

A Multi-Site Comparison of the Validity and Utility of the Static-99 and Static-2002 for
Risk Assessment With Sexual Offenders

Leslie Helmus

Carleton University

100613177

Honours Thesis

CRCJ 4908

Dr. Joanna Pozzulo

March, 2007

Table of Contents

List of Tables.....	v
Acknowledgements.....	vi
Abstract.....	vii
Introduction.....	1
Risk Assessment: History and Various Methods.....	3
Which Method is More Effective?.....	6
Which Method is More Effective With Sex Offenders?.....	8
Development of the Static-99.....	10
Use and Replication of the Static99.....	14
Development of the Static-2002.....	17
Replications of the Static-2002.....	22
Practical Implications: Implementation of Actuarial Tools.....	23
Purpose of the Study.....	24
Study 1.....	25
Study 2.....	26
Study 1.....	26
Methods.....	26
Samples.....	26
1) Dynamic Supervision Project.....	29
2) Canadian federal offenders: Regional Treatment Centre..	30
3) Canadian federal offenders: B.C.....	31
4) Canadian federal offenders: 1995 WED.....	31

5) Canadian federal offenders: Warkworth institution.....	32
6) Canadian federal offenders: Quebec.....	32
7) Bridgewater: Massachusetts Treatment Center.....	33
8) U.K.: Multi-site treatment.....	34
Missing Data.....	34
Plan of Analysis.....	35
Results.....	37
Risk Scores.....	37
Predictive Accuracy Across Samples.....	38
Note on Aggregating Results.....	41
Comparative Accuracy.....	41
Risk Categories for the Static-2002.....	44
Discussion.....	45
Study 2.....	49
Hypotheses.....	49
Methods.....	49
Participants.....	49
Materials.....	52
Procedure.....	52
Plan of Analysis.....	54
Results.....	54
Reasons for Using the Static-99.....	54
Satisfaction With the Static-99.....	55

Willingness to Use the Static-2002.....	59
Discussion.....	62
References.....	67
Appendix A: Static-99 Coding Form.....	75
Appendix B: Static-2002 Coding Form.....	76
Appendix C: Letter Sent to Researchers to Recruit Datasets.....	78
Appendix D: Static-2002 Recidivism Percentages by Risk Level.....	81
Appendix E: Static-99 Questionnaire.....	82

List of Tables

Table 1: Results from Hanson and Morton-Bourgon's (2007) Meta-Analysis.....	9
Table 2: Static-99 Items.....	13
Table 3: Hanson and Morton-Bourgon's (2007) Results for Individual Risk Scales.....	18
Table 4: Static-2002 Items.....	21
Table 5: Study 1 Sample Characteristics.....	28
Table 6: Missing Data for Static-99 and Static-2002 Items.....	35
Table 7: Mean Risk Scores for the Static-99 and Static-2002.....	38
Table 8: ROC Areas for the Static-99 and Static-2002.....	39
Table 9: Meta-Analysis Results.....	42
Table 10: Characteristics of Participants in Study 2 (N = 38).....	51
Table 11: Reasons People Use the Static-99 (N = 38).....	55
Table 12: Advantages of the Static-99 (N = 38).....	56
Table 13: Disadvantages of the Static-99 (N = 32).....	57
Table 14: Overall Satisfaction With the Static-99.....	59
Table 15: Desired Advantages of the Static-2002 (N = 33).....	61
Table 16: Hesitations About Using the Static-2002 (N = 26).....	62

Abstract

The Static-99 (Hanson & Thornton, 2000) is one of the most commonly used and researched risk assessment instruments for sex offenders. Recently, the Static-2002 (Hanson & Thornton, 2003) was developed to increase conceptual clarity, enhance scoring consistency, reduce counter-intuitive scorings, and increase predictive accuracy. Two studies examined the following questions: is the Static-2002 more accurate than the Static-99, and are people willing to use it? The first study aggregates raw data from eight samples from Canada, the U.S., and the U.K. ($N = 3,520$), including offenders from forensic treatment centers, community supervision, and federal prisons. Meta-analysis found that the Static-2002 predicts sexual, violent, and any recidivism significantly better than the Static-99. Also, cut-off scores were developed to create five risk categories for the Static-2002. In the second study, 38 Static-99 users were interviewed. Qualitative analysis suggested that practical reasons (e.g., ease of use) for using an instrument were most important, but empirical issues (e.g., predictive accuracy) were not far behind, and that users may be willing to use the Static-2002 if it is more predictive or if it has other advantages (e.g., easier to use). While the Static-2002 is more predictive than the Static-99, additional factors affect whether people will use it.

Acknowledgments

Firstly, I'd like to thank Karl Hanson for his time, support, patience, and humour. I appreciate your commitment not only to conveying the material, but to ensuring that I understand it and get as much from the experience as possible.

For my co-supervisor, Ralph Serin, I thank you for your assistance. Your helpful editing and prompt response to my concerns enabled my thesis to go about as smoothly as a thesis can go....

My heartfelt thanks go to all the researchers who allowed me to access their data for this project: Karl Hanson, David Thornton, Jan Looman, Alexandra Boer, Andrew Haag, Calvin Langton, Jean Proulx, and Leigh Harkins. Most people put up with a lot of emails from me and I am very grateful for their cooperation. It is encouraging to know that so many researchers remember what it's like to be a student and are willing to help one out.

Thanks to Guy Bourgon for his valuable input and for being a sounding board when I'm trying to work things out. Thanks to my mom and dad for their encouragement. And lastly, to my husband, thanks for your support and for making everything fun.

Introduction

Sexual offending is a serious offence with profound negative consequences for the victim (Paolucci, Genuis, & Violato, 2001; Resick, 1993). Even though sexual recidivism rates are lower than the public generally believes (Harris & Hanson, 2004), the consequences of these offences make it understandable that there is a strong public desire to find methods of effectively managing sex offenders. Such methods require identifying those offenders more likely to sexually reoffend, and considerable research has focused on this area. For example, child molesters with unrelated male victims sexually recidivate at higher levels than incest offenders (Hanson & Bussière, 1998; Harris & Hanson, 2004). Additionally, recidivism is higher for younger offenders (Hanson, 2006; Hanson & Bussière, 1998; Harris & Hanson, 2004) unmarried offenders (Hanson & Bussière, 1998), those with prior sexual offences (Hanson & Bussière, 1998; Harris & Hanson, 2004), and those who exhibit sexual deviancy (Hanson & Morton-Bourgon, 2004). Cumulatively, this research has identified a variety of factors related to sexual reoffending. Because of the profound consequences of sexual offending, it is desirable to combine this information in an attempt to predict which sexual offenders are more likely to reoffend, and to use this information to assist in the management of these offenders.

Risk assessment is evaluating the likelihood of future criminal behaviour, a task that has become embedded in our criminal justice system. Its consequences for public safety, and particularly for the offender, are substantial. There are numerous opportunities for risk assessment throughout the criminal justice system. During sentencing, particularly if the judge is contemplating a non-custodial sentence, he or she may want to

know the offender's risk for reoffending. If an offender is sentenced to prison, a risk assessment will be an important part of their security classification. When the offender is eligible for some form of conditional release, the parole board will evaluate their risk to reoffend. Whether the offender is in prison or on parole or probation, they will likely encounter some form of intervention or treatment. If the principles of effective correctional treatment espoused by Andrews and Bonta (2003) are adhered to, a risk assessment will determine the intensity of the intervention, with higher risk offenders receiving more intensive services and more frequent contacts with a parole or probation officer. Clearly, risk assessment is a pervasive component of the criminal justice system.

Given this reliance on risk assessment, and the implications for public safety and for the offender, it is important to ask questions about this growing practice, such as the following: what is the history of risk assessment? What are the different methods of risk assessment? Which method is most accurate? These questions will be explored further, with a particular emphasis on the prediction of violent and sexual offences.

Ultimately, for an effective risk assessment, two conditions must exist. Firstly, there must be a valid instrument with demonstrated predictive accuracy. Secondly, the practitioners who conduct risk assessments must accept the instrument and agree to use it. This study will examine the accuracy of the Static-2002, which is an instrument designed to assess risk of recidivism for sexual offenders, and has evolved from a previous measure, the Static-99. Additionally, the study will examine whether practitioners are willing to use the Static-2002.

Risk Assessment: History and Various Methods

One of the early milestones in the history of violence risk assessment is the case of the Baxstrom patients. After a 1966 U.S. Supreme Court ruling, 966 patients deemed criminally insane and too dangerous to be released from the hospital, were released or moved to reduced security (Webster, Douglas, Eaves, & Hart, 1997). Of 98 patients followed up for four years, only 20 were rearrested (Andrews & Bonta, 2003). Furthermore, the majority of rearrests were for minor crimes such as intoxication and vagrancy (Webster et al., 1997), and only two were reconvicted for a violent offence (Quinsey, Harris, Rice, & Cormier, 1998), demonstrating that these offenders were not as dangerous as predicted. This case threw into question the validity of violence prediction and it was often concluded that the accurate prediction of violent behaviour was simply not feasible (Hanson, 2005).

In the decades since the Baxstrom case, a radical transformation has occurred. Research on the prediction of violence has flourished, scores of risk assessment instruments have been developed, and validations have found that future violence can be predicted with at least moderate accuracy (Hanson, 2005). Different approaches to risk assessment have emerged over the past decades and their comparative accuracy has been rigorously evaluated and debated, making this a very contentious field.

Bonta (1996) characterizes the development of risk assessment instruments in three generations. The first generation consists of unstructured professional judgment, where assessments are subjective. Typically the clinician interviews the offender to gather information and then combines the information into a meaningful risk assessment.

The weakness in this method is that it relies on personal discretion and lacks accountability, consistency, fairness, and replicability (Bonta, 1996).

The second generation of risk assessment is seen as the opposite of the first. Typical instruments in this generation are empirically based (as opposed to theoretically derived) and consist primarily of static factors (Bonta, 1996). Static factors are unchangeable, historical factors, such as criminal history variables. In these instruments, items are often scored either with a 0-1 dichotomy (absent-present) or with a specified weighting determined by the strength of the item's relationship to recidivism (Bonta, 1996). These instruments are commonly referred to as actuarial. The weakness in this generation is that the focus on static (historical) factors represents a limited view of risk predictors. It precludes an identification of areas to target in treatment to reduce risk and cannot reflect treatment changes or rehabilitation efforts (Bonta, 1996).

Since the early 1980s, a third generation has evolved from the second. Beyond simply predicting the level of risk, this generation identifies criminogenic needs (Bonta, 1996). Criminogenic needs are dynamic, meaning that they can change (e.g., attitudes), and if changed, they can alter the likelihood of reoffending (Andrews et al., 1990). Therefore, these instruments examine both static and dynamic risk factors, making them sensitive to changes. There is also a tendency for these instruments to have a stronger basis in theories of offending (Bonta, 1996). However, similar to the second generation, the third generation also mainly consists of actuarial instruments, whereby scoring rules tend to be explicit.

Recently, Andrews, Bonta, and Wormith (2006) have suggested that the field has reached a new phase: the fourth generation of risk assessment. While the third generation

identified criminogenic targets, the fourth generation provides a comprehensive guide for human service delivery that spans from intake through to case closure.

This characterization of the four generations is a rough sketch. There are further nuances and controversies in the history of risk assessment. One of the main controversies pertains to the status of structured professional judgment (SPJ) in this model. SPJs are risk assessment schemes that contain a list of risk factors (usually both static and dynamic); however, the combination of these items into an overall evaluation of risk is left to the judgment of the clinician. In other words, they are guidelines (based on scientific literature) to assist in clinical judgment (Boer, Wilson, Gauthier, & Hart, 1997). The proponents of SPJ argue that clinical predictions should have a prominent place in risk assessment because actuarial approaches are founded on statistical averages and as such, may not always be appropriate in individual cases (Webster et al., 1997). Although actuarial approaches are a useful starting point, the final prediction should reflect clinical opinion (Webster et al., 1997).

Many researchers, however, dismiss the claims that SPJ is a stronger method (Andrews & Bonta, 2003; Bonta, 2002; Hanson & Morton-Bourgon, 2004; Quinsey et al., 1998). Although some argue that SPJ is where risk assessment should evolve, Andrews et al. (2006) classify SPJ as merely a variation of the first generation. Quinsey et al. (1998) argue against combining actuarial methods with clinical judgment. Their strong opposition to SPJ is evident in their statement that “actuarial methods are too good and clinical judgment too poor to risk contaminating the former with the latter” (p. 171). And so, the debate about the utility of SPJ continues.

Hanson and Morton-Bourgon (2004) have added to the classification of risk assessment instruments. Similar to Bonta, they distinguish between clinical and actuarial approaches, but they also subdivide both actuarial and SPJ methods according to a conceptual-empirical distinction. Essentially, the items included in a risk assessment instrument are selected based on either their empirical association with recidivism, or on a theory of criminal behaviour. Although this distinction is useful in understanding how risk assessment instruments were developed, it may be overly simplistic because some measures (and possibly the better ones) may have empirical *and* conceptual backgrounds. The distinction also becomes more difficult to apply for instruments where specific items have a strong empirical basis but are also a key component of theory (e.g., the LSI).

Which Method Is More Effective?

Although risk assessment can be conceptualized as an evolution with four generations, most of the discussion surrounding the accuracy of risk assessment has focused on the distinction between clinical judgment and actuarial estimates. Because multiple studies are more informative than a single study, this discussion will focus on meta-analytic findings, both for psychology in general, and for predicting criminal recidivism.

A meta-analysis conducted by Grove, Zald, Lebow, Snitz, and Nelson (2000) examined clinical versus actuarial prediction in psychological research. They analyzed 136 studies that included at least one subjective judgment and at least one judgment made through a mechanical prediction scheme (an actuarial measure). All studies predicted a human behaviour outcome (as opposed to non-human events). Most studies were from the domain of medicine (51) and clinical-personality (41). Other domains included

education (18), forensic (10), financial (5), and other (11). On average, the results modestly favoured actuarial prediction, with approximately half of the studies indicating an advantage for actuarial methods, approximately half showing no difference, and with only eight studies (6%) finding that clinical methods outperformed actuarial. Interestingly, there was also a trend ($p < .07$) for findings in medical and forensic settings to show a stronger advantage for actuarial prediction compared to other fields. These results suggest that actuarial methods are superior.

