

Cross-Cultural Validity of Actuarial Risk Assessment Instruments for Individuals
in North America with a History of Sexual Offending: Static-99R and Static-2002R

by

Seung Chan Lee

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University

Ottawa, Ontario

© 2018

Seung Chan Lee

Abstract

Cultural bias in structured risk assessment instruments is an important concern given the overrepresentation of certain racial/ethnic minority groups in the criminal justice systems of Canada and the U.S. The goal of this dissertation was to evaluate the cultural bias in predictive accuracy of widely used actuarial risk assessment instruments, Static-99R and Static-2002R, for the overrepresented ethnic minority groups in the criminal justice system in Canada (Indigenous peoples) and the U.S. (Blacks and Hispanics).

Study 1 evaluated the predictive validity of Static-99R across three major ethnic groups (White, $n = 789$; Black, $n = 466$; Hispanic, $n = 719$) in the State of California. Static-99R was able to discriminate recidivists from non-recidivists among Whites, Blacks, and Hispanics with a history of sexual crimes (all area under the curve [AUC] values $> .70$; odds ratios > 1.39). Base rates (at a Static-99R score of 2) with a fixed 5-year follow-up across ethnic groups were very similar (2.4% - 3.0%) but were significantly lower than the norms (5.6%). The current findings support the use of Static-99R in risk assessment procedures for individuals of White, Black, and Hispanic heritage; however, Static-99R should be used with caution in estimating absolute sexual recidivism rates, particularly for Hispanics with a history of sexual crime, as it may overestimate the absolute recidivism rates.

Study 2 compares the characteristics and risk factors for non-Hispanic Whites ($n = 797$) and Blacks ($n = 788$) who had been convicted of a sexual crime in New Jersey, USA. The results indicated that Whites appeared more paraphilic, whereas Blacks displayed higher anti-sociality. Despite the differences, the Static-99R predicted equally well for both racial groups: Whites (AUC = .76) and Blacks (AUC = .78). The findings suggest that there may be opportunities to improve treatment for

the individuals at risk for sexual offending by tailoring interventions to the distinctive risk-relevant characteristics of Whites and Blacks.

Study 3 examined the predictive accuracy (discrimination and calibration) of Static-99R and Static-2002R for Whites ($n = 1,560$) and Indigenous individuals ($n = 653$) who have a history of sexual crimes from five independent Canadian samples. The results indicated that Static-99R predicted sexual recidivism with similar accuracy for the Indigenous and White study groups whereas Static-2002R predicted sexual recidivism only for Whites. In particular, the domains of persistence/paraphilia and youthful stranger aggression were not as predictive for the Indigenous study group as for White study group. In contrast, the general criminality domain predicted sexual recidivism for both Indigenous and White groups. Furthermore, the White study group showed more indicators of paraphilic interests whereas the Indigenous study group displayed higher general criminality. The findings suggest that the treatment for the individuals at risk for sexual offending may benefit from an increased focus on the distinctive risk-relevant characteristics of Whites and Indigenous peoples.

In conclusion, there is sufficient evidence to justify the use of the Static-99R for assessing the likelihood of sexual recidivism for certain ethnic minority groups (Blacks, Hispanics, and Indigenous peoples), who are overrepresented in the criminal justice systems in Canada and the U.S. (i.e., quantitative information). In contrast, for Static-2002R, the differences between Indigenous and Whites were big enough to make a difference; consequently, considerable caution is needed when interpreting the results of Static-2002R for individuals of Indigenous heritage.

Preface

This dissertation is an integrated article thesis, the core of which are three papers by Seung C. Lee that have been accepted for publication or submitted for peer review. Given that these papers had co-authors, we assert that this dissertation is an original and independent work by Seung C. Lee, who was fully involved in setting up and conducting the research, obtaining data and analyzing results, as well as writing the materials presented in the dissertation.

The three research papers are as follows:

Study 1 (Lee & Hanson, 2017) was published in *Criminal Justice and Behavior*. The final, published version is reproduced in this dissertation with only minor edits

Lee, S. C., & Hanson, R. K. (2017). Similar Predictive Accuracy of the Static-99R Risk Tool For White, Black, and Hispanic Sex Offenders in California. *Criminal Justice and Behavior*, *44*, 1125-1140. doi: 10.1177/0093854817711477

Study 2 (Lee, Hanson, Calkins & Jeglic, in press) has been accepted for publication pending minor revisions in *Sexual Abuse*. The dissertation includes the revised version of the study that has benefited from the editorial feedback on the initial submission. The revision was still under editorial review as of October, 2018

Lee, S. C., Hanson, R. K., Calkins, C., & Jeglic, E. (in press). Paraphilia and Antisociality: Motivations for Sexual Offending May Differ for American Whites and Blacks. *Sexual Abuse*.

Study 3 (Lee, Hanson & Blais, 2018) was submitted for peer review for the first time in October, 2018. The dissertation version is the same as the initial submission, except for nonconsequential edits.

Lee, S. C., Hanson, R. K., & Blais, J. (2018). *Predictive Accuracy of the Static-99R and Static-2002R Risk Tools for Identifying Indigenous and White Individuals at High Risk for Sexual Recidivism in Canada*. Manuscript submitted for publication.

Co-author contributions

R. Karl Hanson (thesis supervisor; co-author) was the thesis supervisor and co-author on Study 1, Study 2, and Study 3. He helped shape Seung Lee's ideas on ethnic bias in risk assessment into a manageable set of research studies for this dissertation. Dr. Hanson also provided overall guidance concerning research designs, statistical strategies, and the reporting of the findings. In Study 2, Cynthia Calkins and Elizabeth Jeglic (co-authors) supervised the original data coding procedures and reviewed the complete manuscript before submission. In Study 3, Julie Blais (co-author) contributed to the literature review and the editing/revision of the manuscript as it progressed towards completion.

Each of the individual research projects that comprise this thesis received research ethics approval from the Carleton University's Research Ethics Boards (CUREB)-B as below,

[Study 1]

File No: 108685

Project Title: "The Predictive Validity of the Static-99R Risk Assessment Scale in the State of California"

[Study 2]

File No: 109215

Project Title: “Motivations for sexual offending for American Whites and Blacks
(New Jersey)”

[Study 3]

File No: 109216

Project Title: “Sexual recidivism risk factors for Canadian Indigenous and Whites”

The articles published in SAGE journals (e.g., *Criminal Justice and Behavior* [Study 1] and *Sexual Abuse* [Study 2]) are allowed to be used in unpublished dissertations without requiring separate permission from SAGE (*Guidelines for SAGE Authors*).

Seung C. Lee, M.A.

Ph.D. Candidate

Psychology Department

Carleton University

R. Karl Hanson, Ph.D.

Thesis Supervisor

Adjunct Research Professor

Psychology Department

Carleton University

Acknowledgments

First and foremost, I would like to sincerely thank my supervisor and mentor, Karl Hanson, for his patience, guidance, support, and encouragement that he has invested in mentoring me throughout my doctorate. Karl, I will never forget the moment we first met in 2013 at the ATSA conference in Chicago. Those five-minutes have changed my life. I also won't forget the time we had at Public Safety and in the library at Carleton University over 4 years. I feel blessed for all the opportunities working with you. I would also like to thank my co-supervisor, Adelle Forth, for her support and kindness. Adelle, I am incredibly fortunate to work with you. Without your support and kindness, I would not make this long journey. Thank you again for all the effort and support you have made for me.

I would like to acknowledge my committee members, Shelley Brown and Julie Blais, who have been very supportive and helpful throughout this process. I would also like to thank my internal examiner, Lara Karaian, and external examiner, Sandy Jung, for agreeing to be my committee members and to evaluate this dissertation. All the feedback from each of you has been invaluable to improve the quality of this dissertation.

I have been very fortunate to work for the California State Committee on Sex Offender Risk (SARATSO). This opportunity made me possible to access to the data (Study 1). I am grateful to Janet Neeley for the opportunity and the data provided. I also thank you, Elizabeth Jeglic and Cynthia Calkins, for your support in providing access to the data file (Study 2). It would also like to thank the following researchers for permitting us to use their data (Study 3): Jim Bonta, Andy Haag, and Mark Olver. I would also like to thank all the people who have taken the time to review each study and provided helpful feedback: Janet Neeley, Daryl Kroner, David D'Amora, Leticia

Gutierrez, and Patti McDonald. Your contributions on the overall narratives and interpretations of the research results for Blacks in the U.S. (David D'Amora) and Indigenous peoples in Canada (Leticia Gutierrez and Patti McDonald) helped us consider in the manuscript the lived experience of these racial/ethnic groups.

I would like to thank my colleagues and friends who have supported my work throughout my Ph.D. student life: Kelly Babchishin, Leticia Gutierrez, Nick Chadwick, Maaike Helmus, Andrew Brankley, Kimberly Mularczyk, Elisabeth Leroux, Rebeca Mugford, and Kayla Wanamaker. With your support and encouragement, I was able to get through this. Please remember I appreciate that you are my colleagues and friends. Thank you all.

I would like to thank my family in South Korea – particularly my Dad (Kwang Bok Lee), Mom (Ok Sun Choi), my older sister (Ji Eun Lee) and brother (Seung Man Lee) – for your love and prayers throughout my life. I love you so much. Finally, I would like to thank my wife, Boram Kim, for her unconditional love and confidence in me. Boram, it is my greatest fortune to have met you in my life. Thank you, and I love you. To my two precious sons, Raymond and Jesse, you guys have always disrupted my work at home and have never cooperated to finish up my dissertation. But, I still love you so much. I cannot imagine my life without you two boys. Thank you for coming to us.

Table of Contents

Abstract.....	i
Preface.....	iii
Acknowledgments.....	vi
Table of Contents	viii
List of Tables.....	xv
List of Figures	xviii
List of Appendices	xix
Chapter 1: Study Background and Literature Review	2
History of Criminal Management Principles and Practices.....	2
The Advent of Structured Risk Assessment Instruments.....	7
Actuarial.....	8
Structured Professional Judgment (SPJ).....	9
Four generations of structured risk assessment instruments	10
Structured Risk Assessment Instruments for Sex Offence Recidivism	12
Current Uses of Structured Risk Assessment Instruments for Sex Offenders.....	15
Criminal sentencing and corrections	15
Civil commitment (SVP) and Dangerous Offender legislation (DO).....	16
Sex offender registration/notification and National Sex Offender Registry	17
Static-99R and Static-2002R	19
Summary.....	21
Cross-Cultural Generalizability of Psychological Testing Instruments.....	21
Racial/Ethnic Minority Groups in Canada and the U.S.....	25
Blacks in the U.S.....	25

Hispanics in the U.S.26

Indigenous peoples in Canada.....27

Overrepresentation of racial/ethnic groups in the criminal justice system29

Cultural Bias in Structured Risk Assessment Instruments30

 Possible explanations of cultural bias in predictive validity34

 Previous Studies Evaluating Cultural bias in Static-99R and Static-2002R.....37

Overview of Purpose40

Chapter 2: Method42

 Participants42

 Study 1.....42

 Study 2.....42

 Study 3.....42

 Measures43

 Static-99R.....43

 Static-2002R.....43

 Analytical Strategy44

 Discrimination.....44

 Area Under the Curve (AUC).44

 Odds ratios45

 Calibration.....45

 E/O index45

 Comparing Logistic regression parameters45

Chapter 3: Study 147

Similar Predictive Accuracy of the Static-99R Risk Tool for White, Black, and Hispanic Sex Offenders in California47

 Current Study54

Method.....54

 Sample.....54

 Measures.....55

 Static-99R55

 Rater Reliability55

 Recidivism55

 Procedure.....56

 Plan of Analysis.....56

 Area Under the Curve (AUC).....56

 Odds ratios57

 E/O index57

 Comparing Logistic regression parameters57

Results58

 Discrimination.....58

 Calibration.....61

Discussion.....64

 Limitations67

 Implications for Research.....67

 Implications for Practice68

Conclusion.....69

Chapter 4: Study 271

Paraphilia and Antisociality: Motivations for Sexual Offending May Differ for

American Whites and Blacks	71
Racial Differences in Sex Offending	72
Predictive Validity of Risk Assessment Instruments across Racial Groups	75
Current Study	79
Method	79
Sample	79
Measures	80
Static-99R	80
MnSOST-R	80
Pervasive Anger Scale	81
General Criminality Scale	81
Sexual Criminality Scale	82
Paraphilia Scale	82
Sexual Recidivism	82
Procedure	83
Plan of Analysis	83
Comparing the risk-relevant characteristics	83
Predictive validity of Static-99R	84
Area Under the Curve (AUC)	84
Odds ratios	84
E/O index	85
Comparing Logistic regression parameters	85
Results	85
Comparing Risk Relevant Characteristics	86

Predictive Validity of Static-99R within a Fixed 5-Year Follow-Up Period.....	96
Discrimination.....	97
Calibration.....	99
Discussion.....	103
Limitations	108
Implications for Research.....	109
Implications for Practice	110
Chapter 5: Study 3	112
Predictive Accuracy of the Static-99R and Static-2002R Risk Tools for Identifying Indigenous and White Individuals at High Risk for Sexual Recidivism in Canada ..	112
Indigenous Overrepresentation in Canadian Corrections.....	113
Risk Assessment and Cultural Bias	116
Risk Assessment with Indigenous Individuals with a History of Sexual Crimes...	118
Current Study.....	121
Method.....	123
Sample.....	123
Bonta & Yessine (2005).	123
Haag (2005)	124
Olver et al. (2018).....	124
Brankley et al. (2017a).....	124
Lee et al. (2018).....	125
Measures.....	125
Static-99R	125
Static-2002R	126

Sexual recidivism.....	126
Psychological constructs of Static-99/R and Static-2002R	126
Plan of Analysis.....	127
Aggregation of findings	127
Comparing the risk-relevant characteristics.....	127
Predictive validity of Static-99R and Static-2002R.....	128
Area Under the Curve (AUC).....	128
Odds ratios.....	129
Comparing logistic regression parameters	129
Results	130
Comparing Risk Relevant Characteristics.....	131
Predictive Accuracy of Three Psychological Structures and Items.....	139
Predictive Validity Static-99R and Static-2002R	143
Static-99R	143
Discrimination	143
Calibration	150
Static-2002R	150
Discrimination	150
Calibration	155
Discussion.....	157
Predictive Accuracy of Static-99R and Static-2002R	157
Psychological Constructs	159
Cultural bias in Static-99R and Static-2002R for Indigenous Canadians	161
Limitations	164
Implications for Research.....	165

Implications for Policy and Practice	167
Conclusion	168
Chapter 6: General Discussion.....	170
How Much Evidence is Sufficient to Detect Cultural Bias?	171
Why Are There Low Average Scores and Low Sexual Recidivism Rates for Hispanics?.....	175
Implications for Future Research	177
Conclusions	179
References.....	181
APPENDIX A	229
APPENDIX B	230
APPENDIX C	232
APPENDIX D.....	233
APPENDIX E	234

List of Tables

Table 1.1: Summary description of previous studies.....	53
Table 1.2: Five-year Sexual Recidivism Rates, Static-99R Scores, and AUC values for Whites, Blacks, and Hispanics.....	59
Table 1.3: Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism across Different Racial/Ethnic Groups.....	60
Table 1.4: Comparison of Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism with Meta-Average (from Hanson et al., 2016).....	61
Table 1.5: Comparison of Expected and Observed 5-year Sexual Recidivism Rates for Static-99R Standardized Risk Levels	63
Table 2.1: Predictive Accuracy of Static-99R for Blacks and Whites.....	78
Table 2.2: General Demographics.....	87
Table 2.3: Socioeconomic Status (Income, Employment, Education).....	88
Table 2.4: Early Childhood Adversity	89
Table 2.5: Hostility (Sadism, Anger, Impulsivity).....	91
Table 2.6: Paraphilia.....	92
Table 2.7: General Criminality.....	93
Table 2.8: Sexual Criminality.....	94
Table 2.9: Mental Health.....	95
Table 2.10: Five-year Sexual Recidivism Rates, Static-99R Scores, and AUC Values for Whites and Blacks.....	97
Table 2.11: Predictive Accuracy (AUC values) of General Criminality, Sexual Criminality, and Paraphilia Dimensions for Whites and Blacks.....	98
Table 2.12: Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism for Whites and Blacks.....	99

Table 2.13: Comparison of Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism with Meta-Average (from Hanson et al., 2016)..... 100

Table 2.14: Comparison of Expected and Observed 5-year Sexual Recidivism Rates for Static-99R Standardized Risk Levels..... 101

Table 3.1: Age at Release, Static-99R Scores, and Static-2002R Scores for Indigenous and White Study Groups..... 132

Table 3.2: Differences in Static-99R and Static-2002R Items between Indigenous and White Study Groups..... 133

Table 3.3: Predictive Accuracy of Static-99R and Static-2002R items for Sexual Recidivism for Indigenous and White Study Groups..... 140

Table 3.4: Sexual Recidivism Rates and AUC values for Static-99R for Indigenous and White Study Groups..... 144

Table 3.5: Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism for Indigenous Study Group..... 145

Table 3.6: Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism for White Study Group..... 146

Table 3.7: Comparison of Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism with Meta-Average (from Hanson et al., 2016)..... 148

Table 3.8: Sexual Recidivism Rates and AUC values for Static-2002R for Indigenous and White Study Groups..... 151

Table 3.9: Logistic Regression Parameters for Static-2002R Predicting Sexual Recidivism for Indigenous Study Group..... 152

Table 3.10: Logistic Regression Parameters for Static-2002R Predicting Sexual Recidivism for White Study Group..... 153

Table 3.11: Comparison of Logistic Regression Parameters for Static-2002R

Predicting Sexual Recidivism with Meta-Average (from Hanson et al., 2016).....156

List of Figures

Figure 1.1: Logistic curve for the overall sample with the norms.....	60
Figure 1.2: Logistic curves for each ethnic group with the norms.....	62
Figure 2.1: Distributions of White and Black racial groups across risk levels.....	86
Figure 2.2: AUCs of Static-99R for White and Black racial groups.....	98
Figure 2.3: Observed and expected recidivism rates based on Static-99R 5-year sexual recidivism rates.....	102
Figure 2.4: Logistic curves for White and Black racial groups with the norms.....	103
Figure 3.1: Logistic curves of Static-99R for Indigenous and White study groups with the High Risk/High Need Norms with a fixed 5-year follow-up.....	149
Figure 3.2: Logistic curves of Static-99R for Indigenous and White study groups with the High Risk/High Need Norms with a fixed 10-year follow-up.....	149
Figure 3.3: Logistic curves of Static-2002R for Indigenous and White Study groups with the High Risk/High Need Norms with a fixed 5-year follow-up.....	155

List of Appendices

Appendix A: Regional Distribution of the Samples (Study 3).....	229
Appendix B: Comparison of logistic regression parameters of Static-99R between Indigenous and White groups (Study 3).....	230
Appendix C: Comparison of logistic regression parameters of Static-2002R between Indigenous and White groups (Study 3).....	232
Appendix D: Static-99R - TALLY SHEET.....	233
Appendix E: Static-2002R - TALLY SHEET.....	234

Cross-Cultural Validity of Actuarial Risk Assessment Instruments for Individuals in North America with a History of Sexual Offending: Static-99R and Static-2002R

Understanding an offender's likelihood to reoffend is vital to making accurate decisions within the criminal justice systems (e.g., informing sentencing and conditional release, determining the intensity of rehabilitation efforts). For the last few decades, many structured risk assessment instruments have been developed and used to estimate the likelihood of reoffending in Western countries. Considerable research has supported the good predictive accuracy of those instruments for different types of offences (e.g., violent, sexual, or general offences; Campbell, French, & Gendreau, 2009; Hanson & Morton-Bourgon, 2009; Tully, Chou, & Browne, 2013; Yang, Wong, & Coid, 2010).

Nevertheless, they are not without their critics. One important line of criticism concerns their potential for cultural bias (Angwin, Larson, Mattu, & Kirchner, 2016; Day, Tamatea, Casey, & Geia, 2018; Haag, Boyes, Cheng, MacNeil, & Wirove, 2016; Hart, 2016; Kehl, Guo, & Kessler, 2017; Shepherd & Lewis-Fernandez, 2016). These debates have continued in a relative absence of evidence, given that the topic has drawn relatively little research attention. The potential for cultural bias in structured risk assessment instruments is a crucial concern for racial/ethnic minority groups who are already overrepresented in the criminal justice systems.

The primary purpose of this dissertation was to evaluate cultural bias in the Static-99R and Static-2002R actuarial risk assessment instruments for individuals with a history of sex offending. The studies focused on these minority groups because they are the most overrepresented in their jurisdictions: Blacks and Hispanics in the U.S., and Indigenous peoples in Canada. As well, this dissertation compares the risk-relevant characteristics of the racial/ethnic minority groups to those of Whites (e.g.,

general criminality and sexual criminality).

Chapter 1 reviews the background history of criminal management principles and structured risk assessment instruments. It also addresses the rationales for evaluating cultural bias in structured risk assessment instruments across different racial/ethnic groups. Chapter 2 briefly reviews the samples, measures, and statistical strategies. Chapter 3 to 5 present three studies examining cultural bias in Static-99R and Static-2002R with different ethnic groups. Chapter 6 provides a general discussion of the results, comments on the implications of the findings, and proposes directions for future research.

Chapter 1: Study Background and Literature Review

History of Criminal Management Principles and Practices

Since the 1970s, the primary goal of criminal sanctions in North America has gone back and forth between retribution/deterrence and rehabilitation/reintegration. This debate was heavily influenced by Martinson's (1974) early work. He questioned and examined the effectiveness of rehabilitation efforts in reducing the potential risk of reoffending and reached the conclusion that none had a meaningful effect on recidivism (i.e., "nothing works"; Martinson, 1974). Martinson's (1974) movement was supported by Von Hirsch (1976) who asserted that the objective of corrections was not crime prevention, rehabilitation, or even deterrence, but rather to provide retribution that was proportional to the crime committed ("just deserts" theory). Consequently, the fundamental rationale for probation and parole was questioned and opportunities for individuals to serve sentences in the community were reduced. Furthermore, the indeterminate sentencing schemes common at that time ("e.g., two to five years") began to shift towards determinate sentencing, which denied

correctional services discretion over the length of the sentence.

Martinson's review and subsequent endorsement by opinion leaders served as the intellectual foundation of the "tough on crime" era. Since then, criminal sanctions, particularly in the United States (U.S.), have become harsher and stricter (Turner, Greenwood, Chen, & Fain, 1999). For example, the "three strikes and you're out law," which is currently enacted in some form by 25 American states and the federal justice system (Chen, 2008), mandates significant sentences (usually 25 years to life in prison) for offenders who are convicted of a third felony. Similarly, "truth-in-sentencing," which has been adopted by almost 40 states since 1998 (Ditton & Wilson, 1999), refers to the requirement that offenders must serve a substantial proportion (e.g., 85% to 100%) of their imposed sentence incarcerated as opposed to on parole (Holt, 1998).

Although the "tough on crime" strategy was meant to prevent crimes, it has largely failed to do so. First, there has been no clear evidence supporting that strict sanctions reduce crime rates or recidivism rates (Chen, 2008; Doob & Webster, 2003; Smith, Goggin, & Gendreau, 2002; Turner et al., 1999). Intensive correctional sanctions (e.g., boot camps, longer and intense incarceration) do not inhibit subsequent criminal behaviours (Aos, Phipps, Barnoski, & Lieb, 2001; MacKenzie, Wilson, & Kider, 2001; Petrosino, 1997; Smith et al., 2002). Furthermore, there is evidence suggesting that increasing the severity of penalties can even increase recidivism rates (Andrews & Bonta, 2010). The gross failure of the "tough on crime" strategy is also evidenced by the tremendous growth in incarceration rates in the U.S. over the last 30 years (Bureau of Justice Statistics, 1987; Guerino, Harrison, & Sabol, 2011).

During the "tough on crime" era of the 1970s and 1980s, new research began

to emerge supporting the effectiveness of offender rehabilitation across a variety of settings and with diverse types of offenders (Gendreau & Ross, 1979, 1987). The advent of a new statistical method, meta-analysis, in the 1980s allowed researchers to aggregate the results of individual studies to identify trends otherwise hidden in the data. During the past few decades, meta-analytic reviews have provided substantial evidence of the positive effects of rehabilitation in reducing recidivism for juveniles and adults, and for males and females (e.g., Andrews, Zinger, et al., 1990; Garrett, 1985; Gobeil, Blanchette, & Stewart, 2016; Lipsey & Wilson, 1998; Petrosino, 1997; Whitehead & Lab, 1989) and for different types of offences (e.g., general offences, sexual offences, violent offences; Babcock, Green, & Robie, 2004; Dowden & Andrews, 2000; Hall, 1995; Hanson, Bourgon, Helmus, & Hodgson, 2009; Schmucker & Lösel, 2015).

For many years, the topic of correctional rehabilitation has been debated from contrasting perspectives aligned with either sociology or with psychology (Andrews & Bonta, 1994). Academic criminologists often took a sociological approach that attributed the cause of crime to social structure (e.g., poverty, social inequality). From this perspective, psychological treatment should not work because it does not address the root causes. Furthermore, resources devoted to offender rehabilitation divert energy from the crucial projects of social change. In contrast, the psychological approach (e.g., social learning and social cognition theory) emphasizes the importance of individual change. Psychological approaches typically focus on understanding the criminal conduct at the individual level (e.g., risk and need factors) and promote interventions addressing the attitudes, values, social relationships, and lifestyle supportive of crime.

From a social psychological perspective, Andrews and colleagues proposed

three principles at the core of effective correctional interventions: the risk, need, and responsivity principles (R-N-R; Andrews, Bonta, & Hoge, 1990). According to the R-N-R model, the most effective interventions are those that provide the most service to the highest risk offenders (Risk), while targetting criminogenic needs (Need), and delivering the treatment in a manner suited to the offender's culture and learning style (Responsivity; Andrews & Bonta, 2010; Andrews, Bonta, et al., 1990; Andrews & Dowden, 2006).

Whereas harsher sanctions have resulted in no effects or adverse outcomes, rehabilitation programs that adhere to the core principles of the R-N-R result in substantial reductions in recidivism. When offender treatment adheres to all three of the R-N-R principles, recidivism rates decrease by up to 35% (Andrews & Bonta, 2010). In contrast, treatment programs that fail to adhere to any of the R-N-R principles show a small increase in recidivism (Andrews & Bonta, 2010). The R-N-R principles apply to intervention programs for different types of offending (e.g., violent offending [Dowden & Andrews, 2000]; prison misconducts [French & Gendreau, 2006]; sexual offending [Hanson et al., 2009]) and in different settings (e.g., community and institutions; Andrews & Bonta, 2010).

In alignment with values common in Westernized democracies, the fundamental purposes of Canadian corrections are to 1) re-integrate individuals who committed crimes into society (i.e., rehabilitation), and 2) promote public safety. Consider the preamble of the Corrections and Conditional Release Act (1992/2012) in Canada,

3(b) The purpose of the federal correctional system is to contribute to the maintenance of a just, peaceful and safe society by assisting the rehabilitation of offenders and their reintegration into the community as law-abiding citizens

through the provision of programs in penitentiaries and in the community (*Corrections and Conditional Release Act*, 1992, c. 20), and

3.1 The protection of society is the paramount consideration for the Service in the corrections process. (*Corrections and Conditional Release Act*, 2012, c. 1, s. 54).

Similarly, according to the Sentencing Reform and Corrections Act (2015) in the U.S.,

Sec. 202. DOJ must review existing recidivism reduction programs and productive activities (e.g., a prison work program). The Bureau of Prisons (BOP) must expand offerings to all eligible prisoners. Certain prisoners who successfully complete a recidivism reduction program or productive activity are eligible to earn time credits and other incentives (e.g., additional telephone or visitation privileges).

Sec. 203. DOJ must develop the Post-Sentencing Risk and Needs Assessment System for use by the BOP to assess prisoner recidivism and violence risk and ensure appropriate housing, grouping, and program assignments.

Sec. 204. The bill amends the federal criminal code to allow pre-release custody for an additional period of time equal to a prisoner's earned time credits for successful completion of recidivism reduction programs or productive activities. A prisoner may serve the additional period of pre-release custody in a residential reentry center, on home confinement, or on community supervision (*Sentencing Reform and Corrections Act*, 2015, S.2123).

Risk assessment is necessary to accomplish both fundamental purposes (i.e., rehabilitation of offenders and public safety). It is only by knowing who is at risk for reoffending and by knowing why they are at risk, that the risk can be effectively

managed. During the last few decades, structured risk assessments have been developed to identify those needing the most intensive interventions (Risk-principle), and to identify the factors that, when changed, reduce the likelihood of criminal recidivism (i.e., criminogenic needs; Need-principle). Consequently, a core function of recidivism risk instruments is guiding interventions intended to reduce the likelihood of reoffending (Heilbrun, 1997).

The Advent of Structured Risk Assessment Instruments

The first step in successful rehabilitation programs is to understand the level of risk posed by the individuals served. Recidivism risk assessment estimates the likelihood of recidivism based on the consideration of intermediate indicators of risk, such as substance use and association with individuals involved with crime (Hanson, 2009). The risk factors need not be problematic in themselves; instead, the risk factors are concerns because they increase or decrease the likelihood of problems in the future.

The factors considered in risk assessment have changed over several decades. Prior to the 1950s, the evaluation of offender risk was conducted based solely on the unstructured judgment of correctional staff (e.g., probation officers) and clinical professionals (e.g., psychologists, psychiatrists, and social workers). Guided by their experiences and intuition, professionals would make judgments as to who was at high risk of reoffending and would then assign them differing levels of security and supervision. Although this method was flexible and specific to individual offenders, it is now widely accepted that more structured approaches to risk assessment provide more accurate risk assessments (Ægisdóttir et al., 2007; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009).

The most accurate structured approaches to risk assessment consider only

empirically identified characteristics (i.e., risk factors) associated with the relevant outcomes (e.g., violent or sexual recidivism). There are two broad categories of risk factors: *static* and *dynamic* risk factors. *Static* risk factors are relatively fixed aspects of the individual's life, such as demographic (e.g., age) and criminal history variables (e.g., the total number of previous convictions) that are not amenable to deliberate intervention. In contrast, *dynamic* risk factors (i.e., criminogenic needs) are psychological or behavioural features that are potentially changeable factors (e.g., attitudes, cognition, impulsivity, and self-control) that when addressed reduce the likelihood of reoffending. The dynamic risk factors can be further divided into stable risk factors (relatively enduring characteristics; e.g., personality disorders) and acute risk factors (rapidly changing features; e.g., emotional collapse; Hanson & Harris, 2000).

Despite a reasonably strong consensus that structured approaches are superior to unstructured approaches, there is still considerable debate as to how risk assessments should be structured (Monahan, 2007). The two main approaches to structured risk assessment are 1) actuarial risk assessment and 2) structured professional judgment (SPJ).

Actuarial. Actuarial assessment is “a formal method... *that* uses an equation, a formula, a graph, or an actuarial table to arrive at a probability, or expected value, of some outcome” (Grove & Meehl, 1996, p. 294). Actuarial risk assessment considers individual items (i.e., risk factors or predictors) associated with the results (e.g., recidivism) and combines them into an overall risk instrument. Each item is given a score (e.g., 0 or 1). The scores on the items can then be summed, and the total score is linked to the estimated recidivism probability. In an early meta-analysis, Grove and Meehl (1996) reported that this actuarial method was almost invariably equal to or

superior to the unstructured clinical method.

Although it was not widely adopted, the earliest documented actuarial risk instrument in corrections was Burgess's (1928) instrument designed to assess the likelihood of success or failure on parole. Burgess identified 21 risk factors associated with success or failure on parole (e.g., marital status; previous criminal record; nature of the offence; age at the time of parole; type of neighborhood; psychiatric prognosis) with 3,000 parolees from the Illinois State Penitentiary with at least a two-and-a-half-year follow-up period. Total scores of the items were associated with expected rates of parole violation ranging from 1.5% to 76%.

Despite Burgess' early success, actuarial risk assessment instruments were not widely adopted by corrections agencies until the 1970s. Those instruments were used to make release decisions and largely replaced unstructured professional judgments. For example, the Salient Factors Score (SFS; Hoffman, 1994; Hoffman & Beck, 1974), which consists of 6 items, has been used since 1973 to assess parole releases from U.S. federal prisons. During the 1980s, the Correctional Service of Canada (CSC) adopted the Statistical Information on Recidivism Scale (SIR; Nuffield, 1982) to assess the risk of reoffending for male offenders. The Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993) is a typical example of actuarial instruments that empirically derive combinations of clinical variables to predict long-term violent recidivism for adult offenders. Although it was not developed as a risk assessment instrument, the Psychopathy Checklist-Revised (PCL-R; Hare, 1980, 2003) has also contributed to the modern era of psychological risk assessment, particularly in forensic mental health.

Structured Professional Judgment (SPJ). Structured Professional Judgment (SPJ) has a relatively short history compared to traditional actuarial instruments. SPJ

involves the assessment of specified risk factors that are usually derived from a broad review of the literature rather than from a particular data set. Evaluators complete an SPJ instrument by rating all the specified factors based on information gained through interviews, collateral interviews, official records, and other sources of information. Then, when estimating overall risk, evaluators are asked to consider the presence of the risk factors and the anticipated intensity of management (e.g., supervision or treatment) needs in reaching an overall conclusion. This risk judgment is typically expressed as “low,” “moderate,” or “high” (Otto & Douglas, 2015). No attempt is made to link these risk categorizations to estimated probabilities of outcomes. Typical examples of SPJ instruments are the Historical Clinical Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) for adult offenders, the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002) for adolescent offenders, and the Spousal Assault Risk Assessment Guide (SARA; Kropp, Hart, Webster, & Eaves, 1995) for intimate partner offenders.

Four generations of structured risk assessment instruments. Another approach to classifying risk assessment instruments is generationally (Andrews, Bonta, & Wormith, 2006). The first generation represents unstructured professional judgments, assessing the risk of reoffending based solely on clinical and professional experience and intuition. As highlighted above, there is strong evidence that this first-generation risk assessment procedure is not as reliable and accurate as the actuarial approaches (Ægisdóttir et al., 2007; Grove et al., 2000).

Second generation risk instruments consist mainly of empirically derived static risk factors (e.g., age, criminal history). In these instruments, there are explicit rules for scoring individual items, and there is typically some algorithm for obtaining a total score (e.g., VRAG; Harris et al., 1993). Although second-generation

instruments have demonstrated overall good predictive accuracy (Andrews et al., 2006), they provide limited information as to what needs to be done to reduce offender risk (i.e., criminogenic needs; Need principle for R-N-R model; Heilbrun, 1997). In response to this limitation, risk assessment instruments evolved in the 1970s and 1980s to include dynamic risk factors (i.e., third-generation).

Third-generation risk instruments are sensitive to changes in an offender's circumstances and also provide correctional staff with information as to what offender needs should be targeted by their interventions (e.g., Level of Service Inventory-Revised; Andrews & Bonta, 1995). Finally, fourth-generation instruments not only include static and dynamic factors but also integrate the assessment with a case management plan. These new risk assessment instruments aim to improve intervention and to monitor a broad range of risk and personal factors that are important to intervention (Andrews et al., 2006; e.g., Level of Service/Case Management Inventory [LC/CMI], Andrews, Bonta, & Wormith, 2004).

SPJ instruments do not fit into any particular generation category. Although they include static and/or dynamic risk factors with specific scoring rules, the final risk decision is left to the professional (i.e., it is not actuarial). It is argued that this professional judgment approach is more appropriate than the strictly statistical approach for evaluating individuals because it takes into account the distinctive characteristics of the offender (Webster et al., 1997). At the same time, SPJ is criticized for being too similar to first-generation unstructured clinical judgment (Andrews et al., 2006).

Several studies have examined and compared the predictive accuracy of actuarial instruments and SPJ instruments. Those findings are consistent with good predictive accuracy for both types of instruments, and the differences of predictive

accuracy are negligible (Catchpole & Gretton, 2003; Dahle, 2006; de Vogel, de Ruiter, van Beek, & Mead, 2004; Douglas, Yeomans, & Boer, 2005).

Structured Risk Assessment Instruments for Sex Offence Recidivism

Measuring the actual sex offender recidivism rate is challenging because only a small portion of sexual crimes is reported (Tjaden & Thoennes, 2006).

Consequently, the observed sexual recidivism rates underestimate the actual rate of reoffending. Nonetheless, the observed sexual recidivism rate (any offence that was considered sexually motivated) of individuals with a history of sexual crime is less than what is believed by the general public, usually ranging from 5%-15% after 5 years (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Hanson, Thornton, Helmus, & Babchishin, 2016; Helmus, Hanson, Thornton, Babchishin, & Harris, 2012).

Sexual recidivism risk is positively associated with deviant sexual interests (e.g., pedophilia) as well as with anti-social propensity (e.g., impulsivity; Hanson & Bussière, 1998). In contrast, violent or general recidivism is well modeled by the single dimension of anti-social propensity (e.g., VRAG; Harris et al., 1993). These findings highlight the need for risk assessment instruments that take into account the relatively low recidivism rates and unique risk factors/propensity of individuals with a history of sexual crime.

Empirical studies have identified specific factors associated with sexual recidivism (Hanson & Bussière, 1998; Hanson, Harris, Scott & Helmus, 2007; Hanson & Morton-Bourgon, 2004, 2005; Knight & Thornton, 2007). Examples of empirically supported static risk factors for sexual recidivism include those related to prior criminal history (e.g., age, marital status, unrelated or stranger victim, previous sexual offences; Hanson & Bussière, 1998). Examples of dynamic risk factors include

sexual preoccupation, emotional congruence with children, lack of emotional intimacy (e.g., never married), general self-regulation problems (e.g., impulsivity), and resistance to rules and supervision (e.g., noncompliance with supervision; Mann, Hanson & Thornton, 2010). Acute risk factors include access to victims, sexual preoccupation, substance abuse, and the collapse of social supports (Hanson & Harris, 2000; Hanson et al., 2007).

Risk assessment instruments for sexual offenders have been developed by combining the items representing these risk factors into an overall assessment of recidivism risk (Hanson & Morton-Bourgon, 2009). The first widely used actuarial risk assessment instrument for sexual recidivism risk (second generation) was the Rapid Risk Assessment for Sex Offence Recidivism (RRASOR; Hanson, 1997). Shortly after this, Hanson and Thornton (2000) developed the Static-99, which contained all of the RRASOR items and six additional static items from Structured Anchored Clinical Judgment (SACJ-Min; Grubin, 1997), later updated to Static-99R, which has a re-weighted age item (Helmus, Thornton, Hanson, & Babchishin, 2012). Other early actuarial risk assessment instruments for sexual offenders are the Static-2002 (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 2006), which is a modified version of the VRAG (Harris et al., 1993), and the MnSOST (Epperson et al., 1998).

Third generation actuarial instruments for sexual offenders, which include sex-offence specific dynamic risk factors, include the Sex Offender Need Assessment Rating (SONAR; Hanson & Harris, 2000), the STABLE-2000/2007 and ACUTE-2000/2007 (Fernandez, Harris, Hanson, & Sparks, 2014; Hanson et al., 2007), and the Violence Risk Scale-Sex Offender Version (VRS-SO; Olver, 2003). In addition, SPJ

instruments for sex offenders have also been developed. These mostly include dynamic risk factors. Examples of such instruments are the Sex Offender Risk Assessment (SORA; Atkinson, Kropp, Laws, & Hart, 1996), the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997), and the Risk for Sexual Violence Protocol (RSVP; Hart, Kropp, Laws, Klaver, Logan, & Watt, 2003).

A large meta-analysis compared the predictive accuracy (discrimination) of unstructured judgment, actuarial instruments, and SPJ instruments for sex offenders ($k = 118$, $n = 45,398$; Hanson & Morton-Bourgon, 2009). Consistent with previous studies, unstructured judgment was significantly less accurate than empirically derived instruments (actuarial and SPJ). In addition, this study also found that actuarial instruments for sex offenders were more accurate than SPJ (Hanson & Morton-Bourgon, 2009).

According to an international survey of 434 forensic mental health experts (Neal & Grisso, 2014), 97% of those experts used at least one structured risk instrument to aid their expert judgment for sex offenders. Seventy-eight different risk assessment instruments for sex offenders were used, but the most frequently used instruments were some version of the Static-99/R (65.9%; Hanson & Thornton, 2000), SVR-20 (13.6%; Boer et al., 1997), RSVP (10.2%; Hart et al., 2003), STABLE-2007 (9.1%; Hanson et al., 2007), and SORAG (7.9%; Quinsey et al., 2006).

A more recent study (Kelley, Ambroziak, Thornton, & Baraha, 2018) reported survey results of instruments used by 119 forensic evaluators mostly from the U.S. Among static risk instruments (second generation), Static-99R (Hanson & Thornton, 2000) was most likely used instruments (80.7%), followed by Static-2002R (30.3%; Hanson & Thornton, 2003), SVR-20 (26.9%; Boer et al., 1997), and VRS-SO Static (12.6; Olver, 2003). Among criminogenic needs instruments (third generation), the

most frequently used instruments were STABLE-2007 (50.4%; Hanson et al., 2007), SVR-20 (22.7%; Boer et al., 1997), RSVP (16.0%; Hart et al., 2003), and VRS-SO (16.0%; Olver, 2003).

Current Uses of Structured Risk Assessment Instruments for Sex Offenders

Structured risk assessment instruments have been used to estimate sexual recidivism risk throughout the criminal justice system. Below is a brief review of some of the decision contexts that rely heavily on the assessment of sexual recidivism risk: 1) criminal sentencing and corrections; 2) civil commitment (U.S.) and dangerous offender hearings (Canada); and 3) sex offender registration and notification (U.S.) and the National Sex Offender Registry (Canada).

Criminal sentencing and corrections. During the “tough on crime” period, the prison population tremendously increased, yet there was no evidence of significant crime reduction in the U.S. (Travis, Western, & Redburn, 2014). Further, the massive growth of incarceration rates caused an economic crisis in several state governments (Scott-Hayward, 2009). Consequently, at least 27 states have started to reform their sentencing strategies moving towards a more cost-effective approach of managing ex-offenders in the community. To promote public safety, these practices require the use of reliable and valid risk assessment instruments to align treatment and monitoring to the risk and needs of the individuals on supervision (Lawrence, 2013). Approximately 20 actuarial risk assessment instruments are currently used in the U.S. correctional settings to assess the risk of general recidivism (Desmarais, Johnson, & Singh, 2016).

In the U.S., the future dangerousness of individuals convicted of murder is a primary consideration in capital sentencing procedures in three states (Oregon, Texas, and Virginia) and considered an aggravating factor in 21 additional states (Claussen-Schulz, Pearce, & Schopp, 2004). Although evaluating future dangerousness is not the

same as a risk assessment, it often includes the findings from actuarial risk assessment instruments. It has been reported that courts in at least 20 states in the U.S. have begun to consider the result of actuarial risk assessment instruments (e.g., The Level of Service Inventory-Revised [LSI-R] scores) to help judges and juries in determining an offender's sentence (Starr, 2014).

In Canada, the Corrections and Conditional Release Act (1992) was instituted with the primary goal of successfully reintegrating offenders into the community by providing appropriate programs to reduce recidivism risk. As a result, assessing individuals' risk for recidivism and identifying criminogenic needs was critical. Risk assessment instruments are currently used by the federal corrections system (CSC) to decide security classification, service intensity, parole release decisions, and the level of community supervision. All jurisdictions in Canada (federal, provincial, and territorial) use one or more recidivism risk instruments (Bourgon, Mugford, Hanson, & Coligado, 2018). Once individuals enter any correctional system in Canada, they can expect to be evaluated multiple times on various risk assessments from the time they are initially charged to when they are finally released from their sentence.

Civil commitment (SVP) and Dangerous Offender legislation (DO). In 1990, Washington State first enacted the Washington sex offender civil commitment law (i.e., sexually violent predator; SVP), which allows confining dangerous sex offenders even after they have served their full criminal sentence. In 1997, the U.S. Supreme Court in *Kansas v. Hendricks* (1997) found that this law did not violate the Constitution (double jeopardy, *ex-post facto*, substantive due process). Currently, 20 states, the District of Columbia, and the federal government have enacted some form of SVP civil commitment statutes (Miller, 2010).

The criteria commonly found in state SVP commitment laws are the

following: (i) history of sexual offending, (ii) a mental abnormality (e.g., antisocial personality disorder, pedophilia), (iii) volitional impairment as the result of the mental abnormality, and (iv) high likelihood of sexual recidivism. In this context, forensic psychologists or psychiatrists offer their expert opinion regarding an offender's potential risk by using actuarial risk assessment instruments. The Static-99/R and Static-2002/R instruments are the most frequently used instruments during civil commitment evaluations (Jackson & Hess, 2007; Neal & Grisso, 2014).

The Canadian law equivalent to the U.S. SVP legislation is the Dangerous Offender (DO) designation, which provides the option of indeterminate incarceration to serious violent and/or sexual offenders who are considered an ongoing threat to society at the sentencing stage. The DO criteria as defined in the *Criminal Code*, 753 (1) includes: (i) a pattern of repetitive behaviour by the offender, (ii) a pattern of persistent aggressive behaviour by the offender, (iii) a brutal nature, or (iv) likelihood of causing injury through a failure to control sexual impulses. Actuarial risk assessment instruments (e.g., Static-99/R, VRAG) are used to predict the likelihood of future violent/sexual offending along with the presence of a diagnosis of antisocial personality disorder and severity of criminal record (Bonta, Harris, Zinger, & Carriere, 1996; Solicitor General Canada, 2001). Similarly, Canada's National Flagging System (NFS) was established in 1995 as an early identification and tracking system for high-risk offenders (violent and sexual offenders). Provincial and territorial NFS coordinators flag high-risk offenders on the national criminal history database based on the DO criteria as defined in the *Criminal Code* (1985).

Sex offender registration/notification and National Sex Offender Registry.

In 1994, a federal law was passed in the U.S. mandating that sex offender register their current address with local law enforcement agencies (Jacob Wetterling Crimes

against Children and Sexually Violent Offender Registration Act; i.e., the Wetterling Act). In 1996, Congress amended the Jacob Wetterling Act with Megan's Law, requiring law enforcement agencies to disseminate the information about registered sex offenders to the public. In 2006, registration and notification requirements were again modified with the enactment of the Adam Walsh Sex Offender Registration and Notification Act (SORNA), which created a national sex offender registry system available online. Currently, all states are mandated to maintain internet websites containing sex offender registration information (U.S. Department of Justice, 2007). Although the requirement of registration and notification are based on the individual's official convicted offence, jurisdictions may use risk assessments (individuals's risk level) to inform decisions concerning the amount of notification information needed (e.g., name, current location, and past offences) and the duration of the order (U. S. Department of Justice, 2007).

