Evaluation of the Performance of the HCR-20 across Genders

by

Diane S. Strub

B.A. (Hon.) Simon Fraser University, 2008

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

SIMON FRASER UNIVERSITY

Fall 2010

APPROVAL

Name:	Diane S. Strub
Degree:	Master of Arts (Department of Psychology)
Title of Thesis:	Evaluation of the Performance of the HCR-20 Across Genders
Examining Committee:	
Chair:	Dr. Kathleen Slaney Assistant Professor
	Dr. Kevin Douglas Senior Supervisor Associate Professor
	Dr. Jodi Viljoen Supervisor Assistant Professor
	Dr. Tonia Nicholls Supervisor BC Forensic Psychiatric Services Commission
External Examiner:	Dr. Margaret Jackson Professor School of Criminology
Date Approved :	December 10, 2010

Declaration of
Partial Copyright Licence

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

STATEMENT OF ETHICS APPROVAL

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
Simon Fraser University
Burnaby, BC, Canada

Last update: Spring 2010

ABSTRACT

Over the past three decades, much has been learned about risk factors associated with violence. Subsequently, significant advances have occurred in terms of the conceptualization as well as communication of violence risk and numerous risk assessment measures have been developed in order to inform violence prevention efforts. However, most such instruments have been validated in male populations and research examining their application across genders is scarce. This study investigated the performance of one of the most established violence risk assessment schemes - the Historical/Clinical/Risk Management-20 (HCR-20) – in a sample of 49 male and 46 female short-term psychiatric inpatients. Results indicate that the HCR-20 as well as its components predict the likelihood and imminence of violent outcomes. Moreover, gender did not moderate that relationship. Exploratory analyses revealed gender differences in the baseline item and scale ratings as well as in the nature of both the predicted and observed violence. Additionally, the HCR-20 demonstrated an association with other negative outcomes, particularly violent victimization.

ACKNOWLEDGMENTS

I feel privileged for having had the opportunity to work with so many inspiring individuals from the criminology and psychology departments at Simon Fraser University as well as from universities abroad. I most particularly want to express my immense gratitude to Kevin Douglas whom I respect so greatly both as a professional and as a person. Not only his academic excellence, patient guidance, and high professional standards, but also his encouragements, kindness and generosity have made my graduate journey under his supervision all I could hope for and more. I am extremely thankful to Jodi Viljoen and Tonia Nicholls for their valuable insights and assistance with this project. I appreciate the gracious efforts of the researchers, both locally and abroad, who made available to me unpublished disseminations and other important sources of information. Thank you to the other members of the Douglas lab whose hard work was instrumental in the completion of this project and whose friendships I am so tremendously grateful for. I am also appreciative of Joan Wolfe's precious help with the formatting of the final document. Finally, I am indebted to the family members and friends, who, whether next door or on the other side of the world, have continued to support me and remained so present, no matter how crazy and consuming my life pursuits, how great the physical distance, or how trying the circumstances.

Funding for this study was awarded to Dr. Kevin Douglas by the Canadian Institute of Health Research (CIHR) and the Michael Smith Foundation for Health Research (MSFHR). The author received scholarships from the Social Sciences and Humanities Research Council of Canada (SSHRC) as well as The Michael Smith Foundation for Health Research (MSFHR).

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Numerous legal, forensic and psychiatric institutions are confronted with the difficult task of determining whether a specific individual may be at risk of harming others. Violence risk assessments are conducted in order to find a way to reduce or manage that risk, either in the community, or within institutions. Because the decisions made by mental health professionals during such assessments have serious consequences, it is essential that the decision-making processes involve conceptually, empirically and clinically sound instruments in order to foster informed and focused case formulations. Indeed, with regard to violence risk predictions, false positives can result in the unjust violation of a patient's rights and freedom (e.g., judicial sentencing, commitment decisions, involuntary medication, refusal of release from institutions, etc.), whereas false negatives might generate devastating and potentially lethal outcomes (Bloom, Webster, Hucker, & De Freitas, 2005).

Because an individual's true risk for violence can never be known, it has to be estimated with regard to various conditions (Hart, 2001). This critical part of decision making in clinical practice and in criminal justice and psychiatric settings involves the gathering of specific information pertaining to recognized risk factors and a classification according to perceived risk (Douglas, Ogloff, Nicholls, & Grant, 1999; Hart, 2001). Moreover violence risk assessments embody dynamic and ongoing processes since risk can change over time, across conditions and in response to various interventions (Hart,

2001; Otto, 2000). Depending upon the evaluated level of risk for a given individual (e.g., low, moderate, high) at a certain time, a judgment can be made about the management of this person in the community or in forensic and psychiatric institutions.

Over the past 25 years, much has been learned about numerous risk factors associated with violence (e.g., demographic, historical, dispositional, clinical, environmental, and contextual factors) and the accuracy of violence prediction has considerably improved (Bloom et al., 2005; Douglas et al., 1999; Nicholls, Ogloff, & Douglas, 2004). Indeed, violence risk assessments seem to perform significantly above chance level and produce at least moderate effect sizes (Mossman, 1994, Otto, 2000). Several such risk assessment instruments have been developed and used in order to predict and, most importantly, prevent future violence.

However, the overwhelming majority of the data inherent to the creation of available risk measures has been generated from research on all or predominantly male samples. Indeed, because men, regardless of age, ethnicity, socioeconomic status and culture, commit the majority of violent crimes, risk assessment and management of aggressive behavior in males has been extensively investigated over the past three decades, whereas female populations have received scarce interest (Nicholls, Garcia-Mansilla, & Risenfeld, 2009; Strand & Belfrage, 2001). Based on the available research addressing gender considerations in violence risk evaluations, women and men have been shown to sometimes be more similar or less so than expected with respect to some specific factors. For example, although the base rate of violence is much lower in women overall, it does not appear to differ between female psychiatric inpatients and their male counterparts (Binder & McNiel, 1990; Lidz, Mulvey, & Gardner, 1993). Moreover,

research has shown that when using unstructured clinical judgment, clinicians tend to underestimate the risk for violence in female psychiatric patients (Lidz et al., 1993; Mc Niel & Binder, 1995) and the use of structured risk assessments has been recommended (Borum, 1996). However, although the known risk factors for violent behavior typically included in such structured assessment measures have been demonstrated to apply to both genders in a general sense (de Vogel & de Ruiter, 2005; Strand & Belfrage, 2001), as stated earlier, most such instruments were developed based on research conducted primarily with males. Professionals in the field have communicated a need to examine in more detail a necessity, or lack thereof, to adopt more gender-tailored assessment processes (e.g., Binder & McNiel, 1990; Monahan et al., 2001; Nicholls, 1997; Nicholls, Garcia-Mansilla, & Rosenfeld, 2009; Odgers, Moretti, & Reppucci, 2005).[1]

The Historical/Clinical/Rick Management-20 (HCR-20) Version 2 (Webster, Douglas, Eaves, & Hart, 1997) embodies one of the most well-established violence risk assessments schemes for community and institutional violence in offenders, civil psychiatric patients, forensic psychiatric patients, as well as, more generally, males and females with mental illness, personality disorder, or substance abuse. Although the HCR-20 has demonstrated good reliability and predictive validity in aggregate results from samples comprising both genders (e.g., Douglas & Reeves, 2010; Douglas, Ogloff, & Hart, 2003; Hart, Douglas, & Webster, 2001; Otto, 2000), a very limited body of research reports detailed and informative gender comparisons (Coid et al., 2009; de Vogel & de

[1] Although de Vogel and colleagues (2010) are currently testing a set of women-particular additional guidelines for existing violence risk measures, to my knowledge, only one structured professional assessment instrument has been developed specifically for the assessment of violence risk in females: The Early Assessment Risk List for Girls (EARL-21G; Levene, Augimeri, Pepler, Walsh, Webster, & Koegl, 2001).

Ruiter, 2005; Nicholls, 2001; Nicholls, Ogloff, & Douglas, 2004; Schaap, Lammers, & de Vogel, 2009; Strand & Belfrage, 2001; Warren et al., 2005). The present study, which examines the specifics of the performance of the instrument across genders in a civil psychiatric sample, represents the first such evaluation of the HCR-20 in a short-term psychiatric setting.

Professional Judgment and Actuarial Decision-Making

Mental health professionals commonly use two approaches to make decisions about violence risk: *actuarial decision-making* and *professional judgment* (Hart, 2001). These two approaches differ in the way information is weighted, combined, and considered. With the actuarial approach, the ultimate conclusions about violence risk are reached according to specific numeric rules. On the other hand, professional judgment procedures imply that the evaluator uses some degree of discretion in the decision-making process.

There are two major types of actuarial decision-making processes: the use of *actuarial risk assessment instruments* and the *actuarial use of psychological tests* (Hart, 2001). The Violence Risk Appraisal Guide (VRAG, Quinsey, Harris, Rice, & Cormier, 1998) and the Violent Offender Risk Assessment Scale (VORAS; Howells, Watt, Hall, & Baldwin, 1997) are examples of an actuarial risk assessment instrument. The Psychopathy Checklist family of instruments (PCL-R, Hare, 1991, 2003; Psychopathy Checklist Screening Version, Hart, Cox, & Hare, 1995) are examples of psychological tests now commonly used as part of structured professional judgment guidelines (e.g., one of the 20 items of the HCR-20), as part of an actuarial risk assessment instrument

(e.g., one of the 12 items of the VRAG), or used independently in an actuarial way for research purposes.

Professional judgment comprises three major procedures. The first, *unstructured professional judgment*, involves an informal and subjective decision-making process with a total absence of structure (Hart, 2001). The second, the *anamnestic risk assessment* procedure, requires that the evaluator identify the personal and situational factors that resulted in violence in the past for a given individual. It therefore imposes a minimal structure on the assessment. In contrast, the third procedure, the *structured professional judgment* or *SPJ model*, following the medical tradition of using clinical guidelines to aid decision-making, involves the use of *aides-mémoires* that reflect the state of empirical and clinical knowledge of the discipline. The HCR-20 violence assessment scheme (Webster et al., 1997) embodies one such guideline.

A Structured Professional Approach to Decision Making: The SPJ Model

In the past decade, researchers have reached the consensus that unstructured professional judgment generates inadequate interrater reliability and predictive validity (Douglas et al., 2003; Douglas, Yeomans, & Boer, 2005; Guy, 2008; Quinsey et al., 1998). As a result, structured decision-making approaches, in the form of actuarial or SPJ models, have been favored. Despite the high levels of statistical accuracy demonstrated by actuarial predictions conducted on samples of psychiatric patients and offenders, concerns have been expressed with regard to the generalizability and clinical applicability of the results, failure to take into consideration information that may be crucial in the

evaluation of individual cases, lack of sensitivity to change, and failure to address risk management or prevention strategies (Douglas et al., 2003; Litwack, 2001). Comprehensive and rationally developed SPJ guides, such as the HCR-20, capitalize on the strengths of both the unstructured clinical and actuarial models while simultaneously avoiding most of their weaknesses, and improving the consistency as well as usefulness of clinical decisions (Bloom et al., 2005; Hart, 2001; Otto, 2000). However, it is worth noting that it is not a combined "clinical-actuarial" or an "adjusted actuarial" approach (Douglas & Reeves, 2010).

In contrast to most actuarial instruments, the set of risk factors considered as part of the SPJ model of risk assessment is selected based on a review of a multitude of studies, rather than derived from investigations conducted on a single sample (Tengström, 2001). In addition, one of the strengths of this approach lies in the variety of risk factors that are examined. Whereas other assessment models tend to emphasize *risk status* or the "interindividual variability in risk" SPJ evaluations additionally take into consideration *risk state* or the "intraindividual variability in violence potential" (Douglas & Skeem, 2005; Skeem & Mulvey, 2002, p. 118). Violence risk assessment has thus come to involve an ongoing process of decision-making which requires an informed estimation of the ever-changing human and situational conditions that might aggravate or mitigate the likelihood and imminence of violent behaviors. Such contemporary approaches to risk assessment allow for the consideration of past fixed risk factors as well as recent more inconstant markers of violence risk. This more recent focus on dynamic risk factors, in addition to static ones, not only fosters a more complete understanding of causal mechanisms for violence, but also facilitates the development of monitoring,

treatment and management plans (Heilbrun, Yasuhara, & Shah, 2010). Indeed, if violent behavior can be predicted based, in part, on variables that are subject to change or changeable, optimized interventions may be developed in order to specifically target these variables. One must not forget that the ultimate clinical goal of violence risk assessment is violence *prevention*, rather than the mere *prediction* of a violent outcome (Douglas & Kropp, 2002; Guy, 2008; Sjöstedt & Grann, 2002). The concept of violence risk assessment has been defined as "the process of identifying and studying hazards to reduce the probability of their occurrence" (Hart, 1998, p. 121). Whereas actuarial scales generate probabilistic statements based on nomothetic considerations, SPJ instruments aid to inform idiographic risk management strategies.

Another strength of the structured professional model pertains to its ability to maintain the structured format of actuarial approaches (i.e., fixed number of operationally defined items, explicit coding rules, explicit instructions to guide final risk decisions) while avoiding the limitations inherent to weighted procedures and final risk judgments based on fixed algorithms. It has been demonstrated that equal-weighting models can produce robust predictions of a criterion (Dawes, 1979) and at the same time enhance generalizability compared to assessment methods involving the assignment of weights calculated based on the statistical relevance of each variable to a particular calibration sample. Indeed, the predictive accuracy of such specifically weighted scales tends to

considerably decrease upon application to other samples and populations (e.g., Grann, Belfrage, & Tengström, 2000; Grann & Långström, 2007).[2]

Moreover, rather than following rigid mathematical rules, the clinician formulates a final professional opinion informed by a structured consideration of empirically-based risk factors, the idiographic relevance of these to the case at hand, all additional available information, and the foreseen level of risk management efforts needed (Douglas & Skeem, 2005; Hart, 2001). In fact, it appears that in some cases, although the assessee presents with minimal number of risk factors, the salient and unalterable character of some of these violence facilitating variables yields a substantial risk for adverse outcomes (Douglas et al., 1999; Douglas & Ogloff, 2003). Whereas procedural guidance maximizes the consistency and transparency of judgments, the use of discretion provides the assessor with the ability to address risk management and prevention strategies tailored to the unique situations of the individuals being evaluated. As a result, the SPJ violence prevention paradigm allows for a uniquely dynamic and pragmatic convergence of both the rigor of scientific research and the flexibility necessary to clinical practice. The aforementioned fit of this comprehensive approach with the demands of clinical reality has been described as a *method-function match* (Hart, 1998). Indeed, the value of the

[2] Grann, Belfrage and Tengström (2000) compared the predictive acumen of the VRAG and the Historical subscale of the HCR-20 Version 2 in a sample completely independent from those inherent to the original work on both measures. The HCR-20 Historical subscale fared better in predictive future violence compared to the VRAG, even though the latter instrument uses optimized weights. Moreover, Grann and Långström (2007) examined the predictive accuracy of the Historical subscale of the HCR-20 Version 2 under five conditions, one unweighted and four weighted with increasing complexity. Non weighted total scores predicted reconviction for a violent crime with a mean AUC of .71 (range = .67 to .78 in five subpopulations of offenders) whereas weighted total scores generated AUCs between .72 (range = .66 to .73) and .64 (range = .58 to .72). These findings suggest that using weighting algorithms does not improve actuarial prediction of risk and rather, tends to result in statistical shrinkage, especially as the weighting procedures become more sophisticated. The authors propose that "weighting risk factor is a premature exercise as too little is still known about the proximal causes and true mechanisms behind violence" (p. 33).

structured professional judgment model of decision-making stems from its ability to simultaneously identify individual levels of risk, be useful in varied contexts and settings, rely on clinical consultation during both the development and testing of measures, be testable or falsifiable, generate rational and extensive evaluations of idiographic risk at all stages of the intervention process, consider both static and dynamic risk factors, imply that the assessors exhibit a considerable level of training and expertise, promote accountability, encourage discussion among and between researchers and clinicians, prove valuable for many practitioners without requiring adherence to any theoretical orientation, and leave room for the consideration of protective factors (Bloom et al., 2005; Douglas & Kropp, 2002; Hart, 1998; Müller-Isberner & Fransson, 2002).

The First SPJ instrument: The HCR-20 Violence Risk Assessment Scheme

In the 1990s, changes in the law concerning mentally disordered offenders made apparent the need for advances in the area of violence risk assessment (Bloom et al., 2005). Subsequently, the Mental Health Law and Policy Institute of Simon Fraser University consulted with numerous professionals across psychiatric, forensic and civil settings in order to identify the crucial factors to consider with regard to both the prediction and prevention of violent behavior. Webster, Eaves, Douglas and Wintrup developed the first SPJ instrument in 1995, the HCR-20. The scheme was revised in 1997 with the input of German and Swedish colleagues (HCR-20 Version 2; Webster et al., 1997). Four years later, a Companion Guide (HCR-20CG; Douglas, Webster, Hart, Eaves, & Ogloff, 2001) was published for the purpose of assisting mental health

professionals in the identification and development or risk management plans and provisions of services to patients and parolees (Bloom et al., 2005; Douglas & Reeves, 2010). The HCR-20 Version 2 has been translated into 16 languages and is used across Europe, South America, Australia and Asia. It can be used in the following contexts: (a) admission decisions upon entry to correctional, psychiatric or forensic facilities, (b) decisions about release from such institutions, (c) risk monitoring within institutions, and (d) risk monitoring when an individual is under community supervision. The HCR-20[3] has been formally adopted for forensic, correctional, and civil use in North America (e.g., Correctional Service of Canada, Ohio Department of Mental Health, New York Office of Mental Health, Forensic Bureau, and the Forensic Services Division of the Department of Mental Health in California) and on other continents (e.g., in Sweden, Germany, Australia, Japan, Canada, the United Kingdom, Norway, the Netherlands, and Denmark) (Guy, 2008).

In the HCR-20 manual, violence is defined as "actual, attempted and threatened harm to a person or persons" (p.24). Threatened or attempted harm fit the definition of violence as would the actual infliction of physical injury. Douglas and Reeves (2010) explain that "there is simply no meaningful logical difference between attempted and completed violence in terms of the behavior and intention of the perpetrator, *other than* he or she failed" (p. 148). They illustrate their statement with the example of a shooter who fires a gun into a crowded room but strikes no-one. Moreover, some actions that primarily cause psychological harm count as violence as well as (e.g., stalking,

[3] For the remaining of the present paper, "HCR-20" will refer to the current version of the scheme (i.e., Version 2) unless otherwise specified.

kidnapping, unlawful confinement, extortion, etc.) since empirical research indicates that psychological damage can be as or more harmful than physical injury. Conversely, some clearly physically aggressive and injurious acts may not count as violence (e.g., acts of self-defense with a degree of force not exceeding what is required for protection, acts inherent to the practice of certain sports, law enforcement acts not exceeding their legal mandate, etc.) (Douglas & Reeves, 2010).

The HCR-20 evaluates 20 variables grouped in three temporal domains (Table 1):

1. the 10 *H* items refer to *historical* aspects of violence risk and address such past events and static individual features as mental disorder and social adjustment. Fall into three general categories: 1) problems in adjustment or living (i.e., H3, H4, and H8), 2) problems with mental health (i.e., H5, H6, H7 and H9), and 3) past antisociality (i.e., H1, H2, and H10). Although historical, these items are essential in providing a context from which to understand someone's present manifestations of both static and dynamic risk factors as well as their relevance, and they are crucial to the overall formulation of both present and future violence risk (Douglas & Reeves, 2010).

2. The five *C* items refer to *clinical* aspects of recent and present psychological functioning of the individual. The latter items embody dynamic variables that can change acutely. They are inferred from recent behavior and need to be re-assessed on a regular basis. Subsequently, risk conceptualization as well as recommended or implemented intervention/management strategies should be re-formulated accordingly (Douglas & Reeves, 2010; Tengström, 2001).

3. The five *R* items focus on potential future adjustment problems and on *risk management.* Because potential obstacles to successful management may change over time, according to varying life circumstances, and following fluctuations in a person's current functioning, the R items should be re-evaluated whenever the C-scale is re-examined.

Table 1. HCR-20 Version 2

Subscale	Item	Risk Factor
Historical	H1	Previous Violence
	H2	Young Age at First Violent Incident
	H3	Relationship Instability
	H4	Employment Problems
	H5	Substance Use Problems
	H6	Major Mental Illness
	H7	Psychopathy
	H8	Early Maladjustment
	H9	Personality Disorder
	H10	Prior Supervision Failure
Clinical	C1	Lack of Insight
	C2	Negative Attitudes
	C3	Active Symptoms of Major Mental Illness
	C4	Impulsivity
	C5	Unresponsive to Treatment
Risk Management	R1	Plans Lack Feasibility
	R2	Exposure to Destabilizers
	R3	Lack of Personal Support
	R4	Noncompliance with Remediation Attempts
	R5	Stress

Note. Adapted from Webster et al. (1997).

The items of the HCR-20 Version 2 (Webster et al., 1997) were based on an evaluation of risk factors for violence selected from a review of numerous studies and each of these items has demonstrated predictive power across multiple subsequent studies and meta-analyses (e.g., Guy, 2008; Hanson & Bussiere, 1998; Heilbrun & Kramer, 2001; Otto, 2000; Tengström, 2001).

Each variable of the HCR-20 receives a unit-weight score of 0, 1, or 2 (0 = *absent or does not apply*; 1 = *the item might be present or is present to a limited extent*; 2 = *the*

item is definitely present) for an HCR-20 maximum total score of 40. Based on the number of risk factors present, their relevance, any other relevant case specific variables, and the degree of anticipated management necessary to mitigate risk, the assessor then formulates a risk estimate of *Low*, *Moderate* or *High* risk. Someone should be considered *High risk* if many relevant risk factors are present and/or he/she requires "frequent intensive, or highly restrictive supervision, monitoring, management or intervention" in order to mitigate violence risk (Douglas & Reeves 2010, p.158). *Low risk* cases are those in which there are few relevant risk factors present and/or which require "minimal or no supervision, monitoring, management or intervention in order to stem violence risk" (p. 158). Individuals who do not fit either of the *High* or *Low* risk definitions should be considered to be at *Moderate* risk of violent behavior.

It is important to remember that these HCR-20 final risk judgments do not represent the final goal of a structured professional assessment of violence risk (Douglas & Reeves, 2010). Indeed, they merely embody a way to summarize the information that has been considered, rated and integrated into a risk decision, and a means to communicate level of needs for services as well as management in order to engage in adequate case prioritization. The clinician then makes recommendations regarding the specific resources and intervention strategies required to minimize the risk for future violence, and outlines the scenarios under which the estimated risk may be reduced or increased.

Research on the HCR-20

The structured professional judgment model has now been tested through over 100 independent empirical studies and the HCR-20 represents one of the most researched and established guides used to assess risk for community and institutional violence in offenders, civil psychiatric patients, forensic psychiatric patients, as well as males and females with mental illness, personality disorder, or substance abuse (e.g., Douglas & Reeves, 2010; Hart et al., 2001; Nicholls et al., 2004; Otto, 2000). This broad-band instrument has demonstrated good interrater reliability, validity, and concurrent validity with other commonly used instruments (e.g., Belfrage, Fransson, & Strand, 2004; Bloom et al., 2005; Douglas & Webster, 1999; Douglas et al., 2003; Douglas et al., 2005; Guy, 2008; Nikolova, Collins, Guy, Lavoie, Reeves, Wilson, & Douglas, 2006). Indeed, the scale generates moderate to large effect sizes with regard to violent recidivism. Moreover, the HCR-20 approach to violence risk assessment provides the evaluator with the flexibility necessary for applications across different settings and populations, and, more specifically, for the assessment of individuals whose risk may vary as a function of phase of illness and clinical context (McNiel, Gregory, Lam, Binder, & Sullivan, 2003; Otto, 2000).

Since the plethora of published and unpublished research on the HCR-20 is too voluminous to describe in details (i.e., over 50 studies), a summary of relevant findings is presented next. Across 36 studies (13 unpublished and 23 published), the total scores on the instrument yield good to excellent *interrater reliability*, with ICCs[4] ranging from .67

[4] ICC refers to Intraclass Correlation Coefficient, an index of reliability that corrects for additive and multiplicative biases (Bartko & Carpenter, 1976).

to .95 (median = .85), but with the majority of studies reporting coefficients of .80 or greater (Douglas & Reeves, 2010). The median interrater value for the H subscale is .86 (range = .58 to .97), .74 for the C subscale (range = .55 to .95) and .68 for the R subscale (range =.47 to .98). The few existing empirical investigations pertaining to the final risk judgments generated a mean interrater coefficient of .65 (range = .41 to .76).

The results of predictive validity analyses across 42 studies indicate a moderate association between the HCR-20 and violence, with a median AUC[5] value of .69 (Douglas & Reeves, 2010). The median effect sizes for the H, C and R subscales are .68, .62, and .65 respectively. Across the few studies that examined the predictive acumen of the final risk judgments, the average AUC was .70. Moreover, in a recent meta-analysis (Guy, 2008), aggregate results for 328 weighted AUC values yielded a largest mean weighted effect size (AUC_w = .79) for such summary judgments when the outcome criterion was physical violence, with or without sexual acting out (AUC_w for summary risk ratings = .70 for a general antisocial index, .76 for a general violence index, and .67 for a non-violent index). It is worth noting however, that the total scores exhibited a no better than chance association with future sexual violence (AUC_w based on the two available effect sizes = .46). This is not surprising since the development of the HCR-20 was informed by empirical knowledge pertaining to risk factors for general rather than sexual violence (Guy, 2008).

After examining the nine sets of bivariate comparisons that investigated the predictive power of the summary risk judgments, seven yielded greater effect sizes for

[5] AUC represents the Area Under the Curve of a receiver operating characteristic analysis. Such analyses and their relevance to the project at hand are described in the *data analytic methods* section of the present document.

the summary risk ratings compared to the total scores on the instrument and, on average, the AUC for the summary risk ratings was considerably larger than the H, C and R subscale scores (Douglas & Reeves, 2010). In the handful of studies in which multivariate comparisons were conducted, the summary risk ratings demonstrated incremental validity above and beyond the HCR-20 numeric total scores. In terms of the predictive performance of the HCR-20 compared to that of other decision-making approaches (PCL-R; VORAS; VRAG; unstructured clinical prediction), the studies conducted to date (i.e., six HCR-20 studies and other SPJ studies) have indicated that the summary risk ratings are either more strongly related to violent outcomes than the other instruments, adds incrementally to them, or possesses unique predictive variance.

Gender, Violence and Violence Risk Assessment

Regardless of age, ethnicity, socioeconomic status or culture, gender is one of the most significant predictors of formally sanctioned violence, with males being convicted of most violent crimes (e.g., Monahan et al., 2001). Moreover, the multifaceted nature of violence should not be forgotten when comparing genders since research has indicated that the context, nature, severity, developmental course and victims of female aggression seem to differ from those of male perpetrated violence (e. g., Binder & McNiel, 1990; de Vogel & de Ruiter, 2005; Odgers et al., 2005). Indeed, the violence committed by women tends to be more reactive (i.e., as opposed to instrumental), less severe, more rarely resulting in injury, more often directed at close others (i.e., as opposed to strangers or acquaintances), more likely to occur in the home, more rarely sexual in nature, and overall less noticeable than of men (McKeown, 2010; Monahan et al., 2001; Nicholls,

1997; Nicholls, Greaves, & Moretti, 2008). In addition, although men and women often evidence equivalent motives behind the perpetration of aggression (e.g. domestic violence; Graham-Kevan, 2007), the purposes violent behaviors serve seem to sometimes diverge across genders (McKeown, 2010). For example, although men and women commit non-familial homicides for similar reasons, females tend to murder intimate partners following the prolonged experience of domestic abuse, whereas for their male counterparts, jealousy, infidelity, desertion, dissolution of the relationship, and a need for control seem to drive the commission of such lethal acts. Moreover, since social bonds are of greater importance to women and because women therefore experience greater distress than men do when these are threatened, interpersonal disruptions may differentially motivate violence across genders (e.g., Funk, 1999; Odgers et al., 2005).

Although mental illness per se is not a strong risk factor for violence, it plays a substantial role in women's propensity to act violently (Manchak, Skeem, Douglas, & Siranosian, 2009). Among psychiatric inpatients the base rate of violent acting out does not appear to differ between males and females (e.g., Lidz, Mulvey, & Gardner, 1993). Indeed, female psychiatric patients may exhibit a greater risk of aggression compared to women in the general population but whether or not this violence also tends to be more severe and/or as severe as that of male psychiatric inpatients remains a source of debate. Although male psychiatric inpatients exhibit more violence prior to hospital admission generally, research indicates that hospitalized women may be more assaultive than hospitalized men, whereas the latter individuals appear to engage in more fear inducing behaviors during the first days of admission (Binder & McNiel, 1990). Other findings suggest that injuries to staff members in psychiatric settings are as likely to be caused by

violence from female as from male patients, even after controlling for other correlates of violence (e.g., Lam, McNiel, & Binder, 2000). The fact that institutional staff generally both underestimates the violence potential of women as well as overestimates that of men, and further fails to notice cues that could help distinguish women at risk to act violently (e.g., Lidz, Mulvey, & Gardner, 1993; Skeem, Schubert, Stowman, Beeson, Mulvey, Gardner, & Lidz, 2005) may in part explain the aforementioned observations as more effort may be directed at the management of men. Other factors may also be at play: For example, acute exacerbations of major mental illness may attenuate the conformity of gender roles pertaining to aggressive displays.

Much research highlights differing pathways to both general as well as violent offending for men and women (for reviews, see McKeown, 2010; Odgers et al., 2005). A growing body of empirical literature underlines the importance of considering risk factors which are not commonly included in current risk assessment practices when dealing with female populations: Victimization, post-traumatic stress, childcare and parenting difficulties, low self-esteem and self-efficacy, poverty, social marginalization, women's unique health, past and recent suicidal behaviors, and specific variables pertaining to the female experience of dysfunctional intimate relationships. In addition, female offenders exhibit higher rates and greater severity of childhood sexual abuse as well as all other types of abuse, repeated victimization both as children and as adults, trauma, depression, and other Axis I syndromes.

Sexual abuse seems to play a pivotal role in the background of those women who break the law (e.g., female victims of sexual abuse appear more likely to be arrested for a violent juvenile offense) (McKeown, 2010; Odgers et al., 2005). Indeed, compared to

males, females victims of child sexual abuse tend to be abused more often by close others and for longer periods of time, which in turn generates greater levels of trauma, psychological distress, depression and hazardous survival strategies such as substance abuse, running away, prostitution, and involvement in risky and/or criminogenic relationships.

Moreover, higher levels of mental illnesses such as schizophrenia, delusional disorder, lifetime depression, PTSD, drug dependence and alcohol disorder have been observed in female offenders, which supports a link between psychopathology and female pathways to violent and general criminal acting out (McKeown, 2010; Odgers et al., 2005).

In addition, although prior violence and early onset of aggressive or antisocial behavior represent some of the most robust predictors for future violent offending in male populations, the predictive power of these variables is less clear for females (Odgers et al., 2005). Further, because women exhibit a greater propensity to engage in relational aggression[6] and act out in private spheres, their violence is less detectable than that of men which is often perpetrated in public contexts and against strangers (Monahan et al., 2001; Odgers et al., 2005). Subsequently, a significant proportion of violent women are not identified as such, and risk factors inherent to their behavioral presentation may not be included in current prediction models or adequately understood (i.e., when included).

[6] It is worth noting that relational forms of aggression appear to be equally as harmful as more overt violence (e.g., Crick & Bigbee, 1998; Paquette & Underwood, 1999).

The HCR-20 and Gender

Below are summarized all the studies which have explicitly investigated gender issues pertaining to assessments conducted with the HCR-20 violence assessment scheme. To my knowledge, no other such empirical enquiry has been disseminated to date. However, findings pertaining to HCR-20-related investigations beyond the scope of the present empirical exercise (e.g., performance of the PCL-R across gender and incremental validity analyses across various risk instruments) were not reported.

Warren and colleagues (2005) documented interesting descriptive results in their sample of 132 female inmates from a maximum security facility. Interrater reliability was good to excellent in this sample. The inmates who had been convicted of murder scored significantly lower on the HCR-20 compared to those who had not such convictions, whereas those who had committed either property or minor crimes exhibited significantly higher scores compared to those who exhibited no such index offenses. No differences were noted for other criminal categories, including violent, potentially violent, sex, and drug crimes. Moreover, there were no differences between those women scoring high and low on the HCR-20 and those who had or had not engaged in institutional violence. In terms of subscale scores, inmates did not differ according to whether they had or had not committed the various types of crimes recorded, to the exception of minor crimes. Women who scored in the lowest quartile of the HCR-20 total score were more likely to be charged with first degree murder and those who scored in the highest quartile were more likely to be charged with robbery, breaking and entering, larceny, probation violation, failure to appear, and prostitution. Results from ROC analyses indicated that

the HCR-20 yielded a no better than chance postdiction for index violent charges in this sample, but accurately forecasted (i.e., in a retrospective fashion) a range of non-violent index crimes. Unfortunately, such *postdictive* observations yield little insight into the adequacy of the instrument across genders for its intended use, namely the *prediction* of *future* violent behavior.

Strand and Belfrage (2001) retrospectively assessed with the HCR-20 all male (N=85) and female (N=63) inpatients from two forensic psychiatric hospitals in Sweden. They reported very good interrater reliability in both subsamples. There were no gender differences in the HCR-20 total scores or on the subscale scores. On the historical factors, women scored higher for *personality disorder* and lower for *previous violence, young age at first violent incident* and *substance use problems* (i.e., the frequency of previous violence and substance abuse was similar across genders, but men engaged in most severe forms of these behaviors compared to women). On the clinical and risk management variables, men exhibited more *negative attitudes* and women more *impulsivity* and *stress*. Once again, although informative in a descriptive sense, this study generates a very limited understanding of the adequacy of the HCR-20 for clinical, forensic and correctional use with female populations since predictive validity enquiries across gender were not undertaken.

Nicholls (2001) pseudo-prospectively examined the performance of the HCR-20 in a sample of 47 women and 47 matched men found *Not Criminally Responsible on Account of Mental Disorder* (NCRMD) in British Columbia, Canada. Some of the analyses were conducted on a subsample of these individuals (i.e., N= 61 to 70 depending

on the analyses) who had been followed up in the community for approximately four years. With respect to the scores on the HCR-20, measures of dispersion as well as subscale and total scores were comparable across genders. The only difference observed at the item level resulted from men scoring higher on *substance abuse problem*s compared to women. Men and women with a past history of aggression had significantly higher H and total scores (with item H1 removed) compared to individuals with no such history.

With regard to post-evaluation institutional events, the HCR-20 strongly predicted both men and women's (AUCs = .75 and .81 respectively) inpatient physical aggression (i.e., incidents not involving the use of weapons and not resulting in physical injuries) as well as women's inpatient violence (i.e., "violence" referred to sexual assaults, physical assaults resulting in physical injury, and verbal threats with a weapon in hand) and property damage (AUCs = .80 and .77). The H subscale yielded a similar relationship with the outcomes and the C subscale correlated with inpatient physical aggression for both genders. The R subscale did not correlate with any of the outcomes in men but did so for all outcomes (verbal and physical inpatient aggression as well as inpatient violence) in women. The HCR-20 was not associated with verbal aggression in either gender. Time to first inpatient aggression did not vary according to whether the women scored above or at/below a cutoff total score of 25 on the HCR-20, whereas, on the other hand, high risk males became aggressive in the institution sooner compared to lower risk men. When the outcome was narrowed down to inpatient *physical* aggression, dichotomized HCR-20 scores did not distinguish physically and non-physically aggressive patients and did not predict time to first aggression.

As it pertains to post-release outcomes, although the H subscale predicted physical aggression and recidivism in men and the C subscale predicted recidivism in women, the HCR-20 failed to predict community physical aggression and violence in both genders. Dichotomized HCR-20 scores (i.e., above or at/below cutoff) did not correlate with criminal recidivism for either gender. The author indicates that the absence of predictive accuracy for many postdischarge outcomes may have resulted both from low power and from having rated the C and R items solely at the first disposition determination (i.e., at the start of the patients' NCRMD status) rather than having these risk factors reassessed prior to release, as is recommended in the HCR-20 manual (Webster et al., 1997). It is worth noting that this study did not investigate the predictive power of the final risk judgments. Although, in a general sense, the greater the number of risk factors present the greater the risk for violence, final risk ratings summarize individual formulations of risk more meaningfully than total scores or characterizations based on numerical cutoffs do and it would have been worthwhile to examine the performance of the HCR-20 across genders at that level.

Nicholls et al. (2004) conducted a pseudo-prospective study in a sample of 117 male and 75 female chronic involuntary psychiatric inpatients from a Canadian psychiatric hospital. Interrater reliability for the HCR-20 ratings was good. Unlike the results from Strand and Belfrage's empirical enquiries into a forensic population (2001), in this civil sample, women tended to exhibit lower H, C and total scores compared to men. Not surprisingly, they were also less likely than males to be classified as high risk using either dimensional or categorical ratings. Overall, in terms of postdischarge community outcomes (i.e., *any violence*, *physical violence*, *any crime*, and *violent crime*),

the HCR-20 total scores generated moderate to large effect sizes for violence and criminal offending in both men (AUCs = 0.67 - 0.75) and women (AUCs = 0.66 – 0.83). In the male subsample, the H and R subscales as well as the total HCR-20 score predicted all four outcomes, whereas the C subscale only predicted *any violence*. In women, the H subscale and the HCR-20 total score predicted all outcomes except physical violence, but the scores on the C and R subscales did not significantly forecast any of the four selected postdischarge behaviors. Interestingly, the predictive accuracy for criminal behavior (including violent crime) was greater for women (AUC > .80) than it was for men (AUC < .80). The authors suggested that the HCR-20 apparent lack of predictive power for physical violence in women may have resulted in part from having relied on official records rather than comprehensive community follow up to document outcomes. Indeed, although official records tend to misrepresent the perpetration of aggression in both genders, it may be the case to a greater extent for women whose violence often goes unnoticed. Another limitation of the present study lies in the omission of final risk judgments as part of its HCR-20–based predictive enquiries.

Coid and colleagues (2009) carried out a prospective study with male (*N* = 1353) and female (*N* =304) prisoners serving a sentence of two years or more for a sexual or violent index offence (excluding life sentence) in England and Wales. Interrater reliability was good to excellent. Men scored significantly higher on the C subscale of the HCR-20, whereas women scored higher on the HCR-20 total, H and R scores. In the male subsample, the HCR-20 total H, C and R scores predicted *violent*, *acquisitive* and *any* reconvictions. The HCR-20 total H, and C scores, predicted all three outcomes in women but the R subscale failed to reach significance for the prediction of violent or acquisitive

reconvictions. In the overall sample, the HCR-20 and its H and C subscales forecasted reconvictions with small to moderate accuracy, but the R subscale evidenced very low predictive power. Formal moderation analyses indicated no impact of gender on the predictive accuracy of the HCR-20 or its subscales for violent, acquisitive and any reconvictions in this correctional sample. The authors indicated that because of methodological constraints (i.e., participants were interviewed during the 6 to 12 months before their release and followed up between 7 and 1317 days postdischarge) and since the timing of ratings is crucial for the clinical and risk management variables (i.e., because of their dynamic nature, these items should be re-evaluated every 6 to 12 months), the lower performance of the C and R subscales in this study is unsurprising. As for the other four aforementioned studies, the failure to examine the predictive power of the HCR-20 final risk judgments represents an important limitation since those ratings, rather than total or subscale scores, should represent the basis for clinical, forensic and legal decisions.

de Vogel and de Ruiter (2005) examined the interrater reliability and predictive validity of the HCR-20 in a matched sample of female (*N*= 42) and male (*N*= 42) inpatients from a Dutch forensic psychiatric hospital. On the HCR-20, women scored higher for *relationship instability* and *impulsivity* and lower for *young age at first violent incident*, *psychopathy*, and *negative attitudes* compared to men, differences generally in line with the results reported by Strand and Belfrage (2001). Men and women did not differ in their subscale and total scores but the female participants were more often assessed to be at moderate risk for violence and less often believed to be at high risk for aggression compared to their male counterparts.

The interrater reliability of HCR-20 ratings was moderate to good, with total scores and final risk judgments performing generally better for both genders compared to subscale scores. In terms of predictive validity, only the final risk judgment was significantly related to violent outcome[7] for females (AUC = .86), whereas, all of subscale scores (AUCs = .83, .75 and .88 for the H, C and R subscales respectively), total score (AUC = .88) and final risk rating (AUC = .91) were predictive for men. There were no differences in predictive accuracy when inpatient aggression or post-discharge violent recidivism were compared for either of the male or female subsamples but the R subscale and total scores appeared to better predict violent recidivism in men than in women (i.e., z-statistic computed on AUCs). It was noted that the most frequent *other considerations* for men were *financial problems*, *lack of prospects for the future*, and *violent fantasies*, whereas for women there were *forming a new relationship*, *care for children*, and *prostitution*. The results from this study suggest that final risk judgments, which likely incorporate risk factors and considerations potentially more specific to female violence, should be considered when making clinical, forensic or correctional decisions with female populations rather than relying on the summation of risk factors. Although the HCR-20 assessment scheme is intended to be used in such a manner with assessees of both genders in order to generate integrated case formulations and not to overlook important idiographic variables, failing to do so may yield particularly erroneous management recommendations with women.

[7] The authors' definition of violence was narrower than that described in the HCR-20 manual since only physical assault or property destruction intended to threaten another person were coded.

In a follow-up pseudo-prospective study including some of the women from the above sample (N = 15) in addition to 30 female forensic patients from another Dutch hospital (Schaap et al., 2009), the HCR-20, its subscales, and the final risk judgments did not predict either violent and general recidivism.

Guy and Douglas (2006) used item-response theory to examine the differential functioning of the HCR-20 items across genders in a sample drawn from 3230 participants (88% men) from forensic psychiatric and correctional settings. The constructs underlying the HCR-20 items were four factors derived by Douglas (2006) via exploratory and confirmatory factor analyses in the same participants. In terms of their relevance to these constructs, the rank ordering of the HCR-20 items differed across genders. However, *plans lack feasibility* (R1), *unresponsive to treatment* (C5), *employment problems* (H4), and *noncompliance with remediation attempts* (R4) were among the most relevant items for both genders. Similarly, those items who were lowest in discrimination were comparable across genders (e.g., *impulsivity* [C4] and *substance use problems* [H5]).*Young age at first violent incident* (H2) and *unresponsive to treatment* (C5) were more discriminative in women, whereas *early maladjustment* (H8) was more discriminative in men. Men tended to endorse *negative attitudes* (C2) and *previous violence* (H1) more often, whereas *early maladjustment* (H8) was more often coded in women. However, overall, most items seemed to perform comparably across genders in terms of their ability to tap into the construct of violence risk.

de Vogel and de Vries Robbé (under review) adapted some of the items of the HCR-20 and created new ones in order to create the Female Additional Manual (FAM; de

Vogel, de Vries Robbé, van Kalmthout, & Place, 2010) which can be considered to represent pilot additional guidelines for use with women. The FAM was developed based on the literature, clinical interviews, and clinical expertise, as well as the results from a two-year pilot study. With regard to the Historical subscale, five new items were created: *Prostitution* (H11), *parenting difficulties* (H12), *pregnancy at young age* (H13), *Suicide attempt/self-harm* (H14), *victimization after childhood* (H15), and *sexuality* (H16). One new clinical item, *overt behavior* (C6), and two new risk management items, *responsibility for children* (R6) and *problematic intimate relationship* (R7), were added. The authors caution that for some of the new items no solid empirical evidence is available in the literature at present but emphasize that these were deemed relevant and important based on theoretical and clinical considerations. In addition, coding instructions were adapted for items H1, H3, H4, H7, H8, H9, H10, C1, C2, and C5 for use with women. For example, for *previous violence* (H1), influencing someone else to commit violence is coded under this item and for *lack of insight* (C1), the evaluators are made aware of the risk for underestimation , given that women tend to have better verbal skills than men.

The pilot study conducted with the adapted HCR-20 on 72 women from a forensic psychiatric facility in the Netherlands yielded good to excellent interrater reliability for the total scores (ICC = .90), final risk judgments (.83) and all the new items (ICC = .74 - .95) to the exception of *responsibility for children*. The total scores did not predict violent incidents during treatment but the final risk judgments did (AUC =.75). The new item *suicide attempt/self harm* (H14) also evidenced predictive validity (AUC = .69).

These results taken together yield a mixed picture of the validity of the HCR-20 assessment scheme across genders. Indeed, findings indicate both comparable and differing predictive accuracy for violent behavior in women as compared to their male counterparts. In the handful of empirical reports presented above, the violence of women appears to be somewhat better captured by HCR-20 assessments when these pertain to the forecasting of inpatient aggression as opposed to violence in the community. This is unsurprising since women's base rates of inpatient aggression tend to mirror or surpass those of men, and given that female violence is much more readily noticeable in institutions as compared to the community. It may also be the case that the exacerbation of certain psychiatric syndromes creates unique intra- and inter-individual contexts within which certain male and female behavioral expressions become more similar. However, the one study (de Vogel & de Ruiter, 2005) that followed the recommended protocol for the use of the HCR-20 and tested the predictive power of the final risk ratings rather than that of the numerical scores reported a moderate to large relationship between those summary judgments and community violence in women.

The present study aimed to expand the limited body of literature on the performance of the HCR-20 violence risk assessment scheme across genders by prospectively testing, for the first time, the reliability and validity of all of the main components of the instrument (including the summary risk ratings) in a sample of men and women from a short-term psychiatric setting. As part of these investigations, analyses were conducted to examine whether gender moderates the predictive validity of the assessment scheme. Only one of the studies presented above explicitly examined the

impact of gender on the relationship between the HCR-20 and violence through such moderation analyses.

Moreover, the relationship between the HCR-20 and other negative outcomes (i.e., violent victimization, self-harm and suicidal behaviors) were explored for both men and women. Indeed violence that is directed outwards evidences a considerable overlap in precipitating and predisposing variables with that which is inwardly oriented (e.g., Hillbrand, 2001; Nicholls, Brink, Desmarais, Webster, & Martin, 2006). For example, substance abuse, impulsivity, child abuse, personality disorders, cognitive distortions, stress, and inadequate coping strategies are common to individuals who tend to hurt themselves or aggress others. Similarly, violence perpetration and violent victimization seem to arise in the context of similar historical, personal, and situational risk factors. Moreover, self- and other-directed violent behaviors as well as victimization tend to both co-occur and also embody precipitating or predisposing events for one another (e.g., Evans, Marte, Betts, & Silliman, 2001). Finally, compared to the general population, psychiatric patients exhibit a greater likelihood not only to be violent towards others, but also to be victimized and/or to engage in self-harming or suicidal behaviors (e.g., Hillbrand, 2001). Researchers and clinicians in the field have come to recognize the importance of a multifaceted understanding of risk and risk overlaps. For example, the Short Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004), which was developed in order to assess multiple risk domains, has demonstrated great promise in pertinently informing the short-term management of mentally ill individuals (e.g., Nicholls et al., 2006). The risk factors evaluated by this structured professional judgment instrument mirror many of those

considered under the HCR-20 scheme. Moreover, in their adaptation of the HCR-20 for use with female populations, de Vogel and de Vries Robbé (under review) invite assessors to not only formulate summary judgments pertaining to the risk for violence but also for that of self-harm, victimization and non-violent offending. The authors comment that "although there is presently no empirical evidence supporting the assumption that the risk factors in the FAM (Female Additional Manual; de Vogel et al., 2010) are indeed related to these specific risks, at least the distinction between the different types of risk may be useful for clinical practice" (p. 14).

For these reasons, it seemed worthwhile to examine the presence of a potential association between the HCR-20 components and the aforementioned other negative outcomes in the present sample. If existing violence risk assessments schemes, such as the HCR-20, are shown to adequately forecast other destructive occurrences, interventions taking this knowledge into consideration are likely to maximize the efficiency of management efforts.

The Present Study and Research Hypotheses

This research project intended to: (a) compare the descriptive features of the HCR-20 across genders; (b) examine the interrater reliability of the scale; (c) evaluate the predictive validity of the instrument for violence in both men and women (i.e., subscale and total scores as well as final risk judgments); (d) investigate whether gender moderates the predictive validity of the HCR-20 for violence; (e) determine whether numerical and categorical ratings on the HCR-20 predict time to first violent event; (f) query into potential gender variations in the type, severity and victims of both forecasted

violence and actual violent outcomes; g) investigate the predictive acumen of the instrument across genders for other negative outcomes which share many of the same risk factors (i.e., suicidal behaviors, self-harm, and violent victimization).

It was hypothesized that: (a) significant gender differences in mean ratings would be observed on some of the items and subscales, as well as for the total scores, and more men would be characterized as *high risk* compared to women; (b) the HCR-20 would exhibit adequate interrater reliability (ICC = .80+); (c) the predictive validity of the instrument for violent outcomes would be equal to or greater than mean meta-analytic validity coefficients for existing risk assessment instruments (AUC = .70+) in terms of both total scores and final risk judgments across genders (although variations in predictive power at the subscale level were expected); (d) gender would not impact the predictive performance of the HCR-20 for violent outcomes; (e) the higher the numerical and categorical risk ratings the faster men and women would be shown to have engaged in new acts of violence after discharge; and (f) women would be rated as being more likely than men to both display milder forms of violence and direct their aggression towards close others, and these predictions will be accurate on average. Because of the dearth of research exploring the predictive validity of the HCR-20 across genders for negative outcomes other than violence, no firm hypothesis was formulated in that regard.

METHOD

Participants

The sample comprises 95 civil psychiatric inpatients from a hospital in the Lower Mainland Area of British Columbia. The participants were recruited both by institutional staff and by graduate students through the acute stay psychiatric ward of that hospital. The inclusion criteria for the individuals who wished to participate were as follows: Participants must be older than 19 years, fluent in English, planning to reside in the Greater Vancouver area for the next 6 months, and must have no known diagnosis of mental retardation. The sample was composed of 49 men and 46 women (descriptives are presented in Table 3 in the results section).

The majority of the sample was Caucasian (85.7% men, 80.4% women), never married (57.1% men, 32.6% women), living alone (supporting self; 36.7% men, 30.4% women) or with family (30.6% men, 37% women), unemployed (69.4% men, 67.4% women), and spoke English as their first language (83.7% men, 89.1% women). Approximately 14% of men and 39% of women were in a long-term relationship at admission and 20.4% of men as well as 54.3% of women had children. Most patients had been admitted involuntarily (73.5% men, 71.7% women) and diagnosed with mood (55.1% men, 73.9% women), psychotic (57.1% men, 45.7% women), and substance abuse/dependence (26.5% men, 19.6% women) disorders. Harm to self was the most

frequent reason for admission for both men (69.4%) and women (82.6%). Unsurprisingly, most individuals (83.7% men and 76.09% women) had been previously committed. On average, men and women had been hospitalized for psychiatric reasons five and three times respectively.

The vast majority of patients had no concern about their own risk for violence (73.5% men, 82.6% women). With regard to their current admission, 18.4 % of the male participants and 2.2% of their female counterparts had been violent (i.e., actual, attempted, or threatened physical harm other than spousal violence or stalking) and a minority of individuals had engaged in spousal violence (no men, 2.2% women) or stalking (2 % men, 2.2% women). A number of patients perpetrated physical (i.e., physical contact with body or objects) and non-physical aggression (i.e., threats or fear-inducing behaviors) during their index hospitalization (12.2% men and 4.3% women; 18.75% men and 2.17 % women respectively) but no incident resulted in serious harm.

At or just prior to admission, approximately half of the participants nourished suicidal ideation (44.9% men, 58.7% women), and some exhibited suicide attempts (4.1% men, 10.9% women), self-harm (2.1% men, 17.4% women), or violent victimization (8.2% men, 13% women). Half of the patients had a history of suicide attempt(s) (53.1% men, 56.5% women), some indicated a history of self-harm (29.2% men, 43.5% women), and a majority reported a prior violent victimization (63.3% men, 71.7% women). A considerable proportion of the sample experienced sexual abuse (16.3% men, 52.2% women), physical abuse (40.8% men, 65.2% women), and emotional abuse or neglect (36.7% men, 82.6%) during childhood.

Sample descriptives and the analyses that compared the HCR-20 scores across genders were based on the baseline data available for 95 participants (i.e., 49 men and 46 women). Because of attrition during the follow-up period, the analyses that tested the predictive validity of the HCR-20 vis-à-vis violent and other negative outcomes were based on a subsample of 73 (i.e., 35 men and 38 women) of the above participants. Of the 22 patients who did not yield any follow-up data, 54.5% could not be reached and 45.5 % did not wish to continue. Of the 28 participants who generated follow-up data but dropped out of the study before completing the last follow-up session (i.e., follow-up 5), 39.3% decided to withdraw from the study, 57.1% could not be reached, and one individual repeatedly failed to attend his scheduled interviews despite making appointments with his assigned research assistant.

Procedure

Data for this prospective study have been gathered as part of a larger research project entitled “The Risk Reduction Study” which involves the collection of extensive information pertaining to psychiatric inpatients over 6 time points. The participants were interviewed before their release into the community (i.e., baseline) and then, again, five times post-discharge (i.e., 5 follow-ups). Although attempts were made at conducting each of the five follow-ups every four weeks, the actual intervals varied due to scheduling challenges. Data were collected by graduate students in clinical psychology experienced (including the author) in conducting interviews and assessments with psychiatric populations. The interviews, which include information to rate pertinent standardized measures, were thorough (about 4 to 6 hours each). In addition, all of the

existing individuals' psychiatric files were reviewed at each of the 6 time points. The HCR-20 was coded and rated at baseline from the extensive material being collected.

Measure

The *Historical/Clinical/Risk Management-20 (HCR-20) Version 2* (Webster et al., 1997) evaluates 20 risk factors grouped in three temporal domains (Douglas et al., 2003). The 10 *H* items refer to *historical* aspects of violence risk and address such past events and static individual features as mental disorder and social adjustment. The five *C* items refer to *clinical* aspects of present psychological functioning of the individual. The latter items embody dynamic variables which are inferred from recent behavior and need to be re-assessed on a regular basis (Tengström, 2001). The five *R* items focus on potential future adjustment problems and on the management of *risk.* The evaluator not only examines the presence of the 20 risk factors, but also that of any other relevant case specific variables and, subsequently, formulates an opinion regarding the person's level of violence risk (low, moderate, high) (Hart et al., 2001). The clinician also makes recommendations regarding the resources and management strategies required to minimize the risk for future violent acts (detailed information pertaining to the instrument was reviewed earlier).

For the purpose of the present study, some additional variables were coded for each participant as part of the baseline HCR-20 assessment: The severity (i.e., *minor, moderate, severe*), target (i.e., *family/friend* or *stranger*), and nature of the forecasted violence (i.e., *weapon use, instrumental, hostile/reactive*), as well as a pilot expanded

version of the summary risk judgments (i.e., alternate risk judgments of *low, low/moderate, moderate, moderate/high, high, very high*).

Outcome Measures

Of the 95 patients interviewed at baseline, 73 generated outcome data (i.e., at least one follow-up contact), and 59, 51, 47 and 41 completed follow-ups 2, 3, 4, and 5 respectively. In other words, 43% of the participants finished all five follow-up sessions.

Violence was understood to represent any "actual, attempted, or threatened harm to a person or persons" following the definition from the HCR-20 manual (Webster et al., 1997, p. 24). Instances of violent behavior post-discharge were recorded after baseline interviews (i.e., over follow-up contacts 1 to 5 for a follow-up period varying from 1 to 5 months depending on the data available for each individual participant). Violence during the follow-up period was recorded both from file information (i.e., psychiatric files) and based on the monthly follow-up semi-structured interviews. Recent violence was assessed according to the nine categories of violence perpetration used in the MacArthur violence risk assessment study (Monahan et al., 2001) and additional violent behaviours such as threatening, stalking, and sexually violent behaviours were inquired about during the post discharge interviews. Additional information (i.e., once again pertaining to the nine categories of violence perpetration used in the MacArthur violence risk assessment study) from collateral interviews with family or friends was included to rate the occurrence of the various violent outcomes. These categories were collapsed for analyses into four dichotomous violence perpetration variables: "any violence," "verbal violence," "physical violence," and "sexual violence."

None of the participants reported having engaged in sexual violence during the follow-up period and, as a result, this category was dropped. The base rates for the violent outcomes considered for the 73 participants with follow-up data were as follows: (a) 26% any violence (*N* = 19; 29% of men and 24% of women); (b); 21% verbal violence (*N* = 15; 23% of men and 18% of women); and (c) 11% physical violence (*N* = 8; 11% of men and 11% of women).

Time to study end (re: *any violence)* was calculated by counting the days from the date of baseline assessment until the date of the first instance of violence for the patients who engaged in violence during the follow-up period (*N* = 19; 10 men, 9 women) and until the date of last interview contact for those patients who remained violence free (*N* = 54; 25 men, 29 women). It ranged from 11 to 288 days (*M* = 119.82, *SD* = 75.87). *Time to study end* (re: *physical violence)* was calculated by counting the days from the date of baseline assessment until the date of the first instance of physical violence for the patients who engaged in violence during the follow-up period (*N* = 8; 4 men and 4 women) and until the date of last interview contact for those patients who remained violence free (*N* = 64; 30 men and 34 women). It ranged from 24 to 163 days (*M* = 96.25, *SD* = 51.35). The data pertaining to one patient who was physically violent could not be included because no date was recorded for that occurrence. *Time to study end* (re: *verbal violence)* was calculated by counting the days from the date of baseline assessment until the date of the first instance of violence for the patients who engaged in violence during the follow-up period (*N* = 11; 6 men, 5 women) and until the date of last interview contact for those patients who remained violence free (*N* = 60; 27 men, 33 women). It ranged from 11 to

163 days (M = 50, SD = 43.162). The data pertaining to two patients who were verbally violent could not be included because no dates were recorded for these occurrences.

In order to contrast the *nature and severity* of violence post-discharge between genders, the victims' gender, victims' ages, victim-perpetrator relationships (i.e., *family, friend, acquaintance, stranger*), degree of inflicted physical harm (i.e., *minor, moderate, severe, death*), and use of weapon (i.e., *none, possession, threatened, used*) were compared across genders (i.e., rated according to the MacArthur study categories; Monahan et al., 2001). However, because not all post-discharge violent incidents were enquired about in such detail upon interview (e.g., only the specifics of the most serious acts of violence and of one instance of each type of those were recorded for each participant), 17 (i.e., 9 perpetrated by men and 8 perpetrated by women) out of 44 identified acts (i.e., 26 for men and 18 for women) yielded sufficient information for the aforementioned qualitative considerations.

Suicide/suicide attempts, *self-harm* and *violent victimization* were recorded during follow-up interviews and coded dichotomously (i.e., *presence during follow-up*: yes/no). Following the protocol from the MacArthur violence risk study (Monahan et al., 2001), participants were asked whether they ever attempted to hurt themselves since last contact with the interviewer. Then, if applicable, they were questioned with regard to method of injury, degree of harm sought, lethal intent, and other details surrounding the incident(s) in order to differentiate self-harm occurrences (e.g., cutting as a self-soothing strategy in the face of intense distress) from true suicide attempts (i.e., deliberate and life-threatening self-infliction of harm). Violent victimization was also assessed following the interview protocol from the MacArthur violence risk study. The various victimization experiences

queried about (e.g., having been shoved, forced into sexual activity, attacked with a knife, etc.) mirrored the questions pertaining to violence perpetration. The base rates of the above outcomes were as follows: 11% for suicide attempts (N = 8; 14% of men and 8% of women), 8% for self-harm incidents (N = 6; 9% of men and 8% of women), and 27% for violent victimization (N = 20; 20% of men and 34% of women).

Ethics Review

The larger research project which yielded data for the present research study was approved by the Simon Fraser University Research Ethics Board and by all the institutions where data collection took place. The present study was approved by the Simon Fraser University Ethics Board on October 14, 2010.

Data Analytic Methods

Descriptive Comparisons

In order to compare the means of two groups, the *t*-test is well-suited for use with numerical data, whereas, the chi-square test is preferred for use with categorical data (Howell, 2002). Two-tailed *t*-tests for independent means were used in order to investigate the presence/absence of significant gender differences on continuous variables such as HCR-20 total scores, whereas chi-square tests were conducted for the same purpose on categorical variables such as *summary risk ratings*.

Reliability

The reliability of a measure refers to the consistency with which the instrument generates scores over repeated applications (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). With regard to the prediction of violence, interrater reliability represents the most important type of reliability. Indeed, because violence risk assessment instruments do not measure single constructs but, rather, embody compilations of them, the issue of internal consistency is not as essential as that of interrater reliability (Douglas & Kropp, 2002). Intraclass correlations (ICCs) were used for interrater reliability analyses. ICCs embody measures of chance-corrected agreement and are, therefore, sensitive to additive and multiplicative biases between raters (Bartko & Carpenter, 1976).

Using mathematically equivalent indexes of reliability, Cicchetti and Sparrow (1981) defined intraclass correlation coefficient values of .40 and below as "poor," .40 to .59 as "fair," .60 to .74 as "good," .75 and above as "excellent". Correspondingly, Landis and Koch (1977) judged values of .00 and below as "poor,".00 to .20 as "slight," .21 to .40 "fair," .41 to .60 as "moderate," .61 to .80 as "substantial," and .81 to 1.00 as "almost perfect" in terms of the reliability of the associated measurements.

Predictive Validity

Point Biserial Correlations

In the domain of violence risk assessment, predictive validity, also called criterion-related validity, refers to the degree to which actual violence matches predicted future violence (Quinsey et al., 1998). The Pearson point-biserial correlation coefficient

(r_{pb}) estimates the degree of relationship between a dichotomous categorical scale and an interval or ratio scale (Brown, 1988). Such analyses were conducted in order to evaluate the bivariate relationship between post-release outcomes (e.g., "violent", "non-violent") and the HCR-20. The r_{pb} can range from -1 to + 1. The more the r_{pb} value departs from zero, the stronger the relationship (negative or positive) between the two variables under investigation. According to Cohen (1992), *r*s of .10, .30 and .50 represent small, moderate and large effect sizes respectively. However, Rice and Harris (1995) recommended that these guidelines, which premise a criterion base rate of .50, be modified when the latter differ from such value. They argued that at a base rate of .25 or .12, *r*s of .40 and .30 respectively reflect large effect sizes. Although, the interpretation of the results were done conservatively (i.e., according to Cohen's criterion), since the base rates of violent or other outcomes ranged from 8% to 34% depending on the (sub)sample and the outcome considered, the reported effect sizes may be understood to be larger than formally described.

Univariate Logistic Regression

In order to investigate the predictive validity of the HCR-20 for the three categories of violence (i.e., *any*, *physical*, and *verbal*), logistic regression analyses were conducted. Those analyses quantify the relationship between a dichotomous dependent variable (e.g. violence *yes/no*) as a function of one or more predictors (e.g., the HCR-20 and gender) (Cohen, Cohen, West & Aiken, 2003). They owe their popularity in the health sciences to their robustness and flexibility (Tabachnick & Fidell, 2001). Indeed, logistic regression implies no assumption about the distribution of its predictors (e.g.,

normal distributions, linear relationships, equal variance within groups) and its predictors may be discrete, continuous, dichotomous, or a mix.

ROC Analyses

Receiver Operating Characteristic (ROC) analyses were also used to test the predictive validity of the HCR-20. These analyses quantify the relationship between a continuous predictor variable (i.e., ratings on the HCR-20) and a dichotomous dependent outcome measure (i.e., the naturally occurring "violence" vs. "no violence" outcome criterion) (Tengström, 2001). "ROC" describes the prediction "characteristics" of the test and the "receiver" of the data can operate at any given point on the curve (Douglas et al., 1999). Therefore, the predictive performance of an instrument can be understood at all possible cut-offs in the test in terms of its sensitivity (true positive rate) and specificity (false positive rate). The hit rate (true positives) is plotted against the false-alarm rate (false positives) for all observed predictor values and the ROC curve graphs the tradeoffs in specificity that occurs as sensitivity increases and vice versa (Grann & Långström, 2007). Because ROC analyses have the advantage of being less dependent on the base rate of the criterion variable (i.e., violence) and the selection ratios compared to other statistical measures of predictive validity frequently used in psychology (e.g., 2 X 2 contingency tables identifying false positives and false negatives), they have been described as representing "a state of the art method for the estimation of predictive validity of a continuous risk measure" (Grann & Långström, 2007, p. 26). Indeed, since the base rate for violence is often lower than 50% in risk assessment research, the use of ROCs seems to yield better estimations of the predictive accuracy of the instruments

under investigation than traditional correlational techniques (Mossman, 1994). However, Guy (2008) noted that ROC analyses regrettably do not allow for considerations pertaining to the frequency, imminence and temporal rate of violence.

ROC analyses generate "areas under the curves" (AUCs) ranging from 0 (perfect negative prediction) to 1.00 (perfect positive prediction) with a value of 0.5 indicating a "no better than chance" prediction (Douglas et al., 1999; Tengström, 2001). A particular "area under the curve" depicts the probability that someone who is violent will actually score higher on the measure than a nonviolent person and therefore, summarizes overall discriminating power (Mossman, 1994). For example, an area of .80 means that there is an 80% chance that an actually violent individual will score above the cutoff for violence on the predictor and that an actually non-violent person will score below that cutoff. In light of both the lack of strong consensus in the field with regard to the interpretation of AUC and the financial and human costs associated with either false positive or false negatives, Sjöstedt and Grann (2002) proposed a conservative understanding of AUCs for predictive validity as follows: <.60 = low accuracy, .60 to .70 = marginal accuracy, .70 to .80 = modest accuracy, .80 to .90 =moderate accuracy and >.90 = high accuracy. Others have suggested a broader interpretation of AUC values, considering values below .70 to be small, those between .70 and .75 to be moderate, and those above .75 to be large (see e.g., Douglas et al., 2005; Guy, 2008).

Moderation Analyses

In order to examine whether gender moderates the predictive validity of the HCR-20 for the three categories of violence (i.e., *any*, *physical*, and *verbal*), multivariate

logistic regression analyses were conducted. In the present study, the *gender, HCR-20* and *gender-HCR-20* interaction predictors were entered sequentially in the model and a logistic regression was run in order to determine whether gender moderates (i.e., significant interaction term) the relationship between the HCR-20 components (i.e., subscales, total scores and final risk judgments) and violent outcomes at follow-up.

Survival Analyses

In order to determine whether summary numerical and categorical ratings on the HCR-20 predict time to first violent event across genders, survival analyses were conducted. These analyses model the time to an event, which can be defined "some type of qualitative change that can be determined as occurring at a relatively specific point in time" (Luke & Homan, 1998, p. 360). Some advantages inherent to these statistical procedures lie in their provision of control for censored data (i.e., cases for which the time of the event being studied is unknown) and individual *time at risk*. Both of these issues are of particular concern in the present study since some outcome data are missing due to attrition ($N = 22$) and because the follow-up period for the patients whose data are available varies in length (i.e., 1, 2, 3, 4, or 5 months).

The Cox Proportional-Hazards Model allows for the development of multivariate predictive models of hazard rates (i.e., prediction of time to violent event) from one or more categorical or continuous predictors) (Luke & Homan, 1998; Tabachnick & Fidell, 2001). These statistical analyses model event rates (e.g., violence) as a log-linear function of predictors called covariates (i.e., time at risk, gender and HCR-20 summary risk ratings) and generate regression coefficients which provide the relative effect (i.e.,

magnitude of effect) of each covariate on that function. This represents a semi-parametric procedure since the effects of the covariates on the hazard function are modeled rather than the shape of that function.

RESULTS

First, background sociodemographic, psychiatric and criminal characteristics of the sample as well as the baseline HCR-20 ratings between male and female patients are compared. Then, the main predictive validity results are presented, with figures pertaining to all participants reported initially in order to contextualize the findings across genders that are described subsequently.

Descriptive Analyses

Sociodemographic, Clinical and Criminal Background

A number of background gender differences emerged in this sample of psychiatric inpatients short-term (see Table 2). Women were significantly more likely than men to be in a long-term relationship and to have children. Moreover, self-harm was more often relevant to their current admission and they reported more sexual, physical, and emotional abuse or neglect during childhood compared to their male counterparts.

On the other hand, more men were currently homeless or ever had no fixed address. They displayed more adult arrests, charges or convictions, and were more likely to have been incarcerated in the past, have breached parole or probation, robbed someone, possessed or sold drugs, committed fraud, and assaulted someone as adults. In addition, violence (i.e., not spousal violence or stalking) was more often relevant to their

current admission and they exhibited more non-physical aggression during their index hospitalization).

No significant statistical gender differences were noted with regard to major diagnostic categories. Because specific diagnoses were not always adequately defined, more detailed statistical comparisons were not conducted. However, men appeared to be diagnosed most often with bipolar and/or substance abuse/dependence disorders, whereas women seemed to present most often with depressive and/or psychotic disorders (see Table 2.). Axis II diagnoses were more prevalent in women (N = 14) as compared to men (N = 7), and more specifically, borderline personality disorder/traits.

Table 2. Background Comparisons by Gender

	Men (N = 49)	**Women (N = 46)**	**t-test/ χ^2**	**Sig. (2-tailed)**
Age	33.94 (SD=9.23)	37.37 (SD=11.70)	1.59	.12
Ever in a long-term relationship	77.6%	87%	1.43	.23
Currently in a long-term relationship	14.3%	39.1%	7.56	.01
Education (years)	12.11 (SD=2.52)	12.98 (SD=2.09)	1.82	.07
First language: English	85.7%	89.1%	.25	.62
Most frequent most recent diagnosis:				
Axis I:				
Mood disorder	55.1%	73.9%	3.65	.06
Bipolar spectrum	30.1%	26.7%		
Depression spectrum	18.4%	51.1%		
Anxiety spectrum	12.2%	11.1%		
Psychotic disorder	57.1%	45.7%	1.25	.26
Schizophrenia	18.4%	8.9%		
Other Psychosis	16.3%	26.7%		
Schizoaffective	16.3%	15.6%		
Substance abuse/dependence	26.5%	19.6%	.65	.42
Axis II:				
Cluster B PD or traits	8.2%	22.2%	3.66	.06
Cluster C PD or traits	2%	8.9%	2.18	.14
Other PD or PD NOS	2%	4.4%	.44	.51
Involuntary admission status	73.5%	71.7%	.04	.85
Admission reason:				
Harm to Others	18.4%	17.4%	.02	.90
Harm to Self	69.4%	82.6%	2.26	.13
Psychiatric Deterioration	67.3%	60.9%	.43	.51

	Men (*N* = 49)	Women (*N* = 46)	t-test/ χ^2	Sig. (2-tailed)
At admission:				
Drugs and/or alcohol	30.6%	19.6%	2.64	.27
Delusions	55.1%	45.7%	.85	.36
Hallucinations	40.8%	37%	.15	.70
Agitation	40.8%	39.1%	.03	.87
Confusion/Disorientation	30.6%	15.2%	3.16	.08
Hostility/Anger/Aggression	24.5%	18.6%	.33	.56
Manic Symptoms	32.7%	26.1%	.49	.48
Currently has no fixed address	10.2%	0%	4.96	.03
Ever had no fixed address	42.8%	15.2%	.87	.00
Unemployed at admission	69.4%	67.4%	.04	.83
Have any children	20.4%	54.3%	12.24	.00
Previous:				
Commitment Decision	83.7%	76.1%	.84	.36
Incompetence Decision	2%	0%	1.92	.38
NCRMD	4.1%	0%	1.92	.17
Number of prior psychiatric hospitalizations	4.73 (*SD*=14.12)	2.70 (*SD*=2.61)	-.96	.34
Arrest/charge/conviction:				
As an adult	44.9%	15.2%	9.86	.00
As a youth	22.4%	13%	1.43	.23
As an adult:				
Breached parole/probation	12.8% (*N* = 47)	0% (*N* = 46)	6.28	.01
Stole something worth $50+	23.4% (*N* = 47)	8.7% (*N* = 46)	3.72	.05
Did break and enter	14.9% (*N* = 47)	4.3% (*N* = 46)	2.96	.09
Robbed someone	12.8% (*N* = 47)	0% (*N* = 46)	6.15	.01
Possessed or sold drugs	72.3% (*N* = 47)	37% (*N* = 46)	11.75	.00
Committed fraud	26.7% (*N* = 45)	8.7% (*N* = 46)	5.07	.02
Assaulted someone	55.3% (*N* = 47)	19.6% (*N* = 46)	12.66	.00
Spousal violence			1.62	.45
Current admission	0%	2.2%		.
Ever	14.3%	19.6%		
Sexual Violence			.00	.95
Current admission	0%	0%		
Ever	4.1%	4.3%		
Stalking			2.46	.29
Current admission	2%	2.2%		
Ever	28.6%	23.9%		
Other Violence			13.69	.00
Current admission	18.4%	2.2%		
Ever	42.9%	23.9%		
No concern about own violence risk	73.5% (*N* = 48)	82.6% (*N* = 46)	1.18	.55

	Men (*N* = 49)	Women (*N* = 46)	t-test/ χ^2	Sig. (2-tailed)
Physical aggression during index hospitalization	12.2% (*N* = 48)	4.3% (*N* = 46)	2.01	.16
Resulted in serious harm	0%	0%		
Non-physical aggression during index hospitalization	18.8% (*N* = 48)	2.17 (*N* = 46)	6.79	.01
Current admission:				
Suicidal ideation	44.9%	58.7%	1.81	.18
Suicide attempt	4.1%	10.9%	1.60	.21
Self-harm	2.1% (*N* = 48)	17.4% (*N* = 46)	6.36	.01
Violent victimization	8.2%	13%	1.41	.47
Suicide attempt history	53.1%	56.5%	1.00	.61
Self-harm history	29.2%	43.5%	2.85	.24
Violent victimization history	63.3%	71.7%	3.54	.32
Raised by both biological parents until 16 years old	61.2%	63%	.03	.86
Sexual abuse during childhood	16.3%	52.2%	13.65	.00
Physical abuse during childhood	40.8%	65.2%	6.18	.05
Emotional abuse/neglect during Childhood	36.7%	82.6%	20.81	.00

Note. *N* = 95 (i.e., 49 men and 46 women, unless specified).

HCR-20 across Genders

Baseline ratings on the HCR-20 were contrasted across genders and detailed in Table 3. The scores on the historical subscale ranged from 3 to 20 for men and from 2 to 15 for women, those on the clinical subscale ranged from 1 to 10 for men and from 0 to 9 for women, and those on the risk management subscale ranged from 1 to 9 for men and 1 to 8 for women. The total scores varied between 7 to 37 for men and 7 to 29 for women. Most individuals of both genders were rated as *low risk* and no female participant was deemed to be at high risk for violence. On an exploratory expanded final risk rating scale comprising six possible summary risk categories (i.e., *low, low-moderate, moderate, moderate-high, high, very-high*) instead of the usual *low, moderate,* or *high* ratings, the majority of men were considered to be at a *low-moderate* level of risk and over half of

the women were classified in the *low* risk category. No woman was believed to be at *high* or *very-high* risk for aggression.

In order to examine potential differences in HCR-20 components across genders, *t*-tests and Chi-Square tests were conducted (see Table 3). With regard to historical and more static items, men exhibited more frequent or severe previous violence, committed violent acts at a younger age, evidenced more substance use problems, were rated as more psychopathic, and had more prior supervision failures on average compared to women. No significant gender differences emerged on the clinical subscale, but men seemed to lack insight and display negative attitudes marginally more so than women did. As pertains to risk management items, men were assessed to be more likely to lack personal support and to be noncompliant with remediation attempts, whereas women were believed to be more likely to find themselves in stressful circumstances in the near future. Overall, men obtained greater historical as well as total HCR-20 scores and were more often classified as high risk as well as less often classified as low risk compared to their female counterparts (i.e., on both the three- and the expanded six-category summary rating scales). In addition, female patients exhibited somewhat restricted ranges of scores on the historical subscale and total scores compared to those of their male counterparts and considerably less variation in their summary risk ratings.

Table 3. HCR-20: Comparative t-tests and Chi-Square Tests across Genders

Item		Men Mean	SD	Women Mean	SD	ES (*d*)	Sig.
H1	**Previous Violence**	**1.12**	.78	**.52**	.72	**.81**	**.00**
	0	24.5%		60.9%			
	1	38.8%		26.1%			
	2	36.7%		13%			
H2	**Young Age at 1st Violence**	**1.10**	.90	**.67**	.90	**.48**	**.02**
	0	34.7%		60.9%			
	1	20.4%		10.9%			
	2	44.9%		28.3%			
H3	**Relationship Instability**	**1.06**	.78	**1.31**[a]	.79[a]	.04	.87
	0	26.5%		27.4%			
	1	40.8%		36.3%			
	2	32.7%		36.3%			
H4	**Employment Problems**	**1.16**	.80	**1.13**	.72	.04	.83
	0	24.5%		19.6%			
	1	34.7%		47.8%			
	2	40.8%		32.6%			
H5	**Substance Use Problems**	**1.47**	.79	**.80**	.89	**.81**	**.00**
	0	18.4%		50%			
	1	16.3%		19.6%			
	2	65.3%		30.4%			
H6	**Major Mental Illness**	**1.92**	.34	**1.83**	.44	.25	.26
	0	2%		2.2%			
	1	4.1%		13%			
	2	93.9%		84.8%			
H7	**Psychopathy**	**.33**	.63	**.09**	.35	**.47**	**.02**
	0	75.5%		93.5%			
	1	16.3%		4.3%			
	2	8.2%		2.2%			
H8	**Early Maladjustment**	**1.14**	.82	**1.17**	.71	.04	.84
	0	26.5%		17.4%			
	1	32.7%		47.8%			
	2	40.8%		34.8%			
H9	**Personality Disorder**	**.33**	.56	**.48**	.72	.25	.26
	0	71.4%		65.2%			
	1	24.5%		21.7%			
	2	4.1%		13%			
H10	**Prior Supervision Failure**	**.61**	.76	**.33**	.60	**.42**	**.04**
	0	55.1%		73.9%			
	1	28.6%		19.6%			
	2	16.3%		6.5%			
C1	**Lack of Insight**	**1.04**	.84	**.72**	.81	.40	.06
	0	32.7%		50%			
	1	30.6%		28.3%			
	2	36.7%		21.7%			
C2	**Negative Attitudes**	**.47**	.62	**.26**	.49	.38	.07
	0	59.2%		76.1%			
	1	34.7%		21.7%			
	2	6.1%		2.1%			

Item		Men Mean	SD	Women Mean	SD	ES (d)	Sig.
C3	**Active Symptoms of MMI**	**1.59**	.61	**1.63**	.57	.07	.75
	0	6.1%		4.3%			
	1	28.6%		28.3%			
	2	65.3%		67.4%			
C4	**Impulsivity**	**.82**	.64	**.78**	.79	.05	.82
	0	30.6%		43.5%			
	1	57.1%		34.8%			
	2	12.2%		21.7%			
C5	**Unresponsive to Treatment**	**.73**	.70	**.67**	.63	.09	.66
	0	40.8%		41.3%			
	1	44.9%		50%			
	2	14.3%		8.7%			
R1	**Plans Lack Feasability**	**.55**	.71	**.65**	.77	.14	.51
	0	57.1%		52.2%			
	1	30.6%		30.4%			
	2	12.2%		17.4%			
R2	**Exposure to Destabilizers**	**.94**	.66	**.87**	.72	.10	.63
	0	24.5%		32.6%			
	1	57.1%		47.8%			
	2	18.4%		19.6%			
R3	**Lack of Personal Support**	**.94**	.75	**.54**	.66	.57	.01
	0	30.6%		54.3%			
	1	44.9%		37%			
	2	24.5%		8.7%			
R4	**Noncompliance**	**.90**	.65	**.57**	.62	.53	.01
	0	26.5%		50%			
	1	57.1%		43.5%			
	2	16.3%		6.5%			
R5	**Stress**	**1.12**	.70	**1.50**	.66	.56	.01
	0	50%		8.7%			
	1	43.5%		32.6%			
	2	6.5%		58.7%			
Historical	**Total out of 20**	**10.24**		**8.07**		**.58**	**.01**
	range	3 to 20		2 to 15			
	(modes)	(10, 13)		(7)			
Clinical	**Total out of 10**	**4.65**	2.17	**4.07**	1.91	.29	.17
	range	1 to 10		0 to 9			
	(modes)	(6)		(3)			
Risk Management	**Total out of 10**	**4.45**	1.83	**4.13**	1.97	.17	.42
	range	1 to 9		1 to 8			
	(modes)	(3)		(3)			
HCR-20	**TOTAL score out of 40**	**19.45**	6.73	**15.97**	5.57	**.57**	**.01**[b]
	range	7 to 37		7 to 29			
	(modes)	(18, 22)		(14, 17)			

		Men	Women	χ^2	Sig.
Summary Rating	**Final Risk Judgments**			6.05	**.05**
	Low	61.2%	71.7%		
	Moderate	26.5%	28.3%		
	High	12.2%	-		
Alternate Rating	**Alternate Final Risk Judgments**			15.75	**.01**
	Low				
	Low-Moderate	24.5%	58.7%		
	Moderate	42.9%	28.9%		
	Moderate-High	18.4%	13%		
	High	2%	4.3%		
	Very High	8.2%	-		
		4.1%	-		

Note. N = 95 (men = 49, women = 46); 2-tailed (unless specified);
[a] N = 45; [b]1-tailed; [c]N = 94; p = .05

Interrater Reliability

With regard to risk assessment instruments, the most crucial type of reliability pertains to the consistency with which raters conduct the assessments and generate individual violence risk decisions. For the purpose of this study, the HCR-20 was independently coded by two different raters for 21 of the 95 cases in order to statistically measure the degree of agreement between assessors (i.e., 16 raters and 21 interrater comparisons in total). The historical subscale and alternate final risk judgments yielded excellent/substantial interrater reliability (ICC_1 = .79 and .78 respectively; see Table 4). The clinical subscale, total scores and original final risk judgments generated good/substantial coefficients (ICC_1 = .65, .74, and .68 respectively), whereas the risk management subscale evidenced fair/moderate interrater reliability (ICC_1 = .43). The overall adequacy of these results allowed for predictive validity analyses to be conducted.

Table 4. HCR-20 Interrater Reliability

	ICC_1	ICC_2
H	.79	.88
C	.65	.79
R	.43	.60
HCR-20	.74	.85
Final Risk Judgments	.68	.81
Alternate Final Risk Judgments	.78	.88

Note. *N*=21 (10 women; 11 men).

Predictive Validity

The predictive validity of the HCR-20 for violent outcomes is described first for all participants in order to contextualize the predictive findings across genders that are presented next. The validity of the scale was examined using four types of analyses: Correlations, Logistic Regressions, Receiver Operating Characteristic, and Survival analyses. Correlation coefficients allow for useful comparisons with effects reported in the literature. However, although relatively robust, they may be biased due to factors such as assumptions violations (e.g., normality) or inadequate base rates of outcomes (i.e., < 50%), which raised some degree of concern in the present data as is often the case for evaluations of risk assessment measures. For these reasons, more robust and less assumptions-tied procedures such as logistic regression and Receiver Operating Characteristic were conducted in order to ascertain predictive findings. Additionally, the ability of the HCR-20 to predict time to violence during the follow-up period was investigating using Cox regression analyses. Moreover, regression analyses (i.e., both logistic regression and Cox regression) provided useful statistical models from which to test potential moderation effects (i.e., the impact of gender).

Point Biserial Correlations among Violent Outcomes and between the HCR-20 and Violent Outcomes on the Whole Sample

In this sample of short-term psychiatric inpatients, all violent outcomes were intercorrelated to a moderate to very large degree according to Cohen's guidelines (1992) (Table 5).

Table 5. Intercorrelations (r_{pb}) between Violent Outcomes for the Whole Sample

	Any Violence	Physical Violence	Verbal Violence
Any Violence	-	.59**	.86**
Physical Violence	.59**	-	.37**
Verbal Violence	.86**	.37**	-

Note. N = 73. 1-tailed; **Significant at .01.

Overall, as expected, the HCR-20 and its components were associated with all violent outcomes (i.e., *any violence*, *physical violence*, and *verbal violence*) to a moderate degree (see Table 6). The relationship between the alternate final risk judgments evidenced the strongest effect size (r = .38, p = .01). However, the clinical subscale did not demonstrate an association with verbal violence and the final risk judgments (i.e., both rating scales) did not correlate with any violence.

Table 6. Correlations (r_{pb}): HCR-20 and Violent Outcomes for the Whole Sample

	Any Violence	Physical Violence	Verbal Violence
H total	.30**	.32**	.34**
C total	.27*	.25*	.17
R total	.26*	.22*	.32**
HCR-20 total	.30**	.37**	.30**
Final Risk Judgments	.16	.30**	.20*
Alternate Final Risk Judgments	.26*	.24*	.38**

Note. N = 73. 1-tailed. **Significant at .01; *Significant at .05.

Logistic Regression:
Predictive Validity of the HCR-20 for Violent Outcomes in the Whole Sample

Logistic regressions were conducted in order to test the predictive validity of the HCR-20 for the three violent outcomes (any violence, physical violence and verbal violence). Table 7 presents a series of logistic regressions with any violence as the dependent variable. The H, C, R, and HCR scores as well as the alternate final risk judgments produced overall significant models. Only the final risk judgments did not significantly predict violence in this sample. The odds of acting out violently increased by approximately 21%, 36%, 38% and 13% for every unit increase in the H, C, R and the total scores respectively. They increased by 64% for every categorical increase in severity of alternate summary risk rating (e.g., from *low* to *low/moderate*).

Table 7. Logistic Regression: HCR-20 and Any Violence

	B	Wald	sig	e^b	Model
H	.191	6.111**	.01	1.210	χ^2= 6.724**, p = .01 R^2 = .130
C	.305	4.817**	.01	1.356	χ^2= 5.197*, p = .02 R^2 = .101
R	.320	4.736*	.02	1.377	χ^2= 5.010* , p = .03 R^2 = .097
HCR	.118	5.905**	.01	1.125	χ^2= 6.685**, p = .01 R^2 = .128
FRJ	.573	1.769	.09	1.774	χ^2= 1.735, p = .19 R^2 = .034
FRJ II	.493	4.262*			χ^2= 4.587*, p = .03 R^2 = .089

Note. N = 73; 1-tailed; R^2= Nagelkerke R Square; *p = .05, **p = .01.

Table 8 presents a series of logistic regressions with verbal violence as the dependent variable. With the exception of the C subscale and the final risk judgements, all scores as well as the alternate final risk judgments produced overall significant

models. The odds of a patient becoming verbally aggressive increased by approximately 26%, 53%, and 13% for every unit increase in the H, R and the total scores respectively. They were multiplied by 1.601 for every categorical increase in the severity of the alternate summary risk ratings.

Table 8. Logistic Regression: HCR-20 and Verbal Violence

	B	Wald	sig.	e^b	Model
H	.232	7.277**	.01	1.261	χ^2= 8.352**, p = .00 R^2 = .169
C	.202	2.025	.08	1.224	χ^2= 2.053, p = .15 R^2 = .043
R	.424	6.640**	.01	1.528	χ^2= 7.351** , p = .01 R^2 = .150
HCR	.125	5.771**	.01	1.133	χ^2= 6.531*, p = .01 R^2 = .134
FRJ	.752	2.737*	.05	2.121	χ^2= 2.669, p = .10 R^2 = .056
FRJ II	.471	3.787*[a]	.03	1.601	χ^2= 3.869*, p = .05 R^2 = .081

Note. N = 73; 1-tailed; R^2= Nagelkerke R Square; *p = .05, **p = .01,[a]sig =.052.

Table 9 presents a series of logistic regressions with physical violence as the dependent variable. With the exception of the risk management subscale, all scores and both versions of the final risk judgments produced overall significant models. The odds of a patient engaging in physical violence increased by approximately 32%, 46%, and 22% for every unit increase in the H, C, and the total scores respectively. For every categorical increase in the severity of summary risk ratings or alternate such risk ratings the odds of becoming physically aggressive were multiplied by 3.705 and 2.213 respectively.

Table 9. Logistic Regression: HCR-20 and Physical Violence

	B	Wald	sig	e^b	Model
H	.277	6.182**	.01	1.319	χ^2= 7.258**, p = .01 R^2 = .190
C	.377	4.034*	.03	1.457	χ^2= 4.301*, p = .04 R^2 = .115
R	.351	3.144*	.04	1.421	χ^2= 3.256 , p = .07 R^2 = .087
HCR	.199	7.202**	.00	1.220	χ^2= 9.506**, p = .00 R^2 = .245
FRJ	1.310	5.494**	.01	3.705	χ^2= 5.527*, p = .02 R^2 = .146
FRJ II	.794	7.143**	.00	2.213	χ^2= 7.886**, p = .01 R^2 = .205

Note. N = 73; 1-tailed; R^2= Nagelkerke R Square; *p = .05, **p = .01.

ROC Analyses: The HCR-20 and Violent Outcomes in the Whole Sample

With ROC analyses, all subscales predicted any violence to a modest degree (Table 10). The Historical subscale demonstrated a moderate predictive ability in terms of both subtypes of violence, whereas this was only true for verbal violence and the Risk Management subscale. The Clinical subscale yielded marginally significant results for any violence and non-significant AUCs when types of violent outcomes were considered individually. The total scores evidenced moderate predictive power for physical as well as any violence, and marginally significant results of modest magnitude with regard to verbal violence. Both versions of the final risk judgments failed to demonstrate statistically significant predictive associations with any of the outcomes under consideration.

Table 10. ROC Analyses: The HCR-20 and Violent Outcomes in the Whole Sample

	Any Violence		Physical Violence		Verbal Violence	
	AUC	**SE**	**AUC**	**SE**	**AUC**	**SE**
H	.69*	.07	.72*	.11	.72*	.07
C	.65[a]	.08	.66	.13	.60	.09
R	.66*	.08	.66	.12	.70*	.08
HCR	.66*	.08	.75*	.11	.66[c]	.09
FRJ	.58	.08	.70[d]	.11	.61	.09
FRJ-II	.64[b]	.07	.70	.12	.64	.08

Note. *Significant at .05; [a]p = .054; [b]p = .066; [c]p = .056; [d]p = .068.

Logistic Regression: Moderation Analyses

After having examined the performance of the HCR-20 over the entire sample, series of sequential logistic regressions were carried out in order to investigate whether gender moderates the observed relationship between the scale and violent outcomes (Tables 11, 12 and 13). The HCR-20 components and *gender* were entered into the first block. In block 1, with regard to *any violence,* the models including the H subscale scores and the total scores were both significant. In terms of verbal violence, the H and R subscales as well as the final risk judgments yielded significant statistical models. As pertains to physical violence, the models inherent to the Historical subscale, the total scores, and the alternate final risk judgments yielded significant statistical fit to the data.

In addition to the HCR-20 components and *gender,* the interaction between those two variables was entered in the second block. In block 2, no interaction term reached significance for any of the HCR-20 predictors with any of the three violent outcomes, indicating an absence of moderation effects in this sample. With regard to *any violence,* no model represented a significant statistical fit to the data. In terms of verbal violence,

the models inherent to the H, R, and HCR predictors were significant. The Historical subscale, the total scores and the alternate final risk judgments yielded significant models as pertains to physical violence.

Table 11. Moderation Analyses: Any Violence

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.196	6.016**	1.217	.184	2.663	1.202	**1.** χ^2= 6.827*, p = .03
Gender	-.156	.072	.855	-.183	.090	.833	R^2 = .131
H*G				.025	.024	1.025	**2.** χ^2= 6.851, p = .08
							R^2 = .131
C	.307	4.582*	1.359	.339	2.608	1.404	**1.** χ^2= 5.200, p = .07
Gender	-.031	.003	.970	-.002	.000	.998	R^2 = .101
C*G				-.061	.046	.941	**2.** χ^2= 5.246, p = .16
							R^2 = .102
R	.317	4.587*	1.373	.173	.776	1.189	**1.** χ^2= 5.046, p =.08
Gender	.106	.036	1.112	-.049	.007	.952	R^2 = .098
R*G				.326	1.124	1.386	**2.** χ^2= 6.211, p = .10
							R^2 = .120
HCR	.125	5.896*	1.133	.092	1.599	.704	**1.**χ^2= 6.861*, p = .03
Gender	-.248	.174	.780	-.350	.306	1.096	R^2 = 131
HCR*G				.066	.390	1.068	**2.** χ^2= 7.260, p = .06
							R^2 = .139
FRJ	.553	1.574	1.739	.321	.378	1.379	**1.** χ^2= 1.781, p =.41
Gender	-.118	.045	.889	-.444	.422	.642	R^2 = .035
FRJ*G				.799	.684	2.224	**2.** χ^2= 2.464, p = .48
							R^2 = .049
FRJ II	.507	4.110*	1.661	.414	2.031	1.512	**1.** χ^2= 4.624, p =.10
Gender	.111	.037	1.118	-.195	.061	.823	R^2 = .090
FRJII*G				.309	.328	1.361	**2.** χ^2= 4.953, p = .18
							R^2 = .096

Note. N = 73; R^2= Nagelkerke R Square; *p = .05.

Table 12. Moderation Analyses: Verbal Violence

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.240	7.266*	1.271	.203	2.719	1.225	**1.** χ^2= 8.477*, p = .01 R^2 = .172
Gender	-.227	.125	.797	-.350	.239	.704	
H*G				.077	.182	1.080	**2.** χ^2= 8.661*, p = .03 R^2 = .175
C	.197	1.826	1.218	.179	.684	1.196	**1.** χ^2= 2.075, p = .35 R^2 = .044
Gender	.089	.022	1.093	.076	.015	1.079	
C*G				.033	.013	1.034	**2.** χ^2= 2.088, p = .55 R^2 = .044
R	.422	6.514*	1.525	.202	.882	1.223	**1.** χ^2= 7.374*, p =.03 R^2 = .151
Gender	.093	.023	1.098	-.312	.191	.732	
R*G				.524	2.099	1.688	**2.** χ^2= 9.668*, p = .02 R^2 = .194
HCR	.132	5.771*	1.141	.070	.826	1.073	**1.**χ^2= 6.704*, p = .04 R^2 = 138
Gender	-.268	.172	.765	-.534	.519	.586	
HCR*G				.126	1.146	1.135	**2.** χ^2= 7.921*, p = .05 R^2 = .161
FRJ	.738	2.498	2.093	.658	1.437	1.931	**1.** χ^2= 2.684, p =.26 R^2 = .057
Gender	-.074	.015	.928	-.202	.070	.817	
FRJ*G				.286	.077	1.332	**2.** χ^2= 2.760, p = .43 R^2 = .058
FRJ II	.483	3.590	1.620	.506	2.765	1.658	**1.** χ^2= 3.892, p =.14 R^2 = .081
Gender	0.95	.023	1.099	.176	.042	1.193	
FRJII*G				-.080	.020	.923	**2.** χ^2= 3.912, p = .27 R^2 = .082

Note. N = 73; R^2= Nagelkerke R Square; *p = .05.

Table 13. Moderation Analyses: Physical Violence

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.298	6.572*	1.347	.157	1.133	1.170	**1.** χ^2= 7.713*, p = .02 R^2 = .201
Gender	-.562	.449	.570	-1.647	1.497	.193	
H*G				.354	1.763	1.425	**2.** χ^2= 9.732*, p = .02 R^2 = .250
C	.393	4.150*	1.482	.357	1.855	1.429	**1.** χ^2= 4.441, p = .11 R^2 = .118
Gender	-.299	.139	.742	-.396	.176	.673	
C*G				.081	.042	1.084	**2.** χ^2= 4.483, p = .21 R^2 = .119
R	.353	3.154	1.424	.089	.111	1.094	**1.** χ^2= 3.267, p =.20 R^2 = .088
Gender	-.080	.011	.924	-.693	.481	.500	
R*G				.604	1.949	1.829	**2.** χ^2= 5.398, p = .15 R^2 = .143
HCR	.224	8.009**	1.251	.162	2.426	1.175	**1.** χ^2= 10.539**, p = .01 R^2 = 269
Gender	-.893	1.000	.409	-1.589	1.414	.204	
HCR*G				.144	.670	1.155	**2.** χ^2= 11.277**, p = .01 R^2 = .287
FRJ	1.421	5.461*	4.142	1.536	4.160*	4.646	**1.** χ^2= 5.814, p =.06 R^2 = .153
Gender	.453	.282	1.573	.691	.312	1.996	
FRJ*G				-.357	.074	.700	**2.** χ^2= 5.888, p = .12 R^2 = .155
FRJ II	.901	7.715**	2.461	1.068	6.116*	2.911	**1.** χ^2= 8.701*, p =.01 R^2 = .225
Gender	.804	.779	2.234	1.527	.1.153	4.603	
FRJII*G				-.498	.480	.608	**2.** χ^2= 9.194*, p = .03 R^2 = .237

Note. N = 73; R^2= Nagelkerke R Square; *p = .05, **p = .01.

Survival Analyses

Another way to examine the predictive validity of the HCR-20 for violence is to model the relationship between the instrument and the timing of violent outcomes while controlling for uneven follow-up periods. Results of survival analyses conducted both with and without measuring the impact of gender on the performance of the scale are presented below.

Survival Analyses for Any Violence

Cox regression analyses were conducted to examine the relationship the HCR-20 components and time to first violent incident. Of the 73 patients with available follow-up data, 19 engaged in violence (i.e., 54 right-censored).

The Historical subscale predicted time to first violent incident (β = .154, Wald = 6.332, p =.01; $\chi 2$ = 6.640, p =.01). The hazard ratio (e^b = 1.166) indicated that each unit increase on this subscale increased the daily hazard of violent acting out by about 17%. With regard to the Clinical subscale, ratings predicted time to first violent incident (β = .226. Wald = 4.276, p = .02; $\chi 2$ = 4.357, p =.02) with each unit increase on the subscale augmenting the daily hazard of violent acting out by about 25% (e^b = 1.254). Risk Management items predicted time to first violent incident (β = .294. Wald = 6.285, p = 01; $\chi 2$ = 6.579, p =.01) with each unit increase on that subscale raising the daily hazard of violent acting out by about 34% (e^b = 1.341). HCR-20 total scores predicted time to first violent incident (β = .091. Wald = 6.816, p = 01; $\chi 2$ = 6.830, p =.01). The hazard ratio (e^b = 1.096) indicated that each unit increase on the overall HCR-20 increased the daily hazard of violent acting out by about 10%.

In terms of final risk judgments, the *low*, *moderate* and *high risk* groups did not appear to significantly differ in their survival probabilities either when considered on a continuum (β = .410, Wald = 1.462, p = .23; $\chi 2$ = 1.492, p =.22) or categorically ($\chi 2$ = 1.527, p =.47). Conversely, the alternate final risk judgments yielded overall significant models when treated both categorically ($\chi 2$ = 13.336, p =.02) or as a continuous variable (β =.344 Wald = 4.365, p = .04; $\chi 2$ = 4.519, p =.03). Both versions of the final risk judgments were examined from a continuous standpoint in order to minimize the bias

generated by the low number of cases for some of the categories (e.g. $N = 1$ for the alternate summary risk rating of *high*).

Survival Analyses for Physical Violence

The Historical subscale predicted time to first physically violent incident (β = .269, Wald = 7.804, p =.01; $\chi 2$ = 9.088, p =.00) with a hazard ratio (e^b = 1.309) indicating that each unit increase on this subscale raised the daily hazard of physically violent acting out by about 31%. The Clinical subscale did not predict time to first physical violence (β = .266. Wald = 2.830, p = .10; $\chi 2$ = 2.896, p =.09). The Risk Management items predicted time to first physical violence (β = .380. Wald = 5.169, p = .02; $\chi 2$ = 5.661, p =.02) with each unit increase on that subscale augmenting the daily hazard of physical violence by about 46% (e^b = 1.462). The HCR-20 total scores also predicted time to first physically violent incident (β = .139. Wald = 9.523, p = 00; $\chi 2$ = 10.171, p =.00) with each unit increase on the overall HCR-20 raising the daily hazard of violent acting out by about 15% (e^b = 1.149).

In terms of final risk judgments, the *low*, *moderate* and *high risk* groups differed in their survival probabilities when the model was processed in a continuous fashion (β =. 907, Wald = 4.231, p = .04; $\chi 2$ = 4.781, p =.03) with each categorical increase in summary risk rating (e.g., from *low* to *moderate*) multiplying the daily hazards of physical violence by approximately 2.5 (e^b = 2.477). However, when examined categorically, the model was no longer significant ($\chi 2$ = 4.899, p =.09). The alternate final risk judgments yielded overall significant models when treated both categorically

($\chi 2$ = 13.336, p =.01) or as a continuous variable (β =.527, Wald = 6.796, p = .01; $\chi 2$ = 7.961, p =.01).

Survival Analyses for Verbal Violence

The Historical subscale was marginally predictive of time to first verbal aggression (β = .148, Wald = 3.420, p =.06, $\chi 2$ = 3.568, p =.06), whereas the Clinical subscale was not (β = .139, Wald = .915, p =.34, $\chi 2$ = .922, p =.34). Risk Management items predicted time to first verbal violence (β = .387. Wald = 6.476, p = 01; $\chi 2$ = 7.056, p =.01) with each unit increase on that subscale raising the daily hazard of verbal aggression by about 47% (e^b = 1.473). HCR-20 total scores did not forecast time to first verbally violent incident (β = .078. Wald = 2.800, p = 09; $\chi 2$ = 2.798, p =.09).

In terms of final risk judgments, the *low*, *moderate* and *high risk* did not differ in their survival probabilities either when the model was processed in a continuous fashion (β =. 463, Wald = 1.109, p = .29; $\chi 2$ = 1.138, p =.28) or when it was examined categorically ($\chi 2$ = 3.577, p =.17). On the other hand, the alternate final risk judgments yielded overall significant models when treated both categorically ($\chi 2$ = 19.103, p =.00) or as a continuous variable (β =.404, Wald = 3.836, p = .05; $\chi 2$ = 4.046, p =.04).

Survival Analyses for all Outcomes across Genders

Time to first violent incident was examined across genders without including the HCR-20 ratings in the model. Men and women did not differ significantly in their survival probabilities for any of the three violent outcomes (any violence: β = .150. Wald

= .106, p = .38, $\chi 2$ = .106, p =.38; physical violence: β = -.039. Wald = .003, p = .96, $\chi 2$ = .003, p =.96; or verbal violence: β = .235. Wald = .150, p = .70, $\chi 2$ = .151, p =.70).

Survival Analyses: Moderation of Gender for Time to Any Violence

A series of Cox regressions was carried out in order to investigate whether gender moderates the relationship between the HCR-20 and time to first violent incident for the three categories of violence under consideration. With regard to *any violence*, the HCR-20 components and *gender* were entered into the first block. In block 1, the models pertaining to all components of the HCR-20, with the exception of the final risk judgments, were significant (see Table 14). In addition to the HCR-20 components and *gender*, the interaction between those two variables was entered in the second block. In block 2, no interaction term reached significance for any of the HCR-20 predictors. The models inherent to the H, R, and HCR predictors were significant.

Table 14. Cox Regression Analyses: Moderation of Gender on the Relationship between the HCR-20 and Time to First Violent Incident

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.156	6.309*	1.168	.195	4.733*	1.215	**1.** χ^2= 6.709*, p = .04
Gender	.082	.031	.861	.225	.176	1.252	**2.** χ^2= 7.349, p = .06
H*G				-.076	.372	.927	
C	.228	4.171*	.04	.230	2.183	1.258	**1.** χ^2= 4.360, p = .11
Gender	.032	.005	.945	.036	.005	1.037	**2.** χ^2= 4.360, p = .23
C*G				-.005	.000	.995	
R	.293	6.200*	1.340	.390	6.277*	1.477	**1.** χ^2= 6.579*, p =.04
Gender	-.036	.006	.964	.128	.063	1.136	**2.** χ^2= 8.555*, p = .04
R*G				-.210	.839	1.688	
HCR	.095	6.760**	1.100	.113	6.365*	1.120	**1.**χ^2= 7.004*, p = .03
Gender	.184	.146	1.202	.296	.315	1.345	**2.** χ^2= 8.212*, p = .04
HCR*G				-.048	.429	.953	
FRJ	.403	1.323	1.497	.249	.348	1.282	**1.** χ^2= 1.503, p =.47
Gender	-.038	.006	.963	-.305	.254	.737	**2.** χ^2= 1.976, p = .58
FRJ*G				.579	.533	1.784	
FRJ II	.364	4.168*	1.439	.312	2.250	1.367	**1.** χ^2= 4.614, p =.10
Gender	.148	.088	1.159	-.106	.023	.899	**2.** χ^2= 4.723, p = .19
FRJII*G				.221	.282	1.248	

Note. N = 73 (35 men, 38 women; 54 censored cases). *p = .05; **p = .01.

Survival Analyses: Moderation of Gender for Time to Physical Violence

The HCR-20 components and *gender* were entered into the first block. In block 1, the models pertaining to the Historical subscale, the total scores and the alternate final risk judgments were significant (Table 15). In block 2, no interaction term reached significance for any of the HCR-20 predictors. The models inherent to the H, R, total scores and alternate risk judgments were significant.

Table 15. Cox Regression Analyses: Moderation of Gender on the Relationship between the HCR-20 and Time to First Physically Violent Incident

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.287	7.969**	1.333	.128	1.044	1.137	**1.** χ^2= 9.770**, *p* = .01
Gender	-.539	.542	.583	-2.103	2.137	.122	**2.** χ^2= 12.870**, *p* = .01
H*G				.371	2.851	1.449	
C	.280	2.957	1.323	.311	1.723	1.365	**1.** χ^2= 3.008, *p* = .22
Gender	-.273	.141	.761	-.206	.065	.814	**2.** χ^2= 3.121, *p* = .37
C*G				-.059	.033	.943	
R	.401	5.260*	1.494	.193	.523	1.213	**1.** χ^2= 5.939, *p* =.05[a]
Gender	-.365	.247	.694	-.781	.930	.458	**2.** χ^2= 7.821*, *p* = .05
R*G				.365	1.038	1.441	
HCR	.165	9.516**	1.180	.146	2.295	1.157	**1.**χ^2= 11.233**, *p* = .00
Gender	-.912	1.219	.402	-1.046	1.050	.351	**2.** χ^2= 11.855**, *p* = .01
HCR*G				.027	.057	1.028	
FRJ	1.060	4.202*	2.888	1.058	1.119	2.882	**1.** χ^2= 5.036, *p* =.08
Gender	-.549	.464	.963	-.551	.230	.576	**2.** χ^2= 5.036, *p* = .17
FRJ*G				.003	.000	1.003	
FRJ II	.675	6.900**	1.965	.616	1.184	1.852	**1.** χ^2= 8.969*, *p* =.01
Gender	-.960	1.140	.383	-1.069	.675	.343	**2.** χ^2= 9.093*, *p* = .03
FRJII*G				.074	.014	1.077	

Note. *N* = 72 (34 men, 38 women; 64 censored cases). *p = .05; **p = .01; [a]sig =. 051.

Survival Analyses: Moderation of Gender for Time to Verbal Violence

The HCR-20 components and *gender* were entered into the first block. In block 1, only the model pertaining to the Risk Management subscale was significant (Table 16). In block 2, no interaction term reached significance for any of the HCR-20 predictors and, once again, only the model inherent to the Risk Management subscale was significant.

Table 16. Cox Regression Analyses: Moderation of Gender on the Relationship between the HCR-20 and Time to First Verbally Violent Incident

	Block 1			Block 2 (enter moderator)			Model
	B	**Wald**	**e^b**	**B**	**Wald**	**e^b**	
H	.148	3.298	1.159	.113	.950	1.119	**1.** χ^2= 3.568, *p* = .17
Gender	.023	.001	1.023	-.096	.019	.909	**2.** χ^2= 3.984, *p* = .26
H*G				.068	.178	1.071	
C	.132	.800	1.141	.004	.000	1.004	**1.** χ^2= .969, *p* = .62
Gender	.130	.044	1.139	.049	.005	1.050	**2.** χ^2= 1.582, *p* = .66
C*G				.220	.489	1.246	
R	.385	6.379*	1.469	.209	.854	1.232	**1.** χ^2= 7.060*, *p* =.03
Gender	.104	.029	1.110	-.243	.113	.784	**2.** χ^2= 10.072*, *p* = .02
R*G				.309	1.062	1.362	
HCR	.078	2.617	1.081	.002	.001	1.002	**1.**χ^2= 2.802, *p* = .25
Gender	-.028	.002	.973	-.187	.068	.830	**2.** χ^2= 5.312, *p* = .15
HCR*G				.117	1.416	1.124	
FRJ	.442	.946	1.556	-.294	.069	.745	**1.** χ^2= 1.174, *p* =.56
Gender	.106	.028	1.112	-.235	.093	.790	**2.** χ^2= 2.080, *p* = .56
FRJ*G				.923	.566	2.516	
FRJ II	.420	3.498	1.523	.282	.298	1.326	**1.** χ^2= 4.087, *p* =.13
Gender	-.125	.036	.883	-.314	.119	.731	**2.** χ^2= 4.585, *p* = .21
FRJII*G				.172	.090	1.188	

Note. *N* = 71 (33 men, 38 women; 60 censored cases). *p = .05; **p = .01.

Supplemental Analyses

Although the regression analyses described earlier did not indicate a moderating effect of gender on the predictive validity of the HCR-20 for violent outcomes, it seemed worthwhile, nevertheless, to examine the predictive validity of the instrument for men and women separately. However, the following empirical investigations should be considered tentative and exploratory compared to the formal tests of moderation since they suffer both from lower power profiles given the limited number of participants in either of the gender groups (i.e., 35 men and 38 women) and from other potential

weaknesses (i.e., differing variances, ranges and errors across groups) which may lead to spurious correlational effects (see Baron & Kenny, 1986).

Point Biserial Correlations between the HCR-20 and Violent Outcomes across Genders

Table 17 presents the correlations between the HCR-20 components and violent outcomes in the male subsample. The historical and risk management subscales as well as the total scores were associated with all violent outcomes (*r* ranging from .37 to .62, *p* = .01). However, the summary risk ratings (i.e., both versions) correlated solely to physical violence (*r* = .39 and .40, *p* = .01) and the clinical subscale did not demonstrate any statistical association with any of the three outcomes. All coefficients indicated moderate to large effect sizes. In the female subsample, none of the HCR-20 components evidenced a statistically significant relationship to violent outcomes. However, a number of effect sizes of moderate magnitude were noted, raising the possibility that the observed absence of statistical significance embodies an artifactual consequence of the low power profile rather than a true absence of effect in the population.

Table 17. Correlations (r_{pb}): HCR-20 and Violent Outcomes by Gender

	Men			Women		
	Any Violence	**Physical Violence**	**Verbal Violence**	**Any Violence**	**Physical Violence**	**Verbal Violence**
H total	.37**	.48**	.45**	.27	.18	.28
C total	.27	.27	.21	.27	.25	.11
R total	.48**	.40**	.62**	.13	.06	.15
HCR-20 total	.45**	.50**	.52**	.34	.26	.30
Final Risk Judgments	.14	.39*	.26	.32	.18	.29
Alternate Final Risk Judgments	.11	.40**	.21	.23	.18	.18

Note. Men: N = 35; Women: N = 38. 1-tailed; **Significant at .01; *Significant at .05.

Fisher's *z*-tests were carried out in order to examine potential differences between correlation coefficients across genders. Taking $z = 1.96$ for $p = .05$ and $z = 2.58$ for $p = .01$ (2-tailed), men and women differed significantly solely on the relationship between the risk management subscale and verbal violence during follow-up ($z = 2.346$) with men evidencing a stronger such relationship.

ROC Analyses: Predictive Validity of the HCR-20 across Genders

ROC analyses were conducted in order to examine the predictive acumen of the HCR-20 for men and women separately, yet once again, these should be considered exploratory and tentative. In the male subsample (Table 18), *any violence* was predicted by the risk management subscale; *physical violence* by the historical subscale, the HCR-20 total scores and the alternate final risk judgments; and *verbal violence* by the risk management subscale as well as the total scores on the instruments . All observed effect sizes were moderate to large in magnitude (see Douglas et al., 2005). The clinical subscale did not appear to predict any of the three outcomes for men and none of the HCR-20 components demonstrated predictive power with regard to violent outcomes in the female subsample (see Table 19). Once again, a number of effect sizes of moderate magnitude were noted for women, raising the possibility that the observed absence of statistical significance for these embodies an artifactual consequence of the low power profile rather than a true absence of effect in the population.

Table 18. Predictive Validity (AUCs) of the HCR-20 for Violence in Men

	Any Violence		Physical Violence		Verbal Violence	
	AUC	**SE**	**AUC**	**SE**	**AUC**	**SE**
H	.66	.11	.81*	.17	.72	.11
C	.62	.11	.68	.16	.58	.13
R	.72*	.11	.77	.16	.80*	.11
HCR	.69	.11	.83*	.14	.74*	.12
FRJ	.54	.11	.77	.15	.61	.12
FRJ-II	.61	.10	.85*	.11	.63	.12

Note. N = 35. 1-tailed; *Significant at .05.

Table 19. Predictive Validity (AUCs) of the HCR-20 for Violence in Women

	Any Violence		Physical Violence		Verbal Violence	
	AUC	**SE**	**AUC**	**SE**	**AUC**	**SE**
H	.71	.08	.67	.12	.72	.09
C	.65	.12	.60	.22	.61	.13
R	.60	.11	.56	.16	.62	.11
HCR	.60	.12	.69	.15	.58	.13
FRJ	.62	.11	.63	.16	.60	.13
FRJ-II	.66	.11	.58	.17	.65	.11

Note. N = 38; 1-tailed.

Descriptives of the Nature and Severity of Violence across Genders

Aside from the predictive validity and moderation enquiries presented above, another aim of the present study was to explore the nature, severity and targets of both the predicted and perpetrated violent acts across genders. First, ratings inherent to the baseline HCR-20 assessments are compared between men and women, then the qualities of the violent acts observed during the follow-up period are described for each subgroup.

Comparison of the Predicted Nature and Severity of Violence across Genders Based on HCR-20 Assessments

Gender differences in the nature and severity of the forecasted violence based on the baseline HCR-20 assessments emerged (Table 20). Women were more often believed to be at low risk of engaging in minor forms of violence, whereas men were more often rated to be at moderate risk for such acts. No women were judged to pose a high risk of committing either moderate of severe aggression. Moreover, all women were perceived to be at low risk of perpetrating instrumental acts of violence, whereas 8.2 % and 4.1% of their male counterparts were judged to pose a moderate and high risk (respectively) for aggression of that nature. Although women were more often predicted to be at moderate or high risk for hostile-reactive forms of aggression compared to instrumental ones, they also obtained risk ratings lower than that that of men for violence of that nature. No women were assessed to be at high risk of targeting strangers or using weapons.

Table 20. HCR-20: Comparative Chi-Square tests across Genders

Predicted			Men	Women	χ^2	Sig.
Violence Severity	Minor:				7.361	.025[a]
		Low	37.5%	65.2%		
		Moderate	50.0%	26.1%		
		High	12.5%	8.7%		
	Moderate:				5.244	.073
		Low	71.4%	87%		
		Moderate	20.4%	13%		
		High	8.25	0%		
	Severe:				3.754	.153
		Low	87.8%	97.8%		
		Moderate	8.2%	2.2%		
		High	4.1%	0%		
Target	Family/Friend:				.710	.701
		Low	59.2%	63%		
		Moderate	38.8%	32.6%		
		High	2%	4.3%		
	Stranger:				5.861	.053[a]
		Low	64.6%	82.6%		
		Moderate	27.1%	17.4%		
		High	8.3%	0%		
Nature of Violence	Weapon Use:				3.456	.178
	Low		83.7%	93.5%		
	Moderate		10.2%	6.5%		
	High		6.1%	0%		
	Instrumental:				6.012	.046
	Low		87.8%	100%		
	Moderate		8.2%	0%		
	High		4.1%	0%		
	Hostile-Reactive:				9.053	.011
	Low		34.7%	65.2%		
	Moderate		49%	28.3%		
	High		16.3%	6.5%		

Note. N = 95 (men = 49, women = 46); [a]N = 94.

Nature and Severity of Follow-Up Violent Incidents across Genders

Detailed data were available for 17 incidents, 9 perpetrated by men and 8 perpetrated by women. Due to this very limited number of data points, no statistical comparisons were carried out across genders and the descriptive frequencies presented in Table 21 should be considered exploratory and interpreted with caution. Most of the participants' aggression seemed to have involved pushing, grabbing or shoving someone.

All of the men's victims were male, whereas women acted violently towards equivalent proportions of men and women. Men aggressed similar proportions of family, friends, acquaintances and strangers, whereas half of the women's victims were family members. If any injury was inflicted during the incidents, only minor harm reportedly occurred across both genders. Weapons were never involved in male perpetrated aggression and never actually used if involved in female perpetrated violence. Approximately half of the participants of both genders uttered threats as part of their aggression.

Table 21. Nature and Severity of Follow-Up Violent Incidents across Genders

		Men *N* (% of incidents or % of victims[a])	Women *N* (% of incidents or % of victims[a])
Type of Violence	Threw something at someone	1 (11.1%)	0
	Pushed, grabbed, or shoved someone	5 (55.5%)	3(37.5%)
	Slapped anyone	0	2 (25%)
	Kicked, bit or choked someone	1 (11.1%)	1(12.5%)
	Hit someone with a fist, or beat someone up	2 (22.2%)	0
	Threatened someone with a lethal weapon	1 (11.1%)	2 (25%)
	Used a knife or fired a gun at someone	0	0
Victim's Gender	Male	9 (100%)	5 (50%)[a]
	Female	0	5 (50%)
Victim's Age	Mean Age	31.6 (*N* = 8)	24.8 (*N* = 8)
Relationship with Victim	Family	2 (22.2%)	5 (50%)[a]
	Friend	3 (33.3%)	1 (10%)
	Acquaintance	2 (22.2%)	2 (20%)
	Stranger	2 (22.2%)	2 (20%)
Harm Inflicted	Minor	7 (77.7%)	4 (40%)[a]
	Moderate	0	0
	Severe	0	0
	Death	0	0
	Not Applicable	2 (22.2%)	6 (60%)
Weapon Used	No	9 (100%)	6 (75%)
	Possession	0	1 (12.5%)
	Threatened	0	1 (12.5%)
	Used	0	0
Use of Threats	Yes	4 (44.4%)	4 (50%)

Note. *N* = 17 incidents (9 perpetrated by men; 8 perpetrated by women, [a] with one of the latter involving three victims).

Exploratory Predictive Validity Analyses for Other Negative Outcomes

The final goal of the present study was to explore the potential relationship between the HCR-20 and other negative outcomes such as violent victimization experiences and deliberate acts of self-injury. Findings are presented both for the whole sample and for men and women separately.

Intercorrelations between all Outcomes for the Whole Sample and by Gender

All violent outcomes (any violence, physical violence and verbal violence) were intercorrelated to a moderate to very large degree, with the stronger association observed between *any violence* and *verbal violence* (Table 22). With regard to other negative outcomes, only self-harm and violent victimization exhibited a statistically significant association. Suicide was not correlated with any of the violent or non-violent outcomes. Self-harm was associated with physical violence and violent victimization displayed a moderate to large relationship with all three violent outcomes.

Table 22. Intercorrelations (r_{pb}) between all Outcomes for the Whole Sample

	Any Violence	Physical Violence	Verbal Violence	Suicide Attempt	Self-Harm	Violent Victimization
Any Violence	-	.59**	.86**	.09	.16	.41**
Physical Violence	.59**	-	.36**	.02	.21*	.37**
Verbal Violence	.86**	.36**	-	.04	.10	.45**
Suicide Attempt	.09	.02	.04	-	.06	-.02
Self- Harm	.16	.21*	.10	.06	-	.26*
Violent Victimization	.41**	.37**	.45**	-.02	.26*	-

Note. N = 73. 1-tailed. **Significant at .01; *Significant at .05;[a]Sig = .054

For both genders, any violence displayed a moderate to very strong relationship to both physical and verbal violence, as well as violent victimization (Table 20). Physical violence was associated with verbal violence, self-harm, and violent victimization in men, but not in women. Verbal violence evidenced a relationship with violent victimization across genders and with physical violence in men. Suicide was not correlated with any of the other five outcomes for either gender. Violent victimization was associated with all three violent outcomes in the male subsample and with any violence, verbal violence and self-harm in its female counterpart.

Table 23. Intercorrelations (r_{pb}) between all Outcomes across Genders

	Any Violence		Physical Violence		Verbal Violence		Suicide Attempt		Self-Harm		Violent Victimization	
	M	**W**	**M**	**W**	**M**	**W**	**M**	**W**	**M**	**W**	**M**	**W**
Any Violence	-	-	.56**	.62**	.86**	.85**	.10	.07	.03	.23	.47**	.38*
Physical Violence	.59**	.62**	-	-	.45**	.23	.11	.10	.21*	.22	.49**	.30
Verbal Violence	.86**	.85**	.45**	.23	-	-	.03	.11	.08	.11	.58**	.37*
Suicide Attempt	.10	.07	.11	-.10	-.03	.11	-	-	.17	-.09	.00	-.01
Self-Harm	.03	.23	.21*	.22	.08	.11	.17	.09	-	-	.10	41*
Violent Victimization	.47**	.38*	.49**	.30	58**	.37*	.00	.01	.10	.41*	-	-

Note. N = 73(35 men, 38 women); 2-tailed; **Significant at .01; *Significant at .05.

Point Biserial Correlations between the HCR-20 and Other Negative Outcomes for the Whole Sample and across Genders

Of all three negative outcomes examined (i.e., suicide, self-harm and violent victimization), only violent victimization correlated with the HCR-20 in this sample of psychiatric patients (Table 24). More specifically, the historical subscale, the total scores

and alternate final risk judgments exhibited a statistically significant association of a moderate magnitude with that outcome. The effect size was large for the relationship between the risk management subscale and violent victimization.

Table 24. Correlations (r_{pb}) with Other Negative Outcomes for the Whole Sample

	Suicide Attempt	Self-Harm	Violent Victimization
H total	.19	.14	.30**
C total	.10	.14	.06
R total	.03	.02	.47**
HCR-20 total	.03	.14	.30**
Final Risk Judgments	.00	.07	.08
Alternate Final Risk Judgments	.06	.15	.20*

Note. N = 73;1-tailed; **Significant at .01; *Significant at .05

In the male subsample, the HCR-20 did not correlate with suicide attempts or self-harm incidents during the follow-up period (see Table 25). However, the historical and risk management subscales, as well as the total scores were associated with violent victimization incidents to a large degree. For women, a different pattern of results was observed. The historical subscale, total scores, and, in a marginal sense, both versions of the final risk judgments correlated with suicide attempts. The clinical subscale was associated with self-harm incidents and the risk management subscale as well as the total scores were associated with violent victimization. All effect sizes observed in the female subsample were of a moderate magnitude.

Table 25. Correlations (r_{pb}): HCR-20 and Other Negative Outcomes across Genders

	Suicide Attempt		Self-Harm		Violent Victimization	
	M	**F**	**M**	**F**	**M**	**F**
H total	.03	.38*	.10	.18	.57**	.21
C total	-.13	.04	-.01	.32*	.04	.11
R total	-.19	.08	.07	-.04	.62**	.38*
HCR-20 total	-.10	.29*	.07	.21	.53**	.32*
Final Risk Judgments	-.17	.27[a]	.07	.04	.26	-.04
Alternate Final Risk Judgments	-.09	.27[b]	.08	.05	.27	-.05

Note. N = 73 (35 men, 38 women); 1-tailed. **α = .01; *α =.05; [a]p=.052; [b]p=.056

Fisher's *z*-tests were conducted in order to examine potential differences between correlation coefficients across genders. Taking $z = 1.96$ for $p = .05$ and $z = 2.58$ for $p = .01$ (2-tailed), men and women did not display significantly different coefficients. However, the relationships between both versions of the final risk judgments and suicide were marginally stronger in women (z=1.83 in both cases), whereas the relationship between the historical subscale and violent victimization appeared marginally stronger in men (z=1.77).

ROC Analyses: Predictive Validity of the HCR-20 for Other Negative Outcomes across Genders

ROC analyses yielded large effect sizes with regard to the predictive power of the historical and risk management subscales as well as the total scores for the violent victimization of men during the follow-up period (see Table 26). However, the HCR-20 did not predict suicide attempts or self-harm incidents in the male subsample. For women (see Table 27), large effect sizes were observed for the prediction of suicide attempts

based on the historical scale and that of violent victimization based on the risk management subscale. However, no other indices demonstrated statistically significant predictive effects with any of the three outcomes.

Table 26. Predictive Validity (AUCs) of the HCR-20 for Other Negative Outcomes in Men

	Suicide Attempt		Self-Harm		Violent Victimization	
	AUC	**SE**	**AUC**	**SE**	**AUC**	**SE**
H	.50	.13	.55	.19	.87**	.07
C	.40	.14	.50	.13	.48	.14
R	.38	.13	.55	.19	.87**	.10
HCR	.44	.13	.50	.20	.84*	.09
FRJ	.39	.12	.53	.20	.65	.13
FRJ-II	.47	.12	.53	.22	.70	.12

Note. N = 35; 1-tailed; **Significant at .01; *Significant at .05.

Table 27. Predictive Validity (AUCs) of the HCR-20 for Other Negative Outcomes in Women

	Suicide Attempt		Self-Harm		Violent Victimization	
	AUC	**SE**	**AUC**	**SE**	**AUC**	**SE**
H	.85*	.08	.76	.07	.67	.10
C	.70	.13	.78	.11	.60	.11
R	.73	.14	.46	.17	.75*	.08
HCR	.60	.24	.74	.10	.70	.10
FRJ	.72	.17	.54	.18	.63	.10
FRJ-II	.80	.08	.52	.21	.58	.10

Note. N = 38; 1-tailed; *Significant at .05.

DISCUSSION

Although the HCR-20 embodies one of the most researched and well-validated violence risk instruments, the literature pertaining to its applicability across genders is scarce. In the present study, the impact of gender on the performance of the scale was formally tested for the first time in a sample of short-term psychiatric inpatients. Supplemental analyses were carried out in order to explore the differential qualities of the violence perpetrated by men and women as well as the potential usefulness of the HCR-20 for the forecasting of other negative events such as violent victimization, self-harm and suicide.

Results indicate that the HCR-20 as well as its components predict the likelihood and imminence of violent outcomes in this sample of short-term psychiatric inpatients. Moreover, gender did not moderate that relationship. Exploratory analyses revealed gender differences in the baseline item and scale ratings as well as in the nature of both the predicted and observed violence. Additionally, the HCR-20 demonstrated an association with other negative outcomes, and more particularly violent victimization.

Sample Descriptives

Overall, socio-demographic as well as clinical and legal variables inherent to the present sample were similar to those reported for other (civil or forensic) psychiatric samples (e.g., de Vogel & de Ruiter, 2005; Lidz, Mulvey, & Gardner, 1993; Nicholls et

al., 2004; Strand & Belfrage, 2001) across Europe and North America. Most patients of both sexes were admitted to the hospital involuntarily, were unemployed, and had a history of prior commitment decisions indicative of ongoing psychiatric struggles. Men appeared to be diagnosed most often with bipolar and/or substance abuse/dependence disorders, whereas women seemed to present most often with depressive and/or psychotic disorders as well as personality disorders, particularly borderline personality disorder/traits. In addition, violence (i.e., not spousal violence or stalking) was more often relevant to their current admission but rates of inpatient physical aggression did not differ across genders. Female patients were more likely than their male counterparts to be in a long-term relationship and to have children, and to report sexual, physical, and emotional abuse or neglect during childhood compared to their male counterparts. With regard to legal variables, the participants also conformed to expectations based on prior research (e.g., Nicholls et al., 2008) with men displaying more adult arrests, charges, convictions, and incarcerations, and being more likely to have committed various types of violent and non-violent crimes.

It is worth noting that a considerable proportion of men had been homeless in the past and 10% of them did not have a fixed address at admission, whereas this seemed to represent less of an issue for women. In the future, it may be worthwhile to investigate further the causes, correlates and consequences of this gender disparity since homelessness embodies an extremely salient risk factor for a vast array of deleterious medical, psychiatric, social and legal outcomes.

HCR-20 Descriptives and Baseline Comparisons

In this sample, most individuals of both genders were rated as *low risk* and no female participant was deemed to be at high risk for violence. In accordance with the results from the two studies which examined summary risk ratings across genders (de Vogel & de Ruiter, 2005; Nicholls et al., 2004), women were both less likely to be classified in higher risk categories and more likely to be rated as lower risk as compared to men.

Unlike the findings reported in forensic populations (de Vogel & de Ruiter, 2005; Nicholls, 2001; Strand & Belfrage, 2001), some subscale- and scale-level gender discrepancies were observed in the present sample. Similarly to the research conducted with chronic involuntary patients (Nicholls et al., 2004) and contrary to that with inmates (Coid et al., 2009), male participants exhibited higher Historical and total scores on the instrument compared to female patients. However, in contrast to results generated with chronic involuntary patients (Nicholls et al., 2004) and inmates (Coid et al., 2009), no gender differences were noted with regard to the clinical subscale total scores. These findings taken together, although tentative, generally suggest that the subscale and total scores obtained by women, compared to those of their male counterparts, tend to be lower in psychiatric populations, equivalent in forensic samples, and higher in inmates convicted of violent or sexual offences.

Generally in line with the research conducted in forensic populations (de Vogel & de Ruiter, 2005; Nicholls, 2001; Strand & Belfrage, 2001) a number of item-level differences were noted. Indeed, men exhibited more frequent or severe previous violence,

committed violent acts at a younger age, evidenced more substance use problems (i.e., sole difference reported by Nicholls, 2001), were rated as more psychopathic, had more prior supervision failures on average and were deemed more likely to lack personal support, to be noncompliant with remediation attempts and to experience lower levels of stress compared to women. Although no comparisons reached statistical significance on the Clinical subscale, male patients can be considered to lack insight and display negative attitudes more so, in a marginal sense, than their female counterparts. However, in this sample, women did not yield higher ratings on the personality disorder, impulsivity and relationship instability risk factors (see de Vogel & de Ruiter, 2005; Strand & Belfrage, 2001).

Interrater Reliability

When coding the HCR-20 components, the raters agreed to a substantial degree and the ICCs for most of the HCR-20 components were well within the ranges described in the quasi-meta-analytic review conducted by Douglas and Reeves (2010), even though they tended to be somewhat lower than the median values calculated on the 36 studies which examined the interrater reliability of the instrument. However, more rating discrepancies were observed as pertains to the risk management subscale and the ICC associated with this susbscale (i.e., .43) fell slightly out of the reported range (i.e., lowest reported ICC for this subscale = .47). Most of the total risk management scores did not vary by more than one or two points (i.e., out of a total subscale score of 10) within pairs of raters but for four cases (i.e., out of 21) the scores differed by three points and for one participant, they did by four points. The cause of those disparities is unclear.

Because forensic and clinical decisions as well as recommendations are based on the summary risk judgments, these embody the most important ratings on the HCR-20. The ICCs inherent to these components were good to excellent.

Outcomes in the Whole Sample and across Genders

In line with the published literature (e.g., Lidz, Mulvey, & Gardner, 1993; Monahan et al. 2001), the base rates of follow-up violent acts did not differ markedly across gender, corroborating the drastic reduction of the gender gap with regard to violence perpetration in mentally ill individuals as compared to the general population (e.g., Stueve & Link, 1998). As pertains to self-harming and suicidal acts, prior research suggests a considerably larger proportion of females engaging in such behaviors in adolescence as compared to males, with the gender gap greatly narrowing in early adulthood, and disappearing completely in later life (e.g., Hawton & Harris, 2008). Consequently, the findings from this sample of 19 to 61 year-old psychiatric patients, which yield no gender differences in the occurrence of deliberate self-injurious behaviors during the follow-up period, did not depart significantly from expected trends. As in prior empirical accounts (e.g., Wood & Edwards, 2005), men and women also displayed similar rates of violent victimization overall. Although variations in the types of such victimization were not examined in the present study, this may embody a worthwhile area of enquiry in the future since male and female have been shown to experience differential rates of victimization depending on the nature of that experience (Kilpatrick & Acierno, 2003; Lauritsen & Heimer, 2008).

In the literature, violence perpetration and violent victimization have been shown to be more common among persons with severe mental illness than among the general population (Choe, Teplin, & Abram, 2008). In this sample, all types of violence perpetration were strongly interrcorrelated and evidenced relationships of non-negligible magnitudes with violent victimization (with the exception of physical violence in women). Interestingly, much of self-directed aggression did not statistically correlate with other-directed violence. However acts of self-harm evidenced a moderate link to physical aggression in male patients and to the experience of violent victimization in their female counterparts. The present observations lend some support to prior research suggesting that, especially in psychiatric populations, when assessing individual risk for negative outcomes such as violence, the likelihood of patients being vulnerable for a wide variety of often interconnected intra- and inter-personal detrimental occurrences needs to be considered and communicated (e.g., Choe et al., 2008; Hillbrand, 200; Nicholls et al., 2006).

Predictive Validity:
Predictive Accuracy of the HCR-20 for Violent Outcomes

In line with the published literature (e.g., Douglas & Reeves, 2010), overall, the HCR-20 components predicted violence to a moderate degree in correlational and regression analyses, including physical and verbal violence when those subtypes were examined separately. The Historical subscale, total scores and alternate final risk judgments predicted all types of violence, whereas the Clinical and the risk management subscales as well as the original version of the final risk judgments varied in their

association to violence depending both on the type of violence examined and on the statistical analyses conducted (i.e., correlations vs. logistic regressions). The total scores exhibited a comparatively strong relationship with physical violence, which was also true for the alternate final risk judgments and their association to verbal violence.

A similar pattern of results was observed when ROC analyses were conducted, although fewer effects reached significance with these less powerful procedures. The HCR-20 total scores predicted physical violence and nay violence with moderate accuracy, but exhibited only marginally significant results for verbal aggression and both versions of the summary risk ratings failed to demonstrate a predictive association with violent outcomes during the follow-up period. The effect sizes inherent to the Historical (.69), Clinical (.65) and Risk Management (.66) subscales as well as the total scores on the instrument (.66) mirrored the median results reported by Douglas and Reeves in their quasi-meta-analytic review of the HCR-20 literature (.68, .62, .65 and .66 respectively), with the Historical subscale outperforming the other numerical scores. They also resembled the weighted values described by Guy (2008) in a recent meta-analysis (.70, .69, .71 and .73 respectively). Although Douglas and Reeves (2010) and Guy (2008) indicated that the final risk judgments were generally most accurate at predicting future violence (average = .70 and AUC_w =.76), and prior research found those ratings to add incrementally to the numerical performance of the HCR-20 (e.g., de Vogel & de Ruiter, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; Douglas et al., 2003), the present study yielded more inconsistent results with regards those variables.

The HCR-20 seemed to generally predict timing to first violent incident. Indeed, as the ratings on most of the HCR-20 components increased, so did the daily odds of perpetration of all types of violence. Comparably to the results described above, the Historical and Risk Management subscales as well as the total scores and alternate final risk judgments yielded significant models for most outcomes, whereas more variability arose with regard to the Clinical subscale and the original final risk judgments. Overall, findings echoed those from prior research (e.g., Douglas et al., 2003), with higher risk patients being more likely to become violent and to act out sooner compared to lower risk patients. In other words, the HCR-20 not only appears to predict the likelihood of violence perpetration but also its imminence.

Inquiries into a Possible Impact of Gender on the Predictive Validity of the HCR-20

Findings were consistent with expectations, suggesting no moderating effect of gender on the relationship between HCR-20 ratings and the occurrence of all types of violent outcomes during the follow-up period. Indeed, although no study had yet formally tested the impact of gender on the predictive validity of the scale in a sample of short-term psychiatric inpatients, prior empirical enquiries have indicated that, generally, men and women share similar risk factors and motivations for violence (see Nicholls et al., 2008, for review), the HCR-20 evidences predictive validity with female assessees in studies with adequate statistical power (Nicholls et al., 2004), and gender may not moderate the prediction of criminal reconvictions (i.e., including violent reoffenses) with inmates (Coid et al., 2009).

When the predictive validity of the HCR-20 was modeled while controlling for time and uneven follow-up periods, similar results were observed. Moreover, the timing of aggression did not differ between men and women. These results taken together indicate that the assessment scheme likely exhibits an equivalent overall performance across genders in terms of its ability to predict both the likelihood and the imminence of violence perpetration and therefore support its use with both male and female psychiatric inpatients.

Given that the present study represents the second formal test of the impact of gender on the predictive validity of the HCR-20, further exploratory analyses were conducted in order to generate a broader and more detailed understanding of the data. However, the discussion presented below is to be considered with due caution. Indeed, it was informed by procedures that were less powerful and less rigorous in a statistical sense than those previously reported.

As pertains to the female subsample none of the HCR-20 components evidenced a statistical relationship with any of the violent outcomes regardless of the analyses conducted (i.e., correlations vs. ROC). These findings mirror those from Schaap and colleagues (2009) who did not find the HCR-20, its subscales or the final risk judgment to predict violent or general recidivism in their sample of female forensic patients. On the other hand, in the present study, the HCR-20 indexes generally exhibited moderate to large correlations with all types of violence in men, with the exception of the Clinical subscale. In the latter subgroup, both versions of the final risk judgments correlated only with physical violence. Direct comparisons between correlation coefficients indicated

that the effect sizes differed across genders solely for the association between the Risk Management subscale and verbal violence. Results from ROC analyses yielded a parallel yet less uniform predictive picture. However, in light of the current power limitations resulting from restricted sample sizes, it seems reasonable to suggest that only larger effect sizes might have been detected, whereas existing effects of more modest magnitude may have been overlooked. The patterns of findings observed in the four methodologically equivalent studies that investigated the predictive validity of the HCR-20 across genders seem to support this conclusion. Indeed, the samples of a similar size (de Vogel & de Ruiter, 2005; Nicholls, 2001) generated comparable results across genders, whereas in the larger samples (Coid et al., 2009; Nicholls et al., 2004) most components of the scheme predicted most outcomes in both men and women. It may be the case that the HCR-20 forecasts violent outcomes in both genders, although it may do so more strongly for male assessees.

Since data collection is ongoing for the present study, it will be important to re-examine the performance of the scale in the future, as sample size increases and with it, the ability to address certain empirical questions with greater statistical accuracy. Based on the findings presented by Nicholls and colleagues (2004), a sample twice the size of the current one (i.e., 75 men and 75 women approximately) should yield adequate power for follow-up analyses.

Gender-Based Variations in the Type, Severity and Victims of the Forecasted and Observed Violence across Genders

Research indicates that the violence committed by women tends to be more reactive (i.e., as opposed to instrumental), less severe, and to occur more often in familial and intimate contexts compared to that of men (e.g., McKeown, 2010; Monahan et al., 2001; Nicholls, 1997; Nicholls et al., 2008). In this sample, at baseline, no women were judged to pose a high risk of committing either moderate or severe aggression, to engage in instrumental aggression, use weapons, or target strangers, whereas some men were. Even for minor forms of violence, more men were perceived to be at risk. Interestingly, female patients were judged less likely to perpetrate hostile-reactive aggression than men, which may have resulted from the fact that women were on the whole rated as embodying less of a violence threat. However, in line with what has been reported in the literature, over a third of the female participants were assessed to be at risk for this type of aggression, whereas none were for instrumental violence.

Of the few participants who actually became violent, men aggressed similar proportions of family, friends, acquaintances and strangers, whereas half of the women's victims were family members. This corroborates findings from prior research indicating that women tend to direct their aggression towards close others and act out in private spheres such as home more so than do men (e.g., Monahan et al., 2001; Nicholls et al., 2008). In this sample, since, allegedly, no or minor harm was inflicted during violent events, no differences in the severity of the perpetrated violence became apparent across genders. Although more men were predicted to be at risk of using weapons, none of them reported having actually used one. However, two out of the eight violent women engaged

in violence involving weapons (i.e., threatened use). This seems to mirror other accounts (e.g., Brennan & Moore, 2009) suggesting that although males tend to carry weapons more often, females are more likely to use weapons in domestic contexts.

One must keep in mind that the aforementioned results were derived from a very limited number of observations and should therefore be considered tentative. As more data are being collected, more conclusive interpretations may be generated. Moreover, as sample size increases, it will be possible to formally test (i.e., rather than broadly describe) the accuracy of individual predictions pertaining to the nature and severity of the forecasted violence across genders. Moreover, it will be worth investigating in the near future whether some HCR-20 components show a consistent association with the nature of the violent acts, the degree of the inflicted harm, and the types of victims.

Predictive Validity of the HCR-20 for Suicide Attempts, Self-Harm Incidents, and Violent Victimization

Given that violence perpetration and violent victimization seem to arise in the context of similar historical, personal, and situational risk factors, individuals at risk for aggression tend to also be at risk to be victimized. According to expectations, the HCR-20 predicted violent victimization in both men and women, although the presence and strength of that association varied across genders and scale components.

However, in the sample taken as a whole, the HCR-20 did not correlate with self-directed aggression. This observation mirrors the findings from Gray and colleagues (2003) who report no association between the HCR-20 and instances of deliberate self-injurious behaviors in forensic patients. Yet it opposes other results linking risk for

violence towards the self and that for violence towards others (e.g., Hillbrand, 2001) and more specifically, the utility of the HCR-20 in capturing risk for self-harm (Daffern & Howells, 2007). In line with the latter research, when data were examined for men and women separately in the present sample, HCR-20 components were associated with suicide attempts and self-harm in women. It may be the case that in female psychiatric patients, the risk factors, for outwardly- and inwardly-directed aggression, as measured by the HCR-20, are more intrinsically linked and/or that the female expression of anger (or other such precipitant for violence) may be more often re-directed towards the self as the result of socialization and societal expectations as compared to men (e.g., Motz, 2008). Undoubtedly, these observations should be viewed as tentative at this point and much more research is needed before firm conclusions be reached.

Limitations and Strengths

Some important limitations must be noted. First, although the sample size yielded adequate statistical power for the main analyses, which were conducted on the entire sample (N = 95 at baseline and N = 73 at follow-up), the power profile was substantially weaker for the supplemental analyses (N = 38 and N = 35) and it seems reasonable to believe that some existing effects likely were not detected. Second, because only a subset of the participants became violent during the follow-up period, few cases yielded detailed information pertaining to the nature, severity and targets of the aggressive acts and no formal comparisons could be carried out across genders. However, even if tentative at this point, the observed qualitative properties of the perpetrated violence are meaningful in a descriptive sense and suggest that further related enquiries will be worth pursuing as

more data are generated. Third, the available data did not allow for investigations into the predictive validity of the HCR-20 with regard to the frequency of violent and non-violent outcomes.

Finally, although the participants lost to attrition did not appear to differ significantly (i.e., general demographic variables and HCR-20 ratings) from those for whom post-discharge data were available, it may be the case that these two groups of patients experienced dissimilar rates of outcomes, thereby generating somewhat biased inferences pertaining to the overall population. Moreover, some of the patients who remained violence-free during the follow-up period exhibited a very short time at risk (e.g., 28 days), which provided less opportunity for recording any of the outcomes under scrutiny should they have occurred during the months following the initial assessment. These limitations will be circumvented to some degree in the near future as new follow-up information is made available to the author (i.e., review of medical files and online court reports) with regard to both the individuals who did not generate any data past baseline and those who completed some but not all follow-up sessions.

In spite of the aforementioned limitations, the present study evidences considerable methodological strengths such as its truly prospective design, the collection of thorough and varied information (i.e., interview, file review, and collateral data), and the inclusion of a similar number of male and female participants. Moreover, this is the first study to both examine the performance of the HCR-20 across genders in a sample of inpatients from a short-term psychiatric unit and to formally test whether gender moderates the predictive validity of the instrument. Further, given that data collection is ongoing, stronger conclusions pertaining to the research questions at hand will be made

possible as more information becomes available as pertains both to the current participants and to new recruits.

Conclusions and Future Directions

Violence risk assessments represent a crucial component in the management of criminal, forensic and psychiatric populations. They are conducted in order to identify both high risk and low risk individuals and subsequently inform adequate management strategies. Because the clinical and legal decisions that are made based on risk evaluations can have tremendous personal and social consequences - such as the unjust violation of someone's rights and freedom or conversely, the failure to prevent harmful or lethal occurrences – it is of utmost importance that such procedures be theoretically, empirically and clinically sound.

Despite rapid advances in the conceptualization and assessment of risk for violent behaviors in male populations, the applicability of current violence risk measures to women remains less well understood (Garcia-Mansilla, Rosenfeld, & Nicholls, 2009). Although men are more likely to be convicted of violent offences, the gender gap seems nonexistent and even sometimes reversed when the observed and self-reported violence perpetration of male and female psychiatric inpatients is examined (e.g., Garcia-Mansilla, Rosenfeld, & Nicholls, 2009; Lidz, Mulvey, & Gardner, 1993; Manchak et al., 2009). Such observations suggest that efforts directed at refining the understanding, assessment and prevention of female violence are far from being futile.

Results from the present study indicate that, generally, the predictive power of the HCR-20 violence risk assessment scheme for the likelihood and imminence violence

does not appear affected by the gender of the assessees and that, consequently the use of the instrument with both male and female psychiatric inpatients is supported. However, in line with prior research (e.g., Hiday, Swartz, Swanson, Borum, & Wagner, 1998; Nicholls et al., 2004; Strand & Belfrage, 2001), some differences emerged with regard to HCR-20 ratings as well as the nature and targets of aggression across genders. Although findings are tentative at this point, it may be the case that the instrument more strongly predicts other-directed aggression for males and self-directed aggression for females. Additionally, the HCR-20 forecasted violent victimization experiences for both men and women in this sample.

Although the aforementioned findings yield important insights into the performance of the HCR-20 across genders, it would be fruitful to pursue more quantitative and qualitative investigations into the application of such violence risk assessments with men and women from various contexts (i.e., psychiatric, forensic, criminal). First, it may be worthwhile to examine whether, across genders, mental health professionals base their item-level ratings on equivalent information and whether they perceive the impact of each risk factor on violence risk to be comparable. Second, more extensive and detailed enquiries into the link between HCR-20 ratings and the severity, nature, and targets of male and female violence are needed in order to refine the pertinence of the management recommendations that result from HCR-20 evaluations. Third, the relationship between violence risk, as measured by the HCR-20, and vulnerability for violent victimization and/or deliberate self-injury should be explored further across genders. Moreover the dynamics underlying the co-occurrence and reciprocal influences of those negative outcomes must be clarified so that the intra- and

inter-individual contexts within which men and women inflict and/or incur harm may be better understood.

Finally, since the main goal of the structured professional approach to violence risk assessment is violence prevention, rather than the mere prediction of violent outcomes (Douglas & Kropp, 2002), it is the demonstrated reduction in harm both to the assessees and to those around them in connection with the results of the assessments that will speak to the success or inadequacy of these procedures. Consequently, it will be important to examine how the HCR-20 may best inform dynamic, individualized and effective interventions targeting the multifaceted, diverse and transactional nature of violence risk for both men and women.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Baron, R. M. & Kenny D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bartko, J. J. & Carpenter, W. T. (1976). On the methods of theory and reliability. *Journal of Nervous and Mental Disease, 163,* 307-317.

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

Bloom, H., Webster, C., Hucker, S. & De Freitas, K. (2005). The Canadian contribution to violence risk assessment: History and implications for current psychiatric practice. *The Canadian Journal of Psychiatry, 50*(1), 3-11.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist, 51,* 945-956.

Brennan, I. R., & Moore, S. C. (2009). Weapons and violence: A review of theory and research. *Aggression and Violent Behavior, 14,* 215-225.

Choe, J. Y., Teplin, L. A., & Abram, K. M. (2008). Perpetration of violence, violent victimization, and severe mental illness: Balancing public health concerns. *Psychiatric Services, 59,* 153-164.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86,* 127-137.

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

Crick, N. R., & Bigbee, M. A. (1998). Relational and overt forms of peer victimization: A multi-informant approach. *Journal of Consulting and Clinical Psychology, 66,* 337-347.

Daffern, M., & Howells, K. (2007). The prediction of imminent aggression and self-harm in personality disordered patients of a high security hospital using the HCR-20 clinical scale and the dynamic appraisal of situational aggression. *International Journal of Forensic mental Health, 6*(2), 137-143.

Dawes, M. D. (1979). The robust beauty of improper linear models. *American Psychologist, 34*(7), 571-582.

de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: A comparison with a matched sample of males. *Clinical Psychology and Psychotherapy, 12*, 226-240.

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

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3,* 149-165.

de Vogel, V., & de Vries Robbé, M. (under review). Working with women: Towards a more gender-sensitive violence risk assessment. In L. Johnstone & C. Logan (Eds.), *Managing clinical risk: A guide to effective practice*.

de Vogel, V., de Vries Robbé, M., van Kalmthout, W., & Place, C. (2010). *Female Additional Manual (FAM): Aanvullende richtlijnen bij de HCR-20 voor het inschatten van gewelddadig gedrag bij vrouwelijke (forensish) patiënten* [Additional guidelines to the HCR-20 for the assessment of violent behavior by women]. Utrecht, The Netherlands: Van der Hoeven Kliniek.

Douglas, K. S. (2006). V*iolence risk assessment covered ground and ground yet to cover.* Invited acceptance address for the Saleem Shah Award for Early Career Excellence in Psychology and Law given at the annual conference of the American Psychology-Law Society (Division 41 of the American Psychological Association), St. Petersbourg, Florida.

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

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

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

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical, 67,* 917-930.

Douglas, K. S., & Reeves, K. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. New York: Routledge/Taylor & Francis Group.

Douglas, K. S., & Webster, C. D. (1999). The HCR-20 Violence Risk Assessment Scheme: Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior, 26*(1), 3-19.

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

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32,* 479-510.

Evans, W. P., Marte, R. M., Betts, S., & Silliman, B. (2001). Adolescent suicide risk and peer-related violent behaviours and victimization. *Journal of Interpersonal Violence, 16*(12), 1330-1348.

Funk, S. J. (1999). Risk assessment for juveniles on probation. A focus on gender. *Criminal Justice and Behavior, 26,* 44-68.

Graham-Kevan, N. (2007). Power and control in relationship aggression. In J. Hamel & T. L. Nicholls (Eds), *A handbook of gender-inclusive theory and treatment: Family Interventions in domestic violence*. New York, NY: Springer Publishing Company.

Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial prediction of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 20,* 315-335.

Grann, M., & Långström, N. (2007). Actuarial assessment of risk for violence: To weigh or not to weigh? *Criminal Justice and Behavior, 34,* 22-36.

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey.* (Unpublished doctoral dissertation). Simon Fraser University, Burnaby, British Columbia, Canada.

Guy, L. S., & Douglas, K. S. (2006, June). *HCR-20 Violence Risk Assessment Scheme: Evaluating item bias with item response theory*. Paper presented at the 6th annual conference of the International Association of Forensic Mental Health Services, Amsterdam, The Netherlands.

Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

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

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

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3,* 121-137.

Hart, S. D. (2001). Assessing and managing violence risk. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 13-25). Burnaby, British Columbia/Tampa, FL: Simon Fraser University, Mental Health, Law, & Policy Institute/University of South Florida, Department of Mental Health Law and Policy, Florida Mental Health Institute.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). The Hare Psychopathy Checklist: Screening Version (PCL:SV). Toronto, Ontario, Canada: Multi-Health Systems. *Psychology, 67*, 917-930.

Hart, S, D., Douglas, K. S., & Webster, C. D. (2001). Risk management using the HCR-20: A general overview focusing on historical factors. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff (Eds.), *HCR-20 violence risk management companion guide* (pp. 27-40). Burnaby, British Columbia/Tampa, FL: Simon Fraser University, Mental Health, Law, & Policy Institute/University of South Florida, Department of Mental Health Law and Policy, Florida Mental Health Institute.

Hawton, K., & Harriss, L. (2008). The changing ratio in occurrence of deliberate self-harm across the lifecycle. *Crisis, 29*(1), 4-10.

Heilbrun, K., & Kramer, G. M. (2001). Update on risk assessment in mentally disordered populations. *Journal of Forensic Psychology Practice, 1*, 55-63.

Heilbrun, K., Yashuara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment*. New York, NY: Routledge/Taylor & Francis Group.

Hiday, V. A., Swartz, M. S., Swanson, J. W., Borum, R., & Wagner, H. R. (1998). Male-female differences in the setting and construction of violence among people with severe mental illness. *Social Psychiatry and Psychiatric Epidemiology, 33,* S68-S74.

Hillbrand, M. (2001). Homicide-suicide and other forms of co-occurring aggression against self and against others. *Professional Psychology: Research & Practice, 32,* 626-635.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Howells, K., Watt, B., Hall, G., & Baldwin, S. (1997). Developing programs for violent offenders. *Legal and Criminological Psychology, 2,* 117-128.

Kilpatrick, D. G., & Acierno, R. (2003). Mental health needs of crime victims: Epidemiology and outcomes. *Journal of Traumatic Stress, 16*(2), 119-132.

Lam, J. N., McNiel, D. E., & Binder, R. L. (2000). The relationship between patients' gender and violence leading to staff injuries. *Psychiatric Services, 51,* 1167-1170.

Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,*159-174.

Lauritsen, J. L., & Heimer, K. (2008). The gender gap in violent victimization, 1973-2004. *Journal of Quantitative Criminology, 24,* 125-147.

Levene, K. S., Augimeri, L. K., Pepler, D. J., Walsh, M. M., Webster, C. D., & Koegl, C. J. (2001). *Early assessment risk list for girls: EARL-21G. Version 1: Consultation version.* Toronto, Ontario, Canada: Earlscourt Child and Family Centre.

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*(2), 409-443.

Luke, D. A., & Homan, S. M. (1998). Time and change: Using survival analysis in clinical assessment and treatment evaluation. *Psychological Assessment, 10*(4), 360-378.

Manchak, S. M., Skeem, J. L., Douglas, K. S., Siranosian, M. (2009). Does gender moderate the predictive utility of the level of service inventory—Revised (LSI-R) for Serious Violent Offenders? *Criminal Justice and Behavior, 36*(5), 425-442.

McKeown, A. (in press). Female offenders: Assessment of risk in forensic settings. *Aggression and Violent Behavior.* Retrieved on September 19, 2010, from http://www.sciencedirect.com.proxy.lib.sfu.ca/science?_ob=ArticleListURL&_method=list&_ArticleListID=1466423167&_sort=r&_st=13&view=c&_acct=C000049301&_version=1&_urlVersion=0&_userid=955653&md5=c9a73962cab4ae4c1332ea75da491eaf&searchtype=a

McNiel, D., Gregory, A., Lam, J., Binder, R., & Sullivan, G. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology, 71*, 945-953.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., Roth, L. H., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of mental disorder and violence.* New York, NY: Oxford University Press.

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

Motz, A.(2008). The psychology of female violence (2nd ed.): Crimes against the body. New York, NY: Routledge.

Nicholls, T. L. (1997). *Comparing risk assessments with male and female psychiatric patients: The utility of the HCR-20 and PCL:SV* (Master's thesis, Simon Fraser University, Burnaby, British Columbia). Retrieved from http://ir.lib.sfu.ca/bitstream/1892/8714/1/b18913970.pdf

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

Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., & Martin, M.-L. (2006). The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric sample. *Assessment, 13*(3), 313-327.

Nicholls, T. L., Garcia-Mansilla, A., Rosenfeld, B. (2009). Risk assessment: Are current methods applicable to women? *International Journal of Forensic Mental Health, 8,* 50–61.

Nicholls, T. L., Greaves, C., & Moretti, M. (2008). Female aggression. In A. Jamieson & A. Moenssens (Eds.), *Wiley Encyclopedia of Forensic Science*. Mississauga, Ontario, Canada: Wiley.

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

Nikolova, N., Collins, M. J., Guy, L. S., Lavoie, J., Reeves, K., Wilson, C. M., & Douglas, K. S. (2006, March). *HCR-20 violence risk assessment scheme: A quantitative synthesis of its application reliability and validity*. Poster presented at the annual conference of the American Psychology-Law Society, St. Petersburg, FL.

Odgers, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior, 29*(1), 7-27.

Otto, R. K. (2000). Assessing and managing violent risk in outpatient settings. *Journal of Clinical Psychology, 56,* 1239-1262.

Paquette, J. A., & Underwood, M. K. (1999). Gender differences in young adolescents' experiences of peer victimization: Social and physical aggression. *Merrill-Palmer Quarterly, 45,* 242-266.

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

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737-748.

Schaap, G., Lammers, S., & de Vogel, V. (2009). Risk assessment in forensic psychiatric patients: A quasi-propspective study into the validity of the HCR-20 and the PCL-R. *The Journal of Forensic Psychiatry & Psychology, 20,* 354-365.

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial "prediction instruments"? *International Journal of Forensic Mental Health, 1,* 179-183.

Skeem, J. L., & Mulvey, E. P. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur Violence Risk Assessment Study. *Journal of Consulting and Clinical Psychology, 69,* 358-374.

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

Strand, S., & Belfrage, H., (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime & Law, 7,* 71-79.

Stueve, A., & Link, B. G. (1998). Gender differences in the relationship between mental illness and violence: Evidence from a community-based epidemiological study in Israel. *Social Psychiatry and Psychiatric Epidemiology, 33,* S61-S67.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics (4th. Ed.).* MA, USA: Allyn & Bacon.

Tengström, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry, 55,* 243-249.

Warren, J. I., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., & Van Patten, I. (2005). Understanding the risk factors for violence and criminality in women: The concurrent validity of the PCL-R and HCR-20. *International Journal of Law and Psychiatry, 28,* 269 -289.

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

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

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short-Term Assessment of Risk and Treatability (START). Version 1.0, Consultation edition*. Hamilton, Ontario/Port Coquitlam, British Columbia: St. Joseph's Healthcare/Forensic Psychiatric Services Commission.

Wood, J., & Edwards, K. (2005). Victimization of mentally ill patients living in the community: Is it a life-style issue? *Legal and Criminological Psychology, 10,* 279-290.