

**Updated 5-Year and New 10-Year Sexual Recidivism Rate Norms for Static-99R with
Routine/Complete Samples**

Seung C. Lee^{1,2} and R. Karl Hanson¹

¹Department of Psychology, Carleton University

²Corrections Research, Public Safety Canada, Ottawa, Ontario, Canada

Law and Human Behavior (in press, 17-DEC-2020)

Author Note

Seung C. Lee  <https://orcid.org/0000-0002-6705-4535>

R. Karl Hanson  <https://orcid.org/0000-0002-0920-884X>

R. Karl Hanson is a co-author and certified Static-99R trainer. The copyright for Static-99R is held by the Government of Canada, and none of the authors receive royalties from this measure. The opinions, findings, conclusions, and recommendations expressed are those of the authors and do not necessarily reflect those of the institutions that supported this research.

The authors would like to thank Kelly Babchishin, Andrew E. Brankley, Yolanda Fernandez, Andrew J. R. Harris., L. Maaike Helmus, and David Thornton for their helpful feedback on an earlier draft of this manuscript. We would also like to thank the following researchers for permitting us to use their data: Darci L. Bartosh, Jacques Bigras, Sasha Boer, Cynthia Calkins, Jackie Craissati, Reinhard Eher, Doug Epperson, Leigh Harkins, Elizabeth

Jeglic, Niklas Långström, Robert Lehmann, Janet Neeley, and Mark Olver.

Correspondence concerning this article should be addressed to Seung C. Lee, Department of Psychology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6, Canada. Email: seungc.lee@carleton.ca

Abstract

Objectives: The purpose of this study was to develop new 10-year recidivism rate norms as well as to update 5-year norms for the Static-99R risk tool for routine/complete samples. We also presented the extrapolated sexual recidivism rates from this new 10-year norms for follow-up periods of 11 to 20 years. **Hypotheses:** We hypothesized that absolute recidivism base rates ($B0_2$; i.e., the intercept centered on the median score of 2) would vary; however, the relative predictive accuracy (i.e., discrimination; $B1$) would be stable across samples. In addition, compared to the estimated sexual recidivism rates with a fixed 5-year follow-up time, the estimated rates with a fixed 10-year follow-up time would be expected to be consistently higher across the Static-99R scores. **Method:** The current study included 12 independent samples ($N = 7,224$ for the 5-year recidivism rate norms and $N = 1,599$ [$k = 6$] for the 10-year norms) classified as routine/complete samples, that is, relatively random samples from a correctional system. Logistic regression parameters ($B0_2$ and $B1$) across the studies were aggregated using fixed-effect meta-analyses. **Results:** There was statistically significant variability in the base rates ($B0_2$), whereas the between-sample variability in the relative risk parameters ($B1$) was no more than would be expected by chance. As expected, the 10-year base rates were approximately 1.5 times higher than the 5-year base rates (7.20% vs. 4.58%), and the extrapolated 20-year sexual recidivism rates were approximately double the observed 5-year sexual recidivism rates. **Conclusions:** The current study provides empirical evidence to estimate 5- and 10-year sexual recidivism rates based on Static-99R total scores. Evaluators who are especially concerned about long term sexual recidivism risk (e.g., civil commitment) can report the expected sexual recidivism risk based on the new 10-year norms and the extrapolated sexual recidivism rates for follow-up periods of 11 to 20 years.

Keywords: sexual recidivism rates, Static-99R, norms, risk assessment

Public Significance Statement

Evaluators can be confident in the relative risk differences implied by Static-99R scores, regardless of the sample or follow-up period. In contrast, it is difficult to make strong statements concerning the absolute recidivism rates associated with Static-99R scores. Absolute risk interpretations of Static-99R scores need to consider the decision threshold as well as local base rate information, if available.

**Updated 5-Year and New 10-Year Sexual Recidivism Rate Norms for Static-99R
with Routine/Complete Samples**

In the last few decades, there has been substantial research indicating structured approaches (e.g., actuarial or structured professional judgment [SPJ]) provide more accurate assessments of re-offending risk than unstructured professional judgment (Ægisdóttir et al., 2007; Grove et al., 2000; Hanson & Morton-Bourgon, 2009). Actuarial approaches combine empirically identified risk factors (e.g., prior criminal history, male victims) into an overall evaluation of risk. Total scores or risk levels are then linked to a nominal risk category, such as “average,” and also to estimated recidivism probabilities (Dawes et al., 1989; Grove & Meehl, 1996).

Forensic practitioners routinely use empirical actuarial risk assessment tools to evaluate recidivism risk for individuals involved in the criminal justice and forensic mental health systems (e.g., sentencing procedure, treatment programs, community supervision programs; Hill & Demetriooff, 2019; Kelley et al., 2020; Neal & Grisso, 2014). Risk assessment tools are criterion-referenced, prognostic measures; that is, they are created to estimate the likelihood of a future event (Moons et al., 2009). Prognostic tasks (where the outcome does not exist at the time of assessment) contrast with diagnostic tasks (where the characteristic exists at the time of assessment). With criterion-referenced measures, the interpretation of scores is not fundamentally based on comparisons with the scores of others; instead, scores support inferences concerning the likelihood of the outcome among similar individuals (i.e., those with the same score).

Both norm- and criterion-referenced assessment instruments can support quantitative inferences about individuals. For norm-referenced measures, the major quantitative descriptors

concern percentile ranks or the degree to which the score is common or unusual in some reference populations (Crawford et al., 2009). The scores for criterion-referenced measures can also be described in terms of percentiles (Hanson et al., 2012); however, criterion-referenced measures can speak to other quantitative indicators, such as risk ratios (Hanson et al., 2013) and absolute likelihoods of the outcome (Hanson et al., 2016). In this paper, we refer to the quantitative information asserted by test developers as *norms*, regardless of whether a measure is primarily norm-referenced or criterion-referenced. Specifically, the focus of this paper is on sexual recidivism rate norms for the Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012), a criterion-reference prediction tool.

Static-99R is the actuarial risk assessment tool most widely used to assess the risk of sexual recidivism among adult males who have a history of sexual crimes (Bourgon et al., 2018; Kelley et al., 2020). It contains 10 static items (e.g., age, prior sex offenses), and the total score (ranging from -3 to 12) is calculated by summing all items. The total scores are routinely used two ways in risk communication: one related to an individual's relative risk (i.e., how likely the individual is to re-offend compared to other individuals; a norm-referenced interpretation), and the other related to absolute risk (i.e., the proportion of individuals with this score expected to re-offend within a defined time period; a criterion-referenced interpretation).

Norms for Static-99R are intended to guide relative comparisons (i.e., percentiles and risk ratios; Hanson et al., 2012, 2013) and to link specific scores to expected sexual recidivism rates (Hanson et al., 2016; Phenix, Helmus, et al., 2016). The most recent recidivism rate norms include 21 independent samples ($N = 8,805$) for two distinct groups (routine/complete samples and high risk/high need samples), and for two different follow-up periods (5 years for both normative groups and 10 years for the high risk/need group only; see details in Hanson et al.,

2016). For a comprehensive communication of the results, both norm-referenced (i.e., relative) and criterion-referenced (i.e., absolute risk) approaches should be considered.

At the time of updating the Static-99R norms in 2016, there were, however, insufficient data to produce reliable 10-year recidivism estimates for routine/complete samples (Hanson et al., 2016). The current study fills that gap, and we took the opportunity to update the 5-year recidivism norms for routine/complete samples using newly available data (Lee et al., 2016; Mercado et al., 2011). Given the criterion-referenced interpretations of Static-99R scores, considerable efforts should be made to obtain empirical evidence concerning the recidivism rates associated with test scores. In addition, the norms should be updated with sufficient frequency to permit continued accurate and appropriate score interpretations (Standard 5.5 and 5.11; Joint Committee, 2014).

Risk assessment tools, including Static-99R, are often used to inform high-stakes decisions concerning long-term detention, such as civil commitment and Sexual Violent Predator (SVP; U.S.) and dangerous offender (D.O.) designation (Canada; Blais & Forth, 2014; Jackson & Hess, 2007; Neal & Grisso, 2014). In some of these legal decisions, the threshold concerns lifetime risk (e.g., Will this individual ever re-offend?). In contrast, the recidivism rate norms for many measures, including Static-99R, are presented for shorter time frames (e.g., 3 years, 5 years, 10 years).

In order to address questions concerning lifetime risk, Thornton and his colleagues (2019) provided an empirical model to extrapolate observed sexual recidivism rates from fixed follow-up times (e.g., 5-year, 10-year) to follow-up times between 1 and 20 years by accounting for the offense-free time in the community (see Model 5 from the Hanson et al. [2018] study). Projections stopped at 20 years because there were almost no new recidivism events after that

time. In Thornton et al.'s calculations, each year offense-free was associated with an approximately 12% decrease in the expectation that individuals would re-offend sexually in each subsequent year. This means that each 5-year period in which the individual was in the community and sexual offense free was associated with approximately a 50% decrease in sexual recidivism. Using these heuristics, Thornton et al. (2019) extrapolated sexual recidivism rates ranging from 6 to 20 years based on the 2016 5-year norms of Static-99R (Hanson et al., 2016). Although 10-year estimates are available through these extrapolations, it is helpful to directly estimate 10-year norms as a check on the credibility of the extrapolation procedure, and 10-year observed rates provide a closer starting point for extrapolating rates for 11 to 20 years (i.e., less extrapolation than starting with observed 5-year rates).

There is a long-standing controversy over what standards should be taken into consideration in deciding whether scientific evidence presented by experts is admitted or not in court. The 1923 *Frye* standard requires only that the method of forming the expert opinion has “*general acceptance*” in the field to judge the admissibility of scientific information (*Frye v. United States*, 1923). Under the *Frye* standard, it is possible for unreliable evidence to be admitted when the method had achieved general acceptance, even if there is no evidence (and no professional consensus) to support the specific conclusion (e.g., unstructured clinical judgment of dangerousness).

In 1993, the U.S. Supreme Court outlined non-exclusive criteria of expert testimony admissibility for federal courts (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*). The *Daubert* standard is more conservative than the *Frye* standard, including 1) whether the scientific theory and technique underlying the testimony have been tested, 2) whether the information has been peer-reviewed and published, 3) whether an error rate associated with the scientific theory and

technique has been established, and 4) whether the scientific theory and technique have been generally accepted in the relevant scientific field. Given that all federal courts and most state courts have adopted *Daubert* standard, publishing updated recidivism rate norms and error rates for empirically-derived risk tools is useful to experts wishing to use such measures in court.

Updating norms is not without risks to users of risk tools. Changes in norms could call into question previous legal decisions that were informed by previous norms existing at that time. It is not hard to imagine an individual who was civilly committed based on a risk scale score that had expected recidivism rates just above the commitment threshold. Now imagine that the revised norms suggest lower recidivism rates such that the case no longer meets the commitment threshold. Should the original commitment decision be reconsidered even reversed it? Could the committed individual raise the issue of liability for wrongful detention? Or should the commitment decision stand if it was based on the best available information *at that time*, even if that information is now known to be incorrect? What strength of evidence is required to revise and update norms?

