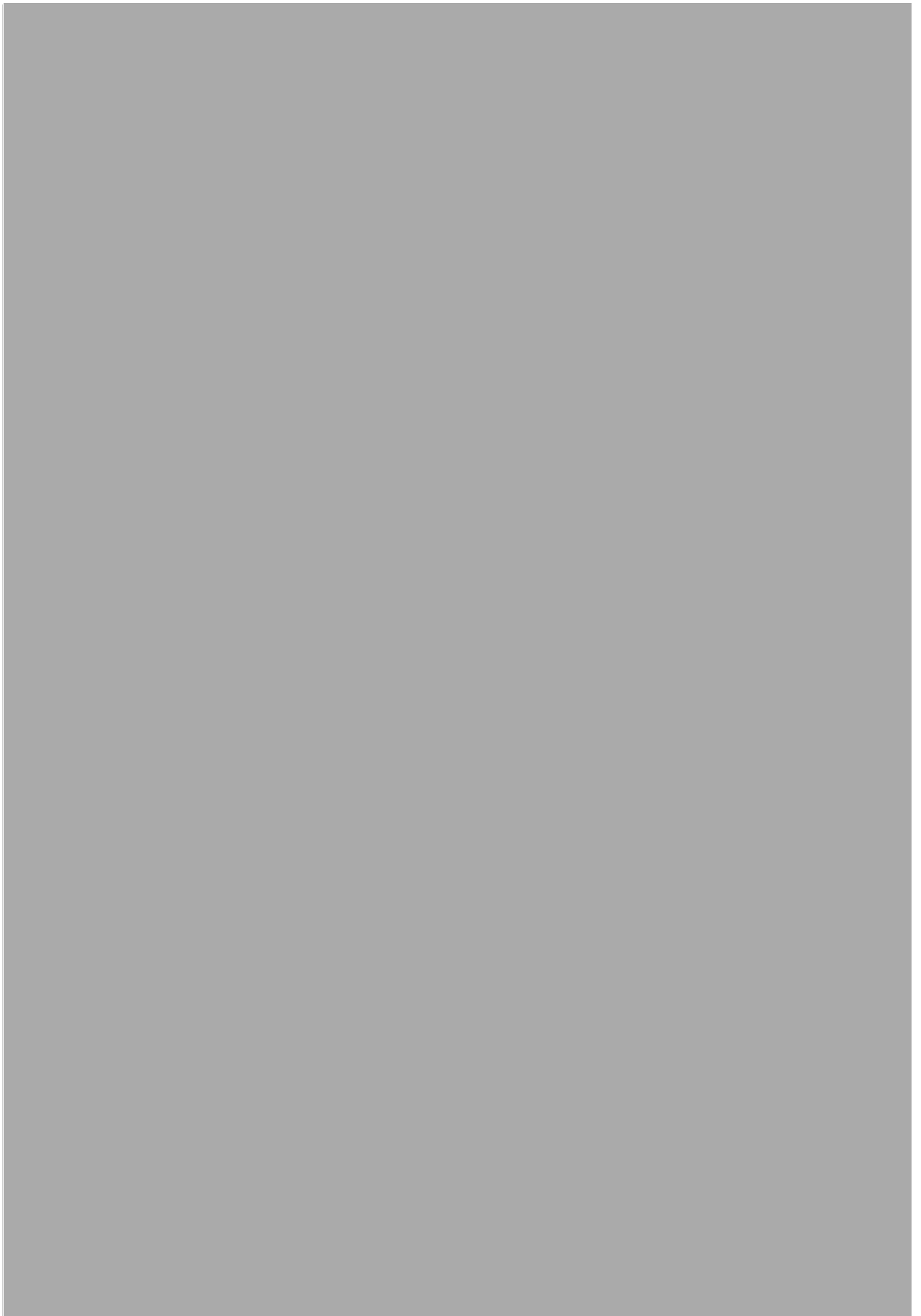


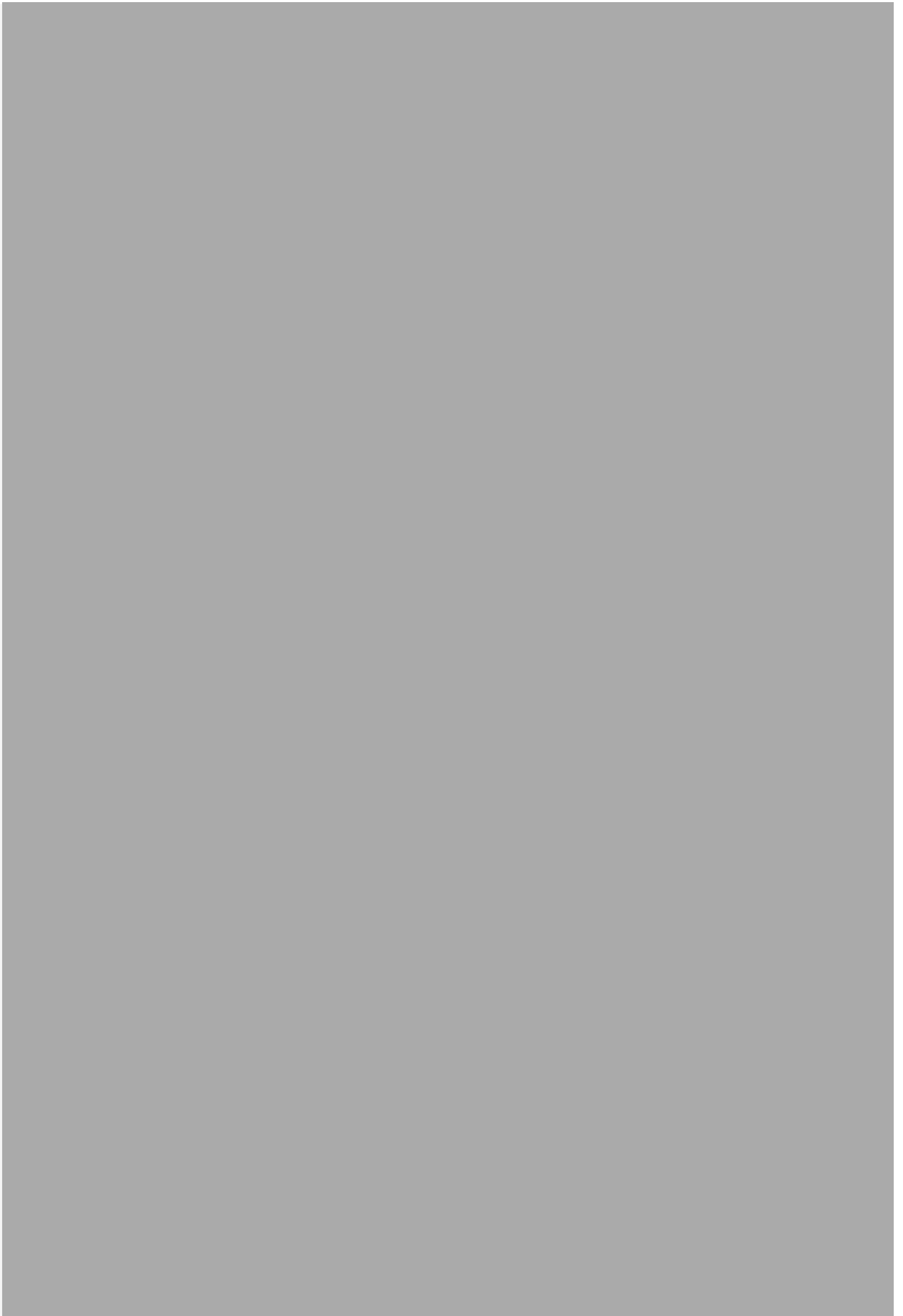


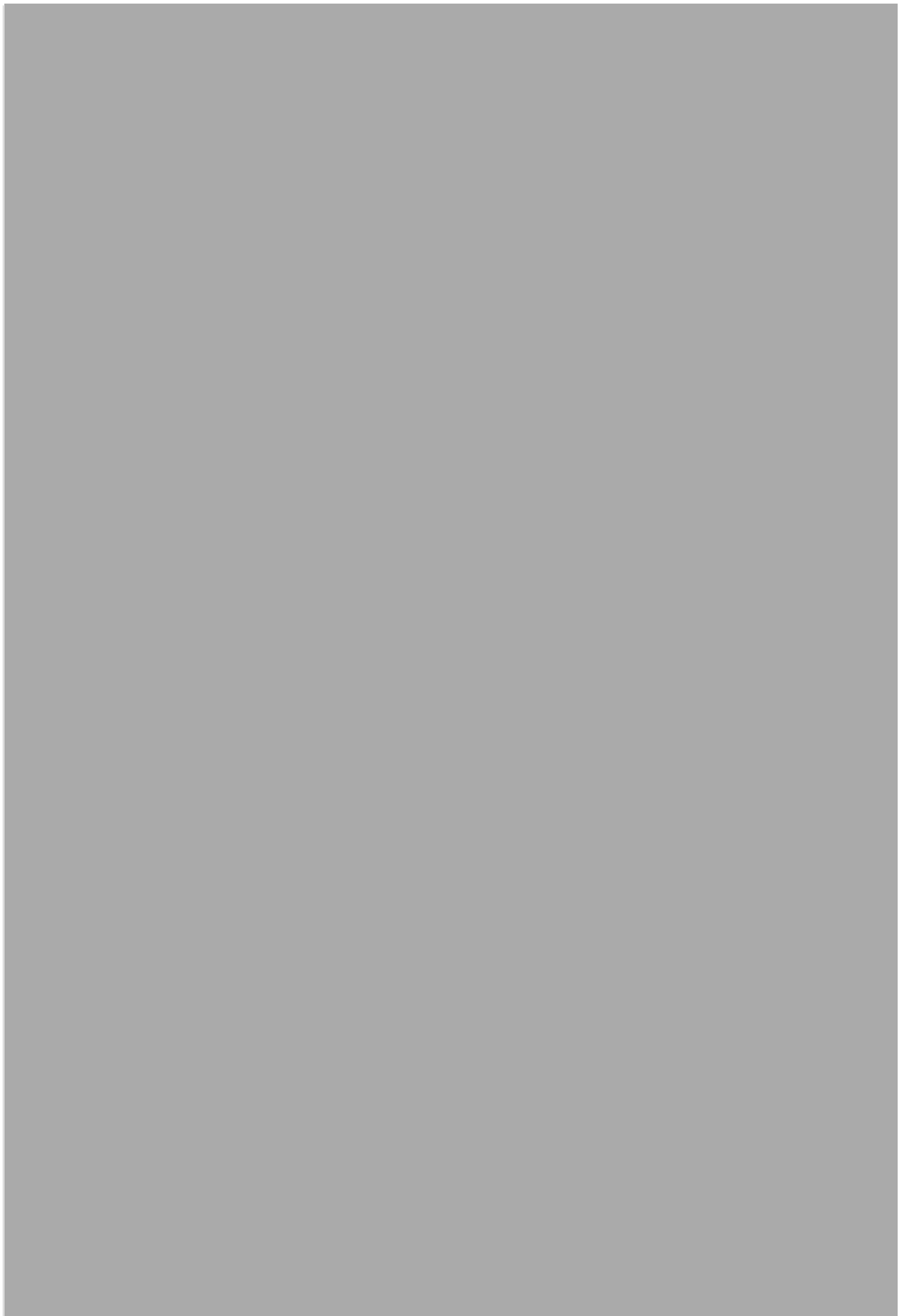


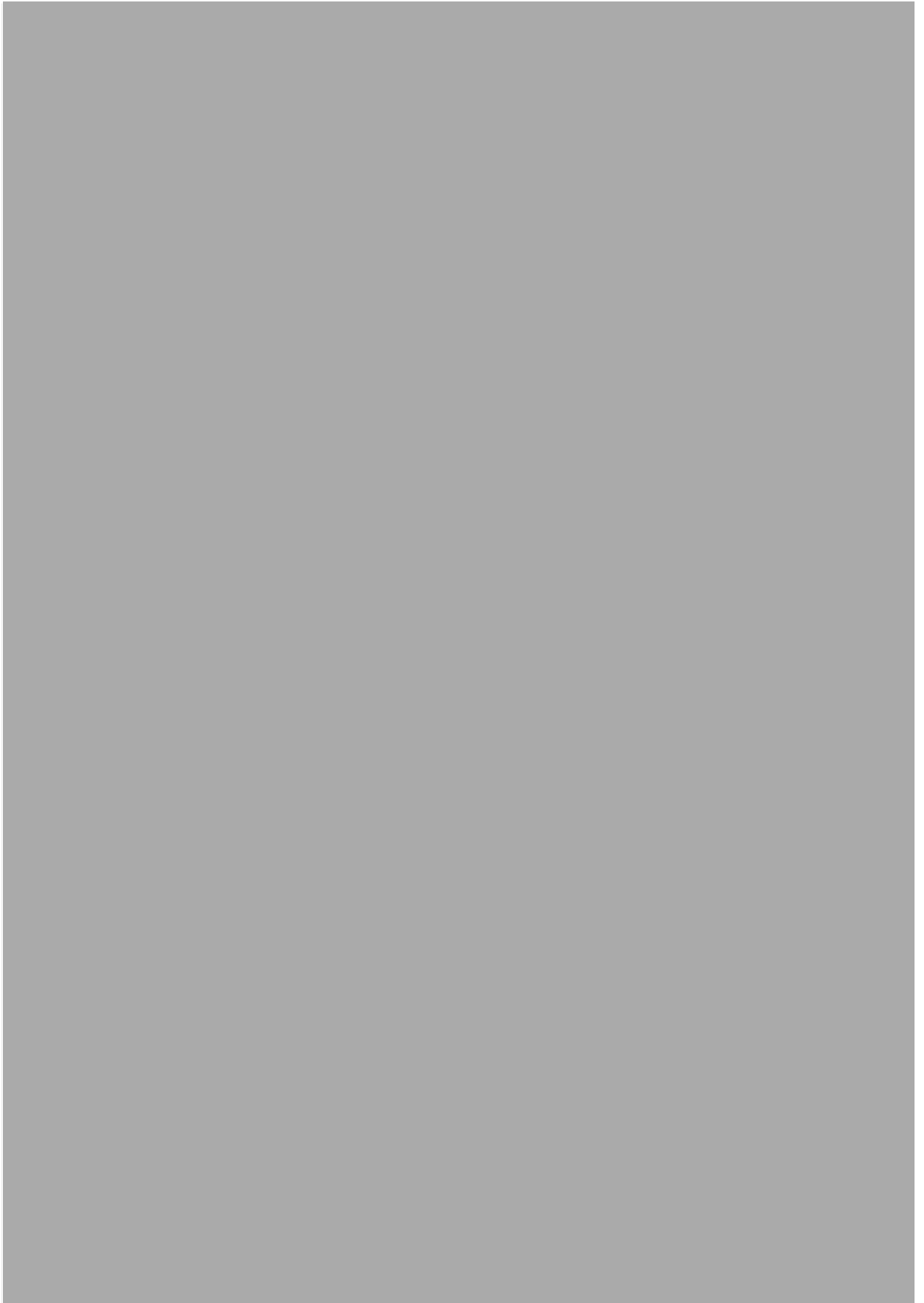


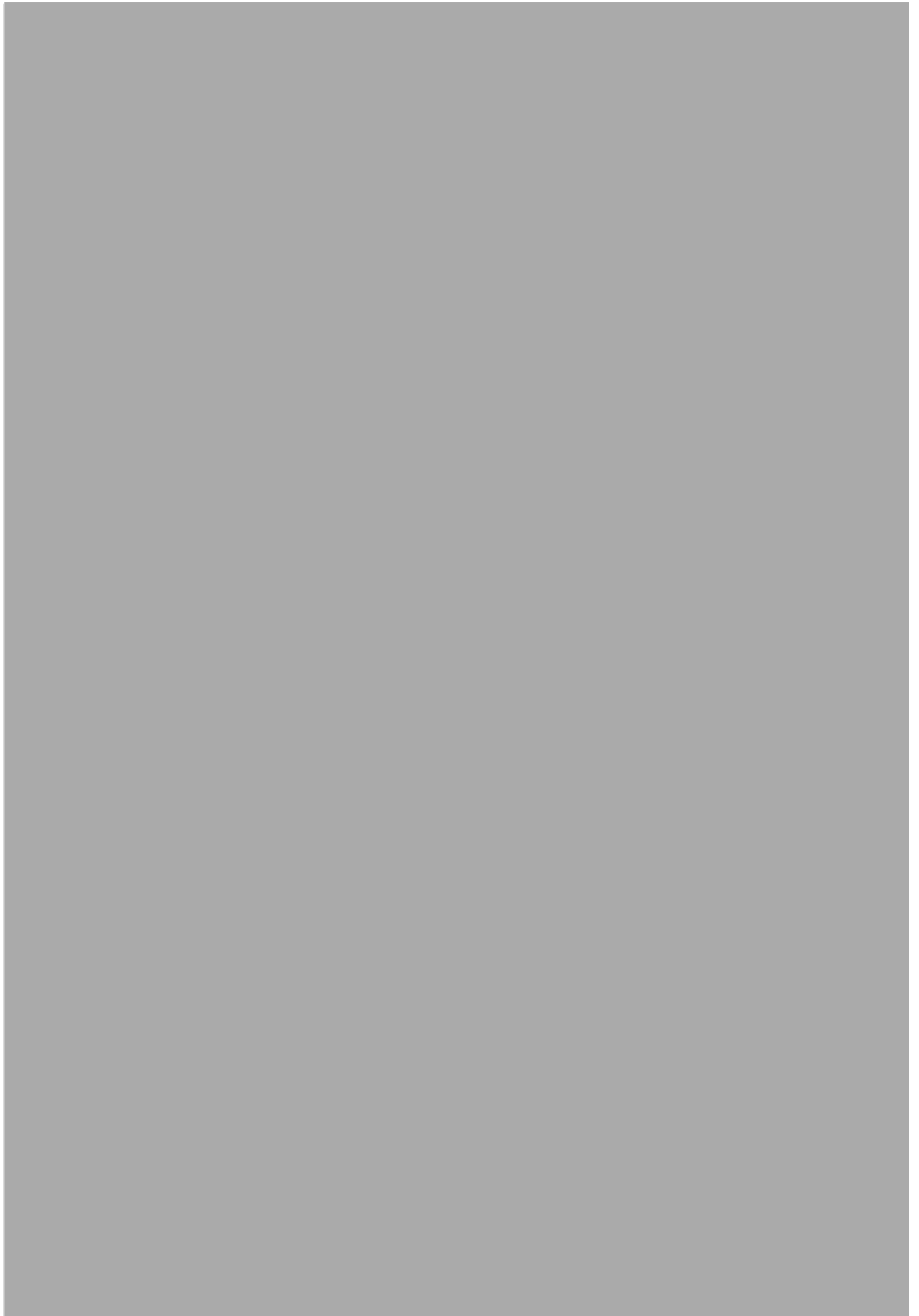


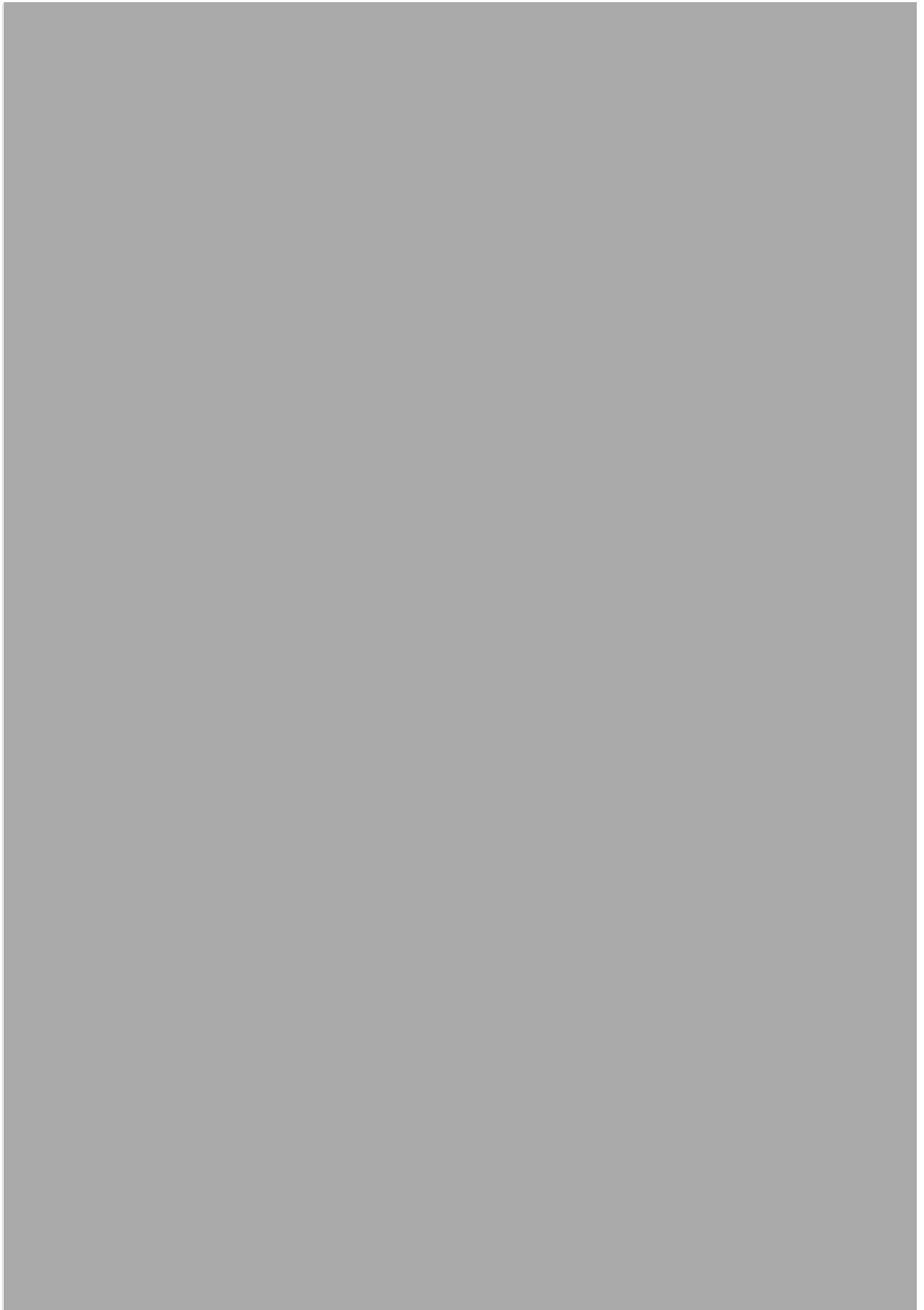














CHAPTER THREE

Critique of a psychometric measure: Examining the psychometric properties of the Historical, Clinical and Risk Management (HCR-20) Version 3 violence risk assessment scheme

Abstract

This paper examines the psychometric value of the HCR-20^{V3}. Kline's (1986) criteria of the characteristics of a good test were used to assess the tools standardisation, reliability and validity. Further, external validity issues are incorporated through assessing how pragmatic the HCR-20^{V3} is. The HCR-20^{V3} appears to meet some criteria for reliability and validity and possesses strong clinical utility given the emphasis on formulation of violence risk. The extent to which the HCR-20^{V3} meets scientific criteria for a sound psychometric critique will need to be further assessed following the accumulation of further peer reviewed literature as well as the publication of independent studies. Additionally research will need to determine its reliability and validity on correctional samples. Implications for practice are also considered.

Introduction

This review critically examines the most recent version of the Historical, Clinical and Risk Management Violence Rating Scheme, HCR-20 Version 3 (hereafter referred to as HCR-20^{V3}), developed by Douglas, Hart, Webster and Belfrage (2013). It is an example of the Structured Professional Judgement (SPJ) approach to risk assessment (Bonta, 1996). Risk assessment guides sentencing practices, release decisions and effective correctional programming through identifying an individual's level of risk and treatment needs. SPJ tools of violence risk have strong clinical utility as they provide direct guidance for decision making in terms of risk management (Hart & Logan, 2011). Thus, rather than focusing solely on prediction SPJ tools emphasis prevention through determining the nature, severity, imminence, frequency, duration and how likely future violence would be. It achieves this through speculating about violence that individuals may plausibly perpetrate (i.e formulation). Management plans are subsequently developed to prevent violence.

The HCR-20^{V3} is an assessment of risk of interpersonal violence and can be used to measure the extent of risk reduction across a range of forensic settings including prison and secure mental health and community settings (Fazel, Singh, Doll & Grann, 2012). The HCR-20 has been used to inform decision making for correctional and psychiatric release, admission to forensic facilities, as well as monitoring risk whilst in the community or within secure settings. As such evaluating risk of interpersonal violence is considered a central task for professionals to help understand and mitigate violence (Douglas et al., 2013). It is also used as an outcome measure within forensic settings. Assessing risk of violence is highlighted as a core task within the standards and policies of forensic practice in the UK (Department of Health [DoH], 2007). It is also used as a key performance indicator within the Commissioning for Quality and Innovation (CQUINS) framework which seeks to ensure optimal standards in the quality of healthcare provision (DoH, 2008).

This chapter provides an overview of the HCR-20^{V3} rating scheme, however, its primary aim is to present a critique of its scientific and psychometric properties. Its applicability to correctional, forensic mental health and community services as well as its research utility is also examined.

Overview of Tool

The HCR-20^{V3} is a standardised schedule for the assessment of risk of future interpersonal violence. It is made up of twenty items which have been shown by research (Douglas et al., 2013) to be associated with risk of future interpersonal violent behaviour. It is completed by professionals who have a high level of knowledge and expertise of violence, including the professional and scientific literature on the nature, causes and management of violence. Further, assessors should have training and experience interviewing clients and reviewing clients' case history as well as in the administration and interpretation of standardised tests. It is also advocated that assessors have experience in the assessment and diagnosis of mental, personality and substance-related disorders (Douglas et al., 2013, p.38).

Development

The first version of the HCR-20, HCR-20^{V1}, was published in 1995 (Webster, Eaves, Douglas & Wintrup, 1995). It was one of the first examples of the SPJ approach to risk assessment and offered a manualised and structured framework to guide clinicians when making decisions about risk of future interpersonal violence. HCR-20^{V1} offered practitioners a systematic manner to assess risk which was grounded in the empirical base but also enabled flexibility whereby a clinician's view could be taken into account (Webster et al., 1997). This was subsequently revised in 1997 to produce HCR-20 Version 2, hereafter abbreviated HCR-20^{V2}, (Webster et al., 1997). Langan and Ludlow (2004) conducted a research study

about the perceptions of risk assessment among mental health professionals. The report found that risk assessment was emphasized at the expense of risk management. Nevertheless, meta-analytic reviews have demonstrated the merits of HCR-20^{V2} through assessing its ability to predict violence when compared to other tools such as the VRAG and the PCL-R (Fazel et al., 2012; Guy, Douglas & Hendry, 2010; Yang et al., 2010). To improve the clinical utility of HCR-20, HCR-20^{V3} was developed. Changes were made to enhance decision making, formulation of risk and risk management as well to revise items in line with the most recent developments in the violence risk assessment literature. Violence under the HCR-20^{V3}, consistent with previous versions, is defined as “actual, attempted or threatened infliction of bodily harm to another person”. (Douglas et al., 2013, p.2).

