STABILITY AND PREDICTIVE AND INCREMENTAL ACCURACY OF THE INDIVIDUAL ITEMS OF STATIC-99R AND STATIC-2002R IN PREDICTING SEXUAL RECIDIVISM

A Meta-Analysis

LESLIE-MAAIKE HELMUS
Forensic Assessment Group

DAVID THORNTON
Research Unit, Sand Ridge Secure Treatment Centre

This study investigated a potential source of variability in actuarial scale performance: the individual items. Using data from 8,053 sex offenders across 22 samples, we examined the predictive and incremental accuracy of the items from Static-99R and Static-2002R, and the stability of predictive accuracy across samples. Generally, all items had significant predictive accuracy and contributed incrementally to predicting sexual recidivism, with few exceptions. Roughly half the items demonstrated significant variability in their predictive accuracy across samples, although this was often variability in the magnitude of predictiveness as opposed to the direction. Some moderator effects were found, with the most common being the country of the study (which influenced accuracy in different directions depending on the item) and whether the offenders were preselected as unusually high risk or need (lower discrimination was found in these samples). The findings support the Static-99R and Static-2002R items with few exceptions, and possibilities for future research are highlighted.

Keywords: risk assessment; sex offenders; Static-99R; Static-2002R; meta-analysis; predictive accuracy

AUTHORS' NOTE: *The views expressed are those of the authors and not necessarily those of the Wisconsin Department of Health Services. We would like to thank the following researchers for granting us permission to use their data: Alfred Allan, Howard Barbaree, Tony Beech, Susanne Bengtson, Jacques Bigras, Sasha Boer, Jim Bonta, Sébastien Brouillette-Alarie, Vivienne de Vogel, Margretta Dwyer, Reinhard Eher, Doug Epperson, Andy Haag, R. Karl Hanson, Leigh Harkins, Andreas Hill, Steve Johansen, Ray Knight, Niklas Långström, Calvin Langton, John Milton, Terry Nicholaichuk, Jean Proulx, Martin Rettenberger, Rebecca Swinburne Romine, Daryl Ternowski, Robin Wilson, and Annie Yessine. Correspondence concerning this article should be addressed to Leslie-Maaike Helmus, Forensic Assessment Group, 11 Aspen Grove, Nepean, ON K2H 8Z9, Canada; e-mail: Lmaaikehelmus@gmail.com.*

CRIMINAL JUSTICE AND BEHAVIOR, 2015, Vol. 42, No. 9, September 2015, 917–937.
DOI: 10.1177/0093854814568891

Professionals seeking to evaluate the risk presented by sexual offenders commonly use empirical actuarial instruments to anchor their assessments. These instruments use items selected on the basis of their observed statistical association with sexual recidivism. Individual items typically have only a small association with recidivism but by combining them into prediction scales, a moderate level of prediction can be obtained. When they were first introduced, the appeal of these scales probably seemed to be that they divided offenders into groups, each of which had its own distinct sexual recidivism rate.

Static-99 (Hanson & Thornton, 2000) is one of the scales that benefited from this perception of actuarial tools and which accordingly has become widely used for pretreatment assessments (Jackson & Hess, 2007; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010), community supervision (Interstate Commission for Adult Offender Supervision, 2007), and civil commitment evaluations (Jackson & Hess, 2007). It is also the most extensively studied of the empirical actuarial instruments used to assess risk of sexual recidivism (Hanson & Morton-Bourgon, 2009). Its ability to sort offenders into groups that differ in their relative risk of sexual recidivism based on relatively simple, easy-to-obtain information has been widely replicated (Hanson, Babchishin, Helmus, & Thornton, 2013; Hanson & Morton-Bourgon, 2009) and also has been demonstrated to be fairly stable across diverse samples (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012), but this close scrutiny has also revealed important limitations of the method.

To begin with, all empirically derived recidivism estimates are based on limited samples and so these estimates are subject to sampling error. Of course it is possible to calculate confidence intervals for these estimates. Examples of these can be seen on www.static99.org, but they are sometimes misunderstood as implying that the risk of the individual offender can be expected to fall within these bounds. This is not the case; these confidence intervals relate only to how well-estimated the average recidivism rate of persons having a given score is. They tell us nothing about how homogeneous, in terms of risk, persons having a given Static-99 score are. Possible sources of heterogeneity could include factors external to the risk scale, as well as differences in how offenders obtained a given score (i.e., offenders can obtain the same Static-99 score based on a completely different set of risk items).

We know that there are important factors external to Static-99 that influence the risk of recidivism. One example is age. Although the optimal way of measuring age (e.g., first arrest, index offense, release) and the reasons for its relationship to recidivism are debated (e.g., Barbaree & Blanchard, 2008; Rice & Harris, 2014), early research was clear that age at release added incremental predictive accuracy to the original version of Static-99 (Barbaree, Langton, & Blanchard, 2007; Barbaree, Langton, Blanchard, & Cantor, 2009; Hanson, 2006; Thornton, 2006; for a notable recent exception to this finding, see Rettenberger, Haubner-MacLean, & Eher, 2013). These findings led to the development of a revised version of the scale (Static-99R; Helmus, Thornton, Hanson, & Babchishin, 2012).[1]

In addition, fairly comprehensive assessment of the kinds of psychological factors identified by Mann, Hanson, and Thornton (2010) have consistently shown incremental prediction beyond that possible from static scales (Allan, Grace, Rutherford, & Hudson, 2007; Beggs & Grace, 2010; Craig, Thornton, Beech, & Browne, 2007; Eher, Matthes, Schilling, Haubner-Maclean, & Rettenberger, 2012; Hanson, Helmus, & Harris, in press; Harkins, Thornton, & Beech, 2009; Olver, Wong, Nicholaichuk, & Gordon, 2007; Thornton, 2002;

Thornton & Knight, 2013). This means that variation in the density of psychological risk factors among offenders with the same Static-99R score will lead to corresponding variation in these offenders' actual risk (e.g., Hanson & Thornton, 2012). This represents a source of variation in risk that is not taken into account in the way confidence intervals are calculated.

A particularly striking finding has been the significant variation between samples in sexual recidivism rates associated with Static-99R scores. For example, the predicted recidivism rate for a Static-99R score of 2 is as low as 3% in some samples and as high as 20% in others (Helmus, Hanson, et al., 2012). This has thoroughly undermined the notion of a single recidivism rate being associated with each actuarial risk score and has led to the development of multiple sets of recidivism norms being offered on www.static99.org, with evaluators having to determine which is the most suitable to the individual case.

The present research is seeking to explore another potential source of variation in the utility of Static-99R and Static-2002R: the contribution of individual items. It is notable that most of the static risk factors included in the scales have not been systematically examined since Hanson and Bussière's (1998) meta-analysis, and the specific definitions and coding rules of the Static-99R and Static-2002R items have not yet been meta-analyzed. Validation of these individual items in more current samples, corresponding to the specific definitions of the Static-99R and Static-2002R coding rules, would be helpful. To the extent that certain items are not predicting recidivism, this would reduce confidence in the results of the scale. Furthermore, even if all items predict recidivism, it is possible that they do not add uniquely to the prediction of recidivism after controlling for the other items.

Beyond these basic questions, we hypothesize that there may be meaningful variation between samples in how predictive these items are. If this is the case, it would have important implications for Static-99R and Static-2002R, as well as other static empirical actuarial scales, which often use similar items. There are at least seven kinds of reason for anticipating variability.

First, variation in how well the outcome variable is measured should affect the association of all predictors with recidivism. There are a number of ways in which this can occur: Use of a fixed follow-up period rather than a ragged follow-up should lead to a more valid outcome measure (although analyses such as Cox regression can alleviate some of the disadvantages of varying follow-up), use of multiple sources may identify recidivism events that have otherwise been missed, and prosecutorial discretion and variability across jurisdictions in charging practices will also affect the validity of official measures of recidivism. In addition, coding sexual recidivism based on the name of the conviction is known to exclude a meaningful portion of sexually motivated reconvictions that are classified as violent offenses on criminal records (e.g., through plea bargains, or focusing on the violent aspect, such as homicide; Rice, Harris, Lang, & Cormier, 2006).