To develop credible norms for criterion-referenced risk assessment tools, some standards should be considered. First, there should be well-established evidence of adequate discriminative accuracy of the targeted risk assessment tools. This evidence must include samples other than the sample used to develop the risk tool (Collins et al., 2016). Second, the norming samples should include representative individuals with whom evaluators will ordinarily wish to compare their own examinees (e.g., adult males who have a history of sexual crimes). Third, the recidivism criteria should be well-defined for accurate interpretations of the test scores, such as arrested, charged with, or convicted of sexual crimes after release to the community. Fourth, there should be sufficient sample sizes (e.g., a minimum of 100 sexual recidivists; Vergouwe et al., 2005)

when developing a prognostic measure or evaluating its performance.

Purpose of the Current Study

This study provides expected long-term sexual recidivism rates (i.e., 10-year follow-up) as well as update 5-year sexual recidivism norms (two additional samples) for the Static-99R risk tool for routine/complete samples. Specifically, we examined the stability of the absolute recidivism base rates (B_0 from logistic regression) and relative predictive accuracy (B_1 from logistic regression) of Static-99R across 12 studies for 5-year norms and six studies for 10-year norms. Further, we calculated the extrapolated sexual recidivism rates (i.e., 11 to 20 years) using Thornton et al.'s (2019) procedures based on the new 10-year norms.

Hypotheses

Based on the previous research findings (Hanson et al., 2016; Helmus, Hanson, et al., 2012), we expect that absolute recidivism base rates (B_0 ; i.e., the intercept, centered on the median score of 2) would vary across samples from different systemic backgrounds (e.g., jurisdictions, recidivism criteria; [H1]). Variability in recidivism base rates complicates inferences concerning the expected recidivism rates associated with specific scores. The relative predictive accuracy (i.e., discrimination; B_1), however, would be stable across samples [H2]. In addition, compared to the estimated sexual recidivism rates with a fixed 5-year follow-up time, the estimated sexual recidivism rates with a fixed 10-year follow-up time would be expected to be approximately 1.5 times higher (Thornton et al., 2019; [H3]).

Method

Samples

The current study included 12 independent samples ($N = 7,244$) classified as routine/complete samples (Table 1), that is, relatively random samples from a correctional

system, representing a broad spectrum of all eligible individuals (e.g., prisoners, probationers, or parolees). The routine/complete samples is, thus, a distinct group from those among the high risk/high need samples, which consists of preselected offenders based on a perceived high level of risk and/or need (e.g., individuals serving the indefinite sentence or under high-intensity treatment programs).

Specifically, 10 samples overlapped with the previous 5-year routine/complete norms ($N = 4,223$; Hanson et al., 2016), one of which was updated recidivism information (i.e., substituting Hanson et al. [2014] study with Lee et al. [2018]). There were, in addition, two new samples ($N = 3,021$; Lee et al., 2016; Mercado et al., 2011; Table 1; Appendix B contains more descriptive information of the new samples).

[Insert Table 1 about here]

All samples were adult males (18 years old or older) who had been convicted of a Static-99R eligible sexual crime (Category A, see Phenix, Fernandez, et al., 2016). In addition, to be included in the current study, the information on sexual recidivism for a 5-year and/or a 10-year follow-up period and Static-99R total scores was required. The routine/complete samples were from the United States ($k = 5$), Canada ($k = 3$), Sweden ($k = 1$), U.K. ($k = 1$), Austria ($k = 1$), and Germany ($k = 1$). Individuals were released between 1976 and 2010 (Table 1).

Measures

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

Static-99R is a 10-item empirical actuarial risk tool designed to assess the risk of sexual recidivism among adult males with a history of sexual offending. Static-99R is identical to Static-99 except for revised age weights. The total score (ranging from -3 to 12) is calculated by summing all item points and places individuals in one of five risk levels: Level I - very low risk

(scores of -3 to -2), Level II - below average risk (scores of -1 to 0), Level III - average risk (scores of 1 to 3), Level IVa - above average risk (scores of 4 to 5), and Level IVb - well above average risk (scores of 6 or higher; Hanson, Babchishin, et al., 2017). When necessary, Static-99R scores were computed from Static-99 scores by using the individual's date of birth to calculate the updated age item. The Static-99R total score was found to have good interrater reliability in previous studies (e.g., ICC = .78, $n = 14$ cases; Hanson, Lunetta, et al., 2014; ICC = .89, $n = 30$ cases; Quesada et al., 2014). Considerable research has found overall moderate predictive accuracy (i.e., discrimination; AUC = .70, $N = 8,106$, $k = 23$; Helmus, Hanson et al., 2012).

Sexual Recidivism

Sexual recidivism was defined as any subsequent contact or non-contact sexual offenses that were classified based on the name of charge/conviction (e.g., sexual assault) or sexual motivation (regardless of what the charge or conviction was) after release. Recidivism criteria, however, varied across the samples: arrest ($k = 2$, $n = 1,075$ [15%]), charge ($k = 4$, $n = 1,997$ [27%]), and conviction ($k = 6$, $n = 4,172$ [58%]) for 5-year normative samples; Table 1). For the 10-year samples, about three quarters of the recidivism criteria were conviction ($k = 4$, $n = 1,225$), followed by charge ($k = 1$, $n = 338$ [21%]) and arrest ($k = 1$, $n = 36$ [2%]).

Results

Overview of Analyses

To develop the norms of Static-99R, we conducted fixed-effect meta-analyses of means, AUC values, and logistic regression coefficients across the samples. The analyses (e.g., means, AUC, and logistic regression) were conducted using version 24 of SPSS, with the exception of the meta-analyses, which were conducted with a package "*metafor*" (Version 2.4-0; Viechtbauer,

2010) for the statistical software R (Version 4.0.0; R Core Team, 2013).

Meta-Analytic Averages

Means of each sample (Static-99R total scores, age, and AUCs) were weighted by the inverse of the variance to calculate the fixed-effect meta-analytic averages (Borenstein et al., 2009).

Area Under the Curve (AUC)

Discrimination predictive accuracy of Static-99R for each sample was evaluated by using AUC analysis (Swets et al., 2000). The AUC values indicate the probability that a randomly selected recidivist would have a higher score on the Static-99R than a randomly selected non-recidivist. The AUC can vary between 0 and 1, with .50 indicating a chance level of prediction. A value of AUC between .50 and 1.00 indicates above chance-level prediction with statistical significance as long as the 95% confidence interval does not include .50. As a rough heuristic, an AUC of .56 corresponds to a small effect size, while .64 reflects a moderate effect, and .71 reflects a large effect size (Rice & Harris, 2005). The AUC has the advantage of insensitivity to base rates and robustness to outliers (Ruscio, 2008).

Logistic Regression Analysis

Logistic regression analysis was conducted to provide two coefficients (B0 – an intercept, and B1 – a slope) with one independent variable (Static-99R total scores) and one binary dependent variable (sexual recidivists vs. non-recidivists). We used B0 coefficients centered on Static-99R scores of 2 (B_{0_2}), which represents the median value in the population of adjudicated individuals with sexual crimes (Hanson et al., 2012). The intercept (B_{0_2}), therefore, estimates the adjusted sexual recidivism rates (as a logit) for individuals with a Static-99R score of 2 (i.e., individuals in the middle of the risk distribution). For ease of interpretation, the B_{0_2}

coefficients and their confidence intervals (CIs) were transformed from logits back into probabilities (p ; see Appendix A from Hanson et al., 2016).

Logistic regression was also used to calculate odds ratios (i.e., the extent to which the recidivism rates vary as a function of Static-99R total scores). Odds ratios (the exponent of a slope coefficient, [exp^{B1}]) indicate the change in relative risk associated with a one-unit change in Static-99R scores. For example, in routine/complete samples, Static-99R scores are associated with a consistent relative risk increase of approximately 1.45 (Hanson et al., 2016), which means the odds of recidivism increase 1.45 times with a one-point increase in the Static-99R score. An odds ratio is statistically significant if the 95% confidence interval does not include 1.

Aggregation of Findings

Logistic regression parameters ($B0_2$ and $B1$) across the studies were aggregated using fixed-effect meta-analyses (Borenstein et al., 2009; Hanson & Broom, 2005). Although the random-effects meta-analyses are preferred conceptually because they aim to permit generalization to a population of studies, the fixed-effect meta-analyses were used in this study due to the small number of studies ($k = 12$ for 5-year norms and $k = 6$ for 10-year norms). The random-effects meta-analyses require a much larger sample of studies (30+; Schulze, 2007) to calculate a stable estimate of a between-study variability term (tau-square, τ^2).

Fixed-effect analyses have the advantage of providing an estimate of between-study variability (Cochran's Q statistic) that can be used to compare the variability across the studies (Borenstein et al., 2009). A significant Q statistic (i.e., p value $< .05$) indicates that there is more variability across studies than would be expected by chance. Q statistics are distributed as a Chi-square, with $k-1$ degree of freedom ($k =$ the number of studies). If a single study accounts for more than 50% of the total variance (Q), the case is considered as an outlier (Hanson & Morton-

Bourgon, 2009).

In addition, the I^2 statistic was used to measure the magnitude of the variability (i.e., an effect size index), which describes the ratio of observed heterogeneity to total variance across the observed effect estimates.

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100$$

As rough benchmarks, the I^2 values of 25%, 50%, and 75% can be considered low, moderate, and high variability, respectively (Higgins et al., 2003).

Moderator Analyses

Given that we found significant variability in the base rates ($B0_2$), we conducted some post hoc comparisons to explore the potential sources of variability (e.g., country, recidivism criteria). These were conducted by a $Q_{between}$ -test based on analysis of variance from fixed-effect analyses (Borenstein et al., 2009). First, we calculated the sum of within-group weighted variations for p subgroups (Q_{within}).

$$Q_{within} = \sum_{j=1}^p Q_j$$

And, $Q_{between}$ were computed as

$$Q_{between} = Q_{total} - Q_{within}$$

The $Q_{between}$ statistics are distributed as a Chi-square, with a $p-1$ degrees of freedom (p = the number of subgroups).

Preselection Effects

Preselection effects were also examined to explore group differences within routine/complete samples. Previous research has found a positive correlation between the adjusted base rates ($B0_2$) and the average Static-99R for the sample (Hanson et al., 2016),

indicating that samples that had been preselected on risk-relevant variables had higher recidivism base rates even after adjusting for Static-99R scores. Although none of the current samples were explicitly preselected based on risk, there was, nonetheless, variability in average Static-99R scores, suggesting some degree of risk-related preselection had occurred. If there is a significant association between the average Static-99R scores and the adjusted base rates (B_0) of each sample, it would provide evidence of preselection effects based on risk-relevant factors external to Static-99R. These were examined using fixed-effect and random-effects meta-regression analyses.

Computation of Expected Rates and Extrapolations to 20-Year Recidivism Rates

The logistic regression coefficients ($B_0 + B_1 \cdot \text{Static-99R score}$; as logits) were used to compute the expected 5-year and 10-year sexual recidivism rates displayed in Figure 3. In addition, the 95% confidence intervals of the expected recidivism rates were computed based on the standard errors of the logits ($S.E._{\text{logit}}$):

$$95\% \text{ CI} = \text{logit} \pm (1.96 * [S.E._{\text{logit}}])$$

The procedures and formulas to calculate recidivism rate estimates and their standard errors are provided in Appendix A of Hanson et al. (2016).