Items, scoring and uses

Items within the HCR-20^{V3} were selected using a rational or logical approach. They are based on risk factors identified by epidemiological and clinical research, which, according to the authors, establish the validity of including such factors (Douglas et al., 2013). In determining the relevant risk factors the authors extensively reviewed the research base. The risk factors in HCR-20^{V3} are consistent with the earlier version of the HCR-20, which were chosen based on factors empirically linked with violence. The items are consistent with major studies in the area, including the MacArthur Violence Risk Assessment Study (Monahan et al., 2001) which highlights the significance of clinical factors in the prediction of future violence including substance misuse and symptoms of mental disorder. A draft of HCR-20^{V3} was produced and presented to professionals in the field which led to the development of Draft 2. Both Draft 1 and Draft 2 were then pilot-tested in different countries to determine their clinical utility and feedback was obtained. This then led to the final version of HCR-20^{V3} which like earlier versions has twenty risk factors for assessing violence risk: ten historical;

five clinical and five risk management items. Further, unlike earlier versions of the HCR-20, HCR-20^{V3} also codes for not only whether a risk factor is present but also whether it is relevant to the individual. It also yields a 'Summary Risk Rating' which is an estimate of an individual's overall level of risk. For a list of full items please see Appendix III.

An issue which has dominated the measurement of personality is the debate of bandwidth fidelity (Cronbach & Gleser, 1965). This relates to breadth of construct measurement. For the purposes of illustration, if applied to the context of violence risk assessment, this would relate to whether items that are more broadly defined are better at predicting violence than more narrowly defined items. In a study of personality traits, Ones and Viswesvaran (1996) have argued that broader items are more likely to provide the most accurate prediction in construct measurement as they are more likely to include the range of behaviours in the outcome being assessed. One of the motivations for updating the HCR-20 was to be more inclusive about the criteria being assessed. Although the number of items within the assessment remains the same, it could be argued that the breadth of construct measurement has increased due to the inclusion of sub-items within HCR-20^{V3}, as well as more guidance about the types of incidents that may qualify as a risk factor.

In terms of scoring, the HCR-20^{V3} does not follow traditional theories of psychological testing. It does not yield a composite score. However, this is not necessarily a weakness of the tool. Classical test theory of measurement focuses on an individual's overall test score where it is inferred that the higher the test score the higher the level of clinical problem or attribute. However, although respondents may achieve the same overall score, their responses to the construct may be different, and it does not inform the clinician about what kinds of violence, at what frequency and under what circumstances may be perpetrated. Individuals may be at risk of perpetrating violence for different reasons and relying on

scoring protocols does not adequately account for the idiosyncratic nature underlying violence risk.

As the HCR-20 is an 'observer rating scale' it is free from issues such as social desirability which plagues self-report measures. Nevertheless, it does depend on how the individual assessor interprets each item, the information used to inform their coding, assessing what items are more relevant than others. Further the individual's overall level of risk is based on their own formulation (Nevatti, 2011). Arguably this can be seen as a limitation as individuals with varying levels of experience may score items differently or place more emphasis on different risk factors. Therefore, the responsibility lies with the clinician to ensure that they are aware of the empirical base and that they incorporate specific aspects of an individual's functioning that are relevant to their violence risk. Whilst assessors are required to provide justification for their scoring, the framework could be said to leave too much room for assessor error which undermines the reliability and the validity of the assessment through introducing bias.

Level of measurement

A ratio scale is deemed to be the most ideal form of measurement as it has a true zero point. However, the majority of psychometrics do not meet this; as such interval scales are considered to be most appropriate for psychometric measures (Kline, 1998). In terms of the HCR-20, numerical scoring is not advocated by the authors, although for the purposes of research codes are converted into numerical scores for evaluation. As such, it may be considered to be an interval scale. However, even with the application of numbers this interval scale is artificially imposed. For instance, an individual with a higher score is not necessarily more likely to be at an elevated risk of violence. Rather the focus is on the constellation of risk factors, including dynamic changeable factors, which are indicative of

risk. This is determined through formulation and supports assessors to consider idiosyncratic factors that may be overlooked through prioritizing 'scores'. Kline (1998) contends that once the test is of practical utility the assumption of an interval scale is reasonable. From a clinician's point of view the HCR-20 framework is inherently practical as it accounts for dynamic risk which subsequently enables the development of management plans to reduce violence risk.

Characteristics of a good test

A psychometric test will be justly described as a good test if it has at least an interval scale; is reliable, valid, discriminating and has appropriate norms (Kline, 1986). Thus from a measurement perspective we want to know if the test is accurate, standardised, valid; reliable; and free from measurement and predictive bias (Shultz & Whitney, 2005).

Reliability

The reliability of a test refers to its tendency to be consistent over time from measurement to measurement. It can be determined through three measures: test retest reliability, internal consistency and inter-rater reliability (Kline, 1998). If a test is shown to be reliable then any difference in an individual's scores can be attributable to changes in the individual rather than attributable to the test. For instance, the HCR-20^{V3} may be applied to the same person pre- and post-treatment to identify any reduction in the risk of interpersonal violence. Test retest reliability assumes that when the same test is applied to the same person on two occasions there will be no substantial change in their scores (Kline, 1998). Internal consistency refers to the consistency between the items within a test and whether they measure the same thing. A common coefficient to determine the internal reliability of an assessment is Cronbach's Alpha (α ; Kline, 1998). The fluid nature of some of the items within the HCR-20, particularly the Clinical and Risk scale items, which can fluctuate rapidly depending on an individual's

presentation make it difficult to assess its retest reliability. Furthermore, it is argued that the HCR-20 does not measure a 'psychological construct' as the items (i.e risk factors) within the test do not contribute to form a construct of 'violence' in the same way other psychometric tools would. Rather, each item could be viewed as a construct. As such internal consistency is suggested as being less of an issue as it would be for other psychometric tools (Douglas and Reeves, 2010, p.162).

Inter-rater reliability refers to the variation across different assessors using the same test. Generally, a minimum correlation coefficient of .70 is considered necessary for a test to be deemed to have good inter-rater reliability. This would indicate that different raters agree on how the test was applied to particular individuals (Kline, 1998). Inter-rater reliability is asserted as being the most important aspect of reliability for the HCR-20 (Douglas & Reeves, 2010). In terms of published literature which has assessed inter-rater reliability, one study thus far has reported coefficient estimates for the three subscales (evaluating both the presence and relevance of the scales) as well as for the summary risk ratings among a male and female inpatient sample (Douglas & Belfrage, 2014). With respect to the presence of risk factors across the scales, for single raters, coefficients of .94 were obtained for the presence of risk items and .62 relevance of items for single raters. For a group of raters, coefficients of .98 for the presence and coefficients of .93 were reported for a group of raters. In terms of the clinical scale, an average ICC of .86 was reported for the presence and estimates of .60 for the relevance among single raters. Coefficients of .95 were reported for the presence and .82 for relevance among several raters. Reliability for the risk scale was slightly lower for the presence of risk factors with coefficients ranging from .69 and .75 for presence and between .69 and .74 for relevance among single raters. Coefficients of .87 and .90 were reported for the presence of items and estimates of .89 and .87 for the relevance of items for a group of raters. With respect to the summary risk ratings, whereby an individual's overall level of risk

of violence is assessed, coefficients of .81 and .75 were found for single raters and estimates of .90 and .93 were reported for several raters (Douglas and Belfrage, 2014). Among a UK sample who discharged from hospital, the reported reliability coefficients for the Historical scale was .91, for the Clinical scale was .90 and the risk management scale was .93 (Doyle et al., 2014).

As such reliability for the presence of risk factors was strong across all the scales, although slightly lower coefficients were reported for the Risk scale. Furthermore, coefficients estimates were also higher across all scales for both the presence and relevance of items as well as the summary risk ratings than has been reported in a study by Douglas and Reeves (2010) which summarized the reliability estimates across studies assessing HCR-20 V². However, lower coefficients were obtained for the relevance of risk items with some items achieving poor or low reliability for relevance. These included past violence and antisocial behaviour (H1 and H2) as well as items that assess poor compliance or responsiveness (C5 and R4). The authors suggest that as the relevant ratings are intended to inform the formulation of an individual's risk (Douglas et al., 2013), based on models of varying theoretical orientation, the relevancy of certain risk factors may be weighted differently under the various formulation models (Douglas & Belfrage, 2014). Therefore, it may be that the subjective nature that underpins formulation may affect whether an item is assessed as relevant or not. Additionally, it may also be that assessing relevance may rely on greater clinical skill from a practitioner in understanding the client's behaviour. Others have suggested that perhaps some of the definition of the items within each of the risk factors may contribute to low reliability scores (Kötter et al., 2014). To overcome the potential reliability issues with assessing relevance, evaluators are encouraged to spend additional time evaluating the relevance of these factors to an individual's interpersonal violence (Douglas and Belfrage, 2014). There is an important consideration for professionals using the

assessment in practice. Given that higher estimates were found when a group of individuals score the risk items, including presence and relevance, practitioners would benefit from consulting other professionals to enhance the accuracy of their assessment.

In summary, it could be argued that the HCR-20 V³ appears to meet some of the criteria for the reliability of a good psychometric, given that test retest reliability and internal consistency cannot be reliably determined (Douglas & Reeves, 2010). Whilst encouraging estimates have been found for inter-rater reliability, it is noted that the reliability estimates yielded were based on hospital samples and therefore the applicability to correctional samples has yet to be determined.

Validity.

The validity of a test refers to the extent to which the test measures what it intends to measure (Kline, 1998). It can be assessed through various forms including concurrent validity, predictive validity, content validity and construct validity. Validity can be characterised as weak, acceptable or strong and as with reliability there are many forms of validity.

Criterion related validity refers to how well a test predicts the criteria being assessed (Kline, 1998). In terms of the HCR-20, as an assessment of risk, the assessor would therefore be interested in the extent to which the tool predicts the risk of interpersonal violence occurring. There are two forms of criterion validity concurrent validity and predictive validity. Concurrent validity refers to how well a test correlates with another validated measure assessing the same construct at the same time (Kline, 1998). For instance, the HCR-20 could be compared with another violence risk assessment tool, such as the VRAG. Predictive validity refers to how well a test predicts an event occurring. In the case of the HCR-20, how well the test predicts an event of interpersonal violence occurring would be assessed. Arguably, this form of validity may be seen as particularly important in the context

in which the HCR-20 is used: stratifying individual's in a risk level (low, medium, high) in order to inform appropriate risk management strategies for prison, forensic mental health and community settings.

As the purpose of the HCR-20^{V3} was to retain continuity with its predecessor HCR-20^{V2} (Douglas et al., 2013), and given that HCR-20^{V2} have demonstrated adequate psychometric properties, assessing concurrent validity of HCR-20^{V3} in the context of HCR-20^{V2} is considered as acceptable (Belfrage & Douglas, 2014). In comparing HCR-20^{V2} and HCR-20^{V3} high coefficients were found in the Historical scale ($r = .89$), Clinical scale ($r = .76$) and Summary Risk Rating ($r = .98$) (Strub et al., 2014). In a Norwegian sample of inpatients, correlations between HCR-20^{V2} and HCR-20^{V3} for the Historical and Risk scales as well as the total scale scores were high ($r = .85, .81$ and $.84$ respectively). A moderate coefficient was found for the Clinical scale ($r = .58$; Eidhammer, Selmer & Bjørkly, 2013). The HCR-20^{V3} manual also describes what empirical testing was undertaken with respect to concurrent validity referencing a series of unpublished studies. These have demonstrated moderate to high correlations between the subscales and summary risk ratings for V² and V³, r s ranging from $.67 - .91$ (Douglas et al., 2013, pp.25-26). All of the studies were based on forensic mental health settings with the exception of one study which was based on a mixed sample of psychiatric patients and criminal offenders. Separate coefficients for both samples were not reported within the manual. Essentially, this would indicate that individuals' who were of high risk on HCR-20^{V2}, are also likely to be high risk on HCR-20^{V3}.

In terms of comparing HCR-20 V³ with other measures of violence one unpublished set of data is referred to within the HCR-20 manual. Smith and Eden (2013 as cited in Douglas et al., 2013) compared V³ to the PCL-R and found moderate correlations between the PCL-R and HCR-20 total scores, individual subscales (r s ranging from $.66 - .70$). Furthermore the data also revealed that Factor 1 (Affective/Interpersonal) was not strongly or

significantly correlated with HCR-20 ($r_s = .05$ -.21, $p > .05$), however, Factor 2 (Lifestyle/Antisocial) was ($r_s = .64$ -.83, $p_s < .001$). Further, in terms of HCR-20^{V2}, there is research which evaluated the concurrent validity of the HCR-20 through comparing it with the PCL-R (Gray et al., 2003). Nonetheless, although the PCL-R has been found to be predictive of violence (Coid et al., 2009), the purpose of the tool is to determine the presence of psychopathic personality disorder. Using a tool not designed to measure violence to compare a violence risk assessment test is arguably questionable as it does not meet Kline's (1998) concurrent validity criteria for comparing tools which measure the same construct. Other tools which assess violence include the Violence Risk Assessment Guide (VRAG; Quinse et al, 1998; 2006), Spousal Assault Risk Assessment (SARA; Kroop et al.,1994), Sexual Violence Risk (SVR)-20 (Boer et al., 1997) and Violence Screening Checklist (VSC; McNeil and Binder, 1994). However, none of these tools are directly comparable with the HCR-20 framework. The VRAG follows the actuarial approach and uses conviction rates and specific forms of violence to determine a violent outcome, the SARA and SVR-20 are measures domestic violence and sexual violence respectively, and thus not the same 'construct' as violence defined under the HCR-20. The VSC has been designed for use on psychiatric populations only. This makes assessing concurrent validity difficult and arguably a weakness in the validity of the HCR-20.

Within the risk assessment literature, predictive validity is typically assessed using Receiver Operating Characteristics (ROC) Area Under the Curve (AUC) estimates. AUCs of $\geq .70$ are considered to be statistically acceptable for determining how well a tool predicts the outcome (Mossman, 1994). In terms of the HCR-20^{V3}, Doyle et al. (2014) conducted a prospective study evaluating the predictive validity of the tool for community violence among a group of mental health patients discharged from hospital. They found that the HCR-20^{V3} significantly predicted community violence at six months (AUC = .73), and twelve

months (AUC = .70). Further, the Clinical scale demonstrated the strongest ability to predict violence at both six months (AUC = .75) and twelve months (AUC = .71). This would support previous research on HCR-20^{V2} which found that changes in the items within the Clinical scale, and Risk scale, scores were predictive of violence (Michel et al., 2013). The HCR-20 total scores were evaluated for predictive validity in Doyle et al.'s study (2014), however, the Summary Risk Ratings (SRRs) were not. Arguably, this is a weakness of the study given the emphasis by the authors on the importance of risk formulation and encouragement to use the SRRs when assessing an individual's risk of violence. In terms of unpublished literature, Douglas and Strub (2013 as cited in Douglas et al., 2013) found that the SRRs were significantly predictive of violence in psychiatric and correctional samples (AUC = .73 and .68 respectively). Further, the study reported that the SRRs were equally predictive between the psychiatric and correctional sample as well as across gender. The manual also references other unpublished literature on earlier versions of the HCR-20^{V3} which demonstrated moderate to large levels of predictive validity (Douglas et al., 2013). Therefore, on the basis of research thus far predictive validity appears to be at levels similar to those in meta-analytic estimates for risk assessment effect sizes (Guy, 2008; Singh et al, 2011; Yang et al., 2010). However, there has been debate concerning how well AUC estimates assess predictive validity. Singh et al. (2013) criticised AUCs for not adequately assessing all aspects of predictive validity, specifically 'calibration' which estimates how well an instruments' prediction of risk coincides with observed risk (as oppose to categorising those who do and do not engage in a particular incident i.e. violence). It would be beneficial for all aspects of predictive validity to be assessed to determine how well the HCR-20^{V3} predicts violence.

Content validity assesses the magnitude of expert agreement regarding how well the item content matches the construct being assessed. This could be argued as being acceptable

for HCR-20^{V3}. As highlighted in the description of the selection of items within the framework, drafts of the HCR-20^{V3} were presented to experts in the field and feedback obtained led to two early drafts before the final version of the assessment was published (Douglas et al., 2013). Face validity is concerned with how well a test measures what it is claiming to measure. Kline (1998) suggests that a test is said to have face validity if those undertaking the test find it acceptable. In terms of the HCR-20 it could be argued that the test appears to measure various forms of interpersonal violence and the items are defined in terms of what the violence literature suggests. Further, within HCR-20^{V3} the definition of violence is more clearly described and there is further guidance on the individual items in comparison to V² to enhance its usefulness for assessors. This makes it easier for assessors to score and interpret the test by improving the decision making process in HCR-20^{V3}. Construct validity refers to how well a tool measures all aspects of the subject being assessed. In terms of the HCR-20 it would refer to how well the HCR-20 measures all aspects of interpersonal violence. A particular difficulty when assessing the construct of violence is that we can never be fully confident of the specific conditions under which violence will or will not occur given the idiosyncratic conditions under which violence occurs. Nevertheless, content validity is a particular strength of the HCR-20 as it includes not just acts of physical violence, but also assesses violence where there is an intention to cause physical or psychological harm (Douglas et al., 2013). Further, the authors have extensively reviewed the empirical base with respect to violence and the items within the tool have been revised. This was guided by the available scientific and professional knowledge to understand individual's potential for engaging in violence in the future and what plans can be put in place to prevent them from doing so. For instance, exposure to trauma as an adult can elevate an individuals' risk of violence (Swanson et al., 2002). Whilst this was not accounted for in HCR-20^{V2}, under HCR-20^{V3} assessors can now evaluate how victimization after childhood may contribute to

violence risk. However, despite the advances of neuropsychology in highlighting neurobiological factors that may predispose individuals to violent behaviour (Creedon, 2009), the risk items do not account for all of these. For instance some developmental risk factors (such as foetal alcohol syndrome, foetal abnormalities, nutrition and brain injury) are not accounted for within the HCR-20 framework. Research has found that traumatic brain injury and nutrition deficiencies have been associated with greater violent behaviour (Liu & Raine, 2006; Williams, Cordan, Mewse, Tonks, & Burgess, 2010).

Further, the HCR-20^{V3} does not have theoretical validity. A theory is an explanation of the causal mechanisms of the problem to help understand it, how specific events happen and why people behave the way they do (Ward, Polaschek & Beech, 2006). The HCR-20 is not underpinned by any theory, rather, it is an amalgamation of factors which have been associated with future risk of violence. Determining risk of violence using the HCR-20 is thus atheoretical. It may advise clinicians about an individual's risk level, but it does not inform practitioners how and why a particular individual is at risk of perpetrating violence.

In summary, the HCR-20 appears to meet most criteria for validity. Given that the purpose of the tool is to predict an outcome where sentencing decisions and progress is determined, further peer reviewed studies assessing its predictive validity would be important. However, it is acknowledged that the measurement has only recently been published and it is anticipated that further studies will evaluate this in the near future.

Appropriate Norms

A criticism of the HCR-20 is the lack of reported norms on which the assessment was developed. This is important as in the absence of a normative sample drawing meaningful conclusions from the test becomes problematic. Norms provide a basis on which test scores can be compared. As such it has been argued that without appropriate norms the HCR-20

cannot be deemed to be a standardised assessment (Nevatti, 2011). The HCR-20 was designed for use with offending populations. The scoring criteria for some of the items (such as history of problems with violence, difficulties with treatment/supervision response and future problems with professional service and plans) are not suitable for assessment on non-criminal populations. Further this may be reflective of the idiographic approach the assessment takes. The authors have stressed that the assessment was not designed as a formal measurement tool and emphasized the importance of formulation within the assessment to account for interindividual variability. Further, Kline (1993) has asserted that standardization becomes less of an issue where individual differences are assessed in measurement.

Conclusion

The HCR-20 does not market itself as a psychometric tool and as such it may appear unusual to evaluate the psychometric value of the framework. However, it has psychometric properties of validity and reliability embedded throughout the assessment. Kline's (1986) criteria were used to evaluate the psychometric qualities of the tool which showed that HCR-20^{V3} meets some of the criteria for reliability and validity. Only one aspect of reliability can be assessed which highlighted that the accuracy of the assessment is improved when a group of people score the tool. The take home message for practitioners is that it is better to consult other professionals when scoring and assessing risk. Further, more research is needed to determine its reliability and consideration to measuring internal consistency despite the argument that interpersonal violence is not a 'construct' in the same way other psychological attributes are. Validity could be argued as being strong in some types of validity (content, construct, predictive). However, the HCR-20 does not have theoretical validity. There is no single theory of violent behaviour. As such the HCR-20 does not necessarily use a theory to identify the causal mechanisms of violence for the person being assessed. The authors of the

HCR-20^{V3} encourage assessors to ensure their formulation is theory driven (Douglas et al., 2013), however, theory does not underpin the HCR-20 assessment framework. While the risk factors are known to be associated to violence there is less known about the causal pathways to violence. The absence of research on causal pathways to violence has been recognised as a failing of the violence research base (Rutter, 2003; Moffitt & Caspi, 2001). Although it is acknowledged that it may be difficult to develop a formal theory of violence (McGuire, 2008) the advancements in the area of sexual offending (Ward et al., 2006) would suggest that this could be feasible. McGuire (2008) stressed the importance of being able to identify a pattern of offending using functional analysis or case formulation prior to determining an intervention approach. The authors of the HCR-20^{V3} also encourage the use of formulation to guide and structure violence risk management (Douglas et al., 2013).

Concurrent validity of HCR-20^{V3} is weak as it has only been assessed in the context of HCR-20^{V2}, and HCR-20^{V3} is an update to Version 2. Studies assessing reliability and validity thus far have focused predominantly on inpatient samples. It is unclear if similar estimates would be obtained among prison samples. Perhaps this is reflective of the realities of conducting research within forensic settings, whereby it may be easier to conduct follow-up studies within secure hospitals than within the prison setting or by following offenders post-release. Nevertheless, it is clear that further research among correctional samples is needed. Furthermore, studies assessing the basic components of reliability and validity of HCR-20^{V3} are also needed on males and females. This would also provide more evidence of the extent to which the tool meets the scientific criteria for a sound psychometric measure.

The perpetration of violence is a pervasive social problem (Hart & Logan, 2011) and therefore evaluating people to understand and mitigate risk for interpersonal violence is crucial. Using tools with strong psychometric properties is important to ensure their defensibility. Tools when developed and implemented competently can provide valuable

input into critical decision making. However, when poorly developed and/or implemented, tools can do more harm than good (Shultz & Whitney, 2005). However, the prediction of any form of human behaviour is an inherently difficult task (Buchanan et al., 2012). This becomes increasingly difficult when trying to predict the spontaneous actions of others. Pragmatism highlights the importance of the practical utility of the 'scientific method' stressing methodological diversity and pluralism (Baert, 2005; Bem & de Jong, 2006). Measurements need to be able to account for this. Best practice within violence risk assessment involves not only grounding the assessment within the research base but also ensuring assessment tools are valued by practitioners, supporting them to identify management strategies and monitor risk (DoH, 2007; Johnstone, n.d.).

As mentioned, assessing risk for interpersonal violence is context dependent. For instance the probability of an individual committing violence is uncertain, as well as the types of violence they may commit, the motivations for doing so as well as likely victims. Violence risk is both dynamic and contextual in nature as hazards arise and exist in specific circumstances (Douglas et al., 2013, p.4). Additionally, the risk for interpersonal violence may also be dependent on where the individual lives, the level of supervision they have, their motivation, adverse experiences as well as what treatment services they will have access to. The importance of an offender's environment in increasing or decreasing future risk of reoffending is increasingly being recognised. Boer (2015) highlighted how risk is in interactive function between offender's risk issues and the environment. He opined that practitioners should consider environmental strengths and risks within risk assessments and evaluate how an individual responds to risk opportunities. The HCR-20^{V3} allows for these aspects to be taken into consideration. Furthermore, from a practical point of view the test should also be cost effective and relatively easy to administer and score (Shultz & Whitney, 2005). HCR-20^{V3} provides professionals with sufficient guidance which, following training

and additional study time, renders it both cost effective and relatively easy to administer and score in practice.

Ultimately, the HCR-20^{V3} does not meet in full all the scientific criteria of a psychometric test as defined by Kline (1986). Nevertheless, it does reflect the pragmatic approach that should embed the study of human behaviour. If violence risk assessments were to strictly adhere to the scientific criteria of what makes a good psychometric measure, they would fall short of best practice guidelines (DoH, 2007; Morgan, 2013) placing less emphasis on formulation. It would thus fail to be practically useful for professionals and lack external validity. Risk assessment tools are developed to guide and support practitioners and adequately capture risk (Morgan, 2014). Traditional risk assessments have been driven by negativity and defensiveness and have thus shaped the design of tools to predict risk. Therefore, the HCR-20^{V3} can be viewed as sound psychometric measure as it balances between two opposing worlds, research and practice, and whilst it may fall short on some aspects of internal validity it does have external validity. The onus is now on practitioners to take these limitations into account when using and defending the tool and for professionals to conduct more research into its reliability and validity with greater attention to correctional populations.

CHAPTER FOUR

The predictive validity of a female-specific violence risk assessment: examining the HCR-20

FAM with female psychiatric offenders

Abstract

Adequately assessing risk of violence in females is critical to ensure public safety and appropriate management of violent female offenders. However, a limitation of violence risk assessment is the failure of risk assessment tools to incorporate research on the unique factors relevant to risk for women. The current study presents a validation study on the predictive validity of a violence risk assessment tool for female psychiatric offenders. The study seeks to compare gender-specific (HCR-20 FAM, De Vogel et al., 2012; 2014) and gender-neutral (HCR-20) risk assessment tools. The study also considers the HCR-20's most recent version, HCR-20^{V3} (Douglas et al., 2013). A prospective design is used with a 12-month follow-up to predict inpatient aggression in a psychiatric facility. A total of 66 participants were recruited to the study. Incidents of violence were recorded using the Overt Aggression Scale (OAS, Yudofsky, Silver, Jackson, Endicott & Williams, 1986) at both six and twelve month follow-up periods. The study also used the newly developed RAGEE checklist (Singh, Yang, Mulvey & The RAGEE Group, 2015) to ensure accuracy in reporting standards. Results indicated that the HCR-20 FAM was slightly better at predicting inpatient violence at both the six and twelve month follow-up periods in both HCR-20^{V2} and HCR-20^{V3}. The overall risk ratings (low, medium, high) were also found to be more predictive of inpatient violence than any of the risk assessment subscales. However, the addition of the female specific risk factors did not appear to improve the prediction of inpatient violence within the sample. HCR-20^{V3} did not appear to be a better predictor of inpatient violence in the study. Based on the results obtained, the study does not support nor refute the adoption of gender-specific violence risk assessment tools. However, it does suggest that continued research on the validity of female-specific risk assessments and continued investigation of the HCR-20^{V3} is warranted before meaningful conclusions can be obtained. Equally, developing a tool that demonstrates sound scientific properties and is applicable to female offenders is needed.

