Gender-sensitive violence risk assessment:
A Preliminary investigation into the predictive accuracy and clinical utility of the Female Additional Manual

by

Duncan G. Greig
B.A., University of British Columbia, 2009

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Arts

in the
Department of Psychology
Faculty of Arts and Social Sciences

© Duncan G. Greig 2014
SIMON FRASER UNIVERSITY
Summer 2014
Approval

Name: Duncan G. Greig
Degree: Master of Arts (Psychology)
Title: Gender-sensitive violence risk assessment: A preliminary investigation into the predictive accuracy and clinical utility of the Female Additional Manual

Examiner Committee: Chair: Dr. Robert Ley, Ph.D.
Associate Professor
Department of Psychology
Simon Fraser University

Ronald Roesch, Ph.D.
Senior Supervisor
Professor
Department of Psychology
Simon Fraser University

Tonia L. Nicholls, Ph.D.
Supervisor
Associate Professor
Department of Psychiatry
University of British Columbia

Denise Zabkiewicz, Ph.D.
External Examiner
Assistant Professor
Faculty of Health Sciences
Simon Fraser University

Date Defended/Approved: July 30, 2014
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files (“Work”) (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU’s own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU’s rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author’s written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author’s knowledge, infringe upon anyone’s copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010
Abstract

Most violence risk assessment measures were developed using male samples. Few measures consider gender in the assessment of risk. Nevertheless, research indicates that tools like the Historical-Clinical-Risk Management-20 (HCR-20) perform as well for women as men. However, some studies reveal poorer predictive accuracy for women. The Female Additional Manual (FAM) was developed to improve upon the accuracy of the HCR-20 for women; however, very little research with the tool has been conducted. This present prospective, longitudinal study compared the predictive accuracies of the HCR-20 and FAM in 62 male and 41 female chronically ill civil psychiatric patients. Data collection involved file review, and patient and treatment staff interviews across three time points. Results showed no gender differences in violent behaviours. Unexpectedly, few HCR-20 variables showed satisfactory predictive accuracy. Certain FAM variables showed better accuracy. This study is among the first to examine the validity of the FAM.

Keywords: violence; risk assessment; gender; women; mental disorder; psychiatry
I dedicate this work to my parents
who have steadfastly supported me
through every twist and turn
that I have taken on my path
Acknowledgements

I wish to thank my senior supervisors Drs. Tonia Nicholls and Ron Roesch for their continuing guidance and encouragement, and for the seemingly endless opportunities with which they have provided me. I thank Dr. Denise Zabkiewicz for contributing her time and helpful feedback. I also owe thanks to Simone Viljoen, whose “encouragement” and good humour helped make this project possible.

I also wish to thank Ashley Pritchard, whose own periods of stress fortuitously alternated with mine, and whose understanding, support, and positive spirit continue to help me find balance.
# Table of Contents

Acknowledgements .......................................................................................................................... vii  
Dedication ......................................................................................................................................... vi  
Abstract ............................................................................................................................................. iv  
Ethics Statement ................................................................................................................................. iv  
Partial Copyright Licence .................................................................................................................... iii  
Partial Copyright Licence ..................................................................................................................... ii  
Ethics approval ................................................................................................................................... ii  
Introduction ......................................................................................................................................... 1  
1.1. Rates and Characteristics of Female-perpetrated Violence ....................................................... 1  
1.2. Gender-specific and Gender-neutral Theories of Female-perpetrated Violence ....................... 3  
1.3. Mental Illness, Gender, and Violence ....................................................................................... 4  
1.4. Violence Risk Assessment and Gender .................................................................................... 7  
1.5. Female Additional Manual ........................................................................................................ 10  
1.6. Objectives .................................................................................................................................. 12  
1.7. Hypotheses ................................................................................................................................. 12  
Method ................................................................................................................................................ 14  
2.1. Participants ................................................................................................................................. 14  
2.2. Procedure ................................................................................................................................... 15  
2.2.1. Assessors .................................................................................................................................. 15  
2.2.2. Baseline assessment ............................................................................................................... 16  
2.2.3. File review .............................................................................................................................. 16  
2.2.4. Participant interviews ........................................................................................................... 16  
2.2.5. Follow-up assessments ......................................................................................................... 17  
2.2.6. Staff interviews ..................................................................................................................... 17  
2.2.7. Coding of the FAM ............................................................................................................... 18  
2.2.8. Ethics approval ...................................................................................................................... 18  
2.3. Measures .................................................................................................................................... 18  
2.3.1. Historical-Clinical-Risk Management-20 ............................................................................ 18  
2.3.2. Female Additional Manual .................................................................................................. 19  
2.3.3. Definition of violence and outcomes measure ..................................................................... 20  
Data Analytic Approach ...................................................................................................................... 22  
3.1. Data Preparation ....................................................................................................................... 22  
3.1.1. Variable Treatment ............................................................................................................... 22  
3.1.2. Diagnostics and Assumption Checks .................................................................................... 22  
3.2. Statistical Analyses ..................................................................................................................... 23  
3.3. Power Analysis ............................................................................................................................ 25  
Results ................................................................................................................................................. 26  
4.1. Sample Description ..................................................................................................................... 26
   4.2.1. Prevalence rates of violence in overall sample ................................................. 26
   4.2.2. Gender differences in violence by type ............................................................. 27
   4.2.3. Gender differences in rates of violence by severity .......................................... 27
   4.2.4. Gender differences in frequency of violent behaviours ................................... 27
4.3. Measure Descriptives and Correlations ...................................................................... 28
4.4. Interrater Reliability .................................................................................................... 29
4.5. Predictive Validity ....................................................................................................... 29
   4.5.1. Any violence .......................................................................................................... 29
   4.5.2. Violence by type .................................................................................................... 30
   4.5.3. Violence by severity .............................................................................................. 31
   4.5.4. Self-destructiveness and victimization .................................................................. 32
4.6. Incremental Validity ..................................................................................................... 32

Chapter 5. Discussion ....................................................................................................... 33
5.2. Gender Differences in Violent Behaviour ..................................................................... 35
5.3. Descriptive Differences between HCR-20 and FAM and Associations with Violence ............................................................................................................................... 35
5.4. Comparing the Predictive Validities of the HCR-20 and the FAM ............................... 36
   5.4.1. HCR-20 .................................................................................................................. 36
   5.4.2. FAM ....................................................................................................................... 37
5.5. Limitations and Strengths of the Present Study ........................................................... 40
5.6. Clinical Implications ................................................................................................. 41

References ......................................................................................................................... 43
Appendices .......................................................................................................................... 54
Appendix A. Table 1 ............................................................................................................. 54
Appendix B. Table 2 ............................................................................................................. 55
Appendix C. Table 3 ............................................................................................................. 56
Appendix D. Table 4 ............................................................................................................. 58
Appendix E. Table 5 ............................................................................................................. 59
Appendix F. Table 6 ............................................................................................................. 60
Appendix G. Figure 1 .......................................................................................................... 61
Appendix H. Figure 2 .......................................................................................................... 62
Appendix I. Figure 3 .......................................................................................................... 64
Chapter 1. Introduction

1.1. Rates and Characteristics of Female-perpetrated Violence

That the vast majority of research on offending and violence risk has been undertaken in correctional samples of males is not surprising given the extent of male overrepresentation in criminal justice systems around the globe. Official justice statistics demonstrate that women are significantly less likely than men to be arrested, charged, convicted, and incarcerated for criminal offences. In North America, between 1980 and 2011, women have, on average, represented just 20% of all individuals arrested for crimes, and comprised only 6-7% of U.S. and Canadian federal inmate populations (Bureau of Justice Statistics, 2011; Federal Bureau of Investigation, 2008; Statistics Canada, 2011). Even more gendered are statistics regarding violent offending, for which ‘maleness’ is widely considered to be one of the most potent predictors (Archer & McDaniel, 1995; Boritch, 1997; Monahan et al., 2001). While there is evidence that this gender gap is narrowing as increases in women’s violent offending have outpaced men’s over the past few decades (Glaze, 2011; Steffensmeier, 1996), important differences continue to be evident when looking at characteristics of offending, such as typology, severity, and victimology.

Whereas men are disproportionately more likely to commit violent crimes (e.g., sexual assault, robbery, serious assault, and homicide), women’s offending is typically characterized by lower severity, non-violent crimes including property offences, theft, drug offences, public-order failures/breaches, embezzlement, and prostitution (FBI, 2005, 2006, 2008; Javdani, Sadeh, & Verona, 2011; McKeown, 2010; Owen & Bloom, 1995; Pollock & Davis, 2005; Statistics Canada, 2011). For the minority of violent offenders who are female (14%; see Greenfield & Snell, 1999), research has shown that their crimes are typically less ‘visible’ (i.e., less likely to be reported), as they tend to
target family members or intimate partners rather than strangers (Monahan et al., 2001; Nicholls, 2001; Robbins, Monahan, & Silver, 2003); their violence tends to cause far less harm to its victims (Bureau of Justice Statistics, 2010; Denno, 1994; Nicholls, Greaves, & Moretti, 2008; Statistics Canada, 2011); and it may be more reactive and less instrumental as compared to their male counterparts (Crick & Grotpeter, 1995; Monahan et al., 2001; Nicholls, 2001; Odgers et al., 2005; Robbins et al., 2003). Further, some research has found that self-defense, jealousy, and perceptions of disrespect by others may play a larger role in motivating violence in women than in men (Kruttschnitt & Carbone-Lopez, 2006); and women are less likely to use weapons than men (Koons-Wit & Schram, 2003).

Although findings with regard to differences in recidivism rates are somewhat mixed, Langan and Levin (2002) found that of 272,111 inmates released from federal detention centers in the United States, women were less likely than men to be re-arrested (57.6% vs. 68.4%); reconvicted (39.9% vs. 47.6%); resented (17.3% vs. 26.2%); and re-incarcerated (39.4% vs. 53.0%) over a period of three years. Further, women recidivate with less serious offences than men (Spjeldnes & Goodkind, 2009), and consistently produce lower scores and different profile patterns on recidivism risk assessment measures (Andrews et al., 2012; Van Voorhis, Wright, Salisbury, & Bauman, 2010). Moreover, women may even be less likely to recidivate than their risk profiles would suggest (Blanchette & Brown, 2006; Hannah-Moffat, 2009; Holtfreter & Cupp, 2007). While women continue to constitute a decided minority of those who offend, rates of female offending have steadily been climbing, and violent offence rates for adult women have nearly tripled over the last three decades (Manchak, Skeem, Douglas, & Siranosian, 2009; Bureau Justice Statistics, 2013, 2011; FBI, 2008; Statistics Canada, 2011). Importantly, these increases have been observed even against a backdrop of decreases in overall and male crime rates, and is seen across offence types including homicide, rape, robbery, and aggravated assault (FBI, 1990, 2007).

The reasons for the apparent surge in women’s conflict with the law are not immediately clear, and have been the source of discussion in the sociological and criminological literatures. Some theories implicate fundamental changes in women’s behaviour secondary to shifting social norms about gendered behaviour, and others point to “net-widening” changes in law and policy that have been eroding a traditional
reluctance to prosecute women (Chunn and Gavigan 1991; Jeffries, Fletcher, & Newbold, 2003; Schwartz, Steffensmeier, & Feldmeyer, 2009). Whichever the cause, it is clear that there are important differences between men and women in terms of the nature and quality of violence, calling into question whether there are differences in etiology.

1.2. Gender-specific and Gender-neutral Theories of Female-perpetrated Violence

Given the significant underrepresentation of women supervised under criminal justice systems internationally, and the associated perspective that violence is largely a male problem, it is not surprising that most of the research into the etiology of violence has focused on men. More traditional sociological literature on pathways to violence notes that men and women who are violent (e.g., offenders) tend to share sociodemographic characteristics such as low socioeconomic status, low education, underemployment, and minority group membership (Chesney-Lind & Shelden, 1994; Denno, 1994; Steffenmeiser & Allan, 1995); and concordantly, community rates of male criminality tend to coincide with rates of female criminality (Steffenmeiser & Allan, 1988; Steffenmeiser, Allan, & Streifel, 1989). Theories emanating from such findings are typically referred to as ‘gender-neutral’, and suggest that aggressive traits and propensity for violence develop early in women (just as they do in men) as a result of macro-level factors related to systemic disadvantage, and continue to impact individuals’ violence risk throughout their lifespans (Moffitt, Caspi, Rutter, & Silva, 2001; Serbin et al., 2004). Some scholars, however, assert that such theories fail to explain more micro-level gender differences in rate, severity, and pattern of violent behaviour, as well as gender differences in prevalence of health- and mental health-related dysfunction in violent populations.