In Canada, there is a national registration system for convicted sex offenders, called the National Sex Offender Registry mandated by the Sex Offender Information Registration Act (SOIRA). The requirement of the national registration is also solely based on the designated sexual offences by law (e.g., rape, child pornography, sexual offence against children; see Criminal Code, 1985, 490.011[1]). The duration of the order (10 years, 20 years, or lifetime) is determined by the maximum term of imprisonment for the offence and the number of offences of which the person is convicted (Criminal Code, 1985, 490.013[2]). All accredited Canadian police agencies can access the database containing information about the registered sex offender (e.g., name, current address, vehicle information). The public does not have access to the National Sex Offender Registry. However, Canada Border Services can have access when an individual is considered at high-risk to commit a sexual offence

against a child (SOIRA, 2018, 16[4][j.3]). Currently, the Royal Canadian Mounted Police (RCMP) use the Static-99R for the purpose of identifying high-risk individuals.

Static-99R and Static-2002R

Static-99 (Hanson & Thornton, 2000) was created from items that had demonstrated empirical associations with sexual recidivism (e.g., Hanson & Bussière, 1998), not based on construct validity. The specific items were selected from two previously existing empirical actuarial risk tools - Rapid Risk Assessment for Sexual Offence Recidivism (RRASOR; Hanson, 1997) and the Structured Anchored Clinical Judgment – Minimum criteria (SACJ-Min; Grubin, 1997). In total 10 items were retained, assessing criminal history, victim characteristics, age, and relationship history.

Given that Static-99 was not intended to have construct validity, Static-2002 was developed with the aims of categorizing items into subscales linked with psychological propensities (e.g., general criminality, deviant sexual interests). Static-2002 has 14 items grouped into 5 main subscales: age at release (1 item), the persistence of sex offending (3 items), sexual deviance (3 items), relationship to victims (2 items), and general criminality (5 items).

The Static-99 and Static-2002 were later updated with a re-weighted age item (i.e., Static-99R and Static-2002R). The Static-99R and Static-2002R items are identical to the original Static-99 and Static-2002 except for updated age weights, but the updated age item improves the predictive accuracy of both instruments, particularly for older individuals who committed sexual crimes (Helmus, Thornton, et al., 2012).

Today, Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) and Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012) are the

most widely used structured (actuarial) risk assessment instrument for adult males with a history of sexual offending in the criminal justice system in Canada and the U.S. (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Bourgon et al., 2018; Jackson & Hess, 2007; Kelly et al., 2018; Neal & Grisso, 2014;). Static-99R (10 items) and Static-2002R (14 items) contain only static items (i.e., second generation) and are readily utilized by forensic evaluators (e.g., psychologists) or field officers (e.g., parole officers) across different stages in the criminal justice system (e.g., sentencing, civil commitment evaluations, or parole evaluations; Archer et al., 2006; Jackson & Hess, 2007; Kelley et al., 2018; Neal & Grisso, 2014).

Static-99R and Static-2002R provide two main estimations regarding recidivism risk: 1) relative risk (i.e., how likely the individual is to reoffend compared to other offenders) and 2) absolute risk (i.e., the proportion of individuals with the same score expected to reoffend within a defined period). The norms of Static-99R and Static-2002R have been updated in 2016 with 21 independent samples ($n = 8,805$; rapists and child molesters) for two distinct groups (routine/complete samples and high risk/high needs samples) and for two different follow-up periods (5 years and 10 years; see details in Hanson et al., 2016).

Although it was not developed to measure psychological constructs, factor analyses of the items of both the Static-99R and Static-2002R, have identified three psychological constructs associated with sexual recidivism: 1) persistence/paraphilia, related to sexual criminality, 2) youthful stranger aggression, related to young age and offence seriousness, and 3) general criminality, related to diversity and magnitude of criminal careers (Brouillette-Alarie, Babchishin, Hanson, & Helmus, 2016). The first construct, persistence/paraphilia, is associated with deviant sexual interests towards atypical objects, but without intent to harm their victims (e.g., pedophilia, voyeurism,

and exhibitionism). The second construct, youthful stranger aggression, is related to the intent to sexually harm the victim (e.g., sexual sadism, hostility towards women). The construct of general criminality is related to the antisocial trait (e.g., impulsivity, lack of remorse, rule violations; Brouillette-Alarie, Proulx, & Hanson, 2017).

Summary

In the modern and present era, rehabilitation of individuals who committed crimes is a central concern of the criminal justice system in Canada and the U.S. The most effective interventions are those that provide the most service to the highest risk offenders (Risk), that target criminogenic needs (Need), and are delivered in a manner suited to the offender's culture and learning style (Responsivity; Andrews & Bonta, 2010; Andrews et al., 1990; Andrews & Dowden, 2006). Accurately estimating an offender's recidivism risk is, therefore, essential and should be conducted by reliable and valid risk assessment procedures.

Considerable research supports that structured approaches provide more accurate estimations than unstructured clinical judgments (e.g., actuarial method and SPJ; (Ægisdóttir et al., 2007; Grove et al., 2000). Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) and Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012) are widely used structured (actuarial) risk assessment instruments for adult males with a history of sexual offending in the criminal justice system in Canada and the U.S.

Cross-Cultural Generalizability of Psychological Testing Instruments

Generalizability is an essential issue for all psychological testing. Specifically, psychological testing instruments should only be used with individuals who are substantially similar to the people from whom the instrument was developed. Given the multicultural nature of societies, however, the issue of generalizability across

individuals with different cultural backgrounds (i.e., race/ethnicity) has been a critical concern in Canada, and the U.S. Racial/ethnic minority groups in Canada represent approximately 28% of the general population (e.g., 17% for Asian, 5% for Indigenous, and 3% for Black; Statistics Canada, 2016). In the U.S., racial/ethnic minority groups make up approximately 40% of the general population (18% for Hispanic, 13% for Black, 6% for Asian, 1.3% for American Indian, and 0.2% for Pacific Islander; U.S. Census Bureau, 2017a).

The terms 'race' and 'ethnicity' are commonly used interchangeably. However, they are different. 'Race' usually refers to the biological group to which an individual belongs to base on genetically transmitted physical features (e.g., skin, eyes, or hair colour; White, Black, or Asian). For example, there are five racial categories used in the U.S.: White (including Hispanic), Black, American Indian, Asian, and Native Hawaiian (U.S. Census Bureau, 2018). There are also multiracial groups, which have a mixed ancestry of two or more races. Approximately 8.5 million people (2.6%) of the total U.S. population are identified as multiracial (U.S. Census Bureau, 2017a).

The term 'ethnicity' has a broader meaning than 'race,' and indicates a specific cultural heritage. In other words, ethnicity refers to the groups to which individuals belong to based on shared culture; culture is defined by a combination of behaviours (e.g., social norms, religion, traditions, and language) and attitudes (e.g., beliefs, values, opinions, and stereotypes; Cornell & Hartmann, 2007). In the Canadian census, ethnic origin information including racial information has been collected (e.g., South Asian origins [Sri Lankan, Bangladeshi], Latin, Central and South American [Costa Rican, Ecuadorian]). A racial group (e.g., White), therefore, comprises multiple ethnic groups (Hispanic, Irish, or Italian). Indigenous peoples in Canada also comprise multiple ethnic groups from different cultural backgrounds (multiple First

Nations, Inuit, and Métis).

Furthermore, racial/ethnic minorities might develop multicultural identities (e.g., Korean Canadian/American) through the acculturation process (LaFromboise, Coleman, & Gerton, 1993). Acculturation is “a phenomenon which results when groups of individuals having a different culture come into continuous first-hand contact with subsequent changes in their original culture patterns of either or both groups” (Redfield, Linton, & Herskovits, 1936, p. 149). Berry (1997) described four broad acculturation strategies: *integration*, *assimilation*, *separation/segregation*, and *marginalization*. Integration refers to individuals of ethnic minority origins adopting the cultural norms of the dominant culture while maintaining key aspects of their culture of origin. In contrast, assimilation is when individuals abandon their identity and original cultural characteristics and fully adopting those of the dominant culture. Separation refers to individuals rejecting the dominant culture while preserving their culture of origin (*separation*). The strategy of rejecting both the dominant culture and the culture of origin is, in Berry’s schema, called marginalization. Generation status and language use are strongly associated with the level of acculturation and predict changes in attitudes, behaviours, and values (Berry & Sabatier, 2010). Also, the perception of discrimination against ethnic minorities plays a role in the rate of adaptation to a new culture (Juang & Cookston, 2009; Knight et al., 2009).

Many studies have found associations between acculturation level and behaviour, such as diet and smoking (Allen et al., 2014), substance use (Chappin & Brook, 2001; Miller, Miller, Zapata, & Yin, 2008; Rosa, 2002), delinquent behaviour (Ventura Miller, Barnes, & Hartley, 2011; Wong, 1999), and sexual behaviour (Lee & Hahm, 2010; Meston & Ahrold, 2010). The level of acculturation also could influence the patterns of risk-relevant characteristics of ethnic minority individuals with a

history of crimes. For example, assimilated immigrants in Canada and the U.S. might have the risk-relevant characteristics most similar to Whites and, thus, may be the least likely to have culturally biased results of structured risk assessment instruments.

In particular, cultural bias in structured risk assessment instruments is an important concern given a overrepresentation of certain racial/ethnic minority groups in the criminal justice systems in Canada and the U.S. There is strong evidence that a more significant proportion of racial/ethnic minorities (e.g., Blacks and Hispanics) than Whites are represented among offender populations in general, and for violent crimes specifically (e.g., theft, robbery, or assault; Carson, 2018; Hartney & Vuong, 2009; Minton & Zeng, 2016). Also, research has demonstrated an overrepresentation of racial/ethnic minorities in sex offender populations. For example, racial/ethnic minority groups in Canada represent approximately 28% of the general population, yet they constitute more than 32% of federal sex offenders in custody and under community supervision (Axford, 2008). In the U.S., racial/ethnic minority groups make up approximately 40% of the general population, yet they constitute more than 50% of arrests, inmates held in prisons, and registered sex offender groups (Carson, 2018; Federal Bureau of Investigation, 2015; Smith-Socar, Perry, & Fox-Mullen, 2006; U.S. Census Bureau, 2017a).

Not all racial/ethnic minorities are, however, overrepresented in the criminal justice system. In the U.S., Blacks are the most overrepresented group followed by Hispanics and Native Americans whereas Whites and Asians are underrepresented (e.g., arrests, incarcerations, probation and parole; Carson, 2018; Hartney & Vuong, 2009; Minton & Zeng, 2016). In Canada, Indigenous peoples are the most overrepresented ethnic group followed by Blacks whereas Whites and Asians are consistently underrepresented (Axford, 2008).

Racial/Ethnic Minority Groups in Canada and the U.S.

As mentioned above, there are three racial/ethnic minority groups who are consistently overrepresented in the offender populations in Canada and the U.S.: Blacks, Hispanics, and Indigenous peoples. The issue of cultural bias in the structured risk assessment is thus particularly important for these overrepresented ethnic groups.

Blacks in the U.S. Blacks make up approximately 13% of the U.S. population of over 40 million people (U.S. Census Bureau, 2017a). The immigration history of Blacks to the U.S. began with the slavery trade through the European colonization of the Americas in the 17th and 18th centuries. Currently, it is reported that more than 90% of the Blacks' ancestors were forcibly taken from Africa as slaves by European settlers during this period (Spickard, 2007). Although the Civil War (1861-1865) freed Blacks by law from slavery, Blacks have continued to experience racial oppression or discrimination (e.g., social or economic inequality by skin colour) to this day.

During the early 1900s, Blacks had been terrorized by European White peoples (e.g., Ku-Klux-Klan), particularly in southern regions of the U.S. (e.g., burning or lynching; Tolnay & Beck, 1992; Tolnay, Deane, & Beck, 1996). There were also widespread segregation rules that reinforced Blacks as second-class citizens. For example, on buses, Black passengers had to sit at the back and leave the front seats to White passengers until 1956 (*Browder v. Gayle*, 1956). Also, until the Supreme Court declared that the "separate but equal" doctrine was unconstitutional (*Brown v. Board of Education of Topeka*, 1954), racial segregation of children in public schools had been legal (*Plessy v. Ferguson*, 1896) and widely practiced.

Although the Civil Rights Act in 1964 mandated constitutional rights without discrimination or segregation by race, colour, religion, or national origin, there is evidence that both subtle and overt racism (social inequality) has continued until

today. In comparison to White Americans, for example, current estimates show that Blacks have lower average annual incomes (\$36,898 versus \$62,950), higher poverty rates (24% versus 9%), lower homeownership rates (42% versus 72%), and lower education levels (87% with a high school diploma versus 93%; Callis & Kresin, 2018; Proctor, Semega, & Kollar, 2016; U.S. Census Bureau, 2017b).

Hispanics in the U.S. Hispanics refer to people having Spanish ancestry, including those from Puerto Rico, Cuba, Central and South American, and Spain. Given that most Hispanics are racially mixed (e.g., Spanish and Indian, *Mestizos*; Spanish and African, *mulattos*), ethnicity is more appropriate as a term rather than race to identify these groups (e.g., Mexican American, Cuban American; Huddy & Virtanen, 1995). Hispanics represent approximately 18% of the U.S. population at over 50 million people (U.S. Census Bureau, 2017a). Mexicans are the largest Hispanic subgroup (60%) in the U.S., followed by Puerto Ricans (10%), Cubans (4%) and Dominicans (2%; Chun, 2007).

The first major influx of Mexican people into the U.S. began in the aftermath of the Mexican-American war (1846-1848). Following the defeat of the Mexican army, the Mexican government signed the Treaty of Guadalupe Hidalgo in 1848, which ceded all or parts of Texas, California, New Mexico, Nevada, Colorado, Arizona, and Utah to the U.S. (approximately 45% of Mexican territory including over 100,000 Mexicans; Bernal & Enchautegui-de-Jesús, 1994). Although the treaty promised citizenship, freedom of religion and language, and maintenance of lands, Mexicans continued to experience social and cultural oppression by European Americans. For example, they were forced to use English and to follow U.S. laws (Spickard, 2007). In addition, Mexican landowners lost their lands and resources to the U.S. government (Takaki, 2008). During the early 1900s, many Mexicans

immigrated to the U.S. to find better job opportunities. The available work was, however, often menial and dangerous (e.g., working on railroads or in mining). Mexicans were a cheap source of labour, as they were paid less than European Americans for the same work (Takaki, 2008).

During the Spanish American War in 1898, Puerto Rico and Cuba were invaded and occupied by the U.S. (Bernal & Enchautegui-de-Jesús, 1994). Since then, Puerto Rico has remained a colony of the U.S. whereas Cuba gained independence from the U.S. in 1902. Large numbers of Puerto Ricans emigrated to the northeastern U.S. in the 1940s and 1950s for economic reasons. Puerto Ricans often took menial and low paying jobs that no one else wanted (Inclán & Quiñones, 2003). In contrast, political reasons were the main motivation for initial waves of Cuban immigrants to the U.S. since 1959. Consequently, unlike other Hispanic groups, Cuban immigrants included a broad spectrum of people regarding race, education, gender, and socioeconomic status (Bernal & Enchautegui-de-Jesús, 1994).

Following the end of the Soviet Union in the 1990s, which had been sending aid to Cuba, Cuba experienced a severe economic recession (Bernal & Shapiro, 2005). Since then, a lot of *balseiros* (i.e., rafters) have attempted to immigrate illegally to Florida. During this period, Cuban migrants came primarily from the lower and middle-class (e.g., unskilled or service workers). Overall, compared to White Americans, Hispanics in the U.S. also have lower average annual incomes (\$45,148 versus \$62,950), higher poverty rates (21% versus 9%), lower homeownership rates (47% versus 72%), and lower education levels (72% with a high school diploma versus 93%; Callis & Kresin, 2018; Proctor et al., 2016; U.S. Census Bureau, 2017b).

Indigenous peoples in Canada. There are 1.7 million Indigenous peoples in Canada, which is 5% of the total population (Statistics Canada, 2016). There are three

categories of Indigenous peoples in Canada: First Nations (registered or non-registered), Inuit, and Métis (Constitution Act, 1982, s 35[1]). First Nations are the largest subgroup (60%) of the Canadian Indigenous peoples, followed by Métis (35%) and Inuit (4%). First Nations and Inuit peoples were the original inhabitants of the land that is now Canada. First Nations occupy territories south of the Arctic whereas Inuit people primarily inhabit the northern regions of Canada (i.e., the Arctic region). Métis in Canada is a group of people with mixed European and First Nations ancestry. They are recognized as one of Canada's indigenous peoples under the Constitution Act of 1982, along with First Nations and Inuit peoples. They currently live across the country, but mostly in the prairie provinces and Ontario.

During the colonization activities of France (1534-1763) and Britain (1763-1931), Indigenous peoples lost their lands and resources to European settlers. Their political, economic, and religious systems were destroyed, and they experienced traditional racism, which is a belief in the genetic or cultural inferiority of Indigenous peoples (Frideres & Gadacz, 2012). Since the colonization era, considerable effort has been made to address and recover the rights of Indigenous peoples (e.g., the Royal Commission on Aboriginal Peoples, 1996; Indian Act, 1985).

Nevertheless, Indigenous peoples still experience racial oppression or discrimination in the European White-dominated society. In comparison to non-Indigenous peoples, for example, Indigenous peoples have lower median incomes (\$25,526 versus \$34,604), lower employment rates (52% versus 60.5%), and lower education levels (62% with a high school diploma versus 80%; Statistics Canada, 2018). Indigenous peoples are also reported to have poorer health status than non-Indigenous peoples (e.g., poor mental health and short life expectancy; Statistics Canada, 2005; Statistics Canada, 2015).

Overrepresentation of racial/ethnic groups in the criminal justice system.

These three overrepresented racial/ethnic minority groups in Canada and the U.S. have been commonly experienced social and economic oppression in White-dominated societies. It has been theorized that racial/ethnic minority individuals in post-colonial societies experience the unfair social strain and social conflict with dominant groups (Whites), which results in criminal delinquency and behaviour. According to *social strain theory*, when an individual is blocked from achieving their goals, it can result in frustration and anger. Consequently, frustrated and angry individuals are, in turn, likely to seek illegitimate means of achieving their goals (Agnew, 1992). Racial minorities experience excessive strain in their social environment (e.g., economic strain, family strain, educational strain, discrimination) and have fewer resources for coping with this strain in conventional ways (Kaufman, Rebellon, & Thaxton, & Agnew, 2008). Experiencing social strain like discrimination of racial minority groups has been shown to significantly increase the likelihood of engaging in delinquent behaviours (Pérez, Jennings, & Gover, 2008; Simons, Chen, Stewart, & Brody, 2003).

Conflict Theory views crime as a result of social conflict between dominant groups and subordinate groups (i.e., power differentials). Authority and influence in politics, economics, and society are not equally distributed among individuals in society with the dominant groups determining the direction of justice administration. If ethnic minorities, notably immigrants who have different conduct norms, fail to follow the laws, values, and standards of society administered by a dominant group (White), their behaviour is deemed criminal (Sellin, 1938). Further, *Conflict Theory* has been expanded by Blalock (1967) and Hawkins (1987) who included the effects of racial discrimination. They argue that as the racial minority population increases

both economically and politically, the dominant group sees it as a threat, so there is increased social control on ethnic minorities to reduce their economic and political growth and to maintain their power. This racial threat theory might be used to explain the overrepresentation of minority groups in the criminal justice system.

Colonial Theory accounts for the etiology of crime and violence in colonial and postcolonial societies. Essentially, the colonizers (European Whites) attempt to dominate economically and politically to control the colonial society. This colonialization affects the structural and cultural status of minority groups, for example, the degradation of indigenous culture, values, and norms (Tatum, 2002). As a result of colonialism, the colonized have low social status and limited access to societal resources. They become alienated from society; a process sometimes referred to as psychological deprivation, which leads to violent behaviour. This violent behaviour can either be directed towards the colonizer (i.e., vertical violence) or their people (i.e., horizontal violence) as a response to the unequal structural relations in the colonial environment. Social support provided by social networks and communities, however, can reduce minority alienation, thereby reducing subsequent crime and violence as well.

Consequently, the overrepresented racial/ethnic minority groups under unfair social structures might have substantially different patterns of criminal behaviours than Whites. Given that these racial/ethnic minority groups were underrepresented in the research samples used to develop recidivism risk instruments, it is possible that these instruments could contain as yet unknown cultural biases.

Cultural Bias in Structured Risk Assessment Instruments

Not only are psychologists and educators concerned about the potential of cultural bias in psychological assessment instruments, so too are the courts. In the

case of *Ewert v. Canada* (2007), the Supreme Court of Canada (June 13, 2018) ruled that the structured risk assessment instruments (e.g., Static-99R, VRS-SO, VRAG) most commonly used in CSC have insufficient evidence to justify their application to Indigenous peoples (i.e., the potential for cultural bias). Consequently, the Supreme Court of Canada ruled that CSC violated the statutory duty requiring taking reasonable steps to ensure those instruments give accurate risk estimations for Indigenous peoples (*Corrections and Conditional Release Act*, 1992, c. 20, s 24 [1]).

There are certain differences in the understanding and use of the term ‘bias’ between laypeople and scientific researchers. Laypeople may use the term ‘bias’ to demonstrate unjustified or incorrect attitudes towards an individual based solely on the individual’s membership to a particular social group (e.g., “You’re biased against Blacks”). Within the context of scientific research, however, the term ‘bias,’ is an empirical term (i.e., statistically estimated quantity) rather than a prejudicial attitude (Reynolds & Suzuki, 2013). If there are any statistically significant differences in the outcomes (e.g., estimated recidivism risk) of structured risk assessment instruments attributable to cultural variables, such as race/ethnicity, cultural bias is empirically present.

There are broadly three possible causes of cultural bias in the use of structured risk assessment instruments. First, the construct measured (e.g., general criminality and sexual criminality) might not be the same across cultural groups (construct bias). Second, risk factors (items) used in structured risk assessment instruments might have a different psychological meaning across racial/ethnic groups (item bias). Lastly, an evaluator might distort the conclusions of the measure (e.g., a total score) based on race or ethnicity of offenders (evaluator bias; Van de vijver & Tanzer, 2004).

The primary role of structured risk assessment instruments is to inform

decision-makers of the likelihood of future recidivism (i.e., a criterion-referenced outcome). The validation process of risk assessment instruments is associated with the scientific evaluation of proposed score interpretations (i.e., recidivism risk; American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). Criterion-related validity evidence (i.e., predictive validity) is the primary justification for the use of risk assessment instruments. According to Standard 3.7. of *The Standards for Educational and Psychological Testing*,

When criterion-related validity evidence is used as a basis for test score-based predictions of future performance, and sample sizes are sufficient, test developer and/or users are responsible for evaluating the possibility of differential prediction for relevant subgroups for which there is prior evidence or theory suggesting differential prediction (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014, p. 66).

Consequently, it is necessary to evaluate if a risk instrument systematically overestimates or underestimates recidivism risk for particular racial/ethnic groups (i.e., cultural bias in predictive validity). The Standards further stipulate that even when there is substantial evidence to support the predictive accuracy of a prediction instrument, it does not diminish the need for evidence to support other types of validity (e.g., construct validity of predictor domains: sexual deviancy and anti-sociality). In addition, a prediction instrument should be supported by a causal model justifying why the indicators (e.g., negative attitudes towards authority) should be associated with the outcomes (i.e., criminal recidivism). Credible causal theories minimize the likelihood that the observed associations between the indicators and the

outcome are confounded by variables that are not appropriate for the assessment task. For example, few evaluators would consider it appropriate to ask about parents' income on a college admissions assessment, even if it predicts educational achievement.

Given that the static risk instruments (i.e., Static-99R and Static-2002R) are the most commonly used risk assessment instruments for individuals with a history of sexual crime, it is important to examine their predictive validity among different racial/ethnic groups. As a criterion-referenced instrument, there are two main interpretations of Static-99R and Static-2002R total scores: relative risk and absolute risk. Relative risk states that sexual offenders with a higher Static-99R score are more likely to reoffend than individuals with a lower Static-99R score. Absolute risk indicates that the estimated probability of sexual offending of individuals with an individual score of Static-99R within a defined period. For a comprehensive evaluation of the predictive validity of an instrument, predictive accuracy concerning both relative and absolute risk should be considered (Hanson et al., 2016; Helmus & Babchishin, 2017; Helmus & Hanson, 2011).

As previously mentioned, actuarial risk assessment instruments use pre-specified risk factors and explicit combination rules to estimate the recidivism rates (Dawes, Faust, & Meehl, 1989; Kleinmuntz, 1990; Meehl, 1954). Such measures aim to minimize cultural bias in evaluations by making the administration and interpretation of actuarial instruments as transparent and mechanical as possible. Nevertheless, actuarial risk assessment instruments can be culturally biased due to the cultural discrepancies of construct or items measuring sexual recidivism across different racial/ethnic groups. The actuarial instruments, thus, might systematically (i.e., non-random error) overestimate or underestimate the sexual recidivism risk for

particular racial/ethnic groups (Reynolds, 2000a; Reynolds & Suzuki, 2013).

Possible explanations of cultural bias in predictive validity. According to the reasoned action approach (Fishbein & Ajzen, 2015), three factors (attitude towards the behaviour, perceived norm, and perceived behavioural control) determine behavioural intentions, which are the most immediate and potent predictor of human behaviour. Those three factors are formed based on certain beliefs: behavioural (positive or negative consequence), normative (approval or disapproval from influential individuals or groups), and control beliefs (feasibility). The beliefs are acquired by experience in real life and are influenced by personal characteristics (e.g., personality, temperament, intelligence, values), and social and cultural factors (e.g., ethnicity, race, religion, education). Culture is significantly associated with those background elements. People who come from different cultural backgrounds, thus, can form different beliefs. Consequently, considering cultural backgrounds may help to predict human behaviour including delinquent behaviours.

Ethnic minorities are usually present in relatively small numbers in developmental or norming samples of structured risk assessment instruments. According to the Standard 5.8. of *The Standards for Educational and Psychological Testing*,

Norms, if used, should refer to clearly described population. These populations should include individuals or groups with whom test users will ordinarily wish to compare their own examinees (2014, p. 104).

In other words, risk instruments should only be used with individuals who are substantially similar to the people from whom the instrument was developed. Consequently, the underrepresentation of ethnic minorities in the research samples might lead to biased risk factors (items) selection, which provides accurate

estimations of sexual recidivism risk only for majority group members (i.e., Whites), but might not for ethnic minority counterparts (e.g., Hispanic, Black, or Indigenous peoples).

Second, the risk factors in sex offender risk assessment instruments are composed broadly of two major propensities associated with sexual recidivism: general criminality (e.g., impulsivity, juvenile delinquency, hostility towards women) and sexual criminality (e.g., sexual preoccupation and pedophilia; Brouillette-Alarie & Proulx, 2018; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann et al., 2010; Whitaker et al., 2008). Racial/ethnic minority individuals might have substantially different patterns of the risk-relevant characteristics related to general and sexual criminality compared to Whites.

There are considerable efforts made to explain the different patterns on the general criminality of racial/ethnic minority in the Western countries based on criminological models (e.g., social strain, colonization theories). Relatively few studies, however, have examined the risk-relevant characteristics associated with sexual criminality (e.g., paraphilia) across different racial/ethnic groups. Broadly, sexual behaviour is regulated by human physiological factors including genetics and hormones, as well as societal factors such as laws, culture, and norms. Every culture has a set of beliefs, attitudes, and norms regarding sexuality. In traditional sex cultures (e.g., Asian), there are more restrictions on the expression of sexuality, premarital sex, extramarital sex, and homosexuality than many contemporary Western cultures, such as Canada. Despite this, Asians (in a patriarchal society) are more tolerant towards sexual offences than other ethnic groups, are more likely to endorse rape myths, and have a higher acceptance of sexual harassment (Kennedy & Gorzalka, 2002). In contrast, Whites (nontraditional sex culture) have more tolerant attitudes towards

sexual behaviour and less acceptance of rape myths, and sexual harassment (Ahrold & Meston, 2010; Kennedy & Gorzalka, 2002).

There has been several integrated models and theories proposed to explain sexual offending (Finkelhor, 1984; Hall & Hirschman, 1991; Malamuth, 1998; Marshall & Barbaree, 1990; Ward & Beech, 2006). Although each model consists of various external and internal variables, they all consider that culture should influence the etiology of sexual offending. For example, the strict social norms regarding child sexual abuse and right child supervision/monitoring level observed in collectivistic societies might inhibit the occurrence of child sexual abuse (Finkelhor, Ji, Mikton, & Dunne, 2013). It is also possible that restricted sexual expression in traditional sex cultures and corresponding low sex education make children more vulnerable to sexual assault in collectivistic societies (Finkelhor, 1984). In addition, there are cultural differences associated with acceptance of interpersonal violence, dominant male culture, and negative attitudes towards women (Hall & Hirschman, 1991; Marshall & Barbaree, 1990). Consequently, sexual offenders of different cultures may demonstrate substantially different patterns of the risk-relevant characteristics associated with sexual criminality.

Third, given that cultural values, such as societal norms and religion, influence our social behaviours (Horens & Poortinga, 2000), different racial/ethnic minority groups might have cultural-specific risk factors, which are directly or indirectly associated with sexual recidivism. Consequently, it is possible that these potential cultural risk factors have not been adequately addressed by risk assessment instruments, resulting in restraint on the predictive ability of these instruments.

Lastly, regardless of social structure and possible cultural-specific risk factors, cultural values or immigrant status could influence the predictive accuracy of certain

risk assessment instruments by restricting the amount of necessary information for proper administration. For example, ethnic minority offenders (e.g., Asians) have a strong tendency to underreport sexual crimes due to the fear of losing face in their community (Hall, 2002) and disgracing their social groups (Hall, Windover, & Maramba, 1998). Further, undetected crimes occurring in the country of origin of recent immigrants might also undermine the predictive accuracy of these instruments.

Previous Studies Evaluating Cultural bias in Static-99R and Static-2002R.

There are, to date, a few published studies in the U.S. that evaluate the cultural bias in predictive accuracy of Static-99R across different ethnic sex offender groups: Whites, Blacks, and Hispanics. First, Varela and colleagues (2013) examined the predictive validity of Static-99R among 1,911 sex offenders released from prison in Texas (White, $n = 912$; Black, $n = 411$; and Hispanic, $n = 588$). They found poor to moderate discrimination for violent sexual recidivism (i.e., any contact sexual offence) across different ethnic groups (AUCs of .57 for Hispanics, .59 for Whites, and .65 for Blacks) and no significant differences among the ethnic groups. No calibration analyses were performed (see Varela, Boccaccini, Murrie, Caperton, & Gonzalez, 2013).

Hanson, Lunetta, Phenix, Neeley, and Epperson (2014) reported good predictive validity with 475 sex offenders under the parole system in California (AUC of .82 [.72, .92]) and overall good and similar discrimination across all ethnic groups (e.g., White, Black, and Hispanic; AUCs $>.74$). The overall calibration between the expected and observed recidivism rate was generally good (4.8% vs. 6.31%; $E/O = 1.30 [0.87, 1.96]$). Calibration analyses were, however, not conducted for each ethnic group due to the small total sample size and a small number of recidivists.

In 2018, Lee and his colleagues followed up the Hanson et al. (2014) study

with a 10-year follow-up period. Within a fixed 10-year follow-up, the overall AUC with any sexual recidivism was 0.75 [0.65, 0.85]. For each ethnic group, White sex offenders showed the highest AUC value of .85 [.74, .96] and Hispanic sex offenders had moderate AUC of .70 [.48, .91]. For Black sex offenders, the values of AUC were lower than the values of the other groups (AUC of .63 [.42, .84]). The differences between ethnic groups were, however, not statistically significant. Calibration could not be tested because the Static-99R user guide does not currently have norms for routine/complete samples at 10-year follow-up (Lee, Hanson, Fullmer, Neeley, & Ramos, 2018).

In a study with 483 Hispanic sex offenders in New Jersey (Leguizamo, Lee, Jeglic, & Clakins, 2017), Static-99R was able to discriminate recidivists from nonrecidivists for Hispanics (AUC of .72 [.59, .91], $n = 268$). Further, Static-99R, worked only for U.S.-born Hispanics (AUC of .82 [.64, .99]), but did not work for those born in Latin America (AUC of .52 [.19, .86]; $n = 215$). The observed 5-year overall recidivism rate was significantly lower than the expected recidivism rates from Static-99R norms (3.0% vs. 6.3%; $E/O = 2.11 [1.10, 4.06]$). Calibration analyses were, however, not conducted for each ethnic group due to the small total sample size and a small number of recidivists.

Boccaccini, Helmus, Murrie, and Harris (2017) examined the predictive validity of Static-99R for three different ethnic groups (White, $n = 7,938$; Black, $n = 9,725$; Hispanic, $n = 8,939$). They found very similar and moderate discrimination for Static-99R for sexual recidivism across the different ethnic groups (AUCs of .65 for White, .64 for Black, and .63 for Hispanic). They also divided the sample into U.S.-born Hispanic ($n = 6,337$) and Hispanic born outside the U.S. ($n = 2,459$). Contrary to Leguizamo et al. (2017), Static-99R worked better for Hispanics born outside the U.S.

than for those born inside (AUC of .65 [.52, .78] and AUC of .61 [.58, .65], respectively), but the difference was not statistically significant. Calibration analyses for each ethnic subgroup were not reported.

Collectively, these studies consistently found that Hispanic sex offenders had relative low mean scores of Static-99R and low sexual recidivism rates, whereas Black sex offenders had relatively high mean scores and high sexual recidivism rates. Findings, however, about the predictive accuracy of Static-99R across different ethnic groups have been inconsistent although there is a pattern of relatively weak discrimination and calibration of Static-99R for Hispanic sex offenders.

In Canada, there is only one published study that has examined the cultural bias in predictive accuracy of Static-99R and Static-2002R for Indigenous peoples. Babchishin and colleagues (2012) examined the predictive validity of Static-99R and Static-2002R among 1,588 sex offenders (Indigenous, $n = 319$; non-Indigenous, $n = 1,269$; Babchishin, Blais, & Helmus, 2012). The total scores of Static-99R and Static-2002R for Indigenous people were significantly higher than for non-Indigenous people. They found very similar and moderate discrimination for Static-99R for sexual recidivism (AUCs of .71 for Indigenous, and .74 for non-Indigenous). Static-2002R, however, was only significantly related to sexual recidivism for non-Indigenous people (AUC of .76 versus AUC of .61 for Indigenous peoples). This study did not evaluate the potential cultural bias in the calibration of Static-99R and Static-2002R.

Although some research on these instruments has been conducted, there is still insufficient evidence to have reliable estimates of the predictive accuracy for each ethnic minority groups; in particular, there has been very limited research on cultural bias in calibration. Also, the previous studies did not consider psychologically

meaningful construct (e.g., general criminality and sexual criminality) for the comparisons across different ethnic groups, which could provide some insight as to why specific ethnic group has a high average score on the instruments, and why the instruments may work better for certain ethnic groups than others. The main purpose of this dissertation is to fill this research gap.

Overview of Purpose

Criminological and psychological theories concerning sexual offending potentially suggest that there could be differences in predictive accuracy of risk assessment instruments among ethnic minorities (i.e., measurement of different constructs; anti-sociality and sexual deviancy; e.g., Agnew, 1992; Hall & Hirschman, 1991; Tatum, 2002). In addition, the small number of ethnic minorities in development samples raises questions concerning the predictive validity of risk assessment instruments for these understudied subgroups (e.g., item-selection bias).

Given that actuarial risk assessment instruments clearly identify the risk factors to consider and how they should be combined into a total score, they should be less vulnerable to cultural/racial bias than unstructured clinical judgment. However, the potential for cultural bias remains. To date, the degree of cultural bias in actuarial risk assessment instruments for individuals with a history of sexual crime has remained mostly unexplored, possibly due to the relatively small number of ethnic minorities in the development and normative samples. Consequently, research is needed to examine whether actuarial risk assessment instruments, specifically Static-99R and Static-2002R, would perform similarly across different ethnic groups.

The current studies evaluated and discussed the following research questions:

1) Do risk levels (i.e., mean scores of Static-99R and Static-2002R) differ across different racial/ethnic groups?; 2) What is the discrimination accuracy of Static-99R

and Static-2002R across the various racial/ethnic groups?; 3) What is the calibration accuracy of Static-99R and Static-2002R across racial/ethnic groups?; and 4) Are there any differences in risk-relevant characteristics across different racial/ethnic groups? These research questions were answered using three separate studies: 1) California sample including White, Black, and Hispanic individuals utilizing the Static-99R, 2) New Jersey sample including White and Black sex offenders using the Static-99R, and 3) Indigenous peoples in Canada using Static-99R and Static-2002R.

Chapter 2: Method

In this dissertation, a series of statistical analyses were conducted to assess the cultural bias in the predictive validity of Static-99R and Static-2002R for different ethnic/racial groups. Predictive validity involves two main components: discrimination (how large are the differences in the instrument's scores between those who reoffend and those who do not reoffend) and calibration (how well the estimated recidivism probabilities from the instrument's norms corresponds with the observed recidivism probability of new samples). If there are statistically significant differences in discrimination and calibration of Static-99R and Static-2002R across individuals with different cultural background, cultural bias is empirically present. More detail is provided in each of the studies. A brief overview of the method is provided below.

Participants

Study 1. This study included adult male sexual offenders ($N = 2,101$) released from the California Department of Corrections and Rehabilitation (CDCR; i.e., parolees; $n = 1,673$) as well as those on probation (i.e., probationers; $n = 428$). Of the overall 2,101 offenders, 37.6 % ($n = 789$) were White, 22.2 % ($n = 466$) were Black, 34.2 % ($n = 719$) were Hispanic and 6% ($n = 127$) were Others/Unknown.

Study 2. All individuals in this study were adult males who were convicted of a sexual offence ($N = 1,585$; 788 were Black, and 797 were non-Hispanic White) in the State of New Jersey. The sample was selected from individuals who had been detained at either the Adult Diagnostic Treatment Center (ADTC) or any of New Jersey State Prisons. Of the total sample, 21% ($n = 325$) was legally designated as sexually violent predators (SVPs).

Study 3. This study was a meta-analysis that included an aggregated sample 653 Indigenous and 1,560 White individuals with a history of sexual crimes from five

independent Canadian samples (Bonta & Yessine, 2005; Brankley, Lee, Babchishin, Hanson, & Harris, 2017a; Haag, 2005; Lee, Mularczyk, Babchishin, Blais, & Bonta, 2018; Olver et al., 2018). All individuals were identified as the high level of risk and/or need for sexual recidivism (e.g., preselected [referred to as National Flagging System], detained until the end of a sentence, or attended high-intensity sex offence specific treatment programs).

In accordance with the Tri-Council Policy Statement 2 (TCPS2, Article 9.17, 2010), we solicited experts' reviews and opinions from individuals of Indigenous heritage who were engaged professionally with the issue of Indigenous overrepresentation in Canada's criminal justice system (Leticia Gutierrez and Patti McDonald). Their contributions help shape the interpretations and narratives of the research findings for Indigenous populations.

Measures

Static-99R. Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) is a 10-item empirical actuarial risk tool designed to assess the risk of sexual recidivism among adult males with a history of sexual offending. Static-99R is identical to Static-99 except that it contains revised age weights. The total score (ranging from -3 to 12) is calculated by summing all items and can be used to place individuals in one of five risk categories: Level I - very low risk (scores of -3 to -2), Level II - below average risk (scores of -1 to 0), Level III - average risk (scores of 1 to 3), Level IVa - above average risk (scores of 4 to 5), and Level IVb - well above average risk (scores of 6 or higher; Hanson, Babchishin, Helmus, Thornton, & Phenix, 2017).

Static-2002R. Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012) is also an empirical actuarial risk assessment tool for adult male sex

offenders. The scale has 14 items grouped into five main subscales: Age at Release (1 item), Persistence of Sex Offending (3 items), Sexual Deviance (3 items), Relationship to Victims (2 items), and General Criminality (5 items). The Static-2002R items are identical to the original Static-2002 except for updated age weights. The total score for Static-2002R can range from -2 to 13 and can be used to place offenders in one of five risk categories: Level I - very low risk (scores of -2 to -1), Level II - below average risk (scores of 0 to 1), Level III - average risk (scores of 2 to 4), Level IVa - above average risk (scores of 5 to 6), and Level IVb - well above average risk (scores of 7 or higher; Hanson et al., 2017).

Analytical Strategy

For discrimination, two statistical methods were used: 1) the area under the curve (AUC) from receiver operating characteristic (ROC) analysis (Swets, Dawes, & Monahan, 2000) and 2) odds ratios from logistic regression (Hosmer & Lemeshow, 2002). For calibration, two indices were used: 1) *E/O* index (the ratio of expected number of recidivists divided by an observed number of recidivists; Hanson, 2017) and 2) fixed-effect meta-analysis of logistic regression parameters (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hanson & Broom, 2005).

Discrimination

Area Under the Curve (AUC). AUC values indicate the probability that a randomly selected recidivist would have a higher score on the Static-99R and Static-2002R than a randomly selected non-recidivist. The AUC can vary between 0 and 1, with .50 indicating a chance level of prediction. A value of AUC between .50 and 1 indicates above chance-level prediction with statistical significance as long as the 95% confidence interval does not include .50. As a rough heuristic, an AUC of .56

corresponds to small effect size, while .64 reflects moderate effect, and .71 reflects a large effect size (Rice & Harris, 2005).

Odds ratios. Odds ratios (a slope parameter [B1] from logistic regression) indicate the change in relative risk associated with one unit change in the Static-99R and Static-2002R scores. For example, in routine/complete samples, Static-99R scores are associated with a consistent relative risk increase of approximately 1.45 (Hanson et al., 2016), which means the odds of recidivism increases 1.45 times as each Static-99R score increases with 5-year follow-up time. An odds ratio is statistically significant if the 95% confidence interval does not include 1. Differences between the parameters (*BI*) in the samples and those of the norms were tested using fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005).

Calibration

E/O index. The *E/O* index is a measure of calibration in which the expected number of recidivists from the instrument's norms is divided by an observed number of recidivists from the validating samples (Hanson, 2017). The expected number of recidivists was based on the 5-year or 10-year sexual recidivism rates for routine/complete samples or high risk/high needs samples (Hanson et al., 2016; Phenix, Helmus, & Hanson, 2016). Perfect calibration is indicated by an *E/O* index of 1.0. *E/O* indices above 1.0 indicate the overestimation of sexual recidivism rates whereas *E/O* indices below 1.0 indicate the underestimation of sexual recidivism rates. An *E/O* index is statistically significant if the 95% confidence interval does not include 1.

Comparing Logistic regression parameters. Another approach to testing calibration was to examine the extent to which logistic regression parameters, such as intercept values (centered on Static-99R scores of 2 [*B0₂*] and centered on Static-

2002R scores of 3 [$B0_3$]) differed from the logistic regression parameters for the routine sample norms or the high risk/high needs sample norms. Differences between the parameters in the current sample and those of the norms were tested using fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005).

Chapter 3: Study 1

Similar Predictive Accuracy of the Static-99R Risk Tool for White, Black, and Hispanic Sex Offenders in California

Structured risk assessments are widely used (Neal & Grisso, 2014) and have been supported by a long tradition of research in the criminal justice and forensic mental health systems (Ægisdóttir et al., 2007; Hanson & Morton-Bourgon, 2009; Yang et al., 2010). Although the overall utility of these tools is generally accepted, there are outstanding questions concerning their generalizability to different ethnic groups (e.g., Black, Hispanic, Asian, Indigenous).

The issue of cultural bias in psychological testing (e.g., intelligence tests for Blacks) has a long history in multicultural Western societies. Risk assessment instruments also cannot avoid the debate concerning potential bias for different ethnic groups. There are several possible causes of cultural bias in the performance of the current risk assessment tools. First, ethnic minorities are usually present in relatively small numbers in development samples, which might lead to biased item (i.e., risk factors) selection associated with recidivism risk. Second, risk assessment tools might not fully measure the core propensities (i.e., anti-sociality and sexual deviancy) when used with minority offenders. Consequently, it is possible that there are distinct risk and protective factors for different ethnic groups that have not been fully addressed by risk scales developed and validated on multi-ethnic samples.

Understanding the predictive accuracy of a risk scale (i.e., criterion-referenced prediction tool) should consider calibration (correspondence between expected and observed recidivism rates) as well as discrimination (how different are recidivists from non-recidivists). In order to check for potential cultural/racial bias for both calibration and discrimination, we can test the equivalence of (logistic) regression

equations across ethnic groups (i.e., B0 for calibration; B1 for discrimination; Reynolds, 2000a). Prediction scales are unbiased when there are no systematic differences across ethnic subgroups in the expected recidivism rates for offenders with the same score or category (Reynolds & Suzuki, 2013).

The purpose of this study was to examine whether cultural/racial bias exists in one of the popular actuarial tools, Static-99R, with different ethnic groups of sexual offenders (White, Hispanics, and Blacks) in the United States. Static-99/R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) is the most widely used actuarial risk assessment tool for adult sexual offenders by forensic experts in the Western countries (e.g., the U.S., Canada; Neal & Grisso, 2014). Considerable research has found overall moderate predictive validity ($AUC = .70$, $n = 8,106$, $k = 23$; Helmus, Hanson, et al., 2012). Relatively little research, however, has addressed the potential cultural bias for the performance of Static-99R with different ethnic minority groups.

The United States is a multicultural and multi-ethnic society. Approximately 35% of the U.S. total population (approximately 320 million) is from ethnic minority groups, with the largest number being Hispanics (18%) followed by Blacks (12%). About 96 million people of Hispanic and Black ancestry reside in the U.S. (U.S. Census Bureau, 2015). Further, Hispanics and Blacks are overrepresented in the U.S. criminal justice system. Although Blacks and Hispanics make up about 30% of the general U.S. population, they constitute more than 60% of state and federal inmates (Harrison & Beck, 2005).

Given the characteristics of actuarial risk assessment tools, such as explicit risk factors and clear combination rules by statistical methods, actuarial tools should be less vulnerable to the cultural/racial bias than unstructured clinical judgment (Garb, 1997). Nevertheless, the potential for cultural bias of the actuarial risk assessment

tools has remained largely unexplored, possibly due to the relatively small numbers of ethnic minorities in development and normative samples. Consequently, research is needed to examine whether Static-99R, which was developed with predominately White offenders, would perform similarly across different ethnic groups.