Second, statistical databases used by researchers may vary in how complete and accurate they are; in particular, this might vary between jurisdictions so that greater prediction should be possible in jurisdictions with more accurate statistical databases. Both Canada and the United Kingdom contain centralized national criminal records. A major limitation of Canadian national criminal records is the substantial delays in recording the data and purging of records over time due to inactivity (Hanson & Nicholaichuk, 2000; Helmus, Hanson, & Morton-Bourgon, 2011). In the United Kingdom, criminal record databases exclude certain regions (e.g., Scotland and Northern Ireland) or may exclude less serious types of

offenses (Friendship, Thornton, Erikson, & Beech, 2001). Despite these limitations, these records are at least national, whereas the United States lacks a comprehensive and centralized criminal record database, and statewide and FBI databases are both known to have incomplete, inaccurate, and ambiguous data (Laudon, 1986). For these reasons, it is possible that predictive accuracy may be lower in the United States. Previous research has found significantly higher predictive accuracy for risk scales in studies from the United Kingdom (Hanson & Morton-Bourgon, 2009).

Third, use of convictions rather than charges will change the validity of different outcomes; presumably a conviction is a more valid indicator of recidivism than a charge, but the absence of a conviction would be a less valid indicator of nonrecidivism than the absence of a charge. Nonetheless, previous research has found that surprisingly, the use of charges (vs. convictions) did not consistently and meaningfully affect the predicted recidivism rates for Static-99 scores (Helmus, 2009).

Fourth, there may be variation in how accurately time at risk is measured. Sexual recidivism is defined relative to years in the community with the opportunity to offend. A starting point in measuring this is to compare the date of release to the date at which the follow-up ends, but time at risk of sexual offending will be reduced whenever offenders are removed from the community for reasons other than the commission of new sexual offenses (e.g., following conviction for nonsexual offenses or revocations for noncompliance with conditions of supervision). This can make accurately determining true time at risk difficult and time-consuming.

Fifth, other methodological decisions may affect predictive accuracy. For example, there may be important differences between retrospective or prospective research designs. Accuracy in retrospective designs could be artificially enhanced through hindsight bias. Such effects can be mitigated by scoring risk items and scales blind to outcome data.

Sixth, samples may vary in how well independent variables are measured. Sometimes victim type variables may only be available for the most recent sexual offense and not recorded for priors. In addition, the file information that researchers' code was commonly created for other purposes and may be incomplete. In some jurisdictions, it may be quite hard to get complete records of prior convictions. In other jurisdictions, convictions may be very reliably recorded but information about charges may be unavailable. Another concern relates to the quality of risk assessment scoring. This can be affected by whether the raters received training from a certified Static-99 trainer and the general conscientiousness of the scoring. For example, there is some evidence that studies with higher interrater reliability show significantly higher predictive accuracy (Hanson & Morton-Bourgon, 2009). Quality of training has also been linked to quality of risk scores (Vincent, Guy, Fusco, & Gershenson, 2012), and scores from field staff who followed all requested steps had higher predictive accuracy than staff who were less thorough (Hanson et al., in press).

Seventh, how the sample was selected might make a difference. It is not uncommon for the samples studied in recidivism research to have been subject to selection processes influenced by concern about the risk presented by those involved. This kind of sample selection is liable to have two kinds of effect. The most obvious is that lower risk offenders are less likely to be included (a classic restriction of range effect) but a more subtle effect is a reduced correlation between the items in static actuarial scales and the underlying psychological risk factors. Specifically, the static scales are not intended to reflect all relevant risk factors, and they are correlated with risk factors external to the scale (Hanson & Thornton,

2012). When samples are preselected, it is possible that offenders with high static risk will be preselected no matter what their external risk factors, whereas low risk offenders will only be preselected if they have a high density of external risk factors. This means that in unselected samples, persons with higher static actuarial scores will also on average tend to have higher levels of psychological risk factors than those with lower static actuarial scores. In contrast, in samples selected for risk indicators of all kinds, those with higher static actuarial scores may not have higher levels of psychological risk factors than those with lower static actuarial scores (because lower-scoring offenders are more likely to be preselected if they have higher psychological risk factors). As a consequence, static actuarial factors should be less predictive in samples strongly selected for risk indicators external to the static risk scales. This would match previous findings that Static-99R and Static-2002R total scores are less predictive in samples strongly preselected based on risk factors external to the scales (Hanson et al., 2014).

Testing the hypothesized variation between samples in the predictive accuracy of the items in empirical actuarial instruments is not straightforward given that observed variability is expected to occur merely from sampling error. Meta-analysis overcomes this difficulty by assessing whether observed variability exceeds what could be expected by chance.

The current study will address the following research questions:

Research Question 1: Do the Static-99R and Static-2002R items predict sexual recidivism?

Research Question 2: Do they predict similarly across samples? If not, are there obvious reasons why they do not predict similarly (i.e., moderator variables such as the factors discussed above)?

Research Question 3: Do the items contribute uniquely to the prediction of sexual recidivism?

METHOD

SAMPLES

This study used Static-99R and Static-2002R data sets collected as part of a larger project for renorming the Static-99 (see Helmus, 2009; Helmus, Thornton, et al., 2012). To be included in this study, the sample required information on sexual recidivism and Static-99 items. One sample was excluded because there were no sexual recidivists (Cortoni & Nunes, 2007). Cases were deleted if there were unresolved coding inconsistencies. In addition, as per the coding rules (A. J. R. Harris, Phenix, Hanson, & Thornton, 2003; Phenix, Doren, Helmus, Hanson, & Thornton, 2009), cases were deleted if more than one Static-2002 item was missing, if any Static-99 item was missing other than *Ever Lived With a Lover* (Item 2), if the offender was less than 18 years old at time of release or less than 16 years old when he or she committed the index offense, or if the offender was female. In addition, cases were deleted if the index sex offense was more than 2 years prior to the current offense because Static-99R and Static-2002R were developed on, and intended for, sex offenders with a current (or recent) sex offense conviction. In total, 22 samples were available with Static-99 item data (N = 8,053). Of these, 19 samples also had age data to compute the revised age item for Static-99R (N = 7,461), and 8 samples had Static-2002R item data (N = 2,951).

Table 1 provides descriptive information for the studies included. For additional information, readers are encouraged to obtain a more detailed report of this project (Helmus, 2009) or to refer to the original studies. Ten samples were from Canada; 4 were from the

TABLE 1: Descriptive Information

Study	*n*	Age M (SD)	Country	Recidivism Criteria	Type of Sample	Follow-up M (SD)	Release Period	Mdn Year Release
Allan, Grace, Rutherford, and Hudson (2007)	493	42 (12)	New Zealand	Charges	Prison treatment	5.8 (2.9)	1990-2000	1994
Bengtson (2008)	311	33 (10)	Denmark	Charges	Forensic psychiatric evaluations	16.2 (4.2)	1978-1995	1986
Bigras (2007)	480	43 (12)	Canada	Charges	Routine CSC	4.6 (1.9)	1995-2004	1999
Boer (2003)	299	41 (12)	Canada	Conviction	Routine CSC	13.3 (2.1)	1976-1994	1990
Bonta and Yessine (2005)	133	40 (10)	Canada	Conviction	Preselected high risk	5.5 (2.4)	1992-2004	1999
Brouillette-Alarie and Proulx (2008)	228	36 (10)	Canada	Conviction	Prison & community treatment	9.9 (4.5)	1979-2006	1996
de Vogel, de Ruiter, van Beek, and Mead (2004)	121	—	The Netherlands	Conviction	Indeterminate treatment order	11.6 (5.9)	1977-2000	1990
Eher, Rettenberger, Schilling, and Pfafflin (2008, 2009)	706	41 (12)	Austria	Conviction	Routine European prison	3.9 (1.1)	2000-2005	2003
Epperson (2003)	178	37 (13)	The United States	Charges	Routine correctional	7.9 (2.4)	1989-1998	1995
Haag (2005)	198	37 (10)	Canada	Conviction	Detained until end of sentence	7.0 (0.0)	1995	1995
Hanson, Harris, Scott, and Helmus (2007)	702	42 (13)	Canada	Charges	Routine community supervision	3.4 (1.0)	2001-2005	2002
Harkins and Beech (2007)	198	43 (12)	The United Kingdom	Conviction	Prison & community treatment	10.4 (1.1)	1994-1998	1995
Hill, Habermann, Klusmann, Berner, and Briken (2008)	84	39 (11)	Germany	Conviction	Sexual homicide perpetrators	12.6 (6.6)	1971-2002	1989
Johansen (2007)	273	38 (11)	The United States	Charges	Prison treatment	9.1 (1.1)	1994-2000	1996
Knight and Thornton (2007)	466	36 (11)	The United States	Charges	Civil commitment evaluation	8.6 (2.6)	1957-1986	1970
Långström (2004)	1,278	41 (12)	Sweden	Conviction	Routine European prison	8.9 (1.4)	1993-1997	1995
Langton (2003)	354	—	Canada	Conviction	CSC treatment	6.4 (3.0)	1990-2001	1995
Milton (2003)	116	—	The United Kingdom	Conviction	Psychiatric hospital treatment	10.2 (5.0)	1978-1998	1986
Nicholaichuk (2001)	280	35 (9)	Canada	Conviction	High intensity treatment	6.4 (4.0)	1983-1998	1992
Swinburne Romine, Dwyer, Mathiowetz, and Thomas (2008)	681	38 (12)	The United States	Conviction	Community treatment	16.8 (7.8)	1977-2007	1988
Ternowski (2004)	247	44 (13)	Canada	Charges	Prison treatment	7.5 (1.0)	1994-1998	1996
Wilson, Cortoni, and Vermani (2007; Wilson, Picheca, & Prinzo, 2007)	227	42 (11)	Canada	Charges	Preselected high risk	5.2 (3.0)	1994-2007	2002
Total	8,053	40 (12)	—	—	—	10.2 (5.0)	1957-2007	1996