‘Gender-specific’ theories on etiology seek to account for such differences by describing the role of factors such as genetic predisposition to antisociality (Carter, 1973; Eme, 1992), social cognitive skill development (Bennett, Farrington, & Huesmann, 2005), normative beliefs (Huesmann & Guerra, 1997), childcare needs and responsibilities (Koons, Burrow, Monash, & Bynum, 1997), and exposure to normalized
criminogenic conditions in childhood (Mears, Ploeger, & Warr, 1998) in moderating women’s risk for violence. Research indicates that women who do commit violence have typically experienced disproportionately high rates of maladjustment resulting from physical/sexual abuse or neglect as a child (25-40% in the general population; 85% in offenders), physical/sexual victimization in adulthood, oppressive and/or dysfunctional social and intimate relationships, major mental illness, and substance abuse/dependence (Abram, Teplin, & McClelland, 2003; Brown, Miller, & Maguin, 1999; Byrne & Howells, 2002; McKeown, 2010; Morash, Bynum, & Koons, 1998; Rowe, Vazsonyi, & Flannery, 1995; Teplin, Abram, & McClelland, 1996; Van Marle & Van der Kroft, 2007). It has been suggested that it is women’s relatively greater exposure to such deleterious lifestyle factors that is (for them) most potently associated with violence and offending behaviour. Further, some scholars posit that these risk factors, particularly substance misuse, interface with women’s heightened vulnerability to emotional distress, emotional dysregulation, underdeveloped coping skills, poor anger management, and low self esteem (all of which may exist to a greater degree than for men even prior to women’s first contact with the law) to form a type of ‘gendered pathway’ into violence and criminality (Salisbury & Van Voorhis, 2009; Sorbello, Eccleston, Ward, & Jones, 2002; Thomas & Pollard, 2001).

Recognizing the problems inherent in the gender-neutral assumption that women’s unique developmental, interpersonal, affective, and behavioural experiences are not important mediators of violent behaviour, Murdoch, Vess, and Ward (2012) proposed a descriptive model of both distal and proximal factors that may facilitate women’s decisions to perpetrate violence. While the specifics of the model are beyond the scope of this discussion, it is notable that the authors provide at least exploratory support for the centrality of risk factors such as emotion dysregulation, dysfunctional family systems, and early and repeated exposure to abuse and interpersonal violence, and suggest that violent women may be a more homogenous population than males with regard to their risks and needs.

1.3. Mental Illness, Gender, and Violence

Given the greater prevalence of these types of pernicious lifestyle/risk actors (e.g., historical/chronic victimization) in women of the general population as well as
populations who come into conflict with the law (Byrne & Howells, 2000; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Sorbello et al., 2002; Substance Abuse and Mental Health Services Administration, 2007), it is not surprising that women exhibit higher rates of psychiatric disorder in both the community and within correctional institutions (Byrne & Howells, 2000; Kessler, Chiu, Demler, Walters, 2005; Sorbello et al., 2002). The American Correctional Association (1990) reported that as many as two-thirds of incarcerated women require treatment for mental health-related problems, and more than three-quarters may meet criteria for any number of psychiatric disorders including schizophrenia-spectrum, bipolar, major depressive, generalized anxiety, panic, posttraumatic stress, and personality disorders (Jordan, Schlenger, Fairbank, & Caddell, 1996; Teplin et al., 1996). Further, Lewis (2006) reported that among incarcerated female offenders, 65.4% presented with substance abuse/dependency issues. While rates of drug and alcohol abuse/dependency in male correctional samples are not significantly different from women’s (Karberg & James, 2005), it has been argued that women’s unique developmental, interpersonal, affective vulnerabilities may be compounded in particularly harmful ways by co-morbid substance abuse, precipitating the onset of mental illness and perpetuating a cycle of behavioural dysfunction and violence (Lewis, 2000; Sorbello et al., 2002).

While some early epidemiological research reported that mental disorder is unrelated (or is even inversely associated) with violent behaviour (see Hafner & Bokey, 1982; Steadman et al., 1979), more recent and methodologically sound work has found that mental illness is among the important risk factors for violence (Hodgins et al., 2002; 2011; Steadman et al., 1998; Swanson et al., 1989), particularly in the context of co-morbid substance use and/or when characterized by certain symptoms of psychosis (threat/control-override delusions; Steadman, Mulvey, Monahan, Robbins, et al., 1998; Teasdale, Silver, Monahan, 2006). Importantly, however, Short, Thomas, Mullen, and Ogloff (2013) noted that diagnosis with schizophrenia predicts violent criminality independently of comorbid substance use. Results from the Epidemiology Catchment Area study demonstrated that the incidence of violence among individuals with mental illness (i.e., schizophrenia, depression, bipolar disorder, mania) was five times higher than in the general population, and 16 times higher when co-morbid with substance misuse (Swanson et al., 1989); and similar results were found in Hodgins et al. (2002)
and Elbogen & Johnson (2009). Given consistent empirical support for the link between mental illness and violence, as well as national and international statistics confirming (across populations) greater prevalence of mental illness in women as compared to men, a question arises as to how the criminological truism that men are more violent than women may function in psychiatric patient populations.

Strikingly, studies of gender and violence in psychiatric populations have repeatedly demonstrated that men and women do not differ in reported rates of violent behaviour whether in-hospital (e.g., Binder & McNiel, 1990; Lam, McNiel & Binder, 2000) or in the community (e.g., Hiday, Swartz, & Swanson, 1998; Lidz, Mulvey, & Gardner, 1993; Robbins et al., 1993; Steadman et al., 1998; Newhill, Mulvey, and Lidz, 1995). Findings from the MacArthur Violence Risk Assessment Study, which tracked 1,136 individuals for one year following their discharge from one of three acute inpatient psychiatry wards in the U.S., demonstrated that 29.7% of men and 24.6% of women engaged in at least one violent act (59.8% and 61.6%, respectively, when counting less serious acts of aggression) following their release (Steadman et al., 1998; Robbins et al., 2003). Moreover, research conducted with psychiatric samples has found that not only do men and women not significantly differ in terms of rate of violence, but also that they are comparable in terms of severity and quality of violent behaviour (Nicholls, Ogloff, & Douglas, 2004). Taken together with other findings from both the criminological and epidemiological literatures, it would appear that mental disorder may indeed bear greater consequences for women than it does for men (Brennan, Mednick, & Hodgins, 2000; Robbins et al., 2003). Therefore studying gender differences in psychiatric patient populations presents researchers with a unique opportunity to explore more deeply the causes of violent behaviour in different groups. Moreover, given that some scholars have deemed the underestimation of violence risk in women to be a significant hindrance to clinical risk assessment validity in general (see Coontz, Lidz, & Mulvey, 1994; de Vogel, de Vries Robbé, van Kalmthout, & Place, 2012), as well as empirical evidence demonstrating that women are assessed differently by clinicians making decisions regarding violence risk (Skeem et al., 2005), advances in knowledge in this field should enable clinicians to maximize the utility of their assessments (Hiday, Swartz, Swanson, Borum, & Wagner, 1998).
1.4. Violence Risk Assessment and Gender

Perhaps the most interesting yet challenging psycho-legal application of the extant knowledge on violence comes when clinicians are asked to conduct assessments of violence risk. Violence risk assessment is central to the protection of the public safety, as well as the provision of services to at-risk individuals across diverse contexts internationally including corrections, forensics, and civil psychiatry. It has thus been the topic of many hundreds of empirical works, all of which have brought about significant refinements to assessment procedures, as well as the development and widespread adoption of empirically derived assessment techniques (Douglas & Reeves, 2010; Monahan et al., 2001). Early studies on the accuracy of violence risk assessment focused on the unstructured approach to assessment, and demonstrated that the predictive validity of such methods was rarely much greater than chance, and varied greatly depending on the research setting (Monahan, 1996; Monahan et al., 2001; Otto, 1994; Litwack, 2001). These findings incited efforts to improve upon the capacity of clinicians to predict violence through the use of empirically based technologies such as the actuarial and structured professional judgment (SPJ) approaches.

The actuarial approach dispenses with clinical judgment altogether in favour of an algorithmic model wherein quantitative values for a finite number of weighted variables computes a probabilistic decision about future risk (e.g., Violence Risk Appraisal Guide [VRAG]; Harris et al., 1993). The estimates generated are referenced against group-based norms, and allow for no adaptation for sub-population considerations (e.g., gender-specific risk factors). Despite this, a study of 1,657 prison inmates awaiting release that explored gender differences in the predictive accuracies of five risk assessment tools including the VRAG found its predictive accuracy for men and women to be equivalently strong (Coid, Yang, Ullruch, & Zhang, 2009).

The SPJ approach, alternatively, acknowledges certain types of information shown in the literature to be associated with increased risk of violence (i.e., risk factors), and serves as an aide-memoire with which a clinician arrives at a research- and clinically-informed determination of risk. The Historical-Clinical-Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997; Douglas, Hart, Webster, & Belfrage, 2013) was one of the first risk assessment tools developed under the SPJ model, and is
now widely considered to be the gold standard in the field. The HCR-20 has been the subject of well over 100 independent empirical studies, and findings reveal strong predictive accuracies in correctional, forensic, and civil psychiatric samples (median area under the curve [AUC] = .91, .82, .79, respectively; Douglas & Reeves, 2010).

Despite considerable advances in violence risk assessment technology, as well as a burgeoning literature in support of the sex specificity of certain risk factors, the bulk of the empirical work that has considered these topics continues to support a gender-neutral approach to risk assessment (Andrews & Dowden, 2006; Heilbrun, Yasuhara, & Shah, 2010; Nicholls & Petrila, 2005). Further, due to the previously noted gender disparities in violent behaviour in all but psychiatric populations, to date the vast majority of the research has focused predominately on males (Garcia-Mansilla, Rosenfeld, & Nicholls, 2011). Proponents of the gender-neutral approach assert that the most important violence-associated risk factors are common to both males and females, and that any pronounced gender differences in base rates of violence are attributable to differential prevalence of these factors between men and women (Cale & Lilienfeld, 2002; Funk, 1999; Loucks & Zamble, 2000; Nicholls, Ogloff, & Douglas, 2004).

Contrastingly, proponents of the gender-sensitive approach to violence risk assessment propose that the prevailing gender differences in terms of the nature, severity, frequency, and etiology of women’s violence are important, and that they are associated with women’s unique person- and context-level experiences to a degree that make them an important consideration in risk assessment (de Vogel et al., 2012; de Vogel & de Ruiter, 2005; Garcia-Mansilla et al., 2011; Link et al., 1998; Moffitt, Caspi, Rutter, & Silva, 2001; Odgers, Moretti, & Reppucci, 2005; Robbins, Monahan, & Silver, 2003). Until only recently, however, relatively few efforts had been made to empirically study whether predictors and moderators of aggression are sex-specific.

A good body of research demonstrates that many ‘static’ risk factors (i.e., variables less amenable to change; e.g., history of child abuse) and ‘dynamic’ risk factors (i.e., changeable; e.g., substance use) are relevant for both men and women across populations (e.g., offenders, civil and forensic psychiatric patients; Garcia-Mansilla et al., 2009; Nicholls et al., 2008). Indeed when Blanchette and Taylor (2007) tested 176 historical and dynamic variables purportedly relevant to security classification
decision making for female offenders, the authors found that the most important risk factors did not differ significantly from those used in classifying male offenders. Importantly, however, the authors suggest that more fine-grained gender differences may exist in the relative weighting of those variables found to be common for men and women.

Some other studies that have examined the sex-specificity of risk factors, however, have found evidence that variables such as history of child abuse, adult victimization, dysfunctional familial and intimate relationships, and economic disadvantage may be more strongly associated with recidivism, violence, and other antisocial outcomes for women than for men. (Benda, 2005; Bottos, 2007; Funk, 1999; Odgers et al., 2005; Widom & Maxfield, 2001).