There are possible theoretical explanations that risk assessment tools that work for White offenders would perform differently among ethnic minority offenders. First, it has been theorized that the unfair social structure under white-dominated and the post-colonial societies toward ethnic minority populations might lead to substantially different patterns on the risk-relevant characteristics from those of White offenders (i.e., more socially disorganized, poverty, less opportunities to achieve their goals; Cernkovich, Giordano, & Rudolph, 2000; Sampson, Morenoff, & Raudenbush, 2005).

Ethnic minority/immigrant groups might have their own cultural-specific risk factors (e.g., acculturation, collectivism, patriarchy, and loss of face; Goldsmith, Hall, Garcia, Wheeler, & George, 2005). These culture-specific risk factors might be directly associated with sexual recidivism or indirectly by prompting the development of other empirical risk factors, such as hostility toward women and general social rejection. Ignoring these potential cultural risk factors in risk assessment tools may restrain the predictive ability of the scales.

Regardless of social structure and possible cultural-specific risk factors, their cultural values or immigrant status could influence the predictive accuracy of the Static-99R by restricting the amount of necessary information for proper administration. For example, the tendencies of underreporting sexual crimes due to the fear of losing face in the community (Hall, 2002) and disgracing their groups (e.g., families; Hall et al., 1998; Moro, 1998) as well as undetected crimes occurring in their country of origin might undermine the predictive accuracy of scales.

Four studies, to date, have been conducted in the United States to evaluate the applicability of Static-99R to different ethnic sex offender groups: Whites, Blacks, and Hispanics. Those studies consistently found that Hispanic sex offenders had relative low mean scores of Static-99R and low sexual recidivism rates, while Black sex offenders had relatively high mean scores and high sexual recidivism rates. Findings, however, about the predictive accuracy of Static-99R across different ethnic groups have been inconsistent although there is a pattern of relatively weak discrimination and calibration of Static-99R for Hispanic sex offenders (Table 1.1).

Varela and colleagues examined the predictive validity of Static-99/R among 1,911 sex offenders released from prison in Texas (White, $n = 912$; Black, $n = 411$; and Hispanic, $n = 588$; Varela et al., 2013). They found poor to moderate discrimination of Static-99/R for violent sexual recidivism (i.e., any contact sexual offense) across different ethnic groups (AUCs of .57 for Hispanics, .59 for Whites, and .65 for Blacks) and no significant differences among the ethnic groups. The base rates of any contact re-offense were very similar across ethnic groups (2.4% to 3.1% in the follow-up of 4.8 years). No calibration analyses were performed.

In the state of California, Hanson and colleagues (2014) reported good predictive validity with 475 sex offenders under the parole system (AUC of .82 [.72, .92]) and overall good and similar discrimination across all ethnic groups (e.g., White, Black, and Hispanic; AUCs $> .74$; odds ratios > 1.47). The overall sexual recidivism base rates were relatively low (4.8% after 5 years), particularly low in Hispanic sex offenders (2.5% after 5 years). The overall fit between the expected and observed recidivism rate was generally good (4.8% vs. 6.31%; $E/O = 1.30$ [0.87, 1.96], with over-prediction only in Low-Moderate risk category (scores of 2 and 3; $E/O = 4.58$ [1.15, 18.31]). Calibration analyses were not conducted for each ethnic

group due to the small number of total sample size and recidivists (e.g., 7 recidivists in a total of 99 Black sex offenders).

In the study with 483 Hispanic sex offenders (9 sexual recidivists) released from prisons in New Jersey (Leguizamo et al., 2017), Static-99R was able to discriminate recidivists from non-recidivists (AUC of .72 [.59, .91]). Further, they divided the sample into two separate sub-groups: U.S. born Hispanic ($n = 268$, 6 recidivists) and those born in Latin America ($n = 215$, 3 recidivists). Static-99R worked only for U.S. born Hispanic (AUC of .82 [.64, .99]). They also found relatively low sexual recidivism rates in this study (1.9%, 9/483 with an average follow-up of 6 years) and the observed 5-year overall recidivism rate in the current sample was lower than the expected recidivism rates from Static-99R norms (3.0% vs. 6.3%; $E/O = 2.11$ [1.10, 4.06]).

In the most recent study, Boccaccini and his colleagues (2017) examined the predictive validity of Static-99/R for three different ethnic groups (White, $n = 7,938$; Black, $n = 9,725$; Hispanic, $n = 8,939$). These individuals were drawn from the Texas Department of Criminal Justice (TDCJ) between 1999 and 2011 (Boccaccini et al., 2017). They found very similar and moderate discrimination for Static-99R for sexual recidivism across the different ethnic groups (AUCs of .65 for White, .64 for Black, and .63 for Hispanic). They also divided the sample into U.S. born Hispanic ($n = 6,337$) and Hispanic born outside the U.S ($n = 2,459$). Contrary to Leguizamo et al., (2017), Static-99R worked better for Hispanic born outside the U.S. than those born inside (AUC of .65 [.52, .78] and AUC of .61 [.58, .65], respectively), but the difference was not statistically significant. Calibration analyses for each ethnic subgroup were not reported.

In summary, previous research has found differences in average Static-99R

scores across ethnic groups, whereas no clear conclusions can be reached concerning differences in predictive accuracy. Although group differences in mean scores on risk scales do not necessarily indicate test bias, this information would be a starting point to investigate potential differences in predictive accuracy (Reynolds & Suzuki, 2013).

There are two major propensities associated with the risk for sexual recidivism: general criminality and sexual criminality. Both types of risk factors are included in the commonly used sexual offender prediction tools, including Static-99R (Brouillette-Alarie et al., 2016). For example, Static-99R contains general crime factors (e.g., prior nonsexual violence and any prior sentencing dates) as well as sexual crime specific risk factors (e.g., prior sex offenses and any male victims). It is possible that there are ethnic differences on either or both of two major dimensions. In particular, relatively higher scores on the Static-99R for Black sex offenders might be attributed to their considerable overrepresentation in the arrest and victimization rates, particularly in violent crimes (e.g., murder, robbery; Federal Bureau of Investigation, 2013; Harrell, 2007). There have been few studies, however, that have examined ethnic differences in sex crime specific risk factors.

Table 1.1

Summary description of previous studies

Study	Group	Base rate (%, n/N)	Follow-up (Year)	Mean Score (<i>SD</i>)		AUC [95% C.I.]	
				Static-99	Static-99R	Static-99	Static-99R
Forbes (2008)	Black	-	-	3.52 (1.80)	-	-	-
	White	-	-	2.36 (1.87)	-	-	-
Varela et al. (2013) ¹	Black	2.7% (11/411)	4.9	3.66 (1.81)	3.29 (2.09)	.58 [.43, .73]	.65 [.51, .78]
	White	2.4% (22/912)	4.9	2.90 (1.87)	2.19 (2.56)	.57 [.45, .70]	.59 [.45, .72]
	Hispanic	3.1% (18/588)	4.6	2.51 (1.74)	2.08 (2.20)	.59 [.45, .73]	.57 [.41, .73]
Hanson et al. (2014) ²	Overall	4.8% (23/475)	5.0	2.6 (2.1)	2.2 (2.2)	.82 [.72, .92]	.82 [.72, .92]
	Black	7.1% (7/99)	5.0	3.2 (2.1)	2.7 (2.1)	.75 [.55, .95]	.77 [.56, .97]
	White	7.1% (10/140)	5.0	2.9 (2.3)	2.3 (2.4)	.86 [.72, .99]	.85 [.72, .98]
	Hispanic	2.5% (5/200)	5.0	2.0 (1.8)	1.8 (2.2)	.75 [.40, .99]	.73 [.41, .99]
Leguizamo et al. (2017) ²	Overall	1.9% (9/483)	6.1	1.86 (1.90)	1.64 (1.88)	.68 [.48, .88]	.72 [.53, .91]
	US-born Hispanic/ Puerto Rican	2.2% (6/268)	6.1	2.03 (1.46)	1.81 (1.91)	.77 [.61, .94]	.82 [.64, .99]
	Other Hispanic	1.4% (3/215)	6.1	1.64 (1.30)	1.43 (1.81)	.47 [.06, .99]	.52 [.19, .86]
Boccaccini et al. (2017) ²	Black	4.5% (437/9,725)	5.2	-	-	.63 [.60, .65]	.64 [.61, .67]
	White	4.9% (389/7,938)	5.2	-	-	.64 [.62, .68]	.65 [.62, .68]
	Hispanic	3.0% (268/8,939)	5.2	-	-	.64 [.61, .68]	.63 [.60, .67]
	Hispanic born in U.S.	3.9% (247/6,337)	5.2	-	-	.62 [.58, .65]	.61 [.58, .65]
	Hispanic born outside U.S.	0.7% (17/2,459)	5.2	-	-	.67 [.54, .79]	.65 [.52, .78]

1. Based on violent sexual recidivism (i.e., any contact sex offense)

2. Based on sexual recidivism (i.e., non-contact and contact sex offenses)

Current Study

This prospective study evaluated the predictive validity of Static-99R across three major ethnic groups (e.g., White, Black, Hispanic) with a total of 2,101 sexual offenders in California. Offenders' sexual recidivism information was examined for 5 years following release from custody and obtained from official criminal records of the California Department of Justice.

The main research questions were the following:

Research Question 1: Do the minority ethnic groups (e.g., Black and Hispanic) score higher or lower on Static-99R than White sexual offender groups?

Research Question 2: How well does Static-99R discriminate between sexual recidivists and non-recidivist for different ethnic groups (e.g., White, Black, Hispanic)?

Research Question 3: Are there any significant differences in sexual recidivism rates (i.e., base rates) within ethnic groups and from the norms (calibration)?

Method

Sample

We combined a new cohort sample ($n = 1,626$) and a previous study sample from California ($n = 475$; Hanson et al., 2014) in order to increase statistical power for ethnic subgroup analyses (e.g., White, Black, and Hispanic). The ethnic information was recorded on their criminal history record, and it was typically self-reported. Of the overall 2,101 offenders, 37.6 % ($n = 789$) were White, 22.2 % ($n = 466$) were Black, 34.2 % ($n = 719$) were Hispanic and 6% ($n = 127$) were Others/Unknown.

On average, the offenders were 42.9 years at release ($SD = 11.6$; range of 19.6

to 86.6). Hispanic sex offenders ($M = 40.5$, $SD = 12.0$) were significantly younger than White ($M = 45.2$, $SD = 13.3$) and Black sex offenders ($M = 43.1$, $SD = 10.49$); the age difference between Black and White sex offenders was also statistically significant.

Measures

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012).

Static-99R is a 10-item empirical actuarial risk tool designed to predict sexual recidivism among adult male offenders. Static-99R is identical to Static-99 with the exception of revised age weights. The total score (ranging from -3 to 12) is calculated by summing all item points and can be used to place offenders in one of five risk categories: I (very low, -3 and -2), II (below average, -1 and 0), III (average, 1 to 3), IVa (above average, 4 and 5), and IVb (well above average, 6+; Hanson et al., 2017). Static-99R scores in this study were later computed from Static-99 scores by using the offender's date of birth to calculate the updated age item.

Rater Reliability. Although rater reliability of the Static-99R was not directly assessed in this study, a previous study (Hanson et al., 2014) found overall good inter-rater reliability (Intraclass correlation coefficient [ICC] = .78 [.64, .90] for absolute agreement using a two-way and random effects model) in a sample of 55 California parole and probation officers (ICC = .81, $n = 30$; ICC = .77, $n = 25$, respectively). The raters were volunteers who scored 14 randomly selected California case files.

Recidivism. We examined three different recidivism outcomes, defined as arrests after release as either parolees or probationers: sexual, violent, and any recidivism. Sexual recidivism included any offense that was considered sexually motivated (contact and non-contact sex offenses). Violent recidivism included all

crimes that involved direct confrontation with the victim. This category included contact sexual offenses but excluded non-contact sex offenses. Finally, any recidivism included all crimes (sexual, violent, non-violent), as well as all technical offenses (e.g., breach of conditional release), regardless of whether they were sexually motivated.

Procedure

Offenders were scored on Static-99/R by CDCR or probation staff as part of routine practice. Beginning in 2006, CDCR policy required that all released parolee sexual offenders were scored on Static-99/R and in 2008, scoring on the Static-99 became mandatory for sex offenders released on probation. Recidivism information was provided by the California Department of Justice as of October 2015. Recidivism was defined as an arrest for a sexual, violent, and any offense.

Plan of Analysis

For discrimination, we used two statistical methods: 1) the area under the curve (AUC) from receiver operating characteristic (ROC) analysis (Swets et al., 2000) and 2) odds ratios from logistic regression (Hosmer & Lemeshow, 2002).

For calibration, we used the *E/O* index (the ratio of expected number of recidivists divided by an observed number of recidivists; Hanson, 2017) as well as a fixed-effect meta-analysis of logistic regression parameters (Borenstein et al., Hanson & Broom, 2005).

Area Under the Curve (AUC). AUC values indicate the probability that a randomly selected recidivist would have a more deviant score than a randomly selected non-recidivist. AUC can vary between 0 and 1, with .50 indicating the level of prediction that would be expected by chance. According to Rice and Harris (2005),

AUCs of .56 would be considered small, .64 would be moderate, and .71 would be large. AUC values are expected to be smaller in prognostic studies than in diagnostic studies because the outcome of interest in prognostic studies does not exist at the time of assessment, and may never happen (Helmus & Babchishin, 2017; Royston, Moons, Altman, & Vergouwe, 2009). It has an advantage of insensitivity to base rates and robustness to outliers (Ruscio, 2008).

Odds ratios. Odds ratios indicate the change in relative risk associated with one unit change in Static-99/R scores. For example, Static-99R scores are associated with a consistent relative risk increase of approximately 1.45 (Hanson et al., 2016), which means the rate of recidivism increases 1.45 times as each Static-99R score increases. The primary advantage of odds ratios is that they are less influenced by the restriction of range compared to AUCs (Hanson, 2008).

E/O index. The *E/O* index is the expected number of recidivists divided by an observed number of recidivists (Hanson, 2017). Perfect calibration is indicated by an *E/O* index of 1.0. Following Rockhill, Byrne, Rosner, Louie, and Colditz (2003), the 95% confidence intervals for the *E/O* indices were computed as follows:

$$95\%CI(E/O) = (E/O)\exp(\pm 1.96\sqrt{1/O})$$

The expected number of recidivists was based on the 5-year sexual recidivism rates for routine/complete samples reported by Hanson et al. (2016).

Comparing Logistic regression parameters. A second method of testing calibration was to examine the extent to which logistic regression parameters, such as intercept values (centered on Static-99R scores of 2) differed from the logistic regression parameters for the norms (Table 7: $B0_2 = -2.827$, $SE = 0.079$; $B1 = 0.368$, $SE = 0.025$; Hanson et al., 2016). Specifically, the $B0_2$ represents the expected recidivism rate for a Static-99R score of 2 (p_2) in logit units ($\ln[p_2/\{1-p_2\}]$).

Differences between the parameters in the current sample and those of the norms were tested using fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005).

Results

Of the total sample, 45.4% (951/2,101) of offenders were arrested with any offense; 4.0% (85/2,101) were arrested with a violent offense; 4.8% (101/2,101) were arrested with a sexual offense during the fixed 5-year follow-up period. Black sex offenders had the highest sexual recidivism rates (6.4%), and Hispanic and Other/Unknown groups had relatively lower sexual recidivism rates than other groups (3.1% and 2.4%, respectively; Table 1.2).

Of the new cohort sample ($n = 1,626$; excluding the 2014 study sample), about 19.2% (5/26) of sexual re-offenses by probationers and 32.7% (17/52) of sexual re-offenses by parolees were committed by offenders who were registered as transients at the time of re-arrest. In comparison, only about 6% (6,316/103,737) of the total registered sex offenders in the community are transient (California DOJ, 2016, June). Collectively, transient status seems to be associated with higher sexual recidivism rates (overall odds ratio = 6.06 [3.70, 9.93]).

Discrimination

Across ethnic groups, there were significant differences in the average Static-99R scores, $F(3, 2,097) = 25.56, p < .001$. As can be seen in Table 1.2, Black sex offenders ($M = 3.06$) scored significantly higher than White, Hispanic, and Other/Unknown groups, all of which had very similar average scores (mean range of 1.97 to 2.04).

Using fixed 5-year follow-up, Static-99R was able to discriminate recidivists

from non-recidivists for all ethnic groups although AUC value of Other/Unknown group was not significant due to low sample size. The White group had the highest AUC value of .817 [.756, .877] and Hispanics had the lowest AUC value of .702 [.589, .814] (Table 1.2).

Table 1.2

Five-year Sexual Recidivism Rates, Static-99R Scores, and AUC values for Whites, Blacks, and Hispanics

Groups	Sexual recidivism rates (%)	Number of recidivists/total	Static-99R <i>M (SD)</i>	AUC	95% CI	
					Lower	Upper
White	5.83	46/789	2.04 (2.44)	.817	.756	.877
Black	6.44	30/466	3.06 (2.32)	.738	.638	.839
Hispanic	3.06	22/719	1.97 (2.17)	.702	.589	.814
Other/ Unknown	2.36	3/127	1.97 (2.15)	.727	.317	1.000
Total	4.81	101/2,101	2.24 (2.35)	.771	.723	.819

Note. Based on fixed 5-year sexual recidivism analysis. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

The relationship between Static-99R scores (centered on a score of 2) and sexual recidivism also acceptably fit a logistic distribution (i.e., Hosmer-Lemeshow test was not significant: $\chi^2 = 3.65$, $df = 5$, $p = .600$; $B0_2 = -3.619$, $SE = .152$; $B1 = .456$, $SE = .044$; *Figure 1.1*).

The 5-year sexual recidivism rates at a score of 2 across all ethnic groups were very similar (2.4% to 3.0%; $Q_{\text{between}} = .47$, $df = 2$, $p = .792$). The discrimination (change in relative risk) was highest for White offenders (odds ratios = 1.39 to 1.65), but the differences between racial groups were not statistically significant ($Q_{\text{between}} = 2.29$, $df = 2$, $p = .318$; Table 1.3).

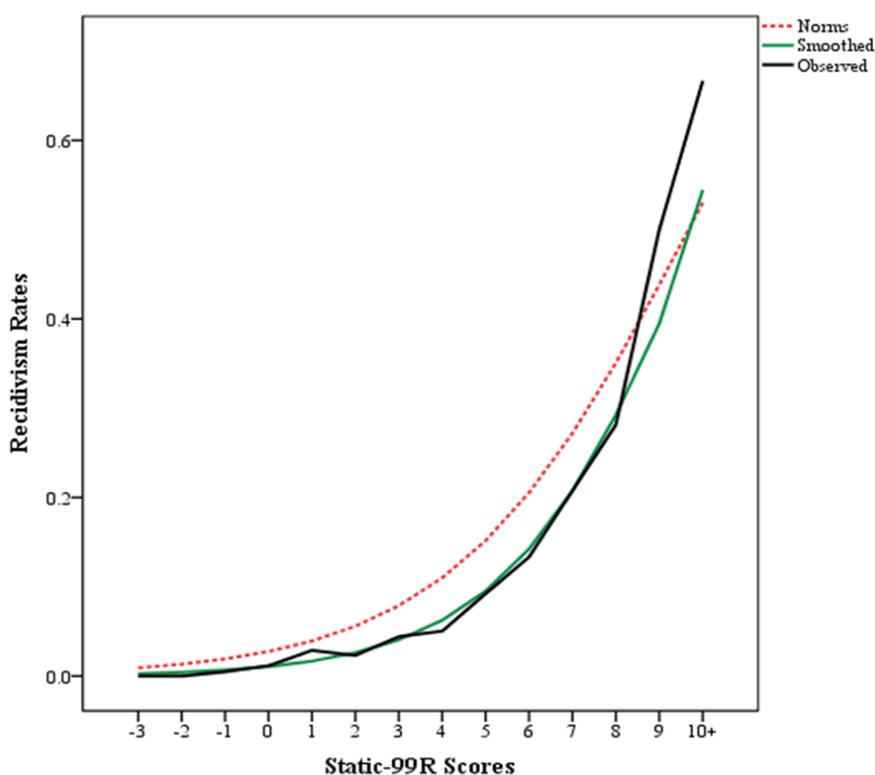


Figure 1.1. Logistic curve for overall sample with the norms

Table 1.3

Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism across Different Racial/Ethnic Groups

Groups	$B0_2$ (Base rate)	$B1$ (Relative risk)	Static-99R Odds Ratio	95% CI	
				Lower	Upper
White	-3.47 (3.0%)	.50	1.65	1.45	1.89
Black	-3.63 (2.6%)	.45	1.58	1.33	1.87
Hispanic	-3.70 (2.4%)	.33	1.39	1.16	1.67
Average (fixed-effect)	- 3.58 (2.7%)	.45	1.56	1.43	1.71
$Q(df = 2)$.47, $p = .792$	2.29, $p = .318$			
I^2	.00	.13			

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Calibration

The overall resulting logistic equation indicated a relative risk increase of 1.58 for each increase in Static-99R score ($e^{.456} = 1.58$, and an adjusted 5-year sexual recidivism rate of 2.6% for a Static-99R score of 2 ($[1/\{1+e^{-(3.619)}\}] = .0261$). When compared to the norms (from Hanson et al., 2016), the adjusted (score of 2) base rate was significantly lower ($B0_2$ of -3.62 vs. -2.83; $Q_{\text{between}} = 21.33$, $df = 1$, $p < .001$), and the discrimination (BI) was larger, but not significantly ($BI = .456$ vs. $.368$; $Q_{\text{between}} = 2.94$, $df = 1$, $p = .086$).

Overall, adjusted base rates ($B0_2$) of each ethnic group were significantly lower than the norms (2.4% to 3.0% vs. 5.6%; all p -values $< .05$). Relative risk rates did not significantly differ from one another or from the norms (Table 1.4).

Table 1.4

Comparison of Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism with Meta-Average (from Hanson et al., 2016)

Static-99R	Norms	Overall	White	Black	Hispanic
<u>Base rate</u>					
$B0_2$ (SD)	- 2.83 (.079) (5.6%)	-3.62 (.170) (2.6%)	-3.47 (.235) (3.0%)	-3.63 (.330) (2.6%)	-3.70 (.260) (2.4%)
Q_{between}		21.33 ***	6.66 **	5.59 *	10.29 **
<u>Relative risk</u>					
BI (SD)	.368 (.025)	.456 (.044)	.503 (.068)	.454 (.087)	.331 (.092)
Q_{between}		2.94	3.45	.89	.16

Note. *** $p < .001$, ** $p < .01$, * $p < .05$

In comparison to norms for routine samples, the observed 5-year overall recidivism rate in this current sample was lower (4.8% vs. 8.1%; E/O index = 1.68 [1.39, 2.04]; Table 1.5). *Figure 1.1* provides a plot of the observed recidivism rates per Static-99R risk score, the rates based on the smoothed logistic curve fitted to this data, and the recidivism rate norms for routine samples (Hanson et al., 2016). As can be seen in *Figure 1.1*, the general pattern is that the recidivism rates in the current sample were lower than expected, except for the IVb (well above average risk) category.

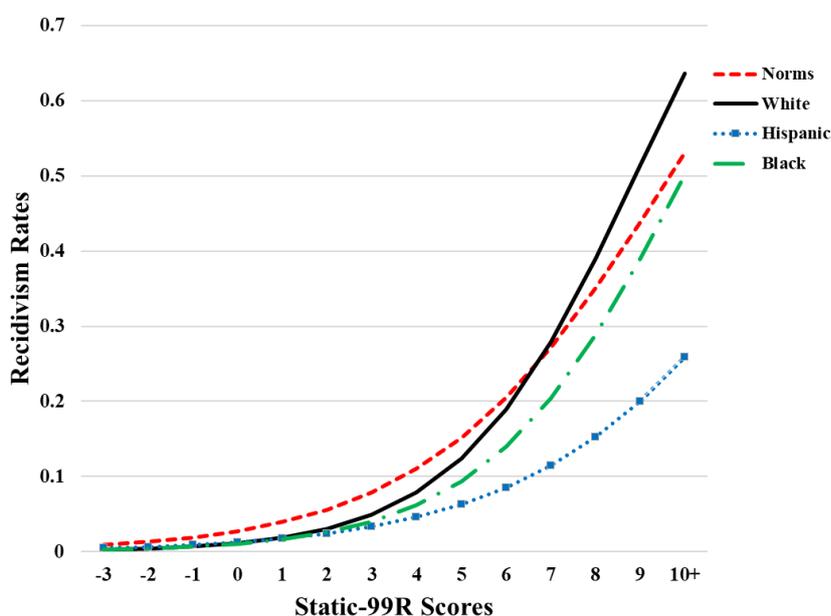


Figure 1.2. Logistic curves for each ethnic group with the norms

For White sexual offenders, the observed 5-year overall recidivism rate was slightly lower than expected rate (5.8% vs. 7.8%; E/O index = 1.34 [1.00, 1.79]; Table 1.5 and *Figure 1.2*). For Black sex offenders, the observed 5-year overall recidivism rate was also lower than the expected rate (6.4% vs. 10.4%; E/O index = 1.61 [1.12, 2.30]), but significantly only in category IVa (scores of 4 and 5; E/O index = 3.15 [1.31, 7.56]; Hanson et al., 2016). For the Hispanic sample, the observed 5-year

overall recidivism rate was lower than the expected rate (3.1% vs. 7.1%; *E/O* index = 2.33 [1.54, 3.55]), specifically in category III and IVa (scores of 1 to 5; *E/O* index = 2.42 [1.26, 4.66] and 2.17 [1.03, 4.55]).

Table 1.5

Comparison of Expected and Observed 5-year Sexual Recidivism Rates for Static-99R Standardized Risk Levels

Static-99R Risk Category	Sample size	Recidivists		E/O index	95% C.I.	
		Observed	Expected		Lower	Upper
Overall sample						
I	60	0	.66	-	-	-
II	461	4	11.0	2.75	1.03	7.33
III	984	31	57.7	1.86	1.31	2.65
IVa	419	27	52.0	1.92	1.32	2.81
IVb	177	39	48.8	1.25	.91	1.71
Total	2,101	101	170.1	1.68	1.39	2.04
White sample						
I	29	0	.30	-	-	-
II	206	1	4.91	4.91	.69	34.86
III	338	12	19.5	1.62	.92	2.86
IVa	148	15	18.1	1.21	.73	2.00
IVb	68	18	18.8	1.04	.66	1.65
Total	789	46	61.5	1.34	1.002	1.79
Black sample						
I	6	0	.06	-	-	-
II	51	0	1.22	-	-	-
III	215	10	12.8	1.28	.69	2.37
IVa	127	5	15.7	3.15	1.31	7.56
IVb	67	15	18.5	1.23	.74	2.05
Total	466	30	48.3	1.61	1.12	2.30

Hispanic sample						
I	22	0	.26	-	-	-
II	171	2	4.10	2.05	.51	8.20
III	370	9	21.8	2.42	1.26	4.66
IVa	120	7	15.2	2.17	1.03	4.55
IVb	36	4	10.0	2.51	.94	6.68
Total	719	22	51.4	2.33	1.54	3.55

Note. I - very low risk, II - below average risk, III - average risk, IVa - above average risk, IVb - well above average risk (Hanson et al., 2017). Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Discussion

The main purpose of this study was to examine how Static-99R predicts sexual recidivism risk (e.g., discrimination and calibration) for two major ethnic minorities (Black and Hispanic) as well as White sex offenders in the U.S. This prospective study found overall good predictive accuracy of Static-99R across the ethnic sex offender groups (White, Black, and Hispanic). Consistent with the findings from the previous studies, Black sex offenders had the highest Static-99R score and highest sexual recidivism rates, while Hispanic had a relatively lower Static-99R score and lower sexual recidivism rates.

The discrimination of Static-99R across ethnic groups (White, Black, and Hispanic) were generally all good (all AUCs $> .70$; odds ratios > 1.39) given the average values of relative accuracy (AUC = .70 and odds ratio = 1.45; Helmus, Hanson, et al., 2012; Hanson et al., 2016). There was, however, a consistent pattern across ethnic groups of the largest discrimination for White and the lowest for Hispanic sex offenders. Although there is weak evidence for significant differences of discrimination between ethnic groups, the consistently lower discrimination for Hispanic sex offenders than other ethnic groups (i.e., White and Black) should be carefully examined by further studies.

In terms of the match between the expected and the observed recidivism rates (calibration), base rates (at a Static-99R score of 2) across ethnic groups were very similar (2.4% to 3.0%). All groups, however, were significantly lower than norms (5.6%; Hanson et al., 2016). In particular, the overall sexual recidivism rate of Hispanic sex offenders was substantially lower than the norms (i.e., poorer calibration) as compared to other groups (3.1% vs. 7.1%; $E/O = 2.33$).

There are several possible explanations for the consistent findings of relatively weak discrimination and calibration among minority ethnic groups, particularly for Hispanic sex offenders as a minority/immigrant group. There might be a different association between the major constructs (general and sexual criminality; Brouillette-Alarie et al., 2016) of Static-99R and sexual recidivism under a white-dominated social structure or different cultural backgrounds. Further, Hispanics might have their own cultural-specific risk factors, not considered in the current risk assessment risk scales. More research is needed to identify unique risk factors or propensities associated with sexual recidivism under a Hispanic cultural background.

One study found that there was a similar level of sexual criminality across different ethnic groups (e.g., Indigenous, Black, Hispanic, Asian) whereas there were ethnic differences in general criminality (Brankley, Lee, Hanson, & Zabarauckas, 2017b). Specifically, Hispanic sex offenders scored significantly lower on risk factors relevant to general criminality than White sex offenders, but Blacks scored much higher than Whites. Further research, consequently, is needed on how well the risk factors measuring general criminality in Static-99R are associated with sexual recidivism among the ethnic minority groups.

It is also possible that a lack of necessary information (e.g., criminal records in their countries of origin) for the proper administration of Static-99R might restrain the

predictive ability of the scales. For example, one-third of Mexicans (the largest group in the Hispanic group) in the United States are foreign-born, compared with 13% of the U.S. population overall (López, 2015). Consequently, lack of criminal history information due to their immigrant status might affect the predictive accuracy of the Static-99R scale (Leguizamo et al., 2017). In addition, given their group-oriented cultural values (e.g., patriarchy, collectivism), the underreporting tendency of sexual crimes/victimizations that occurred in their ethnic community (e.g., families), might also undermine the predictive accuracy of Static-99R.

The overall sexual recidivism base rate of this California sample was significantly lower than the norms (4.8% after 5 years). The reasons for the lower than expected rates are not fully known, but may be related to the research method used (e.g., the accuracy of records), the effectiveness of practices for managing sexual offenders in California, or other factors not fully understood. Further, these low recidivism rates might be attributed to unexpectedly low sexual recidivism rates of the parolee samples (4.5%) over the probationer sample (6.1%; Lee, Restrepo, Satariano, & Hanson, 2016). Further studies are necessary to examine specific factors that may contribute to this low recidivism rate of the parolee sample (e.g., sexual offender treatment, GPS; Gies et al., 2012).

This is the largest prospective study, to date, evaluating whether cultural bias exists in predictive accuracy of Static-99R with White, Black, and Hispanic sex offender groups. Although the average scores of Static-99R scale were significantly different across the ethnic groups, there was no strong evidence that the risk factors/propensities had any differential associations with sexual recidivism for sexual offenders with different cultural backgrounds. Consequently, the current findings support the use of Static-99R for White, Black, and Hispanic sex offenders.

Limitations

Although the overall sample was large (101 recidivists), the sub-analyses were not (e.g., 22 Hispanic recidivists). Additional research with a large number of each ethnic group is recommended for more confident conclusions about minority/ethnic sex offenders.

Recidivism information for this study was provided solely by the California Department of Justice in the first study group (Hanson et al., 2014). This limited recidivism information (nationwide criminal records [FBI] were examined only for the 2016 study group) would affect predictive accuracy, including the validity of the absolute recidivism estimates. This concern is particularly relevant to Hispanic sex offenders whose reoffending may be less likely to be detected (e.g., if they frequently leave the United States).

We did not have item-level data and could not examine if the predictive accuracy of each item or propensities (i.e., sexual deviance, or general criminality) varied across ethnic groups. Although Hispanic and Black populations constitute a large proportion of the California population, there are still other minority ethnicities (e.g., Asians, Native Americans) for which we have very limited information.

Implications for Research

Given the relatively low statistical power, additional studies and meta-analyses are needed with larger samples of these (Black, Hispanic) and other ethnic minorities (e.g., Asians, Native-Americans).

Researchers should also empirically explore potentially unique risk factors for each ethnic group. As well, research should examine the extent to which the two major propensities (general and sexual criminality) are associated with sexual recidivism for non-White sex offenders with different cultural backgrounds. Further,

scoring additional risk measures (e.g., Static-2002R) could also provide insight into potential ethnic differences in the constructs related to sexual recidivism risk (i.e., cultural bias in construct validity; see Brouillette-Alarie et al., 2016).

More research is also necessary to identify the extraneous factors (e.g., immigrant status, underreporting) for ethnic minorities which might moderate the predictive accuracy of risk assessment scales. In other words, do the low sexual recidivism rates in a particular ethnic group (e.g., Hispanics) remain after controlling for those potential extraneous factors? In addition, management practices for sexual offenders implemented in each jurisdiction (e.g., sexual offender treatment, supervision, tracking devices) should be considered to evaluate variation in the correspondence of observed recidivism rates with absolute recidivism rates (calibration; risk band norm designs, see Woodrow & Bright [2011]).

Implications for Practice

The current findings support the use of Static-99R in risk assessment procedures for sex offenders of White, Black, and Hispanic heritage. It is noteworthy that the present study was one of the first studies to examine the performance of Static-99R with Black sex offenders, and the current findings add to the evidence of good performance of Static-99R for this United States ethnic minority.

The current results, however, suggest that evaluators should be cautious when using Static-99R to estimate absolute sexual recidivism rates for Hispanic sex offenders. Although Hispanic sex offenders with low scores had low recidivism rates (as expected), the observed recidivism rates for Hispanic offenders with higher Static-99R scores were lower than expected. Consequently, it is difficult to affirm that high scores justify the same interpretations for Hispanic individuals as they do for offenders of other ethnic groups (e.g., Black, White).

In order to determine whether a high score for Hispanic sexual offenders identifies truly high-risk individuals, evaluators will need to consider factors outside of the Static-99R risk tool. These factors could include their criminogenic needs, protective factors, and any atypical reasons for scoring the items (i.e., artificially high scores).

Evaluators should also be alert to the possibility of immigrant Hispanic sex offenders reoffending in their country of origin. Although intercountry records are often incomplete and difficult to obtain, they may, nonetheless, be worth considering for individuals who spend substantial time out the country or who perceived the opportunities to offend to be greater in their country of origin than in the U.S.

Conclusion

Within psychometrics, bias is an empirical term (i.e., statistically estimated quantity) rather than a principle established through debate and opinion (Reynolds & Suzuki, 2013). If there are any statistically significant differences in the outcomes due to cultural variables, such as ethnicity/race, cultural bias is empirically present.

Potential cultural bias in risk assessment tools has been an issue in the criminal justice and forensic mental health system because of the overrepresentation of ethnic minorities. Although high scores do not indicate bias, they are a sign that bias might be present. Given that there will always be relatively small proportions of ethnic minorities in validity studies of the total sample, specific studies of ethnic subgroups are needed to examine the potential for bias.

In this particular study, cultural bias in the predictive validity of Static-99R was evaluated among White, Black, and Hispanic sex offenders, and the findings indicated that the Static-99R works well to predict the likelihood of sexual recidivism across these ethnic sex offender groups. Further research is needed to explore

potential ethnic differences in the major psychological dimensions associated with sexual crime (antisocial orientation, sexual deviancy).

Chapter 4: Study 2

Paraphilia and Antisociality: Motivations for Sexual Offending May Differ for American Whites and Blacks

Sexual offending is a serious concern across different countries and different cultural groups (Jina & Thomas, 2013; Mason & Lodrick, 2013). For the last few decades, considerable effort has been devoted to understanding and managing individuals with a history of sexual crime. Most of this research, however, has been based on Whites who committed a sexual crime. There is little understanding of the sexual offending behavior of different racial or ethnic groups.

The United States is multicultural and multiethnic society. Approximately 40% of the United States population identifies as a racial or ethnic minority (U.S. Census Bureau, 2017a). The race is a social-construct based on individuals' phenotypic characteristics (e.g., skin, hair, or eye color; White, Black, or Asian). Although the terms of "race" and "ethnicity" are often used interchangeably, ethnicity has a broader meaning than race. Ethnicity is a category of people who are sharing common cultures (e.g., language, social norms, religion, or even racial characteristics; Cornell & Hartmann, 2007). A racial group (e.g., White), therefore, comprises multiple ethnic groups (e.g., Hispanic, Irish, or Italian).

According to the U.S. Census Bureau (2018), there are five racial categories used in the U.S.: White (including Hispanic; 76.9%), Black (13.3%), American Indian (1.3%), Asian (5.7%), and Native Hawaiian (0.2%; U.S. Census Bureau, 2017a). Blacks are the most overrepresented racial group within both the general U.S. criminal justice system and among those convicted of sexual offenses. Currently, Blacks makeup 13% of the U.S. general population (U.S. Census Bureau, 2017a), but represent 22% of state or federal prisoners incarcerated for either rape or sexual

assault offenses (Carson, 2018). Despite considerable efforts made to understand the causes of the disproportionately high rates of criminal convictions among Blacks, to date, little research has examined the sexual offending behavior of Blacks.

The few available studies have found that, compared to Whites, Blacks who had committed a sexual offense showed lower levels of social achievement and less deviant sexual interests (Burton & Ginsberg, 2012; Schaaf, Jeglic, Calkins, Raymaekers, Leguizamo, 2016). There is insufficient research, however, to make strong conclusions concerning racial differences. Consequently, it is not surprising that Blacks with a history of sexual offending are routinely provided with the same correctional services (rehabilitation programs, community supervision) as those provided to Whites. To the extent that racial and ethnic differences exist, it may be possible to improve the effectiveness of correctional intervention by tailoring services to distinctive characteristics of Blacks (i.e., the Need and Responsivity Principles from Andrews and Bonta's [2010] theory of correctional rehabilitation; Hanson et al., 2009).

Racial Differences in Sex Offending

The onset and persistence of sexual offending are broadly associated with two major propensities: general criminality (e.g., impulsivity, juvenile delinquency, hostility toward women), and sexual criminality (sexual crime specific factors; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). The major dimensions of sex crime specific criminality are sexual self-regulation (e.g., sexual preoccupation) and paraphilia (e.g., pedophilia; Brouillette-Alarie & Proulx, 2018; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann et al., 2010; Whitaker et al., 2008). The risk relevant characteristics of sexual offending might differ for Whites and Blacks.

Although there are only a few studies examining the sexual crime specific risk factors for different ethnic groups in the U.S., certain findings have been consistent. First, Whites convicted of sexual offenses appear more paraphilic than Blacks. They show higher sexually deviant arousal (to male children, rape, and exhibitionism; Murphy, DiLillo, Haynes, & Steere, 2001), and are less likely to have committed crimes involving conventional sexual behavior (i.e., adult victim, female victim, vaginal intercourse, less use of pornography; Fix, Falligant, Alexander, & Burkhart, 2017; Forbes, 2007; Kirk, 1975; Leguizamo, Peltzman, Carrasco, Nosal, & Woods, 2010; Waldron, 2012). Second, Whites who commit sexual crimes report a higher rate of childhood sexual and physical abuse than Blacks who commit sexual crimes (Cooper, Murphy, & Haynes, 1996; Fix et al., 2017; Murphy et al., 2001). Similarly, among those who have been abused, the age at first victimization is younger for Whites than for Blacks (the average age of 7 versus 10; Fix et al., 2017).

There is strong evidence that a greater proportion of Blacks than Whites are recorded in official statistics concerning general and violent rule-breaking behavior (e.g., theft, robbery, or assault; Carson, 2018; Hartney & Vuong, 2009). However, theoretical explanations for the high rates of rule-breaking behaviors among Blacks have been widely debated. First, the apparent difference could be attributed to the differential response of the criminal justice system and not on differences in the actual behavior. For example, racial bias and prejudice toward Blacks might directly or indirectly lead police officers to be more likely to stop, search, or arrest Black individuals than Whites in the same circumstances (e.g., racial profiling; Esqueda, 1997; Ioimo, Tears, Meadows, Becton, & Charles, 2007; Weitzer & Tuch, 2002). Supporting this position is research demonstrating that, compared to Whites, Blacks are far more likely to be arrested, and prosecuted, and to receive harsher and longer

sentences (Bales & Piquero, 2012; Barnes & Kingsnorth, 1996; Hartney & Vuoung, 2009).

The second set of explanations focuses on behavioral differences that follow from Blacks' history of slavery, oppression, and social disadvantage. Although officially equal before the law, social inequality persists. In comparison to White Americans, current estimates show that Blacks have lower average annual income (\$36,898 versus \$62,950), higher poverty rates (24% versus 9%), lower homeownership rates (42% versus 72%), and lower education levels (Callis & Kresin, 2018; Proctor et al., 2016; U.S. Census Bureau, 2017b). Consequently, social oppression and discrimination can lead Blacks to seek other or illicit means to achieve common societal goals (e.g., theft, burglary, or robbery; Agnew, 1992; Simons et al., 2003; Tatum, 2002).

Social oppression can increase the risk of developing psychological risk factors associated with general and violent crimes. For example, compared to Whites, Blacks have more antisocial attitudes (e.g., negative attitudes toward society) and antisocial personality (e.g., impulsivity, anger; Agnew, 2006; Arbona & Power, 2003). Further, the disorganized social environments caused by poverty and oppression could inhibit the development of self-control due to the lack of opportunities for conventional bonding (e.g., family, school; Gottfredson & Hirschi, 1990; Reiss, 1951). Self-control is a strong predictor of a variety of delinquent behavior (e.g., vandalism, drug use, assault; Pratt & Cullen, 2000; Vazsonyi & Crosswhite, 2004), and is facilitated by positive, stable relationships with family and friends, and opportunities for success in school and work.

Blacks under an unequal social structure (e.g., poverty or low Socioeconomic status [SES]) might develop a subculture of violence (i.e., norms that support violent

behavior; Anderson, 1999; Brezina, Agnew, Cullen, & Wright, 2004; Wolfgang, 1958). In circumstances where parental monitoring is low, and individuals feel not accepted by society, such individuals may feel insecure in their environments and have little hope for their futures. Individuals with low security may try to protect themselves and learn norms present in the street (e.g., toughness is both a virtue and a necessity; Anderson, 1999; Brezina et al., 2004). Individuals who follow the code of street are more likely to engage in violent acts (e.g., assault, gang fights; Stewart & Simons, 2010).

Predictive Validity of Risk Assessment Instruments across Racial Groups

Even if there are differences in the levels of certain characteristics shown to be associated with sexual crime, further research is required to identify the extent to which these factors have the same meaning across racial/ethnic groups. Over the last few decades, considerable research (on predominantly White samples) has identified the risk factors for sexual recidivism (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann et al., 2010; Whitaker et al., 2008). Many structured risk assessment instruments designed to predict the risk of sexual recidivism have been developed by combining particular risk factors.

The Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) is the most widely used structured (actuarial) risk assessment instrument for adult sexual offenders utilized by forensic experts (Neal & Grisso, 2014). The Static-99R contains both general crime factors (e.g., nonsexual violent offense), and sexual crime-specific risk factors (e.g., non-contact sex offense and any male victims; Brouillette-Alarie et al., 2016; Brouillette-Alarie & Proulx, 2018). Considerable research has found overall moderate predictive validity for the Static-99R (area under the curve [AUC] = .70, $n = 8,106$, $k = 23$; Helmus, Hanson, et al., 2012).

Actuarial risk assessment instruments use pre-specified risk factors and explicit combination rules to estimate the recidivism rates. Consequently, actuarial risk tools provide little room for cultural evaluator bias in administration and interpretation (Dawes et al., 1989; Kleinmuntz, 1990; Meehl, 1954). Nevertheless, actuarial risk tools can be biased when they systematically (i.e., non-random error) overestimate or underestimate the sexual recidivism risk for particular racial/ethnic groups (Reynolds, 2000a; Reynolds & Suzuki, 2013). Given that the development and normative samples of Static-99R included predominantly White samples (an underrepresented racial group in the sex offender population), research is required to evaluate the potential existence of a cultural test bias within Static-99R for racial minority groups (e.g., Blacks, Hispanics, or Asians). Blacks may demonstrate substantially different patterns in the general and sexual criminality due to societal oppression (e.g., poverty), and/or their unique cultural qualities not reflected in predominantly White normative samples (e.g., “Nguzo Saba” [a communitarian African philosophy], religious values, or moral values; Grills & Longshore, 1996; Latzer, 2018; Sanchez, Hamilton, Gilbert, & Vandewater, 2018). Ignoring the potential cultural/racial differences in general and sexual criminality may lead to cultural test bias in the instruments.

There are many possible sources of cultural test bias (e.g., inappropriate content, measuring different constructs, differential predictive validity; Reynolds, 2000a; Reynolds & Suzuki, 2013). Given the main purpose of using Static-99R is assessing the likelihood of sexual recidivism, evaluating cultural bias in predictive validity is necessary and generally sufficient. It is not necessary that the same items function identically for all racial and ethnic groups provided that the results of the overall scale support the same interpretations (i.e., the same inferences from the same

scores). Predictive validity involves two main components: discrimination (how large are the differences in Static-99R scores between those who reoffend and those who do not reoffend) and calibration (how well the estimated recidivism probability from the scale's norms corresponds with the observed recidivism probability of new samples).

Table 2.1 presents five studies that have evaluated the predictive validity of Static-99R for Blacks. Compared to Whites, Black sex offenders are younger and have consistently higher Static-99R scores; however, they have similar sexual recidivism rates. The discrimination (AUC values) of Static-99R for Blacks was moderate but generally lower than for Whites. Concerning calibration, the norms of Static-99R slightly overestimated the sexual recidivism rates for Blacks (Hanson et al., 2014; Lee & Hanson, 2017; Lee et al., 2018). In summary, Static-99R predicts sexual recidivism for Blacks, but perhaps not as well as it does for Whites.