Once the expected rates of the fixed 5-year and 10-year follow-up times were estimated by fixed-effect meta-analyses, the expected rates for subsequent years were calculated by subtracting .130 from the logit of the expected rate (hazard) for the previous year: $\text{logit}(h_{t+1}) = \text{logit}(h_t) - .130$ (Model 5 from Hanson et al., 2018). The logit can be transformed into proportions, $p = 1/(1 + e^{-\text{logit}(h_t)})$. Consequently, the cumulative survival rates are calculated by multiplying the yearly survival rates ($1 - \text{hazard rate}$) for each year at risk based on the estimated yearly hazard rates implied by the observed rates at 5 years and 10 years. In this study, the

cumulative survival rates, S_T , for follow-up times up to 20 years without committing any sexual crimes (i.e., offense-free) were calculated as follows (Formula 2 from Thornton et al., 2019):

$$S_T = \prod_{\text{year}=1}^t \left(1 - \frac{1}{1 + e^{-\text{logit}(h_t)}} \right)$$

Overall calculations in this study were computed by an excel spreadsheet that has been developed for evaluators available on the static99.org website.

Sample Characteristics

The weighted average of age at release was 40 years old ($SD = 4$, ranging from 18 to 86), and the weighted average of Static-99R total scores was 2.4 ($SD = 0.7$; Table 2). The observed sexual recidivism rates were 6.7% for a 5-year follow-up period ($k = 12$) and 11.6 % for a 10-year follow-up period ($k = 6$).

Compared to the 5-year routine/complete samples ($k = 12$), the 10-year normative samples ($k = 6$) showed very similar Static-99R average scores (2.4 vs. 2.5) and the average age at release (40 years old for both normative samples). When the AUC values from each study were meta-analyzed, the Static-99R total scores significantly predicted sexual recidivism with a fixed 5-year follow-up period, $AUC = .76$, 95% CI [.73, .78], $N = 7,244$, $k = 12$, and also a fixed 10-year follow-up period, $AUC = .73$, 95% CI [.69, .77], $N = 1,599$, $k = 6$ (Table 2).

[Insert Table 2 about here]

Absolute Recidivism Base Rates (B0₂)

As hypothesized [H1], a meta-analysis of 12 studies for a fixed 5-year follow-up period found statistically significant and high variability in the base rate associated with the median Static-99R score (i.e., a score of 2), $Q = 58.1$, $p < .001$, $I^2 = 81.1$ (see Table 3; Figure 1). No outlier studies were identified as no single study captured 50% or more of the variance (Hanson & Bussière, 1998).

[Insert Table 3 and Figure 1 about here]

Similarly, in the six Static-99R studies for a fixed 10-year follow-up period, there was also statistically significant and moderate variability in the base rate associated with the median Static-99R score (i.e., a score of 2), $Q = 14.2$, $p = .02$, $I^2 = 64.7$ (see Table 4; Figure 2). One study (Lehmann et al., 2013) appeared to be a statistical outlier in the B0₂ analyses. It was, nonetheless, retained in the 10-year estimates because a) the number of available studies was small ($k = 6$); b) it was not a statistical outlier in the 5-year analyses, and c) careful examination of the research method did not find any unusual features that would limit the generalizability of the findings.

[Insert Table 4 and Figure 2 about here]

As expected [H3], the 10-year base rates (B0₂; a median score of 2) were 1.57 times higher than the 5-year base rates (B0₂; a median score of 2) in the routine/complete samples (7.20% vs. 4.58%; Table 3; Table 4). The estimated 5-year base rates, 4.58%, 95% CI [4.04, 5.19], $k = 12$ (Table 3) were very comparable to the estimated 5-year base rates from the 10-year normative samples, 4.02%, 95% CI [3.39, 4.76], $k = 6$ (Table S1).

Potential Sources of Variability and Preselection Effects

There was a systemic variation of sexual recidivism rates across countries (Canada vs. U.S. vs. European countries), $Q_{between} = 21.7$, $df = 2$, $p < .001$ (Table 5). The estimated 5-year base recidivism rate for a Static-99R score of 2 was highest in the samples from Canada (7.2%) followed by Europe (4.9%) and U.S. (3.3%). We also found the recidivism criteria (arrest/charge vs. conviction) was significantly associated with sexual recidivism rates, $Q_{between} = 4.4$, $df = 1$, $p = .04$ (Table 5). The estimated 5-year base rate was 5.3% in samples using arrests/charges as their recidivism, and 4.0% in samples using convictions.

[Insert Table 5 about here]

In addition, peer-reviewed studies showed significantly higher estimated 5-year base rates (6.1%) than other types of reports, such as government reports, unpublished papers, and thesis papers (3.1%), $Q_{between} = 27.4$, $df = 1$, $p < .001$ (Table 5). There were, however, no significant differences in the base rates between the samples whose median release dates were before and after the year 2000 (5.0% with $k = 6$ and 4.4% with $k = 6$), $Q_{between} = 1.2$, $df = 1$, $p = .28$ (Table 5). There were also no significant differences in the base rates between the studies in which the test developer(s) of the Static-99R tool were involved or not (5.0% with $k = 4$ and 4.2% with $k = 8$), $Q_{between} = 2.1$, $df = 1$, $p = .15$. Although these moderators explained some of the variability, there was still significant variability across studies within each level of moderators (Table 5).

With the fixed-effect model, the average Static-99R scores of the samples were significantly associated with the estimated 5-year base rates ($B0_2$), $b = 0.46$, $SE = 0.15$, $p = 0.002$, $Q = 10.1$. The correlation was the same magnitude, but no longer statistically significant in the random-effects model, $b = 0.48$, $SE = 0.39$, $p = .22$, $Q = 1.5$.

Relative Predictive Accuracy (B1)

Consistent with our hypothesis [H2], the between-sample variability in the relative risk parameters was not statistically significant for both fixed 5-year, $Q = 13.7$, $p = .25$ (Table 3; Figure 1) and fixed 10-year follow-up times, $Q = 5.91$, $p = .32$ (Table 4; Figure 2).

Discrimination (B1) of Static-99R for the routine/complete samples was almost identical between with a fixed 5-year and with a fixed 10-year follow-up times. Specifically, the odds ratios associated with a one-point increase in the Static-99R score were 1.45 ($exp^{[0.372803]}$) from Table 3) for a fixed 5-year follow-up time and 1.44 for a fixed 10-year follow-up time (exp

[0.364911] from Table 4).

Among the 12 studies with 5-year recidivism data, discrimination was higher in the four studies in which a Static-99R author (Hanson) was involved ($B1 = 0.41$) than the other eight studies ($B1 = 0.33$), $Q_{between} = 4.0$, $df = 1$, $p = .046$. Among the six studies with 10-year recidivism rates, discrimination was lower in the two studies in which a Static-99R author was involved ($B1 = 0.33$) than the other four studies ($B1 = .41$), although, in this case, the difference was not statistically significant, $Q_{between} = 1.1$, $df = 1$, $p = .29$.

As hypothesized [H3], the overall sexual recidivism rates with a fixed 10-year follow-up time were higher than those with a fixed 5-year follow-up time (11.6% vs. 6.7%; Table 2; Figure 3). Specifically, the estimated sexual recidivism rates with a fixed 10-year follow-up time were approximately 1.5 times higher, particularly in the Level I, II, III, and IVa, when compared to the estimated sexual recidivism rates with a fixed 5-year follow-up time (see Appendix Table A1 and Table A2). In the well above average risk group (scores of 6 or higher; Level IVb), the 10-year rates were slightly less than 1.5 times the 5-year rates.

[Insert Figure 3 about here]

Extrapolated Sexual Recidivism Rates Across 20-Year Follow-Up Period

The extrapolated 10-year sexual recidivism rates from the 5-year norms (Table S2) were very consistent with the new 10-year norms (Table S3). Based on the extrapolated sexual recidivism rates, the 20-year rates were approximately double the 5-year sexual recidivism rates for Level I, II, III, and IVa; however, the 20-year rates in the well above average risk (i.e., scores of 6 or higher; Level IVb) were less than double the 5-year rates (Table S3).

Table S4 presents the residual risk for individuals who have remained free of sexual crimes while in the community after release. This projected residual risk was calculated based on

the estimated 5-year sexual recidivism rates for routine/complete samples (Table 1). As individuals remain offense free, their risk of a new sexual crime consistently decreases. For example, for individuals whose initial risk was above a high-risk threshold of 35% or more, all individuals drop below this high-risk threshold by 8 years. If they stay offense free long enough (approximately 20 years), all individuals will eventually drop below the very low-risk threshold (<3%), regardless of their initial risk score (lifetime high-risk and low-risk thresholds from Thornton et al., 2019).

Discussion

Accurate actuarial recidivism risk assessment tools require regular updates. The current study updated the 5-year recidivism rate norms (from Hanson et al., 2016) and developed new 10-year norms for the Static-99R for the routine/complete samples. In comparison to previous norms (Hanson et al., 2016), the updated 5-year norms were adjusted slightly downward for Static-99R total scores. For example, the 2016 norms estimated a 5-year rate of 5.6% for a score of 2, whereas the current update estimated a rate of 4.6%. For a score of 6 (the entry score for the well above average risk group), the decline was from 20.5% (the old estimate) to 17.6% (the updated estimate). These changes are unlikely to meaningfully influence decisions concerning whether individuals meet specific risk thresholds; the influence of risk factors external to Static-99R is much greater than the effect of updated norms. For example, the estimated rate for preselected high-risk samples is 11.3% for a score of 2 at 5 years (2016 norms) compared to rates for routine/complete samples of 5.6% (2016 norms) or 4.6% (current best estimates).

As hypothesized, the estimated sexual recidivism rates for a fixed 10-year follow-up period were approximately 1.5 times higher than the updated 5-year norms of the routine/complete samples (adjusted base rates [B0₂] were 4.58% and 7.20%, respectively, a risk

ratio of 1.57). This finding supports the offense-free effect model in which the risk of sexual recidivism declines consistently and predictably when individuals remain sex offense-free in the community after release (Hanson, Harris, et al., 2014, 2018).

Previous research found that some of the variation in sexual recidivism rates across samples can be attributed to systematic differences between the routine/complete samples and high risk/high need samples; consequently, the STATIC Development Team has recommended separate norms of Static-99R for each group (Hanson et al., 2016; Phenix, Helmus, et al., 2016). Even within the routine/complete samples, previous research (Hanson et al., 2016), like the current study, found significant between-study variability in 5-year absolute base rates (B0₂). Significant between-study variability was also found in the 10-year rates in the routine/complete studies in the current meta-analysis. The reasons for this variability remain an active topic of research.

The variability of sexual recidivism rates could be explained by systematic differences in the criminal justice systems across jurisdictions. For example, sexually deviant behavior is tolerated less in certain jurisdictions, thereby more likely to be detected and prosecuted. Some jurisdictions are more likely to plea-bargain a sexual offense to a lesser offense (e.g., a violent offense). Discrepancies in recidivism information used in different jurisdictions (e.g., recidivism criteria or quality of recidivism information) may have contributed to the variation of recidivism rates (Helmus, 2009). The comparatively low rates in the U.S. samples, for example, may be related to the relative difficulty in obtaining national criminal history records in the U.S. (criminal history records are organized at the state level) compared to Canada (which has a well-established system of national criminal history records accessible to researchers).

Although we examined and identified some plausible moderators (e.g., country,

recidivism criteria, and peer-reviewed studies vs. other types), there was still significant variability across studies within each subgroup of moderators. Further research is needed to explore the substantial variability left unaccounted for. Even though none of the current samples were explicitly preselected based on risk, there might be systematic differences in unmeasured risk factors external to Static-99R.