One of the most common ways in which researchers have investigated sex-specificity in risk factors is to simply examine how existing risk assessment measures perform for each gender. Of all the studies that have employed this approach, most support the gender-neutral perspective, including one large meta-analysis examining the predictive validity of nine risk assessment tools across 68 studies and 25,980 individuals. In this study, Singh, Grann, and Fazel (2011) found no evidence suggestive of better predictive validity for either gender using any tool examined. However, the authors caution that women constituted only 8% of the entire sample. Guy and Douglas (2006) used Item Response Theory to examine HCR-20 item equivalence across groups, and found no significant gender differences. Guy (2008) meta-analyzed 113 studies on the subject of the HCR-20, and concluded not only that gender was not a significant moderator of the tool’s predictive validity, but that there was even a trend towards slightly better accuracy for women. Further, the majority of studies looking specifically at gender as a moderator of the predictive validity of the HCR-20 provide no evidence that the tool performs substantially better with men than with women (Douglas & Reeves, 2010; also see de Vogel & de Ruiter, 2006; Nicholls et al., 2004; Strand & Belfrage, 2001). The empirical evidence for gender-neutrality in risk assessment described to date is compelling as it derives from some of the most methodologically rigorous research on the topic, and represents diverse samples and settings.
This is not to say that all research on gender differences in risk assessment tool performance unequivocally supports gender neutrality. A number of studies do provide some evidence of differential performance of the HCR-20 in certain samples. For example, Nicholls et al. (2004) found that, in a sample of 242 civil psychiatric patients (95 of whom were women), predictive accuracy for inpatient violence by women was good (in fact, better than for men); however, accuracy for follow-up physical violence perpetrated in the community was lower for women than for men. De Vogel and de Ruiter (2005) examined predictive validity of the HCR-20 in 42 personality-disordered female forensic patients in a Dutch psychiatric hospital and compared them with a matched male sample. Findings showed significantly poorer prediction, in women, of inpatient violence and violent reconviction post-discharge by all variables except for the final risk judgment. The authors also found significant descriptive differences in scores at the item level such that women scored higher on the Relationship Instability and Impulsivity items, and scored lower on the Young age at first violent incident, Psychopathy, and Negative attitudes items.

Based on these findings as well as those from other studies demonstrating relatively poor performance of the HCR-20 in all-female samples (e.g., Schaap et al., 2009; Warren, South, Burnette, et al., 2005), de Vogel and de Ruiter (2005) echoed calls made by Garcia-Mansilla, Rosenfeld and Nicholls (2009) and McKeown (2010) for continued work in the gendered risk assessment area. While these studies’ findings are intriguing, and such a conclusion is not unwarranted, the state of the field is such that the evidence supports use of the HCR-20 with females (McKeown, 2010). The literature would benefit from continued study of HCR-20 performance in different samples/settings, as thus far support for gender sensitivity in risk assessment using the HCR-20 has only been shown in short-stay, discharging civil psychiatric patients and a somewhat unique sample of forensic patients in the Netherlands.

1.5. Female Additional Manual

Another approach to studying the sex-specificity of risk factors for violence is to examine the performance of gender-informed risk measures – of which there are very few. To address the ambiguity in the literature discussed above, and in response to calls
for better understanding and professional training with regard to the factors specifically relevant for treatment and management of women (e.g., added influence of greater parenting responsibilities), de Vogel and her colleagues recently published the *Female Additional Manual* (FAM; 2012). The FAM (described in detail under Measures) operates as an addendum to the HCR-20 that provides additional guidelines to existing HCR-20 items, as well as additional empirically-based, treatment-relevant items purported to be critical considerations in the assessment of violence risk in women. The additional items include Historical items such as Prostitution and Pregnancy at young age; new Clinical items include Covert/Manipulative Behavior and Low Self-Esteem; and new Risk Management items include Problematic Childcare Responsibilities and Problematic Intimate Relationship (see Figure 2 for a full list of FAM items). In addition to violence, the tool also provides for the determination of risk of other adverse outcomes including self-destructive behaviour, victimization, and non-violent criminal offending.

Research into the use of the FAM is very much in its infancy, as very little (if any) research on the tool has been conducted outside of the Dutch forensic hospital in which it was developed. De Vogel and de Vries Robbe (2011) reported on their prospective study of the use of the FAM with 42 gender-matched female forensic patients. Results showed that females scored higher than men on 7 of the 9 new items (Prostitution, Pregnancy at young age, Suicide attempt/self-harm, Victimization after childhood, Covert/manipulative behavior, Low self-esteem, and Problematic intimate relationship). Males scored higher on Psychopathy and Problematic behavior during childhood. H, C, R, and Total scores predicted violence to others at AUC = .76, .66, .83, and .71, respectively; and they predicted self-destructiveness at AUC = .80, .68, .77, and .81, respectively. Final risk ratings predicted violence to others at AUC = .87; and self-destructive behaviour, victimization, and non-violent criminal offending at AUC = .97, .63, and .99, respectively. Predictive accuracy results for males were not presented.

These findings demonstrate that the FAM is a potentially useful tool, but much more research is needed in other samples/settings, and should include comparisons to established risk measures like the HCR-20 in order to make conclusions about the psychometric properties and clinical utility of the new measure.
1.6. Objectives

The current project was derived from a larger longitudinal follow-up study evaluating the reorganization and redevelopment of tertiary psychiatric services in British Columbia (BC) following the closure of the province’s only tertiary civil psychiatric hospital. Given some residual ambiguity in the gender and violence risk assessment debate, the lack of research using the FAM, and the suitability of a civil psychiatric population to elucidate any existing gender differences in factors associated with violence risk, the current study: a.) examined the psychometric properties (i.e., interrater reliability, predictive validity, incremental validity) of the FAM and HCR-20; b.) compared the predictive accuracies of these measures for a variety of violent and adverse behavioural outcomes over multiple time points (i.e., in hospital, two community follow-ups); and c.) explored gender differences/similarities in predictive accuracy of these measures in an adult residential civil psychiatric sample undergoing transfer from hospital to community care.

1.7. Hypotheses

With consideration for the literature reviewed in preparation for this study, it was anticipated that all metrics of the HCR-20 (i.e., scale scores, total score, final risk judgment) would demonstrate good predictive validity for all types violent outcomes at both follow-up time points; and it was expected that the $H$ scale would provide the highest AUC values, followed by the $C$ scale, and then $R$ scale (Douglas & Reeves, 2010); and it was reasonable to hypothesize that the HCR-20 would perform as well in women as it did in men (Guy, 2008). Given that a more sizeable literature points to gender equivalence in HCR-20 predictive accuracy, the FAM assessment is essentially a modification of the full HCR-20 assessment, and there is yet to be sufficient research to draw strong conclusions about the FAM’s incremental validity, it was anticipated that the FAM would perform as well as the HCR-20 in this sample. Specifically, it was expected that FAM total, scale, and items scores, as well as the summary risk ratings would demonstrate good predictive validity, but that the AUC values would not differ statistically from those on the HCR-20. Moreover, it was expected that the FAM metrics would show no evidence of incremental validity over those of the HCR-20. It was hypothesized that the summary risk ratings for the self-destructiveness, victimization, and non-violent
criminal behaviour on the FAM would show reasonably good prediction of associated outcomes, at least in the short term (i.e., first six month follow-up). It was expected that FAM metrics would show good interrater reliability (e.g., ICC > .80).
Chapter 2.  Method

2.1. Participants

Participants were 103 patients (40% female) transferring from Riverview Hospital (RVH) to smaller regionalized facilities as part of the redevelopment of psychiatric services in BC. Ages of this subsample ranged from 20 to 73 years ($M = 47.1$, $SD = 12.58$). Most patients were Caucasian (76.7%), and diagnosed with a schizophrenia spectrum disorder (85.4%). Thirty-four participants (33%) were diagnosed with a personality disorder or traits of a disorder, and 39 (37.9%) had a diagnosis with a co-morbid substance use disorder. At the time of participant enrolment in the larger study, RVH housed $N = 300$ individuals living with severe mental illness, most of whom were involuntarily civilly committed under the BC Mental Health Act (MHA, 1996). Inclusion criteria included sufficient capability to participate in interviews conducted in English, and receipt of assent from each participant’s psychiatrist, indicating that the study procedures would not cause undue distress and the individual was capable of providing informed consent. For participants deemed incapable of providing informed consent, substitute decision makers (identified by the treatment team) provided assent. Those participants incompetent to consent did not participate in interviews; but rather, data collection was restricted to file review and treatment staff interviews. Of the 230 patients that were approached in hospital by project recruiters, 162 (70.4%) were deemed by their psychiatrist to be competent to consent, and 68 (29.6%) were deemed incompetent (requiring the consent of a substitute decision maker [SDM]). Of these 230, 143 patients and five SDMs agreed to be consented. Thirty-three patients (22.3%) declined to participate, and the remaining three patients were not consented for other reasons. Ultimately, 107 patients (72.3% of those who agreed to be consented) were enrolled in the study. Four patients were not included in the present study due to missing data. Forty-seven (45.6%) participants were assigned to “full procedure” data collection (i.e., patient and staff interviews and file review). Data for the remaining 56 (54.4%) of the
sample were collected from file review only. Full procedure participants were remunerated $25 for baseline assessments, and $20 for each of the follow-up assessments; and those enrolled as file-only participants received $5 for their participation in the study.

2.2. Procedure

2.2.1. Assessors

All patient and staff interviews as well as file reviews were conducted by a team of highly trained research assistants (RAs) who had completed at least a Bachelor degree in a social sciences discipline (e.g., psychology, criminology, public health). These RAs were carefully selected on the basis of their previous academic achievements, research experience, and professionalism. All RAs received extensive training in the study protocol, and were formally trained on the use of the HCR-20 in order to ensure measure adherence to proper scoring standards and reliability of risk data. Training in proper coding of the FAM was provided by a senior clinical psychology graduate student working under the supervision of an author of the measure. Research assistants coding each of the risk measures as well as the outcome measures were all kept distinct across all time points in order to minimize bias.

Three different groups of RAs were assigned to collect data. Those assigned to the “risk” protocol conducted both file review and patient interviews, and primarily collected information needed to code included risk measures (e.g., HCR-20). Those assigned to the “health/outcome” protocol also conducted file review and patient interviews, but were responsible for collecting data primarily pertaining to clinical variables, health service usage and satisfaction, and incidence of adverse behavioural outcomes (e.g., aggression, self-harm, unauthorized leave; measure to be discussed further below). The third group of RAs completed interviews with staff members at the participants’ residences. For the larger project, all three groups collected data for each time point (i.e., baseline, 6 months, 12 months), and data collection extended across four follow up time points (6, 12, 18, and 24 months). However, for the baseline assessment in the present study, only risk assessment data (from file and participant
interview) were used. For the two follow up assessments included in this study (i.e., 6 and 12 months), all three sources of data were used. Only two time points were considered because both the HCR-20 and FAM are generally intended to best assess risk of violence for the antecedent six months to one year.

2.2.2. Baseline assessment

Data collection for the present study began with baseline assessments completed prior to patients' transferring out of hospital, and included file review and participant interview. Following completion of data collection at baseline, 25% of cases were re-coded by a different RA in order to determine interrater reliability. Additionally, official databases were accessed to collect information pertaining to deaths, transinstitutionalization, and criminalization. The vast majority of the measures included in the larger project's data collection protocol were not used in the current study, and will not be described here.

2.2.3. File review

At the outset of the baseline assessment, RAs completing the risk protocol reviewed participants' healthcare files, which included hospital admission and discharge reports, reports by physical and mental health professionals, and interdisciplinary clinical and behavioural progress notes documented by nurses, social workers, psychologists, and physicians. Research assistants recorded relevant file information by hand, and used the information to score the HCR-20. Review of file information for each participant required 1-3 hours to complete.

2.2.4. Participant interviews

Those participants for whom full procedure consent or assent was provided were then interviewed for approximately 2-3 hours. Research assistants assigned the risk protocol collected interview data relevant to the scoring of HCR-20 (and other measures included in the larger project). Again, all other data collected at baseline were not considered in the present study.
2.2.5. Follow-up assessments

Once a given participant was transferred from the hospital to the community facility, follow-up assessments were conducted every six months for two years (four follow-ups) in the patient's new place of residence. In most cases this was one of many smaller, regionalized community-based residential settings around Greater Vancouver for individuals with serious mental illness. These facilities, by-and-large, provide 24-hour staffing by nurses, and also employ physicians, occupational therapists, social workers, and other personnel in the daytime. Staff-patient contact and record keeping in these facilities closely resemble that in Riverview Hospital prior to its closure, which aids in matching the reliability of data collected at baseline and at follow-up.

Data collection for the larger project at follow-up time points consisted of file review and participant and staff interviews, which involved coding most of the same measures included in the baseline assessment (e.g., physical and mental health, life skills, risk variables). For the present study, the follow-up assessments introduced a measure of outcome incidence data (i.e., START-Outcomes Scale) coded for the preceding six-month period using information from file, participant interview, and staff interview. Data collected by this measure included the presence and frequency of a number of outcomes (i.e., four types of aggression, suicide-related behaviours, self harm, substance use, victimization; see Measures section below). Again, two research assistants coded 25% of the cases in order to evaluate interrater reliability.