Literature Review

Presently, female offenders comprise 4.48% of the total prison population in the UK (MoJ, 2015). In 2008/2009 the most common group for which women were arrested was violence against the person (MoJ, 2009). In the US 5% of those sentenced for a violent offence are female (Guerino, Harrison, & Sabol, 2011). Further, females are reported to account for 10% of the population in forensic inpatient settings (de Vogel et al., 2014). It is suggested that the gender disparity between men and women across criminal populations, as well as violent offences, has led to a lack of empirically validated female-centred practices in violence risk assessment and management (de Vogel et al., 2014). Although women comprise a minority of the offending population, violence appears to be increasing in women (de Vogel & de Vries Robbé, 2013). However, it is acknowledged that this increase may be skewed by changes in policies and practices on the detection of crime or changes in tolerance of female crime (Hawkins, Graham, Williams & Zahn, 2009). Conversely, it has been shown that there is a tendency for the criminal justice system to treat women more leniently than men (Jeffries, Fletcher, & Newbold, 2003; MoJ, 2014b), which may underestimate the extent and rate of violence in female offenders. Furthermore, in female correctional facilities the largest offence group is women with violence offences (34%; Guerino et al., 2011). Additionally, the discrepancy between the prevalence rates of violence between men and women is not present within psychiatric populations, with women evincing comparable rates of violence to men (de Vogel & de Ruiter, 2005; Nicholls, 2001).

Violence

Violence is a pervasive social problem which takes many forms (Hart & Logan, 2011). In determining what violence is, this depends on the context in which the question is asked.

Legal definitions are used to define a wide variety of criminal violence (Ware, Ciepulcha & Matsuo, 2011). Violence in the present study will be defined according to the HCR-20 manual as “actual, attempted, or threatened harm to a person or persons” (Webster, Douglas, Eaves & Hart, 1997b, p. 24). The World Health Organisation (WHO) has estimated that the total cost of violence in England and Wales amounts to \$40.2 billion annually (World Health Organisation, 2004).

Inpatient Violence

Violence and aggression within inpatient settings is common, with the UK having one of the highest proportions of patients involved in violence (Bowers, 2011). Further, forensic settings have higher incidents of violence when compared to other inpatient settings, such as the acute hospital setting (Bowers, 2011). Carmel and Hunter (1993) found that 209 staff members in inpatient settings were victims of 236 injuries over the course of a five year period. Other studies have found that between 30-76% of staff within psychiatric settings have been assaulted by patients on at least one occasion (Campbell et al., 2011; Hatch-Maillette et al., 2007; Poster & Ryan, 1994).

Incidents of violence and aggression have a profound negative impact on the safety and well-being of patients and staff and can erode the therapeutic environment within forensic mental health settings (McKenzie & Curr, 2005). Chen, Hwu and Williams’ (2008) study of the impact of aggression and violence on nurses showed that the threat of violence increases anxiety levels among the nursing staff and leads to negative attitudes towards patients. Others have found that violence and aggression can evoke fear in staff members (Ilkiw-Lavelle & Greynder, 2003). Perceptions of risk and safety have been found to influence the decisions and responses of staff (Weyman & Kelly, 1999). Other literature has highlighted the long-term psychological consequences that inpatient violence has on staff including staff burnout,

diminished job satisfaction as well as experiencing symptoms similar to post-traumatic stress (Inoue, Tsukano, Muraoko, Kaneko & Okamura, 2006; Prosser et al., 1997; Dougherty, Bolger, Preston, Jones & Payne, 1992).

Such emotional responses have been found by others to have a negative impact on treatment regimes due to the conflict they are likely to cause among the staff teams (Rossberg & Friis, 2003). In a review of literature on the impact of violence and aggression in inpatient settings, Bock (2011) asserted that it affected not only patients and staff but also the hospital as an organisation. Hospitals can be affected financially and the quality of service delivery may be hindered owing to staff absenteeism, low retention rates and injury while on duty. Bock (2011, p.15) further asserted that nursing staff can experience emotional and physical trauma which can lead to higher rates of absenteeism resulting in less effective treatment delivery which ultimately affects the patient.

Such hazardous working environments for staff, unsafe living environments for patients and the costs incurred from violent incidents are considered to be preventable if accurate prediction of violence is undertaken (Vitacco, Gonzales, Tomony, Smith & Lishner, 2012). Professionals within secure forensic settings are increasingly being asked to opine on the probability that a patient will engage in violent and aggressive behaviours (Vitacco et al., 2009). It is a core task within the National Offender Management Service (NOMS), the National Probation Service (NPS) and a requirement for psychiatric settings in the UK (Department of Health [DoH], 2007; MoJ, 2009). Accurate prediction is therefore crucial. However, predicting any form of human behaviour is difficult and assessing the risk of violent behaviour is no exception to this rule (Buchanan, Binder, Noriko & Swartz, 2012, p.340). Indeed risk as a concept is defined as a mere *estimate* of the likelihood and severity of an undesirable outcome (Yates & Stone, 1992). The principle function of risk assessment is to stratify people into a risk level (low, medium, high) which is instrumental for informing

appropriate risk management strategies for prison, forensic mental health and community settings (Craig et al., 2008).

Implications for getting it wrong

Assessing risk of violence is infused with methodological difficulties and practical considerations. A major practical difficulty when assessing risk is limiting false errors. This occurs when an individual is incorrectly classified as high risk (over-prediction) or wrongly classified as low risk (under-prediction). The consequences of under-prediction may lead to an increased opportunity to commit future violence as risk management strategies may be lower than needed, or removed (Janus & Meehl, 1997). The implications of this within inpatient settings could include placing staff and other patients at increased risk of becoming a victim. On the other hand, over-prediction of future risk of violence may lead to an unnecessary restriction of liberties being placed on the individual (Janus & Meehl, 1997). The impact of this on the patient being assessed could include remaining on a ward which security level is not commensurate with the individual's true level of risk as well as limiting opportunities for the individual's progression.

Assessing risk of violence

Offenders with mental disorders have contributed to the culture of inquiry that dominates the United Kingdom's risk assessment procedure. For instance, the ethical principles (balancing public safety, guiding management and intervention, upholding an evaluatee's constitutional rights, professionalism) in risk assessment have been partly influenced by high profile cases, which have driven research into risk. Examples include the cases of Christopher Clunis (Coid, 1994) who stabbed a member of the public. An inquiry into the death of his victim

highlighted that the organisation of mental health services were inadequate to appropriately assess and manage Mr Clunis' risk of violence (Coid, 1994). Such cases highlight the need for effective risk assessment and risk management. As highlighted in Chapter One, three generations of risk assessment tools have been identified: unstructured clinical judgment; actuarial risk assessments and structured professional judgment (Bonta, 1996). The research base would suggest that structured professional judgment is the most effective way of predicting risk (DoH, 2007; Otto & Douglas, 2010).

Empirical research has established that the presence of a mental disorder, in combination with other risk factors, increases the risk for various forms of violence (Stuart, 2003).

Furthermore, predicting risk in those with mental disorders is asserted as being more complicated than assessing risk in non-psychiatric forensic populations due to the multiplicity and complex interrelation of factors underlying the behaviour in psychiatric patients (Royal College of Psychiatrists [RCPSYCH], 2008). Alongside their duty to the public and patients in their care, mental health practitioners also have a legal and ethical duty to protect healthcare staff and other patients from violence perpetrated by those detained in psychiatric facilities (Hart & Logan, 2011).

A further difficulty in risk assessment is ensuring risk tools are appropriate for the offender being evaluated. This is of particular concern to female offenders (Caulfield, 2010; Chesney-Lind & Pasko, 2013), as will be outlined below.

Gender differences in violence

Gender has been posited to be one of the best predictors of violence, with more males than females engaging in this type of antisocial behaviour (Monahan et al., 2001). However, the

majority of offenders tend to be male (MoJ, 2015) which may account for this. Nonetheless, the nature of female violence is considered to be qualitatively different. It is suggested that female violence is more reactive in nature, less likely to result in serious injuries, less visible and the victim is more likely to be known to the female offender, a factor stable across subtypes of female offenders (de Vogel & de Vries Robbé, 2013). Others have suggested that female violence is more likely to result in more serious injuries (Tjaden & Thonnes, 2000), although this study evaluated intimate partner violence among the general population and therefore, may not be representative of the female offending population. As mentioned previously, the differences in prevalence rates of violence between men and women have not been found in female psychiatric offenders. Research findings show that female psychiatric patients demonstrate comparable levels of violence to their male counterparts (Binder & McNeil, 1990; de Vogel & de Ruiter, 2005; Nicholls, 2001). A Danish study highlighted that 30% of female inpatients and 29% of male inpatients engaged in physical violence (de Vogel & de Ruiter, 2005), while a US study reported that 73.89% of males and 70.21% of females engaged in an act of violence (Krakowski & Czobor, 2004). The increase in rates may be due to the methodology employed as the study included both verbal and physical assaults (Krakowski & Czobor, 2004). However, sex differences in the *nature* and *severity* of violence perpetrated by female psychiatric patients appears stable. For instance, Krakowski and Czobor (2004) found that women tended to have a higher level of verbal assaults than physical assaults when compared with men. Furthermore, the study also found that acute positive psychotic symptoms were more likely to result in physical assault in females.

Although the predictors for male violence have a strong research base across subtypes of offenders (Farrington & West, 1990), the same cannot be said for female offenders (Caulfield, 2010). Felson (2002) suggests that the motives for male and female violence is similar, while others opine that the risk factors that lead to violent behaviour may be

qualitatively different for female offenders (Blanchette & Brown, 2006). Common risk factors for males and female offenders include childhood abuse, adult victimization, relationship instability and low economic status which are often referred to as ‘gender-neutral’ risk factors. However, they may operate in unique ways for female offenders (Chesney-Lind & Sheldon, 2004; de Vogel & de Vries Robbé, 2013). As such the presence of the same risk factor does not mean that the level or etiology of the risk is the same for both males and females (Caulfield, 2010). Risk factors proposed to be specifically relevant for women, commonly referred to ‘gender-specific’ risk factors, include prostitution, pregnancy at a young age and self-harm (Blanchette & Brown, 2006).

It is currently being contested as to whether risk assessment tools adequately capture these nuances: the differences in the relevance of risk factors and different risk factors (Caulfield, 2010; Chesney-Lind & Pasko, 2013; de Vogel & de Vries Robbé, 2013). Policy makers within the UK are also beginning to recognize the importance of a gender-sensitive approach to the assessment of female offenders. This is reflected in the recommendations of the Corston Report (Home Office, 2007) and a Government Green Paper in 2010 (MoJ, 2010). However, there has been little development in the risk assessment research base.

Assessing risk of violence is an extremely complex task due to the multi-factorial nature of this form of crime (Borum, 1996). As highlighted in Chapter one and two improving the prediction of risk can be achieved through basing decisions on a reliable predictor of violence and risk assessment tools may act as this reliable predictor. This can inform both the clinicians’, and the agencies’ decision of the most appropriate tool to use in practice (Bonta, 2002). The reliability of a tool needs to be determined by calculating its predictive validity (its ability to predict future events; Singh, 2013). Typically, recent research uses the Area Under the Curve (AUC) estimate of Receiver Operating Characteristic (ROC) for this purpose (Singh, Desmarais & van Dorm, 2013). A variety of risk assessment tools have been

researched in an attempt to determine their reliability to predict future violence. However, the literature is contradictory in terms of which tools are most accurate, and this becomes even more problematic for female offenders.

Overview of violence risk assessment tools available

Meta-analyses would support the use of both actuarial and SPJ approaches in assessing violence (Farrington, Joliffe & Johnstone, 2008; Guy, 2008). Table 3 outlines some of the most commonly used tools to assess violence in forensic settings.

Table 3

Violence risk assessment tools

Tool	Type of tool	Type of Violence predicted	Authors
(Actuarial vs SPJ)			
HCR-20	SPJ	General violence	Webster et al. (1997)
SARA	SPJ	General violence (specifically spousal violence perpetrated by males)	Kroop & Hart (1994, 1995, 1999)
SAVRY	SPJ	General violence (for juveniles)	Borum, Barter & Forth. (2006)
SVR-20	SPJ	General + sexual violence	Boer, Hart, Kroop & Webster (1997)
VRAG	Actuarial	General violence	Quinsey, Harris, Rice & Cormier (1998, 2006)
VSC	SPJ	General (psychiatric)	McNeil & Binder (1994)

Notes. HCR-20- Historical, Clinical and Risk Management- 20; SARA- Spousal Assault Risk Assessment; SAVRY- Structured Assessment of Violence Risk in Youth; SVR-20- Sexual Violence Risk-20; VRAG- Violence Risk Assessment Guide; VSC- Violence Screening Checklist

Of the tools listed above the SAVRY and SARA are designed for assessment on specific populations (juveniles and males respectively) and therefore, cannot be generalised to the general female offending population. Furthermore, the SVR-20 is designed for the assessment of sexual violence risk (Boer et al., 1997), a specific form of violence. As such, that leaves the VSC, VRAG, PCL-R and HCR-20 as potentially valid for assessing the risk of interpersonal violence among female offenders.

The Violence Screening Checklist (VSC; McNeil & Binder, 1994) was a tool developed specifically to assess inpatient violence. It was validated on a civil psychiatric sample and was found to have below chance ability to predict violence in females (Nicholls, Ogloff and Douglas, 2004). The Psychopathy Checklist-Revised (PCL-R; Hare, 1991) has been used frequently to assess violence among offenders but varying results have been obtained for females offenders in both correctional and psychiatric samples. Among studies conducted on correctional samples, some research (Coid et al., 2009) has demonstrated that the PCL-R was able to predict violence, and that it was better at predicting violence in female samples. Conversely, Warren et al. (2005) did not find the PCL-R to be predictive of violence in females. With respect to psychiatric samples the PCL-R was found to be no better than chance at predicting violence (Nicholls et al., 2004; Schaap, Lammers & de Vogel 2009). The findings in general do not support the use of the PCL-R for predicting female violence as it has been suggested that the construct of psychopathy, as defined in the PCL-R, does not adequately assess psychopathy in females (de Vogel et al., 2012). Additionally, the PCL-R is a measure of psychopathy, not violence, and as such it was never intended to be used to predict violence risk. Assuming the available tools are appropriate for predicting risk, it is important for clinicians to ensure that the most appropriate assessment is used when predicting risk (Lavoie, Guy, & Douglas, 2009).

The Historical, Clinical and Risk- 20 (HCR-20) was developed to bridge the gap between clinical practice and research on violence risk assessment, and specifically between the conflicting evidence between actuarial and unstructured clinical judgment approaches (Webster et al., 1997). It is designed to predict violence in forensic, criminal justice and civil psychiatric settings.

Predictive validity of HCR-20 in assessing violence risk

There is a wealth of research attesting to the ability of the HCR-20 to predict violence in male offending and psychiatric populations (Douglas & Webster, 1999; Grann, Belfrage & Tennstrom, 2000; Gray, Taylor & Snowden, 2008; Witt, 2000; Yang, Wong, & Coid, 2010). Furthermore, the HCR-20, has been found to improve accuracy in violence risk assessment among the psychiatry profession (Tao, Holley, Leary & McNeil, 2012). Although the predictive potency of the HCR-20 for females is less than perfect, based on the tools currently available it would appear that this may be most promising tool for female offenders. Strand, Strand and Belfrage (2001) found no significant differences between male and female inpatients in terms of their HCR-20 scores, although the risk profiles of the sample were not examined. Nicholls et al. (2004) found that the HCR-20 demonstrated good predictive accuracy in female psychiatric patients. However, it is noted that the sample in this study was based on a non-offending population. Other studies have found that the HCR-20 was no better than chance in predicting violence in correctional (Warren et al., 2005) or female inpatients (Schaap et al., 2009). However, de Vogel and de Ruiter (2005) found that the HCR-20 final risk judgment, not the total score or subscales, demonstrated predictive validity in female psychiatric patients. Similar findings have been found elsewhere (Guy, 2008; Garcia-Mansilla, Rosenfeld & Nicholls, 2009). Coid et al. (2009) also found the HCR-20 to have good predictive validity among a female correctional sample. Lidz, Mulvey and Gardner

(1993) found that predicting violence for female psychiatric patients was no better than chance, although this was based within an acute psychiatric facility. McKeown (2010) conducted a review of the potential of the HCR-20 to predict violence in women and concluded that further research is needed to validate its use on female offenders. Skeem et al. (2005) found that the risk of violence in female psychiatric patients tend to be underestimated by mental health professionals including psychiatrists, nurses, clinicians educated to postgraduate level and psychology interns.

Although the HCR-20 demonstrates the most utility for predicting violence, research evidence remains equivocal as to whether gender-responsive or gender-neutral tools would be more predictive of female violence. A meta-analysis by Singh et al. (2011) found that risk assessment tools designed for particular populations (such as young offenders) tend to produce higher predictive accuracy estimates. Conversely, the authors also found that mainstream risk assessment tools (HCR-20, VRAG) produced higher validity estimates for females. However, this was based on 8% of the sample and thus conclusions cannot be drawn from this result. Reisig, Holtfreter and Morash (2006) also found mixed results on the predictive ability of a gender-neutral risk assessment. As such a prevailing question may be whether gender-neutral or gender-specific risk tools are more predictive among female offenders. The ambiguous empirical base suggests that the domination of gender-neutral violence risk assessment tools could mean that practitioners are ill-equipped to assess risk of violence among female offenders. Evaluating the potential value of gender-specific risk factors for violence risk is considered important given that one of the main advantages of the SPJ approach to risk assessment is to understand how risk factors are relevant to the manifestation of violence within the individual (Strub, Douglas & Nicholls, 2014). The Female Additional Manual (FAM) was developed to achieve this in female offenders.

HCR-20 Female Additional Manual (FAM) Guidelines (de Vogel et al., 2012; 2014)

The FAM comprises of 14 items designed to assess risk of violence specifically in female populations. The items were chosen based on a critical review of the literature, clinical expertise and a pilot study conducted in the Netherlands in 2010 (de Vogel et al., 2012, pp.30-31). This led to the inclusion of several additional items and supplementary guidelines within the historical scale as well as the addition of new clinical and risk management items (such as prostitution, pregnancy at a young age, victimization, and self-harm). A copy of the risk factors which have additional guidelines, as well as new risk factors for each scale, is included in Appendix I.

The aim of the FAM is to provide risk assessors with more concrete guidelines for a gender-responsive risk assessment and management for women across various forensic settings (de Vogel et al., 2012; 2014). The FAM can be used across both correctional and psychiatric settings. It is suggested as offering particular potential in assessing risk for inpatient violence (de Vogel & de Vries Robbé, 2013).

HCR-20 Version 3 (HCR-20^{V3}; Douglas, Hart, Webster & Belfrage, 2013)

The HCR-20^{V3} is an update to HCR-20 which incorporates recent research developments and greater consideration to risk formulation and risk management (Douglas et al., 2013). Please see Chapter three for further details on the HCR-20^{V3}. A list of the differences between HCR-20^{V2} and HCR-20^{V3} is included in Appendix II.

Research Aims and Hypotheses

Hypotheses were derived based on the empirical literature reviewed above with respect to the lack of suitability of mainstream violence assessment tools for female offenders (Caulfield,

2010; de Vogel & de Vries Robbé, 2013). They were also derived from research indicating that the SPJ approach, specifically the HCR-20, evinces the most promise with respect to assessing violence. Furthermore, the empirical base identified that a tool designed specifically for adult females (de Vogel et al., 2012) as well as an updated version of the HCR-20 (HCR-20^{V3}; Douglas et al., 2013) have emerged.

Objectives within this empirical study are aimed at evaluating the predictive validity of risk assessment tools for inpatient violence using a prospective research design. Specifically, the research aims to examine the predictive utility of a recently developed risk assessment tool, the HCR-20 FAM, in predicting institutional violence. The project also aims to investigate the predictive validity of the HCR-20's most recent version, HCR-20^{V3}. As such, the predictive validity of the HCR-20^{V2}, HCR-20^{V2} FAM, HCR-20^{V3} and HCR-20^{V3} FAM will be explored.

It is hypothesised that:

- HCR-20^{V2} FAM will be more effective than HCR-20^{V2} in predicting institutional violence.
- HCR-20^{V3} will be a better predictor of inpatient violence than HCR-20^{V2}
- HCR-20^{V3} FAM will be a better predictor of inpatient violence than HCR-20^{V2} FAM

Method

Participants

Purposive sampling was used to recruit participants from both low and medium secure units in a forensic mental health setting in the [REDACTED], UK. Recruitment was undertaken across its two sites. The recruitment of participants did not involve any direct contact with

participants; instead, recruitment was done through reviewing file information. A total of 66 participants were recruited during this phase of the research. The sampling frame for participants included female inpatients aged 18 years or above, who did not have an intellectual disability, who were residing in a forensic mental health setting and who were convicted of an offence.

While sample sizes $n < 200$ can lead to large inaccuracies in estimated population parameters in studies that use ROC analyses (Hanczar et al., 2010), it was anticipated that the sample size for the present study would be 60 females. This number was arrived at by reviewing the sample sizes used in previous studies utilising ROC analyses (which have ranged from 41-1600), recognising that small samples sizes can be a particular problem with respect to researching violence in females (Burman, Batchelor & Brown, 2001; de Vogel & de Vries Robbé, 2013), as well as considering what is feasible for a prospective study at the facility where the study was to take place.

Design

A quantitative approach was used to determine the predictive validity of the HCR-20 FAM guidelines in terms of violent behaviour (physical and verbal). Specifically, a prospective cohort study was undertaken using a 12 month follow-up period.

To reduce the potential for reporting bias and ensure that the study upholds necessary standards for conducting studies on risk assessments, the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE: Singh et al., 2015a) were used in the study. These guidelines have been developed to improve the accuracy of reporting results within risk assessment research and to enable meaningful comparisons across studies measuring the predictive validity of risk assessment tools. (Singh et al., 2013). The RAGEE is a 50-item

reporting checklist which identifies key features (such as reporting sample size, age, inter-rater reliability, analyses) that should be included within validity studies. Using the RAGEE may have the potential for improving the quality of the studies within the violence risk assessment literature (Singh et al., 2015a). It was also hoped that this would reduce the potential for any biased reporting, which is viewed as a form of research misconduct (Fanelli, 2013). Please see Appendix III for a copy of the RAGEE checklist.

Procedure

The data collection period of the project was undertaken in three phases. Phase 1 involved scoring each HCR-20 in March 2014. They were coded for the purposes of this research project only. Each risk assessment undertaken was based on file information and reviewing the daily patient progress notes on the hospital's record system. These sources are detailed, often-voluminous and contain a wide range of information about the patient's psychological, psychiatric, social and criminal history. Whilst an interview is recommended when undertaking a HCR-20 risk assessment, the HCR-20^{V2} and HCR-20^{V3} manuals state that scoring the instruments based on file information is an acceptable form of completing the assessment (Douglas et al., 2013, p.41; Webster et al., 1997, pp.15-17). Each version of the HCR-20 was scored by the author who was trained in completing each of the instruments and who has experience in completing violence risk assessments in forensic settings. To ensure transparency, a second assessor was recruited to assess inter-rater reliability among the overall risk categories. A random sample of assessments, 26% of participants ($n = 17$), were selected for scoring by the independent assessor. The second rater was graduated to Master's level in psychology and had experience conducting violence risk assessments in practice. Differences in scores were resolved by consultation. Inter-rater agreement was assessed using an intra-class correlation coefficient (ICC) for each of the HCR-20's. For HCR-20^{V2} an ICC

of .82 was found, an ICC of .86 was found for HCR-20^{V2} FAM, for HCR-20^{V3} an ICC of .84 was found, while for HCR-20^{V3} FAM an ICC of .86 was found indicating strong agreement between the raters. Any differences in opinion were resolved by consensus.

During Phase 2, in September 2014, the outcome variable (violence) was recorded at the first follow-up period of six months. Phase 3 involved recording the incidents of violence in March 2015, at the 12 month follow-up period. The Overt Aggression Scale (OAS: Yodosky et al., 1986) was used to code for violence. Data on incidents of institutional violence was measured by the hospital's record system where daily information is recorded on salient issues relating to the patients as disruptive incidents and update to patients' medical file.

Measures

Predictive validity was defined as the ability of a structured assessment to accurately predict violence (Singh, 2013). The following risk assessment tools were examined for their ability to predict violence in female offenders:

HCR-20^{V2} (Webster et al., 1997).

This is a structured professional judgment risk assessment intended for assessment of violence risk in those aged 18 years or older within forensic psychiatric, civil psychiatric and offender samples. It is made up of twenty risk factors which are aligned into past (Historical items), present (Clinical items) and future risk factor (Risk management items). They are aimed at highlighting factors that may aggravate or elevate risk. Within the risk management items the HCR-20 allows for the possibility of coding for violence under the premise that the individual will be institutionalized ('In' ratings) and under the premise that individual's will

be in the community ('Out'). As inpatient aggression was being assessed, only the 'In' scenario ratings were recorded.

Scores can be provided for each of the three subscales; the historical, clinical and risk management scales as well as an overall final risk judgment. Higher scores are indicative of higher risk, however, there is also a clinical override option whereby the assessor can elevate or reduce an individual's overall risk. The final risk judgment categorises an individual into a 'low', 'moderate' or 'high' risk of future violence. The final risk rating is arrived at through coding the three subscales of risk factors and integrating all available information. In terms of internal consistency, for the present study HCR-20^{V2} demonstrated low internal consistency with a Cronbach alpha coefficient of .58. However, this is not surprising given the concept that the HCR-20 is measuring, i.e risk of violence. However, violence is not a unitary concept: each factor and scale is heterogenous and the nature of each individual's risk will vary within the measurement.

HCR-20^{V3} (Douglas et al., 2013).

The HCR-20^{V3} is an update to the HCR-20 and places more emphasis on formulating management strategies and reducing risk rather than solely predicting violence. Please see Chapter three for a comprehensive overview of the tool. In the present study, HCR-20^{V3} was approaching acceptable internal consistency, as approaching acceptable internal consistency, $\alpha = .69$. Nevertheless, violence risk is not a unitary concept which makes the interpretation of such estimates difficult.

Female Additional Manual (FAM) Guidelines (de Vogel et al., 2012; 2014).

The FAM guidelines were developed as an additional manual to the HCR-20^{V2}. They are made up of 14 items including nine risk factors considered relevant for female offenders. The guidelines also incorporate new coding aspects such as marking critical items which are factors considered to be particularly relevant for the individual being assessed. Additionally, the guidelines suggest that the final risk ratings be made on a five-point scale: 'low'; 'low-moderate'; 'moderate'; 'moderate-high' and 'high'. This is in contrast with the three-point scale approach in the HCR-20^{V2} and HCR-20^{V3} and the manual suggests that a five-point scale offers more utility for forensic populations where progress may be slow (de Vogel et al., 2012, p.35). The FAM has been validated on a Dutch sample (de Vogel & de Vries-Robbé, 2011). The FAM can also be applied to the HCR-20^{V3} and additional guidelines are provided for this by the authors to the researcher. Since the scoring of the HCR-20's in March 2014, additional FAM guidelines were subsequently published (de Vogel, de Vries Robbé, van Kalmhout & Place, 2014). The guidelines used in the present study for HCR-20^{V3} were consistent with the official guidelines published in September 2014.

In terms of internal consistency, for the present study HCR-20^{V2} FAM and HCR-20^{V3} FAM, $\alpha = .71$, and $.74$ respectively.

Outcome Measure

The outcome measure is 'institutional violence'. Violence was defined as per HCR-20; and included both verbal and physical violence with the intention to harm or hurt others (Douglas et al., 2013; Webster et al., 1997).

Overt Aggression Scale (Yodosky et al., 1986).

The OAS was used to measure the outcome variable. The OAS is a 16-item rating of aggressive behaviour. It has four subscales including: Verbal Aggression; Physical Aggression against self; Physical Aggression against objects and Physical Aggression against other people. The OAS has been validated on adult psychiatric patients and has demonstrated good reliability (ICC = $>.75$). For the present study the scale demonstrated excellent reliability at both the six month and 12 month follow-up point with $\alpha = .94$, and $\alpha = .96$, respectively.

