

Obtaining More Precise Risk Estimates from Static-99R / STABLE-2007 Combinations

R. Karl Hanson & David Thornton

SAARNA

Published in the ATSA CATALYST on 05/14/2024 and reproduced here with the permission of the ATSA Executive Director, Amber Schroeder

STABLE-2007 is a structured approach to assessing psychological and community adjustment factors relevant to the recidivism risk of individuals with a history of sexual offending. Unlike some structured risk tools, STABLE-2007 was not designed to be used on its own; instead, it was designed to supplement other risk tools based on static, historical factors, such as Static-99R. STABLE-2007 contributes incrementally over Static-99R in the assessment of sexual recidivism risk (Brankley et al., 2021), and changes in STABLE-2007 scores are associated with changes in the likelihood of sexual recidivism (Lee et al., 2023, 2024). Although the individual items in STABLE-2007 could inform treatment and supervision efforts, there are no standardized risk levels or recidivism rate estimates associated with STABLE-2007 total scores (Brankley et al., 2017). Instead, the user guidance for STABLE-2007 provides empirically-derived, mechanical rules for combining STABLE-2007 total scores with totals scores from Static-99R, Static-2002R, and Risk Matrix-2000/S (Brankley et al., 2017). Specifically, the Static/STABLE combination rules sort individuals into the standardized risk levels for sexual recidivism (Level I to Level IVb; Hanson et al., 2017). For each risk level, recidivism rate estimates are provided using data from the Dynamic Supervision Project (DSP; Hanson et al., 2015).

The Problem

The above process for combining STABLE-2007 with Static-99R is widely used but is inherently imprecise. Scores on both instruments are first reduced to categories, and then the categories are combined. Although the broadly defined risk levels are useful in many decision contexts, evaluators may desire increased precision for some evaluation questions. In particular, meaningful treatment change in STABLE-2007 scores may have no effect on risk level

placements, thereby limiting the sensitivity of the broad categories for describing treatment change. For most Static-99R scores, it is possible to remain in the same risk level despite improvements of 8 points on the STABLE-2007. An 8 point change is substantial given that the mean of the STABLE-2007 is 7.4 with a standard deviation of 4.7. An 8 point difference is equivalent to a Cohen's d of 1.7, where d values of 0.20, 0.50, and 0.80 are considered small, medium and large, respectively. For the highest risk group, the range is even greater. For example, given a Static-99R score of 6, anyone who, during treatment, reduces their STABLE-2007 score from the extremely high range (19; 98th percentile) to the average range (8; 61th percentile) would still remain in the highest risk level (Level IVb, Well Above Average). A change from 19 to 8 would require improvements in almost all of the need areas measured by STABLE-2007, yet such cases would appear to be unchanged based solely on the categorical risk level placements.

Another inherent problem with all categories is that a difference of a single point can result in meaningfully different recidivism rate estimates when the scores straddle a risk level boundary. Consider, for example, someone with a Static-99R score of 3 and STABLE-2007 scores of 11 (79th percentile) at initial assessment and 12 (83rd percentile) upon reassessment. Using the combination rules, he would appear to have moved from Average Risk to Above Average Risk, even though the amount of apparent change is well within measurement error.

The broad risk levels are also poorly aligned with decisions based on estimated likelihoods. Although most treatment and supervision triage decisions are based on broad categories, some decisions require estimates of likelihoods. Decisions concerning civil commitment as a sexually dangerous person, for example, are often informed by a quantitative threshold of high risk (Knighton et al., 2014); conversely, release from public protection measures (e.g., registration as a sexual offender) can be informed by threshold of very low risk (Kahn et al., 2017; Thornton et al., 2021). Broad categories may not distinguish between persons above or below these thresholds. The average sexual recidivism rate for the category would be expected to accurately represent the risk of individuals in the middle of the risk level; however, it would over-estimate the risk for those at the bottom of the risk level while under-estimating risk for those at the top.