2.2.6. Staff interviews

Research assistants assigned to the staff interview protocol asked treatment staff to provide information about each participants' behaviour (e.g., incidents of aggression and other outcomes) in addition to other information not considered in this study (e.g., participants’ functional independence). Interviews with treatment staff lasted approximately one hour, and staff received $5 gift certificates for their involvement in the study.
2.2.7. Coding of the FAM

Given that the FAM was developed and published well after the present study’s parent project was devised, the decision to introduce the FAM risk assessment into the study protocol occurred after the completion of data collection for the larger project (i.e., after patient interviews had been conducted). As a result, RAs coded the FAM for the 41 females in the sample from file information only. The re-review of the files for the purposes of coding the FAM was performed by different RAs than those who had originally reviewed those same files in coding the HCR-20. Two research assistants (graduate students) coded the FAM. In order to minimize bias, research assistants who previously coded baseline or follow-up risk assessments (including the HCR-20 and the START-Outcomes Scale) for given participants, as part of the larger project, did not code the FAM for those participants.

2.2.8. Ethics approval

Ethical approval had been previously obtained for the larger project from Simon Fraser University (SFU), the University of British Columbia (UBC), Kwantlen Polytechnic University, and the RVH research committee. The principal investigator of the larger project (Dr. Tonia Nicholls, University of British Columbia, BC Mental Health and Substance Use Services) was approved as an affiliated researcher and granted ethics approval by the Fraser and Vancouver Coast Health Authorities. Approval for amendments (including the addition of the FAM to the study protocol) to the existing ethics certificate was granted by SFU, UBC, Kwantlen, and the Fraser and Vancouver Coast Health Authorities. Ethics approval for the completion of this thesis project was granted by the SFU Office of Research Ethics.

2.3. Measures

2.3.1. Historical-Clinical-Risk Management-20

To assess risk of violence, the HCR-20 (Version 2, Webster et al., 1997) was used. As described above, the HCR-20 is a structured assessment of empirically
supported risk factors for interpersonal violence that may be present in a given case. The tool was developed under the SPJ model of risk assessment, and consists of 10 historical factors, five clinical factors, and five risk management factors. The evaluator codes each risk factor as 0 (not present), 1 (possible/less serious), or 2 (definite/serious); the item scores sum to a total possible score of 40. Higher scores indicate higher risk. The evaluator then considers the presence and relevance of the risk factors in order to make a summary risk rating of low, moderate, or high (see Figure 1). The authors of the HCR-20 caution that, as the HCR-20 conforms to the SPJ model, the use of numeric estimates of risk and associated cutoffs is not appropriate, particularly in clinical contexts (Douglas & Reeves, 2010); however, for the purposes of research they may be used. The HCR-20 has consistently demonstrated strong psychometric properties, including good internal consistency (α = .94, .90, .95 for H, C, and R scales; Dunbar, Quinones, & Crevecoeur, 2005); and total score interrater reliability coefficients of over .80 (Douglas & Reeves, 2010). Past research has found high interrater reliability for HCR-20 coding in this sample (i.e., all ICCs above 0.85; Petersen et al., 2012). In terms of predictive validity, as discussed earlier, a large body of empirical work demonstrates strong median AUC values across correctional, forensic, and civil psychiatric samples of .91, .82, and .79, respectively; Douglas & Reeves, 2010). Significantly fewer studies have investigated the predictive validity of the HCR-20 summary risk ratings, and most that have used forensic samples (Douglas & Reeves, 2010). When Guy (2008) meta-analyzed studies investigating the predictive validity of risk ratings, it was found that the average AUC for mixed definitions of violence was .70, and was .76 when the strict HCR-20 definition of violence was used, and .79 for physical violence.

2.3.2. Female Additional Manual

As described earlier, the Female Additional Manual (FAM; de Vogel et al., 2012) provides additional guidelines to five existing HCR-20 items as well as nine additional empirically based items specific to the assessment of violence risk in women (see Figure 2). As with the HCR-20, the evaluator codes each risk factor as 0 (not present), 1 (possible/less serious), or 2 (definite/serious), and then makes a summary rating regarding risk (i.e., low, moderate, high). The measure also includes the option to make
a summary risk judgment using a 5-point scale (i.e., low, low-moderate, moderate, moderate-high, and high), and includes risk ratings for self-destructive behaviour, victimization, and non-violent criminal behaviour. Thus far, the totality of the research on the tool has been conducted by the authoring research team. Preliminary research demonstrated that, not unexpectedly, women scored higher on all but two of the new gender-specific items (i.e., Psychopathy and Behaviour problems in childhood; de Vogel et al., 2012). In de Vogel and de Vries Robbé’s study (2011), adequate to good predictive accuracy was shown for the FAM total score, and $H$, $C$, and $R$ scale scores for violence (AUC = .76, .66, .83, .71, respectively, as well as for self-destructive behaviour (AUC = .80, .68, .77, .81). However, the research has not yielded any evidence supporting the predictive accuracy of the FAM-specific items, and the authors recommend further research.

2.3.3. Definition of violence and outcomes measure

Violence was defined as it is for the HCR-20; that is, “actual, attempted, or threatened harm to a person or persons” (Webster et al., 1997, p. 24). The START-Outcomes Scale (SOS; Nicholls et al., 2006) is a modified version of the Overt Aggression Scale – Modified (OAS-M; Yudofsky, Silber, Jackson, Endicott, & Williams, 1986), and was used to measure a range of adverse behaviours commonly seen in individuals with severe mental illness including self-harm, suicide ideation and behaviour, substance use, self-neglect, victimization, unauthorized leave, and four types of violence (verbal, against others, against property, sexual). Outcome incidence, severity, and chronicity data was coded retrospectively for the preceding six-month period via healthcare files (e.g., nursing/interdisciplinary progress notes), as well as patient and staff interviews. As this measure merely involves tallying instances of various behaviours, no reliability or validity data are available with the exception of one study showing adequate to good mean level of interrater agreement (ICC = .53 - .84; Wilson, Desmarais, Nicholls, & Brink, 2010). As seen in Figure 3, outcomes are rated on an SOS score sheet by type (e.g., sexual aggression, self-neglect, unauthorized leave), by severity (e.g., “Kicks, punches, bites. Actions result in mild-moderate
physical injury [e.g., bruises, sprain, welts]” versus “Attacks others, uses weapons, resulting in severe physical injury [e.g., fracture, loss of teeth or consciousness, lacerations, internal injury”], and chronicity (i.e., tally of events per outcome type). As only the severity levels 3 and 4 for the verbal aggression outcome subtype on the SOS use language consistent with the operational definition of violence used in this study (i.e., “Impulsively threatens violence toward others”, “makes clear threats of violence toward others repeatedly or deliberately”), only these two levels were included in data analyses.
Chapter 3. Data Analytic Approach

3.1. Data Preparation

3.1.1. Variable Treatment

In order to maximize reliability and validity of outcome data, dichotomous variables coding for incidence and severity, and continuous variables for chronicity were created by collapsing across data sources (i.e., patient interview, staff interview, file review; see Steadman et al., 1998). Chronicity data were only obtained from healthcare files. Cases with missing outcome data were excluded from analyses. To account for cases of missing data for the HCR-20 item variables, scales were prorated if no more than one score was missing from that scale; the same was done for the FAM. In total, three HCR-20 assessments and no FAM assessments were missing one value per scale, and one HCR-20 and no FAMs were missing ≥ 2.

3.1.2. Diagnostics and Assumption Checks

In preparation for the description of obtained data and the use of correlational, Receiver Operating Characteristics (ROC), and binary logistic regression analyses to investigate associations between risk assessment metrics and behavioural outcomes, a number of statistical assumptions were acknowledged and/or diagnostic tests performed. First, outlying values from given variables were investigated by inspecting box plots and q-q plots, and for no data were transformations needed. Descriptive statistics for all data analyzed in this study revealed no distributions for which skewness or kurtosis fell outside the acceptable -1.0 to 1.0 interval; therefore, no transformations were required. Descriptive statistics were then checked for univariate normality by inspecting histograms of included variables’ distributions; normality, linearity, and homogeneity of variance for the independent variables. For all correlational analyses
performed, the statistical assumptions of independent populations/groups, bivariate normal data, and independence of cases/observations from each population/group were tested. Data to be subjected to logistic regression analyses were examined for outliers and case influence using Cook’s D and DFBETAS employing cutoffs of ±1.0 (Cohen et al., 2003). Case influence on regression coefficient estimates was assessed using DFFITS and Cook’s D, and case influence on single regression coefficients was examined with DFBETAS. Lastly, multicollinearity amongst the predictors was examined with the variance inflation factor (VIF; acceptable values < 10; Cohen et al., 2003), employing a cutoff score of 10. All diagnostics and analyses were performed using IBM SPSS Statistics, version 21.0 for Mac OS X.

3.2. Statistical Analyses

Descriptive statistics were calculated to examine basic demographic information, prevalence of the SOS-based adverse outcomes during the two follow-up time periods, and metrics of the HCR-20 and the FAM (e.g., mean item, scale, and measure scores; number of low, moderate, high risk ratings, etc.).

Providing for a chance-corrected measure of agreement, interrater reliability of the HCR-20 and FAM scale scores, total scores, and risk estimates was evaluated using intraclass correlations (ICC) calculated using mixed effects for absolute agreement (ICC$_2$ for HCR-20 scores/estimates and ICC$_3$ for the FAM as there were only two raters). As per Cicchetti et al. (2006) values ≥ .75 were taken to represent excellent agreement.

Predictive accuracies of the measures were evaluated using ROC analyses to calculate area under the curve (AUC) values for a number of measure metrics including HCR-20/FAM total scores, scale scores, composite scale scores, item scores, and summary risk ratings for violent behaviours (operationalized by type, severity, and chronicity); and AUCs were calculated for the FAM’s supplemental risk estimates’ (self-destructive behaviour, victimization) predictive accuracies for their associated outcomes. Authors publishing in the risk assessment field have argued that AUC values generated through ROC analysis are appropriate estimates of predictive validity because they are relatively
insensitive to lower base rates (as compared to correlations, for example), and allow for the statistical comparison between thresholds of different predictors (Douglas et al., 1999; Rice & Harris, 1995) using z-tests for AUC values. Area under the curve values of .65 to .70 are considered to be moderate, and above .70 are considered large (Douglas et al., 2005; Rice & Harris, 2005). To accomplish the second objective of the present study (i.e., compare the predictive accuracies of these measures for a variety of violent outcomes), ROC curves calculated for associated metrics of each measure underwent pairwise statistical comparison using the difference between the areas under the curves, 95% confidence interval for the difference, standard error, and p-value. If p was less than 0.05, the two compared AUC values were determined to be significantly different. These calculations were performed using an online calculator available at http://vassarstats.net/roc_comp.html.

To test the incremental validity of the FAM over the HCR-20 in women, first, point biserial correlations were calculated to determine the magnitude of any relationship between HCR-20/FAM total scores and summary risk ratings and violent outcomes; and the Fisher transformation was applied to the sample correlation coefficients r to standardize variable metrics. A Bonferroni type adjustment was conducted to control for the probability of making a Type I error and hold the familywise error rate at ≤ 0.1 across the multiple correlational tests. Given the number of dependent variables tested \( p \), the per-test error rate \( \frac{\alpha}{p} \) was .005.

It was planned that any significant variables would be directly entered in a series of hierarchical logistic regression analyses examining whether FAM total scores and summary risk ratings add to the capacity of the HCR-20 total scores to predict violent outcomes. Each of the variables that would be included in the model would undergo z-score transformation to center them about the mean prior to their analysis. HCR-20 total scores would be added in step 2 of each model predicting the various outcomes of interest, and FAM total scores would be entered the third step to detect any \( R^2 \) change indicative of incremental validity. Adjusted odds ratios would have also examined in order to evaluate any unique predictive power of each independent variable while controlling for the other predictors in the model. However, as none of the correlations
were found to be significant, logistic regressions were not performed (see Results section for details).