The OAS codes for both physical and verbal aggression, is a simple and reliable tool to use and provides an objective measure of measuring violence as defined by the HCR-20. It also measures other forms of aggression (such as aggression towards self, aggression against objects) which may be particularly relevant for female offenders (see de Vogel & de Vries Robbé, 2013) and have clinical utility in terms of future management.

The study also recorded the diagnostic information for personality disorder or mental illness which was taken from each of the participant's file information.

Ethical Considerations

Guidelines highlighted in the organisation's Service Evaluation Policy (February 2013), University of Birmingham's Code of Practice (2013), British Psychological Society (BPS) Code of Ethics (2009) and the Health Care and Professionals Council's (HCPC) Standards of Conduct, Performance and Ethics (2009) were used to guide the research at every stage and ensure ethical standards were adhered to.

The study did not involve any contact with patients and participant consent was not sought. This was due to the study being a service evaluation project for the organization. The administration of the HCR-20 is a compulsory component of care provision within the

organization and is therefore routinely administered. Under the organization's policy such research can be conducted without obtaining participant consent as Section 251 of NHS Act (2006) allows patient data to be used where there is a benefit to existing or future patients which outweighs the potential risks. Under the proposed research this would mean that assessing the validity of a risk instrument routinely used within the organisation would provide invaluable information on the utility of the assessment for female offenders at the hospital. Therefore, consent was obtained from the organization only. However, participant's anonymity was preserved through by assigning each participant a unique code and collecting data on a separate file. This anonymised data was then transferred to an encrypted USB stick that only the researcher had access to.

Approval from the organization was granted through organization's Ethics Committee in January 2014. Ethical approval was granted from the University of Birmingham in February 2014 (ERN_13-0678).

Data analysis

All results were calculated using SPSS 22. Descriptive statistics were used to explore population characteristics, examine HCR-20 risk ratings and incidence of violence and aggression. Differences between the mean HCR-20 scores were assessed using ANOVAs, while differences between violent and non-violent females were explored using a series of MANOVAs. For each HCR-20, the HCR-20 total, subscale totals and overall risk ratings were assessed for predictive validity. Predictive accuracy was assessed using ROC/AUC statistics which were a suitable method for the present study as they are unaffected by base rates (Mossman, 1994) which can be low for violence, and which are even lower among female samples (de Vogel & de Vries Robbé, 2013). ROC curves plot the sensitivity or true

positive rate (those correctly identified as violent) against the specificity or false positive rate (those identified as violent who are not violent) for every possible cut-off score of the risk instruments. This generates an AUC estimate, which can be interpreted as the probability that a randomly selected offender engaged in an antisocial behaviour would score higher on the risk instrument than a randomly selected offender who did not (Altman & Bland, 1994). The AUC is represented from 0-1, with higher estimates representing higher predictive accuracy of an instrument. For instance, if an AUC of .80 is obtained, this would mean that out of 100 times of using the tool on randomly selected recidivists and non-recidivists, 80% of the time the recidivist would score higher. Within the risk assessment literature, an AUC of .50 indicates prediction by chance whereas an AUC of 1 indicates perfect prediction. Typically AUCs of $\geq .70$ are considered to be statistically acceptable for predicting future violence (Mossman, 1994). However, others have suggested that AUC's of .70 are only adequate whereas AUCs of .80 - .90 are more acceptable (Hosmer & Lemeshow, 2000). Assessing associations between individual risk items on the HCR-20^{V2}/FAM and HCR-20^{V3}/FAM and the outcome was analysed using point biserial correlations (r_{pb}). The size of the value of the obtained r value provided an indication of the strength of the relationship, with a perfect correlation being either ± 1 . Rice and Harris (2005) highlight that r_{pb} values of .10, .243 and .371 correspond to small, medium and large effect sizes within risk assessment literature. Where a strong relationship was obtained between risk items and violence (outcome) this would indicate that violence could be determined by knowing the value on the risk factor being analysed.

Results

Participant Characteristics

A total of 66 participants were identified as eligible for inclusion in the study. Over half of the sample were British (62.5%, $n = 43$), 3% were Irish ($n = 2$), 1.5% were Jamaican ($n = 1$), while the nationality of 30.3% of the sample was unknown ($n = 20$). Participants ages ranged from 18-71 years, with the mean age of the sample being 36.48 years ($SD = 13.76$). Over half of the sample resided within low secure wards (65.2%, $n = 43$), with the remaining participants residing in medium secure wards (34.8%, $n = 23$). Table 4 outlines the diagnostic groups and range of offences perpetrated by the sample.

Table 4

Psychiatric Classification, Personality Disorder (PD) and Index Offence

	N	Percentage
Offence Type		
Offences against person	47	76%
Property offences	28	45%
Theft/Robbery	17	27%
Arson	9	15%
Sexual Offences	1	2%
Public disorder offences	13	21%
Personality Disorder		
Any PD ^a	48	73%
Emotionally Unstable PD	41	62%
Dissocial PD	4	6%
Mixed PD	4	6%
Psychiatric Diagnosis		
Schizophrenia, schizotypal, delusional disorders (F20-F29)	29	44%
Mood Disorders (F30-39)	3	5%
Neurotic, stress-related and somatoform disorders (F40-48)	2	3.0%
Behavioural syndromes (F50-59)	4	6%
Comorbid disorders ^b	31	47%

Notes. ^aA diagnosis of any PD includes PD Not Otherwise Specified and/or any other PD, ^bThis includes more than one diagnosis, for example a PD and a mental illness or more than one of either PD and/or mental illness

HCR-20 Risk ratings

For HCR-20^{V2} the total score ranged from 18-36. For HCR-20^{V2} FAM the total score ranged from 19-47. With respect to the HCR-20^{V3} Presence ratings, the total score ranged from 18-38, while the Relevance ratings for HCR-20^{V3} total scores ranged from 18-40. For HCR-20^{V3} FAM Presence ratings, the total score ranged from 17-49, while for HCR-20^{V3} FAM Relevance, the total score ranged from 20-53. Overall risk ratings for each of the risk assessment tools are presented in Table 5.

Table 5

Overall risk ratings for each of risk assessment tools

Risk Tool	Risk Level	N	Percentage
HCR-20^{V2}	Low	29	44%
	Moderate	31	47%
	High	6	9%
HCR-20^{V2} FAM	Low	12	18%
	Low-Moderate	21	32%
	Moderate	16	24%
	Moderate-High	10	15%
	High	7	11%
HCR-20^{V3}	Low	27	41%
	Moderate	29	44%
	High	10	15%
HCR-20^{V3} FAM	Low	16	24%
	Low-Moderate	17	26%
	Moderate	15	23%
	Moderate-High	11	17%
	High	7	11%

A series of between groups ANOVAs were conducted to compare the mean scores across each of the risk levels for each risk assessment tool. For HCR-20^{V2}, the mean scores differed

significantly across each of the overall risk ratings, ($F(2, 63) = 21.43, p = .00$). A Tukey post hoc test revealed that the mean scores were significantly higher in the Moderate ($M = 30.10, SD = 2.46$) and High risk ($M = 32.17, SD = 2.63$) groups when compared with the Low risk ($M = 25.72, SD = 3.50$) group. There was no significant difference found between the Moderate and High risk groups ($p = .27$).

For HCR-20^{V2} FAM, the mean scores differed significantly across the risk levels ($F(4, 62) = 8.75, p = .00$). Post hoc comparisons using the Tukey test indicated that only the Low risk category ($M = 29.92, SD = 6.72$) differed significantly from each of the other risk categories as defined by the HCR-20^{V2} FAM (Low-Moderate; $M = 36.38, SD = 3.72$; Moderate: $M = 39.31, SD = 5.19$; Mod-High: $M = 38.80, SD = 5.07$; High: $M = 41.71, SD = 4.57$). None of the other risk categories differed significantly from each other ($p > .05$).

With respect to the HCR-20^{V3}, a significant difference was found for each of risk categories for the mean scores for the presence of risk factors ($F(2, 63) = 12.30, p = .00$). Scores were significantly higher as the level of risk increased as all risk categories differed significantly from each other (Low, $M = 27.22, SD = 3.46$; Moderate: $M = 30.14, SD = 3.98$; High, $M = 33.70, SD = 3.06$; ($p < .05$). Similarly, for the relevance of risk factors, a significant difference was also obtained, $F(2, 63) = 3.49, p = .03$. Only the Low risk category ($M = 33.59, SD = 4.63$) differed significant from the High risk category ($M = 37.40, SD = 2.76, p = .05$). There was no significant difference between the Moderate risk category ($M = 35.72, SD = 4.26$) and either the Low or High risk categories ($p > .05$).

For HCR-20^{V3} FAM, there was a significant difference between each of the risk ratings for the presence of risk factors, $F(4, 61) = 5.71, p = .01$. Only the Low risk category ($M = 33.06, SD = 7.00$) differed significantly from the Moderate-High ($M = 41.27, SD = 5.16, p = .00$) and High ($M = 43.00, SD = 5.16, p = .00$) risk categories. The Low-Moderate ($M = 39.64,$

$SD = 4.71$) and Moderate ($M = 37.00$, $SD = 4.84$) ratings did not differ from any of the other risk categories ($p > .05$). In terms of the relevance of ratings, a significant difference between the risk categories was also found, $F(4, 61) = 3.22$, $p = .01$. Only the Low ($M = 39.00$, $SD = 8.68$) risk category differed significantly from the Moderate-High ($M = 47.00$, $SD = 4.60$, $p = .01$) risk category. None of the other risk categories differed significantly from each other (Low-Moderate, $M = 43.06$, $SD = 5.47$; Moderate, $M = 42.80$, $SD = 5.10$; High ($M = 45.71$, $SD = 4.11$, $p > .05$).

To assess differences in scores between violent and non-violent offenders, a series of one-way between groups multivariate analysis of variance (MANOVAs) were performed. The four dependent variables included: total HCR-20 scores as well as the total scores for the Historical, Clinical and Risk management scales. Preliminary assumption testing was conducted and revealed that the assumption of equality of variances was violated for the Historical scale on HCR-20^{V2}. As such a more conservative alpha level was set for determining significance for this tool. This was set at .01. Overall, there was a statistically significant difference between violent and non-violent inpatients at both six and 12 month time periods $p = .00$ and $.02$ respectively for HCR-20^{V2} and HCR-20^{V2}FAM. When each of the scales were considered separately, using a bonferroni correction, both the Clinical and Risk management scales were significantly different between violent and non-violent inpatients for HCR-20^{V2} and HCR-20^{V2}FAM at both six and twelve months. Here violent inpatients tended to score higher than non-violent inpatients. Inspection of the mean scores between offenders at both six and twelve months indicates little difference suggesting that the sample was relatively consistent across both follow-up periods, despite the high rate of attrition. Table 7 illustrates the difference between violent and non-violent inpatients for the presence and relevance of ratings in HCR-20^{V3} and HCR-20^{V3}FAM. There were no differences uncovered between the scores of violent and non-violent inpatients for the

relevance of risk items, $p > .05$. Nonetheless, for the presence of ratings the Clinical scale differed significantly between offender groups at both follow-up periods for both HCR-20^{V3} and HCR-20^{V3}FAM.

Table 6

Mean HCR-20^{V2} and subscale scores for violent versus non violent female offenders

Scale	Overall	6 months (<i>n</i> = 54)			<i>F</i>	12 months (<i>n</i> = 30)		<i>F</i>
		Violent	Non-Violent	Violent		Non-Violent		
HCR-20^{V2}								
Total	28.09	29.53	26.00	13.37**	29.95	26.22	6.85	
(s.d)	(3.88)	(2.81)	(4.29)		(3.21)	(4.35)		
H	14.50	14.88	13.95	1.91	14.95	14.78	.03	
(s.d)	(2.42)	(2.08)	(2.82)		(2.31)	(2.64)		
C	7.15	7.81	6.18	13.24**	7.95	6.00	10.04*	
(s.d)	(1.80)	(1.38)	(1.92)		(1.50)	(1.66)		
R	6.43	6.81	5.86	9.32**	6.95	5.44	11.40*	
(s.d)	(1.21)	(1.06)	(1.21)		(1.11)	(1.13)		
HCR-20^{V2} FAM								
Total	36.78	38.25	34.64	4.69	38.57	36.00	1.18	
(s.d)	(6.23)	(5.04)	(7.22)		(4.84)	(8.03)		
H	19.19	19.47	18.77	.34	19.62	20.67	6.91	
(s.d)	(4.07)	(3.98)	(4.88)		(3.68)	(4.50)		
C	9.87	10.66	8.73	10.71*	10.81	8.33	8.24*	
(s.d)	(2.32)	(1.89)	(2.43)		(2.01)	(2.50)		
R	7.70	8.09	7.14	6.96*	8.14	7.00	4.38	
(s.d)	(1.38)	(1.12)	(1.55)		(1.19)	(1.73)		

Note. * $p < .012$, ** $p < .00$

Female Specific Risk Factors

To determine whether the female specific risk factors were present within the sample, the file was split at both the six and twelve month follow-ups. Items were deemed present if they were coded as 'Definitely present' or 'Possibly present' in each HCR-20 and a comparison between violent and non-violent offenders was made. Independent samples *t*-tests revealed that there were no significant differences between violent and non-violent offenders on any of the female specific risk factors for HCR-20^{V2} at either the six or twelve month follow-up periods $p > .05$).

Similarly, for HCR-20^{V3}, there were no significant differences uncovered in the scores for violent and non-violent offenders for the presence or relevance of risk items at either the six or twelve month follow-up periods ($p > .05$). However, at the six month

Table 7

Mean HCR-20^{V3} and subscale scores for violent versus non violent female offenders

Scale	Overall	6 months (<i>n</i> = 54)			12 months (<i>n</i> = 30)		
		Violent	Non-Violent	<i>F</i>	Violent	Non-Violent	<i>F</i>
HCR-20^{V3} P^a							
Total	29.02	30.31	27.14	8.41*	30.90	26.00	8.96*
(s.d)	(4.22)	(3.66)	(4.36)		(4.21)	(3.87)	
H	15.91	16.31	15.32	2.33	16.57	14.44	4.95
(s.d)	(2.38)	(1.99)	(2.80)		(2.14)	(2.96)	
C	7.11	7.63	6.36	10.72*	7.90	6.22	9.36*
(s.d)	(1.51)	(1.24)	(1.59)		(1.34)	(1.48)	
R	6.00	6.38	5.45	5.18	6.43	5.33	2.49
(s.d)	(1.52)	(1.52)	(1.37)		(1.96)	(1.00)	
HCR-20^{V3} R^b							
Total	34.83	35.59	33.73	2.25	35.33	33.11	1.49
(s.d)	(4.54)	(3.25)	(5.85)		(4.23)	(5.23)	
H	19.91	20.34	19.27	1.28	20.43	19.33	1.87
(s.d)	(3.88)	(3.37)	(4.55)		(3.50)	(4.69)	
C	11.59	12.03	10.95	3.64	11.76	10.89	2.09
(s.d)	(1.96)	(1.47)	(2.40)		(1.76)	(2.52)	
R	11.31	11.44	11.14	1.17	11.10	11.67	.00
(s.d)	(1.78)	(1.37)	(2.27)		(1.55)	(2.59)	
HCR-20^{V3} FAM P							
Total	37.11	38.84	34.59	6.25	38.86	34.89	2.25
(s.d)	(6.44)	(5.62)	(6.84)		(6.14)	(7.30)	
H	19.91	20.44	19.14	1.54	20.38	19.56	.30
(s.d)	(3.81)	(3.37)	(4.34)		(3.56)	(4.21)	
C	9.89	10.66	8.77	15.12*	10.81	8.56	9.01*
(s.d)	(1.97)	(1.60)	(1.95)		(1.72)	(2.24)	
R	7.40	7.67	6.78	5.13	7.67	6.78	1.31
(s.d)	(1.96)	(2.03)	(1.72)		(2.03)	(1.72)	
HCR-20^{V3} FAM R							
Total (s.d)	42.81 (6.83)	43.81 (5.37)	41.36 (8.46)	1.70	43.19 (6.02)	41.89 (8.94)	.25
H (s.d)	19.91 (3.88)	20.34 (3.37)	19.27 (4.55)	.99	20.43 (3.50)	19.33 (4.69)	.50
C (s.d)	11.59 (1.96)	12.03 (1.47)	10.95 (2.40)	4.18	11.76 (1.76)	10.89 (2.52)	1.19
R (s.d)	11.31 (1.78)	11.44 (1.37)	11.14 (2.27)	.37	11.10 (1.55)	11.67 (2.60)	.57

Note. P = Presence of ratings, R = Relevance of ratings, $p < .012$,

follow-up period, the presence of C6 Covert/manipulative behaviour was approaching statistical significance, with violent offenders ($M = 1.28, SD = .77$) scoring higher for Covert/manipulative behaviour than non-violent offenders ($M = .86, SD = .77; t(52), p = .05$).

Table 8*Presence of female specific risk factors in HCR-20^{V2} FAM and HCR-20^{V3} FAM*

Risk Tool		Violence at six months (<i>n</i> = 54)				Violence at twelve months (<i>n</i> = 30)			
		<u>Violent</u>		<u>Non-violent</u>		<u>Violent</u>		<u>Non-violent</u>	
		Present	Not present	Present	Not Present	Present	Not present	Present	Not Present
HCR-20^{V2} FAM									
	H11.	7 (21.9%)	25 (78.1%)	7 (31.8%)	15 (68.2%)	5 (23.8%)	16 (76.2%)	4 (44.4%)	5 (55.6%)
	H12	11 (34.4%)	21 (65.6%)	10 (45.4%)	12 (54.5%)	9 (42.9%)	12 (57.1%)	4 (44.4%)	5 (55.6%)
	H13	11 (34.4%)	21 (65.6%)	5 (22.7%)	17 (77.3%)	6 (28.6%)	15 (71.4%)	4 (44.4%)	5 (55.6%)
	H14	31 (96.9%)	1 (3.1%)	19 (86.4%)	3 (13.6%)	20 (95.2%)	1 (4.8%)	8 (88.9%)	1 (11.1%)
	H15	21 (81.2%)	6 (18.8%)	19 (86.4%)	3 (13.6%)	18 (85.7%)	3 (14.3%)	9 (100%)	--
	C6	24 (75%)	8 (25%)	14 (63.6%)	6 (36.4%)	15 (71.4%)	6 (28.6%)	5 (55.6%)	4 (44.4%)
	C7	32 (100%)	--	21 (95.5%)	1 (4.5%)	21 (100%)	--	8 (88.9%)	1 (11.1%)
	R6	8 (25%)	24 (75%)	8 (36.4%)	14 (63.6%)	4 (19%)	17 (81%)	5 (55.6%)	4 (44.6%)
	R7	27 (84.4%)	5 (15.6%)	18 (81.8%)	4 (18.2%)	19 (90.5%)	2 (9.5%)	7 (77.8%)	2 (22.2%)
HCR-20^{V3} FAM Presence									
	H11.	8 (25%)	24 (75%)	6 (27.3%)	16 (72.7%)	6 (28.5%)	15 (71.4%)	3 (33.3%)	6 (66.7%)

Assessing risk in female offenders

HCR-20^{V3} FAM Relevance	H12	13 (40.6%)	19 (59.4%)	10 (45.4%)	12 (54.5%)	11 (52.4%)	10 (47.6%)	4 (44.4%)	5 (55.6%)
	H13	11 (34.4%)	21 (65.6%)	5 (22.7%)	17 (77.3%)	6 (28.6%)	15 (71.4%)	4 (44.4%)	5 (55.6%)
	H14	31 (96.9%)	1 (3.1%)	18 (81.8%)	4 (18.2%)	20 (95.2%)	1 (4.8%)	8 (88.9%)	1 (11.1%)
	C6	26 (81.2%)	6 (18.8%)	14 (63.6%)	8 (36.4%)	15 (71.4%)	6 (28.6%)	5 (55.6%)	4 (44.4%)
	C7	32 (100%)	--	21 (95.5%)	1 (4.5%)	21 (100%)	--	8 (88.9%)	1 (11.1%)
	R6	10 (31.2%)	22 (68.8%)	9 (40.9%)	13 (59.1%)	5 (23.8%)	16 (76.2%)	5 (55.6%)	4 (44.4%)
	R7	27 (84.4%)	5 (15.6%)	18 (81.8%)	4 (18.2%)	18 (85.8%)	3 (14.3%)	7 (77.8%)	2 (22.2%)
	H11.	8 (25%)	24 (75%)	7 (31.8%)	15 (68.2%)	6 (28.6%)	15 (71.4%)	3 (33.3%)	6 (66.7%)
	H12	13 (40.6%)	19 (59.4%)	9 (40.9%)	13 (59.1%)	11 (52.4%)	10 (47.6%)	4 (44.4%)	5 (55.6%)
	H13	11 (34.4%)	21 (65.6%)	5 (22.7%)	17 (77.3%)	6 (28.5%)	15 (71.4%)	4 (44.4%)	5 (55.6%)
	H14	31 (96.9%)	1 (3.1%)	20 (90.9%)	2 (9.1%)	21 (100%)	--	8 (88.9%)	1 (11.1%)
	C6	26 (81.2%)	6 (18.8%)	16 (72.8%)	6 (27.2%)	16 (76.2%)	5 (23.8%)	6 (66.7%)	3 (33.3%)
	C7	32 (100%)	--	21 (95.5%)	1 (4.5%)	21 (100%)	--	8 (88.9%)	1 (11.1%)
	R6	10 (31.3%)	22 (68.8%)	8 (36.3%)	14 (63.6%)	5 (23.8%)	16 (76.2%)	5 (55.5%)	4 (44.4%)
R7	28 (87.5%)	4 (12.5%)	19 (86.4%)	3 (13.6%)	19 (90.4%)	2 (9.5%)	7 (77.7%)	2 (22.3%)	

Notes. H11- Prostitution; H12- Parenting difficulties; H13- Pregnancy at a young age; H14- Suicidality/Self-harm; H15- Victimization after childhood; C6- Covert/manipulative behaviour; C7- Low self-esteem; R6 Problematic childcare responsibility; R6- Problematic intimate relationship

Incidents of aggression and violence perpetrated

Types of aggressive and violent incidents, as measured by the OAS, are highlighted in Table 9. The most common form of aggression perpetrated by the sample at both the six month and twelve month follow-ups was verbal aggression followed closely by aggression towards others. Less than half the sample at both time periods engaged in acts of self-harm.

Table 9

Differences in scores between types of aggression and violence perpetrated

Type of Aggression	Aggression at six months (<i>n</i> = 54)		Aggression at twelve months (<i>n</i> = 30)	
	Yes	No	Yes	No
Verbal Aggression	55.6% (<i>n</i> = 30)	44.4% (<i>n</i> = 24)	66.7% (<i>n</i> = 20)	33.3% (<i>n</i> = 10)
Aggression against self	40.7% (<i>n</i> = 22)	59.3% (<i>n</i> = 32)	36.7% (<i>n</i> = 11)	63.3% (<i>n</i> = 19)
Aggression against objects	29.6% (<i>n</i> = 16)	70.4% (<i>n</i> = 38)	23.3% (<i>n</i> = 7)	76.7% (<i>n</i> = 23)
Aggression against others	44.4% (<i>n</i> = 24)	55.6% (<i>n</i> = 30)	53.3% (<i>n</i> = 16)	46.7% (<i>n</i> = 14)

Note. Missing cases were excluded.

Rates of violence

At the six month follow-up, the rate of attrition was 18.2% (*n* = 12) while at the 12 months further 24 participants were no longer accessible which equated to over half of the original sample (*n* = 36). Reasons for this included participants being moved to other secure facilities or progressing to the community. Rates of violence were calculated by summing three forms of aggression as measured by the OAS; Verbal Aggression; Aggression towards objects and Aggression towards others. These coincide with the definition of interpersonal violence as defined by the HCR-20. The base rate of violence at six months was 59.26% (*n* = 32), while at 12 months 70% (*n* = 21) of the sample were violent.

Predictive validity of HCR-20's

Inpatient Violence.

Table 10 presents point biserial correlations and AUCs for total HCR-20 scores and for the overall risk categories for each of the four risk assessments, HCR-20^{V2}; HCR-20^{V2}FAM; HCR-20^{V3} and HCR-20^{V3}FAM. AUC values ranged from .55-.75. Based on the AUC estimates obtained, results suggest that HCR-20^{V2} total score, HCR-20^{V2} summary judgment, HCR-20^{V2}FAM summary judgment, HCR-20^{V3}FAM summary risk rating and HCR-20^{V2} summary judgment were the best predictors of violence at the six month period. Lower scores were obtained for the twelve month follow-up, however similar trends have been found. It is noted that results indicate wide confidence intervals on each of the measures which should be taken into consideration when interpreting the findings. All of the overall risk ratings were significantly associated with violence at the six month follow-up. However, for the 12 month follow-up only the HCR-20^{V3} summary risk ratings were not significantly associated with violence. Only the HCR-20^{V2}FAM and HCR-20^{V3}FAM risk ratings yielded a strong relationship with violence at twelve months, indicating that those with higher levels of risk also had higher levels of violence. Although HCR-20^{V3}FAM Relevance rating had a very strong positive relationship with violence at six months, at the twelve month follow-up a negative weak relationship was uncovered suggesting that higher scores on the relevance scales were not associated with higher levels of inpatient violence. Thus it would appear that the violence risk assessments were better predictors of inpatient violence at the six month follow-up.

Table 10

Predictive validity of HCR-20^{V2}, HCR-20^{V2}FAM, HCR-20^{V3} and HCR-20^{V3}FAM

	Violence at 6 months				Violence at 12 months			
	<i>r</i> _{pb}	AUC	<i>p</i>	95% CI	<i>r</i> _{pb}	AUC	<i>p</i>	95% CI
HCR-20 ^{V2} Total	.43 ^a	.75	.00	[.61, .88]	.37 ^b	.66	.04	[.51, .81]
HCR-20 ^{V2} SJ	.41 ^a	.72	.00	[.57, .86]	.37 ^b	.54	.59	[.39, .69]
HCR-20 ^{V2} FAM Total	.23	.64	.08	[.48, .79]	.13	.62	.12	[.48, .76]
HCR-20 ^{V2} FAM SJ	.49 ^a	.78	.00	[.65, .91]	.55 ^a	.66	.04	[.52, .80]
HCR-20 ^{V3} Presence Total	.32 ^b	.69	.00	[.54, .84]	.45 ^b	.63	.09	[.47, .78]
HCR-20 ^{V3} Relevance Total	.11	.57	.40	[.41, .73]	.23	.52	.81	[.36, .68]
HCR-20 ^{V3} SRR	.40 ^a	.72	.00	[.57, .85]	.23	.53	.66	[.38, .69]
HCR-20 ^{V3} FAM Presence Total	.28 ^b	.66	.04	[.52, .81]	.24	.60	.21	[.44, .75]
HCR-20 ^{V3} FAM Relevance Total	.86	.55	.53	[.39, .70]	-.02	.50	.95	[.34, .64]
HCR-20 ^{V3} FAM SRR	.44 ^a	.75	.01	[.61, .88]	.50 ^a	.67	.03	[.53, .80]

Notes. SJ = summary judgment; SRR = summary risk rating. ^a*p* < .01; ^b*p* < .05

Figures 2-9 in Appendix IV provide a visual appraisal of the ROC curves for each risk tool. These represent the predictive accuracy of each tool based on the overall risk category at both the six and twelve month follow-up periods. For both HCR-20^{V2} and HCR-20^{V3} it appeared that the tool was better able to discriminate against violent and non-violent patients in those who scored as being at a Low risk of future violence at both follow up period points. For HCR-20^{V2} FAM and HCR-20^{V3} FAM, at six months point the tool was better able to predict those who would be violent than those who would not for patients in the Low-Medium risk categories. At 12 months, both the HCR-20^{V2}FAM and HCR-20^{V3}FAM appeared to be a better predictor of those in the Low and Low-Medium risk categories.