Table 2.1

Predictive Accuracy of Static-99R for Blacks and Whites

Study	Black					White				
	Sexual Recidivism Rates		Follow-up (Year)	<i>M</i> (<i>SD</i>)	<i>AUC</i> [95% C.I.]	Sexual Recidivism Rates		Follow-up (Year)	<i>M</i> (<i>SD</i>)	<i>AUC</i> [95% C.I.]
	%	<i>n/N</i>				%	<i>n/N</i>			
Varela et al. (2013) ¹	2.7%	11/411	4.8	3.3 (2.1)	.65 [.51, .78]	2.4%	22/912	4.8	2.2 (2.6)	.59 [.45, .72]
Hanson et al. (2014) ²	7.1%	7/99	5.0	2.7 (2.1)	.77 [.56, .97]	7.1%	10/140	5.0	2.3 (2.4)	.85 [.72, .98]
Boccaccini et al. (2017) ²	4.5%	437/9,725	5.2	-	.64 [.61, .67]	4.9%	389/7,938	5.2	-	.65 [.62, .68]
Lee & Hanson (2017) ²	6.4%	30/466	5.0	3.1 (2.3)	.74 [.64, .84]	5.8%	46/789	5.0	2.0 (2.4)	.82 [.76, .88]
Lee et al. (2018) ²	10.8%	10/93	10.0	2.9 (2.2)	.63 [.42, .84]	10.6%	14/132	10.0	2.4 (2.4)	.85 [.74, .96]

Note. 1. Based on violent sexual recidivism (i.e., any contact sex offense)

2. Based on sexual recidivism (i.e., non-contact and contact sex offenses)

Current Study

The purpose of this study was twofold: 1) examine whether Whites with a history of sexual offending have different patterns on risk-relevant characteristics from Blacks, and 2) examine whether the Static-99R predicts sexual recidivism risk differently for Whites and Blacks (i.e., cultural test bias). Based on the previous research findings, the main hypotheses were the following:

Hypothesis 1: Blacks will score higher on Static-99R than Whites.

Hypothesis 2: Blacks will be younger and have lower socioeconomic status than the White group.

Hypothesis 3: Blacks will be less likely to report childhood sexual and physical abuse experiences than Whites.

Hypothesis 4: Blacks will have fewer indicators of sexual deviancy (e.g., pedophilia, sexual criminality) than Whites.

Hypothesis 5: Blacks will have more anti-social features than Whites.

Hypothesis 6: Static-99R will show moderate discrimination for Blacks, but less than that observed for Whites.

Hypothesis 7: Static-99R will overestimate recidivism risk for Blacks but not for Whites.

Method

Sample

The sample was from a larger study examining sex offender management, treatment, and civil commitment (Mercado, Jeglic, & Markus, 2011). All individuals in this study were adult males who were convicted of a sexual offense. In total, 1,585 individuals were included in this study; 788 were Black, and 797 were non-Hispanic White. The racial classifications were based on the State of New Jersey Department of

Correction's designations. Of the total sample, 21% ($n = 325$) was legally designated as sexually violent predators (SVPs). The sample was selected from individuals who had been detained at either the Adult Diagnostic Treatment Center (ADTC) or any of New Jersey State Prisons (Table S1). The ADTC is the sex offense-specific treatment center for individuals who are incarcerated as well as who are at highest risk. In practice, most of the clients are individuals whose committed offenses against children, who admit to their crimes, and who are motivated to participate in treatment. Those not classified as SVPs were released from custody between 1996 and 2007.

Measures

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012).

Static-99R is a 10-item empirical actuarial risk tool designed to assess the risk of sexual recidivism among adult males with a history of sexual offending. Static-99R is identical to Static-99 except that it contains revised age weights. The total score (ranging from -3 to 12) is calculated by summing all item points and can be used to place individuals in one of five risk categories: Level I - very low risk (scores of -3 to -2), Level II - below average risk (scores of -1 to 0), Level III - average risk (scores of 1 to 3), Level IVa - above average risk (scores of 4 to 5), and Level IVb - well above average risk (scores of 6 or higher; Hanson et al., 2017). Static-99R scores were computed from Static-99 scores by using the individual's date of birth to calculate the updated age item. The Static-99R total score was found to have good interrater reliability in previous studies ($ICC = .78, [.64, .90]$; Hanson et al., 2014). In this study, the interrater reliability of Static-99 ($n = 30$) was found to be excellent ($ICC = .89$; Quesada, Calkins, & Jeglic, 2014).

MnSOST-R (Epperson et al., 1998). The Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) is a 16-item actuarial risk assessment tool designed to

predict sexual recidivism among adult males with a history of sexual offending. Twelve items are static risk factors (historical), and four items are dynamic risk factors (institutional). In the current study, MnSOST-R total scores were not used; instead, selected items were used to assess sexual criminality (e.g., length of sexual offending history) and general criminality (e.g., adolescent antisocial behavior, substantial drug, and alcohol abuse).

Pervasive Anger Scale. Pervasive Anger Scale was developed as a subscale of Massachusetts Treatment Center: Child Molester Typology - Version 3 (MTC: CM3; Knight, Carter, & Prentky, 1989). It was scored in this study by clinicians at the ADTC. Pervasive anger was measured by five items (e.g., lost temper, verbal aggression, assaults, aggressive fantasies, and cruelty to animals). Each item was scored 0 (absent) or 1 (present). The total score can range from 0 to 5, with higher scores indicating more anger symptoms.

In this study, pervasive anger scores were calculated only if there was no more than one item with missing information ($N = 466$). The five items related to pervasive anger were summed with an average score of 1.01 ($SD = 1.34$) and had good internal consistency (ordinal $\alpha = .84$). For binary and ordinal response scales, ordinal α more accurately estimates reliability by using polychoric correlation matrix than Cronbach's alpha using Pearson correlation matrix (Gadermann, Guhn, & Zumbo, 2012; Zumbo, Gadermann, & Zeisser, 2007).

General Criminality Scale. This scale was developed for this study. General criminality was measured by eight items: three from the Static-99R (index non-sexual violence, prior non-sexual violence, and 4+ prior sentencing dates); and five from the MnSOST-R (sex offense under supervision, adolescent antisocial behavior, substantial drug and alcohol abuse, employment history, and discipline history while

incarcerated).

General criminality scores were calculated only if there was no more than one item with missing information ($N = 1,214$). The eight items related to general criminality were summed with an average score of 2.57 ($SD = 1.90$, range = 0 to 10, $N = 1,214$) and had good internal consistency (ordinal $\alpha = .74$, Gadermann et al., 2012; Zumbo et al., 2007).

Sexual Criminality Scale. This scale was developed for this study. Sexual criminality was measured by six items: five from the Static-99R (prior sex offenses, non-contact sex offenses, the unrelated victim, stranger victim, and male victim) and one from MnSOST-R (length of sexual offending history). Sexual criminality scores were calculated only if there was no more than one item with missing information ($N = 1,581$). The six items related to sexual criminality were summed with an average score of 2.06 ($SD = 1.58$, range = 0 to 8, $N = 1,581$) and had acceptable internal consistency (ordinal $\alpha = .70$, Gadermann et al., 2012; Zumbo et al., 2007).

Paraphilia Scale. This scale was developed for this study. Paraphilia was measured by eight binary items: diagnosis of pedophilia, paraphilia (not otherwise specified), and exhibitionism; any molestation of a child, using pornography in the offense, any exhibitionism/voyeurism offenses, non-contact sexual crime (Static-99R item), and any male victims (Static-99R item). Paraphilia scores were calculated only if there were no more one items with missing information ($N = 1,376$). The eight items related to paraphilia were summed with an average score of 1.40 ($SD = 1.07$, range = 0 to 6) and had acceptable internal consistency (ordinal $\alpha = .69$, Gadermann et al., 2012; Zumbo et al., 2007).

Sexual Recidivism. Sexual recidivism was defined as any subsequent conviction for a sexual offense (contact or non-contact) after release. Recidivism data

was accessed from the New Jersey State Police criminal records database through June 2009. These records include criminal records from the state of New Jersey as well as other states who share their records with the New Jersey State Police.

Procedure

Trained graduate assistants coded the data used in this study, included demographic characteristics (e.g., age and race/ethnicity), offense history (e.g., type and number of past sexual and non-sexual offenses), institutional behavior, victim characteristics (e.g., age and gender), and risk assessment scores (i.e., Static-99 and MnSOST-R).

Plan of Analysis

Comparing the risk-relevant characteristics. For assessing the relationship of characteristics between Blacks and Whites, the AUC analysis (Swets et al., 2000) was used. The AUC can vary between 0 and 1, with .50 indicating no difference in the characteristics between Blacks and Whites. AUCs above .50 indicate that Blacks have higher levels of the risk relevant characteristics compared to Whites. As a rough heuristic, an AUC of .56 corresponds to small effect size, while .64 reflects moderate effect, and .71 reflects a large effect size (Rice & Harris, 2005). In contrast, AUCs below .50 indicate that Whites have higher levels of the risk relevant characteristics compared to Blacks (.44 for a small effect, .33 for moderate effect, and .29 for larger effect). An AUC value is statistically significant if the 95% confidence interval does not include .50.

When a risk factor was a binary variable, odds ratio (instead of AUC) was calculated with 0.5 added to each cell to stabilize the variance (Fleiss & Berlin, 2009). An odds ratio is defined as $p/(1-p)$, where p is the raw proportion of the sample with

the characteristic. Odds ratios above 1 indicate that Blacks have higher levels of the risk relevant characteristics compared to Whites. For example, an odds ratio of 2 can be interpreted that the odds that Blacks have the risk-relevant characteristic were twice as high as the odds for Whites. No association is indicated when the 95% confidence interval of the odds ratio contains 1 (i.e., the odds are equal for both groups).

Predictive validity of Static-99R. Assessing the predictive accuracy of a risk scale requires the consideration of calibration (correspondence between expected and observed recidivism rates) as well as discrimination (how different are recidivists from non-recidivists?). For discrimination, we used two statistical methods: 1) the area under the curve (AUC) from receiver operating characteristic (ROC) analysis (Swets et al., 2000) and 2) odds ratios from logistic regression (Hosmer & Lemeshow, 2002). For calibration, we used two indices: 1) *E/O* index (the ratio of expected number of recidivists divided by an observed number of recidivists; Hanson, 2017) and 2) fixed-effect meta-analysis of logistic regression parameters (Borenstein et al., 2009; Hanson & Broom, 2005).

Area Under the Curve (AUC). AUC values also can be interpreted as the probability that a randomly selected recidivist would have a more deviant score than a randomly selected non-recidivist. AUC values are expected to be smaller in prognostic studies than in diagnostic studies because the outcome of interest in prognostic studies does not exist at the time of assessment, and may never happen (Royston et al., 2009; Helmus & Babchishin, 2017). The AUC has an advantage of insensitivity to base rates and robustness to outliers (Ruscio, 2008). An AUC value is statistically significant if the 95% confidence interval does not include .50.

Odds ratios. Odds ratios indicate the change in relative risk associated with

one unit change in Static-99/R scores. For example, Static-99R scores are associated with a consistent relative risk increase of approximately 1.45 (Hanson et al., 2016), which means the odds of recidivism increases 1.45 times as each Static-99R score increases. The primary advantage is that it is less influenced by a restriction of range compared to AUCs (Hanson, 2008).

E/O index. The *E/O* index is a measure of calibration in which the expected number of recidivists is divided by an observed number of recidivists (Hanson, 2017). Perfect calibration is indicated by an *E/O* index of 1.0. Following Rockhill et al. (2003), the 95% confidence intervals for the *E/O* indices were computed as follows:

$$95\%CI(E/O) = (E/O)\exp(\pm 1.96\sqrt{1/O})$$

The expected number of recidivists was based on the 5-year sexual recidivism rates for routine/complete samples reported by Hanson et al. (2016).

Comparing Logistic regression parameters. A second method of testing calibration was to examine the extent to which logistic regression parameters, such as intercept values (centered on Static-99R scores of 2) differed from the logistic regression parameters for the 5-year routine sample norms (Table 2.2: $B0_2 = -2.827$, $SE = 0.079$; $B1 = 0.368$, $SE = 0.0251$; Hanson et al., 2016). Specifically, the $B0_2$ represents the expected recidivism rate for a Static-99R score of 2 (p_2) in logit units ($\ln[p_2/\{1-p_2\}]$). Differences between the parameters in the current sample and those of the norms were tested using fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005).

Results

The Static-99R total scores for Blacks ($M = 3.3$, $SD = 2.2$, $n = 788$) were significantly higher than for Whites ($M = 2.5$, $SD = 2.5$, $n = 797$; $AUC = .60$, 95% CI [.57, .63]). Compared to the norms of Static-99R (Hanson, Lloyd, Helmus, &

Thornton, 2012), Blacks in this sample were underrepresented in Level I and II, (below average risk groups), whereas White sex offenders were proportional with the norms in Level I and II. On the other hand, in Level IVa and IVb (above average risk groups), both racial groups were overrepresented, but more so for Blacks (Figure 2.1).

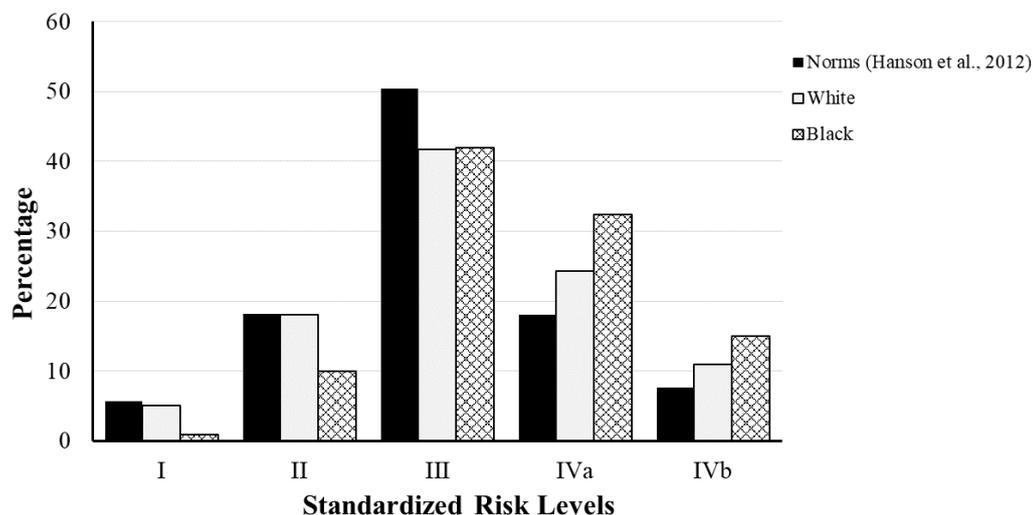


Figure 2.1. Distributions of White and Black racial groups across risk levels

Comparing Risk Relevant Characteristics

Blacks were, on average ($M = 37.5$, $SD = 10.3$, $n = 788$), significantly younger than Whites ($M = 41.8$, $SD = 13.1$, $n = 797$). Blacks displayed lower rates of marriage (odds ratio = 1.24), but were more likely than Whites to have children (odds ratio = 1.82; Table 2.2).

Compared to Whites, Blacks had a significantly lower socioeconomic status level (see Table 2.3). Specifically, Blacks were less likely to earn more than \$20,000 annually (odds ratio = 2.12), less likely to have been employed (odds ratio = 1.91) and less likely to have a high school degree (odds ratio = 1.69).

Table 2.2

General Demographics

	Black		White		Odds Ratio	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Non-U.S. born	6.5	51/787	4.8	38/788	1.36	[0.89, 2.09]	
Never married	45.6	359/788	40.3	321/797	1.24	[1.02, 1.51]	
Having children/stepchildren	71.6	561/783	58.2	460/791	1.82	[1.47, 2.24]	
Age	<i>M</i> = 37.5 (10.3), <i>n</i> = 788		<i>M</i> = 41.8 (13.1), <i>n</i> = 797				.591 [.562, .619]
Location							
New Jersey State prisons	74.2	585/788	62.2	496/797	1.74	[1.41, 2.16]	
Treatment facility	6.6	52/788	15.9	127/797	0.37	[0.27, 0.53]	
Treatment facility (sexually violent predators)	19.2	151/788	21.8	174/797	0.85	[0.67, 1.08]	

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.3

Socioeconomic Status (Income, Employment, Education)

	Black		White		Odds Ratio	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Socioeconomic status							
Lower	79.6	219/275	57.7	164/284			
Middle	16.0	44/275	35.2	100/284			.607 [.561, .654]
Upper	4.4	12/275	7.0	20/284			
Low income (< \$20,000/year)	76.0	225/296	59.9	182/304	2.12	[1.49, 3.01]	
Never employed prior to incarceration	36.3	279/769	22.9	178/776	1.91	[1.53, 2.39]	
Blue collar/Service vs. White collar	83.6	413/494	80.9	486/601	1.20	[0.88, 1.65]	
No high school degree	47.5	371/781	34.9	275/788	1.69	[1.38, 2.07]	

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.4

Early Childhood Adversity

	Black		White		Odds Ratio	95% C.I.
	%	<i>n/N</i>	%	<i>n/N</i>		
Not raised by both parents (up to age 13)	53.1	393/740	32.8	250/762	2.32	[1.88, 2.86]
Child abuse or neglect	28.4	204/719	37.9	285/751	0.65	[0.52, 0.81]
Child physical abuse	14.3	103/719	19.6	147/751	0.69	[0.52, 0.91]
Child sexual abuse	16.6	116/719	23.8	176/751	0.63	[0.49, 0.82]
Immediate family	20.2	20/99	27.7	41/148	0.66	[0.37, 1.22]
Extended family or acquaintance	72.7	72/99	62.8	93/148	1.57	[0.90, 2.71]
Strangers	7.1	7/99	9.5	14/148	0.75	[0.30, 1.89]
Penetration	44.2	42/95	38.1	43/113	1.29	[0.74, 2.24]
Child psychological abuse or neglect	7.2	52/719	7.6	57/751	0.95	[0.64, 1.40]

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Regarding early childhood adversity, Blacks were less likely than Whites to have been raised in a home with both parents up to age 13 (odds ratio = 2.32). Blacks were, however, less likely to have reported childhood physical and sexual abuse (odds ratios = 0.63 to 0.69; Table 2.4).

Blacks were assessed as having higher hostility levels than Whites. Specifically, Blacks were rated to have significantly higher pervasive anger symptoms ($M = 1.3$, $SD = 1.4$, $n = 243$) than Whites ($M = 0.8$, $SD = 1.2$, $n = 223$; $AUC = .58$). In addition, Blacks were noted as more likely to use force, threat, and violent behavior during the sexual offense compared to Whites (odds ratios of 1.80, 1.47, and 2.33, respectively). There were, however, no significant differences in sexual sadism disorder and impulsivity level between Blacks and Whites (Table 2.5).

Blacks showed fewer indicators of paraphilia than did Whites (paraphilia scale total score $AUC = .36$). Compared to Whites, Blacks were significantly less likely to be diagnosed as having pedophilia (odds ratio = 0.42). Similarly, Blacks were less likely to have minor and/or male victims compared to Whites (odds ratios = 0.58 and 0.39). Blacks were less likely to use pornography during the sexual offense than Whites (odds ratio = 0.16). In addition, Blacks were less likely to be involved in non-contact sex offenses (odds ratio = 0.38), specifically for exhibitionism and voyeurism offenses (odds ratio = 0.25; Table 2.6).

Blacks showed more general criminality than Whites ($AUC = .66$). The average age at first non-sexual offense for Blacks ($M = 19.9$, $SD = 5.9$) was significantly lower than for Whites ($M = 21.4$, $SD = 7.7$; $AUC = .56$). Blacks were also more likely to have a history of adolescent antisocial behavior, and more likely to have been charged or convicted as a juvenile. Compared to Whites, Blacks had a greater history of non-sexual violent offending (odds ratio = 2.46; Table 2.7).

Table 2.5

Hostility (Sadism, Anger, Impulsivity)

	Black		White		Odds Ratio	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Sexual sadism (diagnosis)	1.3	9/696	1.5	10/680	0.88	[0.36, 2.14]	
Force in sex offense	85.3	541/634	76.3	439/575	1.80	[1.34, 2.41]	
Pervasive Anger Scale (0 to 5)	<i>M</i> = 1.25 (1.45), <i>n</i> = 243		<i>M</i> = .79 (1.20), <i>n</i> = 223				.584 [.532, .635]
Impulsivity							
High	60.7	148/244	55.8	144/258			
Medium	28.7	70/244	30.2	78/258			.528 [.478, .579]
Low	10.7	26/244	14.0	36/258			
Physically violent (Index offense)	26.3	192/731	13.2	93/702	2.33	[1.77, 3.06]	
Threaten the victim (Index offense)	24.4	179/735	17.9	126/703	1.47	[1.14, 1.90]	

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.6

Paraphilia

	Black		White		Odds Ratio	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Paraphilia (Scale)	<i>M</i> = 1.11 (0.85)		<i>M</i> = 1.69 (1.19)				.362 [.333, .391]
Pedophilia indicators							
Pedophilia (diagnosis)	9.5	66/696	20.0	136/680	0.42	[0.31, 0.58]	
Job involves regular access to children	7.3	42/578	8.9	53/596	0.80	[0.53, 1.22]	
Abuse within the context of this role	57.1	24/42	60.4	32/53	0.88	[0.39, 1.98]	
Molestation of a minor child (Prior)	64.2	147/229	69.6	167/240	0.78	[0.53, 1.15]	
Molestation of a minor child (Index)	67.7	519/766	78.2	615/786	0.58	[0.47, 0.73]	
Pornography involved in the index offense	2.3	17/737	12.8	92/716	0.16	[0.10, 0.28]	
Any male victims (sexual offenses)	12.2	96/786	26.6	212/797	0.39	[0.30, 0.50]	
Other Paraphilias							
Paraphilia NOS (diagnosis)	7.6	53/696	10.1	69/680	0.73	[0.50, 1.06]	
Exhibitionism (diagnosis)	0.9	6/696	1.3	9/680	0.67	[0.24, 1.81]	
Exhibitionism/voyeurism offenses (Prior)	2.6	6/229	9.6	23/240	0.27	[0.11, 0.65]	
Exhibitionism/voyeurism offenses (Index)	1.2	9/799	4.5	35/786	0.25	[0.12, 0.52]	
Non-contact sex offenses	5.0	39/787	12.2	97/797	0.38	[0.26, 0.56]	

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.7

General Criminality

	Black		White		Odd Ratios	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
General criminality (Scale)	<i>M</i> = 3.06 (1.91)		<i>M</i> = 2.03 (1.74)				.657 [.627, .688]
Family members involved in the CJ system	22.7	93/409	12.5	62/497	2.06	[1.45, 2.93]	
Ever charged/convicted as a juvenile	38.3	263/686	26.7	173/649	1.71	[1.35, 2.16]	
Age at first non-sexual offense	<i>M</i> = 19.9 (5.91)		<i>M</i> = 21.44 (7.66)				.555 [.521, .589]
Index non-sexual violence	20.6	162/788	10.8	86/795	2.13	[1.60, 2.82]	
Prior non-sexual violence	45.3	357/788	25.1	200/796	2.46	[1.99, 3.05]	
Prior sentencing dates (4 or more)	44.2	347/785	27.5	219/797	2.09	[1.69, 2.58]	
Alcohol or drugs involved in index offense	37.9	247/651	37.9	269/709	1.00	[0.80, 1.25]	
Substance abuse (diagnosis)	8.9	62/696	12.2	83/680	0.70	[0.50, 0.99]	
APD (diagnosis)	13.1	91/696	9.9	67/680	1.37	[0.98, 1.92]	
Committed sex offenses under supervision	34.0	216/636	23.4	135/576	1.68	[1.30, 2.16]	
Discipline history while incarcerated	30.2	193/639	16.8	97/576	2.13	[1.62, 2.81]	
Adolescent antisocial behavior	25.7	164/637	18.3	105/575	1.55	[1.18, 2.04]	

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.8

Sexual Criminality

	Black		White		Odds Ratio	95% C.I.	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Sexual criminality (Scale)	<i>M</i> = 2.02 (1.49)		<i>M</i> = 2.10 (1.67)				.496 [.467, .524]
Ever charged with a sex crime as a juvenile	15.0	73/488	12.6	54/429	1.22	[0.84, 1.78]	
Age at first sex offense	<i>M</i> = 27.7 (10.16)		<i>M</i> = 30.5 (12.34)				.562 [.532, .593]
Any prior sex offense (> 1)	34.3	270/788	30.4	242/797	1.20	[0.97, 1.48]	
Any unrelated victims (sex)	80.7	634/786	74.3	592/797	1.44	[1.14, 1.83]	
Any stranger victims (sex)	25.0	197/787	20.3	161/793	1.31	[1.03, 1.66]	
Length of sexual offending history (> 1 year)	35.3	225/637	43.5	250/575	0.71	[0.56, 0.90]	
SVP evaluation	41.9	268/639	36.8	231/627	1.24	[0.99, 1.55]	
SVP recommended	16.4	31/189	22.4	28/125	0.68	[0.39, 1.20]	
Not completed sex offender treatment	82.5	527/639	70.6	405/574	1.96	[1.49, 2.57]	

Note. Not completed sex offender treatment (MnSOST-R item). Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 2.9

Mental Health

	Black		White		Odds Ratio	95% C.I.
	%	<i>n/N</i>	%	<i>n/N</i>		
History of psychiatric problems	29.6	228/771	45.8	357/779	0.50	[0.40, 0.61]
Serious medical conditions	14.7	110/748	20.9	157/751	0.65	[0.50, 0.85]
Psychosis (diagnosis)	2.0	14/696	1.6	11/680	1.24	[0.57, 2.70]
Anxiety disorder (diagnosis)	3.0	21/696	2.9	20/680	1.03	[0.55, 1.90]
Mood disorder (diagnosis)	5.8	40/696	8.2	56/680	0.68	[0.45, 1.04]
History of suicide attempts	9.8	75/767	16.5	127/770	0.55	[0.41, 0.75]
Neuropsychological deficits	2.2	16/715	4.2	31/744	0.53	[0.29, 0.98]
Intellectual impairment	11.2	80/712	11.5	87/755	0.97	[0.70, 1.34]

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

There was no significant difference in sexual criminality scale between Blacks and Whites. Blacks engaged in their first sexual offense at a younger age ($M = 27.7$, $SD = 10.2$, Range = 10 to 77) than Whites ($M = 30.5$, $SD = 12.3$, Range = 7 to 73), although the absolute difference was only three years. Blacks were more likely to have unrelated/stranger victims for sexual crimes whereas Whites were more likely to have related victims (e.g., family members; odds ratios = 1.44 and 1.31; Table 2.8).

Blacks were less likely than Whites to have a history of psychiatric problems, serious medical conditions, or suicide attempts (odds ratios of .50 to .65; Table 2.9).

Predictive Validity of Static-99R within a Fixed 5-Year Follow-Up Period

Of the initial 1,585 cases, 676 were eliminated because they had no follow-up information (including 325 civilly committed sex offenders, i.e., SVPs) and 336 were eliminated because they had less than 5 years between their release date and the date of follow-up (June 2009). Of the remaining 573 offenders, 291 were Black, and 282 were non-Hispanic White.

The average age of the Black follow-up group ($M = 37.4$, $SD = 10.3$) was lower than that of the White follow-up group ($M = 39.0$, $SD = 12.4$), but not statistically significant (AUC = .53, 95% CI [.48, .58]). As with the full sample, the Static-99R total scores of Blacks ($M = 2.7$, $SD = 2.1$, $n = 291$) were significantly higher than those of the Whites ($M = 2.2$, $SD = 2.3$, $n = 282$; AUC = .57, 95% CI [.52, .62]; Table 2.10).

Table 2.10

Five-year Sexual Recidivism Rates, Static-99R Scores, and AUC values for Whites and Blacks

Racial groups	Sexual Recidivism Rates (%)	Number of recidivists/total	<i>M (SD)</i>	AUC	95% C.I.	
					Lower	Upper
White	3.90	11/282	2.22 (2.33)	.760	.637	.884
Black	5.50	16/291	2.74 (2.11)	.782	.665	.899
Total	4.71	27/573	2.48 (2.23)	.772	.687	.856

Note. Based on a fixed 5-year follow-up period. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Discrimination. Using the fixed 5-year follow-up, the AUC with sexual recidivism for the full sample was .77 [.69, .86]. Static-99R was able to discriminate recidivists from non-recidivist for both Blacks and Whites (all AUCs $> .76$; Table 2.10). Contrary to expectations, the AUC value of Blacks (AUC of .78) was slightly higher than AUC value of Whites (AUC of .76; Table 2.10; Figure 2.2), although the difference was not significant.

Both major dimensions associated with sexual recidivism, general criminality, and sexual criminality, discriminated recidivists from non-recidivists equally well for both Blacks and Whites (all AUCs $> .65$; Table 2.11). Contrary to expectations, however, the Paraphilia Scale created for this study was unrelated to recidivism for both Blacks and Whites; the direction of the trends was for lower Paraphilia Scale scores to be associated with greater sexual recidivism risk (Table 2.11).

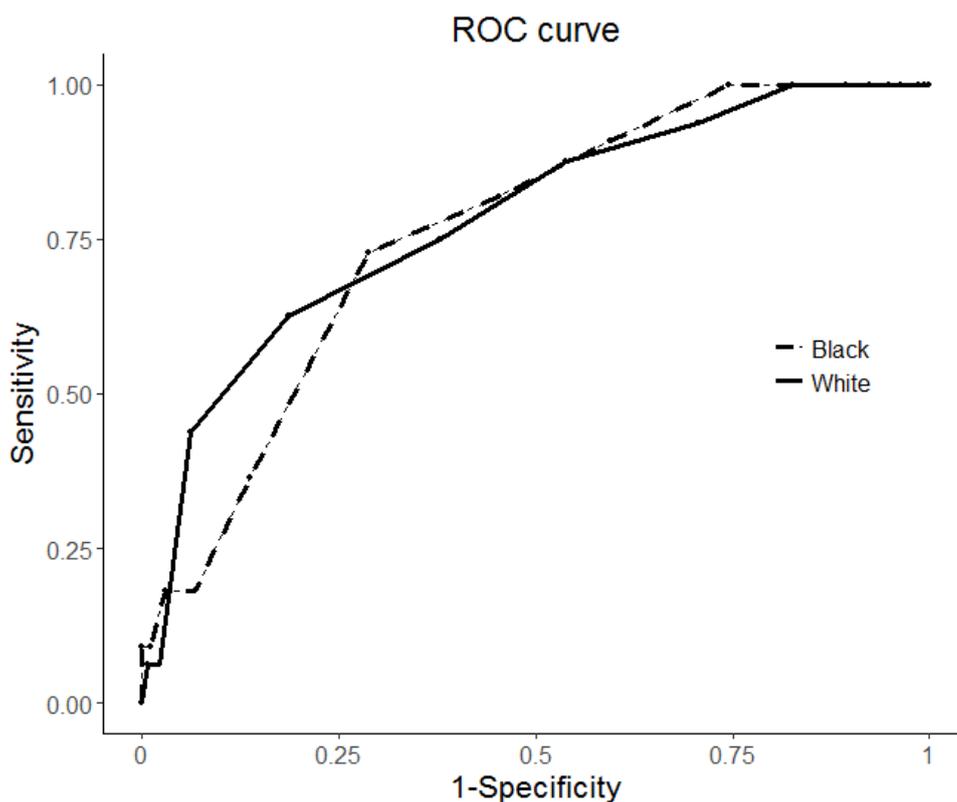


Figure 2.2. AUCs of Static-99R for White and Black racial groups

Table 2.11

Predictive Accuracy (AUC values) of General Criminality, Sexual Criminality, and Paraphilia Dimensions for Whites and Blacks for Sexual Recidivism

Propensities	Black				White			
	n	AUC	95% C.I.		n	AUC	95% C.I.	
			Lower	Upper			Lower	Upper
General Criminality	13/242	.658	.489	.828	10/210	.647	.489	.831
Sexual Criminality	16/291	.664	.516	.811	11/281	.682	.493	.870
Paraphilia	12/238	.334	.152	.517	10/221	.491	.296	.686

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

The 5-year sexual recidivism rates at a Static-99R score of 2 for the two racial groups were very similar (2.1% for Blacks and 2.3% for Whites; $Q_{\text{between}} = 0.02$, $df = 1$, $p = .881$; Table 2.12). The discrimination (change in relative risk) was higher for the Black study group (odds ratios = 1.77) than for the White study group (1.54), but, again, the differences between racial groups were not statistically significant ($Q_{\text{between}} = 0.435$, $df = 1$, $p = .509$; Table 2.12).

Calibration. For the full sample, the logistic equation indicated a relative risk increase of 1.65 for each increase in Static-99R score ($e^{.501} = 1.65$), and an adjusted 5-year sexual recidivism rate of 2.3% for a Static-99R score of 2 ($[1/\{1+e^{(-3.768)}\}] = .0226$). When compared to the norms (from Hanson, et al., 2016), the adjusted (score of 2) base rate was significantly lower ($B0_2$ of -3.77 vs. -2.83; $Q_{\text{between}} = 7.99$, $df = 1$, $p = .004$), and the discrimination was larger, but not significantly ($B1 = .501$ vs. .368; $Q_{\text{between}} = 1.52$, $df = 1$, $p = .218$; Table 2.13).

Table 2.12

Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism for Whites and Blacks

Racial group	Base rate (Static-99R score of 2 in logit units)		Relative risk			95% C.I.	
	$B0_2$	SE	$B1$	SE	OR	Lower	Upper
White	-3.747 (2.3%)	.446	.434	.144	1.54	1.16	2.04
Black	-3.845 (2.1%)	.479	.573	.155	1.77	1.31	2.41
Q ($df = 1$)	0.022, $p = .881$		0.435, $p = .509$				
I^2	.00		.00				
Average (fixed-effect)	-3.793 (2.2%)	.327	.498	.105	1.65	1.34	2.02

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Overall, the adjusted base rates ($B0_2$) of each racial group were significantly lower than the norms (5.6% versus 2.1% for Blacks and 2.3% for Whites). Relative risk rates for each racial group were greater than the norms, but there was no significant difference among those values (all p -values > .05; Table 2.13).

Table 2.13

Comparison of Logistic Regression Parameters for Static-99R Predicting 5-Year Sexual Recidivism with Meta-Average (from Hanson et al., 2016)

	Meta-Average (Norms)	Overall	White	Black
<u>Base rate</u>				
$B0_2$ (SD)	-2.83 (.079)	-3.77 (.323)	-3.75 (.446)	-3.85 (.479)
Q_{Δ} (df = 1)		7.99***	4.13*	4.39*
<u>Relative risk</u>				
$B1$ (SD)	.368 (.025)	.501 (.104)	.434 (.144)	.573 (.155)
Q_{Δ} (df = 1)		1.52	.199	1.69

Note. *** p < .001, ** p < .01, * p < .05.

The E/O index also indicated that the observed 5-year overall recidivism rate in this sample was significantly lower than the norms (4.7% versus 8.4%; E/O index = 1.79 [1.23, 2.61]; Table 2.14). When comparing each of the five Static-99R risk categories, only Level III (average risk; scores of 1, 2, and 3) showed significantly lower observed values than the expected values (2.8% versus 6.1%; E/O index = 2.15 [1.03, 4.51]; Table 2.14), although the direction was the same for all risk levels.

Table 2.14

*Comparison of Expected and Observed 5-year Sexual Recidivism Rates for Static-99R**Standardized Risk Levels*

Risk Level	Sample size	Recidivists		E/O index	95% C.I.	
		Observed (O)	Expected (E)		Lower	Upper
<u>Overall</u>						
I	18	0 (0.0%)	0.19	-	-	-
II	99	0 (0.0%)	2.46	-	-	-
III	254	7 (2.8%)	15.05	2.15	1.03	4.51
IVa	158	11 (7.0%)	19.81	1.80	0.99	3.25
IVb	44	9 (20.5%)	10.81	1.02	0.62	2.31
Total	573	27 (4.7%)	48.32	1.79	1.23	2.61
<u>Whites</u>						
I	14	0 (0.0%)	0.15	-	-	-
II	55	0 (0.0%)	1.39	-	-	-
III	127	3 (2.4%)	7.46	2.49	0.80	7.71
IVa	66	6 (9.1%)	8.14	1.36	0.61	3.02
IVb	20	2 (10.0%)	5.18	2.59	0.65	10.4
Total	282	11 (3.9%)	22.31	2.03	1.12	3.66
<u>Blacks</u>						
I	4	0 (0.0%)	0.05	-	-	-
II	44	0 (0.0%)	1.07	-	-	-
III	127	4 (3.1%)	7.59	1.90	0.71	5.06
IVa	92	5 (8.7%)	11.67	2.33	0.97	5.61
IVb	24	7 (29.2%)	5.63	0.80	0.38	1.69
Total	291	16 (5.5%)	26.01	1.63	0.99	2.65

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

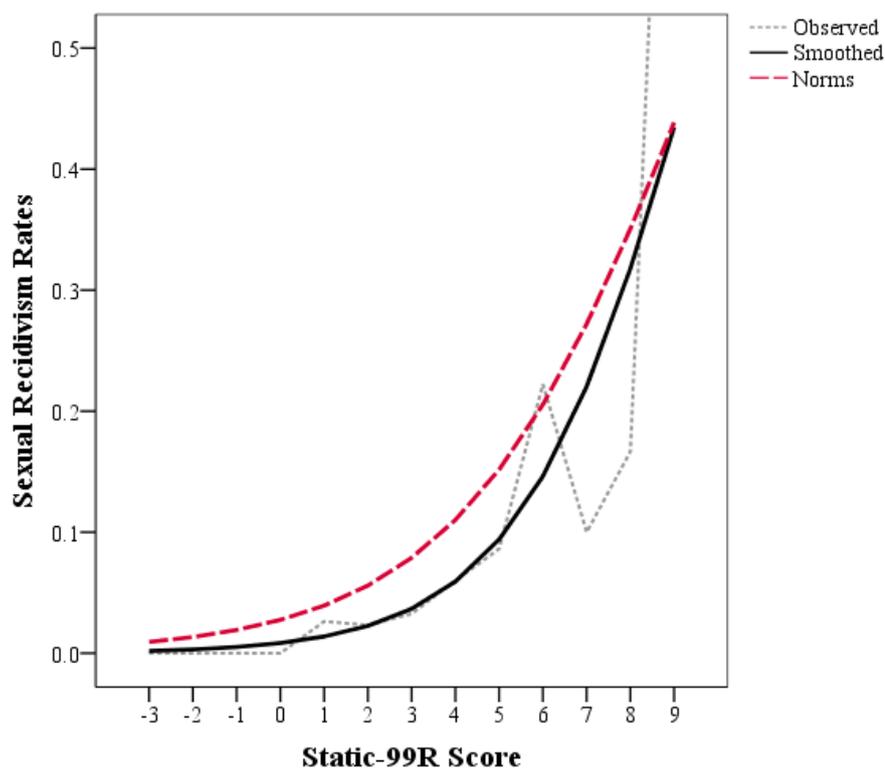


Figure 2.3. Observed and expected recidivism rates based on Static-99R 5-year sexual recidivism rates

Figure 2.3 provides a plot of the observed recidivism rates per Static-99R risk score, the rates based on the smoothed logistic curve fitted to this data, and the recidivism rate norms for routine samples (Hanson et al., 2016). As can be seen in Figure 2.3, the general pattern is that the sexual recidivism rates were lower than expected.

For White sexual offenders, the observed 5-year overall recidivism rate was lower than expected rate (3.9% versus 7.9%; E/O index = 2.03 [1.12, 3.66]; Table 2.14 and Figure 2.4). For Black sex offenders, the observed 5-year overall recidivism rate was also lower than the expected rate, but the difference was not significant (5.5% versus 8.9%; E/O index = 1.63 [0.99, 2.65]; Table 2.14 and Figure 2.4).

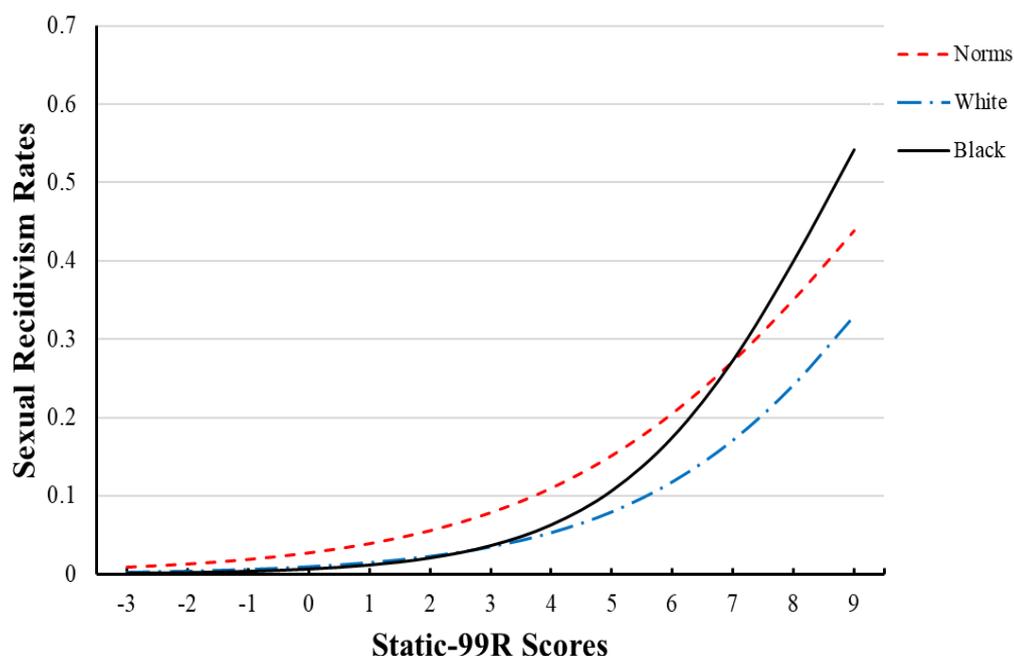


Figure 2.4. Logistic curves for White and Black racial groups with the norms

Although Blacks showed more general criminality than Whites, general criminality similarly predicted sexual recidivism for both Blacks and Whites (odds ratios = 1.30 and 1.35, respectively). There was no significant interaction between general criminality and race with the 5-year sexual recidivism (Wald $\chi^2 = .031$; $p = .860$). In addition, sexual criminality also discriminated recidivists from non-recidivists equally well for Blacks and Whites (odds ratios = 1.68 and 1.56). There was no significant interaction between sexual criminality and race with the 5-year sexual recidivism (Wald $\chi^2 = .093$; $p = .760$).

Discussion

The main purposes of the current study were to examine whether Whites who commit sexual crimes have different patterns of risk-relevant characteristics from Blacks who commit sexual crimes and if so, to test whether these differences influence the predictive validity of a commonly used actuarial risk assessment instrument for sex offenders: Static-99R. The study found that Blacks displayed

higher general criminality (i.e., anti-sociality), whereas Whites appeared more paraphilic (i.e., pedophilic interests). Despite the differences, however, the Static-99R predicted sexual recidivism equally well for both Blacks and Whites. Furthermore, the two major dimensions associated with sexual recidivism risk - sexual criminality and general criminality - both predicted sexual recidivism equally well for both racial groups. The Paraphilia Scale created for this study was, strangely, unrelated to sexual recidivism for either racial group; given the strong relationship between paraphilia and sexual recidivism in previous studies (Hanson & Morton-Bourgon, 2005), this finding is difficult to explain other than by unknown, idiosyncratic features of the current study.

As hypothesized, Blacks who commit sexual crimes had elevated levels of antisocial behavior compared to Whites who commit sexual crimes. For example, Blacks had a greater history of adolescent antisocial behavior and had more criminal charges, convictions, or sentences as a juvenile and as an adult compared to Whites. We also found that Blacks who commit sexual offenses had, on average, lower socioeconomic status, lower levels of education, and lower levels of employment than Whites who commit sexual offenses. These findings are consistent with a large body of empirical work demonstrating an association between social oppression/discrimination, disadvantaged and unstable family environments, and the likelihood of engaging in antisocial behavior and attitudes (Kaufman et al., 2008; Pérez et al., 2008; Simons et al., 2003).

Like previous studies, this study found that Whites showed more sexually deviant interests and behavior than Blacks. Whites were more likely to molest children and more likely to be diagnosed with pedophilia than Blacks. Also, Whites were more likely to commit non-contact sexual offenses (e.g., exhibition and

voyeurism), and more likely to use pornography in their offenses. Despite being a consistent research finding, there are no clear explanations as to why Whites should be more paraphilic than Blacks. It is even difficult to know whether the rates of paraphilia are unusually high among Whites, or unusually low among Blacks.

Their different positions in the racial and social hierarchy could result in differing motivations for Whites and Blacks to commit a sexual offense. Blacks, experiencing discrimination and social inequality as a subordinate group, might commit violent sexual crimes (e.g., rape, sexual assault) along with non-sexual violent crimes as an expression of their frustration with their social position or as a reenactment of power and domination by Whites. Experiencing discrimination and prejudice has been empirically supported to significantly increase the likelihood of engaging in delinquent and criminal behaviors among ethnic minority groups, mediated by anger and depression (e.g., Blacks and Hispanics; Pérez et al., 2008; Simons et al., 2003).

Although we do not know the extent of explicit discrimination experienced by individuals in this study, we can be certain that they were aware of, and influenced by, the widespread social disadvantage and racism against Blacks in the U.S. (Bales & Piquero, 2012; Hartney & Vuong, 2009; Proctor et al., 2016; U.S. Census Bureau, 2017b). In contrast, Whites who already have privileges (social and economic power) in the U.S. would not be acting from a position of systemic disadvantage. White male individuals might feel more sexual entitlement as a dominant class in a family (i.e., patriarchy) as well as a society. They might, thus, have a more distorted belief about sex with children and commit such behavior (e.g., sexual offending against children or incest).

In addition, consistent with the previous findings with men convicted of sexual

crimes (Cooper et al., 1996; Fix et al., 2017; Murphy et al., 2001), Whites who commit sexual crimes reported more experiences of sexual and physical abuse during childhood compared to Blacks. The high rates of childhood sexual and physical abuse among Whites contrasts with population data showing that the rates of child sexual and physical abuse are higher for Blacks males (16.5% and 25%, respectively) than Whites males (7.2% and 20%; Dakil, Cox, Lin, & Flores, 2011; Stoltenborgh, van IJzendoorn, Euser, & Bakermans-Kranenburg, 2011). These findings suggest that the sexually abused-sexual abuser hypothesis may be more relevant for Whites than Blacks. In addition, a prospective study found a history of childhood physical abuse increased the risk of committing a sexual crime as adults (Widom & Massey, 2015). This association, however, also may be more relevant for Whites than Blacks.