A Solution

Mechanically combining Static-99R and STABLE-2007 scores (rather than categories) increases the precision of the estimates. This was not previously done because the available data was too limited to support reliable recidivism rate estimates based on individual scores. There has been, however, additional research on the incremental effect of Static-99R and STABLE-2007 (summarized by Brankley et al., 2021) that can inform mechanical decision rules for combining Static-99R and STABLE-2007 scores.

Logistic regression is a well-established method for estimating likelihoods based on one or more predictor variables (Hanson, 2022, Chapter 11; Kleinbaum & Klein, 2010). Logistic regression can produce estimated recidivism rates for combined Static-99R and STABLE-2007 scores based on three statistics (parameters): a) the recidivism base rate (B_0), b) the incremental effect of Static-99R over STABLE-2007 ($B_{S99R.S07}$), and c) the incremental effect of STABLE-2007 over Static-99R ($B_{S07.S99R}$). The estimated recidivism rate (p) is expressed in the metric of logits ($\text{logit} = \ln[p/(1-p)]$).

$$\text{Logit}(p) = B_0 + (B_{S99R.S07} \times \text{Static-99R}) + (B_{S07.S99R} \times \text{STABLE-2007}) \quad \text{Equation 1}$$

It is common practice to estimate all logistic regression parameters using a single sample. To increase the stability of the estimates, parameters can be aggregated across samples using meta-analysis, which is the basis of the Static-99R recidivism rate norms (e.g., Lee & Hanson, 2021). There are no inherent problems, however, in using different samples to estimate the different logistic regression parameters. Provided that researchers maintain consistent measurement units, one sample (or meta-analysis) could be used to estimate the incremental effects and another sample could be used to estimate the base rate. This is the approach we used to increase the precision of the Static-99R/STABLE-2007 recidivism rate estimates.

The incremental effects of Static-99R over STABLE-2007 ($B_{S99R.S07}$) and STABLE-2007 over Static-99R ($B_{S07.S99R}$) are presented in Brankley et al.'s (2021) meta-analysis. Averaged across 12 studies ($n = 6,825$), we found incremental hazard ratios of 1.24 for Static-99R and 1.07 for STABLE-2007 for the prediction of sexual recidivism. Hazard ratios are not identical to odds

ratios (the metric used in logistic regression); however, for low base rate outcomes, such as sexual recidivism, hazard ratios are sufficiently similar to odds ratios that they can be treated as equivalent (see Hanson et al., 2013). Substituting the logged hazard ratios from Brankley et al. (2021) for the logged odds ratios in Equation 1 results in the following:

$$\text{Logit}(p) = B_0 + 0.215 \text{ Static-99R} + 0.0677 \text{ STABLE-2007} \quad \text{Equation 2}$$

The only parameter missing from Equation 2 is the recidivism base rate: B_0 . An estimate of B_0 can usefully be taken from the Dynamic Supervision Project (DSP) sample (Hanson et al., 2007). The DSP study was unusually careful in collecting sexual recidivism data for a routine sample; furthermore, the DSP sample is the basis of all other STATIC/STABLE recidivism rate estimates presented in the Evaluators' Workbook. For Risk Level III, the 5-year sexual recidivism rate is 7.5%. Level III corresponds to median scores of 2 on Static-99R and 7 on STABLE-2007; these scores (2 and 7) would also be the average scores for Level III, when rounded to whole numbers. When transformed into logits, the following equation is obtained.

$$\text{Logit} = -2.512 + .215 (\text{Static-99R centered on 2}) + 0.0677 \text{ STABLE-2007 (centered on 7)} \quad \text{EQ 3}$$

Centered here means that for Static-99R all scores have 2 points subtracted from them (i.e., a score of 2 becomes 0) and other scores are deviations above or below this. Similarly, all STABLE-2007 scores are centered by subtracting 7 points (a score of 7 becomes 0) and other scores are deviations above or below this. For example, considering an individual with a score of 6 on Static-99R and 14 on STABLE-2007. These scores become +4 and +7, respectively, after centering. The logit for this combination of scores is obtained as follows:

$$\text{Logit}(p) = -2.512 + 0.215(4) + 0.0677 (7) = -1.1781$$

Using the standard transformations (see Hanson, 2022, page 203), a logit of -1.1781 is equivalent to an expected recidivism rate of 23.5%. Note that these calculations are easily

automated in spreadsheets, such as Excel. We have developed such a spreadsheet, and it is available to be downloaded from the SAARNA website (www.saarna.org).