Predictive validity of individual subscales and items.

In order to evaluate how predictive each of the subscales and individual risk items were, a series of further ROC analyses were conducted. These can be seen in Table 11. Total HCR-20 scores and the predictive utility of overall risk categories are presented above.

In terms of HCR-20^{V2}, at six months follow-up point, only item C2 (Negative Attitudes), the overall C subscale and overall R subscale were predictive of inpatient violence at levels identified by Mossman (1994) as acceptable levels. At the twelve months follow-up point

only one historical risk factor, H4 (Employment Problems), was predictive of violence at twelve months. With respect to the Clinical scale, C2 (Negative Attitudes), C4 (Impulsivity) and the overall C subscale were found to be predictive at twelve months. While none of the Risk Management items were predictive of inpatient violence at the six month period, both R3 (Lack of Personal Support) and the overall R Subscale were found to be predictive at twelve months. In terms of the HCR-20^{V2}FAM, both the overall C and R subscales were found to be predictive of inpatient violence at six months. However, inspection of the individual risk factors suggests that none of the female specific risk factors were predictive, showing that the HCR-20^{V2}FAM did not improve the prediction of inpatient violence at either time period.

With respect to HCR-20^{V3}, for the presence of risk factors, similar to HCR-20^{V2} H4 and the overall H subscale was found to have acceptable AUC estimates at the 12 month follow-up only. However, these failed to reach statistical significance ($p > .05$). The overall Clinical scale was found to be predictive at both six and twelve months follow-up periods. While C4, (Impulsivity), has an acceptable AUC estimate at 12 months ($AUC > .70$), this did not reach statistical significance ($p > .05$). In terms of the relevance of risk items within HCR-20^{V3}, none of the risk items of subscales were found to be predictive. The majority were no better than chance in predicting violent outcome at either six or twelve months.

For HCR-20^{V3}FAM, only the presence of items within the Clinical scale was found to be predictive of violence at six months. None of the female specific factors were found to be predictive. Although C6 (Covert/manipulative behaviour) was found to have an AUC of .70 at twelve months follow-up point it did not reach statistical significance. Similar to HCR-20^{V3} none of the Relevance risk ratings were predictive of inpatient violence at either six or twelve months follow-up points for the sample.

Table 11*Predictive validity of HCR-20 subscales and individual risk items*

<u>Risk Tool</u>	Violence at six months (<i>n</i> = 54)			Violence at twelve months (<i>n</i> = 30)		
	<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s	<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s
HCR-20^{V2}						
H Scale						
H1	.52	.78	[.36, .68]	.50	1.0	[.27, .73]
H2	.59	.25	[.43, .75]	.52	.86	[.28, .76]
H3	.52	.80	[.36, .68]	.44	.60	[.21, .66]
H4	.61	.19	[.45, .76]	.75	.03 ^a	[.57, .94]
H5	.49	.94	[.33, .65]	.42	.47	[.20, .63]
H6	.49	.92	[.33, .65]	.67	.14	[.45, .90]
H7	.47	.70	[.31, .63]	.47	.77	[.23, .70]
H8	.54	.65	[.38, .70]	.54	.72	[.31, .78]
H9	.51	.90	[.35, .67]	.43	.56	[.21, .72]
H10	.47	.75	[.32, .63]	.49	.95	[.26, .72]
Total H	.57	.36	 [.41, .74]	.50	.98	 [27, .74]
C Scale						
C1	.62	.12	[.47, .77]	.61	.34	[.39, .84]
C2	.73	.00 ^b	[.59, .87]	.82	.00 ^b	[.67, .99]
C3	.59	.25	[.43, .75]	.49	.93	[.25, .73]
C4	.64	.07	[.49, .80]	.76	.02 ^a	[.54, .97]
C5	.59	.28	[.44, .74]	.67	.15	[.47, .86]
Total C	.74	.00^b	 [.61, .88]	.81	.00^b	 [.66, .97]
R Scale						
R1	.56	.23	[.44, .75]	.66	.16	[.44, .89]
R2	.65	.07	[.50, .76]	.68	.13	[.47, .88]
R3	.57	.41	[.41, .72]	.75	.04 ^a	[.57, .92]
R4	.60	.22	[.44, .76]	.60	.41	[.38, .81]

<u>Risk Tool</u>		Violence at six months <i>(n = 54)</i>			Violence at twelve months <i>(n = 30)</i>		
		<u>AUC</u>	<u>p</u>	<u>CI</u> s	<u>AUC</u>	<u>p</u>	<u>CI</u> s
HCR-20^{V2}FAM	R5	.60	.22	[.44, .75]	.61	.34	[.37, .85]
	Total R	.71	.01^a	[.57, .85]	.82	.01^a	[.66, .98]
	H Scale						
	H11	.46	.58	[.23, .61]	.40	.40	[.17, .63]
	H12	.47	.69	[.31, .63]	.48	.87	[.25, .71]
	H13	.57	.41	[.41, .72]	.43	.59	[.21, .66]
	H14	.55	.55	[.38, .71]	.49	.92	[.26, .72]
	H15	.48	.78	[.32, .63]	.36	.22	[.15, .56]
	Total H FAM	.56	.65	[.37, .70]	.39	.33	[.16, .62]
	C Scale						
	C6	.59	.26	[.44, .75]	.67	.14	[.47, .87]
	C7	.62	.15	[.46, .77]	.46	.75	[.23, .70]
	Total C FAM	.72	.00^b	[.58, .86]	.77	.02^a	[.59, .96]
	R Scale						
	R6	.47	.67	[.31, .62]	.33	.15	[.11, .55]
R7	.51	.87	[.35, .67]	.60	.39	[.38, .82]	
Total R FAM	.69	.02^a	[.54, .83]	.70	.09	[.51, .90]	
HCR-20^{V3}							
H Scale Presence							
H1	.53	.78	[.36, .68]	.50	1.0	[.27, .73]	
H2	.52	.83	[.36, .68]	.67	.09	[.48, .92]	
H3	.58	.34	[.42, .73]	.53	.82	[.29, .76]	
H4	.66	.04 ^a	[.52, .81]	.71	.10	[.52, .91]	
H5	.53	.74	[.37, .69]	.48	.86	[.23, .71]	
H6	.47	.72	[.31, .63]	.66	.17	[.43, .90]	
H7	.48	.83	[.33, .64]	.41	.42	[.19, .62]	

<u>Risk Tool</u> HCR-20 ^{V3}		Violence at six months (n = 54)			Violence at twelve months (n = 30)		
		<u>AUC</u>	<u>p</u>	<u>CI</u> s	<u>AUC</u>	<u>p</u>	<u>CI</u> s
	H8	.58	.30	[.42, .74]	.57	.56	[.33, .81]
	H9	.63	.11	[.48, .79]	.69	.11	[.47, .90]
	H10	.44	.48	[.29, .60]	.52	.89	[.23, .75]
	Total H Presence	.57	.29	 [.43, .75]	.72	.07	 [.51, .91]
	C Scale Presence						
	C1	.64	.08	[.49, .79]	.66	.18	[.44, .88]
	C2	.65	.06	[.50, .80]	.72	.07	[.52, .90]
	C3	.56	.42	[.40, .72]	.50	.93	[.26, .72]
	C4	.59	.26	[.42, .73]	.61	.34	[.37, .85]
	C5	.58	.32	[.42, .73]	.61	.34	[.37, .85]
	Total C Presence	.74	.00^b	 [.60, .88]	.80	.00^b	 [.64, .96]
	R Scale Presence						
	R1	.53	.74	[.37, .69]	.52	.84	[.29, .75]
	R2	.59	.25	[.44, .75]	.59	.46	[.39, .80]
	R3	.60	.24	[.44, .75]	.69	.10	[.49, .89]
	R4	.75	.38	[.42, .72]	.68	.15	[.47, .86]
	R5	.65	.29	[.50, .81]	.54	.73	[.31, .77]
	Total R Presence	.66	.05^c	 [.51, .81]	.68	.13	 [.49, .87]
	H Scale Relevance						
	H1	.52	.78	[.36, .68]	.50	1.0	[.27, .73]
	H2	.57	.37	[.41, .73]	.67	.14	[.45, .90]
	H3	.47	.86	[.32, .64]	.49	.95	[.26, .72]
	H4	.54	.62	[.38, .70]	.48	.86	[.24, .71]
				Violence at six months			Violence at twelve months

Assessing risk in female offenders

<u>Risk Tool</u>		<i>(n = 54)</i>			<i>(n = 30)</i>		
		<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s	<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s
HCR-20^{V3}							
	H5	.52	.82	[.36, .68]	.50	.98	[.27, .74]
	H6	.48	.82	[.32, .64]	.66	.16	[.43, .89]
	H7	.49	.90	[.33, .64]	.42	.53	[.20, .65]
	H8	.59	.26	[.43, .75]	.59	.44	[.35, .82]
	H9	.54	.61	[.38, .70]	.57	.53	[.33, .82]
	H10	.53	.73	[.37, .69]	.46	.73	[.24, .68]
	Total H Relevance	.56	.45	 [.40, .72]	.65	.21	 [.42, .87]
	C Scale Relevance						
	C1	.48	.85	[.33, .64]	.50	1.00	[.27, .73]
	C2	.59	.27	[.43, .75]	.66	.18	[.42, .89]
	C3	.54	.60	[.38, .70]	.49	.93	[.26, .72]
	C4	.57	.52	[.39-.71]	.53	.77	[.30, .77]
	C5	.55	.52	[.39-.71]	.51	.95	[.28, .74]
	Total C Relevance	.60	.22	 [.44, .76]	.66	.17	 [.44, .88]
	R Scale Relevance						
	R1	.54	.65	[.38, .69]	.51	.95	[.28, .74]
	R2	.49	.90	[.33, .65]	.46	.73	[.24, .68]
	R3	.54	.64	[.38-.70]	.51	.91	[.28, .75]
	R4	.52	.78	[.36, .68]	.50	1.00	[.27, .73]
	R5	.55	.57	[.39-.71]	.45	.68	[.23, .67]
	Total R Relevance	.51	.92	 [.35, .67]	.43	.53	 [.20, .65]
HCR-20^{V3}FAM							
	H Scale Presence						
	H11	.49	.90	[.33, .65]	.48	.84	[.25, .71]
		Violence at six months			Violence at twelve months		

Assessing risk in female offenders

Risk Tool		(n = 54)			(n = 30)			
		AUC	p	CI s	AUC	p	CI s	
HCR-20^{V3}FAM								
		H12	.50	.97	[.34, .65]	.52	.87	[.29, .75]
		H13	.56	.47	[.40, .71]	.42	.50	[.19, .65]
		H14	.57	.37	[.40, .71]	.49	.93	[.26, .72]
		Total H Scale FAM Presence	.57	.39	 [.41, .73]	.54	.72	 [.32, .77]
		C Scale Presence						
		C6	.65	.07	[.50, .80]	.70	.09	[.51, .89]
		C7	.56	.29	[.43, .74]	.44	.62	[.21, .67]
		Total C Scale FAM Presence	.77	.00^b	 [.64, .89]	.80	.01^a	 [.64, .96]
		R Scale Presence						
		R6	.48	.78	[.32, .63]	.35	.21	[.13, .58]
		R7	.55	.53	[.40, .71]	.56	.62	[.33, .79]
		Total R Scale FAM Presence	.67	.03^a	 [.52, .81]	.65	.19	 [.46, .85]
		H Scale Relevance						
		H11	.47	.75	[.32, .63]	.45	.68	[.21, .69]
		H12	.53	.75	[.37, .68]	.53	.80	[.30, .76]
		H13	.55	.55	[.39, .71]	.42	.47	[.18, .64]
		H14	.53	.70	[.37, .69]	.49	.95	[.26, .72]
		Total H Scale FAM Relevance	.54	.62	 [.38, .70]	.52	.86	 [.29, .75]
		C Scale Relevance						
		C6	.45	.68	[.21, .69]	.65	.21	[.44, .85]
					Violence at six months			Violence at twelve months

Assessing risk in female offenders

<u>Risk Tool</u>	<i>(n = 54)</i>			<i>(n = 30)</i>		
	<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s	<u>AUC</u>	<u><i>p</i></u>	<u>CI</u> s
HCR-20^{V3}FAM						
C7	.53	.80	[.30, .76]	.34	.18	[.12, .56]
Total C Scale FAM	.42	.47	 [.18, .64]	.62	.29	 [.40, .85]
Relevance						
R Scale						
Relevance						
R6	.46	.64	[.30, .62]	.32	.11	[.09, .54]
R7	.56	.44	[.41, .72]	.54	.72	[.30, .78]
Total R Scale FAM	.50	.98	 [.33, .66]	.37	.27	 [.12, .62]
Relevance						

Notes. H Scale- Historical Items, C Scale- Clinical Items, R Scale- Risk Management Items, HCR-20^{V2}FAM and HCR-20^{V3}FAM total scale scores for H, C and R scales were calculated by summing the original HCR-20s with the HCR-20FAM, ^a*p* < .05, ^b*p* < .01, ^c*p* = .05

Self-harm.

Given the prevalence of self-harm among female offenders (Blanchette & Brown, 2006) and given that it is included as a specific risk factor with the HCR-20 FAM (de Vogel et al., 2012), further ROC analyses were conducted in order to explore the relationship between the Suicidality (H14) risk factor and rates of self-harm as measured by the OAS. At six months, 33.3% ($n = 22$) while at twelve months 36.7% ($n = 11$) of the sample had engaged in acts of self-harm.

Table 12
Predictive validity of Suicidality for rates of self-harm

Risk items	Self-harm at six months ($n = 54$)			Self-harm at twelve months ($n = 30$)		
	AUC	p	CI	AUC	p	CI
Version 2 H14	.47	.72	[.31, .63]	.61	.34	[.40, .80]
Version 3 H14 Presence	.48	.88	[.33, .65]	.61	.34	[.40, .80]
Version 3 H14 Relevance	.48	.83	[.32, .64]	.61	.34	[.40, .80]

As can be seen from Table 12, the risk item Suicidality/ self-harm had poor predictive accuracy in predicting who would engage in acts of self-harm at both the six month and twelve month follow-up periods.

Summary of main findings

In terms of differences in HCR-20 scores for violent and non-violent offenders, significant differences were found for the Clinical Items risk scale across all tools and both follow-up periods. There were no significant differences found between violent and non-violent offenders for the presence of female specific risk factors at either follow-up period. Further, the presence of some of the female risk factors were quite low, with H14 (Suicidality), C6 (Covert/manipulative behaviour); C7 (Low self-esteem) and R7

(Problematic intimate relationship) being more common among violent offenders.

Nonetheless, no significant differences were found between the presence of female specific risk factors between violent and non-violent offenders.

In terms of the predictive validity of the HCR-20's, the overall risk ratings were found to be more predictive of inpatient violence and the risk assessments were better able to predict violence at the six month follow-up. The HCR-20 FAM appeared to be slightly more predictive of inpatient violence at both six and twelve months in both Version 2 and Version 3 of the HCR-20. The Clinical Items for each HCR-20 were found to be more predictive of inpatient violence than the Historical or Risk Management Scales. However, the addition of female specific risk factors did not appear to improve the prediction of violence within the sample. The suicidality risk factor, H14, did not appear to predict incidents of self-harm, although the prevalence of self-harm was low in the sample.

Discussion

The aim of the study was to contribute to the research base on the ongoing debate regarding the suitability of gender-neutral versus gender-specific risk assessments used for female offenders. Specifically, it aimed to evaluate the predictive validity of a gender-neutral (HCR-20) and gender-specific (HCR-20 FAM) violence risk assessment for predicting inpatient violence. It also aimed to evaluate the most recent version of the HCR-20, HCR-20^{V3}.

In terms of whether the HCR-20 FAM would be more effective than the HCR-20 in predicting institutional violence, results from the current study suggest that while the FAM appears to have produced slightly higher estimates, there was little difference in AUC

estimates obtained (AUC range from .55-.78 at six months, and .50- .66 at twelve months). Furthermore, the width of the confidence intervals makes it difficult to draw meaningful conclusions. HCR-20^{V3} was not found to outperform HCR-20^{V2} in the current sample and seemed to be a better predictor of violence at the six month follow-up rather than the twelve month follow-up point. There was very little difference in AUC estimates obtained for HCR-20^{V3} FAM and HCR-20^{V2} FAM. As such, the hypotheses of the present study are not supported. The high prevalence of complex diagnoses within the sample may have affected the results obtained. Over 40% of the sample had a diagnosis of a schizophrenia related disorder and predicting risk of violence can be more difficult in this population (Royal College of Psychiatrists, 2008), as the link between mental illness and violence, and the role that active symptoms of the illness may play, is not universally established in the empirical base (Short, Lennox, Stevenson, Senior & Shaw, 2012). As such, trying to predicting complex behaviours in a complex population may not be feasible.

Overall, the risk ratings appeared to produce the highest AUC estimates, with the exception of HCR-20^{V3}. This is consistent with previous research which has highlighted that the overall risk judgments are more predictive of violence (de Vogel & de Ruiter, 2005; Garcia et al., 2009; Guy, 2008; Strub et al., 2014). However, the Clinical Items scale was found to be most predictive across each of the four risk assessment tools suggesting that dynamic risk factors were more predictive than static factors for inpatient violence. This may be reflective of the statistically significant differences found between violent and non-violent offenders on this scale. Equally, it may also be due to the items that the Clinical scale evaluates. For instance across both HCR-20^{V2} and HCR-20^{V3}, symptoms of major mental illness, insight, negative attitudes/violent ideation, impulsivity/instability are factors that could be particularly relevant for women. Research has shown that mental health disorders are associated with women's use of violence (Logan & Blackburn, 2009;

Silver, Felson, & Vaneseltine, 2008). Additionally, victimization is considered to be more prevalent among female offenders (Greenfeld & Snell, 1999) which has been linked to PTSD (Breslau, Peterson, Kessler, & Schultz, 1999), and female offenders tend to score higher on scales of state and trait anger and anger expression in comparison with male offenders (Suter, Bryne, Bryne, Howells, & Day, 2002). Such factors are common among those experiencing PTSD (Chemtob, Novaco, Hamada, Gross & Smith, 1997).

Experiencing trauma at a young age can adversely affect brain development which can impact emotional and behavioural functioning including problems with self-regulation, aggression against self-and others, affective lability, difficulties in self-concept and rigid and preservative coping styles (Creedon, 2009; Shields & Cicchetti, 1998; van der Kolk, 2003). Dodge, Bates & Petit (1990) found that the experience of trauma can lead to social information being processed as 'deviant' which sets the stage for the development of hostile attributional biases. Arguably such predispositions could affect items such as Violent Ideation and Instability in HCR-20^{V3} and Negative attitudes and Impulsivity in HCR-20^{V2}.

The relevance ratings were found to be no better than chance in predicting violence at either follow-up time period. This is in contrast to previous research that has demonstrated that the relevance ratings can predict violence in psychiatric samples (Strub et al., 2014). However, it is acknowledged that the gender of the sample was not reported in the study, so it is unclear how the results would relate to female offenders. The inability of relevance ratings to predict violence within the current sample may reflect their importance in terms of formulating an individual's risk of violence (Douglas et al., 2013), which is subjective.

The addition of the female specific risk factors did not appear to improve the prediction of inpatient violence which is in contrast with previous research highlighting the

importance of gender-specific risk factors in predicting risk in female offenders

(Blanchette & Brown, 2006; de Vogel et al., 2013). This could be due to the low prevalence of the female specific risk factors, with no differences being found between violence and non-violent offenders. Equally, it could also be due to the nature of violence among female patients. Previous research has highlighted that gender differences in the prevalence of violence disappear among psychiatric samples (de Vogel & de Ruiter, 2005; Nicholls, 2001). As such, it could be that the FAM guidelines may be more relevant for correctional rather than psychiatric populations, although it is acknowledged that this would contest what the authors of the FAM have evidenced (de Vogel et al., 2012).

Conversely, the lack of predictive potency of female specific risk factors may also be due to the items measured by the FAM. For example although many female offenders may be more likely to become involved in prostitution (see Daly, 1994), prostitution, as a proposed risk factor, may not necessarily have a causal link with violence. Currently, there is no theoretical or research evidence to suggest that prostitution is a risk factor for violence. It may well be a risk for victimisation, which is a risk factor for violence, but the role prostitution plays in future violence is not yet evidenced in the empirical base.

The higher prevalence of verbal aggression within the sample in comparison to physical aggression is consistent with previous research which has found high levels of verbal assault in female inpatients (Krakowski & Czobar, 2004). Although the female risk factor 'Suicidality/ self-harm' was found to be a poor predictor of self-harm in the study this may be due to the low rates of self-harm found in the follow-up period points. A further interesting feature of the HCR-20 FAM is the inclusion of both suicidality and self-harm as an inclusive item when the evidence base distinguishes suicidality from some acts of self-harm. For instance, acts of self-mutilation have been viewed as a form of tension reduction (Favazza, 1992) rather than as an intention to die (suicidality). As such, the

motivations underlying various forms of self-harm may be different and how they contribute to future violence, and in what way, would need to be explored. Motz (2001) has suggested that aggression towards the self is a common method of anger expression in females. The internalizing-externalising model of behaviours (Achenbroch, 1991) would suggest that behaviours are directed either towards the self or others. However, it is acknowledged that the internalizing-externalising model of behaviours exists on a continuum, rather than being discrete categories (Beyers & Loeber, 2003). Nevertheless, it could be argued that such acts mean that it is less likely that they will perpetrate violence, at least in the short-term, as they have a vehicle through which to direct their negative emotions.

Additionally the high base rate of violence within the sample should be taken into consideration. For the present study high base rates were found for both follow-up period points: 59% at six months and 70% at twelve months. The importance of base rates in supporting clinicians to make decisions should be taken into consideration as outlined by Barbaree (1997). Applying his example to the present study, and taking the 6 month follow-up period and the HCR-20^{V2} FAM summary judgement as an illustration, this would indicate that of the 32 people that were violent, the HCR-20^{V2} FAM would have a 78% accuracy in predicting who would be violent and who would not. Here, 25 people would be correctly classified as violent whereas 7 violent people would be classified as non-violent. Of those who were not violent, 17 would be correctly classified as non-violent whereas 5 people are predicted to be misclassified as being violent. If decisions within the hospital were solely based on the risk assessment tool, in this example, 5 people may have unnecessary restrictions placed on their liberty while 7 people may engage in violence and not have necessary risk management plans put in place putting staff and other patients at risk.

In summary, the study findings do not support nor refute the suggestion that gender-specific risk assessments are more effective at predicting risk in female offenders. Rather they suggests that further research is needed on developing tools which adequately assess risk for male and female offenders and raise the question of whether assessments should be evaluating more specific types of violence, rather than evaluating violence more generally.

Strengths and Limitations

Some important limitations should be noted which may have affected the results obtained. The sample size in the present study was quite low which may have affected statistical power in predicting inpatient violence. A post hoc power analysis was conducted to test this. The power of the study was 72%, below the 80% level recommended for experimental studies (Hanley & McNeil, 1982), suggesting that it was underpowered. As such it was less likely we would have observed a true effect which may have been why there was little difference between the predictive validity of each risk assessment tool. In order to reach the 80% threshold a sample size of 66 would have been necessary for the statistical analyses. Further, there was a high rate of attrition, particularly at the twelve month follow-up period. As such it is difficult to draw meaningful conclusions. It may be useful for future research to conduct a study with a larger sample size to increase the internal validity of the study. Additionally, the prevalence of female risk factors within the study was low, suggesting a lack of discrimination between each of the four risk assessments. Additionally, there was no control group used in the study. It may be useful for future studies to compare the statistical ability of the risk assessments to predict inpatient violence with a sample of inpatients with no criminal history. A further

consideration is noted. The OAS (Yudosky et al., 1986) was used to code the outcome (inpatient violence) within the study. However, a dichotomous outcome was coded for the presence or absence of each form of violence. Arguably, as this does not account for the severity of each type of violence within the sample the quality of violence is overlooked which may have affected the predictive accuracy. For instance, it may have been useful to evaluate how accurate risk assessments were for more severe forms of violence in comparison with less severe violence.

Nonetheless, the present study had considerable methodological strengths. Firstly, it is a truly prospective study. Further, it is unique in that it explored a new area within the risk assessment research: the predictive validity of a tool designed for female offenders. Issues of diversity, including gender, are often touted as limitations within mainstream forensic psychology (Caulfield, 2010; Chesney-Lind & Pasko, 2013; McKeown, 2010). Although the study is negatively affected by the sample size, and rates of attrition, it contributes to the sparse empirical base that currently exists on female offending and the invisibility of female perpetrators of violence within research. The study also used the RAGEE guidelines (Singh et al., 2015a) to reduce the potential for reporting bias and ensure standards for undertaking risk assessment research were upheld. A further strength of the study is that in comparison with the mainstream psychiatric literature on violence (Strand and Belfrage, 2001) it focuses specifically on female psychiatric patients who have offended. Given the pathways that lead to females engaging in criminal behaviour (Daly, 1994; Simpson, Yahner & Duggan, 2008) the perpetration of violence among offending psychiatric populations may need further examination.

Implications for research

Presently, research is conflicted regarding the predictive validity of risk assessments tools for female offenders. While this may be due to the low numbers of females within the criminal justice system (MoJ, 2015) this does not render the lack of research as acceptable. Some predictive validity estimates were found to be within the acceptable range (i.e. $AUC \geq .70$) which gives encouragement to the HCR-20's predictive power, despite the low sample size. It is recommended that future research continue to evaluate the validity of the FAM. This may provide the impetus for exploring why certain risk factors are gender-specific, and why and in what manner some gender-neutral risk factors may be more relevant for women (e.g. trauma, suicidality). Consistent with the findings of Strub et al. (2014) future research will need to evaluate the role of the relevance ratings of HCR-20^{V3} and how people are formulating their importance for the individual.

Further, it is worth noting the reliability of each of the scales. There was slightly better internal consistency for the FAM on both versions of the HCR-20, suggesting that the FAM is a more reliable assessment of violence within the present sample. However, the estimates were just within the acceptable range (DeVellis, 2003). Although the HCR-20 may not market itself as a psychometric tool, the reliability estimates would suggest that it does not necessarily conform to the basic requirements of a psychometric test or scale. While the factors that make up the scale have some utility, its usefulness as an overall 'scale' is limited. As such the benefit of the HCR-20 may be in terms of its clinical usefulness rather than being beneficial at a group level. However, this reflects the nomothetic and idiographic debate that has dominated the subject of psychology. Risk assessments are designed to support practitioners to accurately capture risk. However, one of the difficulties in predicting risk has been a tendency to overlook what is important at the individual level as they cannot account for the non-linear way in which individuals interact.