3.3. Power Analysis

For all inferential procedures, consistent with the statistical conventions in the field of violence risk assessment and adverse outcome prevention research, Type I error was controlled by setting experimentwise $\alpha$ at .05, and was held constant across hypotheses except where the Bonferroni correction was applied. Given a predetermined sample size (i.e., $N = 103$), power analyses for point biserial correlations (as calculated using G*Power 3.1 for Mac; Faul, Erdfelder, Lang, & Buchner, 2007) for a moderate effect size of $r = .3$ (Cohen, Cohen, West, & Aiken, 2003) yielded an achieved power level of .88. With respect to logistic regression analyses, G*Power was also used to calculate what the achieved power would have been (if the regressions had been performed) given a set sample size of $N = 103$ and a medium effect size (i.e., odds ratio $\geq 2.5$; Rosenthal, 1996); achieved power was .94.
Chapter 4. Results

4.1. Sample Description

A detailed description of sample demographics is presented in Table 1. Participants included 103 civil psychiatric patients, most of whom (n = 98; 95.1%) were involuntarily committed to a large tertiary psychiatric hospital in British Columbia, and who were scheduled for transfer to smaller regionalized facilities following that hospital’s closure in 2013. Sixty-two participants (60.2%) were male and 41 (39.8%) female. Ages ranged from 20 to 73 years (M = 47.1, SD = 12.6). Most patients were Caucasian (79.0%), and diagnosed with a schizophrenia spectrum disorder (85.4%). Thirty four (33%) were diagnosed with a personality disorder or traits of a disorder, and 37.9% had a diagnosis of a co-morbid substance use disorder. The only demographic variables examined that differed as a function of gender were history of childhood sexual abuse (more common among women; χ² = 11.03, p < .001), history of being married (χ² = 8.92, p < .01), and education (χ² = 16.64, p < .05).


4.2.1. Prevalence rates of violence in overall sample

All violent outcome data across follow up periods are presented in Table 2. More than two-thirds of participants (68.9, n = 71) engaged in an aggressive behaviour within either of the follow up periods. In the present study, because verbal aggression was operationalized to more accurately reflect the definition of violence provided in the HCR-20 and FAM manuals (i.e., must have involved explicit threats), physical aggression against others emerged as more prevalent than threatening (56.3% versus 35.0%) over one year. After aggression against others, aggression towards property was seen at a rate of 49.5% (n = 51) and sexual aggression at 32.0% (n = 33). Paired t-tests showed
no significant differences in prevalence between time points for any outcomes. Not surprisingly, there was a trend such that rates were greater for lower severity than higher severity behaviours, and while mean frequency values were generally low (e.g., 3.38 incidents of aggression toward others over a one year period), the ranges and standard deviations of incident counts were large, indicating that there tended to be a relatively smaller group of individuals who were responsible for higher frequency incidence.

4.2.2. Gender differences in violence by type

As is indicated in Table 2, there were no significant gender differences in prevalence rates of any outcome within any time period, with the exception of verbal aggression, for which men demonstrated higher rates during the first six months ($\chi^2 = 4.03, p = .045$).

4.2.3. Gender differences in rates of violence by severity

Chi-square analyses indicated that for no time period or outcome type were there any gender differences as a function of violence severity.

4.2.4. Gender differences in frequency of violent behaviours

For most types of violent behaviour for any given follow up period, $t$-tests revealed that there were no significant gender differences with regard to frequency of violence. Exceptions within the first six months include aggression toward property, which was higher for women ($M = 1.81, SD = 5.53$) as compared to men ($M = .67, SD = 1.65; t = -1.52, p < .01$), and sexual aggression, which was higher for men ($M = 1.44, SD = .53$) as compared to women ($M = .31, SD = .92; t = 1.37, p < .05$). Within the second follow up period, again, aggression toward property was higher for women ($M = 2.00, SD = 6.67$) as compared to men ($M = .46, SD = .92; t = -1.78, p < .01$). Across time periods, women displayed higher incidence of aggression toward property ($t = -1.71, p < .01$).
4.3. Measure Descriptives and Correlations

Descriptive statistics including individual item scores for HCR-20 and FAM scores in women only (n = 41) are presented in Table 3. Total scores on the HCR-20 ranged from 9 to 35 (M = 23.57, SD = 5.43) out of a possible range of 0 to 40, and 15 to 52 (M = 30.33, SD = 9.83) out of a possible range of 0 to 60 on the FAM. Final risk judgments on the HCR-20 classified 26.8% of participants (n = 11) as low, 43.9% (n = 18) as moderate, and 29.3% (n = 12) as high risk, whereas these same three categories on the FAM were 29.3% low, 46.3% moderate, and 24.4% high risk, respectively. There was a significant difference between the classification distributions of the two measures (χ² = 12.13, p = .02), owing to a tendency for participants to be rated as being somewhat higher risk on the HCR-20 as compared to the FAM. However, cell sizes for these data were small. Five-point risk judgments on the FAM classified 5.8% (n = 6) as low, 13.6% (n = 14) as low-moderate, 3.9% (n = 4) as moderate, 14.6% (n = 15) as moderate-high, and 1.9% (n = 2) as high risk. Using the PCL-R cutoff used with the HCR-20 (i.e., >30), no participants met diagnostic thresholds for psychopathy, whereas using the FAM cutoff (i.e., >23), one participant met the threshold for classification.

With regard to item scores, paired samples t-tests showed that there were no items common to both measures for which scores differed significantly from one another. This was true even for those items rated on the FAM using additional criteria (e.g., major mental illness, personality disorder, prior supervision failure).

Between the H scales, which themselves were associated (r = .58, p < .001), each common item pair was strongly associated, with rs ranging from .45 (Item 2 – Young Age at First Violence) to 1.00 (item 6 – Major Mental Illness). All item pairs were significant at a minimum of p < .01. Item 8 on the HCR-20 (Early Maladjustment) is divided into two items on the FAM (Problematic Circumstances in Childhood, Problematic Behaviour in Childhood), and these latter two were associated with the former at r = .50, p < .001 and r = .48, p = .002, respectively. Unexpectedly, the C scale total scores were not significantly associated (r = .20, p = .22), and only the item 1 (Lack of Insight) and 2 (Negative Attitudes) pairs were significantly correlated (r = .34, p < 0.05 and r = .37, p < .05, respectively). The R scales were significantly associated (r = .41, p < .01), but only the item 3 (Lack of Social Support) and item 4 (Noncompliance with
Remediation Attempts) pairs ($r = .43, p < .01$ and $r = .33, p < .05$). HCR-20 and FAM total scores were strongly correlated ($r = .74, p < .001$).

Historical-Clinical-Risk Management-20 total scores were significantly associated with final risk judgments such that higher scores correlated with higher estimations of risk ($r = .54, p < .001$). The same was true for FAM total scores and both 3-point ($r = .62, p < .001$) and 5-point ($r = .70, p < .001$) summary risk judgments.

### 4.4. Interrater Reliability

Interrater reliability was calculated in two ways. First, with regard to the reliability of scores between the two raters of the FAM (for $n = 10$), consistent with the hypotheses, average intraclass correlations (ICC; using two-way mixed methods) were shown to be excellent for the FAM H- (ICC = .90, $p = .007$), C- (ICC = .83, $p = .025$, and R-scale (ICC = .86, $p = .017$) scores, total scores (ICC = .86, $p = .017$), and final summary risk judgments for violence (3-point; ICC = .85, $p = .019$; and 5-point; ICC = .92, $p = .003$). Interestingly, reliability for 5-point risk judgments for self-destructive behaviour and victimization was poor (ICC = .615, $p = .135$ and ICC = .66, $p = .110$, respectively).

Next, interrater reliability was assessed for the HCR-20 in both a subset of the overall sample (i.e., men and women; $n = 15$) and in the women only ($n = 9$). For the overall group, ICCs were shown to be .84 for the H scale, .85 for the C scale, .89 for the R scale, and .87 for total scores. For the women, ICCs were .83, .84, .82, and .84, respectively.

### 4.5. Predictive Validity

#### 4.5.1. Any violence

To examine whether HCR-20 and FAM scale scores, total scores, and final risk judgments predicted violence within and across the three time periods (i.e., 0-6 months, 6-12, 0-12 months), ROC analyses were conducted (AUC values presented in Table 4). Unexpectedly, of all of the HCR-20 test variables examined, only the Clinical scale (for
men in the 6-12 month follow-up period) achieved even a moderate level of predictive accuracy for any violence (AUC = .66). All other HCR-20 test variables fell below the Rice & Harris (2005) criteria for satisfactory prediction, indicating that in this sample the HCR-20 did little to predict violence any better than chance. In contrast, the FAM total score, Clinical and Historical scale scores, three- and five-point final risk judgments showed moderately good (AUC = .70, .69, .68, .68, .66, respectively) prediction of any violence in the first (0-6 month) time period. For the 6-12 month period, the FAM Clinical scale showed good prediction (AUC = .73) and the total score and five-point judgment showed moderate predictive accuracy (both AUC = .67). Considering prediction of violence at any time in the follow up period (i.e., 0-12 months), the five-point final risk judgment produced a large AUC of .87 (p < .001), and Clinical and Historical scales, total score, and three-point final judgment produced moderate but non-significant AUCs (.70, .68, .67, .67, respectively). Despite seeming differences between AUCs within predictor per time period, no differences reached significance. Those differences that approached significance included the Historical scales on the FAM (women; AUC = .69) and the HCR-20 in men (AUC = .52; z = -1.45, p = .15); between the Clinical scales on the FAM (AUC = .68) and the HCR-20 in women (AUC = .47; z = -1.68, p = .09); and between the FAM and HCR-20 total scores (AUC = .70 and .52, respectively) in women (z = -1.39, p = .17). Bivariate correlational analyses revealed that none of the HCR-20 or FAM test variables were associated with incidence (i.e., frequency) of any type of violent outcome (r range -.27 – .20, all ps ns). Thus, no analyses investigating predictive validity were conducted using incidence data.

4.5.2. Violence by type

The AUCs for the HCR-20 and the FAM’s prediction of the different types of violence considered in this study (across the full 0-12 month time period) are presented in Table 5. Again, the HCR-20 generally demonstrated poor prediction of all outcomes (i.e., threatening, against others, against property, sexual aggression) for men and women, with the exception of threatening in women, which resulted in moderately high AUCs (.70 for the C scale and .69 for the R scale). However, these AUCs only approached significance. The Clinical scale on the HCR-20 significantly predicted violence toward property in men (AUC = .65, p < .05). While the FAM did not offer
prediction greater than chance for violence towards others or property (AUCs range from .43-.63), the measure did perform somewhat better for threatening and sexual violence. The FAM’s Clinical scale significantly predicted sexual violence (AUC = .71, p < .05), but did not differ significantly from the HCR-20’s prediction of the same criterion (AUC = .52, p = ns). Similarly, the FAM’s three-point final risk judgment, total score, and five-point judgment showed moderate AUCs for prediction of sexual violence (.68, .66, .66, respectively), but only approached the level of significance. With regard to threatening, the FAM’s five- and three-point final risk judgments, Clinical scale, total score, and Historical scale produced large AUCs of .87 (p < .001), .85 (p < .001), .76 (p < .05), .73 (p < .05), and .72 (p < .05), respectively. Further, the FAM risk judgment’s prediction of threatening produced the only AUC that was found to be significantly different from its corresponding HCR-20 variable in women (z = -2.87, p = .004).

### 4.5.3. Violence by severity

Results from ROC analyses of the HCR-20 and the FAM’s prediction of violence severity by gender are presented in Table 6. Again the HCR-20 variables produced AUCs indicating prediction of both low and high severity violence over one year was no better than chance. The FAM Clinical scale was moderately predictive of low severity violence, although the other variables demonstrated poor predictive accuracy (AUCs ranging from .49-.64). However, all FAM variables except for the R scale produced large and statistically significant AUCs for the prediction of high severity violence [.78 (3-point judgment), .77 (five-point judgment), .76 (H-scale), .76 (total score), and .72 (C-scale)]. Examining the FAM’s predictive accuracy of high severity violence by type revealed large AUCs of .87, .85, .76, .73, and .72 (all significant at p < .05) for five- and three-point risk judgments, C- and H-scales, and total score prediction of threatening, respectively. The FAM total score, as well as H-, C-, and R-scales also produced large AUCs for the prediction of severe sexual violence (.90, .87, .80, .79, respectively). The measure’s three-point final judgment and C-scale also showed good prediction of high severity violence against property at AUC = .71 and .70; and the five-point judgment produced a moderate AUC for violence against others (.66). In comparison, the only the HCR-20’s total score (in women) showed good prediction of severe sexual violence.
(AUC = .80), and moderate-good prediction of the criterion by the final risk judgment (AUC = .70).

4.5.4. Self-destructiveness and victimization

The FAM’s five-point risk judgment for victimization demonstrated relatively poor prediction (AUC = .61). With regard to self-destructiveness (which included self-harm, suicidal behaviours, self-neglect, and substance use), unfortunately, all but one participant had engaged in a self-destructive behaviour (i.e., base rate of ~100%), which precluded analyses of predictive accuracy.