Margin of Error

Evaluators and decision-makers need to recognize that all estimates come with a margin of error. The most common way to describe the margin of error is with confidence intervals for the predicted values from a statistical model. For example, the 2021 Evaluator's Workbook for Static-99R and Static-2002R presents risk estimates with associated confidence intervals, with both the estimates and the confidence intervals derived from logistic regression equations. We are not able to directly apply this approach here because, as a consequence of using different samples to estimate different parameters, we did not have certain statistics needed to calculate the standard errors of the predictive values (the correlations of estimates). After some exploration of alternatives, we decided to report margins of error for the risk estimates based on the range of plausible values (confidence intervals) for the incremental b coefficient for STABLE-2007 controlling for Static-99R (the .0677 in Equation 3). Brankley et al.'s meta-analysis provides the standard error for this b coefficient as 0.0143. Multiplying by ± 1.96 to generate 95% confidence limits gives the lower bound for this coefficient to be .0397 and the upper bound to be .0957. These coefficients were then substituted for .0677 in Equation 3 to give a margin of error for the risk estimates.

It is important to note that this quantifies only one source of error. There are two other sources of error: how accurately the incremental b coefficient for Static-99R is estimated and how accurately the base rate is estimated. As noted above we are unable to incorporate them into our quantification of our margins of error because of missing information (the correlations of estimates). It is, nevertheless, worth considering how important these two unquantified sources of error may be.

The incremental contribution of Static-99R is unlikely to be an important source of error in the present case. The b coefficient for Static-99R in the Brankley et al. meta-analysis is similar to that obtained from much larger collections of samples. Consequently, we believe that the coefficient we used is reasonably close to the population (true) value.

Regarding the base rate, the Dynamic Supervision Project provides a very plausible estimate. Exceptional effort was devoted to collecting the recidivism information, including reports from supervision officers, national and provincial criminal history records, newspaper/internet searches, and direct contact with police in dozens of jurisdictions across Canada. Other samples often give lower estimates of the base rate because they rely on less carefully collected data (Lussier et al., 2023).

Based on these considerations, we believe our quantification of the margins of error captures the main source of error of this mechanical method for integrating STABLE-2007 with Static-99R.

Illustrating the Results

The following illustrations address high scores, which are relevant to evaluations of persons under the Sexually Violent Person laws in the USA or to Dangerous Offender sentencing provisions in Canada. Where someone scores a 6 or higher on Static-99R and 8 or above on STABLE-2007 the existing workbook indicates assignment in the highest risk level (Level IVb - Well Above Average Risk). As the more precise logistic regression results show, however, the range of recidivism rates within this category is quite large and may straddle the high risk thresholds used by decision-makers.

The table below shows the recidivism rates implied by the logistic regression equation (Equation 3) for a **Static-99R score of 6** combined with STABLE-2007 scores ranging from 2 to 18 in the middle column. The rightmost column shows the grouped sexual recidivism rates associated with these scores from the current STABLE-2007 Evaluators' Workbook.

STABLE-2007 Score	Five-Year Sexual Recidivism Rate associated with Static-99R =6	
	From Logistic Regression	From Existing Workbook
2	12.0% (10.6 - 13.6%)	7.5%
4	13.5% (12.6 - 14.5%)	13.6%

6	15.2% (14.8 - 15.6%)	13.6%
8	17.0% (16.6 - 17.4%)	26.8%
10	19.0% (17.8 - 20.3%)	26.8%
12	21.2% (18.9 - 23.6%)	26.8%
14	23.5% (20.2 - 27.3%)	26.8%
16	26.1% (21.5 - 31.2%)	26.8%
18	28.8% (22.9% - 35.5%)	26.8%

The next table shows the recidivism rates implied for a **Static-99R score of 8** combined with different STABLE-2007 scores.