Implications for practice

Although the present study is limited by a number of limitations (such as sample size, low prevalence of female specific risk factors), results suggest that the FAM did not improve the ability of the HCR-20 to predict inpatient violence. On the basis of the results obtained, this would suggest that using the gender-neutral HCR-20's, HCR-20^{V2} or HCR-20^{V3}, would be just as useful as using the HCR-20 FAM. However, as Singh and colleagues (2013) highlight, risk assessment tools are never sufficient on their own to determine risk and that they should only be used to roughly classify individuals into a risk level rather than being used as a prognosis for engaging in offending behaviour. Further, Yang et al. (2010) suggest that the utility of a risk assessment tool should be determined not only on its ability to predict violence but on what other functions it can serve such as managing risk. Both the HCR-20 and FAM guidelines encourage assessors to formulate an individual's risk of violence as well as develop risk management plans.

Nonetheless, effectively assessing risk of violence is crucial for female offenders to ensure adequate risk management plans are implemented to prevent future victimisation. It is only through accurate identification of risk for violence that future violence can be managed or prevented (Ogloff, 2006). As up to 20% of offences committed by reoffenders include violence against the person (MoJ, 2014b), from a preventative perspective effective assessment is paramount. Furthermore, risk assessment professionals have a responsibility to be able to articulate, based on available literature, how existing tools may or may not be relevant to the assessment (Litwack, 2001). However, this is hampered by a lack of a sound empirical base on the validity of risk tools for females. Without an awareness of whether tools are adequately capturing the risk for future violence among

females, professionals are working blind when assessing risk in females. This study will hopefully aid clinicians in determining whether the available tools are effectively predicting violence risk and aid in their decision making regarding which tool to use.

Furthermore, the higher predictive power of the Clinical Items within each of the risk assessments highlights the importance of dynamic factors in predicting inpatient violence. This would suggest that clinicians should prioritise these factors in risk management plans.

Conclusions

The current study contributes towards addressing outstanding questions around the predictive validity of the HCR-20 for females as raised by others (Blanchette & Brown, 2006; Chesney-Lind & Pasko, 2013; de Vogel et al., 2012, 2014; McKeown, 2010). It suggests that gender-neutral risk assessments are just as effective as gender-specific risk tools in predicting inpatient violence. The study also suggests that the HCR-20 was more predictive at the 12 month follow-up period. Practitioners should prioritise clinical risk items in terms of future interventions and risk management to reduce the risk of inpatient violence. The study also indicates that further research is required for the development of female-appropriate risk assessment tool.

CHAPTER FIVE

Discussion

“Prediction of dangerousness is particularly difficult because: dangerousness is the resultant of a number of processes which occasionally may be synergistic amounting to more than the sum of the parts, some within the individual and some in society; it is not static; key factors are the individual’s adaptiveness, resistance to change and his intentions...” (Scott, 1977, p.128)

The aim of the current thesis was to contribute to the sparse literature base on assessing risk in female offenders. It was hoped that this would help reduce the invisibility of female offenders within the empirical base. As has been highlighted in Chapter one, there is little consensus regarding the adoption of female-specific approaches to assessing risk and treatment needs. As such the thesis aimed to determine the suitability of male-oriented risk assessments and evaluate the validity of a female-specific violence risk assessment.

Chapter two presented a systematic review of the predictive validity of risk assessment tools used/available to predict reoffending and violence in female offenders. Specifically, it aimed to contribute to the research base on whether traditional risk assessments tools are valid for female offenders. The systematic review indicated that the current research base continues to present conflicting evidence on whether risk assessment tools are adequately capturing risk in female offenders. Furthermore, it was found that tools were more effective in predicting recidivism than violence. However, the variability in results obtained suggested that risk assessments may not be capturing the relevant risk factors for female offenders. Equally, the quality of studies in the area may have affected the findings. Studies were found to be limited by a number of biases including sampling bias, failing to account for subtypes of female offenders, and differential follow-up periods, in addition to how the outcome was measured as well as how results were reported. Additionally, the dearth of research undertaken in Europe, as well as the lower predictive validity estimates among UK samples, suggest that further research on samples here would be beneficial. As such the review did not support nor contest the adoption of

gender-specific approaches to the assessment of risk in female offenders. Rather, it suggested that research evaluating both gender-neutral and gender-specific risk assessment tools was warranted to resolve the ongoing debate on the suitability of risk assessment tools for female offenders.

Building on these findings, chapter three evaluated the psychometric properties of the HCR-20^{V3} (Douglas et al., 2013). An overview of the tool was presented but the review predominantly considered the reliability and validity of the HCR-20^{V3}. It was also chosen for evaluation as it is a widely used risk assessment tool within correctional and forensic mental health settings as well as being a standardised risk assessment within mentally disordered services (DoH, 2008; Douglas & Reeves, 2010; Fazel et al., 2012). The critique demonstrated that the HCR-20 appears to meet some of the criteria for reliability and validity. One of the benefits of the updated HCR-20 is that there is more breath in terms of the factors measured, making it more inclusive about what each of the factors are measuring. Further, it appears to possess strong external validity as it is cost effective, relatively easy to administer and provides guidance on how to formulate violence risk which subsequently informs risk management and the types of interventions needed to reduce an individual's risk. However, there are some limitations of the assessment. For instance, the reliability of the HCR-20^{V3} improves when a group of raters score the assessment, rather than an individual on his/her. Within clinical settings, particularly within the prison service where there may be less emphasis on multidisciplinary working, it may not always be practical or feasible to have a group of raters score the assessment. Additionally, the lack of an evidence base on how well the HCR-20^{V3} applies to female offenders suggests that it may not be the most appropriate tool for this population. There is also little research conducted on correctional samples with validation studies predominantly focused on forensic mental health populations. Furthermore, the HCR-20^{V3}

is not underpinned by a theory of violence and while the authors suggest that theory should guide the formulation of an individual's violence risk, the assessment does not provide guidance on the causal pathways to violence.

Chapter four presented an empirical research project examining the predictive validity of the HCR-20 FAM (de Vogel et al., 2012; 2014) in predicting inpatient violence in female psychiatric offenders. HCR-20^{V2} and HCR-20^{V3} were evaluated alongside the FAM guidelines for both risk assessments. Results indicated the overall risk ratings (low/medium/high) were found to be more predictive of inpatient violence and the risk assessments were better able to predict violence at the six month follow-up point. The HCR-20 FAM appeared to be slightly more predictive of inpatient violent at both six and twelve months in both Version 2 and Version 3 of the HCR-20. Additionally, the Clinical scale across each of the risk assessments was found to be most predictive of inpatient violence, suggesting that dynamic risk factors are more predictive of future violence. However, the addition of the female-specific risk factors did not appear to improve the prediction of inpatient violence.

Furthermore, it is important to acknowledge that while violence risk assessment is a crucial task within forensic settings it has been argued that it has been overshadowed by a focus on determining probability estimates (Hart & Logan, 2011), which is considered impossible at an individual level (Hart et al., 2007). The research study indicated that although the current literature indicates that the expression of risk factors in female offenders may be different to males (Bonta et al., 1995; Caulfield, 2010; van Voorhis et al., 2010), as outlined in Chapter one, this does not necessarily equate to differences in predictive validity between traditional violence risk assessments and female-specific violence risk assessments. Therefore, the relevancy of the gender-specific literature may be in terms of risk management rather than risk prediction. However, the fair-moderate

predictive validity estimates obtained may be due to the limitations of the study. Equally, it may also be due to the tools not meeting the criteria for a scientific measurement thereby affecting statistical analyses. As such perhaps future research may be directed at developing sound psychometrics which can predict future behaviour with statistical certainty. Furthermore, where the results (i.e. overall risk category) from risk assessments are used solely for the purposes of managing offenders, perhaps such decisions should be made based on detailed formulations of an individual rather than relying on crude estimates from risk assessments.

Any assessment of risk should also consider an individual's protective factors, or strengths, that may reduce his/ her criminogenic risk. It has been suggested that a gender-responsive assessment should consider any strengths or signs of resilience an offender possesses (Meichenbaum, 2006). Similar to the research on risk factors for female offending each sex may respond differently to protective factors (Rumgay, 2004). For instance strong relationships with family and religiosity were found to be protective factors for females but not necessarily as strong for males (Hawkins, Graham, Williams & Zahn, 2009). Such findings would further support the gender-responsive approach as best applied to the management of risk rather than in its prediction.

Despite the limitations and inconsistencies within the current research base on female offenders, the assessment and management of risk remains an area of importance and significance in forensic psychology. It is also of relevance to professionals from other disciplines including psychiatry, nursing, occupational therapy and social work (Hart & Logan, 2011). Professionals working with offenders have a duty to the public to prevent further incidents of violence being committed. There is also a duty to protect staff and other offenders from violence perpetrated. Furthermore, there is a duty to the offenders to assess and manage them appropriately so they can access the interventions needed to

prevent them from the “self-harmful consequences” of perpetrating violence in the future (Hart & Logan, 2011, p.84).

The development of the research base in violence risk assessment (Monahan, 1996) has led to the emphasis on ‘evidence-based practice’ within risk assessment. Evidence based practice is considered to be any clinical action or decision which is guided by the available research on the topic being considered (Sackett & Rosenberg, 1995). The advancements showed that violence is a pervasive social problem, identified risk factors relevant for various forms of violence and also informed clinical practice (with emphasis placed on evidence-based risk assessment procedures where practitioners combine various information so risk assessments), reflect current views on best practice (Hart & Logan, 2011). However, it is still considered limited in identifying the exact critical risk factors, how they relate to each other or the causal roles they play with respect to violence (Hart & Logan, 2011, p.85). It has been argued that forensic psychologists should shift their attention from predicting violence to understanding its causes and preventing future violence (Skeem & Monahan, 2011). The findings within the thesis, would suggest that this argument also extends to female offenders. The debate on whether a gender-specific or gender-neutral approach to female offenders should be adopted has dominated the literature (Pollak, 1950; Heilbrun et al., 2008; Zaplin, 2008). However, there is still an absence on theories of female crime and violence and it is not clear how factors related to crime and violence in women interact with one another and how they differentiate from theories on male offending. The role of gender in theories of offending thus remains unclear (Hannah-Moffat, 2009). Therefore, it is argued that practitioners should be redirecting our attention to a gender-responsive approach to the management of female offenders rather than to a gender-specific assessment of risk in female offenders. Returning to the theories in chapter one on female offending, it would thus seem that gender-neutral

theories such as those outlined in the GPCSL model would be applicable to female offenders, at least in the assessment of risk.

Implications for research

Conclusions drawn from this thesis suggest that perhaps a focus upon a gender-responsive approach to the assessment of risk in female offenders is too narrow. A variety of factors can be relevant to subgroups of offenders (Brennan et al., 2012) and the role that gender plays in offending behaviour is unclear and may influence risk factors in a variety of ways (Hannah-Moffat, 2009; van Voorhis et al., 2010). However, a sound theoretical basis for these differences is yet to emerge. Future research would need to develop a comprehensive theory of female offending, to understand how and in what way traditional or unique risk factors are relevant for women before adopting female-specific risk assessments in practice. Such research would improve our understanding on the nature and etiology of female offending as well as its concomitants. Some scholars opine that there is a tendency to pathologise women's presentation, which is suggested as leading to the over prescription of psychiatric diagnoses including borderline personality disorder (Ussher, 2013). It is argued that although men may have similar presentations, this is accepted by society (Jimenez, 1997). Perhaps similar arguments could extend to the assessment of risk in female offending population. Is there a bias towards searching for differences in female offenders when the nature of their risk may be similar to men? The relevancy of female-specific risk factors has yet to be theoretically or empirically validated. It may well be that women's risk of violence and offending may not be that dissimilar for men. As such, adopting female specific risk assessments without a thorough understanding of how, or whether, differences between both genders influence the expression of criminal

behaviours, particularly violence, may thus be premature. Further, the utility of different theoretical orientations may only be understood when they are applied to a clinical sample. However, that is not to say that female-specific tools, such as the HCR-20 FAM, have little utility in practice. As highlighted above, the factors may not have strong predictive potency but may help guide and shape interventions to reduce risk in female offenders. Equally, it also suggests that once the theoretical base can establish relevant female specific risk factors that these are subjected to psychometric evaluation in predicting risk. This would be particularly important if risk assessments are to be used to make decisions about an offender's management. Furthermore, perhaps research into subtypes of violence would also be a worthy pursuit rather than predicting all forms of violence within one risk assessment tool.

Implications for practice

The management of violent female offenders requires a responsive approach that considers gender differences in mental health, substance use, anger, experiences of victimization and perpetration of violence (Kubiak, Kim, Fedock & Bybee, 2012). The empirical base has been criticised for failing to integrate the current knowledge on the unique offending trajectories and profile of female offenders into risk assessment and risk management (de Vogel et al., 2013). As outlined in chapter two, traditional risk assessment tools may be more predictive of recidivism than violence, however, there is great variability with how recidivism is defined. Chapter four highlighted that the inclusion of female-specific risk factors did not improve the validity of traditional risk assessments at a statistical level. As it has been highlighted within the thesis, and taking account of the limitations within the research study, perhaps the focus for clinicians should be on adopting a responsive

approach to females in the management of risk and interventions delivered to female offenders rather relying on such tools as a sound prediction of risk. In terms of reflecting the differential expression of risk factors between male and female offenders, the onus may thus be on practitioners to adequately reflect this in their formulation of risk and when advising on the appropriate intervention to manage risk. For instance, where trauma has been coded as relevant for a female offender, practitioners may benefit from referring to the literature base on female offenders experience of, and response to, such experiences and direct interventions accordingly.

Additionally, it is important to account for base rates of violence and reoffending which can affect the prediction of risk (Singh et al., 2013). Although the analyses used to predict risk of violence are unaffected by base rates, these should still be incorporated when communicating an individual's risk to others where risk estimates are used for decision making related to the individual's liberty and public safety. Singh et al. (2015b) found that rates of violence among high risk offenders varied greatly across risk assessment instruments and are dependent on geographical location and instrument used. As such general assumptions cannot be made about the probability of violent behaviour without taking such factors into account. Previous research has outlined the importance of considering local base rates of violence when conducting risk assessments where overall risk levels are used to determine detention in prison or hospital and/or discharge or release from custody (Eastman, Adshead, Fox, Latham & Whyte, 2012; Szmukler, 2001) The importance of what information is used to determine base rates is illustrated in the MacArthur risk assessment study (Monahan et al., 2001). Based on officially recorded incidents of violence the base rate for violent acts was reported to be 4.5% whereas when multiple sources of information were used this increased to 27.5%. As such using base rates which are determined by official sources of information, such as rearrest or

reconviction rates, is inadequate (Crighton & Towl, 2008). This suggests that there is a need for research to be conducted on base rates of violence across different populations as well as geographical location. It has been suggested that professionals undertaking risk assessment which includes such information can improve its predictive accuracy (Arkes, 1991).

Conclusions

The current thesis does not support nor refute the adoption of a gender-responsive approach to assessing risk in female offenders. However, it does question the uncritical acceptance of risk assessment tools for female offenders. It also questions whether a gender-responsive approach to risk assessment is needed and whether the adoption of a responsive approach to females is more appropriate for the management of risk. In order to determine the suitability of risk assessment tools for female offenders further research is needed which compares both male and female offenders as well as subtypes of offenders. Additionally, further research into the validity of female specific risk assessments, such as the HCR-20 FAM, is needed before this is adopted within clinical practice. Developing theories on female offending, female violence and violence more generally would be a worthy pursuit. Assessing risk in offenders is an inherently complex process which requires the consideration of a variety of factors, such as the suitability of a tool, whether the tool was validated on the population under question, if it supports clinicians to formulate violence, and how it can inform the management of risk as well as the types of error that can affect the assessment including the base rates of violence/offending of the population under assessment. Such limitations should be reflected within risk assessments.

References

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR and TRF profiles*. Burlington: Department of Psychiatry, University of Vermont.
- Altman, D.G., & Blans, J. M. (1994). Statistics notes: Diagnostic tests 3: receiver operating characteristic plots. *British Medical Journal*, *309*. Retrieved from <http://dx.doi.org/10.1136/bmj.309.6948.188>
- Andrews, D. A. (1982). *The Level of Supervision Inventory (LSI): The first follow-up*. Toronto: Ontario Ministry of Correctional Services.
- Andrews, D. A., & Bonta, J. L. (1995). *The Level of Service Inventory—Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D.A., & Bonta, J. (2003). *The psychology of criminal conduct*. (3rd ed). Cincinnati, OH: Anderson.
- Andrews, D. A., & Bonta, J. (2010). *The Psychology of Criminal Conduct* (5th ed.). New Providence, NJ: LexisNexis.
- Andrews, D., Bonta, J. L., & Wormith, S. J. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with level service general risk and need. *Criminal Justice and Behavior*, *38*, 413-432. doi: 10.1177/0093854811401990
- Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, J., Brews, A., & Wormith, S. (2012). Are the major risk/need factors predictive of both female and male reoffending?: A test with the eight domains of the level of service/case management

- inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56, 113-133. doi: 10.1177/0306624X10395716
- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: a meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, 50, 88-100. doi: 10.1177/0306624X05282556
- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110, 486-498. doi: 10.1037/0033-2909.110.3.486
- Babcock, J. C., Miller, S. A., & Siard, C. (2003). Toward a typology of abusive women: Differences between partner-only and generally violence women in the use of violence. *Psychology of Women Quarterly*, 27, 153-161. doi: 10.1111/1471-6402.00095
- Baert, P. (2005). *Philosophy of social sciences*. Cambridge: Polity Press.
- Bandura, A. (1977). *Social Learning Theory*. General Learning Press.
- Barbaree, H. E. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity of recidivism studies to treatment effects. *Sexual Abuse: A Journal of Research and Treatment*, 9, 111-128. Retrieved from <http://link.springer.com/article/10.1007%2F02674862>
- Batchelor, S. (2005). 'Prove me the bam!': Victimization and agency in the lives of young women who commit violence offences. *Probation Journal*, 52, 358-375. doi:10.1177/0264550505058034
- Becker, S., & McCorkel, J. A. (2011). The gender of criminal opportunity: The impact of male co-offenders on women's crime. *Feminist Criminology*, 6, 79-110. doi: 10.1177/1557085110396501

Belknap, J., & Holsinger, K. (2006). The gendered nature of risk factors for delinquency.

Feminist Criminology, 1, 48-71. doi: 10.1177/1557085105282897

Bem, S., & de Jong, (2006). *Theoretical issues in psychology: an introduction* (2nd Ed.).

London: Sage Publications.

Bennett, S., Farrington, D. P., & Huesman, L. R. (2005). Explaining gender differences in

crime and violence: The importance of social cognitive skills. *Aggression and Violent*

Behavior, 10, 263-228. [doi:10.1016/j.avb.2004.07.001](https://doi.org/10.1016/j.avb.2004.07.001)

Beyers, J. M., Loeber R. (2003). Untangling developmental relations between depressed

mood and delinquency in male adolescents. *Journal of Abnormal Child Psychology*,

31, 247–266. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12774859>

Binder, R. L., & McNeil, D. E. (1990). The relationship of gender to violent behavior in acutely disturbed psychiatric patients. *Journal of Clinical Psychiatry*, 51, 110-114.

Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2307664>

Bishop, C. (1931). *Women and crime*. London, UK: Chatto and Windus Press.

Blanchette, K. (2002). *Classifying female offender for effective intervention: Application of*

the case-based principles of risk and need. Retrieved from

http://205.193.117.157/text/pblct/forum/e141/141h_e.pdf

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of female offenders: an*

integrative perspective. West Sussex: John Wiley & Sons.

Blanchette, K. & Taylor, K. (2005). *Development and field test of a gender-informed security*

reclassification scale for women offenders. Ottawa, ON: Correctional Service

Canada.

- Bock, T. M. (2011). Assessment of attitudes related to the management of aggression and violence in four psychiatric hospitals. (Degree dissertation, Stellenbosch University, South Africa). Retrieved from <http://scholar.sun.ac.za/handle/10019.1/6835>
- Boer, D. P. (2015, July). *The importance of context in assessment, treatment and management*. Paper presented at the Annual Division of Forensic Psychology Conference, Manchester.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (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
- Bonta, J. (1996). Risk-needs assessment and treatment. In A.T. Hartland (Ed). *Choosing correctional options that work: defining the demand and evaluating the supply*. Thousand Oaks, CA: Sage Publications.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice & Behavior*, 29, 355-379. doi: 10.1177/0093854802029004002
- Bonta, J., Pang, B., & Wallace-Capretta, S. (1995). Predictors of recidivism among incarcerated female offenders. *The Prison Journal*, 75, 277 - 294. doi: 10.1177/0032855595075003002
- Borum, R. (1996). Improving clinical practice of violence risk assessment: technology, guidelines and training. *American Psychologist*, 51, 945-956. doi: 10.1037/0003-066X.51.9.945
- Borum, R., Bartel, P., & Forth, A. (2005). Structured Assessment of Violence Risk in Youth (SAVRY). In T. Grisso, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 311-323). New York: Guilford

- Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY)*. Lutz, FL: Psychological Assessment Resources.
- Bottos, S. (2007). Women and violence: Theory, risk and treatment implications. Retrieved from <http://www.csc-scc.gc.ca/research/092/r198-eng.pdf>
- Bowers, L. (2011). Inpatient violence and aggression: a literature review. Report from the Conflict and Containment Reduction Research Programme. Kings College London: Institute of Psychiatry. Retrieved from <http://www.kcl.ac.uk/ioppn/depts/hspr/research/ciemh/mhn/projects/litreview/litreva gg.pdf>
- Bowie, B. H. (2007). Relational aggression, gender, and the developmental process. *Journal of Child and Adolescent Psychiatric Nursing*, 20, 107-115. doi: 10.1111/j.1744-6171.2007.00092.x
- Breakwell, G. M., Fife-Shaw, C., Hammond, S., & Smith, J.A. (2006). *Research methods in psychology* (3rd ed.). London: Sage Publications.
- Brennan, T., & Oliver, W. L. (2000). *Evaluation of reliability and validity of COMPAS scales: National aggregate sample*. Traverse City, MI: Northpointe Institute for Public Management.
- Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, 36, 21-40. doi: 10.1177/0093854808326545
- Breslau, N., Peterson, E. L., Kessler, R. C., & Schultz, L. R. (1999). Short screening scale for DSM-IV posttraumatic stress disorder. *American Journal of Psychiatry*, 156, 908-911. Retrieved from <http://ajp.psychiatryonline.org/article.aspx?volume=156&page=908>

- Brews, A. L. (2009). *The Level of Service Inventory and female offenders: addressing issues of reliability and predictive validity* (Unpublished doctoral dissertation). University of Saskatchewan: Saskatoon, Canada.
- British Psychological Society. (2009). *Code of Ethics and Conduct: guidance published by the Ethics Committee of the British Psychological Society*. Retrieved from http://www.bps.org.uk/system/files/documents/code_of_ethics_and_conduct.pdf
- Buchanan, A., Binder, R., Norko, M., Swartz, M. (2012). Psychiatric violence risk assessment. *The American Journal of Psychiatry*, 169, 340. doi: 10.111176/appi.apj.2012.169.3.340
- Burman, M. J., Batchelor, S. A., & Brown, J. A. (2001). Researching girls and violence: facing the dilemmas of fieldwork. *British Journal of Criminology*, 41, 443-459. Retrieved from <http://eprints.gla.ac.uk/5129/1/Batchelor5129.pdf>
- Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36, 567-590. doi: 10.1177/0093854809333610
- Campbell, J. C., Messing, J. T., Kub, J., Agnew, J., Fitzgerald, S., Fowler, B., Sheridan, D., ...Bolyard, R. (2011). Workplace violence: prevalence and risk factors in the safe at work study. *Journal of Occupational and Environmental Medicine*, 53, 82-89. doi: 10.1097/JOM.0b013e3182028d55
- Carmel, H. & Hunter, M. (1993). "Staff injuries from patient attack: five years' data". *Bulletin of the American Academy of Psychiatry & the Law*, 21, 485-493. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8054678>
- Caulfield, L. (2010). Rethinking the assessment of female offenders. *The Howard Journal*, 49, 315-327. doi: 10.1111/j.1468-2311.2010.00625.x

- Centre for Review and Disseminations. *Systematic Reviews: CRDs guidance for undertaking reviews in health care*. (2009, January). Retrieved from http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
- Chen, S., Hwu, H., & Williams, R. (2005). Psychiatric nurses' anxiety and cognition in managing psychaitric patients' aggression. *Archives of Psychiatric Nursing*, *19*, 141-149. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15991148>
- Chemtob, C. M., Novaco, R. W., Hamada, R. S., Gross, D. M., & Smith, G. (1997). Anger regulation deficits in combat-related posttraumatic stress disorder. *Journal of Traumatic Stress*, *10*, 17-36. doi:10.1023/A:1024852228908
- Chesney-Lind, M., Morash, M. and Stevens, T. (2008). 'Girls' troubles, Girls' delinquency, and gender responsive programming: A review'. *Australian and New Zealand Journal of Criminology*, *41*, 162-89. doi: 10.1375/acri.41.1.162
- Chesney-Lind, M., & Pasko, L. (2013). *The female offender: girls, women and crime* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Chesney-Lind, M., & Shelden, R. G. (2004). *Girls, delinquency, and juvenile justice*, (3rd ed.). Belmont, CA: Thompson Wadsworth.
- Cohen, J. (1988). *Statistical power analyses for the behavior al sciences* (2nd ed). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Coid, J. (1994). The Christopher Clunis enquiry. *The Psychiatric Bulletin*, *18*, 449-452. doi: 10.1192/pb.18.8.449
- Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., Farrington, D. P., & Rogers, R.D. (2009). Gender differences in structured risk assessment: comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, *77*, 337-348. doi: 10.1037/a0015155

- Cooke, N.R. (2008). Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry*, *54*, 17-23. doi: 10.1373/clinchem.2007/096529
- Corston, J. (2007). *The Corston Report: A review of women with particular vulnerabilities in the Criminal Justice System*. London: Home Office.
- Copas, J. B., & Marshall, P. (1998). The Offender Group Reconviction Scale: the statistical reconviction score for use by probation officers. *Journal of Royal Statistical Society*, *47*, 159-171.
- Coulson, G., Ilacqua, G., Nutbrown, V., Giulekas, D., & Cudjoe, F. (1996). Predictive utility of the LSI for incarcerated female offenders. *Criminal Justice and behavior*, *23* (3); 427-439. doi: 10.1177/0093854896023003001
- Craig, L.A., & Beech, A.R. (2010). Towards a guide towards conducting actuarial assessments with sex offenders. *Aggression and Violent Behavior*, *15*, 278-293. doi:10.1016/j.avb.2010.01.007
- Craig, L.A., Browne, K.D., & Beech, A.R. (2008). *Assessing Risk in Sexual Offenders: a practitioners guide*. West Sussex: John Wiley & Sons.
- Creeden, K. (2009). How trauma and attachment can impact neurodevelopment: informing our understanding and treatment of sexual behaviour problems. *Journal of Sexual Aggression*, *15*, 261-273. doi:10.1080/13552600903335844
- Crichton, D.A., & Towl, G.J. (2008). *Psychology in prisons*. (2nd ed.). Oxford, UK: Blackwell Publishing.
- Critical Appraisal Skills Programme (CASP). (May 2013). Critical Appraisal Checklists. Retrieved from <http://www.casp-uk.net/>.
- Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

- Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin*, *122*, 5-37. Retrieved from <http://dx.doi.org/10.1037/0033-2909.122.1.5>
- Daly, K. (1994). *Gender, crime, and punishment*. New Haven, CT: Yale University Press.
- Davidson, J. T., & Chesney-Lind, M. (2009). Discounting women: context matters in risk and need assessment. *Critical Criminology*, *17*, 221-245. doi: 10.1007/s10612-009-9084-x
- Department of Health. (2007). *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. (2008). *Commissioning for quality and innovation (CQUIN) payment framework*. London: Department of Health.
- Deschenes, E. P., Owen, B., Crow, J. (2006). Recidivism among female offenders: Secondary analysis of the 1994 BJS recidivism data set. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/216950.pdf>
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- de Vogel, V. (2005). *Structured risk assessment of (sexual) violence in forensic clinical practice: The HCR-20 and SVR-20 in Dutch forensic psychiatric patients*. (Doctoral dissertation, University of Amsterdam, Netherlands). Retrieved from <http://dare.uva.nl/document/13725>

- de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: a comparison with a matched sample of males. *Clinical Psychology and Psychotherapy*, 12, 226-40. doi: 10.1002/cpp.452
- de Vogel, V., & de Vries Robbé, M. (2013). Working with women: towards a more gender-sensitive violence risk assessment. In L. Johnstone & C. Logan (Eds.), *Managing Clinical Risk: A Guide to Effective Practice* (pp.224-241). London: Routledge.
- de Vogel, V., & de Vries Robbé, M. (2011). *Risk assessment in female psychiatric patients. First results with new gender-sensitive risk assessment guidelines*. Paper presented at the 11th Barcelona, Spain: Conference of the International Association of Forensic Mental Health Services.
- de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2012). *Female Additional Manual (FAM): additional guidelines to the HCR-20 for assessing risk for violence in women*. Amsterdam: Van der HoevenKliniek.
- de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2014). *Female Additional Manual (FAM): additional guidelines to the HCR-20^{V3} for assessing risk for violence in women*. Amsterdam: Van der HoevenKliniek.
- Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science*, 250, 1678-1683.
- [http://www.ncbi.nlm.nih.gov/pubmed?term=\(Dodge%20KA%5BAuthor%5D\)%20AND%20mechanisms%20in%20the%20cycle%20of%20violence%5BTitle%5](http://www.ncbi.nlm.nih.gov/pubmed?term=(Dodge%20KA%5BAuthor%5D)%20AND%20mechanisms%20in%20the%20cycle%20of%20violence%5BTitle%5)
- Dougherty, L. M., Bolger, J. P., Preston, D. G., Jones, S. S., & Payne, H. C. (1992). Effects of exposure to aggressive behaviour on job satisfaction of health care staff. *Journal of Applied Gerontology*, 11, 160-172. doi: 10.1177/073346489201100203

- Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 version 3. *International Journal of Forensic Mental Health, 13*, 130-139. doi: 10.1080/14999013.2014.908429
- Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4*, 149-184. doi: 10.1348/135532599167824
- Douglas, K. S., Guy, L. S., & Reeves, K. A. (2008). HCR-20 violence risk assessment scheme: Overview and annotated bibliography. Retrieved from http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1362&context=psych_cmhsr.
- Douglas, K.S., Hart, S.D., Webster, C.D., & Belfrage, H. (2013). *HCR-20 V3: Assessing risk for violence: user guide*. Simon Fraser University, Canada: Mental Health, Law and Policy Institute.
- Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist— screening version. *Journal of Consulting and Clinical Psychology, 67*, 917–930. <http://dx.doi.org/10.1037/0022-006X.67.6.917>
- Douglas, K. S., & Reeves, K. A. (2010). Historical- Clinical - Risk- Management – 20 (HCR-20) Violence Risk Assessment Scheme: Rationale, application and empirical overview. In R. K. Otto, & K. S. Douglas (Eds.), *Handbook of violence risk assessment: International perspectives of forensic mental health* (1st ed.) (pp. 147-186). New York; London: Routledge.
- Douglas, K. S., & Strub, D. S. (2013). *Predictive validity of HCR-20 (Version 3) amongst civil psychiatric patients and criminal offenders*. Unpublished data analyses. Simon Fraser University, Burnaby, Canada.