Previous research has suggested a link between child sexual abuse, and sexual offending against children or pedophilic interests (e.g., young and male victims; Fedoroff & Pinkus, 1996; Kaufman & Hilliker, 1996; Jespersen, Lalumière, & Seto, 2009; Levenson, Willis, & Prescott, 2016; Simons, Wurtele, & Durham, 2008). Nevertheless, most sexually abused victims do not become sexual offenders against children (Salter et al., 2003). Consequently, individual or cultural differences might play a role in the association between sexual abuse history and later sexual offending against children (pedophilic interests).

In addition, sexual crimes among Whites may be related to psychological dysfunctions. The current study found more evidence of dysfunctional coping among Whites than Blacks. It is possible that some of this dysfunctional coping are related to the risk of sexual offending. For example, privileged Whites might be more likely to use sex as a coping method when they are in emotional distress (use of prostitution, pornography, or masturbation), which, in turn, increase the risk of engagement in

deviant sexual behaviors (Långström & Hanson, 2006). Population surveys of U.S. sexual practices, however, find little difference in the overall rate of conventional sexual behaviors (masturbation, number of sexual partners) between Whites and Blacks (Laumann, Gagnon, Michael, & Michaels, 1994). We are unaware of any U.S. population surveys on the rates of paraphilic sexual interests and behaviors.

Although Blacks showed less sexually deviant interests and behavior, they are, nonetheless, overrepresented among individuals convicted of sexual offenses (Carson, 2018; U.S. Census Bureau, 2017a). The reasons for this overrepresentation are not fully understood. Racial prejudice in the criminal justice system is well documented and would be one contributing factor. Blacks show disproportionately high rates of arrest compared to Whites across all types of crimes (e.g., murder, aggravated assault, robbery, or burglary; U.S. Census Bureau, 2017a; U.S. Department of Justice, 2015). However, it is also possible that there are real, behavioral differences in antisociality between Whites and Blacks.

It is also possible that there are some distinct, culturally specific risk factors for sexual offending that were not measured by the current study. For example, it is possible that some negative elements of their family and school experiences (other than sexual abuse) might inhibit the development of appropriate sexual self-regulation strategies and result in increased rates of sexual offending behavior among Blacks. Consequently, future research should explore cultural differences in sexual socialization across racial and ethnic groups. In particular, it would be important to consider the extent to which positive sexual socialization practices are linked to family income, education, and social disadvantage.

Another purpose of this study was to examine how well Static-99R assessed sexual recidivism risk for Blacks compared to Whites. Consistent with previous

studies, Blacks had a higher Static-99R total score than Whites. The discrimination of Static-99R for both racial groups was good (all AUCs $>.76$; odds ratios >1.54) given the average values of relative accuracy (AUC = $.70$ and odds ratio = 1.45 ; Hanson et al., 2016; Helmus, Hanson, et al., 2012). Contrary to our hypothesis, however, the discrimination of Static-99R for Blacks was slightly better (not worse) than those for Whites.

When compared to the norms for a 5-year routine sample, the observed 5-year sexual recidivism rate (4.7% after 5 years) was significantly lower than the norms (8.34%; $E/O = 1.79$). Both racial groups showed lower than expected recidivism rates, and there were no differences in the calibration of Static-99R for Blacks compared to Whites. The reasons for the low recidivism rates are not fully known. The low sexual recidivism rates of this specific sample may be related to the research method used (e.g., the accuracy of records), sample selection (excluding high-risk sex offender groups; e.g., SVP), or other factors that are not yet fully understood.

Despite the significantly higher score of Static-99R for Blacks than Whites and the different patterns on risk-relevant characteristics between Blacks and Whites, there was no evidence that the risk factors/propensities measured by Static-99R had any differential associations with sexual recidivism for Blacks compared to Whites.

Limitations

The sample used in this study was limited to individuals who were serving a sentence for sexual crimes in prison or a forensic treatment center in the State of New Jersey. The findings of this study, thus, may not generalize to Whites and Blacks convicted of sexual crimes who are being managed in the community, or in the different jurisdictions. Specifically, a higher level of paraphilic interests among Whites in this study (i.e., a higher risk group) may not be generalized to Whites who

were sentenced to community supervision in lieu of prison time (i.e., a lower risk group).

The meaning of racial and ethnic identity is fluid and is expected to have different features in different times and places. Although there are certain features commonly found among Blacks in the U.S., Black culture is not homogeneous, including multiple ethnic groups with distinct cultures (e.g., Jamaicans, West Indians, Somalis, and Nigerians). In addition, given the high rates of undetected or unreported sexual crimes (e.g., rape or sexual assault; Langton, Berzofsky, Krebs, & Smiley-McDonald, 2012), the current sample was unlikely to be representative of all those who engage in sexual offending.

The information regarding childhood abuse in this study was collected via self-report. Given the cultural values or norms of Black society (e.g., emphasizing on family and community relationships; Boyd-Franklin & Lockwood, 2009; Grills & Longshore, 1996; Mattis, & Grayman-Simpson, 2013), relatively low rates of childhood abuse among those who convicted of sexual offences in this study might be related to the underreporting of sexual or violent abuse, especially in family and community members.

Although the overall sample was large ($N = 1,585$), the sample size was significantly reduced when examining the predictive validity of Static-99R ($n = 573$). In sub-analyses, there were 291 Blacks (16 recidivists) and 282 non-Hispanic Whites (11 recidivists). Additional research with larger numbers of Blacks is recommended in order to increase confidence in any conclusions regarding the potential of a cultural test bias of Static-99R for Blacks.

Implications for Research

Further research is needed to explain the different patterns of general criminal

behavior and paraphilia between Blacks and Whites who commit sexual crimes. In particular, are there different patterns of sexual self-regulation problems (e.g., sexual preoccupation and using sex as coping) among Blacks and Whites who commit sexual crimes? In addition, researchers should empirically explore potentially race and culture-specific risk or protective factors for Blacks that influence the likelihood, or continuation of, sexual offending.

Another avenue of research could investigate whether the latent constructs (sexual and general criminality) measured by risk assessment instruments are consistent across different racial groups - particularly minorities who are overrepresented in the criminal justice system (i.e., cultural test bias in construct validity; Brouillette-Alarie et al., 2016). Further, the extent to which the two major propensities (general and sexual criminality) are associated with sexual recidivism for Blacks is worthy of further examination.

Implications for Practice

The current findings support the use of Static-99R in risk assessment for both Black and White individuals to identify those at high risk to reoffend sexually (Risk-principle: Andrews & Bonta, 2010).

Given that Blacks present as more antisocial, whereas Whites present as more paraphilic, it might be desirable to tailor treatment to these differences. The treatment for Blacks who commit sexual crimes may benefit from an increased focus on antisocial attitudes and behaviors (e.g., anger management, cognitive restructuring, vocational training). In contrast, sex crime specific treatment may more likely be prioritized for Whites. Such treatment would address the usual targets of managing deviant sexual interest and developing sexual self-regulation. Treatment providers should also expect that they would have an increased need to address general mental

health concerns for Whites compared to Blacks.

It is important to remember, however, that both general criminality and sexual criminality were present in both racial groups, and both dimensions were related to the risk of sexual recidivism. Consequently, differences in treatment should be a matter of emphasis rather than creating separate programs that address fundamentally different criminogenic needs.

In general, treatment programming will likely be most effective when it takes into consideration the cultural values or norms of Blacks (e.g., maintaining family relationships, community, racial unity, spirituality and religion; Boyd-Franklin & Lockwood, 2009; Grills & Longshore, 1996; Mattis & Grayman-Simpson, 2013) as well as socio-demographic characteristics that influence their response to treatment (e.g., low education and socioeconomic status, systemic oppression, distrust of criminal justice system; Responsivity-principle).

Prevention programs for sexual crimes may also benefit from being tailored to racial and racial differences. For example, sexual violence prevention programs for Blacks may be most effective when they focus on education and vocational training as well as antisocial attitudes and delinquency in adolescence whereas for White individuals, prevention programs may benefit from reducing child sexual abuse and providing treatment for childhood sexual abuse victims. As well, the effectiveness for programs for White individuals may be increased by addressing distorted beliefs about privilege in general, and sexual privilege and entitlement in particular.

Chapter 5: Study 3

Predictive Accuracy of the Static-99R and Static-2002R Risk Tools for Identifying Indigenous and White Individuals at High Risk for Sexual Recidivism in Canada

The purpose of the Canadian correctional system is to administer sentences, promote public safety, and assist in the rehabilitation of offenders (*Corrections Act*, 1988; *Corrections and Conditional Release Act*, 1992/2012; *Ministry of Correctional Services Act*, 1990). Forensic risk assessment has become a ubiquitous practice in achieving these goals, from determining whether offenders should be incarcerated indefinitely (e.g., Dangerous Offender hearings; ss. 752 and 753 of the Criminal Code of Canada) to deciding who should be released on parole. According to the risk principle of the widely accepted Risk, Need, and Responsivity model of offender rehabilitation (R-N-R), treatment programs are most likely to be effective when the most intensive services are provided to the highest risk offenders (Andrews et al., 2006; Andrews & Bonta, 2010; Gendreau, Smith, & French, 2006). Consequently, accurate assessments of recidivism risk are a prerequisite for effective offender rehabilitation.

An ongoing concern with the assessment of recidivism risk is the extent to which measures, developed and validated using predominantly White samples, can be applied to other ethnic and racialized minority groups (i.e., cultural bias). In Canada, the issue of cultural bias has mostly concerned individuals of Indigenous heritage, who are overrepresented at all stages of the criminal justice system (Haag et al., 2016). This issue has received renewed attention with the *Ewert v. Canada* (2015, 2018) case where the Supreme Court of Canada ruled that several well known risk assessment scales (e.g., Static-99; Hanson & Thornton, 2000; Psychopathy Checklist-

Revised [PCL-R], Hare, 2003) had insufficient evidence on cross-cultural validity to justify their use with Indigenous peoples with a history of crime (for an overview of the Ewert case, see the 2016 special issue of the *Journal of Threat Assessment and Management* [Vol. 3, No. 2]).

Given this context, the purpose of this study was twofold: 1) examine the potential for cultural bias in the predictive validity of two widely adopted scales for individuals with a history of sexual crimes (i.e., Static-99R [Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012] and Static-2002R [Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012]), one of which was specifically named in the Ewert case; and 2) assess the presence and predictive validity of the psychological constructs that underlie these scales. This study represents the largest and most up-to-date evaluation of both the Static-99R and Static-2002R with individuals of Indigenous heritage. It is also the only study of these measures to examine both discrimination (i.e., relative risk) and calibration (i.e., absolute risk) as indicators of predictive validity.

Indigenous Overrepresentation in Canadian Corrections

Despite accounting for approximately 4% of the total Canadian adult population (Statistics Canada, 2015), Indigenous peoples (First Nations, Métis, and Inuit) account for more than 20% of the total federal inmate population (Public Safety Canada, 2015) and 23% (444/ 1,898) of individuals serving a federal sentence for a sexual offence (MacDonald, 2014). Furthermore, Indigenous peoples in the criminal justice system are more likely to be placed in higher security prisons (Public Safety Canada, 2015), less likely to be granted statutory release (i.e., released after 2/3 of their sentence; Public Safety Canada, 2015), and more likely to have their parole revoked (Office of the Correctional Investigator, 2015).

There are several proposed reasons for the overrepresentation of Indigenous peoples in Canadian corrections. First, research has consistently found that Indigenous peoples are more likely to be arrested and convicted than non-Indigenous peoples (e.g., Bonta, Laprairie, Wallace-Capretta, 1997; Quann & Trevethan, 2000) and have significantly higher recidivism rates (Gutierrez, Wilson, Rugge, & Bonta, 2013; Rojas & Gretton, 2007; Sioui & Thibault, 2002). An important finding for the current study is that Indigenous individuals in the criminal justice system score higher on the majority of established risk factors for general and violent recidivism than non-Indigenous individuals (i.e., Central Eight risk/needs factors; Andrews & Bonta, 2010). For example, Indigenous individuals in the criminal justice system have lengthier criminal histories (e.g., Holsinger, Lowenkamp, & Latessa, 2006; Perley-Robertson, Helmus, & Forth, 2018; Shepherd, Adams, McEntyre, & Walker, 2014), higher rates of substance abuse (e.g., Ellerby & MacPherson, 2002; Shepherd et al., 2014), more problems with family/marital relationships (e.g., Shepherd et al., 2014; Trevethan, Moore, & Rastin, 2002), and less education/employment (Holsinger, Lowenkamp, & Latessa, 2003; Shepherd et al., 2014; Trevethan et al., 2002) than non-Indigenous individuals in the criminal justice system.

The presence of these risk factors is not surprising when we consider that, in comparison to non-Indigenous peoples, Indigenous peoples have lower median incomes (\$25,526 vs. \$34,604), lower employment rates (52% vs. 60.5%), lower education levels (62% with a high school diploma vs. 80%; Statistics Canada, 2018), and poorer health outcomes (e.g., mental health issues and short life expectancy; Statistics Canada, 2005; Statistics Canada, 2015). It is also well documented that Indigenous peoples report higher rates of childhood maltreatment (neglect and sexual/physical abuse) and intimate partner violence (Boyce, 2016; Scrim, 2010;

Sinha et al., 2011). They are also more likely than non-Indigenous peoples to figure as victims in official crime statistics (Boyce, 2016; Scrim, 2010; Sinha et al., 2011).

These indicators of adversity and social disadvantage need to be understood in the context of Canada's history of colonization, and the devastating effects of racist social policies towards Indigenous peoples (Truth and Reconciliation Commission of Canada, 2015). As the Royal Commission on Aboriginal Peoples (RCAP) makes clear, many of the most pressing problems in Indigenous communities are predictable consequences of the efforts of the dominant social class to eliminate Indigenous cultures (i.e., residential schools, the *Indian Act*, and enfranchisement, wherein the Canadian government would strip Status Indians of their status if they obtained a university degree, served in the armed forces, or if a Status Indian woman married a non-Indigenous or non-status person). Whether it is poverty, substance abuse, low levels of formal education, or alienation and isolation, the criminogenic factors that contribute to higher rates of crime among Indigenous peoples in Canada are rooted in some 500 years of Indigenous-settler relations. To quote the RCAP (1996):

People who endure these disruptions may feel adrift - disoriented and unsure of how to get along in the sometimes hostile non-Aboriginal world. If their aboriginality has been devalued or ridiculed, they may have lost pride and self-esteem and be unable to build these qualities in their children. If they have been damaged in heart and soul, they may turn to alcohol, violence, crime or other forms of anti-social behavior.

The creation of the reserve system has further marginalized Indigenous communities and fostered environments conducive to crime for the following reasons: 1) the reserve system creates large groups of disadvantaged people with limited resources; 2) these communities tend to exist separate from mainstream society

creating the potential for culture clash (e.g., differing views of crime and justice); and 3) individuals in this environment are more likely to experience early and sustained abuse (Laprairie, 1996). Although such conditions do not describe all Indigenous communities, there was no question that the reserve system placed non-Indigenous peoples in control of most of Canada's natural resources – natural resources for which Indigenous people continue to be dependent upon.

Another factor that may lead to Indigenous overrepresentation is the effect of systemic discrimination and bias in how Indigenous individuals are treated at all levels of the criminal justice system (*R v. Gladue*; Rudin, 2009; Truth and Reconciliation Commission of Canada, 2015). For example, there is evidence for the differential treatment of Indigenous peoples in policing (i.e., higher likelihood of arrest), sentencing (i.e., longer sentences), and rehabilitation (i.e., reduced access to culturally appropriate programming; LaPrairie, 2002; Mann, 2009; Rudin, 2009). Underlying this discrimination is the false assumption that Indigenous individuals, just because of their race, are more likely to commit a crime (i.e., higher risk) than non-Indigenous individuals.

Risk Assessment and Cultural Bias

Considerable research has demonstrated that structured approaches to risk assessment, particularly those based on statistical methods (i.e., actuarial), provide a more accurate estimate for the likelihood of recidivism than unstructured judgments (Ægisdóttir et al., 2006; Grove et al., 2000; Hanson & Morton-Bourgon, 2009). Risk assessment scales are criterion-referenced measures because the main goal is to assess the likelihood of a specific outcome in the future. Unlike norm-referenced scales, where the goal is the assessment of a specific construct, predictive validity becomes the most important metric for evaluating the efficacy of criterion-referenced measures

(Gutierrez, Helmus, & Hanson, 2016). A comprehensive understanding of predictive validity considers both discrimination (how likely offenders with a higher score are to reoffend than offenders with a lower score) and calibration (how well the estimated recidivism probability from the instrument's norms corresponds with the observed recidivism probability of a new sample).

The Canadian Code of Ethics for Psychologists (Canadian Psychological Association, 2017) states that psychologists need to ensure that assessment methods are appropriate to the particular cultural and social contexts of the individuals being assessed (p. 20). Nevertheless, relatively little research has examined the cultural bias in structured risk assessment instruments. Cultural bias in risk assessment occurs when either discrimination or calibration is systematically different across ethnic groups (Reynolds & Suzuki, 2013). There are several possible causes for the cultural bias in the performance of risk assessment instruments. When looking at cultural bias within the instrument itself, we can refer to both item bias and construct bias (Geisinger & McCormick, 2013; see Haag et al. [2016] for a more thorough review). Item cultural bias refers to differential responses based solely on the cultural interpretation of the content of the item itself (e.g., the term "unrelated victim" could mean something different to someone from a more collectivistic culture). Construct cultural bias occurs when the constructs measured by the scale do not equally apply to individuals with different cultural backgrounds (Geisinger & McCormick, 2013). This could be an indication that the items meant to measure a specific construct are different for different ethnic groups, or that constructs may only partially apply to members of different groups (Byrne et al., 2009).

Until recently, relatively little research attention had been devoted to the validity of structured risk assessment instruments for individuals of Indigenous

heritage. This, of course, changed with the Ewert case. Mr. Ewert, a Métis, serving two concurrent life sentences for sexually motivated homicide and assault, challenged the use of psychological and actuarial risk assessment instruments (e.g., Static-99) by the Correctional Service of Canada (CSC). The claim was that Mr. Ewert's rights under s. 24(1) of the *Corrections and Conditional Release Act* (i.e., the use of accurate and up to date information) and ss. 7 (i.e., right to life, liberty, and security of the person) and 15 (i.e., equality before and under the law) of the *Canadian Charter of Rights and Freedoms* were violated given that the instruments were not adequately validated for use with Indigenous peoples (*Ewert v. Canada, 2015*).

The original complaint resulted in a victory for Mr. Ewert; however, this decision was overturned on appeal. The case, therefore, proceeded to the *Supreme Court of Canada* (SCC) where it was determined that, although Mr. Ewert's Charter rights were *not* infringed upon, the court made a declaration that CSC had failed in its duties under s. 24(1). More specifically:

The CSC had long been aware of concerns regarding these tools exhibiting cultural bias yet took no action to confirm their validity and continued to use them in respect of Indigenous offenders, despite the fact that research would have been feasible. In doing so, the CSC did not meet the legislated standard set out in s. 24(1) (*Ewert v. Canada, 2018*).

In addition to the Ewert case, longstanding concerns about Indigenous over-representation has motivated repeated calls for more and better research (e.g., Mann, 2009; Office of the Auditor General of Canada, 2016).

Risk Assessment with Indigenous Individuals with a History of Sexual Crimes

Structured risk assessment instruments consist of empirically identified risk factors associated with the recidivism outcomes (e.g., violent or sexual recidivism).

There are two broad categories of risk factors: *static* (unchangeable) and *dynamic* (changeable) risk factors. In the criminal justice system, risk assessment instruments are widely used to estimate the likelihood of recidivism risk (i.e., prediction; *static* and/or *dynamic* risk factors) and to guide the effective intervention strategies (i.e., risk reduction; *dynamic* risk factors; Heilbrun, 1997; Neal & Grisso, 2014).

Within the last few decades, important risk factors associated with sexual recidivism have been identified based on samples of predominantly White people. These risk factors can be organized into two broad categories: indicators of general antisociality (e.g., impulsiveness, supervision failures, criminal history) and sex crime specific indicators (e.g., sexual preoccupation, emotional congruence with children; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann et al., 2010; Whitaker et al., 2008). When examining the risk profile of Indigenous individuals with a history of sexual crime, they tend to score significantly higher on general antisociality while scoring lower on sexual deviancy compared to non-Indigenous individuals (Babchishin, Blais, et al., 2012; Ellerby & MacPherson, 2002; Helmus, Babchishin, Blais, 2012). They also tend to be younger and have higher sexual, violent, and general recidivism rates (e.g., Babchishin, Blais et al., 2012; Rastin & Johnson, 2002; Rojas & Gretton, 2007).

Few studies have examined the predictive validity of risk assessment scales with Indigenous peoples with a history of sexual offending. Olver and his colleagues (2018) demonstrated that the Violence Risk Scale – Sex Offender version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003) significantly predicted sexual recidivism for both Indigenous and non-Indigenous individuals with a history of sexual crimes. They further demonstrated that treatment change was significantly related to reductions in risk for both groups. In contrast, Helmus, Babchishin, et al.

(2012) found that the STABLE-2007, another dynamic scale, did not predict sexual recidivism for Indigenous individuals with a history of sexual crimes. Their results were mostly attributed to the poor performance of the general criminality items; sexual deviancy items performed equally well for Indigenous and non-Indigenous individuals.

Static-99R and Static-2002R are the most commonly used actuarial risk tools for adult males with a history of sexual crimes in Canada (Bourgon et al., 2018) and the U.S. (Kelley et al., 2018). Static-99R contains 10 static (i.e., unchangeable) items (e.g., prior sex offences) and the total score (-3 to 12) is used to estimate the likelihood of sexual offending. Static-2002R was developed to improve the conceptual coherence of the Static-99R scale and, thus, to measure psychological constructs (e.g., deviant sexual interests, general criminality). Static-2002R also contains only static items (8 of 14 items from Static-99R), and the total score (- 2 to 13) is used to estimate the likelihood of sexual offending. Also, the Static-2002R items are categorized into five conceptually meaningful constructs (see Hanson & Thornton, 2003).

A recent study found that considering the items from both Static-99R and Static-2002R improves the assessment of the psychologically meaningful constructs associated with sexual recidivism (Brouillette-Alarie et al., 2016). The three major propensities are the following: 1) persistence/paraphilia (e.g., rate of sexual offending, male victim); 2) youthful stranger aggression (e.g., age at release, non-sexual violence at index, unrelated victim); and 3) general criminality (e.g., breach of conditional release; prior nonsexual violence; Brouillette-Alarie et al., 2016). Furthermore, Static-99R and Static-2002R recently adopted the standardized risk category system (i.e., Level I, II, III, IVa, and IVb; Hanson et al., 2017). Each category demonstrates

quantitative information (e.g., likelihood of sexual reoffending, percentile ranks) as well as psychological characteristics associated with each level (e.g., sexual deviance and general antisociality). Consequently, the standardized risk levels assist evaluators to arrive at the same inferences for individuals assigned to the same risk category, regardless of the risk tool used.

There is only one study to examine the predictive validity (discrimination) of Static-99R and Static-2002R for Indigenous peoples in Canada. Babchishin and colleagues (2012) examined the predictive validity of both scales among 1,588 individuals with a history of sexual crime (Indigenous, $n = 319$; non-Indigenous, $n = 1,269$). Overall, there was moderate discrimination between recidivists and non-recidivists for both groups for the Static-99/R (sexual recidivism; AUCs of .71 for Indigenous individuals and .74 for non-Indigenous individuals). Static-2002R, however, was significantly less predictive of sexual recidivism for Indigenous individuals compared to non-Indigenous individuals. Total Static-2002R scores did not significantly predict sexual recidivism for Indigenous individuals with a history of sexual crime (AUC of .61 vs. AUC of .76 for non-Indigenous individuals). This study did not evaluate the calibration of either Static-99/R or Static-2002R.

Current Study

The Ewert decision (2018) highlighted the lack of research examining cultural bias in several widely adopted risk assessment scales in Canadian corrections - including the Static-99R. The purpose of this study was to provide an updated examination of the predictive validity of both the Static-99R and Static-2002R for Indigenous peoples and Whites with a history of sexual crimes. The development and validation samples of Static-99R and Static-2002R comprised predominantly Whites. Whites - a dominant and privileged group - might have distinctive features associated

with criminal behaviour, features that differ from the factors motivating criminal behaviour among Indigenous peoples.

There are other ethnic groups in Canada with potentially distinctive features (e.g., South Asians, East Asians, Caribbean Blacks). Of these groups, this study focused on Indigenous peoples because of their overrepresentation at all stages of the criminal justice system, and the concern that risk assessment practices could be contributing to this problem. Given legal and public policy concerns, this study also focused on the ability of these risk tools to identify individuals at high risk for sexual recidivism. Finally, this is the first study to assess both discrimination and calibration when evaluating the predictive accuracy of these specific tools.

Based on existing research, we expected Indigenous individuals to score higher on the total scores for both the Static-99R and Static-2002R, largely due to elevated scores on the items indicative of general antisociality. We also expected Indigenous peoples to score lower than Whites on items related to sexual deviancy. Regarding predictive validity, we expected Static-99R to show moderate discrimination for Indigenous individuals, but that discrimination would be lower than that observed for Whites. Given the similarity between Static-99R and Static-2002R, we expected similar results for both risk tools. Although previous studies with Indigenous peoples found that Static-2002R was less accurate than Static-99R, this finding lacks theoretical explanation and may be a sampling anomaly that could disappear with future research. No hypotheses regarding the calibration analyses were made for either instrument because no previous studies examined calibration of either tool for Indigenous peoples.

The second goal of this study was to examine the presence and predictive validity of risk-relevant characteristics for Indigenous and White peoples with a

history of sexual crimes. Although no study to date has applied these specific constructs to Indigenous peoples, we would again expect that Indigenous individuals with a history of sexual crime to score higher on indicators of general criminality and lower on indicators of sexual deviancy than White individuals with a history of sexual crime.

Method

Sample

The current study included an Indigenous study group ($n = 653$) and a White study group ($n = 1,560$) from five independent Canadian samples (Bonta & Yessine, 2005; Brankley et al., 2017a; Haag, 2005; Lee et al., 2018; Olver et al., 2018). The Indigenous study group represented a mixed group of First Nations, Métis, and Inuit individuals. First Nations and Inuit peoples were the original inhabitants of the land that is now Canada. Métis is a group of people with mixed European and First Nations ancestry. In the general population, First Nations are the largest subgroup (60%) of the Indigenous peoples in Canada, followed by Métis (35%) and Inuit (4%). The White study group referred to individuals of European descent.

All samples were individuals who had been convicted of sexual crimes and were classified by an administrative process as having a high level of risk and/or need (e.g., preselected [referred to as National Flagging System], detained until the end of a sentence, or attended high-intensity treatment programs). Although the samples were collected across all Canadian regions, most of the Indigenous study group were from the Prairies (see Appendix A). The current samples partially overlap with the samples from Babchishin, Blais, et al. (2012; i.e., Bonta & Yessine [2005] and Haag [2005]).

Bonta & Yessine (2005). The original sample included three subgroups: (1) individuals flagged as potential Dangerous Offenders (considered for indeterminate

sentences) by the National Flagging System, (2) individuals designated as Dangerous Offenders, and (3) individuals who committed a violent reoffence after being detained until their Warrant Expiry Date (WED). Only the first group (referred to as National Flagging System) was included in this study because Static-99 scores were only available for this group. These individuals were released into the community between 1992 and 2004 and followed until April 2005. Sexual recidivism was defined as a conviction for a sexual offence (excluding prostitution offences, indecent phone calls, and possession of child pornography) after the release into the community. The sexual recidivism information was obtained from the Canadian Police Information Centre (CPIC).

Haag (2005). The original study sample included all male Canadian federal inmates serving a sentence for a sex offence whose warrant expiry date (WED) was in 1995. For the current study, only individuals detained until WED and released into the community (about 25% of the total sample) were included. Sexual recidivism was defined as a conviction for a further sexual offence within seven years after release, and the sexual recidivism information was obtained from the CPIC.

Olver et al. (2018). This study included adult males who served federal sentences and attended sex crime specific treatment programs (e.g., Clearwater programs, National Sex Offender Program) between 1983 and 2008. Sexual recidivism was defined as any criminal code violation for a sexually motivated offence (including child pornography offences). The recidivism information was obtained from the CPIC.

Brankley et al. (2017a). This study was an extended follow-up to the Hanson and Harris (1998/2000) study. Adult males who had served part of their sentences in the community (e.g., probation, parole) were selected from all Canadian federal and

provincial correctional systems (except Prince Edward Island). All individuals had been convicted of a sexual offence involving physical contact with the victim(s). Sexual recidivism was defined as any subsequent offences that were considered sexually motivated (including offence, charge, or conviction) and their recidivism information was obtained from official (CPIC) and unofficial sources (news articles).

Lee et al. (2018). The sample used in this study comprised individuals who had a history of sexual crimes taken from the Blais and Bonta (2015) study. Updated recidivism information was obtained resulting in a longer follow-up period than was available in the original study. Participants were adult males categorized into three separate and mutually exclusive groups: 1) Flagged Offenders, 2) Dangerous Offenders, and 3) Long-Term Offenders. File information for all three groups was gathered from the National Flagging System coordinators across Canada. Sexual recidivism was defined as any subsequent offences that were considered sexually motivated (including offence, charge, or conviction) and their recidivism information was obtained from official (CPIC) and unofficial sources (news articles).

Measures

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012). Static-99R is a 10-item empirical actuarial risk tool designed to assess the risk of sexual recidivism among adult males with a history of sexual offending. Static-99R is identical to Static-99 except that it contains revised age weights. The total score (ranging from -3 to 12) is calculated by summing all items and can be used to place individuals in one of five risk categories: Level I - very low risk (scores of -3 to -2), Level II - below average risk (scores of -1 to 0), Level III - average risk (scores of 1 to 3), Level IVa - above average risk (scores of 4 to 5), and Level IVb - well above average risk (scores of 6 or higher; Hanson et al., 2017). When necessary, Static-99R

scores were computed from Static-99 scores by using the individual's date of birth to calculate the updated age item.

Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012).

Static-2002R is also an empirical actuarial risk assessment tool for an adult male with a history of sexual offending. The scale has 14 items grouped into five main subscales: Age at Release (1 item), Persistence of Sex Offending (3 items), Sexual Deviance (3 items), Relationship to Victims (2 items), and General Criminality (5 items). The Static-2002R items are identical to the original Static-2002 except for updated age weights. The total score for Static-2002R can range from -2 to 13 and can be used to place individuals in one of five risk categories: Level I - very low risk (scores of -2 to -1), Level II - below average risk (scores of 0 to 1), Level III - average risk (scores of 2 to 4), Level IVa - above average risk (scores of 5 to 6), and Level IVb - well above average risk (scores of 7 or higher; Hanson et al., 2017). When necessary, Static-2002R scores were computed from Static-2002 scores by using the individual's date of birth to calculate the updated age item.

Sexual recidivism. Sexual recidivism was defined as any subsequent offences that were considered sexually motivated (offence, charge or conviction). Some studies, however, did not consider Category B sex offences as defined by the Static-99R coding rules (e.g., possession of child pornography, prostitution offenses; Phenix et al., 2017). To assess sexual recidivism, official criminal records were used for all of the studies (e.g., CPIC) and unofficial sources were used for a subset of the studies (e.g., news articles).

Psychological constructs of Static-99/R and Static-2002R. Three psychological constructs were identified from the items of Static-99R and Static-2002R: 1) persistence/paraphilia, related to deviant sexual interests, but without intent

to harm their victims (e.g., pedophilia, voyeurism, and exhibitionism), 2) youthful stranger aggression, related to the intent to sexually harm the victim (e.g., sexual sadism, hostility towards women), and 3) general criminality, related to the antisocial trait (e.g., impulsivity, lack of remorse, rule violations; Brouillette-Alarie et al., 2016; Brouillette-Alarie et al., 2017). These psychological constructs were used to compare risk relevant characteristics between Indigenous and White individuals who had committed sexual crimes.

Plan of Analysis

Aggregation of findings. AUC values and logistic regression parameters across the studies were aggregated using fixed-effect meta-analyses (Borenstein et al., 2009; Hanson & Broom, 2005). Fixed-effect analyses have the advantage of providing an estimate of between-study variability (Q) that can be used to compare the variability across the studies (Borenstein et al., 2009). The Q statistics is distributed as a chi-square, with $k-1$ degree of freedom (k = the number of studies).

Comparing the risk-relevant characteristics. For assessing the relationship of risk-relevant characteristics between Indigenous and White study groups, AUC analyses (Swets et al., 2000) were used. The AUC can vary between 0 and 1, with .50 indicating no difference between the groups. AUCs above .50 indicate that the Indigenous study group has higher levels of the risk relevant characteristic compared to the White study group. As a rough heuristic, an AUC of .56 corresponds to small effect size, while .64 reflects a moderate effect, and .71 reflects a large effect size (Rice & Harris, 2005). In contrast, AUCs below .50 indicate that the White study group has higher levels of the risk relevant characteristic compared to the Indigenous study group (.44 for a small effect, .33 for a moderate effect, and .29 for a large effect). An AUC value is statistically significant if the 95% confidence interval does

not include .50.

When a risk factor was a binary variable, the odds ratio (instead of AUC) was calculated with 0.5 added to each cell to stabilize the variance (Fleiss & Berlin, 2009). An odds ratio is defined as $p/(1-p)$, where p is the raw proportion of the sample with the characteristic. Odds ratios above 1 indicate that the Indigenous study group has higher levels of the risk relevant characteristic compared to the White study group. For example, an odds ratio of 2 would indicate that the Indigenous study group has twice the odds of having the risk-relevant characteristic compared to the White study group. No association is indicated when the 95% confidence interval of the odds ratio contains 1 (i.e., the odds are equal for both groups).

Predictive validity of Static-99R and Static-2002R. Assessing the predictive accuracy of a risk assessment instrument requires considering discrimination (how different are recidivists from non-recidivists?) as well as calibration (correspondence between expected and observed recidivism rates). For discrimination, we used two statistical methods: 1) AUCs from receiver operating characteristic (ROC) analysis (Swets et al., 2000), and 2) odds ratios from logistic regression (Hosmer, Lemeshow, & Sturdivant, 2013). For calibration, we used a fixed-effect meta-analysis to compare logistic regression parameters (i.e., base rates; Borenstein et al., 2009; Hanson & Broom, 2005).

Area Under the Curve (AUC). AUC values can also be interpreted as the probability that a randomly selected recidivist would have a more deviant score than a randomly selected non-recidivist. AUC values are expected to be smaller in prognostic studies than in diagnostic studies because the outcome of interest in prognostic studies does not exist at the time of assessment, and may never happen (Royston et al., 2009; Helmus & Babchishin, 2017). The AUC has the advantage of

being insensitive to base rates and robust to outliers (Ruscio, 2008). An AUC value is statistically significant if the 95% confidence interval does not include .50.

To compare the unpaired AUC values, bootstrapping ($n = 10,000$) was used to obtain the standard error (*S.E.*) of the differences between AUC1 and AUC2 of the two (original) ROC curves with a *pROC* package for the statistical software R (Robin et al., 2011, 2014). The 95% confidence intervals of differences between AUC1 and AUC2 were calculated as follows:

$$95\% \text{ C.I.} = (AUC1 - AUC2) \pm 1.96 * \text{S.E.}$$

The difference of AUC values is statistically significant if the 95% confidence interval does not include 0.

Odds ratios. Odds ratios indicate the change in relative risk associated with one unit change in Static-99/R and Static-2002R scores. For example, Static-99R scores for the 5-year estimates are associated with a consistent relative risk increase of approximately 1.28 (Hanson et al., 2016) in the high-risk/high-needs samples. This indicates that the odds of recidivism increase 1.28 times as each Static-99R score increases. The primary advantage of odds ratios is that they are less influenced by a restriction of range than AUCs (Hanson, 2008).

Comparing logistic regression parameters. Calibration was examined by comparing the extent to which logistic regression parameters, such as intercept values (centered on Static-99R scores of 2 [$B0_2$] and Static-2002R scores of 3 [$B0_3$]), differed from the logistic regression parameters for the high-risk/high-needs sample norms (Hanson et al., 2016). $B0$ coefficients centered on 2 for the Static-99R and centered on 3 for the Static-2002R (median values among those with a sexual offence conviction in Canada) represent the expected recidivism rate for sex offenders in the middle of the risk distribution. Specifically, the $B0_2$ represents the expected recidivism rate for a

Static-99R score of 2 (p_2) in logit units ($\ln[p_2/\{1-p_2\}]$) and the $B0_3$ represents the expected recidivism rate for a Static-2002R score of 3 (p_3) in logit units ($\ln[p_3/\{1-p_3\}]$).

In addition, high-risk/high-needs sample norms were used to evaluate the calibration of Static-99R and Static-2002R given that the current samples were preselected as individuals with a high risk to reoffend (e.g., individuals flagged as potential Dangerous Offenders or who had attended high-intensity treatment). Differences between the parameters in the current sample and those of the norms were tested using Cochran's Q statistic from a fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005). The Q statistic is distributed as a chi-square, with $k-1$ degree of freedom ($k =$ the number of studies, which in this case $k = 2$: Indigenous study group versus the norms, and, separately, the White study group versus the norms).

The index used to assess the differences between the parameters of Indigenous and White study groups was the b_i -difference statistic, which was calculated for each parameter by subtracting the b_i of the White samples from b_i of the Indigenous samples. The variance for the b_i difference was calculated according to the Ley (1972) using the following formula:

$$Var (b_i \text{ diff}) = s_x^2 + s_y^2 - 2r_{xy}s_x s_y$$

where s_x is the standard deviation of the d_i from the Indigenous sample, s_y is the standard deviation of the d_i from the White sample, and r_{xy} is the correlation coefficient estimating the relationship between the average effect size for the Indigenous sample and the White sample.

Results

The Indigenous study group was, on average, approximately 4 years younger

than the White study group (36.1 versus 40.6 years; AUC = .61, 95% CI [.57, .65]). The average Static-99R and Static-2002R total scores for both groups were in the Above Average range (Risk Level IVa, see Table 3.1), and were approximately a 1/2 point higher for the Indigenous study group than the White study group (4.8 versus 4.3 for Static-99R, AUC = .57, 95% CI [.53, .60]; 6.1 versus 5.7 for Static-2002R, AUC = .55, 95% CI [.51, .59]).

Comparing Risk Relevant Characteristics

The Indigenous study group showed less persistence/paraphilia than the White study group (AUC = .39). Compared to the White study group, the Indigenous study group was significantly less likely to have a history of non-contact sex offences (odds ratio = 0.45), male victim (odds ratio = 0.22), unrelated young victim (odds ratio = 0.27), and prior sex offences (AUC = 0.45; Table 3.2).

Regarding youthful stranger aggression, the Indigenous study group showed a significantly higher score than the White study group (AUC of .60). The difference was attributed to two items (age at release and index non-sexual violence). Specifically, the Indigenous study group was significantly younger than the White study group (AUC of .59) and more likely to have been convicted of a violent index offence (odds ratio = 1.61; Table 3.2).

The Indigenous study group showed more general criminality than the White study group (AUC = .72; Table 3.2). All the items associated with general criminality were significantly higher for the Indigenous study group than the White study group. Compared to the White study group, the Indigenous study group was more likely to have a history of community supervision violation, prior non-sexual violent crimes, and prior criminal involvement.

Table 3.1

Age at Release, Static-99R Scores, and Static-2002R Scores for Indigenous and White Study Groups

Study	Indigenous				White			
	<i>n</i>	Age at release	Static-99R	Static-2002R	<i>n</i>	Age at release	Static-99R	Static-2002R
		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Bonta & Yessine (2005)	37	36.8 (7.8)	5.1 (2.1)	-	118	40.4 (10.1)	5.0 (2.2)	-
Haag (2005)	50	33.7 (8.7)	4.2 (1.9)	5.8 (1.9)	141	38.6 (10.1)	3.9 (2.4)	5.5 (2.5)
Olver et al. (2018)	425	-	4.6 (2.4)	-	739	-	3.2 (2.9)	-
Brankley et al. (2017a)	36	35.5 (8.4)	4.7 (2.0)	6.3 (2.1)	346	40.4 (11.6)	4.2 (2.4)	5.9 (2.5)
Lee et al. (2018)	105	38.3 (9.2)	5.4 (2.1)	6.4 (2.1)	216	43.9 (12.3)	4.8 (2.7)	5.8 (2.8)
Average (fixed-effect)	653	36.1 (4.2)	4.8 (0.9)	6.1 (1.2)	1,560	40.6 (5.5)	4.3 (1.1)	5.7 (1.5)

Table 3.2

Differences in Static-99R and Static-2002R Items between Indigenous and White Study Groups

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	n/N	%	n/N			
<u>Persistence/Paraphilia</u>							
Prior sex offences (Static-99R; Item #5)							.453 [.413, .493]
None	36.0	82/228	34.7	285/821			
1-2 charges or 1 conviction	30.3	69/228	22.5	185/821			
3-5 charges or 2-3 convictions	21.9	50/228	19.5	160/821			
6+ charges or 4+ convictions	11.8	27/228	23.3	191/821			
Prior sentencing occasions for sexual offences (Static-2002R; Item #2)							.470 [.424, .515]
None	36.0	82/191	38.8	273/703			
1	26.8	61/191	32.7	230/703			
2-3	18.4	42/191	20.5	144/703			
4 or more	2.6	6/191	8.0	56/703			
Rate of sexual offending (Static-2002R; Item #4)						1.04	[0.73, 1.48]
< 1 sentencing occasion every 15 years	71.7	137/191	72.4	509/703			
≥ 1 sentencing occasion every 15 years	28.3	54/191	27.6	194/703			

Table 3.2 (continued)

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Non-contact sex offences (Static-99R; Item #7)					0.45	[0.24, 0.82]	
No	94.7	215/227	88.6	727/821			
Yes	5.3	12/227	11.4	94/821			
Male victim (Static-99R; Item #10)					0.22	[0.15, 0.35]	
No	88.6	202/228	63.2	519/821			
Yes	11.4	26/228	36.8	302/821			
2+ victims < 12 year-old, one unrelated (Static-2002R; Item # 7)					0.27	[0.18, 0.40]	
No	82.7	158/191	56.2	395/703			
Yes	17.3	33/191	43.8	308/703			
Persistence/Paraphilia	<i>M</i> = 1.48 (<i>SD</i> = 1.53), <i>n</i> = 191		<i>M</i> = 2.18 (<i>SD</i> = 1.81), <i>n</i> = 703				.387 [.343, .430]

Table 3.2 (continued)

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
<u>Youthful stranger aggression</u>							
Age at release (Static-99R; Item #1)							.590 [.546, .635]
18 to 34.9	0.9	2/221	7.0	55/789			
35 to 39.9	33.9	75/221	41.7	329/789			
40 to 59.9	18.6	41/221	18.0	142/789			
60 or older	46.6	103/221	33.3	263/789			
Juvenile sexual arrest prior to age 18 (Static-2002R; Item #3)					1.01	[0.56, 1.81]	
No	92.1	176/191	92.0	647/703			
Yes	7.9	15/191	8.0	56/703			
Ever lived with a lover for +2 years (Static-99R; Item #2)					1.23	[0.91, 1.66]	
Yes	57.7	131/227	62.7	510/814			
No	42.3	96/227	37.3	304/814			
Index non-sexual violence - any convictions (Static-99R; Item #3)					1.61	[1.18, 2.19]	
No	63.0	143/227	73.2	601/821			
Yes	37.0	84/227	26.8	220/821			

Table 3.2 (continued)

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Unrelated victim (Static-99R; Item #8)					0.73	[0.47, 1.13]	
No	13.6	31/228	10.4	85/821			
Yes	86.4	197/228	89.6	736/821			
Stranger victim (Static-99R; Item #9)					1.21	[0.90, 1.62]	
No	50.4	115/228	55.1	452/820			
Yes	49.6	113/228	44.9	368/820			
Youthful stranger aggression	<i>M</i> = 3.26 (<i>SD</i> = 1.62), <i>n</i> = 184		<i>M</i> = 2.68 (<i>SD</i> = 1.77), <i>n</i> = 668				.599 [.553, .644]

Table 3.2 (continued)

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	n/N	%	n/N			
<u>General Criminality</u>							
Any community supervision violation (Static-2002R; Item #12)					3.88	[2.36, 6.36]	
No	9.9	19/191	30.4	214/703			
Yes	90.1	172/191	69.6	489/703			
Years free prior to index sex offence (Static-2002R; Item #13)					3.11	[2.05, 4.72]	
> 36 months prior to sexual offence AND > 48 months prior to sex conviction	15.7	30/191	37.0	260/703			
< 36 months prior to sexual offence OR < 48 months prior to sex conviction	84.3	161/191	63.0	443/703			
Prior criminal involvement (Static-2002R; Item #10)					2.58	[1.37, 4.86]	
No	5.8	11/191	14.1	99/703			
Yes	94.2	180/191	85.9	604/703			
Prior sentencing occasions for anything (Static-2002R; Item #11)							.670 [.628, .712]
0-2	14.7	28/191	42.5	299/703			
3-13	64.9	124/191	50.9	358/703			
14 +	20.4	39/191	6.5	46/703			

Table 3.2 (continued)

Items	Indigenous		White		Odds Ratio	95% CI	AUC [95% C.I.]
	%	<i>n/N</i>	%	<i>n/N</i>			
Prior sentencing dates (Static-99R; Item #6)					3.17	[2.28, 4.41]	
3 or less	24.6	56/228	50.9	418/821			
4 or more	75.4	172/228	49.1	403/821			
Prior non-sexual violence – any convictions (Static-99R; Item #4)					3.25	[2.36, 4.47]	
No	27.6	63/228	55.5	455/820			
Yes	72.4	165/228	44.5	365/820			
Prior non-sexual violence sentencing occasion (Static-2002R; Item #14)					3.96	[2.79, 5.61]	
No	27.7	53/191	60.5	425/703			
Yes	72.3	138/191	39.5	278/703			
General Criminality	<i>M</i> = 4.47 (<i>SD</i> = 1.43), <i>n</i> = 191		<i>M</i> = 3.22 (<i>SD</i> = 1.74), <i>n</i> = 703				.722 [.682, .763]

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Predictive Accuracy of Three Psychological Structures and Items

Persistence/paraphilia predicted sexual recidivism for the White study group (AUC of .65). The AUC value for the Indigenous study group was, however, lower than that of the White study group (AUC of .58), but the difference was not statistically significant. All the items of persistence/paraphilia had lower predictive accuracy for the Indigenous study group compared the White study group. Only two items (prior sex offences and unrelated young victim) significantly predicted sexual recidivism rates for the Indigenous study group. For one item (non-contact sex offences), the direction of the effect was reversed for the Indigenous study group (i.e., Indigenous individuals with a history of non-contact sex offences were less likely to commit another sexual crime in the future compared to Indigenous individuals with a history of only contact sexual offences; Table 3.3).