Hanson et al. (2016) found samples preselected as high-risk/high need had significantly higher Static-99R scores and higher sexual recidivism rates than routine/complete samples, even after controlling for the higher Static-99R scores. The current study demonstrated a similar pattern, despite a restriction of range in mean Static-99R scores. In Hanson et al. (2016), mean Static-99R scores for the samples ranged from 1.6 (Level III, “average”) to 5.1 (between Level IVa, “above average” and Level IVb, “well above average”), whereas in the current study the mean scores only varied from 1.9 (Level III, “average”) and 3.4 (between Level III, “average” and Level IVa, “above average”). Nevertheless, we still found the mean Static-99R scores of the samples ($k = 12$ for the 5-year normative samples) were significantly and positively associated with the adjusted 5-year base rates for a score of 2, using fixed-effect meta-regression. The correlation was of the same magnitude but not statistically significant using random-effects meta-regression, which limits the confidence in the findings. In general, fixed-effect meta-regression analyses tend to detect effects that are not there (Type I error), whereas random-effects analyses fail to detect effects that are there (Type II error; Overton, 1998).

It is possible to empirically combine Static-99R scores with other risk assessment tools to adjust the overall risk levels. For example, there are empirically informed combination rules for Static-99R with STABLE-2007 and ACUTE-2007 (Brankley et al., 2017, 2019), as well as for the Violence Risk Scale – Sexual Offender version (VRS-SO; Olver et al., 2018). In the

combination rules, the risk levels assessed by Static-99R are mechanically adjusted based on the scores of the other structured risk assessment tools that include factors not directly assessed by Static-99R (e.g., attitudes toward women, progress in treatment). There is, however, little research on how the empirically combined measures reduce the variability of sexual recidivism rates across samples.

The persistence of between-study variability in base rates presents a challenge to evaluators who wish to interpret Static-99R scores in terms of absolute recidivism rates. Variability, even if large, may not matter if the full range of expected values falls on the same side of a decision threshold. For example, if the decision threshold for special measures was 35% after 5 years, evaluators could be confident that individuals with Static-99R scores of 2 would be below that threshold. Although the observed range is large (0.30% to 11.37%; see Table 3), the highest value is still substantially below the decision threshold of 35%. Variability in estimates is a greater challenge when the range of estimates spans a decision-threshold.

Current user guidance (Phenix, Fernandez, et al., 2016) presents recidivism rates for two broad reference groups: routine/complete and high risk/high need. Consistent with previous work, individuals in the routine/complete samples in the current study displayed lower sexual recidivism than individuals in the high risk/high need samples (from Hanson et al., 2016; see Figure S1). There is, however, no empirically validated decision rules for selecting between the routine/complete samples and the high risk/high need reference groups.

Consistent with previous research, the between-study variability for Static-99R slope parameters (B1; relative risk) was not significant in either the 5-year and 10-year routine/complete samples. There was some evidence that the B1 coefficients were higher in studies in which one of the test developers was involved (Hanson) than in the other studies. This

pattern, however, was only found in the 12 studies with 5-year recidivism data and not found (reversed) in the six studies with 10-year data. Consequently, further research is needed before making strong statements about the extent to which author involvement is associated with discrimination accuracy for Static-99R. Overall, evaluators can be reasonably confident in the relative risk differences implied by Static-99R scores, regardless of the sample or follow-up period. The magnitudes of relative risk were very similar in the current study to what has previously been reported (e.g., an odds ratio of about 1.4; Hanson et al., 2013; Hanson et al., 2016; Helmus, Hanson, et al., 2012). The stability of the relative risk estimates demonstrates that the risk factor items of Static-99R can discriminate between recidivists and non-recidivists for diverse samples of individuals who committed sexual crimes (e.g., probationers, prisoners, forensic psychiatric patients) across follow-up periods.

We also found consistency between the new 10-year norms generated by this study (Table S3) and previously extrapolated 10-year rates (Thornton et al., 2019). Our current finding reinforces the stability of the observed reductions in recidivism risk for individuals who remain sexual offense-free in the community (Hanson et al., 2018). There were no meaningful discrepancies between the extrapolated rates of 11 to 20 years from the 5-year and 10-year norms (Table S2 and Table S3). Although evaluators might prefer one to another (e.g., Table S2 is based on more samples or Table S3, which requires less extrapolation), it does not make substantial changes to interpretation.

Limitations

Although the overall sample size was large enough (i.e., more than 100 recidivists) for stable logistic regression estimates (Vergouwe et al., 2005), there was an insufficient number of studies (30+; Schulze, 2007) to obtain stable estimates of between-study variability necessary for

random-effects analyses. As a result, the final recidivism estimates were based on fixed-effect analyses, which is conceptually limited in its generalizability (i.e., it describes only the studies included in the analyses). This could limit the generalizability of the estimation to this collection of individuals. When the between-study variability is small, however, the differences between fixed-effect and random-effects results are very small. This limitation is more related to the base rate parameters ($B0_2$) than the relative risk parameters ($B1$).

When computing the 10-year norms, one study (Lehmann et al., 2013) was identified as an outlier, which accounted for more than 50% of the Q statistic. Nevertheless, we included the study in the 10-year norms because we were unable to identify any systematic flaw or difference that would render it unreliable. In addition, the Lehmann et al. (2013) sample was not an outlier within the samples for 5-year norms. Given the small number of samples in the 10-year routine/complete norms ($k = 6$), and the significant variability, it is likely that some changes in the norms will be observed with the next update to the 10-year norms.

A further limitation is that logistic regression estimates may be unstable when extrapolating to very high Static-99R scores where there are reduced numbers of individuals at each score level; consequently, recidivism estimates were not reported for scores where there were less than 10 cases available. For example, there are no estimates provided for Static-99R scores of 11 to 12 for fixed 5-year norms and no estimates for scores of 10 to 12 for fixed 10-year norms.

Implications for Research

To evaluate the predictive validity of Static-99R, more research is required to test how well the estimated recidivism probability, particularly from the new 10-year norms, corresponds with the observed recidivism probability of new samples (i.e., calibration). There are some

promising statistical methods to evaluate the calibration: 1) *E/O* index (the ratio of the expected number of recidivists divided by an observed number of recidivists; Hanson, 2017) and 2) comparing logistic regression parameters (B0) by fixed-effect meta-analysis (Borenstein et al., 2009; Hanson & Broom, 2005).

Further, it is necessary to evaluate if Static-99R systematically overestimates or underestimates sexual recidivism risk for particular racial/ethnic groups (i.e., cultural bias in predictive validity) given the relatively small numbers of racial/ethnic minorities in development and normative samples (i.e., Asian, Latino, Black, Indigenous peoples). Research, however, should be cautious when interpreting unexpectedly low sexual recidivism rates for certain immigrant groups (i.e., poor calibration) because of high rates of deportation (e.g., individuals of Mexican descent in the U.S.). For example, after excluding the deportees from the calibration analysis, the sexual recidivism rates for individuals of Latino descent became comparable with the sexual recidivism rates for the norms of Static-99R (Lee et al., 2018).

Implications for Practice

The current study provides empirical evidence to estimate 5- and 10-year sexual recidivism rates by Static-99R score for individuals who have committed sexual crimes. Although the current norms are slightly lower than the previous 5-year norms (Hanson et al., 2016), it is unlikely that these changes would meaningfully influence decisions concerning whether individuals meet specific risk thresholds (e.g., SVP evaluation, parole decision). The variance due to external (largely unmeasured) risk factors is much greater than the variance between the 2016 norms and those developed in this paper. Nevertheless, updating the norms of risk assessment tools with new samples provides some assurance that risk estimation remains relevant to current practice. As well, it is easier for evaluators and decision-makers to adjust to small, incremental changes than to be faced with potentially substantial changes from updates

based on very different samples spaced many years apart.

Evaluators who are especially concerned about a long term recidivism risk (e.g., civil commitment) can report the expected sexual recidivism risk based on the new 10-year norms. Further, this study also provides estimated sexual recidivism rates for a follow-up period of 11 to 20 years based on the updated 5-year norms and the new 10-year norms. That information would be critical for decision-makers in the criminal justice system when the decision thresholds are related to lifetime risk (e.g., civil commitment, dangerous offender (D.O.) designation).

The general recommendation is to report the sexual recidivism rates for routine/complete samples (Appendix A). This is the most representative population of individuals who have been convicted of sexually motivated crimes. Specifically, we recommend against using the sexual recidivism rates for high risk/ high need samples unless there is strong justification (see Hanson et al., 2016). Note, however, that the significant differences in sexual recidivism rates between routine/complete samples and high risk/high need samples are only in the lower and middle range of Static-99R scores (i.e., the scores between -1 to 6). The differences in risk decrease as the Static-99R scores increase, such that preselection effects make relatively little difference to the recidivism risk estimate when the Static-99R score is 6 or higher (Figure S1).

Conclusion

Risk assessment tools are commonly used by forensic practitioners to inform their estimates of sexual recidivism risk. As criterion-referenced measures, absolute risk levels of actuarial test scores are central to score interpretations and risk communications. Consequently, continued efforts should be made to obtain empirical evidence concerning the soundness of such interpretations. Periodic updates assist evaluators and criminal justice systems to adjust through incremental changes and provide some assurance that the norms continue to be relevant to

current cases.

References

References indicated with * are data contributors

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. C., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341-382. <https://doi.org/10.1177/0011000005285875>
- *Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*(4), 422-438. <https://doi.org/10.1177/0306624X03253850>
- *Bigras, J. (2007). La prediction de la recidive chez les delinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertations Abstracts International, 68*(09). (UMI No. NR30941)
- Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior, 38*(6), 531-543. <https://doi.org/10.1037/lhb0000082>
- *Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders* (Unpublished master's thesis). University of Leicester, U.K.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Bourgon, G., Mugford, R., Hanson, R. K., & Coligado, M. (2018). Offender risk assessment practices vary across Canada. *Canadian Journal of Criminology & Criminal Justice,*

60(2), 167-205. <https://doi.org/10.3138/cjccj.2016-0024>

Brankley, A. E., Babchishin, K. M., Chankin, L., Barsetti, I., & Hanson, R. K. (2019). *ACUTE-2007 Evaluator Workbook*. [Unpublished report]. Public Safety Canada.

Brankley, A., Helmus, L., & Hanson, R. (2017). *STABLE-2007 evaluator workbook: Updated recidivism rates (includes combinations with Static-99R, Static-2002R, and Risk Matrix 2000)* [Unpublished report]. Public Safety Canada.

Collins, G.S., Ogundimu, E.O., & Altman, D.G. (2016). Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. *Statistics in Medicine*, 35(2), 214-226. <https://doi.org/10.1002/sim.6787>

*Craissati, J., Bierer, K., & South, R. (2011). Risk, reconviction, and “sexually risky behaviour” in sex offenders. *Journal of Sexual Aggression*, 17(2), 153-165.