- Douglas, K. S., & Webster, C. D. (1999). The HCR-20 violence risk assessment scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior*, *26*, 3-19. doi: 10.1177/0093854899026001001
- 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-20^{V3}. *International Journal of Forensic Mental Health*, *13*, 140-147. doi: 10.1080/14999013.2014.906517
- Eastman, N., Adshed, G., Fox, S., Latham, R., & Whyte, S. (2012). *Forensic psychiatry*. London: Oxford University Press.
- Effective Public Health Practice Project (EPHPP). (2008). Quality Assessment for quantitative studies. Retrieved from <http://www.ephpp.ca/tools.html>
- Egger, M., Dickerson, K., & Smith, D. G. (2007). Problems and limitations in conducting systematic reviews. In M. Egger, G. D. Smith, & D. G. Altman (Eds.), *Systematic reviews in health care. Meta-analysis in context* (2nd ed., pp. 43-66). London, UK: BMJ Publishing Group.
- Eidhammer, G., Selmer, L.E., & Bjørkly, S. (2013). A brief report from a comparison of the HCR-20 and HCR-20-V3 concerning internal consistency and clinical utility. [Dataset]. Retrieved from <http://www.researchgate.net/publication/266160080>
- Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature*, *14*, 149. doi: 10.1038/494149a.
- Farr, K. A. (2000). Classification for female inmates: moving forward. *Crime and Delinquency*, *46*, 3-17. doi: 10.1177/0011128700046001001.
- Farrington, D. P., Joliffe, D., & Johnstone, L. (2008). *Assessing violence risk: A framework for practice (Final Report)*. Retrieved from

<http://www.rmascotland.gov.uk/files/2112/7263/5786/Assessing%20Violence%20Risk%20-%20A%20Framework%20for%20Practice.pdf>

Farrington, D.P., & West, D.J. (1990). The Cambridge study of delinquent development: a long-term follow-up of 411 London Males, *Kriminalität*, 115-138. doi: 10.1007/978-3-642-75418-0_9

Favazza, A. R. (1992). Repetitive self-mutilation. *Psychiatric Annals*, 22 (2), 60-63.

Retrieved from <http://psycnet.apa.org/psycinfo/1992-28124-001>

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874. Retrieved from

<https://ccrma.stanford.edu/workshops/mir2009/references/ROCintro.pdf>

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 72 samples involving 24, 827 people: systematic review and meta-analysis. *British Medical Journal*, 345, 1-12. Retrieved from <http://dx.doi.org/10.1136/bmj.e4692>

Federal Bureau of Investigation. (1998). *National Incident-Based Reporting System*.

Retrieved from <http://sociology-data.sju.edu/3031/3031cb.pdf>

Felson, R. B. (2002). *Violence and gender re-examined: Law and public policy*. Washington, CD: American Psychological Association.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed). London: Sage Publications.

Finn, P., Bothe, A. K., Bramlett, R. E. (2005). Science and pseudoscience in communication disorders: criteria and applications. *American Journal of Speech Language Pathology*, 14, 172 -186. Retrieved from

<http://www.ncbi.nlm.nih.gov/pubmed/16229669>

- Folsom, J., & Atkison, J. L. (2007). The generalisability of the LSI-R and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior*, *34*, 1044-1056. doi: 10.1177/0093854807300097
- Funk, S. J. (1999). Risk assessment for juveniles on probation: a focus on gender. *Criminal Justice and Behavior*, *26*, 44–68. doi: 10.1177/0093854899026001003
- Garcia-Mansilla, A., Rosenfeld, B., & Nicholls, T. L. (2009). Risk assessment: Are current methods applicable to women?. *International Journal of Forensic Mental Health*, *8*, 50-61. doi: 10.1080/14999010903014747
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the “unparalleled” measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior*, *29*, 397-426. doi: 10.1177/0093854802029004004
- Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high risk federal offenders. *Criminal Justice and Behavior*, *29*, 235-249. doi: 10.1177/0093854802029003001
- Grann, M., Belfrage, H., & Tennstrom, A. (2000). Actuarial measurement of risk of violence: Predictive validity of the VRAG and historical part of the HCR-20. *Criminal Justice and Behavior*, *27*, 97-114. doi: 10.1177/0093854800027001006
- Grann, M., Långstrom, N., Tengström, A., & Gunnar, K. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior*, *23*, 205-217. <http://www.ncbi.nlm.nih.gov/pubmed/10333757>
- 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 HCR-20, PCL-R, and psychiatric symptomatology. *Journal of*

Consulting and Clinical Psychology, 71, 443-451.

<http://www.ncbi.nlm.nih.gov/pubmed/12795569>

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

Greenfeld, L. A., & Snell, T. L. (1999). *Women offenders* (NCJ Pub. No. 175688).

Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

<http://bjs.ojp.usdoj.gov/content/pub/pdf/wo.pdf>

Guerino, P., Harrison, P. M., & Sabol, W. J. (2011). *Prisoners in 2010* (NCJ Pub. No.

236096). Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics

Guy, L.S. (2008). *Performance indicators of the structured professional judgement approach for assessing risk for violence to others: A meta-analytic survey*.

Unpublished Doctoral Dissertation, Simon Fraser University, Burnaby, Canada.

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, 551-580. doi: 10.1521/pedi.2010.24.5.551

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under the curve

receiving operating characteristic (ROC) curve. *Radiology*, 143, 29-36. Retrieved from

http://www.med.mcgill.ca/epidemiology/hanley/software/hanley_mcneil_radiology_82.pdf

Hannah-Moffat, K. (2009). Gridlock or mutability: Reconsidering “gender” and risk assessment. *Criminology and Public Policy*, 8, 209-219. doi: 10.1111/j.1745-

9133.2009.00549.x

Hannah-Moffat, K., & Shaw, M. (2001). *Taking risks: Incorporating gender and culture into the classification and assessment of federally sentenced women in Canada*. Ontario:

Status of Women Canada. Retrieved from
<http://publications.gc.ca/site/eng/293809/publication.html>

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, M.A. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, *26*, 822-830. doi: 10.1093/bioinformatics/btq037

Hanson, R.K., & Bussière, M.T. (1998). Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*, 348-362. doi: 10.1037/0022-006X.66.2.348

Hare, R. D. (1991). *The Hare Psychopathy Checklist - Revised*. Toronto, Ontario: Multi-Health Systems.

Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: an international perspective. *Behavioral Sciences and the Law*, *18*, 623-45. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11113965>

Hart, S.D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgement approach. In P. Sturmey & M. McMurrin (Eds.). *Forensic Case Formulation*. Chichester, UK: Wiley-Blackwell.

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry*, *190*, 60-65. doi: 10.1037/0022-006X.66.2.348

Harris, A. J. R., & Hanson, R. K. (2010). Clinical, actuarial and dynamic risk assessment of sexual offenders: why do things keep changing? *Journal of Sexual Aggression*, *16*, 296-310. doi: 10.1080/13552600.2010.494772

- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335. doi: 10.1177/0093854893020004001
- Hart, S.D. (1998a). Psychopathy and risk for violence. In Cooke, D., Forth, A.E., & Hare, R.D (Eds), *Psychopathy, theory, research and implications for society* (pp. 355-379). Dodrecht, The Netherlands: Kluwer Academic Publications.
- Hart, S. D. (1998b). The role of psychopathy in assessing risk for violence: conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121– 137. doi: 10.1111/j.2044-8333.1998.tb00354.x
- Hart, S. D., Laws, D. R., & Kropp, R. P. (2003). The promise and peril of sex offender risk assessment. In T. Ward, D. R. Laws, & S. M. Hudson (Eds.), *Sexual deviance: Issues and controversies* (pp. 207-205). Thousand Oaks, CA: Sage Publications.
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: evaluating the ‘margins of error’ of group v. individual predictions of violence. *The British Journal of Psychiatry, 190*, s60-s65. doi: 10.1192/bjp.190.5.s60
- Hastings, M. E., Krishnan, S., Tangney, J. P., & Stuewig, J. (2011). Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological Assessment, 23*, 174-183. doi: 10.1037/a0021290
- Hatch-Mailette, M. A., Scalora, M. J., Bader, S. M., & Bornstein, B.H. (2007). A gender-based incidence study of workplace violence in psychiatric forensic settings. *Violence and victims, 22*, 449-462. Retrieved from <http://dx.doi.org/10.1891/088667007781553982>
- Hawkins, S. R., Graham, P. W., Williams, J., & Zahn, M. A. (2009). Girls study group: Understanding and responding to girls’ delinquency. Retrieved from <https://www.ncjrs.gov/pdffiles1/ojdp/220124.pdf>

- Health and Care Professionals Council. (2009). *Standards of conduct, performance and ethics*. Retrieved from <http://www.hpcuk.org/assets/documents/10003B6EStandardsOfconduct,performanceandethics.pdf>
- Heilbrun, K., Yashuhara, K., & Shah, S. (2010). Violence risk assessment tools; Overview and critical analysis. In R.K. Otto, & K.S. Douglas (Eds.), *Handbook of violence risk assessment: International perspectives on forensic mental health* (pp. 1-18). London: Routledge.
- HM Prison Service. (2013). *Democratic Therapeutic Communities in prisons. Assessment manual*. National Offender Management Service: UK Government.
- Hollin, C. R., & Palmer, E. J. (2005). *Criminogenic factors among women offenders. Part 3*. London: HM Prison Service.
- Home Office. (2007). *The Corston Report: the need for a distinctive, radically different, visibly-led, strategic, proportionate, holistic, woman-centred, integrated approach*. London: The Stationary Office.
- Hosmer, .D.W., & Lemeshow, S. (2000). *Applied logistic regression* (4th ed.). New York: Wiley.
- Howitt, D.H. (2009). *Introduction to forensic and criminal psychology* (3rd Ed). Essex: Pearson Education Limited.
- Iikiw-Lavalle, O., & Grenyer, B. F. S. (2003). Differences between patient and staff perceptions of aggression in mental health units. *Psychiatric Service*, 54 (3), 389-393. Retrieved from https://www.researchgate.net/profile/Brin_Grenyer/publication/10880095_Differences_between_patient_and_staff_perceptions_of_aggression_in_mental_health_units/links/0c96052a3f3c131e1c000000.pdf

- Inoue, M., Tsukano, K., Muraoka, M., Kaneko F., & Hitoshi, O. (2006). Psychological impact of verbal abuse and violence by patients on nurses working within psychiatric departments. *Psychiatry and Clinical Neurosciences*, *60*, 29-36. doi: 10.1111/j.1440-1819.2006.01457.x
- Janus, E. A., & Meehl, P. E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy and Law*, *3*, 33-64. doi: 10.1037/1076-8971.3.1.33
- Jeffries, S., Fletcher, G. J. O., & Newbold, G. (2003). Pathways to sex-based differentiation in criminal court sentencing. *Criminology*, *41*, 329-354. doi: 10.1111/j.1745-9125.2003.tb00990.x
- Jimenez, M. A. (1997). Gender and psychiatry: Psychiatric conceptions of mental disorders in women. *Affilia* *12*, 154–175. doi: 10.1177/088610999701200202
- Johnstone, L. (n.d.). *Assessing risk for violence* [Powerpoint slides]. Retrieved from www.rmascotland.gov.uk/index.php/download_file/view/134/189/
- Juni, P., Hoenstein, F., Sterne, J., Bartlett, C., & Egger, M. (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology*, *31*(1), 115-23. <http://www.ncbi.nlm.nih.gov/pubmed/11914306>
- Kim, H. K., Capaldi, D. M., Pears, K. C., Kerr, D. C. R., & Owen, L. D. (2009).
- Kemshall, H. (2002). *Risk assessment and management of serious violence and sexual offenders: A review of current issues*. Edinburgh: Scottish Executive.
- Intergenerational transmission of internalizing and externalizing behaviours across three generations: Gender-specific pathways. *Criminal Behaviour and Mental Health*, *19*, 125-141. doi:10.1002/cbm.708

- Kline, P. (1986). *A handbook of test construction: introduction to psychometric design*. New York, NY, US: Methuen.
- Kline, P. (1993). *The handbook of psychological testing* (2nd ed.). Florence, Kentucky, US: Taylor & Frances/Routledge.
- Kline, P. (1998). *The new psychometrics: science, psychology and measurement*. Florence, Kentucky, US: Taylor & Frances/Routledge.
- Kötter, S., von Franquè, F., Bolzmacher, M., Eucker, S., Holzinger, B., Müller-Isberner, R. (2014). The HCR-20^{V3} in Germany. *International Journal of Forensic Mental Health*, 13, 122-129. doi: 10.1080/14999013.2014.911784
- Krawkowski, M., & Czobar, P. (2004). Gender differences in violent behaviours: Relationship to clinical symptoms and psychosocial factors. *American Journal of Psychiatry*, 161, 459-465. doi: 10.11.1176/appi.ajp.161.3.459
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment guide*. Vancouver, BC: British Columbia Institute on Family Violence.
- Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. World Health Organisation, Geneva. Retrieved from http://whqlibdoc.who.int/publications/2002/9241545615_eng.pdf?ua=1
- Kruttschnitt, C., & Carbone-Lopez, K. (2006). Moving beyond the stereotypes: Women's subjective accounts of their violent crime. *Criminology: An Interdisciplinary Journal*, 44, 321-351. doi: 10.1111/j.1745-9125.2006.00051.x
- Kruttschnitt, C., Gartner, R., & Ferraro, K. (2002). Women's involvement in serious interpersonal violence. *Aggression and violent behaviour*, 7, 529-565. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1359178901000453>

- Kubiak, S., Kim, W. J., Fedock, G., & Bybee, D. (2012). Assessing short-term outcomes of an intervention for women convicted of violent crimes. *Journal of the Society for Social Work and Research*, 3, 197-212. doi:10.5243/jsswr.2012.13
- Langan, J., & Ludlow, V. (2004). Living with risk: Mental health service user involvement in risk assessment and management. Bristol: Joseph Rowntree Foundation/ The Policy Press.
- Lavoie, J. A. A., Guy, L. S., & Douglas, K. S. (2009). 'Violence risk assessment: principles and models bridging prediction to management', In J.L. Ireland, C.A. Ireland and P. Birch (Eds.), *Violent and Sexual Offenders: Assessment, treatment and management*. (pp. 43-58). Devon: Willan Publishing.
- Lidz, C. W., Mulvey, E. P., & Gardner, W. (1993). The accuracy of prediction of violence to others. *The Journal of the American Medical Association (JAMA)*, 269, 1007-1011. doi:10.1001/jama.1993.03500080055032
- Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy and Law*, 7, 409-433. doi: 10.1037/1076-8971.7.2.409
- Liu, J., & Raine, A. (2006). The effect of childhood malnutrition on externalising behavior. *Current Opinion on Pediatrics*, 18, 565-570. doi: 10.1097/01.mop.0000245360.13949.91
- Lloyd, A. (1995). Doubly deviant, doubly damned: society's treatment of violent women: London: Penguin Books.
- Logan, C., & Blackburn, R. (2009). Mental disorder in violent women in secure settings: Potential relevance to risk for future violence. *International Journal of Law and Psychiatry*, 32, 31-38. doi:10.1016/j.ijlp.2008.11.010
- Lombroso, C., & Ferrero, W. (1895). *The Female Offender*: T. Fisher Unwin: London

- McGuire, J. (2008). A review of effective interventions for reducing aggression and violence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 2577-2597. doi: 10.1098/rstb.2008.0035
- McGuire, J. (2004). *Understanding Psychology and Crime: Perspectives on Theory and Action*. Maidenhead: Maidenhead McGraw-Hill International UK Ltd.
- McKenzie, B., & Curr, H. (2005). Predicting violence in a medium secure setting: a study of the historical and clinical scales of the HCR-20. *The British Journal of Forensic Practice*, 7, 22-28. Retrieved from <http://dx.doi.org/10.1108/14636646200500018>
- McKeown, A. (2010). Female offenders: assessment of risk in forensic settings. *Aggression and Violent Behavior*, 15, 422-429. doi: 10.1016/j.avb.2010.07.004
- McNeil, D. E., & Binder, R. L. (1994). Screening for risk of inpatient violence: Validation of an actuarial tool. *Law and Human Behavior*, 18, 579-586. Retrieved from <http://www.jstor.org/stable/1394257>
- Meehl, P.E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.
- Meichenbaum, D. (2006). *Comparison of aggression in boys and girls: A case for gender-specific interventions*. Retrieved from http://schoolviolenceprevention.com/Meich_06_genderdifferences.PDF
- Michel, S.F., Riaz, M., Webster, C., Hart, S.D., Levander, S., Müller-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. *International Journal of Forensic Mental Health*, 12, 1-13. doi: 10.1080/14999013.2012.760182

Ministry of Justice (2009). *Compendium of research and analysis on the Offender*

Assessment System (OASys) 2006-2009. Research Series 16/09. London: Ministry of Justice.

Ministry of Justice. (2010). *Breaking the cycle: Effective punishment, Rehabilitation and Sentencing of Offenders.* Retrieved from

<http://webarchive.nationalarchives.gov.uk/20120119200607/http://www.justice.gov.uk/consultations/docs/breaking-the-cycle.pdf>

Ministry of Justice. (2011). *Statistics on women in the criminal justice system 2011: a ministry of justice report under Section 95 of the Criminal Justice Act.* London: Ministry of Justice.

Ministry of Justice (2012) *Prisoners' childhood and family backgrounds.* London: Ministry of Justice

Ministry of Justice. (2012a). *Statistics on women in the criminal justice system 2011: A Ministry of Justice publication under Section 95 of the Criminal Justice Act 1991.*

Retrieved from

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220081/statistics-women-cjs-2011-v2.pdf.

Ministry of Justice. (2012b). *2012 Compendium of re-offending statistics and analysis-*

Ministry of Justice statistics bulletin. Retrieved from

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220081/statistics-women-cjs-2011-v2.pdf.

Ministry of Justice. (2014a). *Prison Population Figures: 2014.* Population Bulletin-Weekly 25 April 2014. Retrieved from <https://www.gov.uk/government/publications/prison-population-figures-2014>

- Ministry of Justice. (2014b). *Proven Re-offending Statistics Quarterly Bulletin July 2011 to June 2012, England and Wales*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305896/proven-reoffending-quarterly-bulletin-jul11-jun12.pdf
- Ministry of Justice. (2015). *Prison population bulletin: weekly 22 May 2015*. Retrieved from <https://www.gov.uk/government/statistics/prison-population-figures-2015>
- Moffitt, T. E. (1993). Adolescence-limited and life-cycle-persistent antisocial behaviour: A developmental taxonomy. *Psychological Review*, 100, 674-701. Retrieved from http://www.colorado.edu/ibs/jessor/psych7536-805/readings/moffitt-1993_674-701.pdf
- Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychology*, 2, 355-375.
- Monahan, J. (1986). Risk assessment of violence among the mentally disordered: Generating useful knowledge. *International Journal of Law and Psychiatry*, 11, 249-257. Retrieved from <http://www.sciencedirect.com/science/article/pii/016025278890012X>
- Monahan, J. (2006). A jurisprudence risk assessment: Forecasting harm among prisoners, predators and patients. *Virginia Law Review*, 92, 391-435. Retrieved from <http://www.jstor.org/stable/4144948>
- Monahan, J., & Steadman, H. J. (1996). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.

- Monahan, J., Steadman, H. J., Silver, E., Applebaum, P. S., Robbins, P. C., Mulvey, E. P., .& Banks, S. (2001). *Rethinking risk assessment: The McArthur study of mental disorder and violence*. Oxford: Oxford University Press.
- Mossman, D. (1994). Further comments on portraying the accuracy of violence prediction. *Law and Human Behavior, 18*, 587-594. Retrieved from <http://www.jstor.org/stable/1394258>
- Mossman, D. (1994). Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 64* (2), 783-792
- Monahan, J. (1981). *Predicting violent behaviour: An assessment of clinical techniques*. Washington, DC: Sage Publications.
- Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior, 23*, 107-120. doi: 10.1177/009385489602300100
- Morash, M. (2009). A great debate over using the Level of Service Inventory-Revised with women offenders. *Criminology and Public Policy, 8*, 173 - 182. doi: 10.1111/j.1745-9133.2009.00552.x
- Morgan, S. (2013). Risk decision making: Working with risk and implementing positive risk taking. [Monograph] Retrieved from <http://practicebasedevidence.squarespace.com/storage/RD-M%20Manual.pdf>
- Morgan, S. (2014). A challenge to evidence based practice: 'Re-shaping the risk agenda' [Web log message]. Retrieved from http://practicebasedevidence.squarespace.com/storage/pdfs/A_Challenge_to_Evidence_Based_Practice.pdf
- Morris, N., & Miller, M. (1985). Predictions of dangerousness. *Crime and Justice, 6*, 1-50. Retrieved from

http://www2.law.columbia.edu/fagan/courses/law_socialscience/documents/Spring_2006/Class%2023-CrimPro2/Morris_and_Miller.pdf

Motz, A. (2001). *The psychology of female violence. Crimes against the body*. East Sussex: Brunner-Routledge.

Nedopil, N. (2009). Gender specific and structured approach to violence risk assessment- the Munich prognosis project (MMP). *Proceedings from 17th EPA Congress, Lisbon, Portugal: European Psychiatry, 24* (1), S148.

Nevatti, J.R. (2011). The case against the HCR-20. [Monograph]. Retrieved from <http://www.1077evaluations.com/work-samples/The-Case-Against-the-HCR-20.pdf>

Nicholls, T.L. (2001). *Violence risk assessment with female NCRMD acquittees: validity of the HCR-20 and PCL-SV*. Unpublished Doctoral Dissertation, Simon Fraser University, Vancouver, British Columbia, Canada.