Note. CSC = Correctional Service Canada (administers all sentences of at least 2 years).

United States; 2 were from the United Kingdom; and there was 1 each from Denmark, Austria, Sweden, Germany, New Zealand, and the Netherlands. The average age at release was 40 years (*SD* = 12). Offenders were released between 1957 and 2007, with a median release year of 1996. All samples used official criminal records to measure recidivism, but 13 samples used convictions as the recidivism criteria and 9 used charges. Note that both definitions underestimate the true rate of recidivism due to underreporting (e.g., Dobash & Dobash, 1995). Offenders were followed up for an average of 10.2 years (*SD* = 5.0).

Samples were further coded for moderator analyses. One moderator analysis compared samples preselected as unusually high risk or high need (i.e., the offenders would likely represent the top 20% in terms of risk) to other samples. This classification followed the definition and rationale used previously for these samples (see Helmus, 2009). Seven samples were considered preselected as high risk or need: Bengtson (2008); Bonta and Yessine (2005); de Vogel, de Ruiter, van Beek, and Mead (2004); Haag (2005); Knight and Thornton (2007); Nicholaichuk (2001); and finally, data from both Wilson, Cortoni, and Vermani (2007) and Wilson, Picheca, and Prinzo (2007), which were combined into a single sample.

Samples were also classified based on their country and whether they used charges or convictions as their criteria for recidivism (the assumption being that samples using charges may have had access to more detailed criminal records for coding the Static scales; see Table 1). Last, 11 samples included both rapists and child molesters, with sufficient information to allow calculation of effect sizes separately for both subgroups (Bigras, 2007; Boer, 2003; Brouillette-Alarie & Proulx, 2008; Eher, Rettenberger, Schilling, & Pfafflin, 2008; Haag, 2005; Hanson, Harris, Scott, & Helmus, 2007; Knight & Thornton, 2007; Långström, 2004; Nicholaichuk, 2001; Swinburne Romine, Dwyer, Mathiowetz, & Thomas, 2008; Wilson, Cortoni, & Vermani, 2007; and Wilson, Picheca, & Prinzo, 2007).

MEASURES

Static-99R

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) is a 10-item actuarial scale that assesses recidivism risk of adult male sex offenders (see www.static99.org). The items and scoring rules are identical to Static-99 (Hanson & Thornton, 2000) with the exception of updated age weights (Helmus, Thornton, et al., 2012). Offenders can be placed in one of four risk categories based on their total score (ranging from −3 to 12): *low* (−3 to 1), *moderate–low* (2 to 3), *moderate–high* (4 to 5), and *high* (6+). Static-99R contains items assessing age at release, sexual criminality (e.g., prior sex offenses, victim information), and general criminality (e.g., prior sentencing dates, nonsexual violence). All items were scored dichotomously, except for prior sex offenses (which ranges from 0 to 3) and the revised age item (which ranges from −3 to 1).

Static-2002R

Similar to Static-99R, Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012) is an empirical actuarial risk assessment tool for adult male sex offenders (see also www.static99.org). It has 14 items grouped into 5 main subscales: age at release, persistence of sex offending, sexual deviance, relationship to victims, and general criminality.

The total score (ranging from −2 to 14) can be used to place offenders in one of five risk categories: *low* (−2 to 2), *low–moderate* (3 to 4), *moderate* (5 to 6), *moderate–high* (7 to 8), and *high* (9+). The items are identical to Static-2002 with the exception of updated age weights (Helmus, Thornton, et al., 2012). Previous research found that Static-2002 was significantly more predictive of sexual, violent, and any recidivism than Static-99 (Hanson, Helmus, & Thornton, 2010). In contrast, Static-99R and Static-2002R have similar predictive accuracy (Babchishin, Hanson, & Helmus, 2012).

Analyses of individual Static-2002R items examined only those items that were scored differently than the comparable Static-99R item (given that larger sample sizes were available for the Static-99R items).[2] In addition, the items for any prior involvement in the criminal justice system and prior sentencing occasions were summed into a single item (called "prior sentencing occasions") to reflect general criminal history. All items were scored dichotomously, except for prior sexual sentencing occasions (which ranges from 0 to 3), prior sentencing occasions (which ranges from 0 to 3), and age (the original item ranges from 0 to 3; the revised item ranges from −2 to 2).

OVERVIEW OF ANALYSES

Cox regression (Allison, 1984) was used to look at predictive and incremental accuracy. Cox regression estimates relative risk ratios (hazard ratios) associated with one or more predictor variables from survival data with unequal follow-up times. The risk ratio can be interpreted as the relative increase in recidivism for each 1-point increase in the predictor variable, averaged across the follow-up period. For example, a risk ratio of 2 for a dichotomous variable means that offenders with that risk factor are twice as likely to sexually reoffend, averaged across the follow-up period. Risk ratios are tied to the units of the predictor variable, so the magnitude of the effect size cannot be directly compared for items with different scoring systems (e.g., risk ratios for continuously measured age would be expected to have very small risk ratios as it is describing differences between offenders 1 year apart in age). When multiple predictors are added into a single Cox regression model, the risk ratios reflect their incremental effects (i.e., controlling for other variables in the model). Analyses of the incremental effect of Static-99R items used the sample as a stratum variable, which allows separate baseline hazard functions to be estimated for each sample. Another advantage of Cox regression is that it is invariant to restriction of range in the predictor variable (it is not invariant, however, to restriction of range in underlying, unmeasured variables that are correlated with the predictors).

Analyses of the stability of predictive accuracy across samples were conducted using both fixed-effect and random-effects meta-analyses (Borenstein, Hedges, Higgins, & Rothstein, 2009) of the risk ratios from Cox regression. Whereas the results of fixed-effect meta-analysis are conceptually restricted to the particular set of studies included in the meta-analysis, random-effects meta-analysis estimates effects for the population of which the current sample of studies is a part. More specifically, random-effects meta-analysis incorporates variability across samples into the error term, whereas fixed-effect meta-analysis does not. When variability across studies is low ($Q < df$), random-effects and fixed-effect meta-analyses produce identical results. As the variability across studies increases, the confidence intervals for random-effects meta-analysis exceed the fixed-effect estimates, and the random-effects method gives more weight to smaller studies. Conceptually, as variability across studies approaches infinity, the random-effects mean approaches the

unweighted average. Although random-effects meta-analysis is more conceptually appropriate and provides a more conservative test, its method of incorporating variability across studies is less reliable when the number of studies is lower than 30 (Schulze, 2007).

Variability in findings across studies was tested using Cochran's Q statistic and the I^2 statistic (Borenstein et al., 2009). A significant Q statistic indicates that there is more variability across studies than would be expected by chance. The I^2 statistic is an effect size measure that describes the proportion of the overall variability (the Q) that is beyond what you would expect by chance from sampling error (i.e., the proportion of variability that can be considered "true" differences). Specifically, I^2 is $(Q - df) / Q$, where df = the number of studies minus 1. For easier interpretation, I^2 was reported as a percentage. As a rough heuristic, I^2 values of 25%, 50%, and 75% can be considered low, moderate, and high variability, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

Moderator analyses were examined for Static items demonstrating significant variability in predictive accuracy across samples. Analyses of sample type (i.e., whether the sample was preselected as unusually high risk or need), recidivism criteria, and country were conducted using the fixed-effect $Q_{between}$ test (Borenstein et al., 2009). This test partitions the overall variability across studies into variability within each level of the moderator, with the residual representing the variability between levels of the moderator. The $Q_{between}$ is tested on a chi-square distribution (where df are equal to the levels of the moderator minus 1), with significance denoting that the moderator explains a significant portion of the variability in predictive accuracy. All relationships between moderator variables (tested using the chi-square) were nonsignificant ($p > .30$), suggesting the moderator variables were fairly independent of each other.

Moderator analyses of offender type (rapists compared with child molesters) were tested using a within-sample test. This is analogous to traditional repeated-measures tests, but instead of removing individual differences, we are removing differences across studies. Given that rapist and child molester subgroups from the same study share numerous methodological and unique sample features that may contribute additional variability, removing that study-level variability increases statistical power. These analyses required calculating a new effect size (difference in d) by subtracting the effect size of rapists from the effect size of child molesters in the studies with data on both subgroups. The variance of this effect size was the sum of the variance of each d, minus their covariance, which was defined as $2 \times r \times SD_{\text{rapists}} \times SD_{\text{childmolesters}}$, where r is the correlation between the two effect sizes. If the 95% confidence interval for the fixed-effects meta-analysis of the difference between the ds does not include zero, then the difference is statistically significant.

RESULTS

DO THE STATIC ITEMS PREDICT SEXUAL RECIDIVISM?

Table 2 presents the results of the Cox regression meta-analysis for the Static-99R and Static-2002R items, including both the original and revised age items (as well as age measured as a continuous variable). All items significantly predicted sexual recidivism in both fixed-effect and random-effects meta-analysis, with the exception of the Static-99R item for nonsexual violence as part of the index offense, which was a significant predictor in the fixed-effect analyses but not in the random-effects analyses (rate ratio = 1.2 in both analyses).

TABLE 2: Meta-Analysis of Predictive Accuracy for Sexual Recidivism of Static-99 and Static-2002 Items From Cox Regression Analyses

	Fixed Effect		Random Effects						
Item	Rate Ratio	95% CI	Rate Ratio	95% CI	*Q*	*p*	I^2 (%)	*k*	*n*
Age items									
Age (continuous)	0.98	[0.97, 0.98]	0.98	[0.97, 0.98]	30.7	.031	41.4	19	7,461
Static-99 age item	1.31	[1.10, 1.57]	1.31	[1.10, 1.57]	17.7	.666	0.0	22	8,053
Static-2002 age item	1.26	[1.13, 1.39]	1.30	[1.14, 1.49]	10.8	.146	35.4	8	2,950
Static-99R/2002R revised age item	1.26	[1.18, 1.34]	1.28	[1.18, 1.40]	26.5	.089	32.0	19	7,461
Static-99R items									
Ever lived with lover	1.67	[1.46, 1.90]	1.70	[1.45, 2.00]	28.0	.141	24.9	22	8,004
Index nonsexual violence	1.20	[1.03, 1.40]	1.22	[0.92, 1.63]	60.2	<.001	65.1	22	8,053
Prior nonsexual violence	1.52	[1.33, 1.74]	1.55	[1.31, 1.83]	31.1	.073	32.4	22	8,053
Prior sex offenses	1.65	[1.56, 1.75]	1.66	[1.51, 1.83]	55.4	<.001	62.1	22	8,053
Prior sentencing dates	1.84	[1.61, 2.10]	1.84	[1.55, 2.19]	32.0	.059	34.3	22	8,053
Any noncontact sex offense	2.37	[2.03, 2.76]	2.47	[1.98, 3.10]	38.4	.012	45.3	22	8,053
Any unrelated victim	2.29	[1.91, 2.76]	2.29	[1.90, 2.76]	21.2	.450	0.7	22	8,053
Any stranger victim	2.02	[1.77, 2.31]	2.06	[1.68, 2.54]	42.7	.004	50.8	22	8,053
Any male victim	1.45	[1.26, 1.68]	1.43	[1.18, 1.74]	34.3	.034	38.7	22	8,053
Static-2002R items									
Prior sex sentencing dates	1.62	[1.47, 1.80]	1.67	[1.46, 1.92]	11.7	.112	40.0	8	2,951
Juvenile sex arrest	1.67	[1.21, 2.31]	1.67	[1.21, 2.31]	4.8	.680	0.0	8	2,951
High rate of sex offenses	2.43	[1.99, 2.97]	2.92	[2.01, 4.23]	21.5	.003	67.5	8	2,951
Knight and Thornton (2006) removed	3.07	[2.41, 3.91]	3.25	[2.37, 4.47]	9.9	.131	39.2	7	2,485
2+ victims < 12, one unrelated	1.42	[1.14, 1.77]	1.42	[1.14, 1.77]	2.8	.899	0.0	8	2,944
Prior sentencing occasions	1.54	[1.36, 1.73]	1.55	[1.34, 1.79]	9.4	.223	25.7	8	2,951
Breach of conditional release	2.07	[1.69, 2.53]	2.27	[1.70, 3.03]	11.8	.107	40.7	8	2,947
Years free prior to index	2.31	[1.89, 2.82]	2.35	[1.85, 2.98]	9.2	.241	23.6	8	2,949
Prior nonsexual violence	1.55	[1.27, 1.88]	1.55	[1.26, 1.92]	7.9	.343	11.2	8	2,951

Note. All items are scored dichotomously, with the following exceptions: age (measured as a continuous value), Static-2002 age item (ranges from 0 to 3), revised age item (ranges from –3 to 1 in Static-99R and –2 to 2 in Static-2002R, but both items have the same cut-points), prior sex offenses (ranges from 0 to 3), prior sex sentencing dates (ranges from 0 to 3), and prior sentencing occasions (ranges from 0 to 3). CI = confidence interval.

All items are scored dichotomously, with the following exceptions: age (measured as a continuous value), Static-2002 age item (ranges from 0 to 3), revised age item (ranges from −3 to 1 in Static-99R and −2 to 2 in Static-2002R, but both items have the same cut points),

prior sex offenses (ranges from 0 to 3), prior sex sentencing dates (ranges from 0 to 3), and prior sentencing occasions (ranges from 0 to 3).

The continuous age variable had a rate ratio of 0.98. This means that with each 1-year increase in age at release, the recidivism rate declines 2%. For the age items on the original and revised Static-99 and Static-2002 scales, they all demonstrated a rate ratio of roughly 1.3 (i.e., each one-point increase for age increases risk of sexual recidivism by 1.3 times, or 30%). For the remaining items (excluding index nonsexual violence), rate ratios ranged between 1.4 and 3.2. Note that for dichotomous items, a rate ratio can be considered the ratio of recidivism between offenders who had the risk factor compared with those who did not. For example, offenders with stranger victims were twice as likely to reoffend as offenders without stranger victims (rate ratio = 2.0).

DO THE STATIC ITEMS PREDICT SIMILARLY ACROSS SAMPLES?

The meta-analytic results in Table 2 also include the statistics assessing variability in predictive accuracy across studies (the Q statistic, its associated p value, and the I^2 statistic). For the age items, the predictive accuracy of age (as a continuous measure) showed significant variability across studies; however, the predictive accuracy of the categorical age items for the original and revised Static scales all had nonsignificant variability across studies. For Static-99R, there was significant variability across studies in the predictive accuracy of half the items: index nonsexual violence, prior sex offenses, any noncontact sex offense, any stranger victim, and any male victim. For all the items, the magnitude of variability was roughly in the moderate range (I^2 values indicating that between 39% and 65% of the observed variability was beyond what would be expected by chance).

For the Static-2002R items, note that the findings of variability for any noncontact sex offense, any stranger victim, and any male victim would be applicable to this scale as well. In addition, the item for high rate of sex offenses showed significant variability across studies. This item had the highest rate ratio among the dichotomous items (2.4 in the fixed-effect analyses and 2.9 in the random-effects) and interestingly, was a significant and positive predictor of recidivism in all eight studies examined (which is surprising given that many of the individual studies would have much lower statistical power). This variability was explained by a single statistical outlier[3] (Knight & Thornton, 2007), which had the lowest predictive accuracy (in this sample, offenders with a high rate of sex offending were 1.4 times more likely to recidivate than offenders with a lower rate of sex offending). When this outlier was removed, the predictive accuracy of the item increased substantially, such that high-rate sex offenders were roughly 3 times more likely to sexually reoffend than low-rate sex offenders, and the variability in predictive accuracy among the remaining seven studies was no longer significant. As such, variability was not considered further for this item. Subsequently, moderator analyses were explored for the five items demonstrating significant variability across studies (not easily explained by a single outlier; see Table 3).

Index Nonsexual Violence

Among the 22 studies, the effect size for this item was in the expected (positive) direction in 14 studies, but in the negative direction for 8 studies. Based on the moderator analyses (see Table 3), predictive accuracy of this item was not related to offender type or to whether the sample was high risk or need. There was a significant effect based on the country of the

TABLE 3: Moderator Analyses of Selected Static-99R Items (Fixed-Effect Meta-Analysis)

Item	Rate Ratio	95% CI	*Q*	*p*	I^2 (%)	*k*	*n*
Index nonsexual violence	1.20	[1.03, 1.40]	60.2	<.001	65.1	22	8,053
Not high risk or need samples	1.36	[1.09, 1.69]	41.3	<.001	66.1	15	6,317
High risk or need samples	1.07	[0.86, 1.33]	16.6	.011	63.9	7	1,736
$Q_{between}$			2.3	.128			
Charges	1.40	[1.11, 1.77]	31.7	<.001	74.7	9	3,377
Convictions	1.06	[0.87, 1.31]	25.5	.013	52.9	13	4,676
$Q_{between}$			3.0	.082			
Canada	1.37	[1.08, 1.75]	19.2	.024	53.1	10	3,148
The United States	1.70	[1.21, 2.40]	18.6	<.001	83.9	4	1,598
The United Kingdom	0.80	[0.40, 1.60]	6.3	.012	84.2	2	314
Other	0.90	[0.70, 1.16]	4.8	.434	0.0	6	2,993
$Q_{between}$			11.3	.010			
Offender Type: Difference	0.309	[–0.236, 0.854]	14.5	.150	31.2	11	4,686
Prior sex offenses	1.65	[1.56, 1.75]	55.4	<.001	62.1	22	8,053
Not high risk or need samples	1.82	[1.69, 1.96]	31.8	.004	56.0	15	6,317
High risk or need samples	1.44	[1.31, 1.57]	7.5	.280	19.6	7	1,736
$Q_{between}$			16.1	<.001			
Charges	1.48	[1.36, 1.61]	6.4	.600	0.0	9	3,377
Convictions	1.81	[1.68, 1.96]	36.9	<.001	67.5	13	4,676
$Q_{between}$			12.0	<.001			
Canada	1.55	[1.41, 1.70]	5.8	.756	0.0	10	3,148
The United States	1.52	[1.34, 1.73]	1.6	.664	0.0	4	1,598
The United Kingdom	1.77	[1.39, 2.26]	0.1	.729	0.0	2	314
Other	1.85	[1.68, 2.04]	38.9	<.001	87.1	6	2,993
$Q_{between}$			9.0	.030			
Offender Type: Difference	0.028	[–0.112, 0.168]	15.1	.127	34.0	11	4,686
Noncontact sex offenses	2.37	[2.03, 2.76]	38.4	.012	45.3	22	8,053
Not high risk or need samples	2.83	[2.34, 3.43]	13.4	.497	0.0	15	6,317
High risk or need samples	1.71	[1.32, 2.21]	15.5	.017	61.3	7	1,736
$Q_{between}$			9.5	.002			
Charges	1.97	[1.55, 2.49]	18.0	.021	55.5	9	3,377
Convictions	2.71	[2.21, 3.31]	16.4	.176	26.6	13	4,676
$Q_{between}$			4.0	.044			
Canada	2.57	[1.97, 3.36]	7.6	.570	0.0	10	3,148
The United States	2.07	[1.54, 2.77]	10.3	.016	70.8	4	1,598
The United Kingdom	3.33	[1.73, 6.42]	0.3	.578	0.0	2	314
Other	2.30	[1.77, 2.99]	17.9	.003	72.1	6	2,993
$Q_{between}$			2.3	.520			
Offender Type: Difference	–0.064	[–0.522, 0.394]	5.5	.858	0.0	11	4,686
Any stranger victim	2.02	[1.77, 2.31]	42.7	.004	50.8	22	8,053
Not high risk or need samples	2.62	[2.20, 3.13]	15.3	.355	8.8	15	6,317
High risk or need samples	1.41	[1.15, 1.74]	7.4	.285	18.9	7	1,736
$Q_{between}$			19.9	<.001			
Charges	1.83	[1.50, 2.24]	26.4	<.001	69.7	9	3,377
Convictions	2.19	[1.83, 2.62]	14.6	.265	17.7	13	4,676
$Q_{between}$			1.7	.196			
Canada	1.74	[1.41, 2.16]	17.7	.039	49.1	10	3,148
The United States	1.89	[1.45, 2.46]	8.6	.035	65.2	4	1,598
The United Kingdom	3.43	[1.82, 6.47]	0.6	.454	0.0	2	314

(continued)

TABLE 3: (continued)

Item	Rate Ratio	95% CI	*Q*	*p*	I^2 (%)	*k*	*n*
Other	2.40	[1.88, 3.07]	9.1	.104	45.2	6	2,993
$Q_{between}$			6.7	.084			
Offender Type: Difference	0.352	[0.124, 0.579]	13.7	.188	27.0	11	4,686
Offender Type: All subsamples	2.06	[1.69, 2.50]	28.2	.134	25.6	22	4,686
Child molesters	2.37	[1.80, 3.14]	14.4	.154	30.7	11	2,640
Rapists	1.79	[1.36, 2.36]	11.8	.297	15.5	11	2,046
$Q_{between}$			2.0	.162			
Any male victim	1.45	[1.26, 1.68]	34.3	.034	38.7	22	8,053
Not high risk or need samples	1.52	[1.26, 1.84]	23.6	.051	40.8	15	6,317
High risk or need samples	1.36	[1.08, 1.70]	10.0	.123	40.3	7	1,736
$Q_{between}$			0.6	.442			
Charges	1.52	[1.23, 1.86]	8.8	.357	9.4	9	3,377
Convictions	1.40	[1.15, 1.71]	25.2	.014	52.3	13	4,676
$Q_{between}$			0.3	.590			
Canada	1.38	[1.08, 1.75]	11.7	.233	22.8	10	3,148
The United States	1.47	[1.14, 1.90]	4.1	.252	26.6	4	1,598
The United Kingdom	0.57	[0.26, 1.23]	0.3	.560	0.0	2	314
Other	1.73	[1.32, 2.27]	10.6	.059	53.1	6	2,993
$Q_{between}$			7.5	.056			
Offender Type: Difference	0.444	[–0.244, 1.132]	5.5	.856	0.0	11	4,686

Note. The effect size is the rate ratio (the exponent of the B_1 from Cox regression), except for the moderator examining offender type differences, which are reported in the units of differences between the untransformed B_1 from Cox regression, with positive values indicating higher predictive accuracy for child molesters compared with rapists. CI = confidence interval.

study, however. This item was a significant predictor of sexual recidivism in both Canada (rate ratio = 1.4) and the United States (rate ratio = 1.7), but showed no predictive accuracy in the United Kingdom and the countries classified as "other" (rate ratios below 1). This suggests that the item may perform fine in North America, but not in other Western countries. In addition, the moderator for recidivism criteria approached significance ($Q_{between}$ = 3.0, df = 1, p = .082) in the direction of higher predictive accuracy for this item when charges were used as the recidivism criteria.

Prior Sex Offenses

Although the variability in predictive accuracy for this item was significant, the pattern of findings was more encouraging than that found for index nonsexual violence. The rate ratios were in the expected direction (positive) in all 22 studies. In addition, this item was a statistically significant predictor in 19 of the 22 studies. The studies that did not demonstrate statistically significant accuracy for this item (Bonta & Yessine, 2005; Hill, Habermann, Klusmann, Berner, & Briken, 2008; Ternowski, 2004) tended to have smaller sample sizes than most studies, indicating lower statistical power. In other words, the variability observed for this item was more an issue of variability in the magnitude of the predictive accuracy, as opposed to variability in whether the item predicted at all.

Moderator analyses found no difference in the accuracy of this item for rapists compared with child molesters, but did find significant effects for all other moderators examined.

Predictive accuracy for this item was significantly lower in samples that were preselected as high risk or need (although the item was still a significant predictor among these samples, with each one-point increase on this item increasing recidivism rates by 1.4 times). Predictive accuracy was also higher in samples using convictions as the recidivism outcome. Predictive accuracy was also the highest for studies from the United Kingdom and other countries outside North America (rate ratio of approximately 1.8), and only slightly lower for Canada and the United States (rate ratios of approximately 1.5).

Any Noncontact Sex Offense Convictions

The predictive accuracy of this item was in the expected direction in all but one study, and reached statistical significance in half the studies. The aggregate effect of this item was quite large; offenders with a noncontact conviction reoffended roughly 2.4 times more than offenders without a noncontact offense (from the overall fixed-effect meta-analysis). Predictive accuracy was significantly lower for samples of offenders preselected as high risk or need, although even among these samples, offenders with noncontact offenses were still 1.7 times more likely to sexually reoffend (which was a significant effect). In addition, predictive accuracy was significantly lower (but still significant) in studies using charges as a recidivism outcome (rate ratio = 2.0). Predictive accuracy was not significantly related to country or offender type.

Any Stranger Victims

The predictive accuracy of this item was in the expected direction in all but one study, and reached statistical significance in half the studies. Similar to noncontact sex offenses, this item had a large overall effect (rate ratio = 2 from fixed-effect meta-analysis). Predictive accuracy was significantly lower in the samples preselected as high risk or need (rate ratio = 1.4), although still significant. In addition, in the within-sample moderator analysis, predictive accuracy was significantly higher for child molesters. To examine this finding further, results were presented separately for the subsamples of rapists and child molesters for the 11 studies that included data separately for both subgroups. Although predictive accuracy was lower for rapists, those with any stranger victim were still 1.8 times more likely to sexually reoffend (which was a significant effect size), demonstrating that the item is applicable to both subgroups. Note that when a $Q_{between}$ test is applied to compare rapists and child molesters, the difference is not significant (highlighting its lower statistical power compared with the within-samples moderator analysis). Predictive accuracy of this item was not significantly related to the definition of recidivism. Country approached significance as a moderator ($Q_{between} = 6.7$, $df = 3$, $p = .084$), with a tendency for higher effect sizes outside North America.

Any Male Victim

This item had positive effect sizes in all but five studies, although it only reached statistical significance in six studies. None of the moderators were statistically significant, although country approached significance ($Q_{between} = 7.5$, $df = 3$, $p = .056$), with the 2 studies from the United Kingdom showing no accuracy for this item (rate ratio below 1), whereas all other groups had rate ratios of 1.4 or above.

TABLE 4: Tests of Incremental Validity for Sexual Recidivism of Static-99R Items (*N* = 7,457)

Item	Rate Ratio	95% CI	χ^2	*p*
Age (99R item)	1.24	[1.16, 1.32]	43.72	<.001
Ever lived with lover	1.15	[0.99, 1.32]	3.69	.055
Index nonsexual violence	1.10	[0.94, 1.30]	1.34	.247
Index violence in North American samples only[a]	1.27	[1.04, 1.56]	5.21	.022
Prior nonsexual violence	1.15	[0.99, 1.33]	3.50	.062
Prior sex offenses	1.47	[1.38, 1.56]	130.04	<.001
Prior sentencing dates	1.31	[1.13, 1.52]	12.95	<.001
Noncontact sex offense	1.67	[1.42, 1.67]	36.27	<.001
Any unrelated victim	1.52	[1.25, 1.85]	18.26	<.001
Any stranger victim	1.21	[1.04, 1.41]	6.08	.014
Any male victim	1.24	[1.07, 1.44]	7.69	.006

Note. CI = confidence interval.
a. *n* = 4,388 for this analysis.

DO THE STATIC ITEMS CONTRIBUTE UNIQUELY TO THE PREDICTION OF SEXUAL RECIDIVISM?

In addition to examining how the individual items perform across different samples, we also examined whether the items provide unique information in predicting sexual recidivism. Given the low statistical power of incremental tests, these analyses examined only the Static-99R items, which had more data. Each item was tested to see whether it added incremental predictive accuracy to the Static-99R total score (computed without that item). Results are presented for the incremental effect of each item (but not the total score without the item) in Table 4. The Static-99R items added incrementally to the prediction of sexual recidivism, with three exceptions. Ever lived with a lover and prior nonsexual violence approached statistical significance ($p = .055$ and $p = .062$, respectively). Index nonsexual violence, however, did not add incrementally to the prediction of sexual recidivism ($p = .247$). Given our previous finding of a strong moderator effect whereby this item only predicted in North America, a follow-up analysis indicated that this item did provide significant incremental validity in the North American samples ($p = .022$).

DISCUSSION

The overall findings of this study can be summarized as follows: In general, when analyzed one at a time, the items included in the Static scales (Static-99R and Static-2002R) predict sexual recidivism; about half the items vary across samples in how predictive they are in ways that cannot be explained by sampling error; and, in general, when analyzed relative to the scale score calculated without the item, items showed incremental predictive value.

The main exception to these general trends was the index nonsexual violence item. At the aggregate level, it significantly predicted sexual recidivism only in the fixed-effect analysis (not in the random-effects analysis) and did not add incrementally to the other items. However, this item was both individually and incrementally predictive for samples drawn from North America only. Notably, this item did perform better at predicting violent recidivism compared with sexual recidivism (analyses not reported). However, both Static-99R and Static-2002R were specifically designed to predict sexual recidivism. If the goal

is to predict violent recidivism, these scales are suboptimal as they give disproportionate weight to sexual criminality, which is less relevant for predicting violent recidivism (Brouillette-Alarie, Babchishin, Hanson, & Helmus, in press). Other scales, such as the BARR-2002R (Brief Assessment of Recidivism Risk; Babchishin, Hanson, & Blais, in press) or the VRAG-R (Violence Risk Appraisal Guide - Revised; Rice, Harris, & Lang, 2013) would be preferable for assessing risk of violent recidivism.

There are five more detailed aspects of the findings that merit discussion and which have potential implications for practice. First, it is worth distinguishing variation in the degree to which an item is associated with sexual recidivism from variation in the direction of the association. For half the Static-99R items, there was no significant variation in the magnitude of their association with recidivism and for three of the items (prior sex offenses, noncontact sex offenses, stranger victim) where significant variation was present, this was a matter of variation in the magnitude of the effect rather than its direction. For the two items (male victim, index nonsexual violence) where there appeared to be variation in the direction of the effect, this variation seemed to be at least partially accounted for by results from outside North America. Thus, for samples from within North America, it appears that for all items, the effect is best summarized as there being variation between samples in the magnitude of the association with recidivism rather than variation in the direction of the effect. In other words, the variability does not pose concerns about the use of these items in North American samples.

This finding that some items predicted, but varied in magnitude of predictive accuracy, provides additional evidence that simple weighting systems (e.g., present or absent) are likely to be more generalizable than more complicated procedures, such as using regression weights (e.g., Minnesota Sex Offender Screening Tool - 3; Duwe & Freske, 2012) or the Nuffield (1982) method, subsequently used in the VRAG, SORAG (Sex Offender Risk Appraisal Guide; Quinsey, Harris, Rice, & Cormier, 2006) and VRAG-R (Rice et al., 2013). The finding that whether an item predicts (yes or no) is generally more stable across samples than assessing the magnitude of predictive accuracy bolsters previous research demonstrating that scales premised on the exact relative predictiveness of specific items are likely to be less robust (Grann & Långström, 2007).

Second, there was little evidence that offender type (adult vs. child victim choice) moderated the predictive value of the items. The statistical tests for this were more powerful (because they could be done within samples) and the only significant result was for the stranger victim item to be more predictive for child molesters rather than for rapists (although it predicted for both groups). This finding is striking because it seems to contradict some natural theoretical speculations. It is plausible to suppose that items reflecting sexual deviancy would be more predictive for child molesters whereas items reflecting antisociality would be more predictive for rapists based on findings of differences in the presence of these risk domains for both groups (Firestone, Bradford, Greenberg, & Serran, 2000; D. A. Harris, Smallbone, Dennison, & Knight, 2009). Similarly, it is sometimes supposed that age is more predictive for rapists than for extrafamilial child molesters (Hanson, 2002), particularly for violent recidivism (Rettenberger, Briken, Turner, & Eher, 2014). None of these suppositions were supported by the results for the Static items. Our findings are, however, consistent with other research supporting applicability of actuarial risk tools to diverse subgroups of offenders (G. T. Harris & Rice, 2013).

Third, the effect of jurisdiction did not consistently enhance prediction or consistently reduce it but had different effects for different items (i.e., index nonsexual violence and male victims tended to perform better in North America, but prior sex offenses performed better outside North America). However, there was a consistent pattern of the United States and Canada having similar results whereas the results for other jurisdictions sometimes differed from North America. The similarity in findings between Canada and the United States was surprising given stark differences in their correctional practices. This study provides some evidence that findings of relative predictive accuracy are quite generalizable between these two countries. Conversely, however, our results suggest some caution in generalizing results obtained in North America outside that geographical area. Researchers seeking to establish wider generality of findings would do well to at least carry out parallel studies inside and outside North America.

Fourth, using charges rather than convictions as the outcome criterion seemed to make a difference in how predictive two items were (prior sex offenses and noncontact sex offenses) but this finding is particularly hard to interpret for several reasons. In both significant analyses, predictive accuracy was lower in samples using charges as opposed to those using convictions; if recidivism criterion was an indicator of comprehensiveness of the information (i.e., samples using charges reflecting more comprehensive criminal records), the findings should have been in the opposite direction. In addition, the use of charges rather than convictions is liable to be confounded with how predictor variables are measured (i.e., other features of the criminal history records utilized). Future studies should code predictors and outcomes separately for charges and convictions and then directly compare differences in predictive accuracy. Nonetheless, this moderator variable was significant for only two Static items and the proportion of the overall variability explained by this moderator was lower than for the sample type moderator. In other words, this moderator is likely less important than sample preselection.

Fifth, the effect of sample type was fairly consistent: Three items (prior sex offenses, noncontact sex offenses, stranger victim) were less predictive in the high risk or need samples. This effect was nonsignificant in the other two items but was in the same direction. This result was consistent with our prediction based on the expected reduction in the correlation between psychological and static actuarial predictors. This result is also consistent with the finding for static actuarial scales as a whole (Hanson et al., 2014) and reinforces the value of developing separate norms for populations selected in different ways or of norms that incorporate direct assessments of psychologically meaningful risk factors in conjunction with static risk factors (Hanson et al., 2014; Hanson et al., in press; Helmus, Hanson, Babchishin, & Thornton, 2014; Olver et al., 2007).

Although this was the first meta-analysis of the individual items of the Static-99R and Static-2002R and the first examination of moderator variables for these items, there were limitations to the present study. The most important of these are the limited number of different samples included in the analysis and the limited number of moderator variables examined. In addition, for some moderators, it could be possible to examine their effect using a within-sample design (e.g., charges vs. convictions, type of information used), which allows a better controlled and more powerful examination of their effect. We applied this for offender type, but in principle, it would be possible to apply it for other moderators. Given that none of the moderators significantly explained the variability in the male victim item, future research should further explore the conditions under which it best predicts recidivism.

CONCLUSION

This study supports the utility of the Static-99R and Static-2002R items, with very few exceptions. Outside North America, index nonsexual violence did not predict sexual recidivism, nor did it add incremental predictive accuracy to the rest of the scale. In countries outside Canada and the United States, evaluators may want to exercise caution if an offender's score on this item would make a meaningful difference in how the offender is treated or supervised. Another viable alternative is to use Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012), which does not include this item. The remaining items predicted sexual recidivism overall and generally added incremental value to the rest of the scale (or at least approached significance in the incremental effect). If predicting violence is of concern, we recommend using the BARR-2002R instead of Static-99R or Static-2002R (Babchishin et al., in press).

NOTES

1. Although Rettenberger, Haubner-MacLean, and Eher (2013) is an important exception to the findings of most studies on age and sexual recidivism, these data were included in the data for revising the Static-99 and are also included in the current study.

2. Although the item for any noncontact sexual offense conviction has slightly different coding rules for Static-2002R compared with Static-99R, this item was scored the same way in the Static-2002R data sets. This is because these data were scored prior to the release of the 2009 Static-2002 coding rules (Phenix, Doren, Helmus, Hanson, & Thornton, 2009), which first made some minor changes in the scoring of this item. Consequently, for the purposes of this study, this item was considered the same in the two scales.

3. Statistical outliers were identified following the criteria established by Hanson and Bussière (1998).

REFERENCES

References marked with an asterisk indicate studies included in the analyses.

*Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 347-367. doi:10.1007/s11194-007-9052-5

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: SAGE.

Babchishin, K. M., Hanson, R. K., & Blais, J. (in press). Less is more: Using Static-2002R subscales to predict violent and general recidivism among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*.

Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to actuarial risk prediction. *Assessment*, *19*, 442-461. doi:10.1177/1073191112458312

Barbaree, H. E., & Blanchard, R. (2008). Sexual deviance over the lifespan: Reduction in deviant sexual behavior in the aging sex offender. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (pp. 37-60). New York, NY: Guilford Press.

Barbaree, H. E., Langton, C. M., & Blanchard, R. (2007). Predicting recidivism in sex offenders using the VRAG and SORAG: The contribution of age-at-release. *International Journal of Forensic Mental Health*, *6*, 29-46.

Barbaree, H. E., Langton, C. M., Blanchard, R., & Cantor, J. M. (2009). Aging versus stable enduring traits as explanatory constructs in sex offender recidivism: Partitioning actuarial prediction into conceptually meaningful components. *Criminal Justice and Behavior*, *36*, 443-465. doi:10.1177/0093854809332283

Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual offender version. *Sexual Abuse: A Journal of Research and Treatment*, *22*, 234-251. doi:10.1177/1079063210369014

*Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law*, *14*, 85-106. doi:10.1080/10683160701483104

*Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders] (UMI No. NR30941). *Dissertations Abstracts International*, *68*(09).

*Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders* (Unpublished master's thesis). University of Leicester, UK.

*Bonta, J., & Yessine, A. K. (2005). [Recidivism data for 124 released sexual offenders from the offenders identified in the National Flagging System: Identifying and responding to high-risk, violent offenders] (User Report 2005-04). Unpublished raw data. Ottawa, Ontario: Public Safety and Emergency Preparedness Canada].

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley.

Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L. M. (in press). Latent constructs of static risk scales for the prediction of sexual recidivism: A 3-factor solution. *Assessment*.

*Brouillette-Alarie, S., & Proulx, J. (2008, October). *Predictive and convergent validity of phallometric assessment in relation to sexual recidivism risk*. Poster presented at the annual conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Cortoni, F., & Nunes, K. L. (2007). *Assessing the effectiveness of the National Sexual Offender Program* (Research Report No. R-183) (Unpublished report). Ottawa, ON: Correctional Service of Canada.

Craig, L. A., Thornton, D., Beech, A., & Browne, K. D. (2007). The relationship of statistical and psychological risk markers to sexual reconviction in child molesters. *Criminal Justice and Behavior*, *34*, 314-329. doi:10.1177/0093854806291416

*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.

Dobash, R. P., & Dobash, R. E. (1995). Reflections on findings from the Violence Against Women Survey. *Canadian Journal of Criminology*, *37*, 457-484.

Duwe, G., & Freske, P. J. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota Sex Offender Screening Tool 3 (MnSOST-3). *Sexual Abuse: A Journal of Research and Treatment*, *24*, 350-377. doi:10.1177/1079063211429470

Eher, R., Matthes, A., Schilling, F., Haubner-Maclean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 5-28. doi:10.1177/1079063211403164

*Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment*, *3*(1), 1-20. Retrieved from http://www.sexual-offender-treatment.org/sot-1-2008.html

*Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2009). [Data from sex offenders released from prison in Austria]. Unpublished raw data.

*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.

Firestone, P., Bradford, J. M., Greenberg, D. M., & Serran, G. A. (2000). The relationship of deviant sexual arousal and psychopathy in incest offenders, extrafamilial child molesters, and rapists. *Journal of the American Academy of Psychiatry and the Law*, *28*, 303-308.

Friendship, C., Thornton, D., Erikson, M., & Beech, A. (2001). Reconviction: A critique and comparison of two main data sources in England and Wales. *Legal and Criminological Psychology*, *6*, 121-129.

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior*, *34*, 22-36. doi:10.1177/0093854806290250

*Haag, A. M. (2005). [Recidivism data from 198 offenders detained until their warrant expiry date. From: Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism] (UMI No. NR05662). *Dissertations Abstracts International*, *66*(08), 4531B. Unpublished raw data.

Hanson, R. K. (2002). Recidivism and age: Follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, *17*, 1046-1062. doi:10.1177/088626002236659

Hanson, R. K. (2006). Does Static-99 predict recidivism among older sexual offenders? *Sexual Abuse: A Journal of Research and Treatment*, *18*, 343-355. doi:10.1007/s11194-006-9027-y

Hanson, R. K., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse: A Journal of Research and Treatment*, *25*, 482-515. doi:10.1177/1079063212469060

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

*Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Ottawa, Ontario: Public Safety Canada.

Hanson, R. K., Helmus, L., & Harris, A. J. R. (in press). *Can community supervision officers assess the criminogenic needs of sexual offenders?* Unpublished manuscript.

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior*, *34*, 198-211. doi:10.1007/s10979-009-9180-1

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*, 1-21. doi:10.1037/a0014421

Hanson, R. K., & Nicholaichuk, T. (2000). A cautionary note regarding Nicholaichuk et al. (2000). *Sexual Abuse: A Journal of Research and Treatment*, *12*, 289-293.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*, 119-136. doi:10.1023/A:1005482921333

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002* (Corrections Research User Report No. 2003-01). Ottawa, Ontario: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2012, October). *Preselection effects can explain variability in sexual recidivism base rates in Static-99R and Static-2002R validation studies*. Presentation at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Denver, CO.

Hanson, R. K., Thornton, D., Helmus, L., & Babchishin, K. M. (2014). *What sexual recidivism rates are associated with Static-99R and Static-2002R scores?* Unpublished manuscript.

*Harkins, L., & Beech, A. R. (2007). *Examining the effectiveness of sexual offender treatment using risk band analysis*. Unpublished manuscript.

Harkins, L., Thornton, D., & Beech, A. (2009, September). *The use of dynamic risk domains assessed using psychometric measures to revise relative risk assessment using RM 2000 and Static 2002*. Paper presented at the 28th Annual Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa, Ontario: Solicitor General Canada.

Harris, D. A., Smallbone, S., Dennison, S., & Knight, R. A. (2009). Specialization and versatility in sexual offenders referred for civil commitment. *Journal of Criminal Justice*, *37*, 37-44. doi:10.1016/j.jcrimjus.2008.12.002

Harris, G. T., & Rice, M. E. (2013). Bayes and base rates: What is an informative prior for actuarial violence risk assessment? *Behavioral Sciences & the Law*, *31*, 103-124. doi:10.1002/bsl.2048

Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MR58443).

Helmus, L., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2014). Sex offender risk assessment with the Risk Matrix 2000: Validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression*. Advance online publication. doi:10.1080/13552600.2013.870241.

Helmus, L., Hanson, R. K., & Morton-Bourgon, K. E. (2011). International comparisons of the validity of actuarial risk tools for sexual offenders, with a focus on Static-99. In D. P. Boer, L. A. Craig, R. Eher, M. H. Miner, & F. Pfafflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice, and research* (pp. 57-83). Chichester, UK: John Wiley.

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior*, *39*, 1148-1171. doi:10.1177/0093854812443648

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment*, *24*, 64-101. doi:10.1177/1079063211409951

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557-560.

*Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology*, *52*, 5-20. doi:10.1177/0306624X07307450

Interstate Commission for Adult Offender Supervision. (2007). *Sex Offender Assessment Information Survey* (ICAOS Documents No. 4–2007). Lexington, KY: Author.

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 409-448. doi:10.1007/s11194-007-9062-3

*Johansen, S. H. (2007). Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods (UMI No. 3255527). *Dissertations Abstracts International*, *68*(03-B), 1929.

*Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Washington, DC: US Department of Justice.

*Långström, 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. doi:10.1177/107906320401600202

*Langton, C. M. (2003). Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy (UMI No. NQ78052). *Dissertations Abstracts International*, *64*(04-B), 1907.

Laudon, K. C. (1986). Data quality and due process in large interorganizational record systems. *Communications of the ACM*, *29*(1), 4-11.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, *22*, 191-217. doi:10.1177/1079063210366039

McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American survey*. Brandon, VT: Safer Society Foundation.

*Milton, J. (2003). *Care pathways and outcome of mentally disordered sex offenders referred to high secure psychiatric hospitals* (Unpublished master's thesis). University of Nottingham, UK.

*Nicholaichuk, T. (2001, November). *The comparison of two standardized risk assessment instruments in a sample of Canadian Aboriginal sexual offenders*. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

Nuffield, J. (1982). *Parole decision-making in Canada: Research toward decision guidelines*. Ottawa, Ontario: Solicitor General of Canada. Available from http://www.ps-sp.gc.ca

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale–Sexual offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, *19*, 318-329. doi:10.1037/1040-3590.19.3.318

Phenix, A., Doren, D., Helmus, L., Hanson, R. K., & Thornton, D. (2009). *Coding rules for Static-2002*. Ottawa, Ontario: Public Safety Canada. Retrieved from www.static99.org

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

Rettenberger, M., Briken, P., Turner, D., & Eher, R. (2014). Sexual offender recidivism among a population-based prison sample. *International Journal of Offender Therapy and Comparative Criminology*. Advance online publication. doi:10.1177/0306624X13516732.

Rettenberger, M., Haubner-MacLean, T., & Eher, R. (2013). The contribution of age to the Static-99 risk assessment in a population-based prison sample of sexual offenders. *Criminal Justice and Behavior*, *40*, 1413-1433. doi:10.1177/0093854813492518

Rice, M. E., & Harris, G. T. (2014). What does it mean when age is related to recidivism among sex offenders? *Law and Human Behavior*, *38*, 151-161. doi:10.1037/lhb0000052

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide–Revised (VRAG-R). *Psychological Assessment*, *25*, 951-965. doi:10.1037/a0032878

Rice, M. E., Harris, G. T., Lang, C., & Cormier, C. (2006). Violent sex offenses: How are they best measured from official records? *Law and Human Behavior*, *30*, 525-541. doi:10.1007/s10979-006-9022-3

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift fur Psychologie/Journal of Psychology*, *215*, 90-103. doi:10.1027/0044-3409.215.2.90

*Swinburne Romine, R., Dwyer, S. M., Mathiowetz, C., & Thomas, M. (2008, October). *Thirty years of sex offender specific treatment: A follow-up study*. Poster presented at the conference for the Association for the Treatment of Sexual Abusers, Atlanta, GA.

*Ternowski, D. R. (2004). Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program (UMI No. NR03201). *Dissertations Abstracts International*, *66*(06-B), 3428.

Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 139-153. doi:10.1023/A:1014620214905

Thornton, D. (2006). Age and sexual recidivism: A variable connection. *Sexual Abuse: A Journal of Research and Treatment*, *18*, 123-135. doi:10.1007/s11194-006-9007-2

Thornton, D., & Knight, R. A. (2013). Construction and validation of SRA-FV need assessment. *Sexual Abuse: A Journal of Research and Treatment*. Advance online publication. doi:10.1177/1079063213511120.

Vincent, G. M., Guy, L. S., Fusco, S. L., & Gershenson, B. G. (2012). Field reliability of the SAVRY with juvenile probation officers: Implications for training. *Law and Human Behavior*, *36*, 225-236. doi:10.1037/h0093974

*Wilson, R. J., Cortoni, F., & Vermani, M. (2007). *Circles of support and accountability: A national replication of outcome findings* (Report No. R-185). Ottawa, Ontario: Correctional Service of Canada.

*Wilson, R. J., Picheca, J. E., & Prinzo, M. (2007). Evaluating the effectiveness of professionally-facilitated volunteerism in the community-based management of high-risk sexual offenders: Part two—A comparison of recidivism rates. *The Howard Journal*, *46*, 327-337. doi:10.1111/j.1468-2311.2007.00480.x

Leslie-Maaike Helmus is a PhD student in forensic psychology at Carleton University in Ottawa, Ontario, Canada. Her research interests focus on the assessment and treatment of sexual offenders. She has received numerous grants and academic awards including the Association for the Treatment of Sexual Abusers Pre-Doctoral Research Grant.

David Thornton practiced as a forensic psychologist in the United Kingdom until 2001, developing assessment procedures and treatment programs for sexual and violent offenders. Since 2001, he has been based in Wisconsin's Department of Health Services, working first to develop treatment services for high-risk mentally disordered sexual offenders, and latterly as the director of research for Wisconsin's Sexually Violent Persons' program. He also works part-time as a professor in the department of clinical psychology at the University of Bergen in Norway.