STABLE-2007 Score	Five-Year Sexual Recidivism Rate associated with Static-99R =8	
	From Logistic Regression	From Existing Workbook
2	17.4% (15.4 – 19.5%)	13.6%
4	19.4% (18.1 – 20.7%)	13.6%
6	21.6% (22.1 – 22.1%)	13.6%
8	24.0% (23.5 – 24.5%)	26.8%
10	26.5% (24.9 – 28.2)	26.8%
12	29.2% (26.4 – 32.2)	26.8%
14	32.1% (28.0 – 36.5%)	26.8%
16	35.1% (29.6 – 41.1%)	26.8%
18	38.3% (31.3 – 45.8%)	26.8%

It is apparent that for some score combinations the rates from the existing Evaluators' Workbook are too high whereas for other score combinations they are too low. These differences could be meaningful when decisions are based on expected recidivism rates, and the decision threshold falls within the range provided by the risk level.

Recommendations

The recommended method of combining static actuarial risk scores with STABLE-2007 scores will depend on the purpose of the assessment. The methods described in the current STABLE-2007 Evaluators' Workbook are reasonable when the purpose is triaging a large population for more or less intensive treatment and supervision services. For evaluating response to treatment, however, the methods described here provide a more sensitive measure of change than would the categories. Combining specific scores (not categories) should be used whenever the range of recidivism rate estimates spanned by the category includes the recidivism rate used as the decision threshold.

References

- Brankley, A. E., Babchishin, K. M., & Hanson, R. K. (2021). STABLE-2007 Demonstrates Predictive and Incremental Validity in Assessing Risk-Relevant Propensities for Sexual Offending: A Meta-Analysis. *Sexual Abuse, 33*(1), 34–62. <https://doi.org/10.1177/1079063219871572>
- Brankley, A. E., Helmus, L. M., & Hanson, R. K. (2017). *STABLE-2007 evaluator workbook: Revised 2017*. SAARNA: The Society for the Advancement of Actuarial Risk Needs Assessment.
- Hanson, R. K. (2022). *Prediction statistics for psychological assessment*. American Psychological Association.
- 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, 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., 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. doi:10.1177/0093854815602094
- Kahn, R. E., Ambroziak, G., Hanson, R. K., & Thornton, D. (2017). Release from the sex offender label. *Archives of Sexual Behavior*, 46(4), 861-864. <https://doi.org/10.1007/x10508-017-0972-y>
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic regression : A self-learning text* (3rd ed.). Springer.
- Knighton, J. C., Murrie, D. C., Boccaccini, M. T., & Turner, D. B. (2014). How likely is “likely to reoffend” in sex offender civil commitment trials? *Law and Human Behavior*, 38(3), 293-304. <https://doi.org/10.1037/lhb0000079>
- Lee, S. C., Babchishin, K. M., Mularczyk, K. P., & Hanson, R. K. (2023). *Dynamic risk scales decay over time: Evidence for reassessment*. Research Report 2023-R003. Ottawa: Public Safety Canada. <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2023-r003/index-en.aspx>
- Lee, S. C., Babchishin, K. M., Mularczyk, K. P., & Hanson, R. K. (2024). Dynamic risk scales degrade over time: Evidence for reassessments. *Assessment*, 31(3), 698-714. <https://doi.org/10.1177/10731911231177227>
- Lee, S. C., & Hanson, R. K. (2021). Updated 5-year and new 10-year sexual recidivism rate norms for Static-99R with routine/complete samples. *Law and Human Behavior*. 45(1), 24-38. <https://doi.org/10.1037/lhb0000436>
- Lussier, P., McCuish, E., & Jeglic, E. L. (2023). Against all odds: The unexplained sexual recidivism drop in the United States and Canada. *Crime and Justice*, 52(1), 125-196. <https://doi.org/10.1086/727028>