Youthful stranger aggression predicted sexual recidivism for the White study group (AUC of .64), but not for the Indigenous study group (AUC of .53), which was significantly lower than that of the White study group. Specifically, the item of age at release for the Indigenous study group did not predict sexual recidivism and was significantly lower than the value for the White study group (AUCs of .51 versus .62). In other words, sexual recidivism risk for the Indigenous study group was unrelated to age at release. In addition, it was found that the relationship between the Indigenous study group and their victims (i.e., unrelated/stranger victim) were not associated with sexual recidivism risk. Only one item (index non-sexual violence) predicted sexual recidivism for the Indigenous study group (AUC of .58) whereas it did not predict for the White study group (AUC of .51). The difference was, however, not statistically significant (Table 3.3).

Table 3.3

Predictive Accuracy of Static-99R and Static-2002R items for Sexual Recidivism for Indigenous and White Study Groups

Items	Indigenous		White		Diff.	95% C.I.
	AUC [95% C.I.]	n/N	AUC [95% C.I.]	n/N		
<u>Persistence/Paraphilia</u>						
Prior sex offences (Static-99R; Item #5)	.608 [.528, .688]	59/219	.626 [.582, .670]	187/759	- .018	[-.108, .072]
Prior sentencing occasions for sexual offences (Static-2002R; Item #2)	.571 [.482, .660]	50/182	.628 [.582, .674]	171/641	- .057	[-.159, .045]
Rate of sexual offending (Static-2002R; Item #4)	.573 [.496, .650]	50/182	.606 [.564, .647]	171/641	- .033	[-.120, .054]
Non-contact sex offences (Static-99R; Item #7)	.474 [.448, .499]	59/219	.522 [.494, .551]	187/759	- .048	[-.086, -.010]
Male victim (Static-99R; Item #10)	.495 [.449, .541]	59/219	.538 [.498, .579]	187/759	- .043	[-.103, .017]
2+ victims < 12 year-old, one unrelated (Static-2002R; Item # 7)	.572 [.502, .641]	50/182	.573 [.530, .617]	171/641	- .001	[-.050, .048]
Persistence/Paraphilia	.581 [.489, .673]	50/182	.645 [.597, .693]	171/641	- .127	[-.333, .079]

Table 3.3 (continued)

Items	Indigenous		White		<i>Diff.</i>	95% C.I.
	AUC [95% C.I.]	<i>n/N</i>	AUC [95% C.I.]	<i>n/N</i>		
<u>Youthful stranger aggression</u>						
Age at release (Static-99R; Item #1)	.514 [.436, .592]	59/219	.616 [.574, .659]	186/757	-.102	[-.189, -.015]
Juvenile sexual arrest prior to age 18 (Static-2002R; Item #3)	.502 [.458, .546]	50/182	.534 [.507, .561]	171/641	-.032	[-.084, .020]
Ever lived with a lover for +2 years (Static-99R; Item #2)	.470 [.396, .543]	59/218	.531 [.490, .571]	186/753	-.061	[-.143, .021]
Index non-sexual violence - any convictions (Static-99R; Item #3)	.580 [.506, .654]	59/218	.512 [.476, .549]	187/759	.068	[-.013, .149]
Unrelated victim (Static-99R; Item #8)	.516 [.466, .566]	59/219	.537 [.516, .558]	187/759	-.021	[-.074, .032]
Stranger victim (Static-99R; Item #9)	.513 [.438, .588]	59/219	.577 [.536, .618]	187/758	-.064	[-.148, .020]
Youthful stranger aggression	.529 [.439, .620]	50/182	.644 [.597, .690]	170/636	-.115	[-.217, -.013]

Table 3.3 (continued)

Items	Indigenous		White		<i>Diff.</i>	95% C.I.
	AUC [95% C.I.]	<i>n/N</i>	AUC [95% C.I.]	<i>n/N</i>		
<u>General Criminality</u>						
Any community supervision violation (Static-2002R; Item #12)	.572 [.542, .602]	50/182	.616 [.582, .649]	171/641	- .044	[-.089, .001]
Years free prior to index sex offence (Static-2002R; Item #13)	.517 [.459, .576]	50/182	.579 [.539, .619]	171/641	- .062	[-.133, .009]
Prior criminal involvement (Static-2002R; Item #10)	.500 [.461, .539]	50/182	.548 [.522, .573]	171/641	-.048	[-.095, -.001]
Prior sentencing occasions for anything (Static-2002R; Item #11)	.578 [.499, .578]	50/182	.592 [.549, .635]	171/641	- .014	[-.105, .077]
Prior sentencing dates (Static-99R; Item #6)	.568 [.509, .626]	59/219	.566 [.525, .607]	187/759	.002	[-.096, .100]
Prior non-sexual violence – any convictions (Static-99R; Item #4)	.563 [.501, .625]	59/219	.541 [.500, .583]	187/758	.022	[-.051, .095]
Prior non-sexual violence sentencing occasion (Static-2002R; Item #14)	.545 [.475, .616]	50/182	.515 [.472, .558]	171/641	.030	[-.051, .111]
General Criminality	.589 [.501, .677]	50/182	.613 [.568, .658]	171/641	-.024	[-.122, .074]

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Overall, general criminality significantly and similarly predicted sexual recidivism for both Indigenous and White study groups (AUC of .59 and .61). The item of prior criminal involvement did not predict sexual recidivism for the Indigenous study group (AUC of .50), and it was also significantly lower than that of the White study group (AUC of .55; Table 3.3). Most AUC values between Indigenous and White study groups, however, did not differ significantly.

Predictive Validity Static-99R and Static-2002R

Overall, sexual recidivism rates for the Indigenous study group were higher than those of White study group (Table 3.4). For the Indigenous study group, 16.2% sexually reoffended within a fixed 5-year follow-up period and 24.0% sexually reoffended within a fixed 10-year follow-up period. For the White study group, 11.9% sexually reoffended within a fixed 5-year follow-up period and 19.3% sexually reoffended within a fixed 10-year follow-up period.

Static-99R

Discrimination. Using the fixed 5-year follow-up, Static-99R was able to discriminate recidivists from non-recidivists for both Indigenous and White study groups. Whereas the White study group had the moderate AUC value of .70, the AUC value for the Indigenous study group was smaller (AUC = .61). With the fixed 10-year follow-up, the AUC value of the Indigenous study group increased to moderate effect size (AUC of .66) but was still lower than the AUC value of the White study group (AUC of .72; Table 3.4).

Table 3.4

Sexual Recidivism Rates and AUC values for Static-99R for Indigenous and White Study Groups

Study	Indigenous			White		
	Sexual Recidivism		AUC [95% C.I.]	Sexual Recidivism		AUC [95% C.I.]
	%	<i>n/N</i>		%	<i>n/N</i>	
Fixed 5-year						
Bonta & Yessine (2005)	19.2	5/26	.60 [.25, .96]	14.4	13/90	.71 [.58, .85]
Haag (2005)	28.0	14/50	.60 [.42, .77]	17.7	25/141	.70 [.59, .81]
Olver et al. (2018)	14.3	57/400	.62 [.55, .70]	9.6	64/665	.71 [.65, .77]
Brankley et al. (2017a)	11.4	4/35	.62 [.27, .97]	11.4	35/308	.70 [.61, .80]
Lee et al. (2018)	19.3	17/88	.56 [.41, .72]	15.8	27/171	.69 [.59, .79]
Average (fixed-effect)	16.2	97/599	.61 [.55, .67]^{ns}	11.9	164/1,375	.70 [.66, .74]^{ns}
Fixed 10-year						
Olver et al. (2018)	23.4	64/274	.67 [.59, .74]	16.7	72/432	.71 [.65, .77]
Brankley et al. (2017a)	20.0	7/35	.60 [.34, .87]	21.7	65/299	.74 [.67, .80]
Lee et al. (2018)	28.8	19/66	.66 [.51, .81]	23.1	27/117	.70 [.59, .82]
Average (fixed-effect)	24.0	90/375	.66 [.60, .72]^{ns}	19.3	164/848	.72 [.68, .76]^{ns}

Note. *ns* indicates no significant variability across studies (*Q* statistics). Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 3.5

Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism for Indigenous Study Group

Study	Base rate (Static-99R score of 2 in logit units)		Relative risk		Odds Ratio	95% C.I.		
	$B0_2$	SE	$B1$	SE		Lower	Upper	
Fixed 5-year								
Bonta & Yessine (2005)	- 1.876	(13.3%)	.954	.162	.234	1.18	0.72	1.63
Haag (2005)	- 1.384	(20.0%)	.551	.187	.181	1.21	0.85	1.56
Olver et al. (2018)	- 2.393	(8.4%)	.265	.201	.067	1.22	1.09	1.35
Brankley et al. (2017a)	- 2.366	(8.6%)	.991	.114	.283	1.12	0.57	1.68
Lee et al. (2018)	- 2.030	(11.6%)	.623	.168	.148	1.18	0.89	1.47
$Q (df = 4)$	2.93, $p = .570$		0.14, $p = .998$					
I^2	0.0		0.0					
Average (fixed-effect)	- 2.175	(10.2%)	.212	.190	.056	1.21	1.09	1.35
Fixed 10-year								
Olver et al. (2018)	- 2.001	(11.9%)	.276	.274	.071	1.32	1.18	1.45
Brankley et al. (2017a)	- 1.736	(15.0%)	.783	.126	.226	1.13	0.69	1.58
Lee et al. (2018)	- 2.195	(10.0%)	.756	.342	.175	1.41	1.06	1.75
$Q (df = 2)$	0.18, $p = .914$		0.58, $p = .750$					
I^2	0.0		0.0					
Average (fixed-effect)	- 1.995	(12.0%)	.246	.271	.063	1.31	1.16	1.48

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 3.6

Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism for White Study Group

Study	Base rate (Static-99R score of 2 in logit units)			Relative risk		Odds Ratio	95% C.I.	
	<i>B0₂</i>		<i>SE</i>	<i>B1</i>	<i>SE</i>		Lower	Upper
Fixed 5-year								
Bonta & Yessine (2005)	- 3.305	(3.5%)	.863	.442	.188	1.56	1.19	1.92
Haag (2005)	- 2.343	(8.8%)	.392	.324	.106	1.38	1.17	1.59
Olver et al. (2018)	- 2.889	(5.3%)	.212	.287	.054	1.33	1.23	1.44
Brankley et al. (2017a)	- 3.047	(4.5%)	.366	.340	.089	1.40	1.23	1.58
Lee et al. (2018)	- 2.539	(7.3%)	.401	.279	.092	1.32	1.14	1.50
<i>Q</i> (<i>df</i> = 4)	2.75, <i>p</i> = .600			0.90, <i>p</i> = .925				
<i>I</i> ²	0.0			0.0				
Average (fixed-effect)	- 2.797	(5.7%)	.151	.306	.038	1.36	1.26	1.46
Fixed 10-year								
Olver et al. (2018)	- 2.149	(10.4%)	.228	.326	.059	1.38	1.27	1.50
Brankley et al. (2017a)	- 2.437	(8.0%)	.292	.407	.075	1.50	1.36	1.65
Lee et al. (2018)	- 2.070	(11.2%)	.406	.300	.099	1.35	1.15	1.54
<i>Q</i> (<i>df</i> = 2)	0.78, <i>p</i> = .677			1.02, <i>p</i> = .600				
<i>I</i> ²	0.0			0.0				
Average (fixed-effect)	- 2.227	(9.7%)	.164	.346	.042	1.41	1.30	1.53

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Before comparing predictive accuracies across groups, we first examined the degree of consistency within each group. For the Indigenous study group, both the 5-year (5 studies) and 10-year sexual recidivism rates (3 studies) at a Static-99R score of 2 ($B0_2$) were very similar across the studies included in this meta-analysis (both $p > .50$, see Table 3.5). The relative risk (BI) of Static-99R across the studies was also very similar with a fixed 5-year follow-up (5 studies) as well as with a fixed 10-year follow-up (3 studies; both $p > .70$, see Table 3.5).

For the White study group, both the 5-year (5 studies) and 10-year sexual recidivism rates (3 studies) at a Static-99R score of 2 ($B0_2$) across the studies were very similar (both $p > .60$, see Table 3.6). The relative risk (BI) of Static-99R across the studies was also very similar with a fixed 5-year follow-up (5 studies; $p = .93$) as well as with a fixed 10-year follow-up (3 studies; $p = .60$; Table 3.6).

The 5-year sexual recidivism rates at a Static-99R score of 2 for the Indigenous study group were significantly higher than that of the White study group (10.2% versus 5.7%; Table 3.7; Figure 3.1). The 10-year sexual recidivism rates at a Static-99R score of 2 for the Indigenous study group were, however, similar to that of the White study group (11.9% versus 9.7%; Table 3.7; Figure 3.2). The discrimination (change in relative risk) was higher for the White study group than for the Indigenous study group with both 5-year (odds ratios = 1.21 versus 1.36) as well as 10-year follow-up periods (1.31 versus 1.41), but these differences in discrimination between racial groups were not statistically significant (Table 3.7; Figure 3.1; Figure 3.2).

Table 3.7

Comparison of Logistic Regression Parameters for Static-99R Predicting Sexual Recidivism with Meta-Average (from Hanson et al., 2016)

	Meta-Average (HRHN Norms)	Indigenous	White	B _{Indigenous-White}	95% C.I.
<u>Fixed 5-year</u>					
<u>Base rate</u>					
$B0_2$ (SD)	- 2.06 (.153) (11.3%)	- 2.18 (.212) (10.2%)	- 2.80 (.151) (5.7%)	0.623	[0.179, 1.066]
Q_{Δ} (df = 1)		0.18	11.60^{***}		
<u>Relative risk</u>					
$B1$ (SD)	.250 (.042)	.190 (.055)	.306 (.038)	- 0.116	[- 0.245, 0.013]
Q_{Δ} (df = 1)		0.755	0.971		
<u>Fixed 10-year</u>					
<u>Base rate</u>					
$B0_2$ (SD)	- 1.44 (.186) (19.2%)	- 2.00 (.246) (11.9%)	- 2.23 (.164) (9.7%)	0.232	[- 0.307, 0.770]
Q_{Δ} (df = 1)		3.64	9.99^{**}		
<u>Relative risk</u>					
$B1$ (SD)	.231 (.056)	.271 (.063)	.346 (.042)	- 0.075	[- 0.222, 0.071]
Q_{Δ} (df = 1)		0.28	2.72		

Note. ^{***} $p < .001$, ^{**} $p < .01$, ^{*} $p < .05$. r (between Indigenous and White) = 0.260 for $B0_2$ and 0.208 for $B1$. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

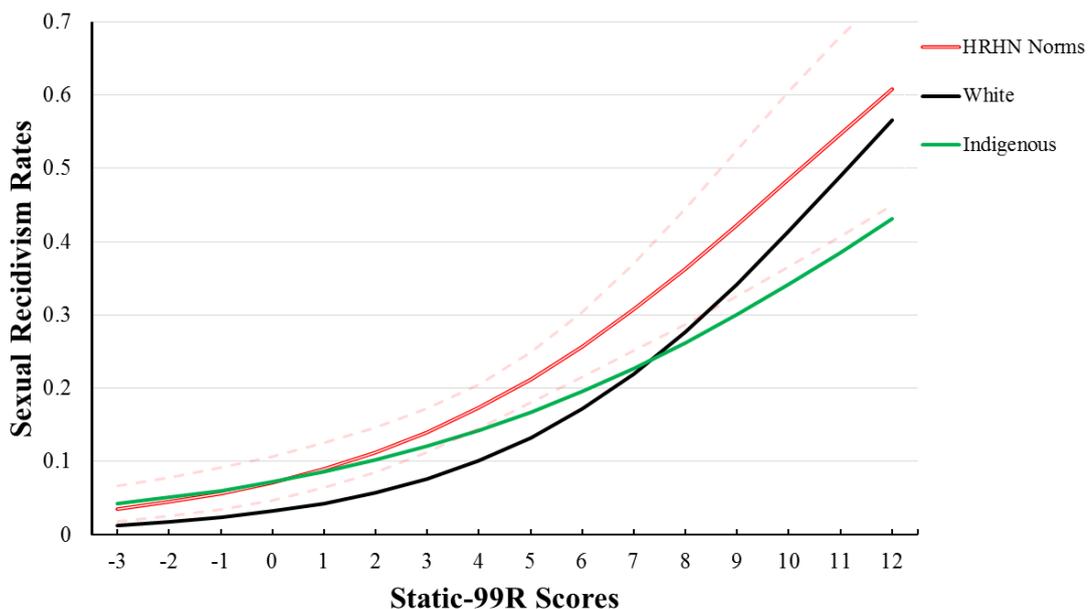


Figure 3.1. Logistic curves of Static-99R for Indigenous and White study groups with the High Risk/High Need Norms with a fixed 5-year follow-up

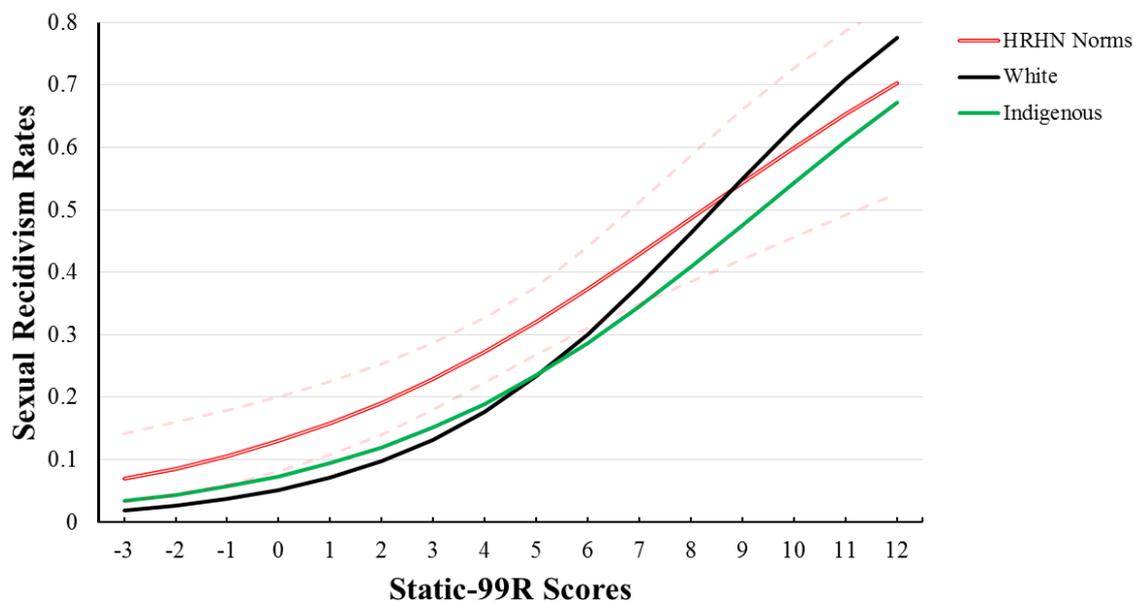


Figure 3.2. Logistic curves of Static-99R for Indigenous and White study groups with the High Risk/High Need Norms with a fixed 10-year follow-up

Calibration. For the Indigenous study group, when compared to the high-risk/high-needs norms (from Hanson, et al., 2016), the adjusted (score of 2) base rate with a fixed 5-year follow up period was slightly lower (11.3% versus 10.2%), and the discrimination was smaller (odds ratios = 1.28 versus 1.21), but those differences were not statistically significant (Table 3.7; Figure 3.1). With a fixed 10-year follow-up period, the adjusted (score of 2) base rate for the Indigenous study group was lower (19.2% versus 11.9%), and the discrimination was larger for the Indigenous study group compared to the norms (odds ratios = 1.26 versus 1.31), but those differences were not statistically significant (Table 3.7; Figure 3.2).

For the White study group, when compared to the high-risk/high-needs norms (from Hanson, et al., 2016), the adjusted (score of 2) base rate with a fixed 5-year follow up period was significantly lower (11.3% versus 5.7%), and the discrimination was larger (odds ratios = 1.28 versus 1.36), but not significant (Table 3.7; Figure 3.1). With a fixed 10-year follow-up period, the adjusted (score of 2) base rate for the White study group was significantly lower (19.2% versus 9.7%), and the discrimination was larger (odds ratios = 1.26 versus 1.41), but not significantly so (Table 3.7; Figure 3.2).

Static-2002R

The sample size of the Indigenous study group was reduced to 173 (3 studies) with a 5-year follow-up period and 101 (2 studies) with a 10-year follow-up period compared to those of Static-99R.

Discrimination. Static-2002R was able to discriminate recidivists from non-recidivists for the White study group (AUCs > .69). The predictive accuracy of Static-2002R for the Indigenous study group was, however, lower (AUCs < .61) and not statistically significant (Table 3.8).

Table 3.8

Sexual Recidivism Rates and AUC values for Static-2002R for Indigenous and White Study Groups

Study	Indigenous			White		
	Sexual Recidivism		AUC [95% C.I.]	Sexual Recidivism		AUC [95% C.I.]
	%	<i>n/N</i>		%	<i>n/N</i>	
Fixed 5-year						
Haag (2005)	28.0	14/50	.56 [.36, .75]	17.7	25/141	.69 [.59, .80]
Brankley et al. (2017a)	11.4	4/35	.57 [.27, .87]	11.4	35/308	.70 [.61, .79]
Lee et al. (2018)	19.3	17/88	.56 [.42, .70]	15.8	27/171	.69 [.59, .79]
Average (fixed-effect)	20.2	35/173	.56 [.45, .67] ^{ns}	14.0	87/620	.69 [.64, .75]^{ns}
Fixed 10-year						
Brankley et al. (2017a)	20.0	7/35	.64 [.40, .88]	21.7	65/299	.74 [.67, .80]
Lee et al. (2018)	28.8	19/66	.60 [.45, .75]	23.1	27/117	.69 [.58, .81]
Average (fixed-effect)	25.7	26/101	.61 [.49, .74] ^{ns}	22.1	92/416	.72 [.67, .78]^{ns}

Note. *ns* indicates no significant variability across studies (*Q* statistics). Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Table 3.9

Logistic Regression Parameters for Static-2002R Predicting Sexual Recidivism for Indigenous Study Group

Study	Base rate (Static-2002R score of 3 in logit units)		Relative risk		Odds Ratio	95% C.I.			
	$B0_3$	SE	$B1$	SE		Lower	Upper		
Fixed 5-year									
Haag (2005)	- 1.212	(22.9%)	.600	.092	.171	1.09	0.76	1.43	
Brankley et al. (2017a)	- 2.540	(7.3%)	1.124	.142	.268	1.15	0.63	1.68	
Lee et al. (2018)	- 1.768	(14.6%)	.563	.093	.132	1.10	0.84	1.36	
$Q (df = 2)$		1.20, $p = .549$		0.03, $p = .985$					
I^2		0.0		0.0					
Average (fixed-effect)		- 1.630	(16.4%)	.386	.099	.097	1.10	0.91	1.34
Fixed 10-year									
Brankley et al. (2017a)	- 2.437	(8.0%)	.997	.293	.229	1.34	0.89	1.79	
Lee et al. (2018)	- 1.560	(17.4%)	.591	.178	.137	1.19	0.93	1.46	
$Q (df = 1)$		0.42, $p = .449$		0.19, $p = .667$					
I^2		0.0		0.0					
Average (fixed-effect)		- 1.788	(14.3%)	.509	.208	.117	1.23	0.97	1.55

Table 3.10

Logistic Regression Parameters for Static-2002R Predicting Sexual Recidivism for White Study Group

Study	Base rate (Static-2002R score of 3 in logit units)		Relative risk		Odds Ratio	95% C.I.			
	$B0_3$	SE	$B1$	SE		Lower	Upper		
Fixed 5-year									
Haag (2005)	- 2.506	(7.5%)	.427	.313	.100	1.37	1.17	1.56	
Brankley et al. (2017a)	- 3.200	(3.9%)	.407	.314	.085	1.37	1.20	1.53	
Lee et al. (2018)	- 2.561	(7.2%)	.410	.248	.085	1.28	1.11	1.45	
$Q (df= 2)$		1.76, $p = .415$		0.37, $p = .829$					
I^2		0.0		0.0					
Average (fixed-effect)		- 2.765	(6.0%)	.239	.289	.052	1.34	1.21	1.48
Fixed 10-year									
Brankley et al. (2017a)	- 2.686	(6.4%)	.332	.391	.072	1.48	1.34	1.62	
Lee et al. (2018)	- 2.040	(11.5%)	.406	.258	.089	1.29	1.12	1.47	
$Q (df= 1)$		1.52, $p = .220$		1.36, $p = .248$					
I^2		0.0		0.0					
Average (fixed-effect)		- 2.427	(8.1%)	.257	.338	.056	1.40	1.26	1.56

Note. Numbers in bold indicate statistical significance (i.e., $p < 0.05$).

Similar to the outcomes of Static-99R, there were very consistent base rates and relative risk levels of Static-2002R across studies within each racial group. Specifically, for the Indigenous study group, both the 5-year (3 studies) and 10-year sexual recidivism rates (2 studies) at a Static-2002R score of 3 ($B0_3$) across the studies were very similar (both $p > .40$; Table 3.9). The overall outcomes of Static-2002R, however, did not seem to be stable given the small number of studies in the meta-analysis. For example, the average base rate ($B0_3$) within a fixed 5-year follow-up (16.4%) was higher than that within a fixed 10-year follow-up (14.3%) because there were a different number of studies in each analysis. The relative risk ($B1$) of Static-2002R across the studies was also very similar with a fixed 5-year follow-up (3 studies) as well as with a fixed 10-year follow-up (2 studies; both $p > .60$; see Table 3.9). Nevertheless, the relative risk (i.e., odds ratios) of Static-2002R was not statistically significant (i.e., poor discrimination; Table 3.9).

For the White study group, both the 5-year (3 studies) and 10-year sexual recidivism rates (2 studies) at a Static-2002R score of 3 ($B0_3$) across the studies were very similar (both $p > .20$; Table 3.10). The relative risk ($B1$) of Static-2002R across the studies was also very similar with a fixed 5-year follow-up (3 studies) as well as with a fixed 10-year follow-up (2 studies; $p > .20$; Table 3.10).

The 5-year sexual recidivism rates at a Static-2002R score of 3 for the Indigenous study group were significantly higher than that of the White study group (16.4% versus 6.0%; Table 3.11; Figure 3.3). The 10-year sexual recidivism rates at a Static-2002R score of 3 for the Indigenous study group were also higher than that of the White study group, but the difference was not statistically significant (14.3% versus 8.1%). The discrimination (change in relative risk) was higher for the White study group than the Indigenous study group for both 5-year (odds ratios = 1.10

versus 1.34) and 10- year follow-up periods (1.23 versus 1.40); however, the difference between racial groups was statistically significant only for the 5-year follow-up period (Table 3.11; Figure 3.3).

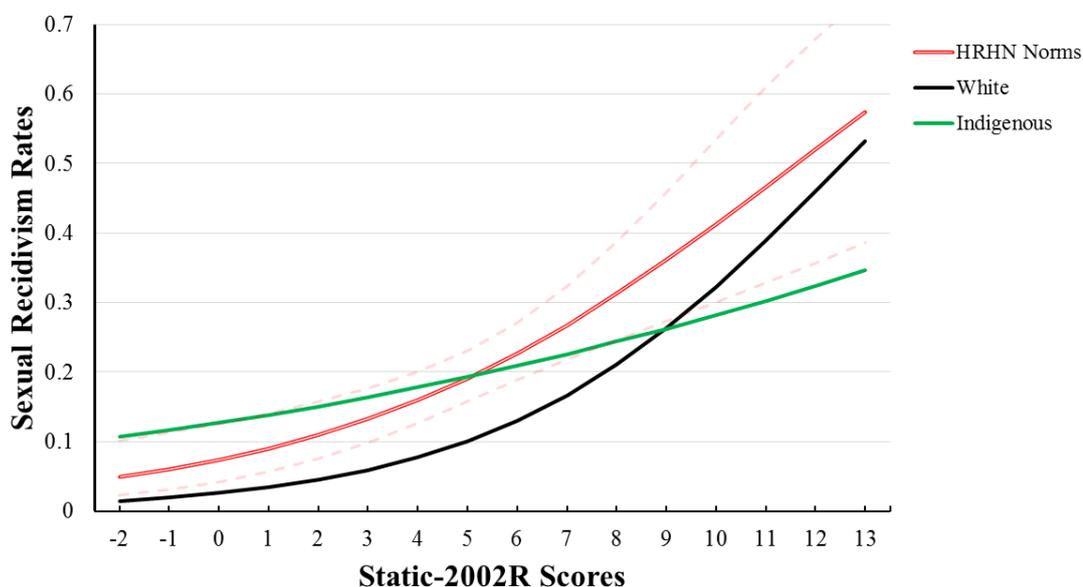


Figure 3.3. Logistic curves of Static-2002R for Indigenous and White Study groups with the High Risk/High Need Norms with a fixed 5-year follow-up

Calibration. For the Indigenous study group, when compared to the high-risk/high-needs norms (from Hanson, et al., 2016), the adjusted (score of 3) base rate with a fixed 5-year follow up period was slightly higher (13.2% versus 16.4%; $p = .55$), and the discrimination was smaller, but not significantly so (odds ratios = 1.24 versus 1.10; $p = .27$; Table 3.11). Nevertheless, the overall calibration was poor at the high-risk level (e.g., 6 or higher - Level IVb; Figure 3.3). There were meaningful discrepancies between expected and observed recidivism rates at the highest risk levels (i.e., overestimation).

Table 3.11

Comparison of Logistic Regression Parameters for Static-2002R Predicting Sexual Recidivism with Meta-Average (from Hanson et al., 2016)

	Meta-Average (HRHN Norms)	Indigenous	White	B Indigenous-White	95% C.I.
<u>Fixed 5-year</u>					
<u>Base rate</u>					
$B0_3$ (SD)	- 1.88 (.153) (13.2%)	- 1.63 (.386) (16.4%)	- 2.76 (.239) (6.0%)	1.135	[0.402, 1.868]
Q_{Δ} (df = 1)		0.36	9.75**		
<u>Relative risk</u>					
$B1$ (SD)	.218 (.042)	.099 (.097)	.289 (.051)	- 1.190	[- 0.379, - 0.001]
Q_{Δ} (df = 1)		1.24	1.15		
<u>Fixed 10-year</u>					
<u>Base rate</u>					
$B0_3$ (SD)	-	- 1.78 (.509) (14.3%)	- 2.43 (.257) (8.1%)	0.639	[- 0.303, 1.581]
Q_{Δ} (df = 1)		-	-		
<u>Relative risk</u>					
$B1$ (SD)	-	.208 (.117)	.338 (.056)	- 0.130	[- 0.356, 0.095]
Q_{Δ} (df = 1)		-	-		

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. r (between Indigenous and White) = 0.359 for $B0_2$ and 0.280 for $B1$. Numbers in bold indicate statistical significance (i.e., $p < 0.05$)

For the White study group, when compared to the high-risk/high-needs norms (from Hanson, et al., 2016), the adjusted (score of 3) base rate with a fixed 5-year follow up period was significantly lower (13.2% versus 6.0%), and the discrimination was larger, but not significant (odds ratios = 1.24 versus 1.34; Table 3.11; Figure 3.3).

Discussion

Professional and legal concerns, and in particular, the Ewert decision (2018), motivates research on the extent to which widely utilized risk assessment tools, such as the Static-99R and Static-2002R, are valid for use with individuals of Indigenous heritage in the criminal justice system. To this end, the current study examined the potential cultural bias in the predictive validity (both discrimination and calibration) of the Static-99R and Static-2002R for Indigenous individuals with a history of sexual crimes in Canada. Furthermore, we examined whether Indigenous peoples who commit sexual crimes have different patterns of risk-relevant characteristics from Whites and whether these differences could potentially help to explain any observed cultural bias.

Predictive Accuracy of Static-99R and Static-2002R

Consistent with previous findings (i.e., Babchishin, Blais, et al., 2012), Indigenous peoples with a history of sexual crimes had higher total scores on both Static-99R and Static-2002R than Whites. When examining discrimination, Static-99R was able to discriminate recidivists from non-recidivists for both Indigenous and Whites. However, AUCs tended to be lower for Indigenous individuals (e.g., AUCs of .66 for Indigenous peoples versus .72 for Whites with a 10-year follow-up period). Results were not as positive for Static-2002R: total scores significantly discriminated recidivists from non-recidivists for Whites (e.g., AUC of .72 with a 10-year follow-up

period); however, neither of the AUCs for the 5-year or the 10-year follow-up period was significant with Indigenous individuals (AUCs = .56 and .61, respectively).

Regarding calibration, there were no significant differences in adjusted base rates (at a score of 2) for Indigenous individuals compared to the high-risk/high-needs norms for Static-99R. In addition, although the overall rates across the Static-99R scores were lower than expected rates (from the norms), the estimated recidivism probability from the norms corresponded well with the observed recidivism probability of the current sample. In contrast, the estimated recidivism probability from the norms of Static-2002R did not correspond with the observed recidivism probability of the current sample. For Whites, the adjusted 5-year base rates of sexual recidivism were significantly lower than the high-risk/high-needs norms for Static-99R and Static-2002R. This finding also extended to the 10-year follow-up period for Static-99R. This indicates that the high-risk/high-needs norms may over-estimate risk for Whites.

Differences in discrimination between Static-99R and Static-2002R for Indigenous peoples are somewhat surprising given that the two scales contain many of the same items and are highly correlated (r s ranging from .70 to .92; Babchishin, Hanson, & Helmus, 2012). It is possible that these findings reflect chance variation because the sample size for the Static-2002R studies was small (e.g., the 10 year estimates were based on only two samples with a total of 101 Indigenous individuals, of which 26 were sexual recidivists). Babchishin, Blais, et al.'s (2012) similar finding of relatively low discrimination for Static-2002R compared to Static-99R cannot be interpreted as an independent replication because of the substantial overlap in the samples included in both meta-analyses. Nevertheless, the current results support extreme caution when using Static-2002R with individuals of Indigenous heritage.

Psychological Constructs

It is possible for a scale to significantly predict a desired outcome between two different cultural groups (i.e., measurement equivalence) while exhibiting cultural bias in the identified underlying constructs of that same scale (i.e., structural non-equivalence; Byrne et al., 2009). Across both Indigenous and White individuals with a history of sexual crimes, we examined the presence and predictive validity of the underlying psychological constructs identified for Static-99R and Static-2002R. Indigenous individuals had significantly fewer indicators of persistence/paraphilia (e.g., less prior sex offences, non-contact offences, male victims, and young victims), and significantly more indicators of youthful stranger aggression (e.g., younger at release, more non-sexual violence at index) and general criminality (e.g., community supervision violations, prior criminal involvement, prior non-sexual violence, etc.) compared to Whites. This pattern of results is consistent with the general research on risk assessment with Indigenous peoples (e.g., Babchishin, Blais et al., 2012; Helmus et al., 2012; Olver et al., 2018; Perley-Robertson et al., 2018; Shepherd et al., 2014).

Neither the persistence/paraphilia nor the youthful stranger aggression constructs significantly predicted sexual recidivism for Indigenous individuals. Given that these constructs did predict sexual recidivism for Whites, it is possible that the same items are measuring different constructs for individuals of different racial/ethnic backgrounds. In contrast, general criminality similarly predicted sexual recidivism for both groups. Several specific findings across the constructs are worth highlighting.

When examining individual items of the persistence/paraphilia construct, the difference in discrimination could be attributed to two items: a) non-contact sex offences and b) any male victim. For White individuals, those items are well-established indicators of general sexual deviancy and paraphilic interests (e.g.,

pedophilia, exhibitionism, voyeurism). Further research is required to determine whether these items measure similar constructs for Indigenous peoples. Although we believe that the construct of deviant sexual interests applies to individuals of Indigenous heritage, Static-99R and Static-2002R might not include items that specifically address how sexual deviancy is expressed among Indigenous peoples.

In particular, it was surprising that a history of non-contact sex offences predicted less, not more, sexual recidivism for Indigenous individuals. A history of noncontact sexual offending is strongly related to increased sexual recidivism among predominately White samples (Hanson & Bussière, 1998; Helmus & Thornton, 2015). Although all explanations are highly speculative, it may be that exhibitionism, which is the strongest driver of the increased risk for noncontact offending among Whites, could be a culturally specific disorder. For Whites, exhibitionism is often a highly repetitive and persistent disorder. It could also be a disorder of White males (like serial killing). Noncontact offending was rare among the Indigenous individuals in the current study, and it is similarly rare for African Americans (Lee, Hanson, Calkins, & Jeglic, 2018, Chapter 4/Study 2).

Within the youthful stranger aggression construct, the lack of relationship between age and sexual recidivism for Indigenous peoples was unexpected. Among White males, there is strong empirical support for declining sexual recidivism risk with advanced age (Barbaree & Blanchard, 2008; Hanson, 2002), supported by credible theory, including decreased sexual drive, predictable increases in sexual self-regulation, and reductions in opportunity (Hanson, 2002). Some cultural values of Indigenous society might suppress the effect of age on sexual recidivism. For example, there might be unique cultural values that promote sexual self-regulation and suppress sexually deviant interests from a young age. As well, sexual offending in

middle age may mean something different from youthful offending for Indigenous peoples compared to Whites. In particular, sexual offending in the 40s and 50s may be a sign of persistent problems with sexual self-regulation for Indigenous individuals.

Overall, the differences in the presence and predictive validity of risk relevant characteristics between Indigenous and White peoples may be indicative of different trajectories or pathways that lead to sexual offending. Recently, Brouillette-Alarie and Proulx (2018) developed an etiological model that included two broad paths to sexual offending, one characterized by social isolation and sexual deviance (e.g., sexual interest in children) and the other characterized by antisociality (e.g., impulsivity), physical and psychological victimization and sexual promiscuity. Whereas the first path was related to sexual recidivism, especially crimes against children, the second was more related to non-sexual violence and sexual offending against adult women. Although the characteristics of the Indigenous individuals in the current study more closely resembled the second, antisocial path than the first, they did not clearly fit that pattern either. It is possible that there are distinctive pathways to sexual offending for Indigenous people that are not well represented by any of our current etiological models. Identifying cultural differences in the etiology of sexual offending could, therefore, add tremendously to our knowledge of Indigenous individuals with a history of sexual crimes.

Cultural bias in Static-99R and Static-2002R for Indigenous Canadians

There were three main findings in this study: 1) Static-99R showed acceptable discrimination and calibration for both Whites and Indigenous peoples, 2) Static-2002R poorly discriminated recidivists from non-recidivists for Indigenous individuals and showed overall poor calibration, and 3) risk-relevant characteristics that predict sexual recidivism for Whites occur at different levels and predict

differently for Indigenous peoples compared to Whites. These findings, however, raise another central question about how much similarities are sufficient to justify the use of risk assessment tools for any different racial or ethnic group, including the use of the STATIC risk tools with Indigenous peoples. To respond to the Supreme Court's decision in *Ewert*, we need to know what evidence would be sufficient to support the use of the tools for Indigenous peoples. Conversely, how different do the results need to be before evaluators should not use a risk tool developed on Whites with a specific racial/ethnic minority?

The scientific and professional communities do not currently have clear conventions or standards to evaluate the extent to which observed cultural variation in assessment tools is tolerable. One approach to setting a limit for cultural bias is to estimate the likelihood that the observed differences are primarily due to chance variation. From this perspective, the a priori expectation is that there are no real cultural differences and that any observed differences are the results of sampling (random or biased) from the same population. For many individuals concerned with Indigenous overrepresentation, the null hypothesis is a very questionable assumption. Instead, some degree of racial differences would be expected, which, from a Bayesian statistical perspective, would lower the amount of evidence required to conclude that real differences were present. The primary problem with relying solely on statistically significant differences, however, is that it is possible to detect differences that are too small to matter. With large samples, differences of a couple of percentage points could be statistically significant but result in the same substantive conclusions.

Another approach to evaluating cultural bias is whether the evidence supports the same inferences for the same test score. From this perspective, the primary concern is the consistency of the decisions that would follow from the test scores. If

there are real cultural differences in test performance, these differences are tolerable if the decision maker would nonetheless be justified in making the same decision regardless of the race/ethnicity of the case at hand. For example, parole would be equally denied to a White inmate whose risk was estimated at 30% and to an Indigenous inmate whose risk was estimated at 20% if the acceptable risk threshold was 5 recidivists out of 100 parole releases. In this example, the same score was associated with meaningfully different recidivism rates for Whites and Indigenous, but despite racial differences, the test results would justify the same decision. In this case, the test supports the valid inference that the individual's risk is greater than 5% for both these racial group. In contrast, the cultural bias in test result would be problematic if the decision threshold was 25 out of 100.

The above example suggests that it would be difficult to have a single standard for tolerable cultural bias: the amount and nature of tolerable bias would depend on the decisions that are required. Consequently, evaluators need to consider carefully the extent to which the specific inferences they make from a test score are justified based on the available evidence, and whether any observed racial/cultural differences would influence the decision-at-hand.

For the Static-99R and Static-2002R, there are several different types of information associated with test scores by the tools' developers (Hanson et al., 2017; Phenix et al., 2016). These meanings have been organized into standardized risk levels (i.e., Level I, II, III, IVa, and IVb), which are intended to convey quantitative information (percentile ranks, risk ratios, recidivism rates) as well as psychological characteristics (e.g., severity of criminogenic needs, developmental history, treatment needs). The current study focused only on certain quantitative information provided by these risk tools (i.e., risk ratios, recidivism rates). Further research is needed to

determine the extent to which other risk relevant characteristics (e.g., sexual self-regulation problems, response to treatment) have the same associations with STATIC scores for Indigenous and White individuals.

Advancing our understanding of cultural bias in risk assessment tools requires examining the meanings of constructs (i.e., the inferences of risk levels) across different racial/ethnic groups. This study focused on the ability of these risk tools to identify individuals as high risk for sexual recidivism, given its direct relevance for the Ewert case. Overall, the relative risk (odds ratios) and absolute sexual recidivism rates of Static-99R for Indigenous peoples were very similar to the norms asserted by the test developers (Phenix et al., 2016). Consequently, evaluators can support inferences about sexual recidivism rates based on Static-99R risk levels for both Indigenous and White peoples. In contrast, the findings of the current study did not support the Static-2002R relative risk estimates nor the absolute recidivism rate estimates for Indigenous peoples, for unknown reasons.

Although this study was not intended to directly examine the risk relevant psychological constructs associated with the standardized risk levels, there were enough significantly different patterns on risk-relevant characteristics between Indigenous and White peoples to suggest that cultural differences were likely. In particular, Whites showed higher levels of deviant sexual interests whereas Indigenous showed higher levels of general criminality. Consequently, future research should examine the extent to which individuals assessed as an above average risk by the STATIC risk tools may present with different psychological profiles depending on their racial/cultural background.

Limitations

Despite several strengths of the current study (e.g., largest sample to date;

consideration of discrimination and calibration), there are several aspects that limit confidence in the conclusions. First, it is important to note that Indigenous peoples in Canada are not a homogenous group. They comprise the First Nations (58%; the original inhabitants in North America), Métis (35%; mixed European and Indigenous ancestry), and Inuit people (4%; live primarily in the northern regions of Canada; 4%; Statistics Canada, 2018). Each subgroup has its own language and culture. Consequently, it is difficult to determine the extent to which the results generalize to each subgroup.

Furthermore, individuals living off-reserve might show more similar risk relevant characteristics with Whites than those living on reserve (e.g., 66% of the First Nations and almost all Métis and Inuit are living off reserve; Statistics Canada, 2018). Those variables were, however, not available in this study. In addition, all individuals with a history of sexual crimes in this study were already identified as high-risk (e.g., from federal prisons, National Flagging System, sex offender treatment programs) and the findings may not generalize to lower risk Indigenous and White individuals (e.g., individuals serving sentences in the community).

Finally, despite having larger samples of Indigenous individuals than previous STATIC studies, the numbers were still well below the sample size required for stable logistic regression estimates (i.e., 100 recidivists and 100 non-recidivists; Vergouse, Steyerberg, Eijkemans, & Habbema, 2005).

Implications for Research

The current results reinforce the importance of considering both discrimination (i.e., relative risk) and calibration (i.e., absolute risk) when evaluating risk tools. For example, whereas the Static-99R showed lower discrimination for Indigenous individuals than Whites, calibration for Indigenous peoples seemed to be

similar with or better than Whites (e.g., the high-risk/high needs norms did not significantly differ from the adjusted recidivism rate of Indigenous individuals but did for Whites).

Second, there are several important variables that could not be examined in this study that would help in further understanding the differences in the predictive validity of risk assessment scales. As previously mentioned, we were unable to categorize the Indigenous samples into more specific types (i.e., First Nations, Métis, or Inuit) or to consider those who live on reserve versus those who live off-reserve. Bonta, LaPrairie, and Wallace-Capretta (1997) found significant differences between Indigenous peoples living on and off reserve (e.g., off-reserve had higher risk scores), highlighting the importance of considering this variable. Future research should, therefore, try to gain a higher level of specificity when defining the Indigenous group as this would provide a more nuanced appraisal of differences in the predictive validity of risk assessment scales.

Third, no study to date has examined whether the inclusion of culturally specific variables (e.g., loss of native language; effect of residential schools; Mann, 2009) would add incremental validity to the prediction of recidivism outcomes for Indigenous peoples. Relatedly, the inclusion of such factors could potentially help to explain differences in the underlying psychological constructs that have been identified for the Static-99R and Static-2002R. More specifically, there may be constructs that are distinctive to Indigenous individuals in the criminal justice system, and the exclusion of items tapping into these constructs may help explain the relatively poor predictive accuracy of certain risk tools for Indigenous individuals (Gutierrez et al., 2016). As well, there may be patterns of sexual offending that are distinctive to Whites, such as exhibitionism and the possession of child abuse images

(both coded as noncontact sexual offences in the STATIC risk tools).