In a meta-analysis examining the prediction of violence, Mossman (1994) examined 58 datasets from 44 published studies and found that clinical methods were moderately predictive ($ROC = .67$), but actuarial methods demonstrated greater predictive accuracy ($ROC = .71$).

Additionally, Bonta, Law, and Hanson (1998) conducted a meta-analysis of recidivism predictors for mentally disordered offenders. While not the primary purpose of their meta-analysis, they examined the comparative accuracy of objective risk assessments (actuarial) versus clinical judgment. For the prediction of general recidivism, six studies of objective risk assessment yielded a significant correlation (r) of .39, while five studies examining clinical judgment produced a significant correlation of .11. However, when two outliers were removed, the correlation for clinical judgment was reduced to .03 and was no longer significant. For violent recidivism, eight studies of objective risk assessments had an average correlation of .27, or .30 when one outlier study was removed. Of only three studies examining clinical judgment, the average correlation with violent recidivism was .09. Regardless of outcome variable (general or violent recidivism), actuarial methods consistently outperformed clinical judgment.

Furthermore, these results indicate a larger gap between the accuracy of clinical and actuarial methods than found in Mossman's (1994) meta-analysis.

Meta-analyses in the area of psychology, violence prediction, and recidivism prediction in mentally-disordered offenders have all found that actuarial methods outperform clinical judgment, with some meta-analyses showing stronger differences than others. With this general trend established, it is now possible to look more directly at the state of risk assessment for sexual offenders.

Which Method is More Effective With Sex Offenders?

Meta-analytical studies of sex offender risk assessment have largely been carried out by R. K. Hanson, through cumulative publications. The meta-analysis was first published in 1998 (Hanson & Bussière, 1998), updated in 2004 (Hanson & Morton-Bourgon, 2004), and again in 2007 (Hanson & Morton-Bourgon, 2007). Only the 2007 meta-analysis will be considered here because it includes more studies and the earlier meta-analyses are subsumed under it.

Hanson and Morton-Bourgon (2007) examined four categories of risk assessment instruments: conceptually (theoretically) derived actuarial, empirically derived actuarial, SPJ, and clinical judgment. Furthermore, risk assessments were divided based on whether they were developed for the prediction of sexual, violent, or any recidivism. Additionally, three recidivism outcomes were used: sexual, violent (including sexual) and any (including violent and sexual). Results are summarized in Table 1, where d refers to the average effect size (Cohen's d), weighted by the inverse of the variance, and k is the number of studies in that category. Outlier studies are excluded from this table. A study

Table 1

Results from Hanson and Morton-Bourgon's (2007) Meta-Analysis

Form of Risk Assessment	Recidivism Outcome Criteria					
	Sex		Violence		Any	
	<i>d</i>	<i>k</i>	<i>d</i>	<i>k</i>	<i>d</i>	<i>k</i>
Designed for sexual recidivism						
Empirical-actuarial	.70	55	.52	30	.56	25
Conceptual-actuarial	.66	22	.55	9	.53	13
Structured professional judgment	.42	5	.31	3	.24	8
Designed for violent recidivism						
Empirical-actuarial	.54	12	.85	10	.79	6
Conceptual-actuarial	.22	4	-	-	.36	3
Designed for any recidivism						
Empirical-actuarial	.52	4	-	-	1.13	3
Conceptual-actuarial	-	-	-	-	1.08	3
Unstructured professional judgment	.43	9	.18	4	.22	8

Note. Categories are left blank when there were less than three findings.

was considered an outlier if it was the most extreme value and it contributed at least 50% of the total variance.

For the prediction of sexual recidivism, actuarial methods, regardless of whether they were conceptually or empirically derived, and regardless of whether they were

designed to predict sexual, violent, or any recidivism, outperformed both SPJ and clinical judgment, with one exception. The four studies examining conceptually derived actuarial instruments designed to predict violent recidivism were less accurate than SPJ and clinical judgment in predicting sexual recidivism. For the prediction of violent recidivism, actuarial methods, regardless of whether they were empirical or conceptual, or designed for sexual or violent recidivism, outperformed both SPJ and clinical judgments. The same trend emerged for the prediction of any recidivism. Overall, actuarial methods consistently outperformed SPJ and unstructured clinical judgment. SPJ and clinical judgment performed equally well for predicting sexual and any recidivism; however, SPJ had greater accuracy in predicting violent recidivism.

Overall, the findings from meta-analyses in multiple fields indicate that actuarial risk assessment methods are the most accurate. Based on these findings, actuarial risk assessment instruments should form the basis of effective risk assessment practices for sex offenders. This review will now move to a more specific discussion of two instruments for predicting the risk of sexual recidivism for sexual offenders: the Static-99 and the Static-2002.

Development of the Static-99

The Static-99 is an empirically derived actuarial risk assessment tool designed to predict sexual recidivism (Hanson & Thornton, 2000). It was developed by combining two actuarial risk assessment instruments for sex offenders: the RRASOR and the SACJ-Min (Hanson & Thornton, 2000). The RRASOR (Rapid Risk Assessment for Sexual Offence; Hanson, 1997) was developed in Canada by R. Karl Hanson as a short and easy-to-use risk assessment. From the meta-analytic research on the predictors of sexual

recidivism (Hanson & Bussière, 1998), four items were chosen based on their contribution to a regression equation (Hanson & Thornton, 2000). The four items are prior sex offences, any unrelated victims, any male victims, and age less than 25.

The SACJ-Min (Structured Anchored Clinical Judgement – Minimum) was developed by David Thornton in the U.K. through exploratory analyses on multiple U.K. datasets (Hanson & Thornton, 2000). The original version, known as the SACJ, consisted of multiple steps with risk categories increasing or decreasing based on the existence of protective or aggravating factors. The first two steps included static items and the third step focused on dynamic factors, mostly relating to treatment performance. The SACJ-Min is a later version referring to the minimum amount of information needed to score the instrument. The first step remains the same, but the second has fewer items, and the third step was removed. The items in the SACJ-Min are the following: any current sexual offences, any prior sex offences, any current nonsexual violent offences, any prior nonsexual violent offences, four or more prior sentencing occasions, stranger victims, male victims, never married, and noncontact sex offences.

The RRASOR's four items target sexual deviancy, while the SACJ-Min includes sexual deviancy but also has a strong focus on criminal history. Since the two measures were different enough in focus, the developers of both measures combined them to determine if the resulting combination could perform better than either of the two scales individually (Hanson & Thornton, 2000). The resulting combination was called the Static-99 (Appendix A) because it consists of static risk factors and because the version was developed in 1999, with the possibility of modifications at a later point. In the development study, Hanson and Thornton (2000) combined three Canadian samples and

one U.K. sample ($n = 1,208$). The Static-99 predicted sexual recidivism significantly better than either the RRASOR or the SACJ-Min (ROC of .71, compared to .68 for the RRASOR and .67 for the SACJ-Min). For violent recidivism (including sexual), the ROC area for the Static-99 was .69, which was significantly better than .64 for both the RRASOR and SACJ-Min.

After combining the RRASOR and the SACJ-Min into the Static-99, there were ten items in the measure (see Table 2). All items are coded as either a 0 or a 1, except for prior sexual offences which is scored as 0, 1, 2, or 3. Total scores (obtained by summing all the items) can range from 0-12. Based on an offender's total score, they are placed in one of four risk categories: low (0-1), moderate-low (2-3), moderate-high (4-5), and high (6+). Advantages of the Static-99 are that it is easily scored and can be administered without an interview with the offender (Harris, A., Phenix, Hanson, & Thornton, 2003). While it is designed to be easy to score, the authors still recommend training before it is used. However, while some instruments can only be coded by psychologists, the Static-99 can be used by parole and probation officers, psychologists, treatment providers, and even police officers (Harris, A. et al., 2003).

The revised coding rules (Harris, A. et al., 2003) detail how and on whom the Static-99 should be used. It also provides 5, 10, and 15 year sexual and violent recidivism rates for each score on the Static-99 (with all offenders scoring 6 and above grouped in one category, due to small numbers), which enables someone conducting an assessment to provide recidivism estimates. According to the coding manual (Harris, A. et al., 2003), the Static-99 can be used to estimate sexual and violent recidivism for adult males who have been charged with or convicted of at least one sexual offence. It is inappropriate to

Table 2

Static-99 Items

Static-99 Item	Included in RRASOR?	Included in SACJ-Min?
Prior sex offences	Yes	Yes
Four or more prior sentencing dates	-	Yes
Any convictions for a non-contact offence	-	Yes
Index non-sexual violence	-	Yes
Prior non-sexual violence	-	Yes
Any unrelated victims	Yes	-
Any stranger victims	-	Yes
Any male victims	Yes	Yes
Young (less than 25)	Yes	-
Single (never lived with lover for 2+ years)	-	Yes

use the Static-99 on female sex offenders, young offenders (less than 18 at time of release), or for offenders whose only sexual offence(s) has/have been category B offences. Category B offences refer to illegal sexual or indecent behaviour where the parties are consenting or there are no specific victims. Examples include consenting sex in public places, possession of child pornography, urinating in public, and prostitution-related offences.

Use and Replication of the Static-99

Since its development, the Static-99 has quickly become one of the most widely used and widely replicated risk assessment instruments for sexual offenders. A 2002 nation-wide survey of sex offender treatment providers in the United States documents the widespread use of the Static-99 (McGrath, Cumming, & Burchard, 2003). Of the 520 surveys from community treatment programs for adult male sex offenders, 63% reported the use of at least one risk assessment instrument for sex offenders. The most commonly used instrument was the Static-99 (54%), followed by its predecessor, the RRASOR (35%). The third most commonly used instrument was the MnSOST-R, used by 20% of treatment providers. Similar trends were found for the 93 residential programs. This rapid adoption is particularly impressive given that the survey was conducted only two years after the Static-99 was published. The Static-99 is also used in countries as diverse as Israel, Singapore, and Taiwan (personal communication with Karl Hanson, 2007). Its use is even mandated in the U.S. jurisdictions of Virginia, California, Massachusetts, and Georgia (personal communication with Karl Hanson, 2007). In Canada, probation officers in British Columbia and Newfoundland are required to use the Static-99, and Ontario is also moving towards a mandatory policy (personal communication with Karl Hanson, 2007).

Not only is it one of the most commonly used, but the Static-99 is also the most researched risk assessment instrument for sex offenders. In Hanson and Morton-Bourgon's (2007) meta-analysis, there were 42 replications of the Static-99, which was substantially more than any other instrument. The RRASOR had 28 replications, and all other instruments had less than ten. Before discussing specific replications of the Static-

99, a statistical note is necessary. To maintain consistency of metrics, all effect sizes will be reported as ROC values. Most studies report ROCs; however, for those that do not, they were calculated using basic statistical formulae (Hanson & Morton-Bourgon, 2004; Rice & Harris, 2005). The meaning of ROC values is discussed in greater detail in the methods section.

Many Static-99 replications have been conducted in Canada, the United States, and the United Kingdom. However, additional replications have been conducted in Sweden (Langstrom, 2004), New Zealand (Beggs & Grace, 2005), Belgium (Ducro & Pham, 2006), Holland (de Vogel, de Ruiter, van Beek, & Mead, 2004), and Germany (Stadland et al., 2005). Replications have been reported in a variety of settings. Effect sizes from psychiatric settings have varied from .62 to .71 (de Vogel et al., 2004; Ducro & Pham, 2006; Harris, G. et al., 2003; Nunes, Wexler, Firestone, & Bradford, 2002). Considerable replications on prison samples have yielded more variability in effect sizes, with a range of .57 to .92 (Beggs & Grace, 2005; Brown, 2003; Craig, Beech, & Browne, 2006; Epperson, 2003; Friendship, Mann, & Beech, 2003; Harris, A. et al., 2003; Hood, Shute, Feilzer, & Wilcox, 2002; McGrath, Hoke, Livingston, & Cumming, 2001; Langstrom, 2004; Saum, 2005; Ternowski, 2004; Thornton, 2002). Additionally, some replications have examined probation samples and found effect sizes of .70 (Craissati, Webb, & Kenn, 2005) and .81 (Epperson, 2003).

While the developers of the Static-99 recommend against its use with juvenile sex offenders, three studies have examined it with this population. Poole, Liedecke, and Marbibi (2000) used the Static-99 with 49 juvenile sex offenders in Texas with a three year follow-up and found a large effect size (ROC = .95), but they raised the ethical issue

that the Static-99 may disproportionately designate juveniles as high-risk offenders because most would score points for being young and for not living with a lover for at least two years. However, their concerns seem somewhat unfounded. Their sample had a mean Static-99 score of 3.4, which is not much higher than the average score of 3.2 in the development sample (Harris, A. et al., 2003). In a U.K. replication with 77 offenders, Beech (2005) found an effect size of .87 with a follow-up of at least five years. Again, the juveniles do not appear to be disproportionately classified as higher risk, as only 1 out of the 77 offenders scored above a six. In the third replication, Morton (2003) found a moderate effect size ($ROC = .63$) with 80 juveniles from a treatment program in Ontario, and her sample obtained scores ($M = 3.7$) only slightly above the average of 3.2. This preliminary research has found effect sizes between .63 and .95, suggesting that the Static-99 may be useful with juvenile sex offenders.

Other replications with more specialized samples have added to the knowledge concerning what populations are appropriate for the use of the Static-99. The Static-99 is predictive with both child molesters and rapists (Bartosh, Garby, & Lewis, 2003; Ducro & Pham, 2006). However, one study suggests that it may not be useful for non-contact sex offenders ($ROC = .39$; Bartosh et al., 2003), although the small sample size ($n = 17$) limits confidence in those results. Tough (2001) replicated the Static-99 with 76 sex offenders displaying developmental disabilities with significant cognitive deficits. In this population, the effect size was small ($ROC = .56$). In another study, Montana and Thompson (2003) used the Static-99 with a sample of 108 priest sex offenders and found moderate predictive accuracy ($ROC = .69$). While its accuracy with non-contact offenders appears questionable, and the accuracy with developmentally disabled

offenders was small, the Static-99 seems to be robust with offenders from multiple countries, multiple settings, juvenile sex offenders, rapists, child molesters, and even priest sex offenders.

As indicated earlier, Hanson and Morton-Bourgon's (2007) meta-analysis included 42 replications of the Static-99. The average effect size is moderate (ROC = .69). Table 3 shows that the Static-99 performs reasonably well compared to other risk assessment instruments. Results for the Static-2002 will be discussed after a description of its development. While effect sizes vary, the instruments are not significantly different from each other. An interesting finding was that the variability among the effect sizes for the Static-99 was greater than the variability expected by chance, which indicates that the Static-99 is not predicting sexual recidivism consistently across different samples.