4.6. Incremental Validity

Given that none of the HCR-20 or FAM variables (i.e., item, scale, total scores; risk ratings) were significantly associated with aggression in the bivariate analyses, the planned hierarchical logistic regression analyses (to determine incremental validity) had to be abandoned.
Chapter 5. Discussion

The current study was the first independent investigation into the psychometric properties of the FAM, with the aim of evaluating the usefulness of a novel, female-targeted risk assessment for use in diverse settings. In doing so, this study also explored gender differences in type and severity of violent behaviour in a rather unique civil psychiatric sample recently transferred out of a tertiary psychiatric hospital and into community settings following that hospital’s closure. Some have argued that as tertiary psychiatric hospitals close, the less severely mentally ill and easier to place patients are transferred first, leaving a sort of “remnant” group of patients with particularly intractable symptoms and behavioural problems (Ford, 1987; Petersen et al., 2013). The particularly high prevalence rate of violence (and other adverse outcomes; Greig et al., 2013) seen in this study suggests that the current sample may indeed pose greater clinical challenges to care and management than other severely mentally ill (SMI) samples, and thus it is important to recognize that the results from this study might not generalize entirely to other psychiatric populations. Despite this, the current study capitalized on the narrowed gender gap in prevalence and typology of violent behaviour previously seen in diverse psychiatric samples (e.g., McNeil & Binder, 1995; Monahan et al., 2001; Nicholls et al., 2004) in order to tease out any gender differences in terms of how violence risk should best be assessed. Specifically, this study’s aims were first to document gender differences in rates, type, and severity of violence; then to compare the novel FAM measure to the gold standard HCR-20 in terms of any descriptive differences, as well as any differences in interrater reliability and predictive validity, and finally to determine the incremental validity of the FAM over the HCR-20. The findings with respect to these aims will be discussed in this order.

As alluded to above, with regard to overall prevalence of violent behaviours, results indicated that rates of inpatient aggression were approximately double that which has been found in similar but earlier studies in the same hospital (Nicholls et al., 2004). There was a general trend of decreased prevalence of violent behaviour coinciding with patient transfer from hospital to community, suggesting an equivalent or superior level of patient management and care between the two settings that was as effective for women as for men. While the finding that physical violence against others was more prevalent than verbal aggression is inconsistent with past research, this is likely attributable to the operationalization of verbal aggression in this study – only those incidents in which explicit threats of harm to others were made were considered. This was done in order to better align the definition of violence used in this study with that specified in the HCR-20 manual. Even still, the higher rate of physical violence versus threatening is a notable general finding, and hints that violence in these settings may more often occur spontaneously and without warning (i.e., without an antecedent threat). Alternatively, given that problematic behaviour was relatively common within this sample, it may be that when a patient made a verbal threat but was not physically violent, hospital staff recorded the threat in nursing notes, and thus it was coded by study evaluators; whereas, if a patient made threats before, during, or after a physically violent behaviour, staff recorded only the most serious aspect of the incident (i.e., the physical violence), failing to mention the verbal threats. If this were the case, staff records of incidents involving co-occurring threatening and violence would lead to an underestimate of the rate of patient threats. Indeed past research has shown that clinical notes may underreport violence incidents in hospital by as much as 45% (Ehmann, Smith, Yamamoto, et al., 2001). Future research could more closely examine gender differences in the specific characteristics of violent incidents (e.g., evaluate differences in the co-occurrence of behaviours such as threatening and physical violence) and staff documentation of inpatient aggression among men and women.

34
5.2. Gender Differences in Violent Behaviour

Lending further support to others’ findings of a narrowed gender gap in psychiatric patients’ violence, results from the present study demonstrate that, by and large, there were no gender differences in rate, timing, type, severity, or frequency of violent behaviours in this sample of persistently mentally ill tertiary psychiatric patients. However, there were a few exceptions. Not surprisingly, and consistent with previous research conducted in diverse samples, men were more likely than women to engage in sexual violence, regardless of severity. Within the first six months of transfer, women were somewhat more likely than men to engage in threatening. The reasons for this difference are not readily clear, but it may be that something about women’s adjustment to a new setting disproportionately increased their tendency to enact their aggression across the continuum of behaviours (i.e., verbal and physical); but again, this is difficult to ascertain without examining the co-occurrence of such behaviours. With regard to incidence of violence, results indicated that of those patients who aggressed toward property (e.g., slammed doors, kicked walls, destroyed furniture), women did so more frequently regardless of the period of assessment. Taken together these findings may suggest that women generally tended to express their aggression using a somewhat more diverse set of behaviours than men, who tended to direct their aggression physically and toward another person. While the present study yielded few novel findings with regard to gender differences in prevalence of violence, the task of evaluating the FAM’s performance relative to the HCR-20 is arguably made simpler, given a level playing field.

5.3. Descriptive Differences between HCR-20 and FAM and Associations with Violence

The HCR-20 and the FAM differed little descriptively, with no significant differences between item scores common to both measures and no important differences between the proportions of patients classified as low, moderate, and high risk. Counter to hypotheses, and despite strong interrater reliability both between and within measures, no correlations between risk assessment variables and outcomes were found to be significant in the present study. It is interesting that neither measure
performed as well as the HCR-20 has in a multitude of previous studies (Douglas & Reeves, 2010), this finding unfortunately precluded further regression analyses, thus preventing examination of the FAM’s incremental validity over the HCR-20. Future research would benefit from larger sample sizes in order to maximize statistical power to test models of incremental validity.

5.4. Comparing the Predictive Validities of the HCR-20 and the FAM

5.4.1. HCR-20

It is interesting that the HCR-20 showed performance somewhat below to that seen in other psychiatric samples. Regardless of follow up period, the predictive validity of the HCR-20 variables was underwhelming; only the Clinical items and total score approached moderate prediction of general violence, and then it did so for men, only. When looking at prediction of different types of violence, the HCR-20’s dynamic Clinical and Risk Management items showed better prediction of verbal threatening, but did not perform above chance for physical violence. This will be discussed shortly. With regard to the HCR-20’s overall mediocre performance, given adequate interrater reliability, there are a few possible explanations that one could consider.

First, it is possible that an unknown issue or issues related to study design influenced the validity of the risk assessments; for example, it is possible (though seemingly unlikely) that the quality of the evaluators’ assessments, as a whole, was suboptimal. While great care was taken to provide evaluators with high quality, expert training in risk assessment using the HCR-20, for the majority of evaluators, this study was their first exposure to the measure (and indeed the risk assessment activity). Nevertheless, many previous studies on the HCR-20 have achieved greater predictive validity using fewer, equivalently trained evaluators to those used in the present study. Another possibility (related to sample characteristics) is that this type of “remnant” group of long-stay inpatients with chronic, treatment-refractory mental illness is not impacted in the same way (or to the same extent) by the risk-relevant characteristics identified as important in other populations more commonly studied using the HCR-20 (e.g.,
offenders, forensic patients, shorter stay civil psychiatric patients). Notably, much of the literature on the HCR-20’s use in psychiatric populations is characterized by community follow up designs. Those studies that have examined assessment of risk for inpatient violence have yielded some interesting findings. For example, Nicholls et al. (2004) used the HCR-20 and found somewhat poorer predictive validity for inpatient violence as compared to community violence. The reasons for such findings are not obvious; but it has been suggested that factors related to the hospital environment, such as increases in structured activities and demands on patients (Haller & Deluty, 1988; Ionno, 1983) among many others moderate the expression of violent behaviour in inpatients. As Rabinowitz and Mark (1999) noted, “there appears to be a lack of systematic information about which variables are associated with violence among long-stay psychiatric patients” (p.342). Perhaps then, in the present sample, it may be that while HCR-20 items capture important markers of risk for violence shown to generalize well across populations, other yet unidentified factors may be of greater causal relevance to this group’s risk, and thus the assessment of that risk (e.g., an especially high prevalence of active symptoms of mental illness).

5.4.2. FAM

Despite the suggestion that the characteristics of the present sample may have somehow moderated the accuracy of the HCR-20, counter to expectations, the FAM variables (particularly the C scale and total score) did tend to show greater (but still moderate) predictive validity for general violence over all time periods; although, none were significantly different from their associated HCR-20 indices. Also counter to hypothesis, the additional risk judgment domains of self-destructiveness and victimization failed to predict those behaviours any better than chance within any time period. This is not entirely surprising, given the HCR-20 (upon which the FAM wholly relies) is intended to assess risk of violence directed toward others. Arguably, the most confounding finding in this study is that although the FAM showed excellent prediction of high severity general violence by all variables except for the Risk Management scale, closer examination revealed that this effect was seen only for prediction of threatening. In fact, with the exception of moderate predictive validity of the risk judgments and C scale for sexual violence, the FAM’s accuracy with regard to threatening behavior in
women likely accounts for the slight superiority that the FAM variables demonstrate over HCR-20 scores for the prediction of general violence.

Again, there are a variety of possible explanations for the finding of the FAM’s modestly better prediction of general violence. First, with regard to study design, it is worth noting that the evaluators who coded the FAM were senior graduate students who had previous experience conducting risk assessments using the HCR-20 for both clinical and research purposes. This contrasts somewhat to the pool of evaluators who worked to score the HCR-20, which was comprised of research assistants at various levels of academic training (although all had completed an undergraduate degree), and having had varied prior exposure to risk assessment. Therefore, again, it is possible that the quality of the assessments was somewhat higher for the FAM. Alternatively, whereas the HCR-20 was scored using information from file and patient interview, the FAM was coded solely from file. It thus remains unclear as to how design factors may have influenced the results.

In a review of the literature on the prediction of inpatient violence, Steinart (2002) notes that the nature of (and immediate consequences of) inpatient violence is so different from community violence that the predictive accuracy of risk factors captured by common measures can falter, as aspects of clinical variables such as psychopathology become more important (e.g., thought disorder, ideas of persecution, hostility). In addition to patient characteristics, the author cites findings that characteristics of the hospital environment (e.g., crowding) as well as the behaviour of staff (behavioural intervention, patient restraint and seclusion) affect the predictive accuracy of risk measures; although, there seems to be little consensus on exactly how these factors do so. It could be argued that the combined characteristics of the present “remnant” group of severely ill, highly medicated, highly managed individuals housed in a structured environment in which behaviour is closely monitored simply attenuated the accuracy of the HCR-20, and to a lesser degree the FAM (“a lesser degree” as a result of rater differences). Further, given the unique situation of members of the present sample, it is also possible that the stress experienced as a result of the hospital’s closure and patients’ transfer during the period of assessed risk affected the validity of the measures. Of course, all of this may help to explain findings for physical violence (including that
toward others, property, and sexual violence), but does little to explain the HCR-20’s moderate validity and the FAM’s excellent validity for assessing threatening behaviours.

A few studies have found similar results with regard to threatening behaviour in similar samples. In a study of 216 forensic inpatients in Germany, Tengstrom, Hodgins, Müller-Isberner, et al. (2006) found that the HCR-20 performed notably better for the prediction of threats than it did for violence across groups of schizophrenic, personality disordered, and cognitively impaired individuals. The authors, however, do not offer any thoughts as to why this effect might have been seen; and it is important to note that this was a forensic psychiatric sample. Grevatt (2004), in a sample of civil psychiatric inpatients, found the Clinical scale of the HCR-20 demonstrated excellent prediction of verbal abuse (the Historical scale performed poorly), whereas no HCR-20 variable achieved even moderate prediction of physical violence for individuals causing fewer than three violent incidents.

While results of the present study may echo these findings, it should also be noted that the results with respect to threatening were based on just a small subset of nine women who engaged in the behaviour, and thus the extent to which this is relevant to those samples is unclear. This could be responsible for inflating the accuracy of both measures for threatening only, which might explain the vast discrepancy in validity relative to the other types of violence for which base rates were higher. After all, predictive validity of the HCR-20 for threatening was poor for men, of whom there were more who actually engaged in the behaviour. Thus, the small size of the subsample of women may explain gender differences seen in the accuracy of the HCR-20 for threatening. Future research would do well to examine the predictive validity of the FAM on a larger sample of women, perhaps from a population more commonly seen in the literature.

In sum, consistent with much previous research, dynamic clinical factors appear to hold the greatest weight when predicting diverse forms of violence in both men and women; and this may particularly be the case in chronic, residential psychiatric settings in which patients’ psychopathology interacts with unit/ward characteristics in such a way so as to complicate the risk assessment activity.
With regard to the ongoing debate about the need for gender-sensitive risk assessment, it appears that the jury is still out. The FAM’s modestly better prediction as compared to the HCR-20 (especially in light of the aforementioned concerns regarding the validity of the assessments used in this study) is far from proof that a novel measure is warranted. The literature on the HCR-20 is vast, and far more often than not, the HCR-20 has demonstrated that it is as effective for violence risk assessment in women as it is for men. While there are indisputable differences between men and women in terms of the rate, type, severity, and targets of the violence that they perpetrate, the research community should be yet unconvinced that such differences call for differential assessment of risk.