<https://doi.org/10.1080/13552600.2010.490306>

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, 23(7), 1173-1195. <https://doi.org/10.1080/13854040902795018>

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674. <https://doi.org/10.1126/science.2648573>

Daubert v. Merrell Dow Pharmaceuticals Inc., 507 U.S. 579 (1993)

*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, Bismarck, ND.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293-323.
<https://doi.org/10.1037/1076-8971.2.2.293>

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19-30.
<https://doi.org/10.1037/1040-3590.12.1.19>

Hanson, R. K. (2017). Assessing the calibration of actuarial risk Scales: A primer on the E/O Index. *Criminal Justice and Behavior*, 44(1), 26-39.
<https://doi.org/10.1177/0093854816683956>

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(5), 482-515. <https://doi.org/10.1177/1079063212469060>

Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment*, 29(5), 582-597. <https://doi.org/10.1037/pas0000371>

Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: Journal of Research and Treatment*, 17(4), 357-373. <https://doi.org/10.1177/107906320501700402>

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender

- recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362.
<https://doi.org/10.1037/0022-006X.66.2.348>
- Hanson, R. K., Harris, A. J. R., Helmus, L., & Thornton, D. (2014). High-risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence*, 29(15), 2792-2813.
<https://doi.org/10.1177/0886260514526062>
- Hanson, R. K., Harris, A. J. R., Letourneau, E., Helmus, L. M., & Thornton, D. (2018). Reductions in risk based on time offense-free in the community: Once a sexual offender, not always a sexual offender. *Psychology, Public Policy, and Law*, 24(1), 48-63.
<http://doi.org/10.1037/law0000135>
- *Hanson, R. K., Helmus, L.-M., & Harris, A. J. R. (2015). Assessing the risk and needs of supervised sexual offenders: A prospective study using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior*, 42(12), 1205-1224.
<https://doi.org/10.1177/0093854815602094>
- Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk scales. *International Journal of Forensic Mental Health*, 11(1), 9-23.
<https://doi.org/10.1080/14999013.2012.667511>
- Hanson, R. K., Lunetta, A., Phenix, A., Neeley, J., & Epperson, D. (2014). The field validity of Static-99/R sex offender risk assessment tool in California. *Journal of Threat Assessment and Management*, 1(2), 102-117. <https://doi.org/10.1037/tam0000014>
- Hanson, R. K., & Morton-Bourgon, K. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1-21. <https://doi.org/10.1037/a0014421>

- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.
<https://doi.org/10.1023/A:1005482921333>
- Hanson, R. K., Thornton, D., Helmus, L., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sexual Abuse: A Journal of Research and Treatment, 28*(3), 218-252. <https://doi.org/10.1177/1079063215574710>
- Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* (UMI No. MR58443) [Master's thesis, Carleton University]. ProQuest Dissertations and Theses database.
- 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*(9), 1148-1171. <https://doi.org/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*(1), 64-101.
<https://doi.org/10.1177/1079063211409951>
- Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560.
<https://doi.org/10.1136/bmj.327.7414.557>
- Hill, D., & Demetriooff, S. (2019). Clinical-forensic psychology in Canada: A survey of practitioner characteristics, attitudes, and psychological assessment practices. *Canadian Psychology/Psychologie canadienne, 60*(1), 55-63. <https://doi.org/10.1037/cap0000152>

- 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*(4), 425-448. [https://doi.org/ 10.1177/107906320701900407](https://doi.org/10.1177/107906320701900407)
- Joint Committee on the Standards for Educational and Psychological Testing of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*. American Educational Research Association.
- Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2020). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse*. Advance online publication. <https://doi.org/10.1177/1079063218800474>
- *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*(2), 107-120. <https://doi.org/10.1177/107906320401600202>
- *Lee, S. C., Hanson, R. K., Fullmer, N., Neeley, J., & Ramos, K. (2018). *The predictive validity of Static-99R over 10 years for sexual offenders in California: 2018 update*. State Authorized Risk Assessment Tools for Sex Offenders (SARATSO), http://saratso.org/pdf/Lee_Hanson_Fullmer_Neeley_Ramos_2018_The_Predictive_Validity_of_S_.pdf
- *Lee, S. C., Restrepo, A., Satariano, A., & Hanson, R. K. (2016). *The predictive validity of Static-99R for sexual offenders in California: 2016 update*. State Authorized Risk Assessment Tools for Sex Offenders (SARATSO), http://www.saratso.org/docs/ThePredictiveValidity_of_Static-99R_forSexualOffenders_inCalifornia-2016v1.pdf.

- *Lehmann, R. J. B., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K.-P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment, 25*(3), 1019-1024.
<https://doi.org/10.1037/a0033098>
- *Mercado, C. C., Jeglic, E., & Markus, K. (2011). *Sex offender management, treatment, and civil commitment: An evidence based analysis aimed at reducing sexual violence (2007-IJ-CX0037)*. National Institute of Justice.
- Moons, K. G., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Prognosis and prognostic research: What, why, and how? *British Medical Journal, 338*, 1317-1320.
<https://doi.org/10.1136/bmj.b375>
- Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior, 41*(12), 1406-1421. <https://doi.org/10.1177/0093854814548449>
- Olver, M. E., Mundt, J. C., Thornton, D., Beggs, D., Christofferson, S. M., Kingston, D. A., Sowden, J. N., Nicholaichuk, T. P., Gordon, A., & Wong, S. C. (2018). Using the Violence Risk Scale-Sexual Offense version in sexual violence risk assessments: Updated risk categories and recidivism estimates from a multisite sample of treated sexual offenders. *Psychological Assessment, 30*(7), 941-955.
<https://doi.org/10.1037/pas0000538>
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological methods, 3*(3), 354-379.
<https://doi.org/10.1037/1082-989X.3.3.354>
- Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, L. M., Hanson, R. K., & Thornton, D.

- (2016). *Static-99R coding rules - Revised 2016*. Public Safety Canada,
<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/sttc-2016/sttc-2016-en.pdf>
- Phenix, A., Helmus, L., & Hanson, R. K. (2016). *Static-99R & Static-2002R evaluators' workbook*. <http://www.static99.org/>
- Quesada, S. P., Calkins, C., & Jeglic, E. L. (2014). An examination of the interrater reliability between practitioners and researchers on the Static-99. *International Journal of Offender Therapy and Comparative Criminology*, 58(11), 1364-1375.
<https://doi.org/10.1177/0306624X13495504>
- R Core Team. (2013). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>
- 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.
<https://doi.org/10.1007/s10979-005-6832-7>
- Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, 13(1), 19-30. <https://doi.org/10.1037/1082-989X.13.1.19>
- Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie/Journal of Psychology*, 215(2), 90-103.
<https://doi.org/10.1027/0044-3409.215.2.90>
- 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.
<https://doi.org/10.1111/1529-1006.001>
- Thornton, D., Hanson, R. K., Kelley, S. M., & Mundt, J. C. (2019). Estimating lifetime and residual risk for individuals who remain sexual offense free in the community: Practical

applications. *Sexual Abuse*. Advance online publication.

<https://doi.org/10.1177/1079063219871573>

Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J. C., & Habbeman, J. D. F. (2005). Substantial effective samples sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology*, *58*(5), 475-483.

<https://doi.org/10.1016/j.jclinepi.2004.06.017>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. <https://doi.org/10.18637/jss.v036.i03>

Table 1*Descriptive Information on Routine/Complete Samples*

Study	Source	Country	Recidivism criteria	Type of sample	Release Period
Bartosh et al. (2003)	Published paper	U.S.	Charges	Corrections	1996
Bigras (2007)	Dissertation	Canada	Charges	CSC Reception Centre	1995-2004
Boer (2003)	Thesis	Canada	Conviction	CSC release	1976-1994
Craissati et al. (2011)	Published paper	U.K.	Conviction	Community supervision	1992-2005
Eher et al. (2009)	Unpublished data	Austria	Conviction	European prison	2000-2005
Epperson (2003)	Government report	U.S.	Charges	Prison and Probation	1989-1998
Hanson et al. (2015)	Published paper	Canada	Charges	Community supervision	2001-2009
Långström (2004)	Published paper	Sweden	Conviction	National prison release	1993-1997
Lee et al. (2016)	Government report	U.S.	Arrest	Probationers/parolees	2009-2010
Lee et al. (2018)	Government report	U.S.	Arrest	Parolees	2006-2007
Lehmann et al. (2013)	Published paper	Germany	Conviction	Berlin police registry	1994-2009
Mercado et al. (2011)	Government report	U.S.	Conviction	Prison and treatment group	1996-2004
Overall	-	-	-	-	1976-2010

Table 2*Descriptive Information for Routine/Complete Samples for 5-Year and 10-Year Norms of Static-99R*

	Static-99R	Age at release	Age range	Sexual recidivism					
				5 years		10 years		5 years	10 years
				<i>M (SD)</i>	<i>M (SD)</i>	<i>n/N</i>	%	<i>n/N</i>	%
Bartosh et al. (2003)	3.0 (2.7)	38.8 (12.4)	18-73	12/90	13.3%	-	-	.64 [.48, .81]	-
Bigras (2007)	2.1 (2.3)	41.4 (12.9)	21-77	18/206	8.7%	-	-	.75 [.62, .87]	-
Boer (2003)	2.8 (2.8)	41.1 (12.5)	20-80	11/299	3.7%	23/295	7.8%	.83 [.71, .94]	.76 [.65, .86]
Craissati et al. (2011)	2.1 (2.3)	37.6 (12.0)	18-72	15/200	7.5%	-	-	.71 [.56, .86]	-
Eher et al. (2009)	1.9 (2.1)	41.2 (12.6)	18-77	3/151	2.0%	-	-	.94 [.86, .99]	-
Epperson (2003)	2.6 (2.6)	37.0 (13.3)	19-77	16/150	10.7%	8/36	22.2%	.74 [.59, .88]	.90 [.78, .99]
Hanson et al. (2015)	2.5 (2.5)	41.2 (13.2)	18-84	77/629	12.2%	-	-	.72 [.65, .78]	-
Långström (2004)	2.0 (2.4)	41.5 (12.0)	18-77	69/1,278	5.4%	26/351	7.4%	.70 [.64, .77]	.75 [.64, .86]
Lee et al. (2016)	2.3 (2.4)	43.2 (11.8)	19-85	78/1,626	4.8%	-	-	.76 [.70, .81]	-
Lee et al. (2018)	2.4 (2.3)	42.9 (10.8)	20-86	23/371	6.2%	35/338	10.4%	.81 [.70, .91]	.75 [.65, .85]
Lehmann et al. (2013)	3.4 (2.2)	37.9 (11.8)	18-78	114/849	13.4%	78/443	17.6%	.73 [.68, .78]	.65 [.58, .72]
Mercado et al. (2011)	2.1 (2.2)	39.4 (11.9)	18-83	47/1,395	3.4%	16/136	11.8%	.69 [.60, .77]	.70 [.57, .83]
Overall	2.4 (0.7) *	40.3 (3.5) *	18-86	483/7,244	6.7%	186/1,599	11.6%	.76 [.73, .78]*	.73 [.69, .77]*

Note. * Fixed-effect meta-analytic averages. Numbers in bold indicate statistical significance.

Table 3*Fixed-Effect Meta-Analysis of Logistics Regression Parameters for 5-Year Norms of Static-99R*

	Parameter ($B0_2$)	%	S.E.	Parameter ($B1$)	Odds ratio	S.E.
Routine/Complete Samples						
Bartosh et al. (2003)	-2.053648	11.37	0.366938	0.137374	1.15	0.114597
Bigras (2007)	-2.708519	6.25	0.321044	0.398792	1.49	0.116650
Boer (2003)	-4.412028	1.20	0.605171	0.477777	1.61	0.139421
Craissati et al. (2011)	-2.820339	5.62	0.332409	0.340817	1.41	0.112092
Eher et al. (2009)	-5.801289	0.30	1.521597	1.013485	2.76	0.401447
Epperson (2003)	-2.651020	6.59	0.371366	0.346714	1.41	0.107381
Hanson et al. (2015)	-2.389223	8.40	0.162232	0.348774	1.42	0.054406
Långström (2004)	-3.118151	4.24	0.150795	0.308408	1.36	0.050260
Lee et al. (2016)	-3.575291	2.72	0.169968	0.429812	1.54	0.050165
Lee et al. (2018)	-3.449	3.08	0.347	0.506	1.66	0.105
Lehmann et al. (2013)	-2.773053	5.88	0.176890	0.425752	1.53	0.051912
Mercado et al. (2011)	-3.613641	2.62	0.182016	0.328338	1.39	0.069058
Overall (Fixed-effect)	-3.035568	4.58	0.066871	0.372803	1.45	0.021437
	$Q_{between} (df = 11)$	58.13, $p < .001$		13.72, $p = .249$		
	I^2	81.07		19.81		

Note. Numbers in bold indicate statistical significance.

Table 4*Fixed-Effect Meta-Analysis of Logistics Regression Parameters for 10-Year Norms of Static-99R*

	Parameter ($B0_2$)	%	S.E.	Parameter ($B1$)	Odds ratio	S.E.
Routine/Complete Samples						
Boer (2003)	-3.211018	3.88	0.357829	0.378009	1.46	0.094347
Epperson (2003)	-3.650418	2.53	1.236742	0.794208	2.21	0.290197
Långström (2004)	-2.979032	4.84	0.280137	0.405512	1.50	0.086117
Lee et al. (2018)	-2.740288	6.06	0.255408	0.447993	1.57	0.083868
Lehmann et al. (2013)	-2.056133	11.34	0.186182	0.271571	1.31	0.060165
Mercado et al. (2011)	-2.627089	6.74	0.415495	0.414826	1.51	0.160784
Overall (Fixed-effect)	-2.556535	7.20	0.118530	0.364911	1.44	0.037367
	$Q_{between} (df = 5)$	14.17, $p = .015$		5.91, $p = .315$		
	I^2	64.72		15.46		

Note. Numbers in bold indicate statistical significance.

Table 5

Moderator Analyses to Explore the Significant Variability of Sexual Recidivism Rates (adjusted 5-year base rates; B0₂) within Routine/Complete Samples

	B0 ₂	%	<i>Q</i>	<i>p</i>	<i>I</i> ²	<i>n/N</i>	<i>k</i>
<i>Q</i> _{total} from Table 3	-3.035	4.58	58.132	< .001	81.08	483/7,244	12
<u>Country</u>							
Canada	-2.560	7.18	10.687	0.005	81.28	106/1,134	3
U.S.	-3.371	3.32	19.920	< .001	79.92	176/3,632	5
European ¹	-2.971	4.88	5.870	0.118	48.89	201/2,478	4
			<i>Q</i>_{between}	21.655	< .001		
<u>Recidivism Criteria</u>							
Arrest/Charge	-2.887	5.28	34.307	< .001	85.42	224/3,072	6
Conviction	-3.168	4.04	19.403	.002	74.23	259/4,172	6
			<i>Q</i>_{between}	4.422	0.035		
<u>Source of study</u>							
Peer-reviewed articles	-2.740	6.07	14.556	0.006	72.52	287/3,046	5
Other ²	-3.449	3.08	16.230	0.013	63.03	196/4,198	7
			<i>Q</i>_{between}	27.346	< .001		
<u>Year of release (a median year)</u>							
Prior to 2000	-2.938	5.03	14.402	0.013	65.28	141/2,223	6
After 2000	-3.090	4.35	42.543	< .001	88.25	342/5,021	6
			<i>Q</i>_{between}	1.187	0.276		

Table 5 continues

Table 5 continued.

	B0 ₂	%	<i>Q</i>	<i>p</i>	<i>I</i> ²	<i>n/N</i>	<i>k</i>
<u>Author Involved</u>							
Yes	-2.940	5.02	28.539	<.001	89.49	292/3,475	4
No	-3.133	4.18	27.495	<.001	74.54	191/3,769	8
	<i>Q</i> _{between}		2.099	0.147			

Note. Numbers in bold indicate the statistical significance of *Q*_{between} analyses.

¹European countries included Austria, Germany, Sweden, and U.K.

²Other included government reports, thesis, dissertation, and unpublished data.

Figure 1

Estimated 5-Year Recidivism Rates for a Static-99R Score of 2 (B0₂; Upper) and its Relative Risk (B1; Lower) for Routine/Complete Samples

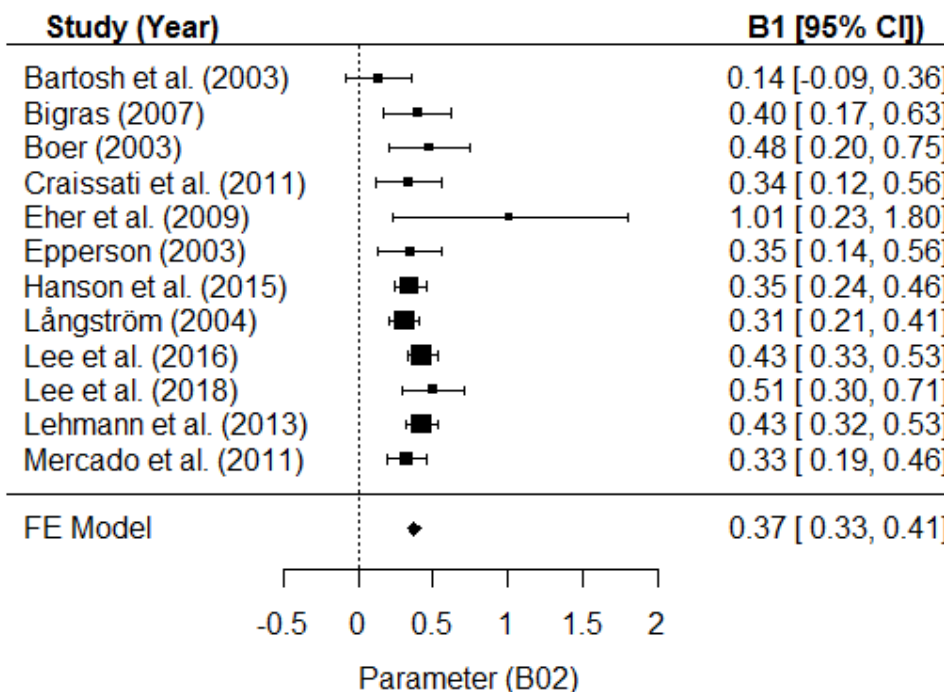
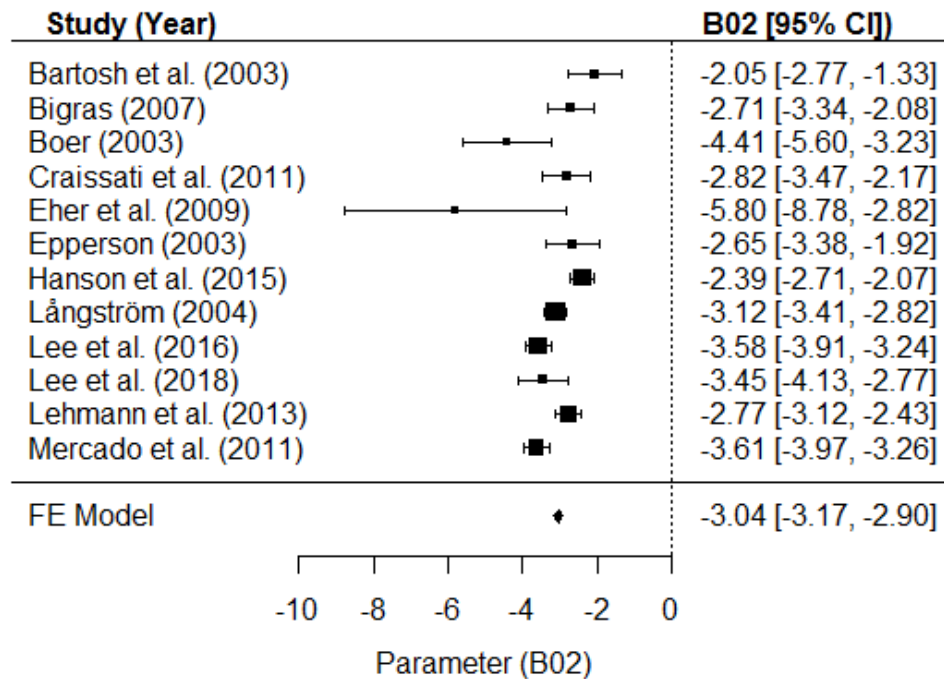


Figure 2

Estimated 10-Year Recidivism Rates for a Static-99R Score of 2 (B0₂; Upper) and its Relative Risk (B1; Lower) for Routine/Complete Samples

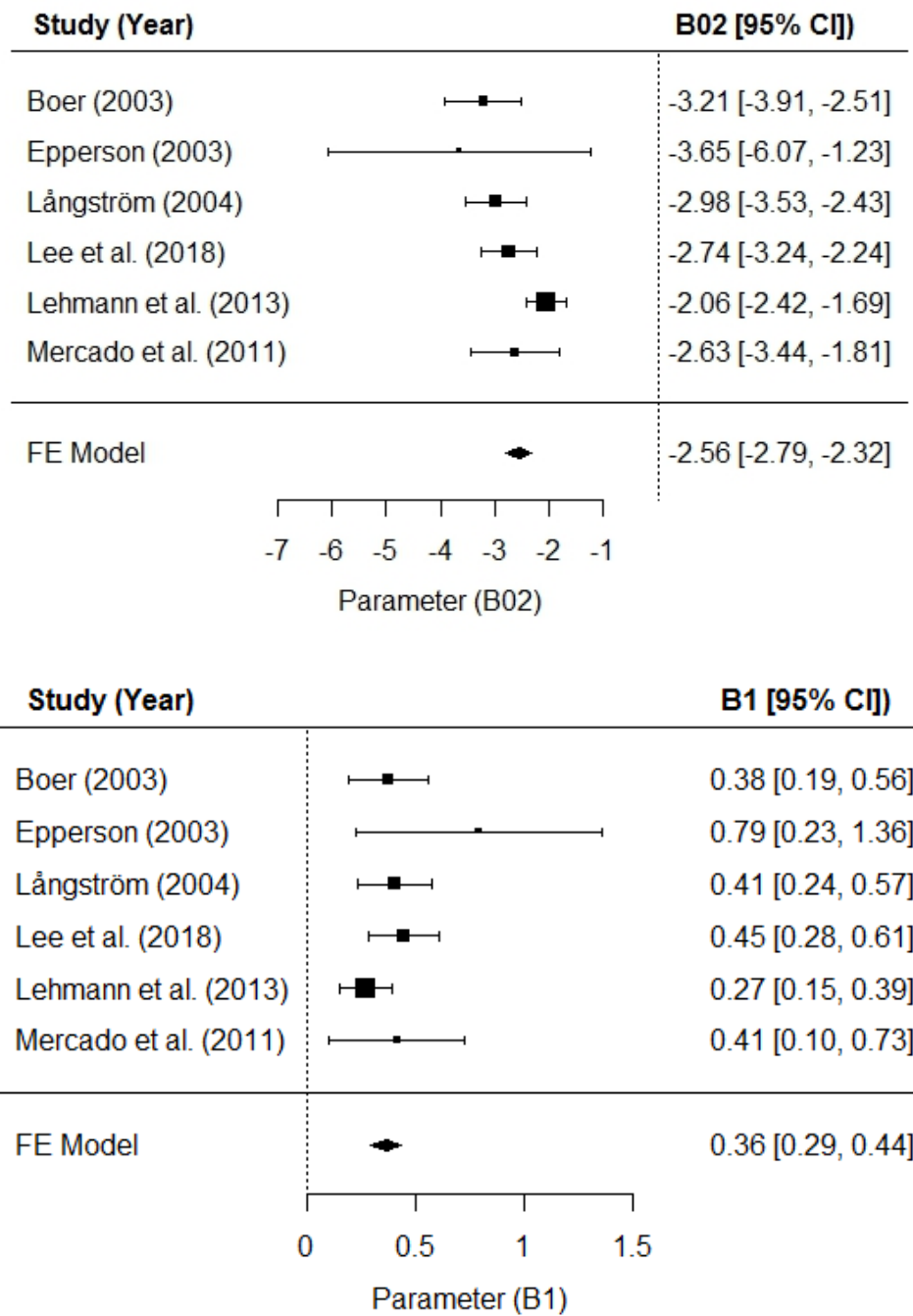
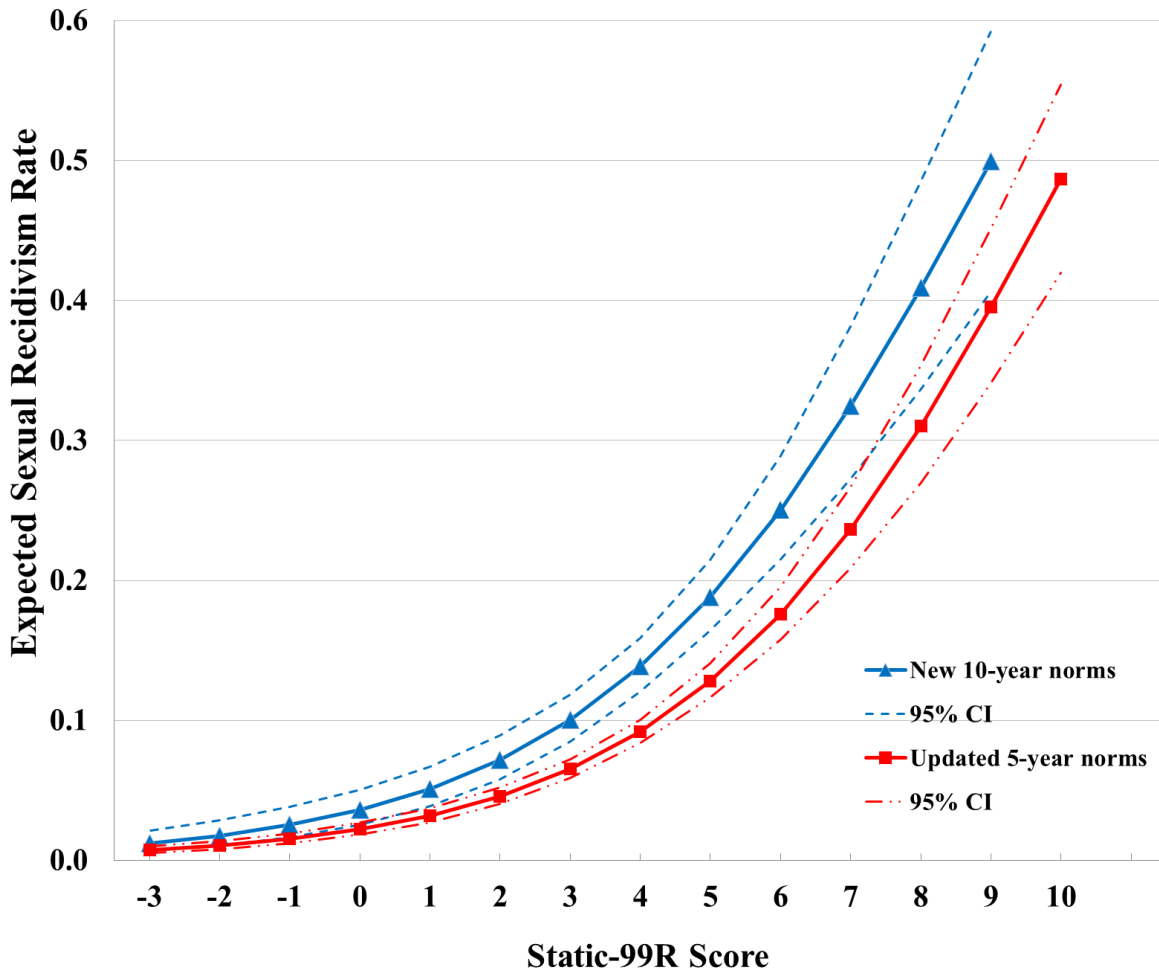


Figure 3

5-Year and 10-Year Sexual Recidivism Rates for Static-99R Routine/Complete Normative

Samples



Appendix

Table A1

*Observed and Estimated 5-Year Sexual Recidivism Rates for Static-99R: Routine/Complete**Samples*

Score	Risk Level	Fixed follow-up		Logistic regression estimates	
		Recidivists/total	Observed recidivism rate (%)	Estimated recidivism rate (%)	95% CI
-3	I	0/119	0.0	0.7	[0.5, 1.0]
-2	I	1/130	0.0	1.1	[0.8, 1.4]
-1	II	13/612	2.1	1.6	[1.2, 1.9]
0	II	18/848	2.1	2.2	[1.8, 2.7]
1	III	36/987	3.7	3.2	[2.7, 3.7]
2	III	35/1,155	3.0	4.6	[4.0, 5.2]
3	III	68/1,152	5.9	6.5	[5.8, 7.2]
4	IVa	74/965	7.7	9.2	[8.4, 10.1]
5	IVa	69/578	11.9	12.8	[11.7, 14.1]
6	IVb	61/332	18.4	17.6	[15.8, 19.6]
7	IVb	47/201	23.4	23.7	[20.9, 26.7]
8	IVb	36/112	32.1	31.0	[27.0, 35.4]
9	IVb	17/38	44.7	39.5	[34.1, 45.2]
10	IVb	6/12	50.0	48.7	[42.0, 55.4]
11	IVb	2/3	66.7	-	-
12	IVb	-	-	-	-
Total		483/7,244	6.7		

Note. The median values of the correlation of estimates from routine/complete samples ($k = 12$) were (r) = $-.6423$ (Static-99R 5-year estimates).

Table A2*Observed and Estimated 10-Year Sexual Recidivism Rates for Static-99R: Routine/Complete**Samples*

Score	Risk Level	Fixed follow-up		Logistic regression estimates	
		Recidivists/total	Observed recidivism rate (%)	Estimated recidivism rate (%)	95% CI
-3	I	0/16	0.0	1.2	[0.7, 2.1]
-2	I	0/23	0.0	1.8	[1.1, 2.9]
-1	II	5/111	4.5	2.5	[1.7, 3.8]
0	II	5/154	3.2	3.6	[2.6, 5.1]
1	III	13/194	6.7	5.1	[3.9, 6.7]
2	III	10/233	4.3	7.2	[5.8, 8.9]
3	III	26/249	10.4	10.1	[8.5, 11.9]
4	IVa	36/261	13.8	13.9	[12.1, 15.9]
5	IVa	18/147	12.2	18.8	[16.4, 21.5]
6	IVb	25/96	26.0	25.0	[21.5, 28.9]
7	IVb	21/59	35.6	32.5	[27.3, 38.2]
8	IVb	15/37	40.5	40.9	[33.7, 48.6]
9	IVb	8/14	57.1	50.0	[40.7, 59.2]
10	IVb	3/4	75.0	-	-
11	IVb	1/1	100.0	-	-
12	IVb	-	-	-	-
Total		186/1,599	11.6		

Note. The median values of the correlation of estimates from routine/complete samples ($k = 6$)

were (r) = $-.738354$ (Static-99R 10-year estimates).

Online Supplemental Material A

Table S1

Fixed-Effect Meta-Analysis of Logistics Regression Parameters for a Fixed 5-Year Follow-up Time of Static-99R from 10-year Normative Samples

	Parameter (B0 ₂)	%	S.E.	Parameter (B1)	Odds Ratio	S.E.
Routine/Complete Samples						
Boer (2003)	-4.412028	1.20	0.605171	0.477777	1.61	0.139421
Epperson (2003)	-2.651020	6.59	0.371366	0.346714	1.41	0.107381
Långström (2004)	-3.118151	4.24	0.150795	0.308408	1.36	0.050260
Lee et al. (2018)	-3.449	3.08	0.347	0.506	1.66	0.105
Lehmann et al. (2013)	-2.773053	5.88	0.176890	0.425752	1.53	0.051912
Mercado et al. (2011)	-3.613641	2.62	0.182016	0.328338	1.39	0.069058
Overall (Fixed-effect)	-3.172975	4.02	0.089656	0.372846	1.45	0.028801
	<i>Q</i> _{between} (<i>df</i> = 5)	17.91, <i>p</i> = .003		5.33, <i>p</i> = .377		
	<i>I</i> ²	72.07		6..22		

Table S2

Extrapolated Sexual Recidivism Rates for Follow-up Periods of 6 to 10 Years for Routine/Complete Samples Based on Estimated 5-year Recidivism Rates

Follow-up year	Initial risk (based on Static-99R scores)													
	Level I		Level II		Level III			Level IVa		Level IVb				
	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
5	0.7	1.1	1.5	2.2	3.2	4.6	6.5	9.2	12.8	17.6	23.7	31.0	39.6	48.7
6	0.8	1.2	1.8	2.5	3.6	5.2	7.4	10.4	14.4	19.7	26.4	34.3	43.5	53.2
7	0.9	1.3	1.9	2.8	4.0	5.7	8.1	11.4	15.8	21.6	28.7	37.2	46.8	56.8
8	1.0	1.4	2.1	3.0	4.3	6.2	8.7	12.3	17.0	23.1	30.7	39.6	49.6	59.7
9	1.1	1.5	2.2	3.2	4.6	6.6	9.3	13.0	18.0	24.5	32.4	41.6	51.9	62.2
10	1.1	1.6	2.3	3.4	4.8	6.9	9.8	13.7	18.9	25.6	33.8	43.3	53.8	64.2
11	1.2	1.7	2.5	3.5	5.1	7.2	10.2	14.3	19.6	26.6	35.1	44.8	55.5	65.9
12	1.2	1.8	2.5	3.7	5.3	7.5	10.6	14.8	20.3	27.5	36.2	46.1	56.9	67.3
13	1.3	1.8	2.6	3.8	5.4	7.7	10.9	15.2	20.9	28.3	37.1	47.2	58.0	68.5
14	1.3	1.9	2.7	3.9	5.6	7.9	11.2	15.6	21.4	28.9	37.9	48.1	59.1	69.6
15	1.3	1.9	2.8	4.0	5.7	8.1	11.4	16.0	21.9	29.5	38.6	48.9	60.0	70.4
16	1.4	1.9	2.8	4.0	5.8	8.3	11.6	16.2	22.3	30.0	39.2	49.6	60.7	71.2
17	1.4	2.0	2.9	4.1	5.9	8.4	11.8	16.5	22.6	30.5	39.8	50.2	61.4	71.8
18	1.4	2.0	2.9	4.2	6.0	8.5	12.0	16.7	22.9	30.8	40.2	50.8	61.9	72.4
19	1.4	2.0	2.9	4.2	6.1	8.6	12.1	16.9	23.2	31.2	40.6	51.2	62.4	72.9
20	1.4	2.1	3.0	4.3	6.1	8.7	12.3	17.1	23.4	31.5	41.0	51.7	62.8	73.3

Note. Bolded values are the 5-year logistic regression estimates for routine/complete samples (483 recidivists, $N = 7,244$)

Table S3

Extrapolated Sexual Recidivism Rates for Follow-up Periods of 5 and 11 to 20 Years for Routine/Complete Samples Based on Estimated 10-year Recidivism Rates

Follow-up year	Initial risk (based on Static-99R scores)													
	Level I		Level II		Level III			Level IVa		Level IVb				
	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
5	0.8	1.2	1.7	2.4	3.4	4.8	6.7	9.3	12.8	17.2	22.6	29.0	36.3	-
10	1.2	1.8	2.5	3.6	5.1	7.2	10.0	13.8	18.8	25.0	32.4	40.9	49.9	-
11	1.3	1.8	2.6	3.8	5.3	7.5	10.5	14.4	19.6	26.0	33.7	42.3	51.5	-
12	1.3	1.9	2.7	3.9	5.5	7.8	10.8	14.9	20.2	26.9	34.7	43.5	52.9	-
13	1.4	2.0	2.8	4.0	5.7	8.0	11.2	15.4	20.8	27.6	35.6	44.6	54.0	-
14	1.4	2.0	2.9	4.1	5.9	8.2	11.5	15.8	21.3	28.3	36.4	45.5	55.0	-
15	1.5	2.1	3.0	4.2	6.0	8.4	11.7	16.1	21.8	28.8	37.1	46.3	55.9	-
16	1.5	2.1	3.0	4.3	6.1	8.6	11.9	16.4	22.2	29.3	37.7	47.0	56.7	-
17	1.5	2.2	3.1	4.4	6.2	8.7	12.1	16.7	22.5	29.8	38.2	47.5	57.3	-
18	1.5	2.2	3.1	4.5	6.3	8.9	12.3	16.9	22.8	30.1	38.7	48.1	57.9	-
19	1.6	2.2	3.2	4.5	6.4	9.0	12.5	17.1	23.1	30.5	39.0	48.5	58.3	-
20	1.6	2.2	3.2	4.6	6.5	9.1	12.6	17.3	23.3	30.8	39.4	48.9	58.8	-

Note. Bolded values are the 10-year logistic regression estimates for routine/complete samples (186 recidivists, $N = 1,599$)

Table S4

Projected Residual Risk (Sexual Recidivism Rates [%]) From Time of Release Up to 20 Years Offense Free in the Community for

Routine/Complete Samples

Follow-up year	Initial risk (based on Static-99R scores)													
	Level I		Level II		Level III			Level IVa		Level IVb				
	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
At release	1.4	2.1	<u>3.0</u>	4.3	6.1	8.7	12.3	17.1	23.4	31.4	41.0	51.8	62.8	73.2
1	1.2	1.8	2.6	3.7	5.3	7.6	10.8	15.0	20.7	28.0	<u>36.8</u>	47.0	57.7	68.4
2	1.1	1.6	2.2	<u>3.2</u>	4.6	6.6	9.4	13.2	18.2	24.8	32.9	42.4	52.7	63.3
3	0.9	1.3	1.9	2.8	4.0	5.8	8.2	11.5	16.0	21.9	29.2	<u>38.1</u>	47.8	58.1
4	0.8	1.2	1.7	2.4	3.5	5.0	7.1	10.0	14.0	19.2	25.8	33.9	43.0	53.0
5	0.7	1.0	1.4	2.1	<u>3.0</u>	4.3	6.2	8.7	12.1	16.8	22.7	30.0	<u>38.5</u>	47.9
6	0.6	0.9	1.2	1.8	2.6	3.7	5.3	7.5	10.5	14.6	19.8	26.4	34.1	42.9
7	0.5	0.7	1.1	1.5	2.2	<u>3.2</u>	4.6	6.5	9.1	12.6	17.2	23.1	30.1	<u>38.1</u>
8	0.4	0.6	0.9	1.3	1.9	2.7	3.9	5.5	7.8	10.8	14.9	20.0	26.3	33.6
9	0.4	0.5	0.8	1.1	1.6	2.3	<u>3.3</u>	4.7	6.6	9.3	12.7	17.3	22.7	29.3
10	0.3	0.4	0.6	0.9	1.3	1.9	2.8	4.0	5.6	7.8	10.8	14.7	19.5	25.3
11	0.3	0.4	0.5	0.8	1.1	1.6	2.3	<u>3.3</u>	4.7	6.6	9.1	12.4	16.5	21.6
12	0.2	0.3	0.4	0.6	0.9	1.3	1.9	2.7	3.9	5.4	7.6	10.4	13.8	18.2
13	0.2	0.2	0.4	0.5	0.8	1.1	1.6	2.2	<u>3.1</u>	4.4	6.2	8.5	11.4	15.0
14	0.1	0.2	0.3	0.4	0.6	0.9	1.2	1.8	2.5	<u>3.6</u>	5.0	6.8	9.2	12.2
15	0.1	0.2	0.2	0.3	0.5	0.7	1.0	1.4	2.0	2.8	<u>3.9</u>	5.3	7.2	9.6
16	0.1	0.1	0.2	0.2	0.3	0.5	0.7	1.0	1.5	2.1	2.9	<u>4.0</u>	5.4	7.2
17	0.1	0.1	0.1	0.2	0.2	0.3	0.5	0.7	1.0	1.4	2.0	2.8	<u>3.8</u>	5.1
18	0.0	0.0	0.1	0.1	0.2	0.2	0.3	0.4	0.6	0.9	1.3	1.8	2.4	<u>3.2</u>
19	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.3	0.4	0.6	0.8	1.1	1.5
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

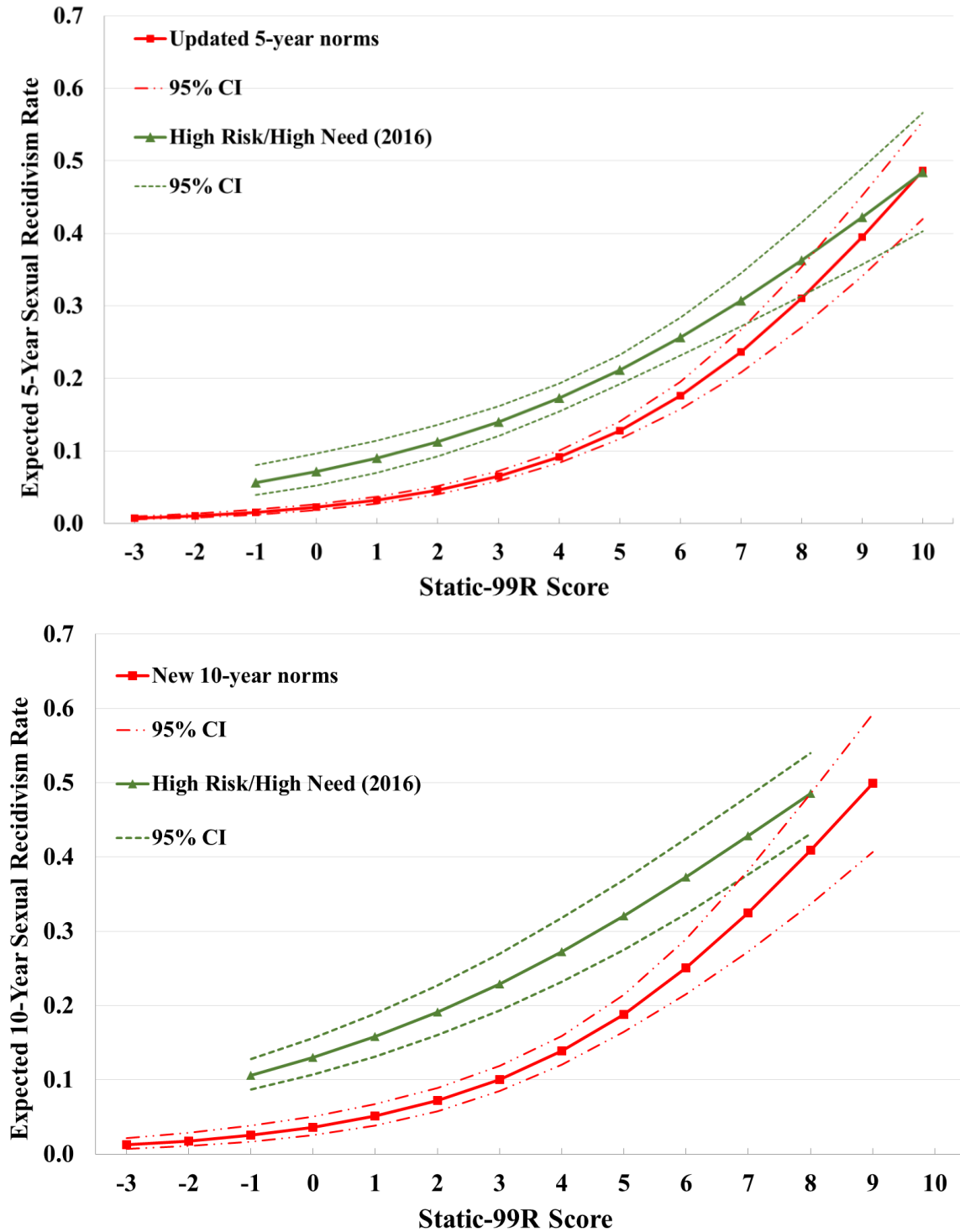
Note. Recidivism rate projections based on 5-year logistic regression estimates. Underlined values mark the transition out of Level

IVb (above 35%) and into Level I (less than 3%).

Figure S1

Comparisons of Sexual Recidivism Rates Between Routine/Complete and High Risk/ High Need

Normative Samples with 5-Year (Upper) and 10-Year (Lower) Follow-up Times



Online Supplemental Material B*Brief Descriptions of New or Updated Samples*

Lee et al. (2016). This study included adult male sexual offenders released from the California Department of Corrections and Rehabilitation (CDCR; i.e., parolees) as well as those on probation (i.e., probationers) between 2009 and 2010. All sex offenders in both groups had been convicted of a sexually-motivated offense against an identifiable victim (Category A offenses). Recidivism was defined as any subsequent arrest for a sexual offense (contact or non-contact) after released on community supervision as parolees or probationers. Recidivism information was provided by the California Department of Justice as of October 2015.

Lee et al. (2018). This study is an update on the previous study (Hanson et al., 2014) with additional information (e.g., death and deportation information). The samples included adult male sexual offenders released from the California Department of Corrections and Rehabilitation (CDCR; i.e., parolees) between 2006 and 2007. Recidivism was defined as any subsequent arrest for a sexual offense (contact or non-contact) after released on community supervision as parolees from the California Department of Corrections and Rehabilitation (CDCR). The California Department of Justice provided recidivism information.

Mercado et al. (2011). The sample was from a larger study examining sex offender management, treatment, and civil commitment (Mercado et al., 2011). All individuals in this study were adult males who were convicted of a sexual offense. The sample was selected from individuals who had been detained at either the Adult Diagnostic Treatment Center (ADTC) or any of New Jersey State Prisons. Sexual recidivism was defined as any subsequent conviction for a sexual offense (contact or non-contact) after release. Recidivism data were accessed from the New Jersey State Police criminal records database. These records include criminal records from

the state of New Jersey as well as other states who share their records with the New Jersey State Police.