Nicholls, T. L., Greaves, C., & Moretti, M. (2008). Female aggression. In A. Jamieson & A. Moenssens (Eds.). *Wiley Encyclopedia of Forensic Science*. West Sussex: John Wiley & Sons.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among male and female psychiatric patients: The HCR-20, PCL-SV, and VSC. *Behavioral Sciences and the Law, 22*, 127-58. doi: 10.1002/bsl.579

Ogloff, J. R. (2006). The violent client: Advances in violence risk assessment. Retrieved from <http://www.psychology.org.au/publications/inpsych/risk/>

Ogloff, J. R., & Daffern, M. (2006). The dynamic appraisal of situational aggression: An instrument to assess risk for imminent aggression in psychiatric patients. *Behavioral Sciences and the Law, 24*, 799-813. doi: 10.1002/bsl.741

- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 609-626. doi: 10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K
- O'Shea, L.E., Mitchell, A.E., Picchioni, M.M., Dickens, G.L. (2012). 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:10.1016/j.avb.2012.11.016
- Otto, R. K., & Douglas, K. S. (2010). *Handbook of violence risk assessment tools*. Milton Park, UK: Routledge.
- Pallant, J. (2009). SPSS survival manual. *A step by step guide to data analysis using SPSS version 15* (3rd Ed.). Berkshire, England: Open University Press.
- Pollak, O. (1950). *The criminality of women*. Philadelphia: University of Pennsylvania Press
- Poster, E. C. and Ryan, J. A. (1994). A multiregional study of nurses' beliefs and attitudes about work safety and patient assault. *Hospital and Community Psychiatry 45*, 1104-1108. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7835857>
- Prosser, D., Johnson, S., Kuipers, E., Szmukler, G., Bebbington, P., & Thornicroft, G. (1997). Perceived sources of work stress and satisfaction among hospital and community mental health staff, and their relation to mental health, burnout and job satisfaction. *Journal of Psychosomatic Research, 43*, 51-59. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9263931>
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Raynor, P. (2007). Risk and need assessment in the British probation: the contributions of the LSI-R. *Psychology, Crime and Law, 13*, 125-138. doi: 10.1080/10683160500337592

- Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly*, 23, 384-405. doi: 10.1080/07418820600869152
- Rettinger, J.L. (1998). *A recidivism follow-up study investigating risk and need within a sample of provincially sentenced women* (Unpublished doctoral dissertation). Carleton University: Ontario, Canada.
- Rettinger, J.L., & Andrews, D.A. (2010). Gender risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and behavior*, 37 (1), 29-46. doi: 10.1177/0093854809349438
- Rettenberger, M., Matthes, A., Boer, D.P., Eher, R. (2010). Prospective actuarial risk assessment: a comparison of five risk assessment instruments in difference sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology*, 54 (2), 169-186. doi: 10.1177/0306624X08328755
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748. doi: 10.1037/1082-989X/1/2/178
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29, 615-620. doi: 10.1007/s10979-005-6832-7
- Robbins, P. C., Monahan, J., & Silver, E. (2003). Mental Disorder, Violence, and Gender. *Law and Human Behavior*, 27, 561-571. doi: 10.1023/B:LAHU.0000004886.13268.f2

- Rossberg, J. L., & Friis, S. M. D. (2003). Staff members' emotional reactions to aggressive and suicidal behaviour of inpatients. *Psychiatric Services*, 54, 1388-1394. Retrieved from <http://ps.psychiatryonline.org/doi/abs/10.1176/appi.ps.54.10.1388>
- Royal College of Psychiatrists (2008). *Rethining risk to others in mental health services. Final report to a scoping group*. Retrieved from <http://www.rcpsych.ac.uk/files/pdfversion/cr150.pdf>
- Rumgay, J. (2004). Scripts for safer survival: Pathways out of female crime. *The Howard Journal of Criminal Justice*, 43, 405-419. doi: 10.1111/j.1468-2311.2004.00338.x
- Rutter, M. (2003). Commentary: causal pathways to violence. *Development and Psychology*, 39, 372-378. doi: 10.1037/0012-1649.39.2.372
- Sackett, D. L., & Rosenberg, W. M. C. (1995). On the need for evidence-based medicine. *Journal of Public Health Medicine*, 17, 330-334. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8544145>
- Salisbury, E.J., Van Voorhis, P., & Spiropoulos, G.V. (2009). The predictive validity of a gender-responsive needs assessment: an exploration study. *Crime & Delinquency*, 55 (4), 550-585. doi: 10.1177/0011128707308102
- Serbin, L. A., Copperman, J. M., Peters, P.L., Lehoux, P. M., Stack, D.M., & Schwartzman, A. E. (1998). Intergenerational transfer of psychosocial risk in women with childhood histories of aggression, withdrawal, or aggression and withdrawal. *Developmental Psychology*, 34, 1246-1262. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9823509>
- 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.

- Journal of Forensic Psychiatry and Psychology*, 20, 354-365. doi:
10.1080/14789940802542873
- Schultz, K. C., & Whitney, D. J. (2005). *Measurement theory in action: Case studies and exercises*. Thousand Oaks, CA: Sage Publications.
- Scott, P. (1977). Assessing dangerousness in criminals. *British Journal of Psychiatry*, 131, 127-142. Retrieved from http://www.heron.dmu.ac.uk/2009-09-01/0007-1250_131%28127-142%2949039.pdf
- Shields, A., & Cicchetti, D. (1998). Reactive aggression among maltreated children: The role of attention and emotion dysregulation. *Journal of Clinical Child Psychology*, 27, 381-395. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9866075>
- Short, V., Lennox, C., Stevenson, C., Senior, J., & Shaw, J. (2012). Mental illness, personality disorder and violence: A scoping review. Offender Health Research Network: University of Manchester. Retrieved from <http://www.ohrn.nhs.uk/OHRNResearch/MIviolence.pdf>
- Silver, F. E., Felson, R. B., & Vaneseltine, M. (2008). The relationship between mental health problems and violence among criminal offenders. *Criminal Justice and Behavior*, 35, 405-426. doi:10.1177/0093854807312851
- Simpson, S. S., Yahner, J. L., & Duggan, L. (2008). Understanding women's pathways to jail: Analysing the lives of incarcerated women. *American and New Zealand Journal of Criminology*, 41, 85-108. doi: 10.1375/acri.41.1.84
- Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: a methodological primer. *Behavioral Sciences and the Law*, 31, 8-22. doi:
10.1002/bsl.2052

- Singh, J. P., Desmarais, S. L., & van Dorn, R. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences and Law, 31*, 55-73. doi: 10.1002/bsl/2053
- Singh, J. P., Fazel, S., Gueorguieva, R., & Buchanan, A. (2015b). Rates of violence in patients classified as high risk by structured risk assessment instruments. *British Journal of Psychiatry, 204*, pp.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 metaaggression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*, 499-513. doi: 10.1016/j.cpr.2010.11.009
- Singh, J. P., Yang, S., Mulvey, E. P., & The RAGGE Group. (2015a). Reporting guidance for violence risk assessment predictive validity studies: The RAGEE statement. *Law and Human Behavior, 39*, 15-22. doi: 10.1037/lhb0000090
- Simpson, S. S., Yahner, J. L., & Dugan, L. (2008). Understanding women's pathways to jail: Analysing the lives of incarcerated women. *The Australian and New Zealand Journal of Criminology, 41*, 84-108. Retrieved from <http://crim.umd.edu/sites/ccjs.umd.edu/files/pubs/Pathways.pdf>
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Psychological Science, 20*, 38-42. doi: 10.1177/0963721410397271
- Skeem, J. L., Monahan, J., & Mulvey, J. (2002). Psychopathy, treatment involvement and subsequent violence among civil psychiatric patients. *Law and Human Behavior, 26*, 577-603. Retrieved from <http://download.springer.com/static/pdf/863/art%253A10.1023%252FA%253A1020>

[993916404.pdf?auth66=1399707108_4fd2ea534747d5d9a9150c2c757e1a15&ext=.pdf](#)

- Skeem, J., Schubert, C., Stowman, S., Beeson, S., Mulvey, E., Gardner, W., & Lidz, C. (2005). Gender and risk assessment accuracy: Underestimating women's violence potential. *Law and Human Behavior, 29*, 173-86. doi: 10.1007/s10979-005-3401-z
- Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,373 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy, 8*, 183-208. doi: 10.1111/j.1745-9133.2009.00551.x
- Snyder, H., & Mulako-Wangota, J. (04-Apr-14). Arrest Data Analysis Tool at www.bjs.gov. Washington, DC: Bureau of Justice Statistics.
- Song, F., Parekh, S., Hooper, L., Loke, Y.K., Ryder, J., Sutton, A.J., Hing, C, Kwok, C.S., Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health Technology Assessment, 14*, S(8). doi: 10.3310/hta14080
- Steffensmeier, D., & Schwartz, J. (2004). Contemporary Explanations of Female Offending. In B. Raffel Price & N. J. Sokolo (Eds.), *The Criminal Justice System and Women: Offenders, Victims, and Workers*. New York: McGraw Hill, Inc.
- Strand, S., & Belfrage, H. (2001). *Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities*. *Psychology, Crime and Law, 7*, 71-79. doi: 10.1080/10683160108401784
- Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2014). The validity of version 3 of the HCR-20 violence assessment scheme among offenders and civil psychiatric patients. *International Journal of Forensic Mental Health, 31*, 148-159. doi: 10.1080/14999013.2014.911785

- Stuart, H. (2003). Violence and mental illness: An overview. *World Psychiatry*, 2, 121-124.
Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1525086/>
- Strub, D. S., Douglas, K. S., Nicholls, T. L. (2014). The validity of version 3 of the HCR-20 violence risk assessment rating scheme amongst offenders and civil psychiatric patients. *International Journal of Forensic Mental Health*, 13, 148-159. doi: 10.1080/14999013.2014.911785
- Suter, J. M., Bryne, M. K., Bryne, S., Howells, K., & Day, A. (2002). Anger in prisoners: Women are different than men. *Personality and Individual Differences*, 32, 1087-1100. doi:10.1016/S0191-8869(01)00105-2
- Swanson, J. W., Swartz, M. S., Essox, S. M., Osher, F. C., Wagner, R., Goodman, L.A. . . . Meador, K.G. (2002). The social-environmental context of violent behavior in persons treated for severe mental illness. *American Journal of Public Health*, 92, 1523-1531. doi: 10.2105/AJPH.92.9.1523
- Szmukler, G. (2001). Violence risk prediction in practice. *The British Journal of Psychiatry*, 178, 84-85. doi: 10.1192/bjp.178.1.84
- Tao, A. R., Holley, S. R., Leary, M., & McNeil, D. E. (2012). The relationship between level of training and accuracy of violence risk assessment. *Psychiatry Service*, 63, 1089-1094. doi: 10.1176/appi.ps.201200019
- Thomas, W. I. (1923). *The unadjusted girl*. New York: Harper and Row.
- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risk for sexual and violent recidivism. *Annals of New York Academy of Sciences*, 989, 225-235. doi: 10.1111/j.1749-6632.2003.tb07308.x

- Tjaden, P., & Thonnes, N. (2000). Extent, nature and consequences of intimate partner violence. (NCJ Pub. No. 181867). Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/181867.pdf>
- Underwood, M. K., Kupersmidt, J. B., & Coie, J. D. (1996). Childhood peer sociometric status and aggression as predictors of adolescent childbearing. *Journal of Research on Adolescence*, 6, 201-223. Retrieved from <http://psycnet.apa.org/psycinfo/1996-03765-004>
- United States Department of Justice. (2010). National Incident-Based Reporting System. Retrieved from <http://sociology-data.sju.edu/3031/3031cb.pdf>
- University of Birmingham. (2013). *Code of Practice*. Retrieved from <http://www.birmingham.ac.uk/Documents/university/legal/uob-code-of-ethics.pdf>
- Ussher, J. M. (2013). Diagnosing difficult women and pathologising femininity: Gender bias in psychiatric psychology. *Feminism and Psychology*, 23, 63-69. doi: 10.1177/0959353512467968
- Valliant, P.M., Gristey, C., Pottier, D., & Kosmyna, R. (1999). Risk factors in violence and nonviolent offenders. *Psychological Reports*, 85, 675-680.
- Van der Knaap, L. M., Alberda, D. L., Oosterveld, P., & Born, M. P. (2012). The predictive validity of criminogenic needs for male and female offenders: comparing the relative impact of needs in predicting recidivism. *Law and Human Behavior*, 36, 413-422. doi: 10.1037/h0093932
- van der Kolk, B. (2003). The neurobiology of childhood trauma and abuse. *Child and Adolescent Psychiatric Clinics of North America*, 12, 293-317. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12725013>

- Van Montfoort, A., & Relcasserig Nederland. (2004). *RISc version 1.0 Risk Assessment Scales*. Harderwijk, the Netherlands: Flevodruk.
- Van Voorhis, P., Salisbury, E., Wright, E., & Bauman, A. (2008). Achieving accurate pictures of risk and identifying gender responsive needs: two new assessments for women offenders. Retrieved from <http://www.uc.edu/content/dam/uc/womenoffenders/docs/NIC%20Summary%20Report.pdf>.
- van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk needs assessment. The current status of a gender-responsive approach supplement. *Criminal Justice and Behavior*, 37, 261-288. Retrieved from <http://www.uc.edu/content/dam/uc/womenoffenders/docs/CJB%202010.pdf>
- Vitacco, M. J., Gonsalves, V., Tomony, J., Smith, B. E. R., & Lishner, D. A. (2012). Can standardized measures of risk predict inpatient violence? Combining static and dynamic variables to improve accuracy. *Criminal Justice and Behavior*, 39, 589-606. doi:10.1177/0093854812436786
- Vitacco, M. J., van Rybroek, G. J., Rogstad, J. E., Yahr, L.E., Tomony, J. D., & Saewert, E. (2009). Preidcting short-term institutional aggression in forensic patients: A multi-trait method for understanding subtypes of aggression. *Law and Human Behavior*, 33, 308-319. Retrieved from <http://dx.doi.org/10.1007/s10979-008-9155-7>
- Vose, B., Lowenkamp, C.T., Smith, P., & Cullen, F.T. (2009). Gender and predictive validity of the LSI-R: a study of parolees and probationers. *Journal of Contemporary Criminal Justice*, 25 (4), 459-471. doi:10.1177/1043986209344797
- Ware, J., Cieplucha, C., & Matsuo, D. (2011). *The violent offenders therapeutic programme (VOTP) - Rationale and effectiveness*. Australasian Journal of Correctional Staff

Development. Retrieved from http://www.bfcsa.nsw.gov.au/_data/assets/pdf_file/0004/278041/The-Violent-OffendersTherapeutic-Programme.pdf

Ward, T., Polaschek, D. L. L., & Beech, A. (2006). *Theories of sexual offending*. West Sussex, John Wiley & Sons.

Warner, R. M. (2008). *Applied Statistics: From Bivariate Through to Multivariate Techniques*. California: Sage Publications, Inc.

Warren, J. L., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., & van Pattern, 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, 269-289. doi: 10.1016/j.ijlp.2003.09.012

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

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

Weyman A. K. & Kelly C. J. (1999). *Risk perception and risk communication: A review of the literature (Report CRR 248/1999)*. Sheffield: Health and Safety Laboratory, Health and Safety Executive. Retrieved from http://www.hse.gov.uk/research/crr_pdf/1999/crr99248.pdf

Williams, W. H., Cordan, G., Mewse, A. J., Tonks, J., & Burgess, C. N. W. (2010). Self-reported traumatic brain injury in male young offenders: A risk factor for re-offending, poor mental health and violence? *Neuropsychological Rehabilitation*, 20, 801-812. doi: 10.1080/09602011.2010.519613

- Witt, P. H. (2000). A practitioners view of risk assessment: The HCR-20 and SVR-20. *Behavioral Science and Law, 18*, 791-798. doi: 10.1002/bsl.420
- World Health Organisation. (2004). *The economic dimensions of interpersonal violence* (Report No. HV 6625). Retrieved from <http://whqlibdoc.who.int/publications/2004/9241591609.pdf?ua=1>
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767. doi: 10.1037/a0020473
- Yates, J. F., & Stone, E. R. (1992). In: Yates, J.F. (Ed.), *The risk construct in risk-taking behavior* (pp. 1-25). Chichester, UK: Wiley.
- Yudofsky, S. C., Silver, J. M., Jackson, W., Endcott, J., & Williams, D. (1986). The Overt Aggression Scale for the objective rating of verbal and physical aggression. *The American Journal of Psychiatry, 143*, 35-39.
- Zaplin, R. T. (2008). *Female Offenders: Critical perspectives and effective interventions*. Sudbury, MA: Jones and Bartlett Publishers.

Appendices

Appendix I
HCR-20^{V2} and HCR-20^{V2} FAM risk factors



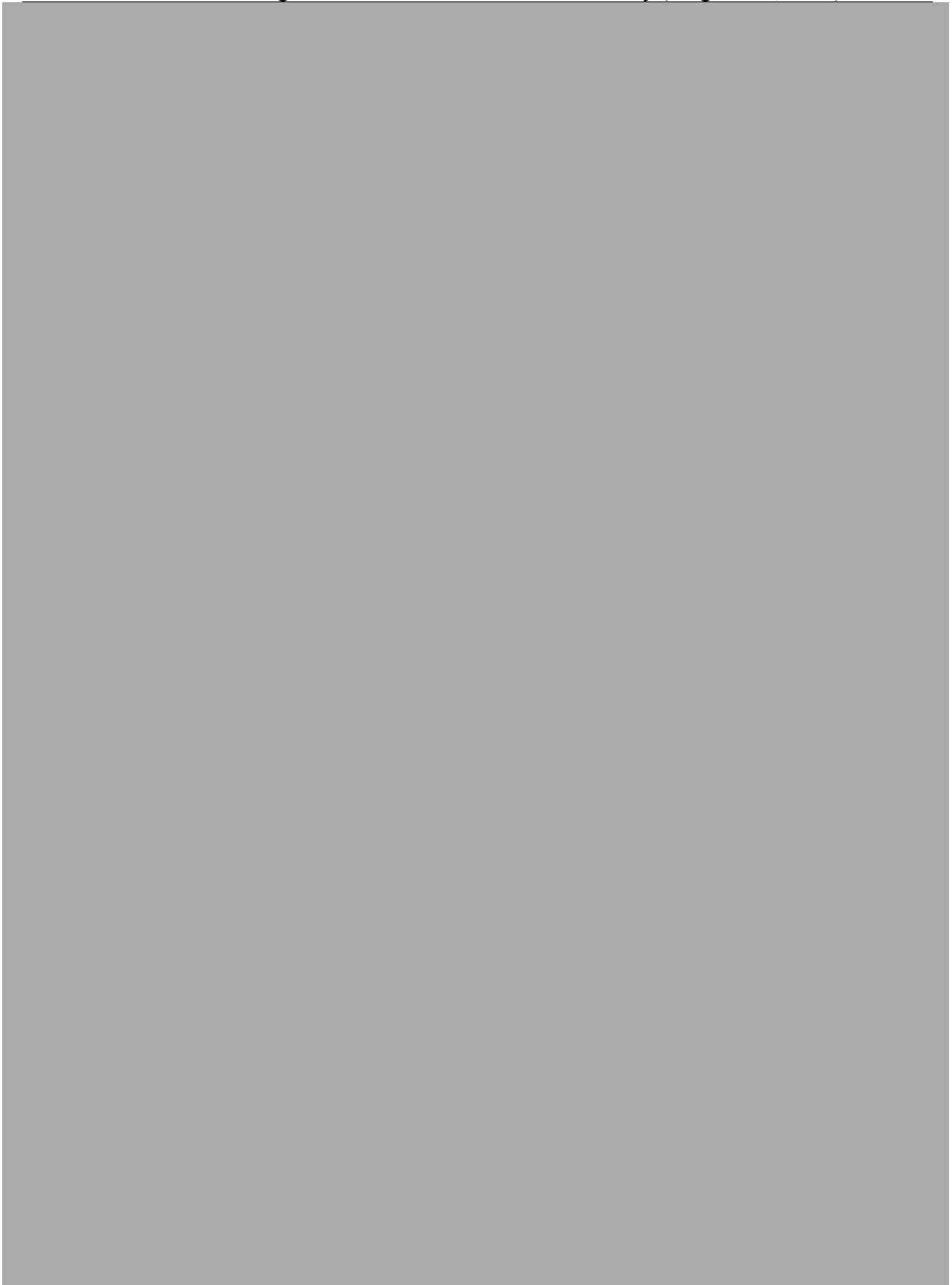
Taken from De Vogel et al. (2011, pp.28-29).

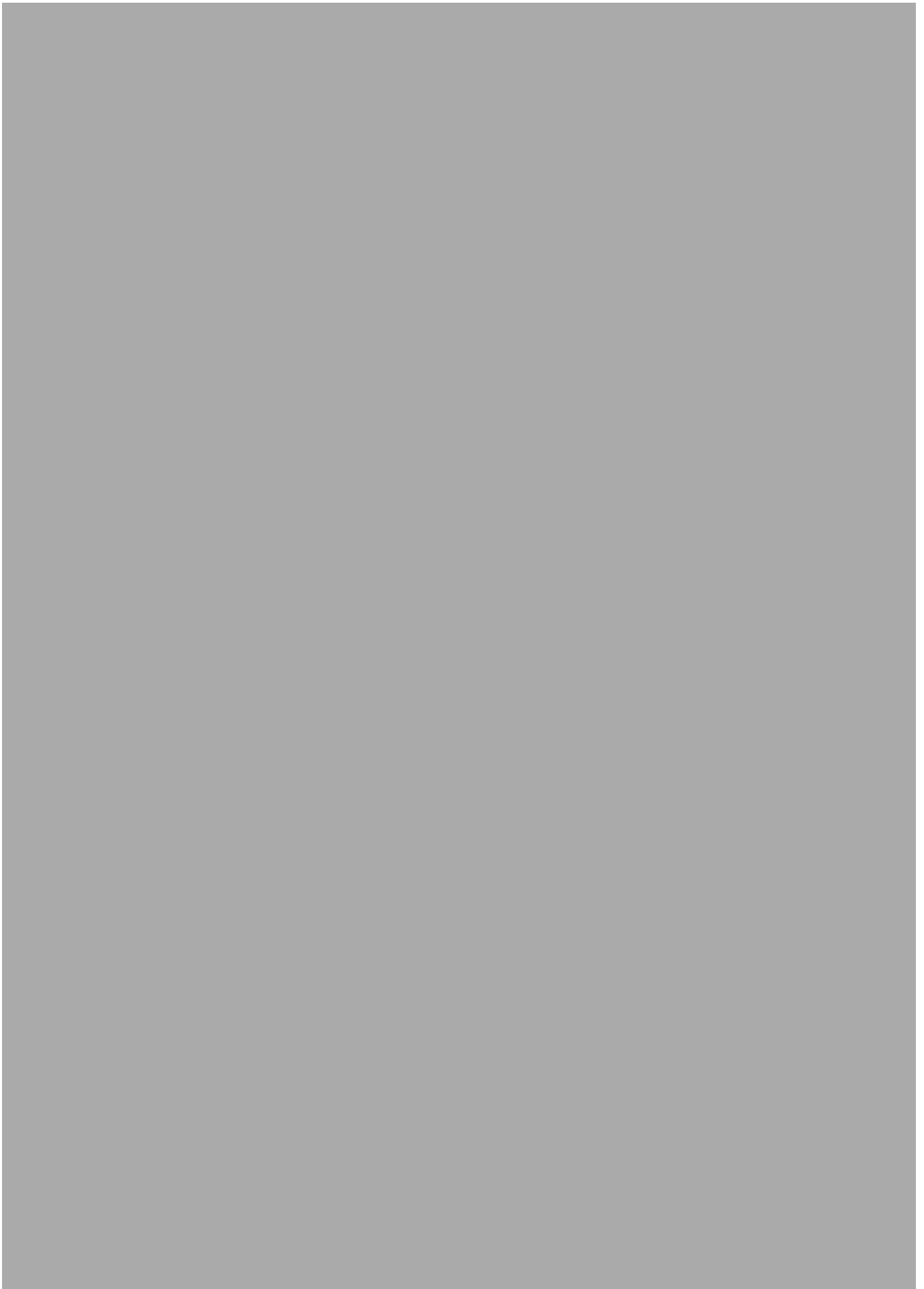
Appendix II
HCR-20^{V2} and HCR-20^{V3} risk factors



Appendix III
RAGEE Checklist

RAGEE Risk assessment guidelines for the evaluation of efficacy (Singh et al., 2015)







Taken from Singh et al. (2015).

Appendix IV ROC curves of predictive accuracy of each risk tool

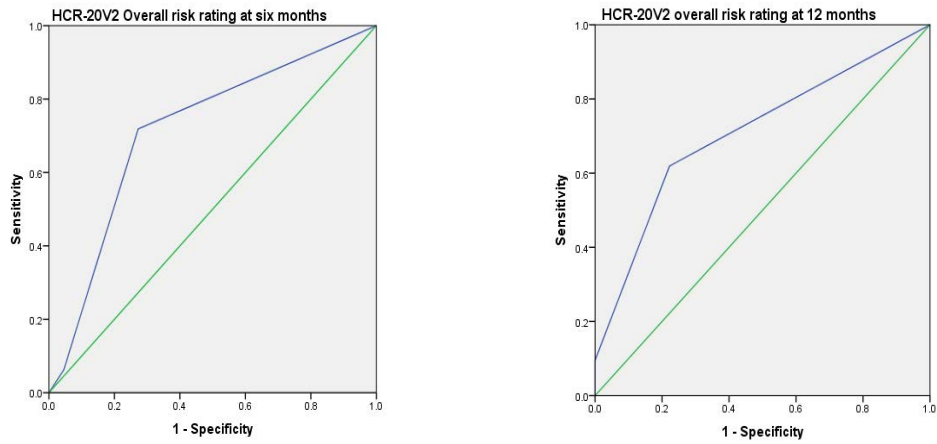


Figure 2-3. Predictive accuracy of HCR-20^{V2}

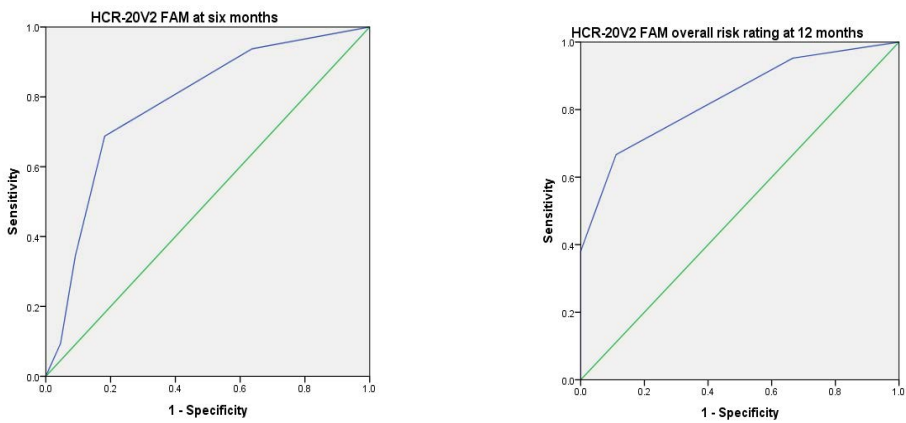
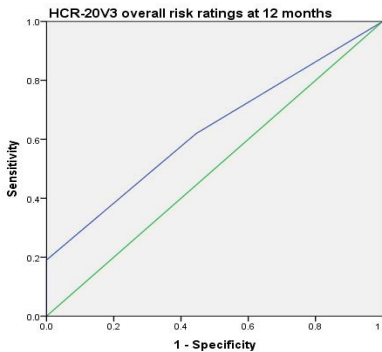
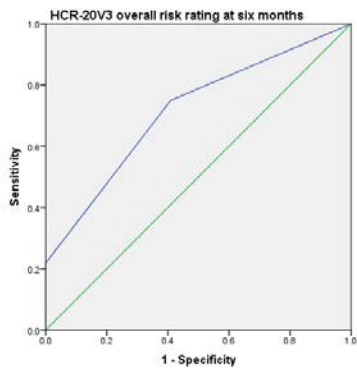


Figure 4-5. Predictive validity of HCR-20^{V2} FAM

Assessing risk in female offenders



Figures 6-7. Predictive accuracy of HCR-20^{V3}

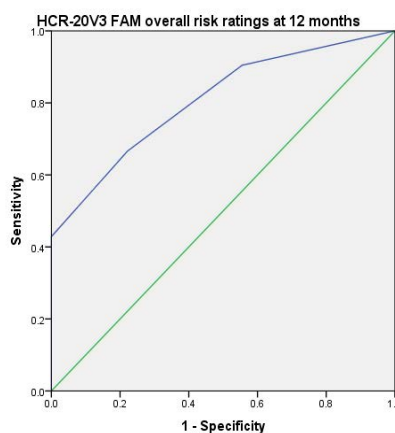
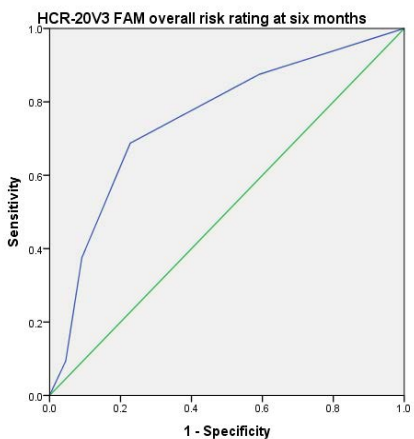


Figure 8-9. Predictive accuracy of HCR-20^{V3}FAM