The present research contributes to the gender and violence risk literature by helping to clarify the value of population- (i.e., female-) targeted risk assessments (see Singh et al., 2011 for a large-scale meta-analysis on this topic), and is among the first studies to specifically examine the psychometrics of the FAM, at least in psychiatric patient populations. The results provide further evidence that the use of empirically validated measures in both clinical and research settings is beneficial, and reaffirms that psychiatric samples are unique in their scarcity of gender differences in violence. They are thus useful settings for continued work in the gender and violence risk area. Unexpectedly, the present findings highlight the need for better understanding of how (in secure inpatient settings) contextual and environmental characteristics can influence patient behaviour to a degree that might be relevant for accurate violence risk assessment.

5.5. Limitations and Strengths of the Present Study

As with any study, the research was subject to some limitations. While the reliability of most data collected as part of this study is maximized through the consideration of multiple sources of information (i.e., healthcare files, official databases, patient and staff interviews, see Steadman et al., 1998), in order to make use of an admittedly small sample of female participants, as noted, the FAM was coded solely from file. This is significant as the FAM (like the HCR-20 and many other risk assessment tools) was intended to include an interview component as a matter of best
practice. Consequently, the results of the FAM assessments completed in this study may not optimally generalize to risk assessment as it has thus far been conducted in the field. Further, as the FAM was based on file information only, it was coded using less information than was the HCR-20. This risks exposing those assessments to biases due to inconsistencies or errors in data recording, collection, or coding, and complicates interpretations of findings from comparisons with the HCR-20. This limitation could be appropriately addressed in future research by ensuring that the assessments being examined are coded in similar fashions and using equivalent information.

The prospective, longitudinal design was a notable strength of this study, and is among few that have followed up with civil psychiatric patients while still housed in an inpatient treatment setting. Relatively little research in the violence risk assessment field has drawn on tertiary/civil psychiatric samples, despite the fact that similar decision making occurs (i.e., Review Panels vs. Review Boards) and the number of patients in the civil psychiatric system far surpasses the number of forensic psychiatric inpatients. Specifically, in BC there are 190 inpatient beds at the Forensic Psychiatric Hospital and 916 tertiary psychiatric beds allocated across the province. Another notable strength is that, unlike many studies comparing the predictive accuracy of risk assessment measures, the proposed project had the distinct advantage of having separate raters code the HCR-20 and the FAM, reducing bias substantially.

5.6. Clinical Implications

Although the results of this study were largely unexpected, and further research will certainly be needed to evaluate the FAM, the findings hint at important implications for the assessment and management of psychiatric inpatients – particularly those who are severely and persistently ill. First, the performance of the FAM in this sample (especially with regard to threatening behaviour) suggests that it may indeed be a useful tool in civil psychiatric populations, and that consideration of gender-informed clinical variables (such as manipulative behaviour or self-esteem) may be relevant considerations in the management of women’s violence, or even untargeted outward displays of aggression (i.e., toward property). At the very least, it will continue to be important for clinicians to be mindful of the research evidence that women’s risk is often
underestimated (see Coontz et al., 1994), that risk assessment should be tailored to the individual in order to maximize management success, and that attention to gender-informed variables may very well assist in accomplishing this.

More generally, the finding that dynamic, clinically relevant variables on both measures did emerge as at least somewhat predictive of threatening behaviour for both genders (but did less to predict other forms of aggression) suggests two things. The first is that for long-stay, severely mentally ill patients, the influence of variables related to psychopathology (e.g., persistent psychotic symptoms) may outweigh other types of variables (e.g., historical) in the prediction of verbal outbursts. If this is indeed the case, the repeated use of a short structured screening tool focused on present symptomology may be more useful than more comprehensive measures for the assessment and management of verbal displays of aggression (which are destabilizing for other patients, and run the risk of inciting physical aggression). Second, it may be that the rather unique constellation of environmental- and person-level behavioural influences present in samples/setttings such as these may simply make physical violence more difficult to predict. Clinicians working in these types of settings would be wise to remain flexible in their assessment approaches, and be open to considering other factors not included in existing gold standard measures – be they gender-sensitive or otherwise.
References


## Appendices

### Appendix A.

### Table 1

Sample Demographics \((N=103)\)

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Men ((n=62; 60.2%))</th>
<th>Women ((n=41; 39.8%))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>46 (75.4)</td>
<td>33 (78.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>4 (6.6)</td>
<td>5 (11.9)</td>
</tr>
<tr>
<td>Aboriginal</td>
<td>4 (6.6)</td>
<td>3 (7.1)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (12.9)</td>
<td>1 (2.4)</td>
</tr>
<tr>
<td><strong>Diagnosis (primary Axis I)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia-spectrum</td>
<td>54 (88.5)</td>
<td>34 (81.0)</td>
</tr>
<tr>
<td>Bipolar</td>
<td>4 (6.6)</td>
<td>2 (4.8)</td>
</tr>
<tr>
<td>Other psychotic</td>
<td>1 (1.6)</td>
<td>3 (7.1)</td>
</tr>
<tr>
<td>Cognitive/organic</td>
<td>2 (3.3)</td>
<td>3 (7.1)</td>
</tr>
<tr>
<td>Personality disorder/traits diagnosis</td>
<td>21 (34.4)</td>
<td>13 (31.0)</td>
</tr>
<tr>
<td>Substance use disorder</td>
<td>30 (32.8)</td>
<td>19 (45.2)</td>
</tr>
<tr>
<td><strong>Years of education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>35 (57.4)</td>
<td>19 (45.2)</td>
</tr>
<tr>
<td>High school only</td>
<td>15 (24.6)</td>
<td>14 (33.3)</td>
</tr>
<tr>
<td>High school and greater</td>
<td>10 (16.2)</td>
<td>9 (21.9)</td>
</tr>
<tr>
<td>Ever married</td>
<td>11 (18.0)</td>
<td>19 (45.2)</td>
</tr>
<tr>
<td>Employed</td>
<td>2 (3.3)</td>
<td>1 (2.4)</td>
</tr>
<tr>
<td><strong>Abuse in childhood</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual (yes)**</td>
<td>2 (3.3)</td>
<td>11 (26.2)</td>
</tr>
<tr>
<td>Physical (yes)</td>
<td>9 (14.8)</td>
<td>8 (19.0)</td>
</tr>
<tr>
<td>Neglect (yes)</td>
<td>11 (18.0)</td>
<td>10 (23.8)</td>
</tr>
</tbody>
</table>

\(a \text{n=97}^* \quad p < .05 \quad ^*^* p < .01 \quad ^* * p < .001 \) (chi-square).
### Appendix B.

**Table 2**

Prevalence, Incidence, Nature, & Severity of Violence

<table>
<thead>
<tr>
<th>behaviours</th>
<th>Overall</th>
<th>0-6 months</th>
<th>6-12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Violent Frequency</td>
<td>Violent Frequency</td>
<td>Those violent Frequency</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M(SD)</td>
</tr>
<tr>
<td>Threatening</td>
<td>36</td>
<td>35.0</td>
<td>1.20(3.10)</td>
</tr>
<tr>
<td>Physical - others</td>
<td>58</td>
<td>56.3</td>
<td>3.38(8.25)</td>
</tr>
<tr>
<td>Low</td>
<td>57</td>
<td>55.3</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>20</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>Physical - property</td>
<td>51</td>
<td>49.5</td>
<td>2.22(7.89)</td>
</tr>
<tr>
<td>Low</td>
<td>50</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>Sexual</td>
<td>33</td>
<td>32.0</td>
<td>1.98(8.43)</td>
</tr>
<tr>
<td>Low</td>
<td>33</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>Violence - any</td>
<td>71</td>
<td>68.9</td>
<td>9.42(21.50)</td>
</tr>
<tr>
<td>Low</td>
<td>65</td>
<td>63.1</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>51</td>
<td>49.5</td>
<td></td>
</tr>
</tbody>
</table>

Note. No significant gender differences in frequency * Significant gender difference (chi-square), p<.05.
### Appendix C.

### Table 3

Mean (SD) item and scale scores for HCR-20 and FAM

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Historical items</strong></td>
<td><strong>HCR-20</strong></td>
<td><strong>FAM</strong></td>
</tr>
<tr>
<td>H1 Previous violence</td>
<td>1.24(.86)</td>
<td>1.47(.75)</td>
</tr>
<tr>
<td>H2 Young age at first violent incident</td>
<td>.70(.76)</td>
<td>.73(.70)</td>
</tr>
<tr>
<td>H3 Relationship instability</td>
<td>1.55(.68)</td>
<td>1.48(.78)</td>
</tr>
<tr>
<td>H4 Employment problems</td>
<td>1.56(.60)</td>
<td>1.55(.75)</td>
</tr>
<tr>
<td>H5 Substance use problems</td>
<td>1.05(.95)</td>
<td>1.23(.95)</td>
</tr>
<tr>
<td>H6 Major mental illness†</td>
<td>1.95(.39)</td>
<td>2.00(.00)</td>
</tr>
<tr>
<td>H7 Psychopathy†</td>
<td>.24(.49)</td>
<td>.51(.68)</td>
</tr>
<tr>
<td>H8 Early maladjustment†</td>
<td>.88(.85)</td>
<td></td>
</tr>
<tr>
<td>H8a Problematic circumstances during childhood</td>
<td></td>
<td>.90(.90)</td>
</tr>
<tr>
<td>H8b Problematic behavior during childhood</td>
<td></td>
<td>.55(.82)</td>
</tr>
<tr>
<td>H9 Personality disorder†</td>
<td>.51(.68)</td>
<td>.70(.91)</td>
</tr>
<tr>
<td>H10 Prior supervision failure†</td>
<td>1.41(.74)</td>
<td>1.45(.64)</td>
</tr>
<tr>
<td>H11 Prostitution</td>
<td></td>
<td>.57(.87)</td>
</tr>
<tr>
<td>H12 Parenting difficulties</td>
<td></td>
<td>.68(.94)</td>
</tr>
<tr>
<td>H13 Pregnancy at young age</td>
<td>.32(.73)</td>
<td></td>
</tr>
<tr>
<td>H14 Suicidality / self-harm</td>
<td></td>
<td>1.10(.90)</td>
</tr>
<tr>
<td>H15 Victimization after childhood</td>
<td></td>
<td>1.00(.88)</td>
</tr>
<tr>
<td><strong>H-scale total</strong></td>
<td>10.65(4.35)</td>
<td>16.2(6.50)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Clinical items</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Lack of insight</td>
<td>1.66(.58)</td>
<td>1.55(.75)</td>
</tr>
<tr>
<td>C2 Negative attitudes</td>
<td>.63(.62)</td>
<td>.60(.78)</td>
</tr>
<tr>
<td>C3 Active symptoms of major mental illness</td>
<td>1.71(.51)</td>
<td>1.71(.60)</td>
</tr>
<tr>
<td>C4 Impulsivity</td>
<td>1.39(.74)</td>
<td>1.50(.72)</td>
</tr>
<tr>
<td>C5 Unresponsive to treatment</td>
<td>1.10(.70)</td>
<td>1.20(.79)</td>
</tr>
<tr>
<td>C6 Covert / manipulative behavior</td>
<td></td>
<td>.75(.74)</td>
</tr>
<tr>
<td>C7 Low self-esteem</td>
<td></td>
<td>.40(.63)</td>
</tr>
<tr>
<td><strong>C-scale total</strong></td>
<td>6.49(1.63)</td>
<td>7.7(2.63)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Risk management items</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 Plans lack feasibility</td>
<td>.98(.72)</td>
<td>1.38(.71)</td>
</tr>
<tr>
<td>R2 Exposure to destabilizers</td>
<td>.98(.57)</td>
<td>1.25(.74)</td>
</tr>
<tr>
<td>R3 Lack of personal support</td>
<td>1.07(.76)</td>
<td>.77(.83)</td>
</tr>
<tr>
<td>R4</td>
<td>Noncompliance with remediation attempts</td>
<td>.98(.69)</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>R5</td>
<td>Stress</td>
<td>1.49(.60)</td>
</tr>
<tr>
<td>R6</td>
<td>Problematic child care responsibility</td>
<td>.03(.16)</td>
</tr>
<tr>
<td>R7</td>
<td>Problematic intimate relationship</td>
<td>.43(.71)</td>
</tr>
<tr>
<td>R-scale total</td>
<td></td>
<td>5.49(1.79)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>22.77(5.29)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final judgment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>11(26.8)</td>
<td>12(29.3)</td>
</tr>
<tr>
<td>Moderate</td>
<td>18(43.9)</td>
<td>19(46.3)</td>
</tr>
<tr>
<td>High</td>
<td>12(29.3)</td>
<td>10(24.4)</td>
</tr>
</tbody>
</table>

Note. $n = 41$ as FAM only scored in women † Considers guidelines additional to those of HCR-20 item.
### Appendix D.