Development of the Static-2002

As discussed earlier, the Static-99 received its name because it was the 1999 version of the instrument, foreshadowing later modifications. The new iteration is the Static-2002. It was hoped that the Static-2002 would retain the advantages of the Static-99, namely that it would be applicable to numerous samples and could be easily and reliably scored without extensive information (Hanson & Thornton, 2003). The new instrument has four main goals behind its development (Hanson & Thornton, 2003).

- 1) To increase coherence and conceptual clarity – The Static-99 is a second-generation, atheoretical, and actuarial risk assessment instrument. It is part of the era referred to by Andrews and Bonta (2003) as “dustbowl empiricism,” (p. 238) meaning items are thrown in the mix purely based on their statistical association with recidivism and

Table 3

Hanson & Morton-Bourgon's (2007) Results for Individual Risk Scales

Measure	Mean ROC	<i>k</i>
SVR-20 (SPJ)	.78	3
Risk Matrix – 2000 sex	.72	6
SRA (Structure Risk Assessment)	.71	3
Static-2002	.71	5
MnSOST-R	.70	8
Beech Deviance	.69	3
Static-99	.69	42
SVR-20 (adding the items)	.68	8
SORAG	.68	8
RRASOR	.66	28
SIR	.64	4
VRAG	.64	7
SACJ-Min	.63	5
Unstructured Professional Judgment	.62	9
HCR-20 (adding the items)	.61	3
JSOAP	.59	4

Note. This table was adapted from data in Table 1 from Hanson & Morton-Bourgon (2007) with the authors' permission

without regard to the construct being measured. The Static-2002 was therefore designed with more conceptual clarity in mind.

- 2) To improve the consistency of the scoring criteria – Since the items and the relevant scoring rules came from two separate instruments (the RRASOR and the SACJ-Min), the Static-99 has inconsistent coding rules. Depending on the item, either convictions, charges, or sentencing occasions are counted. The Static-2002 was designed to reduce these inconsistencies and thereby hopefully facilitate training and increase inter-rater reliability.
- 3) Reduce counter-intuitive scorings – With the Static-99, there are rare cases where it is possible for an offender to be scored on the Static-99, commit a new sexual offence, and receive a *lower* score when scored again. This can occur if the first coding includes a point for index non-sexual violence, but the reoffence does not include a non-sexual violence conviction. For most cases, the incident previously counting as index non-sexual violence would become a point for prior non-sexual violence. But, if they already had a point for prior non-sexual violence, then their score upon reoffending can conceivably be lower. This counter-intuitive result is related to the Static-99's origins in dustbowl empiricism, whereby items were not considered in any depth beyond their predictive accuracy. The Static-2002 eliminates this coding possibility.

- 4) Increase the predictive accuracy – In the Static-2002, more items were added and definitions were changed. It was hoped that these modification might increase the predictive accuracy of the measure.

The Static-2002 has 13 items (see Table 4; see Appendix B for a coding form), and scores can range from 0-14. Unlike the Static-99, the Static-2002 currently does not have risk categories, nor does it have norms from which a risk assessor can derive specific recidivism probabilities. This is due to a lack of Static-2002 replications. Of the 13 items in the Static-2002, some are the same as in the Static-99 (either with the same or modified coding rules) and some are new items. Notably, two Static-99 items were not included in the Static-2002. The item regarding intimate relationships (never lived with a lover for at least two years) was removed because it was often difficult to score (Hanson & Thornton, 2003). Additionally, the item for non-sexual violence conviction during the index offence was also deleted because it was the source of the counter-intuitive scorings discussed earlier (Hanson & Thornton, 2003). Similar to the Static-99, items were included based on their empirical relationship with recidivism; however, increased attention was paid to scoring consistency (e.g., convictions versus sentencing occasions). Also, to increase coherency and make it clear what is being measured, the items in the Static-2002 are grouped into five main domains: age, persistence of sex offending, deviant sexual interests, relationship to victims, and general criminality.

In the development study, Hanson and Thornton (2003) compared the predictive accuracy of the Static-99 and Static-2002 by combining eight samples. However, this should be considered a rough comparison because each sample was missing information to code some of the items. In particular, none of the samples had information to code

Table 4

Static-2002 Items

Static-2002 Item	Included in Static-99		New
	Same Coding	Modified Coding	
Age			
Age	-	Yes	-
Persistence of sexual offending			
Prior sex offences	-	Yes	-
Juvenile arrest for sex offence	-	-	Yes
High rate of sex offending	-	-	Yes
Sexual deviance			
Non-contact convictions	Yes	-	-
Male victims	Yes	-	-
2+ victims, at least one unrelated	-	-	Yes
Relationship to victim			
Unrelated victims	Yes	-	-
Stranger victims	Yes	-	-
General criminality			
Prior arrest/sentencing occasions	-	Yes	-
Breach of conditional release	-	-	Yes
4 years free prior to index	-	-	Yes
Prior non-sexual violence	-	Yes	-

breach of conditional release and time free prior to index. Despite this missing data, the Static-2002 performed slightly better than the Static-99 for predicting sexual recidivism (ROC values of .71 and .70, respectively). Interestingly, similar to Hanson and Morton-Bourgon's later (2007) meta-analysis, the Static-99 showed significant variability across samples, whereas the Static-2002 was more consistent (variability was not significant). For the prediction of violent recidivism, the Static-2002 outperformed the Static-99 (ROC values of .71 and .69, respectively), and the difference was statistically significant.

The Static-2002 was designed to increase coherency and conceptual clarity, reduce inconsistent and counter-intuitive codings, and improve predictive accuracy, while remaining easy to code with basic information. Preliminary research suggests that it predicts sexual recidivism slightly better than the Static-99, and with more consistency across samples, and that it predicts violent recidivism much better than the Static-99. This improvement is likely due to the increase in items measuring general criminality.

Replications of the Static-2002

In Hanson and Morton-Bourgon's (2007) meta-analysis, there are only five replications of the Static-2002. These replications will be discussed further in the methods section, as all of the original datasets are included in this study. From the meta-analysis, the average effect size of the Static-2002 is .71 for the prediction of sexual recidivism, compared to .69 for the Static-99. For the prediction of violent recidivism, the Static-2002 showed a stronger advantage over the Static-99 (ROC = .71 versus .58). However, it should be noted that the average effect size of the Static-99 (ROC = .58) in the meta-analysis is substantially lower than it was in the Static-2002 development study

(ROC = .69). For both outcomes, they were not significantly different from each other. There were insufficient studies of the Static-2002 (less than three) to evaluate its ability to predict any recidivism.

Practical Implications: Implementation of Actuarial Tools

It is important to consider how actuarial instruments are used in applied settings. All the research can strongly indicate that one instrument is superior to another, but this is useless unless people in the field are willing to use the instrument. Andrews and Bonta (2003) note that there has been resistance to risk prediction instruments in the clinical field. Therefore, it is possible that clinicians are not interested in which instrument has empirical support. Perhaps they have different reasons for using or not using a risk assessment instrument.

Schneider, Ervin, and Snyder-Joy (1996) examined the reactions of probation and parole officers after the Wisconsin risk/need instrument became mandatory in their Oklahoma jurisdiction. A sample of 179 probation and parole officers (60% response rate) filled out a questionnaire regarding the implementation of the instrument. The attitudes of probation and parole officers were generally negative or neutral towards the risk/need instrument. Less than half (47%) believed the instrument was helpful, and two thirds believed that their own knowledge was superior to the risk assessment instrument. When asked why they use these instruments, only 24% agreed with the statement that research has shown these instruments to be effective (44% were neutral, and 32% disagreed). The strongest reason they gave for using the instrument was because negative evaluations were given if they failed to properly complete the instrument (83% agreed). However, despite their generally negative responses, when presented with the statement

that the system would be better off without the risk/need instrument, 52% disagreed, while 25% were neutral and 23% agreed. Overall, these probation and parole officers were not strongly supportive of the instrument, and most believed that their judgment was good enough, or better than the instrument.

These findings are puzzling given the rapid adoption of actuarial instruments that was evidenced in the 2002 survey of sex offender treatment providers. Schneider et al.'s (1996) study was conducted earlier in the history of the adoption of actuarial instruments so it is possible that perspectives have changed since then. However, it is also possible that these instruments are being mandated despite the resistance of practitioners. It is therefore important to examine how these instruments are perceived by those who use them.

Purpose of the Study

The Static-99 is the most validated risk assessment instrument for sex offenders, and it predicts sexual recidivism with moderate accuracy. The Static-2002 is designed to be more conceptually clear and consistent, less counter-intuitive, and more predictive than the Static-99. Preliminary research has suggested that it shows stronger predictive accuracy for sexual recidivism, and is particularly stronger in predicting violent recidivism; however, these results are based on relatively few replications.

The purpose of this study falls under two themes: empirical and applied. The empirical purposes are to compare the Static-99 and the Static-2002 and develop risk categories for the Static-2002. The applied purposes are to explore the perceptions of those who use the Static-99, and assess whether the disadvantages of the Static-99, as

cited by the developers, are perceived as such by the users, and to assess their willingness to use the Static-2002. Given the different purposes, two studies were conducted.

Study 1

The first study uses eight samples to compare the Static-99 and the Static-2002. While a meta-analysis has already been done on this topic (Hanson & Morton-Bourgon, 2007), this study includes more samples and also utilizes raw data which allows a broader range of analyses to be performed. Most importantly though, this comparison is more direct. Hanson and Morton-Bourgon's (2007) meta-analysis included 42 replications of the Static-99 and 5 of the Static-2002, meaning that different offenders and samples would be included in the two effect sizes. In this study, the two measures are completed for all offenders, so the instruments are compared on the same samples.

Hypotheses were based primarily on the results of the Static-2002 development study (Hanson & Thornton, 2003) and the Hanson and Morton-Bourgon (2007) meta-analysis. Hypotheses were that:

- 1) The Static-2002 will predict sexual recidivism with more accuracy than the Static-99.
- 2) The Static-2002 will predict violent (including sexual) recidivism with more accuracy than the Static-99.
- 3) The Static-2002 will predict any recidivism with more accuracy than the Static-99.
- 4) The Static-2002 will predict sexual, violent, and any recidivism more consistently across different samples than the Static-99.

- 5) No hypotheses were made regarding the optimal cut-off scores for risk categories on the Static-2002, as this analysis was exploratory.

Study 2

The second study utilizes interviews with users of the Static-99 to gauge their opinions regarding the original scale and their willingness to use the Static-2002. Hypotheses will be discussed after Study 1 is presented.

Study 1

Methods

Samples

To be included, a study required both Static-99 and Static-2002 scores for a sample of sex offenders, and data on sexual recidivism. The authors of all five Static-2002 replications included in Hanson and Morton-Bourgon's (2007) meta-analysis were contacted and asked if they would share their raw data for this study (see Appendix C for the letter that was sent). All authors agreed. Additionally, the co-supervisor for this study (Karl Hanson) had completed a new Static-2002 replication, which was also included. During the fall of 2006, I learned of three more Static-2002 replications that were either ongoing or had just been completed. These authors were contacted and two shared their data. The third author was in Denmark and was willing to share her unpublished data, but the information-sharing policies of Denmark could not be completed in time for this study. In total, eight samples were obtained.

Once the datasets were received, they were cleaned and merged. During this process, cases were deleted if:

- a) There was no follow-up (either because the information was not available or the offender had not yet been released from prison).
- b) More than one item on the Static-2002 was missing.
- c) ANY item on the Static-99 (except for Item 2: Single) was missing. Since the Static-99 has fewer items than the Static-2002, it was seen as necessary for all items to be included for a good assessment. It was decided that Item 2 of the Static-99 could be missing without strongly impacting the assessment.
- d) Static-99 or Static-2002 codings were illogical/impossible (e.g., coded as having a stranger victim on the Static-99 but not the Static-2002). When these cases arose, the authors of the datasets were contacted so the coding could be reviewed and fixed. However, in a few cases the authors could not find the source of the error, or the authors did not respond in time for this study.

Because of these criteria for the inclusion of cases, it is possible that the sample sizes reported in this study may vary slightly from the sample sizes reported in other publications using these samples.

Table 5 displays characteristics of the eight samples that were included. The total number of sex offenders is 3,520. Six samples are Canadian, one is American, and one is from the United Kingdom. Offenders were released from institutional settings in six samples. The multi-site U.K. sample contained a mixed group of offenders from institutional and community settings. In the Dynamic Supervision Project (DSP) sample, the offenders were on community supervision, although some may have served prison

Table 5

Sample Characteristics

Sample	<i>n</i>	Age (<i>SD</i>)	Offender Type: Rape/CM (%)	Average Follow- up (Years)	Recidivism Rate			Outcome Criteria
					Sexual	Any Violent	Any	
DSP (Canada)	706	42 (13)	36 / 50	3.5	8.1	16.4	27.9	Charge
Canadian Federal: RTC	362	38 (10)	43 / 39	6.9	11.6	31.2	51.4	Conv.
Canadian Federal: B.C.	299	41 (13)	41 / 55	12.6	8.7	23.4	48.5	Conv.
Canadian Federal: 1995	669	41 (11)	46 / 52	7.0	14.3	-	-	Conv.
Canadian Federal: Warkworth	364	40 (11)	37 / 53	5.6	11.3	24.5	39.3	Conv.
Canadian Federal: Quebec	487	40 (12)	37 / 46	4.5	6.6	15.4	24.2	Charge
Bridgewater: MTC (Massachusetts)	432	-	41 / 49	5, 10*	19.2	31.9	-	Charge
Multi-Site Treatment (UK)	201	43 (13)	82 / 10	10.4	13.9	20.4	36.8	Conv.
Total	3,520	41 (12)	42 / 47	6.7	11.5	22.5	35.7	

Note. Age refers to age at release.

* Follow-up period was fixed. 335 offenders had a ten year follow-up, and 97 had a five year follow-up.

sentences before their supervision period. All offenders are adult males, except for the DSP sample which contains six females. While ethnicity was rarely provided, it can be assumed that most offenders would be Caucasian, based on the demographics of the countries sampled. All eight samples include information on sexual recidivism, while seven studies also include violent recidivism, and six studies include any recidivism. Each sample will now be described individually.

1) Dynamic Supervision Project (Hanson, 2006; Harris & Hanson, 2003).

This prospective study followed offenders on community supervision between 2001-2005 in Canada, Alaska, and Iowa. However, only Canadian offenders are included in this study because the offenders from Alaska and Iowa did not have sufficient information to score the Static-2002. The original purpose of this study was to validate two dynamic risk assessment instruments. The sample consists mostly of provincial offenders, although there were a few federal offenders (sentenced to two or more years in prison). Exposure to treatment is unknown. From the original project, 171 probation officers participated; they coded the Static-99 and sent the data to the DSP researchers. The Static-2002 was coded in 2005 by myself and a M.A. student, using the original Static-99 coded by the probation officers, and using Canadian Police Information Centre (CPIC) records, maintained by the Royal Canadian Mounted Police (RCMP). CPIC records contain basic criminal history information; namely, date of conviction, offence title (according to the Canadian Criminal Code), the sentence, and the police jurisdiction that reported the incident. Information on charges that were stayed or for which the offender received an acquittal are inconsistently recorded on CPIC records, and offence details are not recorded. Recidivism information came from several sources: probation

officers, CPIC records, and police jurisdictions. Two cases were deleted because they had not been released yet, which left 706 offenders from this sample.

Inter-rater reliability for the Static-2002 was calculated for 25 cases. The Intraclass Correlation Coefficient (ICC) for single measures was .98. This is exceptionally high and should not be considered representative of the typical circumstances where the Static-2002 would be coded. In this study, all victim information was already identified and coded by the probation officer, so there was no need to apply and interpret the coding rules distinguishing between relatives, acquaintances, and strangers. Additionally, because the CPIC records were used, criminal history was coded artificially based on the Criminal Code offence, and not on the circumstances of the offence. These sources simplified the task and contributed to the high reliability.

2) Canadian federal offenders: Regional Treatment Centre (RTC) Ontario (Looman, 2006).

This sample followed offenders either assessed or treated at the RTC's Sex Offender Treatment Program at Kingston Penitentiary, a maximum security federal prison in Kingston, Ontario. Offenders were sentenced between 1970 and 2003, and released between 1990 and 2006. Of the initial sample of 455, 12% and 4% of offenders were serving life or indeterminate sentences, respectively. This, along with the setting (maximum security federal institution) suggests a relatively high-risk group of offenders. Recidivism information was collected using CPIC records. 85 cases were deleted either because there was no follow-up, the offender had been deported, or had died. Three cases were deleted because of coding inconsistencies. One case was deleted because of missing

information on the Static-99, and another case was deleted because of an inconsistency in recidivism information, leaving 362 offenders from this sample.

3) Canadian federal offenders: B.C. (Boer, 2003).

This study used archival data from the Offender Management System (OMS) maintained by Correctional Service Canada (CSC) to identify all federal male sex offenders released in B. C. whose Warrant Expiry Date (WED; the end of their sentence) was between January 1990 and May 1994. Of 410 offenders from that time period, 299 had follow-up data and sufficient information in their institutional files to score the Static-99 and Static-2002. Since many offenders are on some form of conditional release before their WED, offenders were released as early as 1986. Exposure to treatment is unknown. Recidivism information was collected using CPIC records, and Category B sexual offences were excluded from the definition of sexual recidivism. The Static-99 and Static-2002 have different coding rules for the item of prior non-sexual violence, but this item was coded in the same way for both measures in this sample. No cases had to be deleted, leaving 299 offenders from this sample.

4) Canadian federal offenders: 1995 WED (Haag, 2005).

This study used OMS records to identify and collect data on all federal sex offenders whose warrant expiry date was in 1995. The original purpose of this study was to assess whether treatment modifies the predictive accuracy of the Static-99 and Static-2002. Most of the sample (69%) received sex offender treatment in prison, 6% dropped out of treatment, and 25% did not receive any treatment. This study used a fixed follow-up period of seven years after the WED, but it should be noted that most offenders (75%) were in the community on some form of release before their WED. Recidivism

information was collected from CPIC records. Only sexual recidivism was coded. Five cases were deleted because of coding inconsistencies, leaving 669 offenders from this sample. Inter-rater reliability was assessed by having 66 cases coded by the main researcher and a psychologist from Bowden Institution. For the Static-99 and Static-2002, the inter-rater reliability (r) was .92 and .84, respectively.

6) Canadian federal offenders: Warkworth institution (Langton et al., 2007) .

This study followed sex offenders offered treatment at Warkworth Sexual Behaviour Clinic (WSBC) between 1989 and 2001. WSBC is located in a medium security federal institution in Ontario. Most (86%) of the offenders completed treatment, 8% dropped out, and 6% refused to enter treatment. Information to code the Static-99 and Static-2002 was extracted from Warkworth's clinical files, OMS records, and CPIC records. Recidivism information was coded from CPIC records. One case was deleted because of a coding inconsistency, leaving 364 offenders from this sample. Inter-rater reliability was assessed by having two researchers code 25 cases for the Static-99, and 58 cases for the Static-2002. Inter-rater reliability (r) for the Static-99 and Static-2002 was .88 and .90, respectively.

6) Canadian federal offenders: Quebec (Bigras, 2007).

This study followed a sample of Quebec federal offenders sentenced between 1995-2000. Assessment information was extracted from file data and interviews with offenders. Recidivism data was collected using CPIC records. 67 cases were deleted because there was no follow-up, and 48 cases were deleted because of missing information on the Static-99 or Static-2002, leaving 487 offenders from this sample.

7) Bridgewater: Massachusetts Treatment Center (MTC) (Thornton, 2006).

This study followed offenders who were either assessed or treated at MTC between 1959 and 1984. MTC is a treatment center for sexually dangerous persons. The risk measures were retrospectively coded from file data by raters who were blind to recidivism status. Recidivism information was obtained from four sources: Massachusetts Board of Probation records, Massachusetts Parole Board records, MTC Authorized Absence Program records, and FBI records. Recidivism was coded at five and ten year follow-ups. Five year follow-up data was available for 97 cases, and ten year data was available for 335 cases. Only sexual and violent recidivism were obtained. Sexual recidivism was defined as a serious (contact) sexual offense. This was the only dataset that did not include scores on each item; only total scores were available. This meant that the dataset could not be cleaned for errors. 112 cases with no follow-up were deleted, and 64 cases with no score for either or both the Static-99 and Static-2002 were deleted, leaving 432 offenders from this sample.

Inter-rater reliability was evaluated by having 232 Static-99 and 258 Static-2002 cases coded by two raters. Correlations (r) were .87 for the Static-99 and .89 for the Static-2002. As opposed to resolving inconsistencies between the two raters and establishing a consensus rating, the average of the two discrepant scores was entered as the offender's risk score. This was the only dataset where risk scores potentially went to one decimal point, so these scores were rounded. Rounding was done to the nearest even number to prevent artificial inflation of scores.

8) *U.K.: Multi-site treatment (Harkins & Beech, 2007).*

The original purpose of this study was to measure the effectiveness of sex offender treatment. This sample included offenders from three other studies in the United

Kingdom. The first sample consisted of offenders from the mid to late 1990s in the prison service's Sex Offender Treatment Programme. The second sample was drawn from a community treatment program in the early 1990s, and the third sample included offenders who underwent community treatment in the late 1990s in West Midlands. The intensity of the treatment received by the offenders varied, and treatment dropouts were retained in the sample. Recidivism data was collected from the Home Office Offenders Index (OI) and Police National Computer (PNC). 72 cases were deleted because of missing information on the Static-99 or Static-2002, and 49 cases were deleted because there was no follow-up information, leaving 201 offenders from this sample.

Missing Data

Table 6 summarizes the amount of missing information for Item Two (single) on the Static-99, and for all items on the Static-2002. Bridgewater cases are not included because the sample did not provide item scores. Because cases were deleted if there was more than one item missing, the overall amount of missing information is trivial. For the Static-99, only 1% of cases were missing Item 2 (single). Most items on the Static-2002 had no missing information. The item with the most missing data was for two or more victims under twelve, at least one of whom is unrelated. However, this item was missing in only 7 cases (.23%). The remaining items with missing information were age (1 case), breach of conditional release (4 cases), and years free prior to index (2 cases). For the total sample with item information ($N = 3,088$), only 47 cases (1.5%) had any information missing.

Table 6

Missing Data for Static-99 and Static-2002 Items

Measure	Item	Cases missing	% of total
Static-99	Single	33	1.07
Static-2002	Age	1	.03
	2+ victims <12, one unrelated	7	.23
	Breach of conditional release	4	.13
	Years free prior to index	2	.06
Total (<i>N</i> = 3,088)		47	1.52

Plan of Analysis

Effect sizes were coded as Receiver Operating Characteristic (ROC) areas, which is one of the most commonly used and commonly recommended effect sizes for risk prediction (Rice & Harris, 2005). ROC areas are typically preferred to correlations or Cohen's *d* because they are not affected by the base rate of the event (Rice & Harris, 2005), and in the case of sexual recidivism, the base rate is typically low. ROC curves plot the false positive rate by the correct positive rate for each possible cut-off score in the prediction scheme, which creates a curve (Swets, Dawes, & Monahan, 2000). ROC areas refer to the Area Under the Curve (AUC) and can vary between 0 and 1. If an ROC

value is .5, this is the level of prediction that would be expected by chance. An ROC value less than .5 indicates prediction at *less* than chance. ROC values between .5 and 1 indicate prediction exceeding chance levels, with numbers closer to 1 showing stronger prediction. Since .5 indicates chance level, then a confidence interval that does not include .5 demonstrates predictive accuracy significantly greater (or less) than chance. Another way to interpret the ROC area is that it represents the probability that a randomly selected recidivist will have a higher risk score than a randomly selected non-recidivist (Rice & Harris, 2005). Also worth noting is that ROCs of .56 correspond to a small effect size, while .64 reflects a moderate effect, and .71 reflects a large effect size, as these values correspond to effect sizes of .2, .5, and .8 using Cohen's *d* (Rice & Harris, 2005).

There are two general ways of aggregating data from different samples (Hanson & Thornton, 2003). The first way would be to combine all the datasets and analyse them as one sample, which ignores differences between samples. The other option is to use what is called a nested approach, where the results are examined across the samples. Meta-analysis is the most common nested approach. According to Hanson and Thornton (2003), when the samples are similar, the two approaches yield similar results. However, if there are differences across samples, the meta-analytic method is more appropriate. As will be discussed in the results section, there were significant differences between samples in the survival rates, so the nested approach (meta-analysis) was used. Standard meta-analytic methods were used as described in Hanson and Morton-Bourgon (2004). Average effect sizes (ROCs) were calculated by weighting each sample's ROC value by the inverse of its variance. To test whether the effect sizes across studies showed more variability than would be expected by chance, the *Q* statistic was used (Hanson &

Morton-Bourgon, 2004). A significant Q would mean that the measure showed significant variability in predictive accuracy across different samples. To test whether the Static-99 and Static-2002 differed in their level of predictive accuracy, Hanley and McNeil's (1983) test of correlated ROC areas was used (as cited in Hanson & Thornton, 2000).

Results

Risk Scores

Table 7 shows the mean Static-99 and Static-2002 scores for the eight samples. When combined ($N = 3,520$), the average Static-99 score was 3.4 ($SD = 2.2$). However, the average Static-99 score was significantly different between the samples, $F(7, 3512) = 92.94, p < .001$. Average scores from the samples ranged between 2.7 and 5.1. This variability is understandable because the samples came from different settings. For example, the highest average (5.1) came from the RTC sample at a maximum security federal institution in Canada, which would be more likely to have high-risk offenders. For the Static-2002, the average score across the eight samples was 5.0 ($SD = 2.6$). Again, there were significant differences between samples, $F(7, 3512) = 74.11, p < .001$, which can be explained by the different settings from which offenders were selected. Average scores across the samples range between 4.1 and 7.0. As expected based on Static-99 scores, the highest average ($M = 7.0$) was in the RTC sample.

Predictive Accuracy Across Samples

Table 8 summarizes the predictive accuracy (ROC areas) for the Static-99 and Static-2002 across all samples for sexual, violent, and any recidivism. All confidence intervals do not include .5, indicating that all Static-99 and Static-2002 predictions for

Table 7

Mean Risk Scores

Sample	<i>n</i>	Static-99		Static-2002	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
DSP	706	2.9	2.0	4.1	2.3
Canadian Federal: RTC	362	5.1	2.0	7.0	2.2
Canadian Federal: B.C.	299	3.4	2.3	4.6	2.5
Canadian Federal: 1995	669	2.8	2.0	4.5	2.4
Canadian Federal: Warkworth	364	3.6	2.1	5.4	2.4
Canadian Federal: Quebec	487	2.7	2.0	4.3	2.4
Bridgewater	432	4.4	2.2	6.6	2.5
U.K. Multi-Site	201	2.8	2.2	4.3	2.6
Total	3,520	3.4	2.2	5.0	2.6

sexual, violent, and any recidivism, were significant. For sexual recidivism, ROCs for the Static-99 varied between .60 and .77, while ROCs for the Static-2002 varied between .66 and .79. In two of the eight samples, the Static-99 performed slightly better than the Static-2002 (B.C.: ROC = .73 versus .71, respectively; Bridgewater: ROC = .67 versus .66, respectively). In the DSP sample, the two measures performed the same (.76). In the

Table 8

ROC Areas for the Static-99 and Static-2002

		Sexual Recidivism			Violent Recidivism			Any Recidivism		
Sample	<i>n</i>	ROC	95% C. I.		ROC	95% C. I.		ROC	95% C. I.	
DSP										
Static-99	706	.76	.70	.83	.73	.68	.78	.72	.68	.76
Static-2002	706	.76	.70	.82	.76	.72	.81	.76	.72	.80
Canadian Federal: RTC										
Static-99	362	.60	.52	.69	.59	.53	.65	.61	.55	.66
Static-2002	362	.68	.60	.76	.65	.59	.70	.67	.61	.72
Canadian Federal: B.C.										
Static-99	299	.73	.61	.84	.69	.62	.76	.72	.66	.77
Static-2002	299	.71	.61	.82	.72	.66	.79	.78	.73	.83
Canadian Federal: 1995 WED										
Static-99	669	.72	.66	.77	-	-	-	-	-	-
Static-2002	669	.73	.68	.78	-	-	-	-	-	-
Canadian Federal: Warkworth										
Static-99	364	.62	.54	.70	.61	.55	.67	.63	.57	.69
Static-2002	364	.69	.61	.77	.68	.62	.74	.68	.62	.74
Canadian Federal: Quebec										
Static-99	487	.63	.52	.73	.63	.57	.69	.63	.57	.68
Static-2002	487	.66	.56	.76	.67	.61	.73	.68	.63	.74

Table 8 continued

		Sexual Recidivism			Violent Recidivism			Any Recidivism		
Sample	<i>n</i>	ROC	95% C. I.		ROC	95% C. I.		ROC	95% C. I.	
Bridgewater										
Static-99	432	.67	.61	.72	.62	.57	.68	-	-	-
Static-2002	432	.66	.60	.72	.62	.56	.68	-	-	-
U.K. Multi-Site Treatment										
Static-99	201	.77	.68	.86	.73	.64	.81	.73	.66	.80
Static-2002	201	.79	.69	.88	.77	.69	.85	.76	.69	.83

other five samples, the Static-2002 predicted sexual recidivism with greater accuracy than the Static-99, with the largest difference found in the RTC sample, where the Static-99 and Static-2002 obtained effect sizes of .60 and .68, respectively.

For violent recidivism, only seven studies are included because the sample of 1995 federal offenders did not obtain violent outcome data. ROCs for the Static-99 varied between .59 and .73, while ROCs for the Static-2002 varied between .62 and .77. In one sample (Bridgewater), both the Static-99 and Static-2002 performed the same (.62). In the other six samples, the Static-2002 predicted violent recidivism better than the Static-99, with the largest differences found in the Warkworth sample, where the Static-99 and Static-2002 obtained effect sizes of .61 and .68, respectively.

For any recidivism, only six studies are included because the Bridgewater and 1995 federal offender samples did not obtain data on non-sexual, non-violent recidivism.

ROCs for the Static-99 varied between .61 and .73, while ROCs for the Static-2002 varied between .67 and .78. In all six samples, the Static-2002 predicted any recidivism better than the Static-99, with the largest differences found in the RTC and B.C. samples, where the values for the Static-99 and Static-2002 were .61 and .67, respectively, in the RTC sample, and .72 and .78, respectively, in the B.C. sample.

Note on Aggregating Results

As was discussed in the methods section, aggregating data can be done by either treating them as one sample, or performing a meta-analysis. Hanson and Thornton (2003) suggest that when there are differences between samples, it is preferable to use the meta-analytic approach. Cox regression, which controls for time at risk, revealed significant differences in recidivism rates across the samples (Wald = 20.3, $df = 6$, $p = .002$). Six degrees of freedom were used because one sample did not provide survival times. When controlling for pre-existing risk levels using the Static-99, differences between the datasets increased (Wald = 38.0, $df = 6$, $p < .001$). These differences were not easily explained by sample differences such as recidivism criteria, so it was decided that the samples should be analysed separately through meta-analysis.

Relative accuracy

The results of the meta-analysis are displayed in Table 9, which shows the average effect size weighted by the inverse of the variance of the samples, as well as the confidence intervals, the number of studies included, sample size, and Q statistic. The table also shows the difference between the Static-99 and Static-2002 for the three outcome variables, with confidence intervals, sample sizes, and variability across samples (measured by Q). All effect sizes have confidence intervals that do not include .5,

Table 9

Meta-analysis Results							
Recidivism Outcome	Measure	Weighted ROC	95% Confidence Interval		k	n	Q
Sexual							
	Static-99	.69	.67	.72	8	3,520	17.97*
	Static-2002	.71	.69	.74	8	3,520	10.12
Violent							
	Static-99	.66	.63	.68	7	2,851	20.63*
	Static-2002	.70	.68	.72	7	2,851	24.08*
Any							
	Static-99	.67	.65	.70	6	2,419	19.09*
	Static-2002	.73	.70	.75	6	2,419	17.83*
Difference Between Static-99 and Static-2002							
Sexual		.016	.001	.029	8	3,520	12.81
Violence		.036	.024	.048	7	2,851	12.15
Any		.049	.037	.060	6	2,419	13.71*

* $p < .05$, on a Chi-Square distribution with $k - 1$ degrees of freedom.

indicating that both the Static-99 and Static-2002 predicted sexual, violent, and any recidivism significantly better than chance.

In terms of comparative accuracy, the Static-2002 predicted sexual recidivism better than the Static-99 (.71 versus .69, respectively). The difference between the two measures was small (.016); however, the confidence interval (.001 to .029) of the difference score does not contain zero, so the difference is statistically significant.

Variability in the predictive accuracy across samples can be examined through the Q statistic, which is distributed on a Chi-Square distribution with $k - 1$ degrees of freedom. The Q was significant ($Q = 17.97, p < .05$) for the Static-99, indicating that the predictive accuracy across the eight samples showed greater variation than would be expected by chance. The Static-2002 did not show significant variability in predicting sexual recidivism ($Q = 10.12, p > .05$). For the difference between the measures, the difference scores across the eight samples did not show significant variability ($Q = 12.81, p > .05$).

For the prediction of violent recidivism, the Static-2002 had greater predictive accuracy than the Static-99 (.70 versus .66, respectively). The difference between the two measures was .036, with a confidence interval between .024 and .048, which was statistically significant. Both the Static-99 and Static-2002 showed significant variability in predictive accuracy ($Q = 20.63, p < .05$, and $Q = 24.08, p < .05$, respectively). However, the difference scores did not show significant variability ($Q = 12.15, p < .05$).

For the prediction of any recidivism, the Static-2002 had greater predictive accuracy than the Static-99 (.73 versus .67, respectively). The difference between the two measures is .049, with a confidence interval between .037 and .060, which was statistically significant. Both the Static-99 and Static-2002 showed significant variability

in predictive accuracy ($Q = 19.09, p < .05$, and $Q = 17.83, p < .05$, respectively), and there was also significant variability in the difference scores ($Q = 13.71, p < .05$).

Risk Categories for the Static-2002

Because there is no articulated scientific method for developing risk categories, two guiding principles were followed: it is desirable to have large numbers in each risk category (approximately 10% of the total sample, at a minimum), and also to have meaningful differences in the recidivism rates across categories. To identify possible cut-points that met the two guiding principles, risk categories that had been suggested by previous authors in Static-2002 replications were identified. Additionally, more cut-offs were identified based on a cross-tabulation of the frequencies of Static-2002 scores and recidivism (sexual, violent, and any). As a result, six plausible arrays of cut-off scores were identified. These plausible combinations were examined through ROC analysis. One combination of cut-off scores demonstrated the best predictive accuracy for sexual, violent, and any recidivism. This combination was therefore chosen and it has five categories of risk: low (0, 1, 2), low-moderate (3, 4), moderate (5, 6), moderate-high (7, 8), and high (9 and above). To enable the translation of risk scores into recidivism estimates, Appendix D displays the five and ten year sexual, violent, and general recidivism rates associated with each risk score. However, because there were inexplicable differences in recidivism rates across the samples, these estimates should be interpreted with caution.

Discussion

The first three hypotheses were that the Static-2002 would predict sexual, violent, and any recidivism with greater accuracy than the Static-99. All three hypotheses were

supported. While the Static-2002 showed significantly better predictive accuracy for all three outcomes, the difference for sexual recidivism was quite small (ROC = .71 versus .69), while the differences for violent (.70 versus .66) and any recidivism (.73 and .67) were larger. The greater difference in predictive accuracy for violent and any recidivism is likely because the Static-99 focuses almost exclusively on factors related to sexual reoffending, while the Static-2002 includes more items indicative of general criminality. The current findings are similar to the findings in the Static-2002 development study (Hanson & Thornton, 2003) and Hanson and Morton-Bourgon's (2007) meta-analysis. However, compared to the development study, the differences between the measures were larger in the current study. This could be because of more complete information in the present study, which will be discussed later.

The fourth hypothesis was that the Static-2002 would show less variability across samples in predicting sexual, violent, and any recidivism. This hypothesis was only supported for the prediction of sexual recidivism. The Static-99 showed significant variability in the prediction of sexual recidivism, while the Static-2002 did not. This is similar to the findings from the Static-2002 development study (Hanson & Thornton, 2003) and the previous meta-analysis (Hanson & Morton-Bourgon, 2007). However, for the prediction of violent and any recidivism, *both* the Static-99 and Static-2002 showed significant variability. These results are in contrast with the findings in the development study and the meta-analysis, where the Static-99 showed significant variability in the prediction of violent recidivism, and the Static-2002 did not (general recidivism was not examined in those two studies). It is unclear why the Static-2002 showed significant variability in these samples. It is possible that treatment effects in some samples mediated

the predictive accuracy of the measures; however, this is difficult to evaluate because few of the samples provide detailed information on the type and extent of treatment that the offenders may have received. It is also possible that the variability in accuracy was due to different jurisdictions, or other sample characteristics. More research should examine how treatment, jurisdiction, and sample characteristics (such as offender type) impact the predictive accuracy of these measures.

No hypotheses were made regarding the optimal cut-off scores for risk categories of the Static-2002. The analysis resulted in the creation of five risk categories, whereas the Static-99 had only four categories. Also, five and ten year recidivism probabilities were developed based on the eight samples in this study (Appendix D). These probabilities represent recidivism norms that practitioners can refer to when interpreting risk scores. Interestingly, the overall norms for sexual recidivism in this sample are lower than the norms in the Static-99 coding manual. In this sample of 3,520 offenders, the five year sexual recidivism rate is 11%, whereas in the Static-99 development sample ($n = 1,086$), offenders had an 18% sexual recidivism rate at five years. The lower norms in the current study cannot be due to risk scores because the average Static-99 score from both studies are very similar, with scores being slightly *higher* in the current sample ($M = 3.4$, compared to $M = 3.2$ in the development sample). The reasons for this apparent decrease in the norms for sexual recidivism are unclear. The offenders in the Static-99 development sample were released primarily in the 1960s, 70s, and 80s, while in the current study, the vast majority of offenders were released in the 1990s and later. Perhaps the difference is due to a general decline in recidivism rates, or the improvement of treatment and/or supervision methods, or maybe it is because the population is aging. Or

it is possible that the decline is illusory; perhaps the difference is due to more of the samples using conviction data as opposed to data on charges. Further research should examine if there is in fact a trend for declining recidivism rates, and investigate possible contributing factors.

There are numerous strengths to this study. Firstly, comparing the same offenders on both instruments offers a more direct comparison than examining one measure on one sample and the other measure on a separate sample, which is what many meta-analyses do. Additionally, the main strength is that eight different samples were included. Combining multiple samples results in a larger sample size, and can potentially eliminate sample characteristics as an alternative explanation for the findings. Multiple samples also enhance the generalizability of the findings to three countries and multiple settings.

An additional advantage of this study is the comprehensive information for the Static-99 and Static-2002 items. In the Static-2002 development study, none of the samples had sufficient information to code two of the items (breach of conditional release, and time free prior to index), and many samples had additional missing information. In that study, cases were included if the Static-99 had three missing items or less, and the Static-2002 had five missing items or less. This is still a substantial amount of missing information. In the present study, only 1.5% of cases had any missing information, and of that missing information, only a single item was incomplete. This trivial amount of missing information increases the confidence in the accuracy levels obtained from the measures.

There are also a few limitations to this study. Because the Static-99 and Static-2002 were coded in eight separate studies, it is impossible to determine the extent of

deviations from the coding manual. Particularly for the Static-2002, which only has a preliminary coding manual, it is difficult to assess the consistency of coding across the different samples. Additionally, the different samples used different information to code the measures and to define and assess recidivism. These differences could influence the results.

Despite these limitations, this study provided a rigorous comparison of the Static-99 and Static-2002. Future research is necessary to examine how other factors (e.g., treatment, jurisdiction, offender type) can influence predictive accuracy, and to determine if recidivism norms have changed over time. Also, future replications of the Static-2002 on specialized samples (e.g., developmentally delayed, internet sex offenders, juveniles) can build on current knowledge. The results of this study indicate that the Static-2002 predicts sexual, violent, and any recidivism with greater accuracy than the Static-99, and this difference is more noticeable for violent and any recidivism. These results suggest that the Static-2002 could replace the Static-99 in applied risk assessments. However, a valid and accurate instrument is only one condition necessary for effective risk assessment. As was mentioned earlier, the second condition is an audience that is willing to use it. The Static-2002 outperforms the Static-99, but the Static-99 is already in widespread use; consequently, it is necessary to examine whether practitioners are willing to make the change.

Study 2

In this study, users of the Static-99 were interviewed to assess their opinions regarding the original scale and their willingness to use the Static-2002.

Hypotheses

- 1) Individuals who use the Static-99 because it is mandated will show less favourable attitudes towards it than individuals who use it because of choice.
- 2) Practical reasons for using an instrument, as opposed to research evidence, will be seen as more important.
- 3) No hypotheses were made regarding whether users of the Static-99 would list the same disadvantages of the Static-99 that the developers listed when they developed the Static-2002.
- 4) No hypotheses were made regarding whether users of the Static-99 would consider using the Static-2002, and why they would consider using the Static-2002.

Methods

Participants

Thirty-eight participants completed a questionnaire about the Static-99 and the Static-2002 between November 2006, and January 2007. To participate, a person had to have used the Static-99 in either a research or applied capacity. Participants were recruited either by phone or by email through a snowball method. First, Karl Hanson (Static-99 developer) provided a list of names and phone numbers of people known to use the Static-99. These people were contacted and asked if they would participate. Those who participated were asked to recommend 1-2 more people suitable for the study. This method generated several participants; however, because many were acquaintances of Karl Hanson, well-established researchers were being disproportionately sampled. To recruit more people who use the Static-99 in applied settings and who do not use it by

choice, probation officers in Newfoundland and British Columbia who had participated in the Dynamic Supervision Project (one of the samples from Study 1) were contacted.

These jurisdictions were chosen because the Static-99 is mandated there. Because both snowball and convenience sampling methods were used, this study does not purport to offer a random sample of users of the Static-99. Instead, the aim is to sample a variety of user types.

Table 10 provides basic characteristics of the 38 participants. Participants were mainly from Canada (71%) and the U.S. (24%). Half (50%) were either psychologists or professionals (had obtained a PhD). Other participants included probation officers (32%), and other occupations, such as police officers or researchers. For the type of risk assessment user, the percentages add up to more than 100 because participants were permitted to indicate more than one description. Most of the participants used the Static-99 in some sort of applied capacity, such as in clinical or supervision settings (87%), while 26% used it for research purposes, and 24% were managers or policymakers. Half of the participants (50%) were mandated to use the Static-99, and half were not (50%); additionally, nine of the participants mandated that others use the Static-99.

Materials

Participants completed a questionnaire either by phone or email. The questionnaire is attached as Appendix E. It asks for information about their involvement in sex offender risk assessment, their opinion on the Static-99, whether they have heard of the Static-2002, and under what conditions they would be willing to use the Static-2002.

Table 10

Characteristics of Participants in Study 2 (N = 38)

	<i>N</i>	%
<i>Country</i>		
Canada	27	71.1
U.S.	9	23.7
Other	2	5.2
<i>Occupation</i>		
Psychologist/Professional	19	50.0
Probation Officer	12	31.6
Other/Unknown	7	18.4
<i>Type of User</i>		
Policymaker	9	23.7
Applied/clinical Assessment	33	86.8
Research	10	26.3
<i>Mandated to use Static-99?</i>		
Yes	19	50.0
No	19	50.0
<i>Method of Interview</i>		
Telephone	16	42.1
Electronic	22	57.9

Procedure

Most participants were contacted by phone, although some of the B.C. and Newfoundland probation officers were contacted via email. The script that was used when contacting potential participants is included with the questionnaire (Appendix E). Participants were told the purpose of the project and asked if they would consider completing the questionnaire. In total, 78 people were contacted and 38 participated, which is a 49% response rate. Initially, it was planned that all questionnaires would be administered over the phone. However, early on, some participants expressed a preference to be sent an electronic copy of the questionnaire to complete independently, so this was made an option. When participants chose this option, they were encouraged to contact me by phone or email if they required any clarification. Overall, 58% of participants elected to have the questionnaire sent electronically. The two procedures for questionnaire completion (telephone and electronic) will be described separately.

If participants agreed to do the questionnaire over the telephone, it was either completed immediately, or a mutually convenient appointment was established. If no one answered the phone, a message was left providing information about the study and asking the individual to call back. If they did not respond after three days, they were called again. If three messages were not returned, the potential participant was not pursued. When the questionnaire was administered, the informed consent was read to participants over the phone and they were asked to verbally consent. Debriefing was similarly done over the phone (see Appendix E for script). Including the informed consent and debriefing, telephone interviews typically lasted between 10-20 minutes.

If participants chose to complete the questionnaire independently, their email address was obtained and the questionnaire was sent to them. The informed consent and debriefing were included in the questionnaire package. When the questionnaire was sent back via email, it was printed and the email deleted to maintain the confidentiality of the participants. If it was sent via fax or mail, any cover pages or envelopes were destroyed. For the few participants who were initially contacted over email, the same script was used, and the questionnaire was attached. With all questionnaires that were sent via email, participants were encouraged to contact me if they had any questions or concerns, and they were also invited to schedule a telephone interview if they preferred. After a questionnaire was sent to a participant, it took anywhere from a few hours to over two months before the completed questionnaire was received. Participants were typically sent reminder emails once a week. However, if two reminders were not responded to, the participant was not contacted again.

The electronic questionnaires yielded less data about the participants' willingness to use the Static-2002 than the telephone interviews. The most likely reason is because the beginning of the section asks whether they have heard of the Static-2002 and the next two questions start with "if yes...", which indicates that if they had not heard of the Static-2002, they should skip those questions. However, many participants skipped the entire section on the Static-2002. It is difficult to determine whether this was done mistakenly or if they chose not to answer those questions. Ultimately, this resulted in greater missing information for that section.

Plan of Analysis

This study was primarily exploratory and cannot be considered a random sample of the population of Static-99 users. Instead, it is meant to provide some preliminary observations regarding how the Static-99 is viewed by risk assessment consumers, and to describe their willingness to use the Static-2002. For this reason, descriptive statistics and content analysis were used; no significance testing was done. Descriptive statistics were used for forced-choice items, and content analysis was used for open-ended questions. For the open-ended questions, all participants' responses were read through and themes were identified. Responses were then re-read and coded according to which themes were present in the response.

Results

Reasons for Using the Static-99

For all results, responses are only reported if they were mentioned by more than one participant. Table 11 lists the reasons participants use the Static-99. Percentages add up to more than 100 because participants could offer more than one reason for using the Static-99. The most commonly reported reason was because it is quick and/or easy to administer (55%). Other major reasons were because it is the policy at their workplace (50%), because of its predictive accuracy (42%), and because it has the most empirical support (40%). So while practical reasons (ease, policy) were the most commonly cited for its use, empirical reasons (accuracy, empirical support) were mentioned often.

Satisfaction with the Static-99

Because the participants use the Static-99 in different capacities, a variety of responses were given when asked what about the advantages of the Static-99 (see Table

Table 11

Reasons People Use the Static-99 (N = 38)

Reason	<i>n</i>	%
Quick/easy to administer	21	55
It is the policy at my workplace	19	50
Predictive accuracy	16	42
Has most empirical support	15	40
High reliability	12	32
I have been trained on it	11	29
It is popular/widely used	5	13
It has been recommended by others/predecessors	5	13
Can be used by non-psychologists	2	5

12). Percentages add up to more than 100 because multiple advantages could be listed. The majority of participants (68%) said it is quick and easy to administer. The next most common advantage was that it has the most research support or the most validations, which was mentioned by a quarter of participants (26%). Other commonly mentioned advantages were that it is predictive (24%), it is commonly used and has become a widely accepted standard (21%), and it is reliable (18%).

Regarding the disadvantages of the Static-99, six participants did not list any, leaving a sample size of 32. Disadvantages are listed in Table 13. The disadvantage cited by the most participants (41%) was that the Static-99 does not include dynamic risk

Table 12

Advantages of the Static-99 (N = 38)

Advantage	<i>n</i>	%
Quick/easy to administer	26	68
Most research support/most validations	10	26
Predictive/accurate	9	24
Commonly used/it is the accepted standard	8	21
Reliable	7	18
Assists in management of offenders	4	10
Has norms/numbers to compare to	4	10
Good coding manual	4	10
Does not require extensive information/interview to code	4	10
Good across jurisdictions	3	8
Designed particularly for sex offenders	2	5
Overall package good (Static, Stable, Acute)*	2	5
Provides estimates for sexual AND violent outcomes	2	5
Can contact developers if questions arise	2	5
Training program is good	2	5
It is an actuarial measure	2	5

*Karl Hanson has also developed risk assessment instruments that assess stable

Table 13

Disadvantages of the Static-99 (N = 32)

Disadvantage	<i>n</i>	%
No dynamic risk factors/not good for managing offenders	13	41
Missing important risk factors	9	28
Scoring criteria: harder than it looks/confusing/stupid	8	25
Predictive accuracy only moderate	4	12
Excludes sexual deviance (e.g., phallometric data)	3	9
Excludes category B offenders (e.g., internet offenders)	3	9
Results do not reflect professional judgment/goes against instinct	3	9
Excludes psychopathology (e.g., psychopathy)	2	6
Does not reflect treatment	2	6
Should not be the only thing considered	2	6
Excludes females	2	6

Factors and therefore is not good for managing offenders, or for intervening to change the risk level. Two other main issues were raised. Firstly, that the Static-99 is missing important risk factors (28%), although what these factors are was rarely mentioned. The other major issue (mentioned by 25%) was that the scoring criteria was in some way unsatisfactory, with reasons being that it is harder than it looks, can be confusing, or that some rules are even stupid. Of the remaining disadvantages cited by participants, most related to specific items that are not included in the measure (sexual deviance,

psychopathology, treatment gains) or specific offender types that the Static-99 is not appropriate for (category B offenders, female offenders).

Participants were also asked to indicate their overall satisfaction with the Static-99. These results are summarized in Table 14. Overall, satisfaction was very high: 84% of participants indicated that they were satisfied with the Static-99, and only 5% said they were not. The remainder (10%) were somewhat satisfied. When examined according to whether the participant is mandated to use the Static-99, differences emerge. While 95% of participants who used the Static-99 by choice were satisfied with it, only 74% of the mandated participants were satisfied.

To determine satisfaction with specific aspects of the Static-99, participants were asked to rate it on five features, using a scale from 1 (very poor) to 5 (very good). The five features were: easy to learn/train others, easy to administer, predictive accuracy, conceptual clarity, and consistency of scoring criteria. The Static-99 was rated

Table 14

Overall Satisfaction with the Static-99

Group	Yes, Satisfied		Somewhat Satisfied		Not Satisfied	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Mandated (<i>n</i> = 19)	14	74	4	21	1	5
Not Mandated (<i>n</i> = 19)	18	95	0	0	1	5
Overall (<i>n</i> = 38)	32	84	4	10	2	5

high on all items. The lowest rating was for predictive accuracy ($M = 3.8$), and the highest was for ease of administration ($M = 4.3$). The other items scored in between ($M = 4.0$ for easy to learn/train others, and consistency of scoring criteria; $M = 4.1$ for conceptual clarity).

Willingness to use the Static-2002

When asked about the Static-2002, 71% of participants indicated that they had heard of it, although only 22% claimed to know enough about it to form an opinion. When asked how much more accurate the Static-2002 would have to be compared to the Static-99 in order for them to consider using it, participants were given three options: 1) any difference, no matter how small; 2) noticeable difference, or; 3) substantially better. Of the 29 participants who answered this question, 38% said “any difference, no matter how small,” while 48% said “a noticeable difference,” and 14% said “substantially better.”

Participants were also asked if, given the same predictive accuracy, they would consider using the Static-99 if it had other advantages. Additionally, they were asked what other advantages they would be looking for. Thirty-three participants responded to this question. Responses are listed in Table 15. Overall, one third of participants said they would still consider using it if it was easier to score or administer. Other advantages would be if it included dynamic risk factors (18%), reflected additional risk factors (15%), had greater conceptual clarity (15%), or if the scoring manual was clear and consistent (12%).

Additionally, participants were asked if there were any factors that would make them hesitate about using the Static-2002 (see Table 16). Twenty-six responded, and the

Table 15

Desired advantages of the Static-2002 (N = 33)

Advantage	<i>n</i>	%
Easier to score/administer	11	33
Includes dynamic factors/areas to target	6	18
Reflects additional, important risk factors	5	15
Greater conceptual clarity	5	15
Scoring manual good/clear/consistent	4	12
Ease of training/learning	3	9
Includes category B offenders	2	6
Better account of age/maturation effects	2	6

most common hesitation listed was if it was not validated (23%). Another common concern, mentioned by 19% of participants, was that the Static-99 is widely accepted and has become part of a commonly understood language of risk assessment. Other reasons, listed by 15% of participants, were the following: financial/logistical reasons (the amount of time and money required to train people on the Static-2002), if it was less predictive, or if it was difficult or complicated to administer. Another 8% of participants said they would hesitate if it did not have norms or probability estimates to compare to.

Table 16

Hesitations About Using the Static-2002 (N = 26)

Hesitation Factor	<i>n</i>	%
If it is not validated	6	23
Static-99 is widely accepted, people are familiar with it	5	19
Financial/logistical issues (cost/amount of training)	4	15
If it is less predictive	4	15
If it is difficult/long/complicated to administer	4	15
If it does not have norms to compare to	2	8

Discussion

This study solicited the opinions and perceptions of the people who use the Static-99. When interpreting whether the hypotheses were confirmed, it should be noted that the data were analysed qualitatively. Conclusions should therefore be considered preliminary and interpreted with caution. Based partly on research by Schneider et al. (1997), it was hypothesized that individuals who were mandated to use the Static-99 would demonstrate less favourable attitudes towards it than those who used it by choice. This hypothesis was supported. However, even among those mandated to use it, satisfaction with the Static-99 was quite high (74%). So while satisfaction was lower amongst those mandated to use the instrument, the negative attitudes towards actuarial instruments found by Schneider et al. (1997) were not replicated in this study. There are many reasons why that could be the case. Schneider et al.'s (1997) study looked at a different risk assessment instrument (the

Wisconsin Risk/Need Instrument), so it is difficult to compare attitudes to different instruments. Additionally, attitudes were examined shortly after the instrument was implemented, whereas this study examined the attitudes of people who may have been using the Static-99 for several years, so it is possible that preliminary negative attitudes towards actuarial instruments have since been tempered. Or, because this study was conducted ten years after Schneider et al. (1997), the discrepant findings may reflect general changing attitudes towards actuarial risk assessment.

The second hypothesis was that users of the Static-99 would cite practical issues as being more important than empirical issues in influencing why they use particular instruments. This hypothesis appears to be supported. When asked about why they use the Static-99, the advantages of the Static-99, and what advantages (aside from predictive accuracy) they would be looking for in order to use the Static-2002, the most common responses consistently related to how quick and easy-to-use the instrument was. This demonstrates that practical reasons are among the most important for people who use risk assessment instruments in applied settings. However, while practical issues were frequently raised, empirical issues (predictive ability, validations, item content) were not far behind. This suggests that risk assessment users are seeking a balanced approach; they want instruments that are easy to administer and have evidence to support their use.

While no hypotheses were made, another purpose of this study was to see whether the disadvantages of the Static-99 that prompted the development of the Static-2002 were actually perceived as disadvantages by Static-99 users. One of the stated purposes behind the development of the Static-2002 was to increase the conceptual clarity of the measure. When participants were asked to rate the Static-99 on conceptual clarity, their average

score was 4.1 out of 5, suggesting that they did not view it as particularly problematic. However, that does not mean it cannot be improved upon. When asked what advantages the Static-2002 would require for them to consider using it, 15% of participants said greater conceptual clarity. Together, these findings suggest that conceptual clarity is not viewed as a major disadvantage of the Static-99, but it could be improved upon.

Another goal behind the development of the Static-2002 was to increase the consistency of the scoring criteria. While participants rated the consistency of scoring criteria high ($M = 4.0$ out of 5), there are some indications of dissatisfaction. When asked about the disadvantages of the Static-99, 25% of participants cited some concern with the scoring guidelines being confusing, complicated, or stupid. Similar to the findings for conceptual clarity, the results suggest that the consistency of the coding rules are seen as relatively good, but can be improved.

The fourth research question, regarding whether people are willing to use the Static-2002, merges the purposes of the two studies. The first study took a purely empirical look at the Static-99 and Static-2002, and it found that the Static-2002 is more predictive of recidivism (sexual, violent, and any). However, the implications of these findings in applied settings must be considered. Ideally, sentencing decisions, security classifications, parole decisions, detainment decisions, supervision practices, and dangerous offender hearings, should all be guided by the best possible research on risk assessment. However, as the current study found, risk assessment users are also guided by practical issues. So while the research literature appropriately focuses on issues of empirical validity, it is also necessary to acknowledge the other considerations that affect which instruments are used. In other words, the Static-2002 predicts recidivism

significantly better than the Static-99, but will people actually use it? The findings from this study are unclear, but suggest that users may adopt the Static-2002. Eighty-six percent of participants said they would consider using the Static-2002 if it showed either any improvement in predictive accuracy, or a noticeable improvement. Since the difference was statistically significant, this suggests that many of the participants would be willing to use the Static-2002. When asked what advantages they would be looking for to use the Static-2002 (given the same predictive accuracy), the most common advantage was if it was easier to administer. Since the Static-2002 was designed to increase conceptual clarity and consistency of the coding rules, it is possible that Static-99 users would find the Static-2002 easier to administer. However, the fact that the Static-2002 has more items than the Static-99 may give users the perception that it is more complicated.

Participants also mentioned some hesitations about using the Static-2002. Common hesitations were that the Static-99 has already become a familiar part of the language of risk assessment, and that using a new instrument would involve considerable time and cost because people would require training. Overall, it appears that the Static-2002 offers what people are looking for (increased accuracy and ease of use); however, certain obstacles may interfere with its widespread adoption. This study serves as a reminder that while predictive accuracy may seem like the most important thing to discuss, a myriad of factors affect whether an instrument will be used.

While this study was primarily exploratory, a few strengths and limitations can be discussed. The main strength is the diversity of the sample, which included academics, field workers, and policymakers. A diverse sample allows for the incorporation of

different perspectives and priorities. The primary limitation of this study is the haphazard method of participant selection. Because selection was not random, these results cannot be generalized to the entire population of Static-99 users. For example, while the levels of satisfaction with the Static-99 were high, it is possible that this finding is an artifact of the sample selection procedures, because many participants were acquaintances of one of the Static-99's developers. Even among the participants who were mandated to use the Static-99, they were selected based on their previous participation in the Dynamic Supervision Project, which suggests that they might have had stronger support for empirical methods of risk assessment than some of their colleagues. An additional limitation of this study was the small sample size. Both the small sample size and the haphazard sample selection should raise skepticism in the results. While they offer preliminary insight into the perspectives of those who use risk assessment instruments, further studies should use random selection procedures and larger samples to assess the opinions of this population.

References

- Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct*. Cincinnati, OH: Anderson Publishing.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7-27.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28(3), 369-404.
- Bartosh, D. L., Garby, T., & Lewis, D. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy & Comparative Criminology*, 47, 422-438.
- Beech, A. R. (2005, November). *An evaluation of the effectiveness of Static-99 with juveniles*. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.
- Beggs, S. M., & Grace, R. C. (2005, November). *Improving sex offender risk assessments: The validity of incorporating measures of treatment outcome*. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.
- Bigras, J. (2007). *La prédiction de la récidive chez les délinquants sexuels*. Unpublished doctoral thesis, Université de Sherbrooke.
- Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders*. Unpublished MSc. Thesis, University of Leicester.

- Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: guidelines for clinical practice. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 326-342). New York: Guilford Press.
- Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, California: Sage Publications.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice & Behavior*, 29(4), 355-379.
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, 123(2), 123-142.
- Brown, J. A. (2004). *A comparison of actuarial methods of predicting sexual dangerousness*. Unpublished doctoral dissertation, Fielding Graduate Institute, Santa Barbara, CA.
- Craig, L. A., Beech, A. R., & Browne, K. D. (2006). Evaluating the predictive accuracy of sex offender risk assessment measures on UK samples: A cross-validation of the Risk Matrix 2000 scales. *Sexual Offender Treatment*, 1(1). Retrieved January 23, 2007, from <http://www.sexual-offender-treatment.org/19.98.html>
- Craissati, J., Webb, L., & Keen, S. (2005). *Personality disordered sex offenders*. Unpublished report. London, UK: Bracton Centre (Oxleas NHS Trust), London Probation Area and Home Office.

- 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.
- Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: A Journal of Research and Treatment, 18*, 15-26.
- Epperson, D. L. (2003). *Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota prison and probation samples*. Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation.
- Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence, 18*, 744-759.
- 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*(1), 19-30.
- Haag, A. M. (2005). 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. *Dissertation Abstracts International, 66*, 8B.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Report 1997-04). Ottawa, ON: Department of the Solicitor General of Canada.

- Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20(2), 212-217.
- Hanson, R. K. (2006). Does Static-99 predict recidivism among older sexual offenders? *Sexual Abuse: A Journal of Research and Treatment*, 18, 343-355.
- Hanson, R. K. (2006, September). *Dynamic risk assessment for sexual offenders on community supervision*. Paper presented at the annual conference of the International Association for the Treatment of Sexual Offenders, Hamburg, Germany.
- Hanson, R. K. (personal communication, February 5, 2007).
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362.
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, 27(1), 6-35.
- Hanson, R. K., & Morton-Bourgon, K. (2004). Predictors of sexual recidivism: An updated meta-analysis (User Report 2004-02). Ottawa, ON: Public Safety and Emergency Preparedness Canada.
- Hanson, R. K., & Morton-Bourgon, K. (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis* (User Report 2007-01). Ottawa, ON: Public Safety and Emergency Preparedness Canada.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119-136.

- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of the Static-2002* (User Report 2003-01). Ottawa, ON: Solicitor General Canada.
- Harkins, L., & Beech, A.R. (2007). *Examining the effectiveness of sexual offender treatment using risk band analysis*. Manuscript submitted for publication.
- Harris, A., & Hanson, R. K. (2003, August). The Dynamic Supervision Project: Improving the community supervision of sex offenders. *Corrections Today*, 60-65.
- Harris, A. J. R., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question* (User Report 2004-03). Ottawa, ON: Public Safety and Emergency Preparedness Canada.
- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa, ON: Solicitor General Canada.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413-425.
- Hood, R., Shute, S., Feilzer, M., & Wilcox, A. (2002). Sex offenders emerging from long-term imprisonment: A study of their long-term reconviction rates and of parole board members' judgments of their risk. *British Journal of Criminology*, 42, 371-394.
- Langstrom, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment*, 16, 107-120.

- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior*, 34(1), 37-59.
- Looman, J., & Abracen, J. (2006, September). *Comparison of measures of risk for recidivism in sexual offenders*. Presentation at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.
- McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 nationwide survey*. Brandon, VT: Safer Society Press.
- McGrath, R. J., Hoke, S. E., Livingston, J. A., & Cumming, G. F. (2001, November). *The Vermont Assessment of Sex-Offender Risk (VASOR): An initial reliability and validity study*. Presentation at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.
- Montana, S., & Thompson, G. (2005, November). *Risk assessment for priest sex offenders: Developing a risk assessment tool to measure relapse risk*. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.
- Morton, K. E. (2003). *Psychometric properties of four risk assessment measures with male adolescent sexual offenders*. Unpublished master's thesis, Carleton University, Ottawa, ON.

- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62(4), 783-792.
- Nunes, K. L., Wexler, A., Firestone, P., & Bradford, J. M. (2003, October). *Incarceration and recidivism in a sample of sexual offenders*. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, St. Louis, MO.
- Paolucci, E. O., Genuis, M. L., & Violato, C. (2001). A meta-analysis of the published research on the effects of child sexual abuse. *The Journal of Psychology*, 135, 17-36.
- Poole, D., Liedecke, D., & Marbibi, M. (2000). *Risk assessment and recidivism in juvenile sexual offenders: A validation study of the Static-99*. TX: Texas Youth Commission.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Resick, P. A. (1993). The psychological impact of rape. *Journal of Interpersonal Violence*, 8, 223-255.
- 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(5), 615-620.
- Saum, S. (2006). *A comparison of an actuarial risk prediction measure (Static-99) and a stable dynamic risk prediction measure (Stable-2000) in making risk predictions for a group of sexual offenders*. Unpublished doctoral dissertation, Fielding Institute, Santa Barbara, CA.

- Schneider, A. L., Ervin, L., & Snyder-Joy, Z. (1996). Further exploration of the flight from discretion: The role of risk/need instruments in probation supervision decisions. *Journal of Criminal Justice*, 24(2), 109-121.
- Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology*, 16, 92-108.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1-26.
- Ternowski, D. R. (2004). Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program. *Dissertation Abstracts International – B*, 64(06).
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, 14, 139-153.
- Thornton, D. (chair) (2006, September). *Long term predictive accuracy*. Symposium conducted at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.
- Tough, S. E. (2001). *Validation of two standardized risk assessments (RRASOR, 1997; Static-99, 1999) on a sample of adult males who are developmentally disabled with significant cognitive deficits*. Unpublished master's thesis, University of Toronto, Toronto, ON.
- Webster, C. D., Douglas, K. S., Eaves, D. & Hart, S. D. (1997). Assessing risk of violence to others. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 251-277). New York: Guilford Press.

Appendix A

STATIC-99 Coding Form

Question Number	Risk Factor	Codes	Score										
1	Young (S9909)	Aged 25 or older Aged 18 – 24.99	0 1										
2	Ever Lived With (S9910)	Ever lived with lover for at least two years? Yes No	0 1										
3	Index non-sexual violence - Any Convictions (S9904)	No Yes	0 1										
4	Prior non-sexual violence - Any Convictions (S9905)	No Yes	0 1										
5	Prior Sex Offences (S9901)	<table><tr><th>Charges</th><th>Convictions</th></tr><tr><td>None</td><td>None</td></tr><tr><td>1-2</td><td>1</td></tr><tr><td>3-5</td><td>2-3</td></tr><tr><td>6+</td><td>4+</td></tr></table>	Charges	Convictions	None	None	1-2	1	3-5	2-3	6+	4+	0 1 2 3
Charges	Convictions												
None	None												
1-2	1												
3-5	2-3												
6+	4+												
6	Prior sentencing dates (excluding index) (S9902)	3 or less 4 or more	0 1										
7	Any convictions for non-contact sex offences (S9903)	No Yes	0 1										
8	Any Unrelated Victims (S9906)	No Yes	0 1										
9	Any Stranger Victims (S9907)	No Yes	0 1										
10	Any Male Victims (S9908)	No Yes	0 1										
	Total Score	Add up scores from individual risk factors											

TRANSLATING STATIC 99 SCORES INTO RISK CATEGORIES

<u>Score</u>	<u>Label for Risk Category</u>
0,1	Low
2,3	Moderate-Low
4,5	Moderate-High
6 plus	High

Appendix B

Static-2002 Coding Form

Static-2002

Coding of Static-2002

Items	Range of scores	
<u>Age at release</u>		
18 24.9 = 3, 25 to 34.9 = 2, 35 to 49.9 = 1, 50 or older = 0		0, 1, 2, 3
<u>Persistence of sexual offending</u>		
Sentencing occasions for sexual offences		
0 = no prior sentencing dates for sexual offences		
1 = 1	0 - 3	
2,3 = 2		
4 or more = 3		
Juvenile arrest for a sexual offence (and convicted as an adult for a separate offence)		
0 = no	0, 1	
1 = yes		
High rate of sexual offending		
0 = rate less than once every 15 years	0, 1	
1 = rate greater than once every 15 years		
Persistence Subscore raw score	0 - 5	
0 = 0		
1 = 1		
2,3 = 2		0, 1, 2, 3
4,5 = 3		
<u>Deviant sexual interests</u>		
Any convictions for non-contact sex offences	0, 1	
0 = no		
1 = yes		
Any male victims : 0 = no, 1 = yes	0, 1	
Two or more victims < 12 years, one unrelated	0, 1	
0 = no		
1 = yes		
Total	0 - 3	0, 1, 2, 3

<u>Relationship to victims</u>		
Any unrelated victims 0 = no; 1 = yes	0, 1	
Any stranger victims 0 = no; 1 = yes	0, 1	
Total		0, 1, 2
<u>General Criminality</u>		
Arrest/Sentencing Occasions		
0 = no prior charges for anything 1 = any prior charges or convictions, but less than 3 prior sentencing occasions 2 = 3 – 13 prior sentencing occasions 3 = 14 or more sentencing occasions	0 – 3	
Any breach of conditional release: 0 = no; 1 = ycs	0, 1	
Years free prior to index offence. 0 = 4 or more years 1 = less than 4 years	0, 1	
Any convictions for non-sexual violence 0 = no 1 = yes	0, 1	
General Criminality Subscore raw score	0 – 6	
0 = 0 1, 2 = 1 3, 4 = 2 5, 6 = 3		0, 1, 2, 3
Total		0 – 14

Appendix C

Letter Sent to Researchers to Recruit Datasets

Dear _____,

My name is Leslie Helmus and I am doing my undergraduate thesis in criminology/psychology at Carleton University in Ottawa. I am being supervised by Dr. Karl Hanson and Dr. Ralph Serin. For my thesis I plan to conduct replications of the Static-2002 and compare its predictive accuracy to the Static-99 (see below for a more detailed description).

I am contacting you to request your Static-99/Static-2002 data set to include in my thesis. The information I need in the data set is as follows: scores for each item on the Static-99 and the Static-2002, total scores for the Static-99 and the Static-2002, the at-risk date, the survival end date, and the dates of sexual, violent, and any recidivism. Please note that names or any other personal identifiers are not required. Karl Hanson, from the Department of Public Safety and Emergency Preparedness Canada, will be responsible for storing the data. It will be kept on a secure computer network, on a secure floor of a secure federal government building. Karl Hanson and I will be the only people with access to the data. Karl's security level with the federal government is "secret," and my security level is "enhanced reliability." We can send copies of our security clearance certificates on request.

Please let me know if it is possible for me to use your data set, and if so, what steps you need me to go through in order to access your data.

Thank you very much for your time. I look forward to hearing from you soon.

Sincerely,

Leslie

Thesis Proposal
Leslie Helmus

Actuarial risk assessment instruments are increasingly being used with sex offenders to evaluate their risk to reoffend. It is necessary to demonstrate the predictive accuracy of these tools because risk assessments influence important decisions such as sentencing, release, treatment intensity, and/or level of supervision.

The Static-99 is a popular and widely used measure of recidivism risk. Numerous replications have demonstrated that it predicts sexual recidivism with moderate accuracy. Recently, however, the Static-2002 has been developed in an attempt to improve upon the Static-99. Firstly, the Static-99 was designed atheoretically, meaning that variables were included solely based on their predictive accuracy. The Static-2002, while still including

variables demonstrated to be predictive of sexual recidivism, organizes them in a way that is intended to be more conceptual and therefore to have greater construct validity. The Static-2002 was also designed to have more consistent coding rules compared to the Static-99 (which in turn should increase the inter-rater reliability), and it is also intended to increase the predictive accuracy. Given that it has multiple purported advantages, even if the predictive accuracy of the two instruments is the same, it might still be beneficial to switch to the Static-2002 because it has greater construct validity and is easier to score.

Despite the expectation that it is an improvement over the Static-99, it is not yet widely used. This caution in adopting the Static-2002 is appropriate because there have not been sufficient replications to evaluate its predictive accuracy compared to the Static-99. This thesis aims to fill this gap in the research. The thesis will address 6 questions related to the Static-99 and the Static-2002.

- 1) How does the predictive accuracy of the Static-2002 compare to the predictive accuracy of the Static-99? This will be examined separately for sexual recidivism, violent recidivism, and any recidivism.
- 2) How do the Static-99 and the Static-2002 compare in terms of consistency in predictive accuracy across different samples?
- 3) What is the predictive accuracy of the Static-99 and Static-2002 when items concerning marital status or victim information are unknown? In these cases, how should the risk categories be re-scored? This question is relevant because risk assessments using only official criminal records often do not contain this information.
- 4) What is the best way to assign Static-2002 scores into risk categories?
- 5) Does the Static-2002 have construct validity? Within the Static-2002 there are two sections that are intended to measure enduring characteristics: deviant sexual interests and general criminality. If the measure has construct validity, it is expected that scores in the general criminality subsection will correlate with certain items from the STABLE-2000 (e.g., impulsive acts, poor cognitive problem-solving) and will also correlate with general/any recidivism. It is also expected that scores in the deviant sexual interests subsection will correlate with certain items from the Stable-2000 (e.g., lovers/intimate partners, emotional identification with children, sexual pre-occupations, sex as coping, deviant sexual interests).
- 6) Are people who currently use the Static-99 satisfied with it? And what sorts of evidence/advantages would they be looking for in order to switch to the Static-2002?

These questions will be addressed in three stages.

Stage 1: Data from the Dynamic Supervision Project

The Dynamic Supervision Project is a research initiative aimed at improving the community supervision of sexual offenders. This sample has data on approximately 800 offenders from across Canada. The offenders in this sample have been assessed on both the Static-99 and the Static 2002. Research questions 1-5 will be explored with this sample.

Stage 2: Aggregating samples

The author will aggregate data from the Dynamic Supervision Project and 3-5 other samples from Canada and other countries. These combined samples should provide a sufficiently large and diverse sample of sex offenders for a thorough comparison of the predictive validity of the Static-99 and Static-2002. These combined samples will be used to explore research questions 1-4.

Stage 3: Interviewing the "users" of the Static-99.

To answer the 6th research question, the author will conduct telephone interviews with people who use the Static-99. The interview will be designed to determine what these users like/dislike about the Static-99, to ascertain whether some of the issues the Static-2002 was designed to rectify (e.g., inconsistent coding rules, lack of construct validity) are perceived as "problems," and to determine whether these people would be interested in switching to the Static-2002, and what evidence/advantages they would require in order to make this switch. Interviews will be sought from three different categories of Static-users:

- 1) Those who independently choose to use the Static-99 (e.g., psychologists).
- 2) Those who are required to use the Static-99 as part of an official policy they are subject to (e.g., prison classification officers).
- 3) Those who are in a policy-making position who are able to mandate other individuals to utilize the Static-99.

Appendix D

Static-2002 Recidivism Percentages by Risk Level

Static-2002 Score	Sample Size <i>n</i> (%)	Sexual Recidivism		Violent Recidivism		Any Recidivism	
		5 years	10 years	5 years	10 years	5 years	10 years
0	97 (2.8)	.0	.0	.0	.0	.0	.0
1	229 (6.5)	2.5	4.0	4.6	8.0	8.6	12.4
2	348 (9.9)	2.6	3.7	6.0	10.8	14.5	19.2
3	411 (11.7)	4.1	5.5	11.0	14.7	18.9	25.7
4	474 (13.5)	6.8	9.5	13.2	20.0	28.4	37.5
5	483 (13.7)	11.8	19.1	22.8	41.7	42.2	59.4
6	468 (13.3)	10.2	14.2	25.2	37.7	45.1	64.9
7	340 (9.6)	14.4	24.7	27.6	43.6	55.9	65.3
8	328 (9.3)	19.8	25.4	35.0	47.1	63.0	71.9
9+	342 (9.7)	29.4	38.9	41.1	55.3	67.2	79.0
Average							
5.0	3,520 (100)	11.1	16.1	21.6	33.0	36.7	47.6

Appendix E

Static-99 Questionnaire

Note: text in italics is not read to participants. It is an administration aide.

Hi! My name is Leslie Helmus. I'm an undergraduate student at Carleton University and I'm doing my thesis under the supervision of Dr. Karl Hanson, who developed the Static-99. _____ gave me your name and contact information as someone whose occupation involves risk assessment for sexual offenders.

(Fill in the blank with the person who provided the potential participant's name)

For my thesis I'm comparing the predictive accuracy of the Static-99 and the Static-2002. I'm also conducting interviews with people involved in the risk assessment of sexual offenders to assess their opinions on the Static-99. The goal of my research is to improve the risk assessment of sexual offenders and also to gain information on client satisfaction with the Static-99.

I was wondering if you would consider participating in a telephone questionnaire. Your participation is completely voluntary and also anonymous. Your answers will be recorded by myself and your name will not be included anywhere on your questionnaire. Karl Hanson will not know whether you participated, nor will he be able to connect any responses to any participants.

We can do the questionnaire right now, or if you prefer we can schedule another time that works best for you, either in the daytime, during evenings, or on weekends.

Do you need approval from a supervisor before responding to a survey?

(If so, ask them if you can call them back after they have had time to discuss it with their supervisor)

If they're open to it, go to the next section.

The questionnaire asks your opinion on the Static-99, the Static-2002, and it also asks for some demographic information. This questionnaire does not contain any sensitive questions and it is not necessary for you to discuss any specific cases you have dealt with.

Your participation is completely voluntary. If you agree to participate, you have the right to change your mind at any point and withdraw from this study without any consequences. If you do participate, you may also choose not to answer a question and there are no consequences for exercising this right. For your own records, I can also email you a copy of this informed consent that I have just gone through.

Do you agree to participate? _____

Do you understand that your participation is voluntary and that you may withdraw at any point? _____

Would you like to be sent a copy of this informed consent? _____

Static-99

1. What is your involvement in sexual offender risk assessment? *(Let them describe it in their own words, and then check off the category that applies, asking them clarifying questions if necessary)*

- ☐ supervisor/policy-maker
- ☐ applied/clinical assessment
- ☐ research

2. *If not a manager/policy-maker:*

Do you use the Static-99 to assess risk of recidivism in sexual offenders?

- ☐ Yes ☐ No

3. *If a manager/policy-maker:*

Do you recommend/mandate the use of the Static-99 to assess risk of recidivism in sexual offenders?

- ☐ Yes ☐ No

4. Why do you use/recommend the Static-99? *(Let them offer reasons off the top of their heads. Check off the boxes according to what reasons they give. Reasons not covered by the items below should be written in the space provided)*

- ☐ It is the policy at my workplace / my supervisor tells me to
- ☐ Quick/easy to administer
- ☐ Predictive ability
- ☐ Popular
- ☐ High reliability
- ☐ I've been trained on it
- ☐ It's been recommended by others/predecessors

5. In your overall evaluation of risk for sexual recidivism, how much weight do you place (or recommend that others place) on the results of the Static-99?

- 1) No weight, or almost none
- 2) A little
- 3) Some
- 4) A lot (primary consideration)
- 5) Exclusive consideration – it is almost the only thing considered

6. *Skip this question if they are a manager/policy-maker*

How frequently do you use the Static-99?

- | | |
|--|--|
| <input type="checkbox"/> Less than once a year | <input type="checkbox"/> About once a month |
| <input type="checkbox"/> Once a year | <input type="checkbox"/> A few times a month |
| <input type="checkbox"/> 2-5 times a year | <input type="checkbox"/> About once a week |
| <input type="checkbox"/> 6-10 times a year | <input type="checkbox"/> A few times a week |

7. *If not a manager/policy-maker:*

Do you use any other risk assessment instruments when dealing with sexual offenders?

- ☐ Yes ☐ No

If yes, what: _____

8. *If a manager/policy-maker:*

Do you recommend/mandate the use of other risk assessment instruments when dealing with sexual offenders?

- ☐ Yes ☐ No

If yes, what: _____

9. What do you see as some of the advantages of the Static-99?

10. What do you see as some of the disadvantages of the Static-99?

11. With the Static-99, are there any items that you occasionally find difficult to obtain the information to rate?

12. With the Static-99, are there any items that you find difficult to achieve consensus or high rater reliability?

13. Overall, are you satisfied with the Static-99?

☐ Yes ☐ No

14. On a scale of 1-4, how important are the following items to you when it comes to risk assessment instruments?

	1	2	3	4	9
	Not Important	Somewhat Important	Important	Very Important	Unsure
Conceptual clarity (you know what you're measuring)	1	2	3	4	9
Consistency of scoring criteria	1	2	3	4	9
Easy to learn/train others	1	2	3	4	9
Easy to administer	1	2	3	4	9
Predictive accuracy	1	2	3	4	9

15. On a scale of 1-5, how do you think the Static-99 rates on the following items?

	1	2	3	4	5
	Very poorly	Poorly	Unsure	Good	Very Good
Conceptual clarity (you know what you're measuring)	1	2	3	4	5
Consistency of scoring criteria	1	2	3	4	5
Easy to learn/train others	1	2	3	4	5
Easy to administer	1	2	3	4	5
Predictive accuracy	1	2	3	4	5

Static-2002

16. Have you ever heard of the Static-2002? *(If not, briefly describe it)*
☐ Yes ☐ No

If yes, do you know enough about the Static-2002 to have an opinion about it?
☐ Yes ☐ No

If yes, please answer the following two questions:

- a) What do you see as the strengths of the Static-2002 (particularly in comparison to the Static-99)?

- b) What do you see as some of the weaknesses of the Static-2002?

17. How much more accurate would the Static-2002 have to be in order for you to consider switching?
☐ Any difference, no matter how small (e.g., ROC of .72 compared to .71)
☐ A noticeable difference (e.g., ROC of .73 - .75)
☐ Substantially better (e.g., ROC of .76 or higher)

18. If the predictive accuracy of the Static-2002 was the same as the Static-99, would you switch if it had other advantages?
☐ Yes ☐ No

If yes, what advantages would be needed in order for you to consider switching?

19. What factors would make you hesitate about switching to the Static-2002?

Demographic Information

Highest level of education achieved:

- | | |
|---|--|
| <input type="checkbox"/> Some high school | <input type="checkbox"/> University Degree |
| <input type="checkbox"/> High school | <input type="checkbox"/> Masters |
| <input type="checkbox"/> Some college | <input type="checkbox"/> PhD |
| <input type="checkbox"/> College diploma | <input type="checkbox"/> LLB |
| <input type="checkbox"/> Some university | <input type="checkbox"/> Other: _____ |

Area of education: _____

Occupation: _____

Briefly describe the nature of your employment.

Any additional comments?

Date of Interview: _____

Interviewer: _____

Debriefing

Thank you very much for participating!

I'd like to take this opportunity to tell you a bit more about my thesis.

My thesis has three components. The main goal is to compare the Static-99 and the Static-2002. The first component involves using a sample of approximately 800 sex offenders from an archival data set that Karl Hanson used for the Dynamic Supervision Project. The offenders in this project were all coded on both the Static-99 and the Static-2002, and we have recidivism information with a follow-up time that varies between 1-5 years. Using this data set, I will compare the predictive accuracy of the Static-99 and the Static-2002. I will also look at construct validity, how the Static-2002 scores should be divided into risk categories, and how predictive these measures are when there is missing information.

The second component of my thesis involves aggregating the data from the Dynamic Supervision Project with the data sets from a few other researchers who have also done Static-2002 replication studies. Using this combined data set, I will look at the same things as I mentioned for the first component, and I will also compare how consistent the Static-99 and the Static-2002 are at predicting recidivism across samples.

The third component of my thesis is the part you participated in. I'm interviewing users of the Static-99 to look at several things. I want to know why people use the Static-99. Also, the Static-2002 was created in response to what was perceived by the developers as some shortcomings of the Static-99. I'm interested in knowing if the perceptions of the strengths and weaknesses of the Static-99 are similar for the users of the measure and the developers of the measure. I also want to know if professionals have heard of the Static-2002 and whether they are interested in the new measure.

If you would like further information about this study, the Static-99, or the Static-2002, I can provide that to you right now. If you have any ethical or other concerns about this study, I can also give you contact information.

Also, I'm looking for additional participants. If you know of anyone who uses the Static-99 who might be interested in participating, you can provide their names and contact information to me either over the phone, or by email.

*Address any questions they may have.
If necessary, offer to send references or contact information.*

Resources:

This study:

Leslie Helmus
Principal Investigator, Carleton University
leslie.helmus@psepc.gc.ca

Dr. Ralph Serin
Faculty Sponsor, Carleton University
613-520-2600 ext. 1557
ralph_serin@carleton.ca

Dr. Karl Hanson
Adjunct Faculty Sponsor, Carleton University
Developer of Static-99 and Static-2002
Senior Research Officer, Department of Public Safety & Emergency Preparedness
613-991-2840
Karl.Hanson@psepc.gc.ca

Ethical Concerns:

Dr. Janet Mantler
Chair of the Carleton University Ethics Committee for Psychological Research
613-520-2600 ext. 4173
janet_mantler@carleton.ca

Other concerns:

Dr. Mary Gick
Chair of Psychology Department at Carleton University
613-520-2600 ext. 2648
mary_gick@carleton.ca

Information on the Static-99 and the Static-2002:

- Unpublished coding rules and reports on the Static-99 and the Static-2002 can be obtained by request from either Leslie Helmus or Karl Hanson.

Information on the Dynamic Supervision Project:

Harris, A., & Hanson, R. K. (2003). The Dynamic Supervision Project: Improving the community supervision of sex offenders. *Corrections Today* (August 2003), 60-64.