Implications for Policy and Practice

Static-99R was one of the impugned measures identified in *Ewert v. Canada* (2015, 2018). Consistent with previous findings (e.g., Babchishin, Blais et al., 2012), the current study supports the use of Static-99R for Indigenous peoples to identify individuals at high risk for sexual recidivism (Risk-principle). Specifically, Static-99R was able to discriminate recidivists from non-recidivists for Indigenous peoples at acceptable levels of accuracy. Although the discrimination was lower for Indigenous than those of Whites, the differences are small and, even considering the differences, the Static-99R scores support similar inferences for both groups.

Regarding calibration, the base rates of Static-99R for Indigenous individuals were very similar to the high-risk/high-needs norms asserted by the test developers (Phenix et al., 2016). In contrast, Static-2002R had significantly poorer predictive accuracy for Indigenous peoples (discrimination and calibration) than for either the norms or the Whites in the current study. Consequently, our findings do not support the use of Static-2002R for assessing the likelihood of sexual recidivism for Indigenous individuals with a history of sexual crimes.

Given that Indigenous individuals with a history of sexual crimes scored significantly higher on indices of general criminality (e.g., criminal history, non-sexual violence) and significantly lower on indices of sexual criminality (e.g., paraphilia), the treatment for Indigenous peoples who commit sexual crimes may benefit from an increased focus on general criminality (e.g., antisocial attitudes and behaviours). In addition, treatment programming will likely be most effective when it takes into consideration the cultural values or norms of Indigenous peoples (e.g., spirituality) as well as socio-demographic characteristics that influence their response

to treatment (e.g., low education and socioeconomic status, systemic oppression, distrust of criminal justice system; Responsivity-principle; Gutierrez, Chadwick, & Wanamaker, 2018).

Conclusion

In *Ewert v. Canada* (2018), the Supreme Court of Canada ruled that CSC's use of certain recidivism risk tools with Indigenous individuals violated s. 24(1) of the *Corrections and Conditional Release Act*. Specifically, this section requires that the information CSC uses in making decisions about an offender is as accurate and as up to date as possible. The court did not determine that the use of the expunged tools actually caused harm. Instead, it concluded that CSC had not done its due diligence in verifying the validity for Indigenous peoples of certain risk tools commonly used by the service.

Due to the rules of evidence, the Supreme Court justices hearing the Ewert case could only consider the scientific research deemed credible by Judge Phelan in the original, 2015 Federal Court hearing. The only expert relied on in that case, Dr. Stephen Hart, asserted in his 2012 report that the validity of the risk tools for Indigenous peoples was unknown because of insufficient research on the topic. Consequently, Chief Justice McLachlin and colleagues were unaware of any subsequent research, including the 2012 meta-analysis of Babchishin and colleagues on the predictive accuracy of the Static-99R and Static-2002R for Indigenous peoples.

Would the courts have made a different decision if they had access to the results of the current study? The answer is far from certain. As previously stated, we lack clear criteria by which to determine whether any observed amount of cultural bias is tolerable. With regards to the Ewert case, we side with the minority opinion of Judges Rowe and Côté who argued that it is not the place of the courts to determine

which risk tools psychologists should use. Decisions concerning the scientific credentials of psychological tests or criminological assessment tools are rightly made by professional communities of practice.

Based on the accumulated evidence and the current results, we believe that there is sufficient research to justify the use of Static-99R, but not Static-2002R, for assessing the likelihood of sexual recidivism among individuals with Indigenous heritage. Our position accepts that some cultural bias in Static-99R is likely, but interprets the available evidence as indicating that Static-99R scores will most often support the same general inferences concerning relative risk and absolute sexual recidivism rates regardless of whether the individual being assessed is White or Indigenous. Nevertheless, given that risk assessment tools are used at every point of contact with the criminal justice system (e.g., policing, sentencing, rehabilitation programming, parole), continued efforts to validate risk assessment scales with Indigenous peoples remains a research priority.

Chapter 6: General Discussion

Since their inception in the early 1900s, cultural bias in psychological measures (e.g., intelligence testing; Binet & Simon, 1916) has been a concern (e.g., Eells, 1953; Fagan & Holland, 2007; Gutkin & Reynolds, 1981; West & Macarthur, 1964, Zeidner, 1988). The issue of cultural bias, particularly for IQ tests, has drawn considerable attention from the public, courts, and lawmakers (Brown, Reynolds, & Whitaker, 1999). Structured risk assessment instruments (actuarial and SPJ) share many of the features of psychological tests, i.e., they are designed to estimate the likelihood of recidivism based on a set of predetermined personal characteristics (i.e., criterion-referenced tests). Given that culture is one characteristic that influences the development of human behaviour (Fishbein & Ajzen, 1975; Fishbein & Ajzen, 2015), individuals from different cultural backgrounds (e.g., White, Hispanic, Black, or Indigenous peoples) might have different indicators for determining or predicting human behaviour.

Psychometric approaches to examining cultural bias in psychological measures were introduced into the study of psychology during the 1960s (e.g., Jensen, 1980, Reynolds, 1995, 1998, 2000b; Reynolds & Brown, 1984). Nevertheless, relatively little attention has been paid to cultural bias in structured risk assessment instruments compared to other types of psychological measures (e.g., aptitude, personality, intelligence test). The purpose of this dissertation was to evaluate cultural bias in structured risk assessment instruments with different ethnic groups in Canada and the U.S. In particular, the predictive validity of the Static-99R and Static-2002R was assessed for the most overrepresented ethnic populations in the Canadian and U.S. criminal justice systems (i.e., Black, Hispanic, and Indigenous peoples). Furthermore, this dissertation examined whether there were different patterns on risk-

relevant characteristics of individuals with a history of sexual crimes between different racial/ethnic groups (e.g., general criminality and sexual criminality).

How Much Evidence is Sufficient to Detect Cultural Bias?

Given the ongoing debates about cultural bias in psychological measures, it is difficult to arrive at a common definition of cultural bias. Laypeople tend to believe that any differences in the mean level of total scores between racial or ethnic groups demonstrate cultural bias in the measures. It is, however, a fallacious inference (*Egalitarian fallacy*; Jensen, 1980). If the different scores reflect the true differences between racial or ethnic groups, the measures are not culturally biased. Traditional statistical tests have limitations for assessing bias because the scientific and professional community do not agree about the a priori expectation. For some, the a priori expectation (the null hypothesis) is that there are no ethnic/racial differences whereas other experts expect differences to exist (Reynolds, 2000a). Consequently, simply examining statistically significant differences in mean scores between racial or ethnic groups do not fully address the issue of cultural bias (Thorndike, 1971).

Consistent with previous research (Adebimpe, Gigandet, & Harris, 1979; Hilliard, 1979; Padilla, 1988), the overall findings of the current dissertation found mean differences in Static-99R and Static-2002R scores across ethnic groups; however, such differences did not automatically imply cultural bias in predictive accuracy of risk assessment instruments. For example, Study 2 (New Jersey) found that Static-99R scores of Blacks were significantly higher than scores of Whites; however, discrimination and calibration were comparable. In contrast, the Hispanic and White individuals in this study had similar mean scores, but this did not translate into a similar level of predictive accuracy for the two groups (i.e., Hispanics demonstrated consistently lower discrimination and calibration than Whites).

A significant concern is systematic underestimation or overestimation of recidivism risk depending on racial or ethnic groups. For example, the same score of a risk assessment instrument could be associated with lower observed recidivism rates for Blacks than for Whites (overestimation of recidivism risk for Blacks). In this dissertation, a psychometric definition of cultural bias was adopted to estimate the probability that the observed differences between different ethnic groups are primarily due to chance (i.e., statistical bias; p -value). According to Cleary, Humphreys, Kendrick, and Wesman (1975),

A test is considered culturally biased in the predictive validity if the inference drawn from the test score is not made with the smallest feasible random error and if there is a constant error in the inference as a function of membership in a particular racial/ethnic group (Reynolds, 2000a, p. 273).

Multiple statistical tests were, therefore, used to evaluate cultural bias in predictive accuracy (discrimination and calibration) of Static-99R and Static-2002R. For example, whenever the slope (relative risk) or the intercept (absolute base rate) of logistic regressions differed significantly across ethnic groups, this was taken as evidence of cultural bias in prediction. Consequently, cultural bias was determined solely by statistically significant differences (i.e., p -value < .05; typical convention in social sciences). In other words, if the p -value is smaller than .05 (events improbable by chance), the null hypothesis that any observed difference between ethnic groups was assumed to be explained by random error. Based on the accumulated statistical evidence of this dissertation, there is sufficient evidence to justify the use of the Static-99R for assessing the likelihood of sexual recidivism for ethnic minority groups (Blacks, Hispanics, and Indigenous peoples), who are overrepresented in the criminal justice systems in Canada and the U.S.

Nevertheless, there remains an important question whether statistical evidence provides sufficient evidence to justify using or not using risk assessment instruments for any different racial or ethnic groups. The primary problem with relying solely on statistically significant differences is that it is possible to detect differences that may not be practically meaningful. With large samples, minor differences between ethnic groups could be statistically significant (i.e., Type I error), but arrive at the same conclusions (i.e., the same risk level). In contrast, it is also possible that the true differences between ethnic groups are not detected due to the small sample size (i.e., Type II error; low statistical power). Consequently, evaluating cultural bias is a more comprehensive procedure than relying solely on statistical results. Evaluators or users of risk assessment instruments should take account of all available scientific evidence when justifying the inferences they make from a test score for specific racial/ethnic groups.

In the criminal justice system, offenders' risk is communicated using various methods: total score, risk category, percentile, risk ratios, or absolute risk. Forensic experts are highly likely to communicate risk assessment results (e.g., PCL-R, Static-99, VRAG) using categorical terms (e.g., low, moderate, or high risk; Blais & Forth, 2014). In particular, evaluators frequently report category labels of the Static-99R for SVP evaluation (Chevalier, Boccaccini, Murrie, & Varela, 2015). In addition, prospective jurors are substantially influenced by the category labels compared to other information format associated with Static-99R scores (e.g., risk ratios, absolute risk rates; Varela, Boccaccini, Cuervo, Murrie, & Clark, 2014). Validity is defined as the degree to which scientific evidence support the interpretations of a test score. Given the prevalent use and substantial influence of categorical risk levels, it is essential to examine the extent to which the categorical risk levels derived from risk

assessment instruments provide the same inferences regardless of racial or ethnic group (i.e., cultural bias in construct validity).

For the Static-99R and Static-2002R, the standardized risk categories (i.e., Level I, II, III, IVa, and IVb) have been adopted to provide quantitative information (percentile ranks, risk ratios, recidivism rates) as well as psychological characteristics (e.g., severity of criminogenic needs, developmental history, treatment needs) for each risk level (Hanson et al., 2017). Consequently, the use of standardized risk levels should increase the likelihood that the same inferences for individuals assigned to the same category regardless of racial or ethnic groups (i.e., cultural bias in construct validity). There are five risk levels for individuals with a history of sexual crime: Level I (very low risk), II (below average risk), III (average risk), IVa (above average risk), and IVb (well above average risk). For example, Level III identifies individuals in the middle of the risk distribution for sexual recidivism (i.e., average risk). The estimated sexual recidivism rates for individuals classified as Level III are expected to be between 5% to 10% within 5 years after release. Level IVa (above average risk) identifies individuals who are higher risk than the average risk groups. The expected recidivism rates are twice (10% to 20%) as high as for those in the Level III. Also, the offenders in this level are slightly younger than average and are more likely to be rapists than child molesters than individuals in the Level III (see Hanson et al., 2017 for more details).

Although the current study mainly focused on specific quantitative information provided by these risk tools (i.e., risk ratios, recidivism rates), there were also consistently different patterns on risk-relevant characteristics between ethnic minority groups (Blacks, Hispanics, Indigenous peoples) and White individuals. Psychologically meaningful constructs associated with sexual recidivism were

examined by considering the items of the Static-99R and Static-2002R (persistence/paraphilia, youthful stranger aggression, general criminality; Brouillette-Alarie et al., 2016). There are some benefits to consider the psychological constructs in the evaluation of cultural bias of structured risk assessment instruments. First, psychological constructs provide some insight as to why the instruments predict sexual recidivism risk better for certain ethnic groups than others. For example, the construct of paraphilic interests predicted sexual recidivism for Whites but not for Indigenous peoples. Second, psychological constructs provide answers as to why a certain ethnic group has a high average score on the instruments (e.g., high paraphilic interests, but low general criminality).

Further, understanding the distinct psychological characteristics of each ethnic group is useful to develop effective treatment strategies. In the dissertation studies, Whites showed higher levels of sexual deviancy than other ethnic groups, whereas Blacks and Indigenous peoples showed higher levels of general criminality than Whites. Consequently, there might be difference psychological characteristics of individuals with different cultural heritages in the same risk levels. Evaluators should be aware of the distinct psychological characteristics for each ethnic groups when considering intervention strategies.

Why Are There Low Average Scores and Low Sexual Recidivism Rates for Hispanics?

Hispanics are one of the overrepresented populations in the U.S. criminal justice system. In addition, Hispanics have experienced social strain and discrimination similar to Black Americans. Hispanics, however, had unexpectedly lower average Static-99R scores than Blacks (the average score for Hispanics was similar to that of White individuals). It is hypothesized that the unexpectedly low

scores on the Static-99R for Hispanics might be a product of missing information regarding criminal history due to their immigrant status. Several studies found that Static-99 scores of Hispanics born outside the U.S. were significantly lower than scores of those born in the U.S. (Leguizamo et al., 2017; Rice, Boccaccini, & Varela, 2013). Specifically, Hispanics born outside the U.S. (i.e., immigrants) were less likely than those born in the U.S. to have prior non-sexual violence convictions, prior sex offences, and prior sentencing dates (Leguizamo et al., 2017). Consequently, evaluators should be cautious when scoring risk assessment instruments for recent immigrants to maintain good predictive accuracy.

In addition, previous and current studies consistently found that the sexual recidivism rates of Hispanics are also unexpectedly lower than for Whites and Blacks (Boccaccini et al., 2017; Hanson et al., 2014; Leguizamo et al., 2017). Difficulties following their criminal history due to their immigrant status (e.g., deportees) might lead to relatively low sexual recidivism rates of Hispanics. Several studies found that Hispanics born in the U.S. had higher sexual recidivism rates than those born outside the U.S. (Boccaccini et al., 2017; Leguizamo et al., 2017). Further, Lee and his colleagues (2018) found that the low sexual recidivism rates could be explained by the high rates of deportation among the Hispanics. The original study (Hanson et al., 2014) found unexpectedly low sexual recidivism rates for Hispanics (i.e., poor calibration) compared to other ethnic groups (Whites or Blacks). After excluding the deportees from the calibration analysis, however, the sexual recidivism rates for Hispanics became comparable with the sexual recidivism rates for White and Black ethnic groups, as well as the norms of Static-99R (Lee et al., 2018). Consequently, undetected recidivism rates of recent immigrants (i.e., deportees) should be carefully considered when estimating their true recidivism rates.

Implications for Future Research

The findings of the current dissertation suggest that future research on cultural bias could improve the predictive accuracy of a risk assessment instrument for ethnic minority groups. First, researchers should empirically explore the association between identified risk factors and sexual recidivism risk for each ethnic group (item level; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). It is important to note that empirically supported risk factors for certain ethnic groups might not be associated with sexual recidivism for another ethnic group (e.g., non-contact sexual offences for Indigenous peoples).

Another approach to detecting the cultural bias at the item level is differential item functioning (DIF) derived from item response theory (IRT). IRT is a statistical method to examine a probability of occurrence of a particular response to a measurement item (binary, ordinal, or continuous items) as a function of the offender's relative position on the psychological latent constructs (e.g., paraphilia). IRT (by the item characteristic curve [ICC]) consists of two or three parameters: α - discrimination power, b - difficulty level, and, if forced choice, c - a guessing parameter or asymptote. Differential item function (DIF) refers to the degree to which the item measures different abilities for members of different ethnic groups. Technically, DIF represents the area between two ICCs of different ethnic groups. If the DIF index is significant, further analysis of this item with significant DIFs are needed to determine if they are biased. For example, the item 'male victim' can discriminate high paraphilic interests from low paraphilic interests for White, but probably not for Indigenous peoples (α parameter; slope of the ICC). The interpretation of the significant DIF is that the 'male victim' item does not appropriately measure the latent construct of paraphilic interests for Indigenous peoples.

Further, research should examine the extent to which two psychological constructs (general and sexual criminality) are shared across different ethnic groups. Factor analysis is one of the widely used empirical approaches to investigating construct validity (Cronbach, 1990). Briefly, factor analysis identifies groups of psychological indicators (i.e., items) that are highly correlated with one another. When the same psychological constructs are identified by factor analysis across ethnic groups, it is strong evidence to support that the instrument (or subsets) measures the same construct for the different ethnic group. In particular, it is worth examining the extent to which the two major propensities (general and sexual criminality) are replicated among individuals with different cultural backgrounds, and whether they are similarly associated with sexual and general recidivism.

Further research is also needed to examine patterns of general criminal behaviour and sexual crime specific criminal behaviour depending on the standardized risk levels. If there are distinct characteristics on individuals with a different cultural background in the same risk level, taking account of these differences could improve the effectiveness of treatment strategies (Need principle; Andrews et al., 1990; Andrews & Bonta, 2010). In addition, researchers should empirically explore potential race and culture-specific risk or protective factors for ethnic minority groups (e.g., loss of face is a protective factor against sexual aggression in the Asian Americans; Hall, Teten, DeGarmo, Sue, & Stephens, 2005).

More research is also necessary to identify the extraneous factors (e.g., immigrant status, underreporting) that could degrade the predictive accuracy of risk assessment instruments. For example, undetected sexual recidivism rates of immigrants (e.g., deportees) should be considered when estimating absolute recidivism rates. In addition, it is also important to empirically examine the degree to

which missing information (prior criminal history) of recent immigrants (e.g., Mexicans) influence the information value of risk assessment instruments based primarily on static, historical factors. When reliable criminal histories are not available, evaluators may need to focus instead on empirically validated risk factors related to recent psychological and community functioning.

Racial groups in this dissertation are not homogeneous, including multiple ethnic groups with distinct cultures (e.g., Blacks - Jamaicans, West Indians, Somalis, or Nigerians; Hispanics - Mexico, Ecuador, or Puerto Rico; Indigenous peoples - First Nations, Métis, or Inuit). Future research should, therefore, attempt to gain more specific information about the ethnic background of individuals with a history of sexual crimes. Given the association between acculturation level and criminal behaviour (Lee & Hahm, 2010; Meston & Ahrold, 2010; Ventura Miller et al., 2011; Wong, 1999), acculturation level of immigrants should be considered to help to understand the patterns of risk-relevant characteristics of ethnic minority individuals with a history of crimes.

Conclusions

The ethnic minority groups (particularly for Blacks and Indigenous individuals) with a history of sexual crimes showed the significantly higher level of general criminality (e.g., antisocial behaviour, non-sexual violent offences, impulsivity) and significantly lower level of sexual criminality (e.g., exhibitionism/voyeurism, pedophilia). Consequently, treatment for ethnic minority groups who commit sexual crimes may benefit from an increased focus on general criminality (e.g., antisocial attitudes and behaviours) whereas Whites may benefit from an increased emphasis on sexual criminality.

The results of the studies in this dissertation found that the discrimination and

calibration for Static-99R and Static-2002R were not fully consistent across racial/ethnic groups, i.e., some evidence of ethnic bias. For Static-2002R, the differences between Indigenous and Whites were big enough to make a difference; consequently, considerable caution is needed when interpreting the results of Static-2002R for individuals of Indigenous heritage. For Static-99R, on the other hand, the differences between ethnic and racial groups were small. The overall evidence would support the same general inferences implied by the standardized risk levels (e.g., relative risk and absolute sexual recidivism rates) regardless of ethnic groups (White, Blacks, Hispanics, and Indigenous peoples). Consequently, there is sufficient evidence to justify the use of the Static-99R for assessing the likelihood of sexual recidivism for ethnic minority groups (Blacks, Hispanics, and Indigenous peoples), who are overrepresented in the criminal justice systems in Canada and the U.S. (i.e., quantitative information).

References

- Adebimpe, V. R., Gigandet, J., & Harris, E. (1979). MMPI diagnosis of black psychiatric patients. *The American Journal of Psychiatry*, *136*, 85-87.
doi:10.1176/ajp.136.1.85
- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction stefanía ægisdóttir. *The Counseling Psychologist*, *34*, 341-382.
doi:10.1177/0011000005285875
- Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, *30*, 47-87. doi:10.1111/j.1745-9125.1992.tb01093.x
- Agnew, R. (2006). General strain theory: Current status and directions for further research. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory, advances in criminological theory* (pp.101–123). New Brunswick, N.J.: Transaction Publishers.
- Ahrold, T. K., & Meston, C. M. (2010). Ethnic differences in sexual attitudes of US college students: Gender, acculturation, and religiosity factors. *Archives of sexual behavior*, *39*, 190-202. doi:10.1007/s10508-008-9406-1
- Allen, J. D., Caspi, C., Yang, M., Leyva, B., Stoddard, A. M., Tamers, S., . . . Sorensen, G. C. (2014). Pathways between acculturation and health behaviors among residents of low-income housing: The mediating role of social and contextual factors. *Social Science & Medicine*, *123*, 26-36.
doi:10.1016/j.socscimed.2014.10.034

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: AERA Publications.

Anderson, E. (1999). *Code of the street: Decency, violence, and the moral life of the inner city*. New York: Norton.

Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct* (1st ed.). New York, NY: Routledge.

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory – Revised*. Toronto: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16, 39-55.
doi:10.1037/a0018362

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

Andrews, D. A., Bonta, J., Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI): User's Manual*. Toronto, ON: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52, 7-27.
doi:10.1177/0011128705281756

- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, *50*, 88-100.
doi:10.1177/0306624X05282556
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, *28*, 369-404.
doi:10.1111/j.1745-9125.1990.tb01330.x
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias: There's software used across the country to predict future criminals. And it's biased against Blacks. *ProPublica*. Retrieved from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001). *The comparative costs and benefits of programs to reduce crime*. Olympia, WA: Washington State Institute for Public Policy.
- Arbona, C., & Power, T. G. (2003). Parental attachment, self-esteem, and antisocial behaviors among African American, European American, and Mexican American adolescents. *Journal of Counseling Psychology*, *50*, 40.
doi:10.1037/0022-0167.50.1.40
- Archer, R. P., Buffington-Vollum, J., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*, 84-94. doi:10.1207/s15327752jpa8701_07

- Atkinson, R. L., Kropp, P. R., Laws, D. R., & Hart, S. D. (1996). *Scoring manual for the Sexual Offender Risk Assessment (SORA) guide*. Vancouver, BC: British Columbia Institute Against Family Violence & the Mental Health, Law, and Policy Institute.
- Axford, M. (2008). *Homicide, sex, robbery and drug offenders in Federal Corrections: An end-of-2008 review* (Report NO. 2011 N° R-234). Retrieved from Correctional Service of Canada website: <http://www.csc-scc.gc.ca/005/008/092/005008-0234-01-eng.pdf>
- Babchishin, K. M., Blais, J., & Helmus, L. (2012). Do static risk factors predict differently for Aboriginal sex offenders? A multi-site comparison using the original and revised Static-99 and Static-2002 scales. *Canadian Journal of Criminology & Criminal Justice*, *54*, 1-43. doi:10.1037/t04644-000
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, *19*, 442–461. doi:10.1177/1073191112458312
- Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review*, *23*, 1023-1053. doi:10.1016/j.cpr.2002.07.001
- Bales, W.D., & Piquero, A.R. (2012) Racial/ethnic differentials in sentencing to incarceration, *Justice Quarterly*, *29*, 742-773.
doi:10.1080/07418825.2012.659674
- Barbaree, H. E., & Blanchard, R. (2008). Sexual deviance over the lifespan: Reduction in deviant sexual behavior in the aging sex offender. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (pp. 37-60). New York, NY: Guilford.

- Barnes, C. W., & Kingsnorth, R. (1996). Race, drug, and criminal sentencing: hidden effects of the criminal law. *Journal of Criminal Justice*, *24*, 39-55.
doi:10.1016/0047-2352(95)00051-8
- Bernal, G., & Enchautegui-de-Jesús, N. (1994). Latinos and Latinas in community psychology: A review of the literature. *American Journal of Community Psychology*, *22*, 531-557. doi:10.1007/BF02506892
- Bernal, G., & Shapiro, E. (2005). Cuban families. In M. McGoldrick, J. Giordano, & N. Garcia-Preto (Eds.), *Ethnicity and family therapy* (3rd ed., pp. 202-215). New York, NY: Guilford.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology*, *46*, 5-34. doi:10.1111/j.1464-0597.1997.tb01087.x
- Berry, J. W., & Sabatier, C. (2010). Acculturation, discrimination, and adaptation among second generation immigrant youth in Montreal and Paris. *International Journal of Intercultural Relations*, *34*, 191-207.
doi:10.1016/j.ijintrel.2009.11.007
- Binet, A., & Simon, T. (1916). *The development of intelligence in children*. Baltimore, MD: Williams & Wilkins Co.
- Blais, J., & Bonta, J. (2015). Tracking and Managing High Risk Offenders: A Canadian Initiative. *Law & Human Behavior (American Psychological Association)*, *39*, 253-265. doi:10.1037/lhb0000109
- Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior*, *38*, 531–543. doi:10.1037/lhb0000082
- Blalock, H. M. (1967). *Toward a theory of minority group relations*. New York, NY: Wiley.

- Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual offenders. *Psychological Assessment, 29*, 611-623.
doi:10.1037/pas0000377
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional guidelines for assessing risk of sexual violence*. Vancouver, BC: British Columbia Institute Against Family Violence.
- Bonta, J., Harris, A., Zinger, I., & Carriere, D. (1996). *The Crown files research project: A study of dangerous offenders* (Report No. 1996-01). Ottawa, Canada: Ministry of the Solicitor General of Canada.
- Bonta, J., LaPrairie, C., & Wallace-Capretta, S. (1997). Risk prediction and re-offending: Aboriginal and non-aboriginal offenders. *Canadian Journal of Criminology, 39*, 127-144.
- Bonta, J., & Yessine, A. K. (2005). [Recidivism data for 155 released sexual offenders from the offenders identified in *The National Flagging System: Identifying and responding to high risk, violent offenders* (User Report 2005-04). Ottawa, ON: Public Safety and Emergency Preparedness Canada.] Unpublished raw data.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, U.K.: John Wiley & Sons, Ltd.
- Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY)*. Consultation version. Tampa: Florida Mental Health Institute, University of South Florida.
- Bourgon, G., Mugford, R., Hanson, R. K., & Coligado, M. (2018). Offender risk assessment practices vary across Canada. *Canadian Journal of Criminology & Criminal Justice, 60*, 167-205. doi:10.3138/cjccj.2016-0024

- Boyce, J. (2016). *Victimization of Aboriginal people in Canada, 2014*. (Catalogue no. 85-002-X). Retrieved from the Statistics Canada website:
<https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2016001/article/14631-eng.pdf?st=4WKR8AE>
- Boyd-Franklin, N., & Lockwood, T. W. (2009). Spirituality and religion: Implications for psychotherapy with African American families. In F. Walsh (Ed.), *2nd ed.; spiritual resources in family therapy (2nd ed.)* (2nd ed., pp. 141-155). Guilford Press, New York, NY.
- Brankley, A., Lee, S. C., Babchishin, K. M., Hanson, R. K., & Harris, A. J. R. (2017a). Updated recidivism information for 409 individuals from the Dynamic Predictors Project (1998). Unpublished raw data.
- Brankley, A. E., Lee, S. C., Hanson, R. K., & Zabarauckas, C. (2017b). *Not all ethnic groups are overrepresented in sex offender population*. Manuscript submitted for publication.
- Brezina, T., Agnew, R., Cullen, F. T., & Wright, J. P. (2004). The code of the street: A quantitative assessment of Elijah Anderson's subculture of violence thesis and its contribution to youth violence research. *Youth Violence and Juvenile Justice*, 2, 303-328. doi:10.1177/1541204004267780
- Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L. M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution. *Assessment*, 23, 96-111. doi:10.1177/1073191114568114
- Brouillette-Alarie, S., & Proulx, J. (2018). The etiology of risk in sexual offenders: A preliminary model. *Sexual Abuse*. Advance online publication. doi:10.1177/1079063218759325

Brouillette-Alarie, S., Proulx, J., & Hanson, R. K. (2017). Three central dimensions of sexual recidivism risk: Understanding the latent constructs of Static-99R and Static-2002R. *Sexual Abuse*. Advance online publication.

doi:10.1177/1079063217691965

Browder v. Gayle, 142 F. Supp. 707 (1956)

Brown v. Board of Education of Topeka, 347 U.S. 483 (1954)

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since bias in mental testing. *School Psychology Quarterly*, *14*, 208-238.

doi:10.1037/h0089007

Bureau of Justice Statistics. (1987). *Prisoners in 1986* (Report NO. NCJ 104864).

Retrieved from Bureau of Justice Statistics website:

<http://www.bjs.gov/content/pub/pdf/p86.pdf>

Burgess, E. (1928). Factors determining success or failure on parole. In A. A. Bruce (Eds.), *The workings of the indeterminate sentencing law and the parole system in Illinois* (pp. 205-248). Springfield, IL: Illinois State Board of Parole

Burton, D. L., & Ginsberg, D. (2012). An exploration of racial differences in deviant sexual interests among male adolescent sexual offenders. *Journal of Forensic Social Work*, *2*, 25-44. doi:10.1080/1936928X.2011.609766

Byrne, B. M., Oakland, T., Leong, F. T. L., van, d. V., Hambleton, R. K., Cheung, F. M., & Bartram, D. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, *3*, 94-105.

doi:10.1037/a0014516

California Department of Justice (2016, June). *Minutes of the June 15, 2016 meeting of California Sex Offender Management Board (CASOMB)*. Sacramento, CA.

Callis, R., & Kresin, M. (2018). Quarterly residential vacancies and homeownership, first quarter 2018 (Report No. CB18-57). Retrieved from the United States Census Bureau website:

<https://www.census.gov/housing/hvs/files/currenthvspress.pdf>

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36, 567-590.

doi:10.1177/0093854809333610

Canadian Charter of Rights and Freedoms, ss. 7 & 15, Part I of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), 1982, c 11.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada (2010). *Tri-council policy statement: Ethical conduct for research involving humans*. Ottawa, ON: Secretariat on Responsible Conduct of Research. Retrieved from http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf

Canadian Psychological Association. (2017). *Canadian code of ethics for psychologists* (4th ed.). Retrieved from

https://www.cpa.ca/docs/File/Ethics/CPA_Code_2017_4thEd.pdf

Carson, E. A. (2018). *Prisoners in 2016*. (Report No. NCJ 251149). Retrieved from the Bureau of Justice Statistics website:

<https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187>

- Catchpole, R. E. H., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior, 30*, 688-708.
doi:10.1177/0093854803256455
- Cernkovich, S. A., Giordano, P. C., & Rudolph, J. L. (2000). Race, crime, and the American dream. *Journal of Research in Crime & Delinquency, 37*, 131-170
- Chappin, S. R., & Brook, J. S. (2001). The influence of generational status and psychosocial variables on marijuana use among black and Puerto Rican adolescents. *Hispanic Journal of Behavioral Sciences, 23*, 22-36.
- Chen, E. Y. (2008). Impacts of "Three Strikes and You're Out" on crime trends in California and throughout the United States. *Journal of Contemporary Criminal Justice, 24*, 345-370. doi:10.1177/1043986208319456
- Chevalier, C. S., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior, 39*, 209-218.
doi:10.1037/lhb0000114
- Chun, S. (2007). The "Other Hispanics"-- What are their national origins? Estimating the Latino-origin populations in the United States. *Hispanic Journal of Behavioral Sciences, 29*, 133-155. doi:10.1177/0789986306297441
- Claussen-Schulz, A., Pearce, M. W., & Schopp, R. F. (2004). Dangerousness, risk assessment, and capital sentencing. *Psychology, Public Policy, and Law, 10*, 471-491. doi:10.1037/1076-8971.10.4.471
- Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30*, 15-41.
doi:10.1037/0003-066X.30.1.15

- Constitution Act, 1982, Schedule B to the Canada Act 1982 (UK), 1982, c 11,
- Cooper, C. L., Murphy, W. D., & Haynes, M. R. (1996). Characteristics of abused and nonabused adolescent sexual offenders. *Sexual Abuse: Journal of Research and Treatment, 8*, 105-119. doi:10.1007/BF02258004
- Cornell, S., & Hartmann, D. (2007). *Ethnicity and Race: Making Identities in a Changing World*. Thousand Oaks: Pine Forge Press.
- Corrections Act, R.S.N.W.T. 1988, c.C-22.
- Corrections and Conditional Release Act (S.C. 1992, c. 20; S.C 2012, c. 1). Retrieved from Justice Laws website: <http://laws-lois.justice.gc.ca/eng/acts/C-44.6/>
- Criminal Code, RSC (1985) c C-46
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: Harper & Row.
- Dahle, K. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry, 29*, 431-442. doi:10.1016/j.ijlp.2006.03.001
- Dakil, S. R., Cox, M., Lin, H., & Flores, G. (2011). Racial and ethnic disparities in physical abuse reporting and Child Protective Services interventions in the United States. *Journal of the National Medical Association, 103*, 926-931. doi:10.1016/S0027-9684(15)30449-1
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674. doi:10.1126/science.2648573

- Day, A., Tamatea, A. J., Casey, S., & Geia, L. (2018). Assessing violence risk with Aboriginal and Torres Strait Islander offenders: Considerations for forensic practice. *Psychiatry, Psychology and Law*, *25*, 452–464.
doi:10.1080/13218719.2018.1467804
- Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services*, *13*, 206-222. doi:10.1037/ser0000075
- de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, *28*, 235-251. doi:10.1023/B:LAHU.0000029137.41974.
- Ditton, P. M., & Wilson, D. J. (1999). *Truth in sentencing in state prisons* (Report no. NCJ 170032). Retrieved from Bureau of Justice Statistics website:
<http://bjs.gov/content/pub/pdf/tssp.pdf>
- Doob, A. N., & Webster, C. M. (2003). Sentence severity and crime: Accepting the null hypothesis. In M. Tonry (Ed.), *Crime and justice: A review of research* (pp. 143–195). Chicago, IL: University of Chicago Press.
- Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, *32*, 479-510. doi:10.1177/0093854805278411
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*, 449-467.
- Eells, K. (1953). Some implications for school practice of the Chicago studies of cultural bias in intelligence tests. *Harvard Educational Review*, *23*, 284-297.

- Ellerby, L. A., & MacPherson, P. (2002). *Exploring the profiles of Aboriginal sexual offenders: Contrasting Aboriginal and non-Aboriginal sexual offenders to determine unique client Characteristics and potential implications for sex offender assessment and treatment strategies* (research report no. R-122). Ottawa, ON: Correctional Service of Canada. Retrieved from: http://www.csc-scc.gc.ca/research/092/r122_e.pdf
- Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1998). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R)*. St. Paul, MN: Minnesota Department of Corrections.
- Esqueda, C. W. (1997). European American students' perceptions of crimes committed by five racial groups. *Journal of Applied Social Psychology, 27*, 1406-1420. doi:10.1111/j.1559-1816.1997.tb01605.x
- Ewert v. Canada, 2007 FC 13
- Ewert v. Canada, 2015 FC 1093
- Ewert v. Canada, 2018 SCC 30
- Fagan, J. F., & Holland, C. R. (2007). Racial equality in intelligence: Predictions from a theory of intelligence as processing. *Intelligence, 35*, 319-334. doi:10.1016/j.intell.2006.08.009
- Federal Bureau of Investigation. (2013). *Crime in the United States, 2013*. Retrieved from <https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/tables/table-43>
- Federal Bureau of Investigation. (2015). *Crime in the United States, 2015*. Retrieved from <https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-43>

- Fedoroff, J. P., & Pinkus, S. (1996). The genesis of pedophilia: testing the "abuse-to-abuser" hypothesis. *Journal of offender Rehabilitation, 23*, 85-101.
doi:10.1300/J076v23n03_06
- Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual: Revised 2014*. Unpublished manual. Ottawa, Ontario: Public Safety Canada
- Finkelhor, D. (1984). *Child sexual abuse: New theory and research*. New York, NY: Free Press.
- Finkelhor, D., Ji, K., Mikton, C., & Dunne, M. (2013). Explaining lower rates of sexual abuse in china. *Child Abuse & Neglect, 37*, 852-860.
doi:10.1016/j.chiabu.2013.07.006
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fishbein, M., & Ajzen, I. (2015). *Predicting and changing behavior*. New York, NY: Routledge.
- Fix, R. L., Falligant, J. M., Alexander, A. A., & Burkhart, B. R. (2017). Race and victim age matter: Sexual behaviors and experiences among confined African American and European American youth with sexual and non-sexual offenses. *Sexual Abuse*. Advance online publication. doi:10.1177/1079063217720926
- Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237-253). New York, NY: Russell Sage Foundation.

- Forbes, S. M. (2007). *Race differences in scores of actuarial measures of sex offender risk assessment* (Unpublished doctoral dissertation). University of Louisville, Louisville, KY.
- French, S. A., & Gendreau, P. (2006). Reducing prison misconducts: What works! *Criminal Justice and Behavior, 33*, 185-218. doi:10.1177/0093854805284406
- Frideres, J. S., & Gadacz, R. R. (2012). *Aboriginal peoples in Canada*. Toronto, ON: Pearson.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*, 1-13.
Retrieved from <http://pareonline.net/getvn.asp?v=17&n=3>
- Garb, H. N. (1997). Race bias, social class bias, and gender bias in clinical judgment. *Clinical Psychology: Science and Practice, 4*, 99-120. doi:10.1111/j.1468-2850.1997.tb00104.x
- Garrett, C. J. (1985). Effects of residential treatment on adjudicated delinquents: A meta - analysis. *Journal of Research in Crime & Delinquency, 22*, 287-308. doi:10.1177/0022427885022004002
- Geisinger, K. F., & McCormick, C. (2013). Testing and assessment in cross-cultural psychology. In J. R. Graham, J. A. Naglieri & I. B. Weiner (Eds.), *Handbook of psychology (Vol. 10). Assessment psychology (2nd ed.)* (pp. 114-139). Hoboken, N.J.: John Wiley & Sons Inc.
- Gendreau, P., & Ross, B. (1979). Effective correctional treatment: Bibliotherapy for cynics. *Crime & Delinquency, 25*, 463-489. doi:10.1177/001112877902500405

- Gendreau, P., & Ross, B. (1987). Revivification of rehabilitation: Evidence from the 1980s, *Justice Quarterly*, 4, 349-407. doi:10.1080/07418828700089411
- Gendreau, P., Smith, P., & French, S. A. (2006). The theory of effective correctional intervention: Empirical status of future directions. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory* (pp. 419-446). Piscataway, NJ: Transaction Publishers.
- Gies, S., Gainey, R., Cohen, M. I., Healy, E., Duplantier, D., Yeide, M., Bekelman, A., Bobnis, A., & Hopps, M. (2012). *Monitoring High-Risk Sex Offenders with GPS Technology: An Evaluation of the California Supervision Program, Final Report*. Retrieved from National Criminal Justice Reference Service website: <https://www.ncjrs.gov/pdffiles1/nij/grants/238481.pdf>
- Gobeil, R., Blanchette, K., & Stewart, L. (2016). A meta-analytic review of correctional interventions for women offenders: Gender-neutral versus gender-informed approaches. *Criminal Justice and Behavior*, 43, 301-322. doi:10.1177/0093854815621100
- Goldsmith, R. E., Hall, G. N., Garcia, C., Wheeler, J., & George, W. H. (2005). Cultural aspects of sexual aggression. In K. H., Barrett & W. H., George (Eds.), *Race, culture, psychology, and law* (pp. 403-418). London, U.K: Sage.
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Grills, C., & Longshore, D. (1996). Africentrism: psychometric analyses of a self-report measure. *Journal of Black Psychology*, 22, 86-106. doi:10.1177/00957984960221007

- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293-323. doi:10.1037/1076-8971.2.2.293
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30. doi:10.1037/1040-3590.12.1.19
- Grubin, D. (1997). *Sex offending against children: Understanding the risk*. Police Research Series paper 99. London Home Office.
- Guerino, P., Harrison, P. M., & Sabol, W. J. (2011). *Prisoners in 2010* (Report NO. NCJ 236096). Retrieved from Bureau of Justice Statistics website: <https://www.bjs.gov/content/pub/pdf/p10.pdf>
- Gutierrez, L., Chadwick, N., & Wanamaker, K. A. (2018). Culturally relevant programming versus the status quo: A meta-analytic review of the effectiveness of treatment of indigenous offenders. *Canadian Journal of Criminology and Criminal Justice*, *60*, 321-353. doi:10.3138/cjccj.2017-0020.r2
- Gutierrez, L., Helmus, L. M., & Hanson, R. K. (2016). What we know and don't know about risk assessment with offenders of indigenous heritage. *Journal of Threat Assessment and Management*, *3*, 97-106. doi:10.1037/tam0000064
- Gutierrez, L., Wilson, H. A., Ruge, T., & Bonta, J. (2013). The prediction of recidivism with aboriginal offenders: A theoretically informed meta-analysis. *Canadian Journal of Criminology and Criminal Justice*, *55*, 55-99. doi:10.3138/cjccj.2011.E.51

- Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for white and black children from the standardization sample. *Journal of Educational Psychology, 73*, 227-231. doi:10.1037/0022-0663.73.2.227
- Haag, A. M. (2005). Recidivism data from 191 offenders detained until their warrant expiry date. Unpublished raw data. From: Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism. *Dissertations Abstracts International, 66*(08), 4531B. (UMI No. NR05662).
- Haag, A. M., Boyes, A., Cheng, J., MacNeil, A., & Wirove, R. (2016). An introduction to the issues of cross-cultural assessment inspired by Ewert v. Canada. *Journal of Threat Assessment and Management, 3*, 65-75. doi:10.1037/tam0000067
- Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology, 63*, 802-809. doi:10.1037/0022-006X.63.5.802
- Hall, G. C. N. (2002). Culture-specific ecological models of Asian American violence. In G. C. N. Hall & S. Okazaki (Eds.), *Asian American psychology: the science of lives in context* (pp. 153–170). Washington: American Psychological Association. doi:10.1037/10473-006.
- Hall, G. C. N., & Hirschman, R. (1991). Toward a theory of sexual aggression: a quadripartite model. *Journal of Consulting and Clinical Psychology, 59*, 662-669. doi.org/10.1037/0022-006X.59.5.662
- Hall, G. C. N., Teten, A. L., DeGarmo, D. S., Sue, S., & Stephens, K. A. (2005). Ethnicity, culture, and sexual aggression: Risk and protective factors. *Journal of Consulting and Clinical Psychology, 73*, 830-840. doi:10.1037/0022-006X.73.5.830

- Hall, G. C. N., Windover, A. K., & Maramba, G. G. (1998). Sexual aggression among Asian Americans: risk and protective factors. *Cultural Diversity and Mental Health, 4*, 305–318. doi:10.1037/1099-9809.4.4.305.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (Corrections User Report No. 1997-04). Ottawa, Canada: Department of the Solicitor General of Canada. Retrieved from Public Safety Canada Website: <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/dvlpmnt-brf-ctrl/dvlpmnt-brf-ctrl-eng.pdf>
- Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence, 17*, 1046-1062. doi:10.1177/088626002236659
- Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 15(1)*, 15-17. Retrieved from <http://www.cpa.ca/cpsite/UserFiles/Documents/Criminal%20Justice/Crime%20Scene%202008-04.pdf>
- Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie Canadienne, 50*, 172-182. doi:10.1037/a0015726
- Hanson, R. K. (2017). Assessing the calibration of actuarial risk Scales: A primer on the E/O Index. *Criminal Justice and Behavior, 44*, 26-39. doi:10.1177/0093854816683956
- Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment, 29*, 582-597. doi:10.1037/pas0000371

- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36*, 865-891.
doi:10.1177/0093854809338545
- Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: Journal of Research and Treatment, 17*, 357-373. doi:10.1177/107906320501700402
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362. doi:10.1037/0022-006X.66.2.348
- Hanson, R. K., & Harris, A. J. (1998). *Dynamic Predictors of Sexual Recidivism*. (Corrections Research User Report No. 1998-01). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K., & Harris, A. J. (2000). Where should we intervene? Dynamic predictors of sex offense recidivism. *Criminal Justice & Behavior, 27*, 6-35.
doi:10.1177/0093854800027001002
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (User Report No. 2007-05). Ottawa, ON: Public Safety Canada.
Retrieved from <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf>
- Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *The International Journal of Forensic Mental Health, 11*, 9-23. doi:10.1080/14999013.2012.667511

- Hanson, R. K., Lunetta, A., Phenix, A., Neeley, J., & Epperson, D. (2014). The field validity of Static-99/R sex offender risk assessment tool in California. *Journal of Threat Assessment and Management, 1*, 102-117. doi:10.1037/tam0000014
- Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Corrections User Report No. 2004-02). Ottawa, Ontario, Canada: Public Safety Canada.
- Hanson, R. K., & Morton-Bourgon K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163. doi:10.1037/0022-006X.73.6.1154
- Hanson, R. K., & Morton-Bourgon, K. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21. doi:10.1037/a0014421
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119-136. doi:10.1023/A:1005482921333
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. (Corrections Research User Report No. 2003-01). Ottawa, ON: Department of the Solicitor General of Canada. Retrieved from Public Safety Canada website: <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/nts-dvlpmnt-sttc/nts-dvlpmnt-sttc-eng.pdf>
- Hanson, R. K., Thornton, D., Helmus, L., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with static-99R and static-2002R scores? *Sexual Abuse: Journal of Research and Treatment, 28*, 218-252. doi:10.1177/1079063215574710

- Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1*, 111-119.
doi:10.1016/0191-8869(80)90028-8
- Hare, R. D. (2003). *Manual for the Psychopathy Checklist-Revised* (2nd ed.). Toronto, Ontario: Multi-Health Systems.
- Harrell, E. (2007). *Black victims of violent crime*. Washington, D.C: Bureau of Justice Statistics. Retrieved from <http://www.bjs.gov/content/pub/pdf/bvvc.pdf>
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335.
doi:10.1177/0093854893020004001
- Harrison, P. M., & Beck, A. (2005). *Prisoners in 2005*. Washington, D.C: Bureau of Justice Statistics. Retrieved from <http://www.bjs.gov/content/pub/pdf/p05.pdf>.
- Hart, S. D. (2016). Culture and violence risk assessment: The case of *Ewert v. Canada*. *Journal of Threat Assessment and Management, 3*, 76-96.
doi:10.1037/tam0000068
- Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Hartney, C., & Vuong, L. (2009). *Created equal: Racial and ethnic disparities in the U.S. criminal justice system*. Retrieved from <http://jjie.org/wp-content/uploads/2018/03/created-equal.pdf>
- Hawkins, D. F. (1987). Beyond anomalies: Rethinking the conflict perspective on race and criminal punishment. *Social Forces, 65*, 719-745.

- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*, 347-359. doi.org/10.1023/A:1024851017947
- Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*, 8-25. doi:10.1177/0093854816678898
- Helmus, L., Babchishin, K. M., & Blais, J. (2012). Predictive accuracy of dynamic risk factors for Aboriginal and non-Aboriginal sex offenders: An exploratory comparison using STABLE-2007. *International Journal of Offender Therapy & Comparative Criminology, 56*, 856-876. doi:10.1177/0306624X11414693
- Helmus, L., & Hanson, R. K. (2011). More fun with statistics! How to use logistic regression to predict criminal recidivism risk. *Crime Scene, 18*(2), 8-12.
Retrieved from
<http://www.cpa.ca/docs/file/Sections/Criminal%20Justice%20Psychology/Crime%20Scene%202011-10.pdf>
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39*, 1148-1171.
doi:10.1177/0093854812443648
- Helmus, L., & Thornton, D. (2015). Stability and predictive and incremental accuracy of the individual items of Static-99R and Static-2002R in predicting sexual recidivism: A meta-analysis. *Criminal Justice and Behavior, 42*, 917-937.
doi:10.1177/0093854814568891

- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment, 24*, 64-101. doi:10.1177/1079063211409951
- Hilliard, A.G. (1979). Standardization and cultural bias impediments to the scientific study and validation of intelligence. *Journal of Research and Development in Education, 12*, 47-58.
- Hoffman, P. B. (1994). Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's Salient Factor Score. *Journal of Criminal Justice, 22*, 447-494. doi:10.1016/0047-2352(94)90090-6
- Hoffman, P. B. & Beck, J. L. (1974). Parole decision-making: A Salient Factor Score. *Journal of Criminal Justice, 2*, 195-206. doi:10.1016/0047-2352(74)90031-2
- Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2003). Ethnicity, gender, and the Level of Service Inventory-Revised. *Journal of Criminal Justice, 31*, 309. doi:10.1016/S0047-2352(03)00025-4
- Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006). Exploring the predictive validity of the Level of Service Inventory-Revised with Native American offenders. *Journal of Criminal Justice, 34*, 331-337. doi:10.1016/j.jcrimjus.2006.03.009
- Holt, N. (1998). The current state of parole in America. In J. Petersilia (Ed.), *Community corrections: Probation, parole, and intermediate sanctions* (pp. 28-41). New York, NY: Oxford University Press.
- Horens, V., & Poortinga, Y. (2000). Behavior in social context. In K. Pawlik & M. Rosenzweig (Eds), *International handbook of psychology* (pp. 40-63). London, U.K: Sage.

- Hosmer, D. W., & Lemeshow, S. (2002). *Applied logistic regression* (2nd ed.). New York, NY: Wiley. doi:10.1002/0471722146
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). New York, NY: Wiley.
- Huddy, L., & Virtanen, S. (1995). Subgroup Differentiation and Subgroup Bias Among Latinos as a Function of Familiarity and Positive Distinctiveness. *Journal of Personality & Social Psychology*, 68, 97-108.
- Inclán, J. E., & Quiñones, M. E. (2003). Puerto Rican children and adolescents. In J. T. Gibbs & L. N. Huang (Eds.), *Children of color: Psychological interventions with culturally diverse youth*. (2nd ed., pp. 382-408). San Francisco: Jossey Bass.
- Indian Act, RSC 1985, c I-5
- Ioimo, R., Tears, R. S., Meadows, L. A., Becton, J. B., & Charles, M. T. (2007). The police view of biased-based policing. *Police Quarterly*, 10, 270-287. doi:10.1177/1098611106296479
- Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, 19, 425-448. doi: 10.1177/107906320701900407
- Jensen, A. R. (1980). *Bias in mental testing*. New York, NY: Free Press
- Jespersen, A. F., Lalumière, M. L., & Seto, M. C. (2009). Sexual abuse history among adult sex offenders and non-sex offenders: A meta-analysis. *Child Abuse & Neglect*, 33, 179-192. doi:10.1016/j.chiabu.2008.07.004
- Jina, R., & Thomas, L. S. (2013). Health consequences of sexual violence against women. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 27, 15-26. doi:10.1016/j.bpobgyn.2012.08.012

- Juang, L. P., & Cookston, J. T. (2009). Acculturation, discrimination, and depressive symptoms among Chinese American adolescents: A longitudinal study. *The Journal of Primary Prevention, 30*, 475-496. doi:10.1007/s10935-009-0177-9
- Kansas v. Hendricks, 521 U.S. 346 (1997).
- Kaufman, K. L., & Hilliker, D. R. (1996). Subgroup differences in the modus operandi of adolescent sexual offenders. *Child Maltreatment, 1*, 17.
doi:10.1177/1077559596001001003
- Kaufman, J. M., Rebellon, C. J., Thaxton, S., & Agnew, R. (2008). A general strain theory of racial differences in criminal offending. *Australian & New Zealand Journal of Criminology, 41*, 421-437. doi:10.1375/acri.41.3.421
- Kehl, D., Guo, P., & Kessler, S., (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Retrieved from Digital Access to Scholarship at Harvard website:
<https://dash.harvard.edu/handle/1/33746041>
- Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2018). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse*. Advance online publication. doi:10.1177/1079063218800474
- Kennedy, M. A., & Gorzalka, B. B. (2002). Asian and non-Asian attitudes toward rape, sexual harassment, and sexuality. *Sex Roles, 46*, 227-238.
doi:10.1023/A:1020145815129
- Kirk, S. A. (1975). The sex offenses of Blacks and Whites. *Archives of Sexual Behavior, 4*, 295-302. doi:10.1007/BF01541629

- Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*, 296-310. doi:10.1037/0033-2909.107.3.296
- Knight, R. A., Carter, D. L., & Prentky, R. A. (1989). A system for the classification of child molesters: Reliability and application. *Journal of Interpersonal Violence*, *4*, 3-23. doi:10.1177/088626089004001001
- Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Washington, DC: U.S. Department of Justice. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/217618.pdf>
- Knight, G. P., Vargas-Chanes, D., Losoya, S. H., Cota-Robles, S., Chassin, L., & Lee, J. M. (2009). Acculturation and enculturation trajectories among Mexican-American adolescent offenders. *Journal of Research on Adolescence*, *19*, 625-653. doi:10.1111/j.1532-7795.2009.00614.x
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide*. Vancouver: British Columbia Institute on Family Violence.
- LaFromboise, T., Coleman, H. L., & Gerton, J. (1993). Psychological impact of biculturalism: evidence and theory. *Psychological Bulletin*, *114*, 395. doi:10.1037/0033-2909.114.3.395
- Långström, N., & Hanson, R. K. (2006). High rates of sexual behavior in the general population: Correlates and predictors. *Archives of Sexual Behavior*, *35*, 37-52. doi:10.1007/s10508-006-8993-y

Langton, L., Berzofsky, M., Krebs, C., & Smiley-McDonald, H. (2012).

Victimizations not reported to the police, 2006–2010 (Report No. NCJ

238536). Retrieved from Bureau of Justice Statistics website:

<http://www.bjs.gov/content/pub/pdf/vnrp0610.pdf>

LaPrairie, C. (1996). *Examining aboriginal corrections in Canada*. (Catalogue no.

JS5-1/14-1996E). Ottawa, ON: Ministry of the Solicitor General. Retrieved

from: [https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/xmnng-brgnl-](https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/xmnng-brgnl-crrctns/xmnng-brgnl-crrctns-eng.pdf)

[crrctns/xmnng-brgnl-crrctns-eng.pdf](https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/xmnng-brgnl-crrctns/xmnng-brgnl-crrctns-eng.pdf)

LaPrairie, C. (2002). Aboriginal over-representation in the criminal justice system: A

tale of nine cities. *Canadian Journal of Criminology*, *44*, 181–208.

Latzer, B. (2018). Subcultures of violence and African American crime rates. *Journal*

of Criminal Justice, *54*, 41-49. doi:10.1016/j.jcrimjus.2017.12.006

Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social*

organization of sexuality: Sexual practices in the United States. Chicago, IL:

University of Chicago Press.

Lawrence A. (2013). *Trends in Sentencing and Corrections: State Legislation*.

Denver: National Conference of State Legislatures. Retrieved from National

Conference of State Legislatures website:

<http://www.ncsl.org/Documents/CJ/TrendsInSentencingAndCorrections.pdf>

Lee, J., & Hahn, H. C. (2010). Acculturation and sexual risk behaviors among Latina

adolescents transitioning to young adulthood. *Journal of Youth and*

Adolescence, *39*, 414-27. doi:10.1007/s10964-009-9495-8

Lee, S. C., & Hanson, R. K. (2017). Similar predictive accuracy of the Static-99R risk

tool for White, Black, and Hispanic sex offenders in California. *Criminal*

Justice and Behavior, *44*, 1125-1140. doi:10.1177/0093854817711477

- Lee, S. C., Hanson, R. K., Calkins, C., & Jeglic, E. (2018). Paraphilia and antisociality: Motivations for sexual offending may differ for American Whites and Blacks. Manuscript submitted for publication.
- Lee, S. C., Hanson, R. K., Fullmer, N., Neeley, J., & Ramos, K. (2018). The predictive validity of Static-99R over 10 years for sexual offenders in California: 2018 update. Retrieved from SARATSO website:
http://saratso.org/pdf/Lee_Hanson_Fullmer_Neeley_Ramos_2018_The_Predictive_Validity_of_S_.pdf
- Lee, S. C., Mularczyk, K. P., Babchishin, K. M., Blais, J., & Bonta, J. (2018). Updated recidivism information for 360 individuals from the National Flagging System Project (2015). Unpublished raw data.
- Lee, S. C., Restrepo, A., Satariano, A., & Hanson, R. K. (2016). *The predictive validity of Static-99R for sexual offenders in California: 2016 update*. Retrieved from SARATSO website:
http://www.saratso.org/docs/ThePredictiveValidity_of_Static-99R_forSexualOffenders_inCalifornia-2016v1.pdf.
- Leguizamo, A., Lee, S. C., Jeglic, E. L., & Calkins, C. (2017). Utility of the Static-99 and Static-99R with Latino Sex Offenders. *Sexual Abuse, 29*, 765-785.
doi:10.1177/1079063215618377
- Leguizamo, A., Peltzman, B., Carrasco, N., Nosal, M., & Woods, L. (2010, April). *Ethnic differences among incarcerated sex offenders*. Paper presented at the 12th Annual Joint Conference of the Massachusetts Adolescent Sexual Offender Coalition and the Massachusetts Association for the Treatment of Sexual Abusers. Marlborough, MA.

- Levenson, J. S., Willis, G. M., & Prescott, D. S. (2016). Adverse childhood experiences in the lives of male sex offenders: Implications for trauma-informed care. *Sexual Abuse: Journal of Research and Treatment, 28*, 340-359. doi:10.1177/1079063214535819
- Ley, P. (1972). *Quantitative aspects of psychological assessment: An introduction*. London: Duckworth.
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber, & D. P. Farrington (Eds.), *Serious & violent juvenile offenders: Risk factors and successful interventions*. (pp. 313-345). Thousand Oaks, CA: Sage
- López, G. (2015). *Hispanics of Mexican origin in the United States, 2013*. Washington, D.C: PewResearchCenter. Retrieved from <http://www.pewhispanic.org/2015/09/15/hispanics-of-mexican-origin-in-the-united-states-2013/#fn-22745-1>
- MacDonald, S. F. (2014). *Profile of Aboriginal men offenders: Custody and supervision snapshots* (Report No. 2014 N° R-321). Retrieved from the Correctional Service Canada website: <http://publications.gc.ca/site/eng/9.810320/publication.html>
- MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of correctional boot camps on offending. *The ANNALS of the American Academy of Political & Social Science, 578*, 126-143. doi:10.1177/000271620157800108

- Malamuth, N. M. (1998). The confluence model as an organizing framework for research on sexually aggressive men: Risk moderators, imagined aggression, and pornography consumption. In R. G. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 229-245). San Diego, CA: Academic Press.
- Mann, M. (2009). *Good Intentions, Disappointing Results: A Progress Report on Federal Aboriginal Corrections*. Ottawa, ON: Office of the Correctional Investigator. Retrieved from <http://www.oci-bec.gc.ca/cnt/rpt/pdf/oth-aut/oth-aut20091113-eng.pdf>
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: Journal of Research and Treatment*, 22, 191-217.
doi:10.1177/1079063210366039
- Marshall, W. L., & Barbaree, H. E. (1990). An integrated theory of the etiology of sexual offending. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 257-275). New York, NY: Plenum Press.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, 35, 22-54.
- Mason, F., & Lodrick, Z. (2013). Psychological consequences of sexual assault. *Clinical Obstetrics & Gynaecology*, 27, 27-37.
doi:10.1016/j.bpobgyn.2012.08.015

- Mattis, J. S., & Grayman-Simpson, N. (2013). Faith and the sacred in African American life. In K. I. Pargament, J. J. Exline & J. W. Jones (Eds.), *APA handbook of psychology, religion, and spirituality (vol 1): Context, theory, and research; APA handbook of psychology, religion, and spirituality (vol 1): Context, theory, and research* (pp. 547-564). American Psychological Association, Washington, DC. doi:10.1037/14045-030
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Mercado, C. C., Jeglic, E., & Markus, K. (2011). *Sex offender management, treatment, and civil commitment: An evidence based analysis aimed at reducing sexual violence (2007-IJ-CX0037)*. Washington, DC: National Institute of Justice.
- Meston, C. M., & Ahrold, T. (2010). Ethnic, gender, and acculturation influences on sexual behaviors. *Archives of Sexual Behavior, 39*, 179-189.
doi.10.1007/s10508-008-9415-0
- Miller, J. A. (2010). Sex offender civil commitment: The treatment paradox. *California Law Review, 98*, 2093-2128.
- Miller, J. M., Miller, H. V., Zapata, J. T., & Yin, Z. (2008). Mexican-American youth drug use and acculturation: A note on the mitigating effects of contextual dynamics. *Journal of Drug Issues, 38*, 199–214.
- Ministry of Correctional Services Act, R.S.O. 1990, c. M.22.
- Minton, T. D., & Zeng, Z. (2016). *Jail Inmates in 2015*. (Report No. NCJ250394). Retrieved from the Bureau of Justice Statistics website:
<https://www.bjs.gov/content/pub/pdf/ji15.pdf>

- Monahan, J. (2007). Clinical and actuarial predictions of violence. In D. Faigman, D. Kaye, M. Saks, J. Sanders, & E. Cheung (Eds.), *Modern scientific evidence: The law and science of expert testimony* (pp. 112-147). St. Paul, MN: West Publishing Company.
- Moro, P. E. (1998). Treatment for Hispanic sexual offenders. In W. L. Marshall, Y. M. Fernandez, S. M. Hudson, & T. Ward (Eds.), *Sourcebook of treatment programs for sexual offenders*. (pp. 445-456). New York, NY: Plenum Press.
- Murphy, W. D., DiLillo, D., Haynes, M. R., & Steere, E. (2001). An exploration of factors related to deviant sexual arousal among juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 91-103.
doi:10.1177/107906320101300203
- Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, *41*, 1406-1421.
doi:10.1177/0093854814548449
- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, Canada: Solicitor General of Canada.
- Office of the Auditor General of Canada. (2016). *Preparing Indigenous offenders for release*. Ottawa, ON. Retrieved from http://www.oag-bvg.gc.ca/internet/English/parl_oag_201611_03_e_41832.html
- Office of the Correctional Investigator (2015). *Annual report of the office of the correctional investigator, 2014-2015*. (Catalogue no. PS100E-PDF). Retrieved from the Office of the Correctional Investigator website: <http://www.ocibec.gc.ca/cnt/rpt/pdf/annrpt/annrpt20142015-eng.pdf>

- Olver, M. E. (2003). *The development and validation of the Violence Risk Scale: Sexual Offender Version (VRS:SO) and its relationship to psychopathy and treatment attrition* (Unpublished doctoral dissertation). University of Saskatchewan, Saskatoon, Saskatchewan.
- Olver, M. E., Sowden, J. N., Kingston, D. A., Nicholaichuk, T. P., Gordon, A., Beggs Christofferson, S. M., & Wong, S. P. (2018). Predictive accuracy of Violence Risk Scale - Sexual Offender Version Risk and change scores in treated Canadian Aboriginal and non-Aboriginal sexual offenders. *Sexual Abuse, 30*, 254-275. doi:10.1177/1079063216649594
- Otto, R. K., & Douglas, K. S. (2015). *Handbook of Violence Risk Assessment*. New York, NY: Routledge.
- Padilla, A. M. (1988). Early psychological assessments of Mexican-American children. *Journal of the History of the Behavioral Sciences, 24*, 111-117. doi:10.1002/1520-6696(198801)24:1
- Pérez, D. M., Jennings, W. G., & Gover, A. R. (2008). Specifying general strain theory: An ethnically relevant approach. *Deviant Behavior, 29*, 544-578. doi:10.1080/01639620701839385
- Perley-Robertson, B., Helmus, L. M., & Forth, A. (2018). Predictive accuracy of static risk factors for Canadian indigenous offenders compared to non-indigenous offenders: Implications for risk assessment scales. *Psychology, Crime & Law*. Advance online publication. doi:10.1080/1068316X.2018.1519827
- Petrosino, A. J. (1997). *'What works?' revisited again: A meta-analysis of randomized field experiments in rehabilitation, deterrence and prevention* (Unpublished doctoral dissertation). Rutgers State University, Newark, New Jersey.

Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M, Hanson R. K. & Thornton D.

(2017). *Static-99R Coding Rules – Revised 2016*. Research Report 2017-R012.

Ottawa: Public Safety Canada.

Phenix, A., Helmus, L., & Hanson, R. K. (2016). *Static-99R & Static-2002R*

evaluators' workbook. Retrieved from: <http://www.static99.org/>

Plessy v. Ferguson, 163 U.S. 537 (1896)

Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, 38, 931-964.

doi:10.1111/j.1745-9125.2000.tb00911.x

Proctor, B. D., Semega, J. L., & Kollar, M. A. (2016). *Income and poverty in the*

United States: 2015. (Report No. P60-256). Retrieved from the United States Census Bureau website:

<https://www.census.gov/library/publications/2016/demo/p60-256.html>

Public Safety Canada (2015). *Corrections and Conditional Release Statistical*

Overview. (Catalogue No. PS1-3E-PDF). Retrieved from the Public Safety Canada website:

<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ccrso-2015/ccrso-2015-en.pdf>

Quann, N. L., & Trevethan, S. (2000). *Police-reported Aboriginal crime in*

Saskatchewan. (Catalogue no.85F0031XIE). Ottawa, ON: Statistics Canada.

Retrieved from:

<http://www.publications.gc.ca/site/eng/9.685779/publication.html>

Quesada, S. P., Calkins, C., & Jeglic, E. L. (2014). An examination of the interrater

reliability between practitioners and researchers on the Static-99. *International Journal of Offender Therapy and Comparative Criminology*, 58, 1364-1375.

doi:10.1177/0306624X13495504

- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A., (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.
- R. v. Gladue, [1999] 1 S.C.R. 688.
- Rastin, C. J., & Johnson, S. L. (2002). Inuit sexual offenders: Victim, offence and recidivism characteristics. *Forum on Corrections Research*, 14(3). Ottawa, ON: Correctional Service of Canada. Retrieved from: http://www.csc-scc.gc.ca/research/forum/e143/143n_e.pdf
- Redfield, R., Linton, R., & Herskovits, M. J. (1936). Memorandum for the study of acculturation. *American Anthropologist*, 38, 149-152.
doi:10.1525/aa.1936.38.1.02a00330
- Reiss, A. J. (1951). Delinquency as the failure of personal and social controls. *American Sociological Review*, 16, 196-207. doi:10.2307/2087693
- Reynolds, C. R. (1995). Test bias in the assessment of intelligence and personality. In D. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 545-576). New York, NY: Plenum.
- Reynolds, C. R. (1998). Cultural bias in testing of intelligence and personality. In A. Bellack & M. Hersen & C. Belar (Eds.), *Comprehensive clinical psychology: Cross-cultural psychology* (pp. 53-92). New York, NY: Elsevier Science.
- Reynolds, C. R. (2000a). Methods for detecting and evaluating cultural bias in neuropsychological tests. In E. Fletcher-Janzen, T. L. Strickland, & C. R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 249-285). New York, NY: Kluwer Academic/Plenum Publishers.

- Reynolds, C. R. (2000b). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, 6, 144-150.
doi:10.1037/1076 8971.6.1.144
- Reynolds, C. R., & Brown, R. T. (1984). Bias in mental testing: An introduction to the issues. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 1-39). New York, NY: Plenum.
- Reynolds, C. R., & Suzuki, L. A. (2013). Bias in psychological assessment: An empirical review and recommendations. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 82-113). Hoboken, NJ: John Wiley.
- Rice, A. K., Boccaccini, M. T., & Varela, J. G. (2013, March). *How do Static-99 scores differ among White, Black, and Latino offenders?* Poster presented at the meeting of the American Psychology-Law Society, Portland, OR.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29, 615-620.
doi:10.1007/s10979-005-6832-7
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77-84. doi:10.1186/1471-2105-12-77
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., Müller, M., & Siegert, S. (2014). *Package 'pROC'*. Retrieved from: <https://cran.r-project.org/web/packages/pROC/pROC.pdf>
- Rockhill, B., Byrne, C., Rosner, B., Louie, M. M., & Colditz, G. (2003). Breast cancer risk prediction with a log-incidence model: Evaluation of accuracy. *Journal of Clinical Epidemiology*, 56, 856-861. doi:10.1016/S0895-4356(03)00124-0

- Rojas, E. Y., & Gretton, H. M. (2007). Background, offence characteristics, and criminal outcomes of aboriginal youth who sexually offend: A closer look at aboriginal youth intervention needs. *Sexual Abuse: Journal of Research and Treatment, 19*, 257-283. doi:10.1177/107906320701900306
- Rosa, D. L. (2002). Acculturation and Latino adolescents' substance use: A research agenda for the future. *Substance Use & Misuse, 37*, 429-456. doi:10.1081/JA-120002804
- Royal Commission on Aboriginal Peoples (1996). *Report of the Royal Commission on Aboriginal Peoples*. Ottawa: Minister of Supply and Services.
- Royston, P., Moons, K., Altman, D., & Vergouwe, Y. (2009). Prognosis and prognostic research: Developing a prognostic model. *British Medical Journal, 338*, 1373-1377. doi:10.1136/bmj.b604
- Rudin, J. (2009). *Aboriginal Peoples and the Criminal Justice System*. Retrieved from:
http://www.archives.gov.on.ca/en/e_records/ipperwash/policy_part/research/pdf/Rudin.pdf
- Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*, 19-30. doi:10.1037/1082-989X.13.1.19
- Salter, D., McMillan, D., Richards, M., Talbot, T., Hodges, J., Bentovim, A., Hastings, R., Stevenson, J., & Skuse, D. (2003). Development of sexually abusive behaviour in sexually victimised males: A longitudinal study. *The Lancet, 361*, 471-476. doi:10.1016/S0140-6736(03)12466-X

- Sampson, R. J., Morenoff, J. D., & Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health, 95*, 224-231. doi:10.2105/AJPH.2004.037705
- Sanchez, D., Hamilton, E. R., Gilbert, D. J., & Vandewater, E. A. (2018). Examining Africentric cultural values, ethnic identity, and substance use abstinence in low-income, early adolescent, African American girls. *Journal of Black Psychology, 44*, 74-100. doi:10.1177/0095798417746265
- Schaaf, S., Jeglic, E.L., Calkins, C., Raymaekers, L., Leguizamo, A.(2016). Examining ethnoracial related differences in child molester typology: An MTC: CM3 Approach. *Journal of Interpersonal Violence*. Advance online publication. doi:10.1177/0886260516653550
- Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: An international meta-analysis of sound quality evaluations. *Journal of Experimental Criminology*. doi:10.1007/s11292-015-9241-z
- Scott-Hayward, C. S. (2009). *The fiscal crisis in corrections: Rethinking policies and practices*. New York, NY: Center on Sentencing and Corrections, Vera Institute of Justice.
- Scrim, K. (2010). Aboriginal victimization in Canada: A summary of the literature. *Victims of Crime Research Digest, 3*. Retrieved from <http://www.justice.gc.ca/eng/rp-pr/cj-jp/victim/rd3-rr3/p3.html>
- Sellin, T. (1938). Culture conflict and crime. *American Journal of Sociology, 44*, 97-103.
- Sentencing Reform and Corrections Act, 2015, S.2123

- Sex Offender Information Registration Act, S.C. 2004, c. 10 (2004). Retrieved from Justice Laws website: <http://laws-lois.justice.gc.ca/eng/acts/S-8.7/page-1.html>
- Shepherd, S. M., Adams, Y., McEntyre, E., & Walker, R. (2014). Violence risk assessment in Australian aboriginal offender populations: A review of the literature. *Psychology, Public Policy, and Law, 20*, 281-293.
doi:10.1037/law0000017
- Shepherd, S. M., & Lewis-Fernandez, R. (2016). Forensic risk assessment and cultural diversity: Contemporary challenges and future directions. *Psychology, Public Policy, and Law, 22*, 427-438. doi:10.1037/law0000102
- Simons, R. L., Chen, Y. F., Stewart, E. A., & Brody, G. H. (2003). Incidents of discrimination and risk for delinquency: A longitudinal test of strain theory with an African American sample. *Justice Quarterly, 20*, 827-854.
doi:10.1080/07418820300095711
- Simons, D. A., Wurtele, S. K., & Durham, R. L. (2008). Developmental experiences of child sexual abusers and rapists. *Child Abuse & Neglect, 32*, 549-560.
doi:10.1016/j.chiabu.2007.03.027
- Sinha, V., Trocmé, N., Fallon, B., MacLaurin, B., Fast, E., Prokop, S. T., ... Richard, K. (2011). *Kiskisik Awasisak: Remember the children. Understanding the overrepresentation of First Nations children in the child welfare system*. Ontario: Assembly of First Nations. Retrieved from the Canadian Child Welfare Research Portal website: <http://cwrp.ca/publications/2280>
- Sioui, R., & Thibault, J. (2002). Examining reintegration potential for Aboriginal offenders. *Forum on Corrections Research, 14*, 49-51.

- Smith, P., Goggin, C., & Gendreau, P. (2002). *The effects of prison sentences and intermediate sanctions on recidivism: General effects and individual differences*. (Report No. JS42-103/2002). Retrieved from Public Safety Canada website: <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ffcts-prsn-sntncs/ffcts-prsn-sntncs-eng.pdf>
- Smith-Socaris, C., Perry, R. A., & Fox-Mullen, L. (2006). Racial and ethnic demographics of the New York state level 3 sex offender population. Retrieved from New York Civil Liberties Union website: https://www.nyclu.org/sites/default/files/sexoffender_analysis_121106.pdf
- Solicitor General Canada (2001). *High-Risk offenders: A handbook for criminal justice professionals* (No. JS42-94/2000). Ottawa, Ontario, Canada: Solicitor General Canada.
- Spickard, P. R. (2007). *Almost all aliens: Immigration, race, and colonialism in American history and identity*. New York: Routledge.
- Starr, S. B. (2014) Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, 66, 803-872.
- Statistics Canada (2005). *Projections of the Aboriginal populations, Canada, provinces and territories- 2001 to 2017*. (Catalogue no. 91-547-XIE). Retrieved from the Statistics Canada website: <https://www150.statcan.gc.ca/n1/pub/91-547-x/91-547-x2005001-eng.pdf>
- Statistics Canada (2015). *Aboriginal statistics at a glance* (Catalogue no. 89-645-X2015001). Retrieved from the Statistics Canada website: https://www150.statcan.gc.ca/n1/en/pub/89-645-x/89-645-x2015001-eng.pdf?st=_GMvln8d

Statistics Canada (2016). *Census Profile, 2016 Census*. Retrieved from Statistics

Canada website: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=PR&Code2=01&Data=Count&SearchText=Canada&SearchType=Begins&SearchPR=01&B1=All&TABID=1>

Statistics Canada (2018). *Aboriginal Population Profile, 2016 Census*. (Catalogue

no. 98-510-X2016001). Retrieved from the Statistics Canada website:

http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/abpopprof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Data=Count&SearchText=Canada&SearchType=Begins&B1=All&SEX_ID=1&AGE_ID=1&RESGEO_ID=1

Stewart, E. A., & Simons, R. L. (2010). Race, code of the street, and violent

delinquency: A multilevel investigation of neighborhood street culture and individual norms of violence. *Criminology: An Interdisciplinary Journal*, 48, 569-605. doi:10.1111/j.1745-9125.2010.00196.x

Stoltenborgh, M., van IJzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M.

(2011). A global perspective on child sexual abuse: Meta-analysis of prevalence around the world. *Child Maltreatment*, 16, 79-101.

doi:10.1177/1077559511403920

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve

diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26.

doi:10.1111/1529-1006.001

Takaki, R. (2008). *A different mirror: A history of multicultural America*. New York,

NY: Little, Brown.

- Tatum, B. (2002). The colonial model as a theoretical explanation of crime and delinquency. In S. L. Gabbidon, H. T. Greene, & V. D. Young (Eds.) *African American Classics in Criminology and Criminal Justice* (pp. 307-322). Thousand Oaks, CA: Sage.
- Thorndike, R. L. (1971). Concepts of culture-fairness. *Journal of Educational Measurement*, 8, 63-70. doi:10.1111/j.1745-3984.1971.tb00907.x
- Tjaden, P. G., & Thoennes, N. (2006). *Extent, nature, and consequences of rape victimization: Findings from the National Violence against Women Survey*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
- Tolnay, S. E., & Beck, E. M. (1992). Racial violence and Black migration in the American South, 1910 to 1930. *American Sociological Review*, 57, 103-116.
- Tolnay, S. E., Deane, G., & Beck, E. M. (1996). Vicarious Violence: Spatial Effects on Southern Lynchings, 1890-1919. *American Journal of Sociology*, 102, 788. doi:10.1086/230997
- Travis, J., Western, B., & Redburn, S. (2014). *The growth of incarceration in the United States: Exploring causes and consequences*. Washington, DC: National Academies Press
- Trevethan, S., Moore, J.-P., & Rastin, C. J. (2002). A profile of Aboriginal offenders in federal facilities and serving time in the community. *Forum on Corrections Research*, 14, 17-19. Retrieved from: http://www.csc-scc.gc.ca/text/pblct/forum/e143/143f_e.pdf
- Truth and Reconciliation Commission of Canada (2015). *Honouring the truth, reconciling for the future*. Winnipeg, Manitoba, Canada: Author.

- Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33*, 287-316. doi:10.1016/j.cpr.2012.12.002
- Turner, S., Greenwood, P. W., Chen, E., & Fain, T. (1999). The impact of truth-in-sentencing and three strikes legislation: Prison populations, state budgets, and crime rates. *Stanford Law & Policy Review, 11*, 75-91.
- United States Census Bureau (2015). *Annual estimates of the resident population by sex, race, and Hispanic origin for the United States, states, and counties*. Retrieved from <http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>
- United States Census Bureau (2017a). *Annual estimates of the resident population by sex, age, race, and Hispanic origin for the United States and States: April 1, 2010, to July 1, 2016*. Retrieved from <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bk>
- United States Census Bureau (2017b). *Educational attainment in the United States: 2017*. Retrieved from <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>
- United States Census Bureau (2018). *Race*. Retrieved from <https://www.census.gov/topics/population/race/about.html>

- United States Department of Justice (2007). *The national guidelines for sex offender registration and notification*. Retrieved from U.S. Department of Justice website:
https://www.justice.gov/archive/tribal/docs/fv_tjs/session_3/session3_presentations/Sex_Offender_Guidelines.pdf
- United States Department of Justice, Federal Bureau of Investigation. (2015). *Crime in the United States, 2006*. Retrieved from <https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-43>
- Van de vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology / Revue Européenne De Psychologie Appliquée, 54*, 119-135.
doi:10.1016/j.erap.2003.12.004
- Varela, J. G., Boccaccini, M. T., Cuervo, V. A., Murrie, D. C., & Clark, J. W. (2014). Same score, different message: Perceptions of offender risk depend on Static-99R risk communication format. *Law and Human Behavior, 38*, 418 – 427.
doi:10.1037/lhb0000073
- Varela, J. G., Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Gonzalez, E., Jr. (2013). Do the Static-99 and Static-99R perform similarly for White, Black, and Latino sexual offenders? *International Journal of Forensic Mental Health, 12*, 231-243. doi:10.1080/14999013.2013.846950
- Vazsonyi, A. T., & Crosswhite, J. M. (2004). A test of Gottfredson and Hirschi's general theory of crime in African American adolescents. *Journal of Research in Crime and Delinquency, 41*, 407-432. doi: 10.1177/0022427803262060

- Ventura Miller, H., Barnes, J. C., & Hartley, R. D. (2011). Reconsidering Hispanic gang membership and acculturation in a multivariate context. *Crime & Delinquency*, *57*, 331-355. doi:10.1177/0011128709348460
- Vergouwe, Y., Steyerberg, E. W., Eijkemans, J. C., & Habbema, J. D. F. (2005). Substantial effective sample sizes are required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology*, *58*, 475-483. doi:10.1016/j.jclinepi.2004.06.017
- Von Hirsch, A. (1976). *Doing justice: The choice of punishments*. New York, NY: Hill and Wang.
- Waldron, M. O. (2012). *Comparing static-99 scores of incarcerated White, Black, and Latino sex offenders* (Unpublished doctoral dissertation). Antioch University New England, Keene, NH.
- Ward, T., & Beech, A. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior*, *11*, 44-63. doi:10.1016/j.avb.2005.05.002
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (Version 2)*. Vancouver, B.C.: Simon Fraser University.
- Weitzer, R., & Tuch, S. A. (2002). Perceptions of racial profiling: Race, class, and personal experience. *Criminology*, *40*, 435-456. doi:10.1111/j.1745-9125.2002.tb00962.x
- West, L. W., & Macarthur, R. S. (1964). An evaluation of selected intelligence tests for two samples of Metis and Indian children. *Alberta Journal of Educational Research*, *10*, 17-27.

Whitaker, D. J., Le, B., Hanson, R. K., Baker, C. K., McMahon, P. M., Ryan, G., . . .

Rice, D. D. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse & Neglect*, *32*, 529-548.

doi:10.1016/j.chiabu.2007.08.005

Whitehead, J. T., & Lab, S. P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime & Delinquency*, *26*, 276-295.

doi:10.1177/0022427889026003005

Widom, C. S., & Massey, C. (2015). A prospective examination of whether childhood sexual abuse predicts subsequent sexual offending. *JAMA Pediatrics*, *169*,

e143357. doi:10.1001/jamapediatrics.2014.3357

Wolfgang, M. (1958). *Patterns in criminal homicide*. Philadelphia: University of Pennsylvania.

Wong, S. K. (1999). Acculturation, peer relations, and delinquent behavior of Chinese-Canadian youth. *Adolescence*, *34*, 108-119.

Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale: Sexual Offender version (VRS-SO)*. Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and the University of Saskatchewan.

Woodrow, A. C., & Bright, D. A. (2011). Effectiveness of a sex offender treatment programme: A risk band analysis. *International Journal of Offender Therapy and Comparative Criminology*, *55*, 43-55. doi:10.1177/0306624X09352162

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*, 740-767. doi:10.1037/a0020473

Zeidner, M. (1988). Cultural fairness in aptitude testing revisited: A cross-cultural parallel. *Professional Psychology: Research and Practice, 19*, 257-262.

doi:10.1037/0735-7028.19.3.257

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scale. *Journal of Modern Applied Statistical Methods, 6*, 21-29. doi:10.22237/jmasm/117799218

APPENDIX A

Regional Distribution of the Samples (Study 3)

Study	<i>N</i>	Pacific		Prairies		Ontario		Quebec		Atlantic	
		<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)
Bonta & Yessine (2005) ¹	256	51	(20.0)	104	(40.6)	65	(25.4)	0	(0.0)	36	(14.1)
Indigenous	37	-	-	-	-	-	-	-	-	-	-
White	118	-	-	-	-	-	-	-	-	-	-
Haag (2005)	191	28	(14.7)	57	(29.8)	69	(36.1)	22	(11.5)	15	(7.9)
Indigenous	50	8	(16.0)	30	(60.0)	6	(12.0)	4	(8.0)	2	(4.0)
White	141	20	(14.2)	27	(19.1)	63	(44.7)	18	(12.8)	13	(9.2)
Olver et al. (2018) ²	1,164	-	-	-	-	-	-	-	-	-	-
Indigenous	425	-	-	-	-	-	-	-	-	-	-
White	739	-	-	-	-	-	-	-	-	-	-
Brankley et al. (2017a)	382		112 (29.3)			109 (28.5)		102 (26.7)		59 (15.4)	
Indigenous	36		22 (61.1)			7 (19.4)		1 (2.8)		6 (16.7)	
White	346		90 (26.0)			102 (29.5)		101 (29.2)		53 (15.3)	
Lee et al. (2018)	321	72	(22.4)	107	(33.3)	140	(43.6)	0	(0.0)	2	(0.6)
Indigenous	105	19	(18.1)	66	(62.9)	20	(19.0)	0	(0.0)	0	(0.0)
White	216	53	(24.5)	41	(19.0)	120	(55.6)	0	(0.0)	2	(0.9)

Note. 1. In this study, only 155 individuals who have a history of sexual crimes were used.

2. More than 80% of the total sample was from the Prairies (e.g., the Clearwater Program).

APPENDIX B

Comparison of logistic regression parameters of Static-99R between Indigenous and White groups (Study 3)

Study	Indigenous	White	B _{Indigenous-White}	95% C.I.
A fixed 5-year follow-up period				
B₀				
Bonta & Yessine (2005)	- 1.876 (.954)	- 3.305 (.863)	1.43	[- 1.09, 3.95]
Haag (2005)	- 1.384 (.551)	- 2.343 (.392)	0.96	[- 0.37, 2.28]
Olver et al. (2018)	- 2.393 (.265)	- 2.889 (.212)	0.50	[- 0.17, 1.16]
Brankley et al. (2017a)	- 2.366 (.991)	- 3.047 (.366)	0.68	[- 1.39, 2.75]
Lee et al. (2018)	- 2.030 (.623)	- 2.539 (.401)	0.51	[- 0.94, 1.96]
B₁				
Bonta & Yessine (2005)	.162 (.234)	.442 (.188)	- 0.28	[- 0.87, 0.31]
Haag (2005)	.187 (.181)	.324 (.106)	- 0.14	[- 0.55, 0.27]
Olver et al. (2018)	.201 (.067)	.287 (.054)	- 0.09	[- 0.25, 0.08]
Brankley et al. (2017a)	.114 (.283)	.340 (.089)	- 0.23	[- 0.81, 0.36]
Lee et al. (2018)	.168 (.148)	.279 (.092)	- 0.11	[- 0.45, 0.23]

APPENDIX B (continued)

Study	Indigenous	White	B _{Indigenous-White}	95% C.I.
A fixed 10-year follow-up period				
B₀				
Olver et al. (2018)	- 2.001 (.276)	- 2.149 (.228)	0.15	[- 0.55, 0.85]
Brankley et al. (2017a)	- 1.736 (.783)	- 2.437 (.292)	0.70	[- 0.94, 2.34]
Lee et al. (2018)	- 2.195 (.756)	- 2.070 (.406)	- 0.13	[- 1.81, 1.56]
B₁				
Olver et al. (2018)	.274 (.071)	.326 (.059)	- 0.05	[- 0.23, 0.13]
Brankley et al. (2017a)	.126 (.226)	.407 (.075)	- 0.28	[- 0.74, 0.18]
Lee et al. (2018)	.342 (.175)	.298 (.099)	0.04	[- 0.35, 0.44]

APPENDIX C

Comparison of logistic regression parameters of Static-2002R between Indigenous and White groups (Study 3)

Study	Indigenous	White	B _{Indigenous-White}	95% C.I.
A fixed 5-year follow-up period				
B0₃				
Haag (2005)	- 1.212 (.600)	- 2.506 (.427)	1.29	[- 0.15, 2.73]
Brankley et al. (2017a)	- 2.540 (1.124)	- 3.200 (.407)	0.66	[- 1.68, 3.00]
Lee et al. (2018)	- 1.768 (.563)	- 2.561 (.410)	0.79	[0.57, 2.16]
B1				
Haag (2005)	.092 (.171)	.313 (.100)	- 0.22	[- 0.61, 0.17]
Brankley et al. (2017a)	.142 (.268)	.314 (.085)	- 0.17	[- 0.72, 0.38]
Lee et al. (2018)	.093 (.132)	.248 (.085)	- 0.15	[- 0.46, 0.15]
A fixed 10-year follow-up period				
B0₃				
Brankley et al. (2017a)	- 2.437 (.997)	- 2.686 (.332)	0.25	[-1.81, 2.31]
Lee et al. (2018)	- 1.560 (.591)	- 2.040 (.406)	0.48	[- 0.93, 1.89]
B1				
Brankley et al. (2017a)	.293 (.229)	.391 (.072)	- 0.10	[- 0.57, 0.37]
Lee et al. (2018)	.178 (.137)	.258 (.088)	- 0.08	[- 0.40, 0.24]

APPENDIX D

Static-99R - TALLY SHEET

Assessment date: _____ Date of release from index sex offence: _____

Item #	Risk Factor	Codes		Score
1	Age at release from index sex offence	Aged 18 to 34.9		1
		Aged 35 to 39.9		0
		Aged 40 to 59.9		-1
		Aged 60 or older		-3
2	Ever lived with a lover	Ever lived with lover for at least two years?		0
		Yes		1
		No		
3	Index non-sexual violence - Any convictions	No		0
		Yes		1
4	Prior non-sexual violence - Any convictions	No		0
		Yes		1
5	Prior sex offences	<u>Charges</u>	<u>Convictions</u>	
		0	0	0
		1,2	1	1
		3-5	2,3	2
		6+	4+	3
6	Four or more prior sentencing dates (excluding index)	3 or less		0
		4 or more		1
7	Any convictions for non-contact sex offences	No		0
		Yes		1
8	Any unrelated victims	No		0
		Yes		1
9	Any stranger victims	No		0
		Yes		1
10	Any male victims	No		0
		Yes		1
	Total Score	Add up scores from individual risk factors		

Nominal Risk Levels (2016 version)	<u>Total</u>	<u>Risk Level</u>
	-3, -2,	I - Very Low Risk
	-1, 0,	II - Below Average Risk
	1, 2, 3	III - Average Risk
	4, 5	IVa - Above Average Risk
6 and higher	IVb - Well Above Average Risk	

There [was, was not] sufficient information available to complete the Static-99R score following the coding manual (2016 version). I believe that this score [fairly represents, does not fairly represent] the risk presented by Mr. XXXX at this time. Comments/Explanation: _____

(Evaluator name)_____
(Evaluator signature)_____
(Date)

APPENDIX E

Static-2002R - TALLY SHEET

ITEMS	Raw Score	Subscore
AGE		
1. Age at Release 18 to 34.9 = 2 35 to 39.9 = 1 40 to 59.9 = 0 60 or older = -2		
PERSISTENCE OF SEXUAL OFFENDING		
2. Prior Sentencing Occasions for Sexual Offences: No prior sentencing dates for sexual offences = 0 1 = 1 2, 3 = 2 4 or more = 3		
3. Any Juvenile Arrest for a Sexual Offence and Convicted as an Adult for a Separate Sexual Offence: No arrest for a sexual offence prior to age 18 = 0 Arrest prior to age 18 and conviction after age 18 = 1		
4. Rate of Sexual Offending: Less than one sentencing occasion every 15 years = 0 One or more sentencing occasions every 15 years = 1		
Persistence Raw Score (subtotal of Sexual Offending) 0 = 0 1 = 1 2, 3 = 2 4, 5 = 3		
Persistence of Sexual Offending SUBSCORE		
DEVIANT SEXUAL INTERESTS		
5. Any Sentencing Occasion For Non-contact Sex Offences: No = 0 Yes = 1		
6. Any Male Victim: No = 0 Yes = 1		
7. Young, Unrelated Victims: Does <u>not</u> have two or more victims < 12, one of them unrelated = 0 Does have two or more victims < 12 years, one must be unrelated = 1		
Deviant Sexual Interest SUBSCORE		
RELATIONSHIP TO VICTIMS		
8. Any Unrelated Victim: No = 0 Yes = 1		
9. Any Stranger Victim: No = 0 Yes = 1		
Relationship to Victims SUBSCORE		

GENERAL CRIMINALITY		
10. Any Prior Involvement with the Criminal Justice System No = 0 Yes = 1		
11. Prior Sentencing Occasions For Anything: 0-2 prior sentencing occasions for anything = 0 3-13 prior sentencing occasions = 1 14 or more prior sentencing occasions = 2		
12. Any Community Supervision Violation: No = 0 Yes = 1		
13. Years Free Prior to Index Sex Offence: • More than 36 months free prior to committing the sexual offence that resulted in the index conviction AND more than 48 months free prior to index conviction = 0 • Less than 36 months free prior to committing the sexual offence that resulted in the index conviction OR less than 48 months free prior to conviction for index sex offence = 1		
14. Any Prior Non-sexual Violence Sentencing Occasion: No = 0 Yes = 1		
General Criminality raw score (subtotal General Criminality items) 0 = 0 1, 2 = 1 3, 4 = 2 5, 6 = 3		
General Criminality SUBSCORE		
TOTAL -2 to 13		

<u>Score</u>	<u>Labels for Standardized Risk Categories</u>
-2, -1	= Level I, Very Low Risk
0, 1	= Level II, Below Average Risk
2, 3, 4	= Level III, Average Risk
5, 6	= Level IVa, Above Average Risk
7 plus	= Level IVb, Well Above Average Risk