**Table 4**

Area Under the Curve [95% CI] for HCR-20 and FAM Scale and Total Scores, and Final Risk Judgments Predicting Any Violence by Gender

<table>
<thead>
<tr>
<th></th>
<th>0-12 months</th>
<th></th>
<th>0-6 months</th>
<th></th>
<th>6-12 months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td><strong>HCR-20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical</td>
<td>.58 [.42, .74]</td>
<td>.52 [.37, .68]</td>
<td>.50 [.35, .65]</td>
<td>.57 [.40, .75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>.64 [.46, .82]</td>
<td>.61 [.45, .76]</td>
<td>.66 [.51, .81]</td>
<td>.63 [.46, .80]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Management</td>
<td>.60 [.43, .76]</td>
<td>.60 [.45, .75]</td>
<td>.61 [.46, .76]</td>
<td>.48 [.30, .66]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>.63 [.46, .80]</td>
<td>.58 [.42, .74]</td>
<td>.57 [.43, .72]</td>
<td>.59 [.42, .77]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment</td>
<td>.59 [.42, .76]</td>
<td>.54 [.38, .70]</td>
<td>.57 [.39, .74]</td>
<td>.60 [.46, .74]</td>
<td>.47 [.29, .64]</td>
<td></td>
</tr>
<tr>
<td><strong>FAM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical</td>
<td>.67 [.49, .84]</td>
<td>.69 [.53, .85]</td>
<td>.62 [.49, .80]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>.70 [.53, .86]</td>
<td>.68 [.52, .85]</td>
<td>.73 [.58, .89]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Management</td>
<td>.54 [.36, .72]</td>
<td>.58 [.40, .76]</td>
<td>.56 [.38, .74]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>.68 [.51, .86]</td>
<td>.70 [.53, .86]</td>
<td>.67 [.50, .84]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment (3-pt)</td>
<td>.59 [.41, .77]</td>
<td>.68 [.51, .84]</td>
<td>.61 [.44, .78]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment (5-pt)</td>
<td>.67 [.50, .84]</td>
<td>.66 [.50, .83]</td>
<td>.67 [.50, .83]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Men \((n = 61)\) Women \((n = 41)\); No significant differences between any AUCs within predictor per time period.
Appendix E.

Table 5

Area Under the Curve [95% CI] for HCR-20 and FAM Test Variables Predicting Type of Violence by Gender (Over 12 Months)

<table>
<thead>
<tr>
<th></th>
<th>Threatening</th>
<th>Physical - others</th>
<th>Physical - property</th>
<th>Sexual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HCR-20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical</td>
<td>.47 [.32, .62]</td>
<td>.59 [.37, .82]</td>
<td>.47 [.33, .62]</td>
<td>.53 [.35, .71]</td>
</tr>
<tr>
<td>Clinical</td>
<td>.61 [.47, .75]</td>
<td>.70 [.48, .92]</td>
<td>.64 [.49, .79]</td>
<td>.50 [.33, .68]</td>
</tr>
<tr>
<td>Risk</td>
<td>.55 [.40, .70]</td>
<td>.69 [.47, .91]</td>
<td>.57 [.42, .72]</td>
<td>.44 [.26, .61]</td>
</tr>
<tr>
<td>Total score</td>
<td>.51 [.36, .66]</td>
<td>.64 [.40, .89]</td>
<td>.54 [.39, .69]</td>
<td>.52 [.34, .70]</td>
</tr>
<tr>
<td>Judgment</td>
<td>.53 [.39, .68]</td>
<td>.60 [.38, .82]</td>
<td>.60 [.46, .75]</td>
<td>.50 [.32, .68]</td>
</tr>
<tr>
<td><strong>FAM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical</td>
<td>.72’ [.55, .88]</td>
<td>.54 [.35, .72]</td>
<td>.53 [.35, .71]</td>
<td>.64 [.46, .82]</td>
</tr>
<tr>
<td>Clinical</td>
<td>.76’ [.57, .95]</td>
<td>.57 [.39, .75]</td>
<td>.58 [.41, .76]</td>
<td>.71’ [.54, .86]</td>
</tr>
<tr>
<td>Risk</td>
<td>.61 [.39, .83]</td>
<td>.52 [.34, .70]</td>
<td>.43 [.24, .61]</td>
<td>.53 [.33, .74]</td>
</tr>
<tr>
<td>Total score</td>
<td>.73’ [.56, .90]</td>
<td>.55 [.37, .73]</td>
<td>.53 [.35, .72]</td>
<td>.66 [.48, .84]</td>
</tr>
<tr>
<td>Judgment (3-pt)</td>
<td>.85*** [.72, .99]</td>
<td>.50 [.32, .68]</td>
<td>.63 [.46, .81]</td>
<td>.68 [.51, .85]</td>
</tr>
<tr>
<td>Judgment (5-pt)</td>
<td>.87*** [.76, .98]</td>
<td>.56 [.38, .74]</td>
<td>.64 [.47, .81]</td>
<td>.66 [.48, .84]</td>
</tr>
</tbody>
</table>

Note. Men (n = 61) Women (n = 41) * p < .05 *** p < .001.
## Appendix F.

### Table 6

Area Under the Curve [95% CI] for HCR-20 and FAM Variables Predicting Violence Severity by Gender

<table>
<thead>
<tr>
<th></th>
<th>Low severity</th>
<th></th>
<th>High severity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HCR-20</td>
<td></td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Historical</td>
<td>.53 [.37, .69]</td>
<td>.52 [.34, .70]</td>
<td>.50 [.35, .65]</td>
<td>.66 [.48, .84]</td>
</tr>
<tr>
<td>Clinical</td>
<td>.58 [.41, .75]</td>
<td>.52 [.35, .70]</td>
<td>.62 [.48, .76]</td>
<td>.55 [.35, .74]</td>
</tr>
<tr>
<td>Risk Management</td>
<td>.56 [.40, .73]</td>
<td>.48 [.30, .65]</td>
<td>.59 [.45, .74]</td>
<td>.45 [.27, .64]</td>
</tr>
<tr>
<td>Total score</td>
<td>.56 [.39, .72]</td>
<td>.51 [.33, .70]</td>
<td>.54 [.39, .69]</td>
<td>.63 [.45, .81]</td>
</tr>
<tr>
<td>Judgment</td>
<td>.57 [.41, .73]</td>
<td>.43 [.25, .61]</td>
<td>.59 [.45, .73]</td>
<td>.64 [.47, .81]</td>
</tr>
<tr>
<td></td>
<td>FAM</td>
<td></td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Historical</td>
<td>.60 [.41, .78]</td>
<td></td>
<td>.76” [.61, .91]</td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>.68 [.51, .84]</td>
<td></td>
<td>.72” [.56, .88]</td>
<td></td>
</tr>
<tr>
<td>Risk Management</td>
<td>.49 [.31, .68]</td>
<td></td>
<td>.59 [.41, .78]</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>.63 [.44, .81]</td>
<td></td>
<td>.76** [.60, .91]</td>
<td></td>
</tr>
<tr>
<td>Judgment (3-pt)</td>
<td>.58 [.41, .75]</td>
<td></td>
<td>.78** [.64, .93]</td>
<td></td>
</tr>
<tr>
<td>Judgment (5-pt)</td>
<td>.64 [.46, .81]</td>
<td></td>
<td>.77** [.63, .92]</td>
<td></td>
</tr>
</tbody>
</table>

Note. Men (n = 61) Women (n = 41)” p < .01.
Appendix G.

Figure 1

Historical-Clinical-Risk Management-20 Coding Form

<table>
<thead>
<tr>
<th>Historical Items</th>
<th>Code (0, 1, 2)</th>
<th>Critical?</th>
<th>Criminogenic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Previous Violence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2 Young Age at First Violent Incident (Specify Age: ___)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3 Relationship Instability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H4 Employment Problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H5 Substance Use Problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H6 Major Mental illness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H7 Psychopathy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H8 Early Maladjustment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H9 Personality Disorder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H10 Prior Supervision Failure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Historical Item Total:</strong></td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical Items</th>
<th>Code (0, 1, 2)</th>
<th>Critical?</th>
<th>Criminogenic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Lack of Insight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite: Coded from current or past mental health evaluations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provisional: Refer for confirmation by mental health evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2 Negative Attitudes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3 Active Symptoms of Major Mental illness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite: Coded from current or past mental health evaluations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provisional: Refer for confirmation by mental health evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4 Impulsivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5 Unresponsive to Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Clinical Item Total:</strong></td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Risk Management items In / Out                       | Code (0, 1, 2) | Critical? | Criminogenic? |
| R1 Plane Lack Feasibility                            |                |           |              |
| R2 Exposure to Destabilizers                         |                |           |              |
| R3 Lack of Personal Support                          |                |           |              |
| R4 Noncompliance with Remediation Attempts           |                |           |              |
| R5 Stress                                            |                |           |              |
| **Risk Management Item Total:**                      | 10             |           |              |

**HCR-20 Total Score:** 40
Appendix H.

Figure 2

Female Additional Manual Coding Form
| C5 | Unresponsive to Treatment | 0 | 0 |
| C6 | Covert / manipulative behavior | 0 | 0 |
| C7 | Low self-esteem | 0 | 0 |

**HCR-20 + FAM Clinical Item Total:** 14

<table>
<thead>
<tr>
<th>Risk Management Items</th>
<th>In</th>
<th>Out</th>
<th>Code (0, 1, 2)</th>
<th>Critical?</th>
<th>Criminogenic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Plans Lack Feasibility</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>Exposure to Destabilizers</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>Lack of Personal Support</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>Noncompliance with Remediation Attempts</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>Stress</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>Problematic child care responsibility</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>Problematic intimate relationship</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HCR-20 + FAM Risk Management Item Total:** 14

**HCR-20 + FAM Total Score:** 82

<table>
<thead>
<tr>
<th>Other Considerations (if necessary)</th>
<th>Code (0, 1, 2)</th>
<th>Critical?</th>
<th>Criminogenic?</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>O3</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Overall Final Risk Judgment:**

- Low
- Moderate
- High

**Confidence of Final Risk Judgment:**

a) Low
b) 1-10: ______ (indicate number)

**Additional FAM Final Risk Ratings**

| Violence to others | Code (0, 1, 2) | Low
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Moderate</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Moderate-High</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

**Confidence of Final Risk Judgment:**

a) Low
b) 1-10: ______ (indicate number)
Appendix I.

Figure 3

START Outcomes Scale Coding Form

![START Outcomes Scale Coding Form Image]
<table>
<thead>
<tr>
<th>Suicidal Behavior</th>
<th>Self-Injurious</th>
<th># of times</th>
<th># of times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Makes impulsive gesture or attempts with low level of lethality.</td>
<td>99 DU</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>2. Makes attempt using potentially lethal means in non-suicidal situations with a high probability of success.</td>
<td>99 DU</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>3. Makes potentially serious attempt using lethal means in suicidal environment.</td>
<td>99 DU</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>4. Attempts without likelihood of successful suicide or recovery or completed suicide.</td>
<td>99 DU</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Substance Abuse</th>
<th># of times</th>
<th># of times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Returns late from unsupervised leave without prior notification or adequate explanation.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>2. Uses illegal substances to reduce or eliminate anxiety, stress results in moderate adverse effects for self or others.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>3. Absconds from unsupervised leave or is returned by police from unsupervised leave or does not return. Attempted escape from secure setting.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>4. Escapes from secure setting.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Being Victimized</th>
<th># of times</th>
<th># of times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bullied or intimidated by others on occasion (e.g., verbal threats, psychological, financial, injury, property damage, fear, or intimidation).</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>2. Physical assault results in mild to moderate physical injury (e.g., bruises, scratches, or swelling) or non-invasive medical intervention.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
<tr>
<td>3. Physical assault results in severe physical injury (e.g., broken bones, deep lacerations, internal injuries); or violent or obscene sexual assault.</td>
<td>99 DU</td>
<td>99 DU</td>
</tr>
</tbody>
</table>

Coding Notes: