STATIC RISK AND OFFENDER NEEDS GUIDE-REVISED FOR SEX OFFENDERS (STRONG-R-SO)

Ming-Li Hsieh & Zachary K. Hamilton
Background

In 1990, the passage of the Community Protection Act (CPA)\(^1\) required the Washington State Department of Corrections (DOC) to authorize the release of information to the community regarding dangerous sex offenders. The subsequent amendments\(^2\) require sex offenders to register in their county of residence following release from facilities. Offenders also have to update their registration in a moving circumstance.

Following the 1997 Legislature,\(^3\) the End of Sentence Review Committee (ESRC) employed risk assessment tools\(^4\) as part of its sex offender notification risk classification process as a statewide approach. The risk classification is, briefly, based upon an offender’s criminal history, demographic characteristics and other relevant information and identifying the level of risk posed by convicted sex offenders.

The classification tool used by the ESRC, however, was not found to accurately classify offender’s level of risk through comprehensive analysis and evaluations.\(^5\) Given these findings, the DOC set in motion a plan to create a risk assessment model for improving the recidivism prediction accuracy of these risk assessment instruments.\(^6\)

In 2013, the Washington State Institute for Criminal Justice (WSICJ) contracted with the DOC to develop a new risk assessment tool—the Static Risk and Offender Needs Guide-Revised (STRONG-R)\(^7\). The STRONG-R has the highest predictive accuracy of criminal recidivism (AUC = .72) for a Washington-based population.\(^8\)

The STRONG-R for sex offenders (STRONG-R-SO) is subproject of the STRONG-R development. The STRONG-R-SO aims to tailor to the needs of the local offender population by employing sex offender specific items to a newly constructed general offender assessment. Besides predicting reoffending, the DOC also expected the new instrument could apply for case supervision purposes, concerned with sexual felony offenders, misdemeanors, as well as failure to register convictions.

---

\(^1\) RCW 4.24.550.
\(^2\) RCW 9A.44.130.
\(^3\) RCW 4.24.5502.
\(^4\) Washington State Sex Offender Risk Level Classification Tool.
\(^8\) *Id.* Drake, E. (2014).
STRONG-R-SO

Goal. The primary goal for the STRONG-R-SO is to integrate a general risk assessment with sex offender specific items and to predict 1) sex offense (any felony or misdemeanor), 2) failure to register (FTR), and 3) any sex or sex-related offense (a combination of 1 and 2).

Sample. The STRONG-R-SO includes male subjects who: a) possessed and instant offense of a felony by a Washington Stat Court, b) were supervised by the DOC, c) received the STRONG-R general recidivism assessment, d) received the Static-99R and MnSOST-R assessment, e) had a prior sex offence conviction, and f) reentered the community between August 2008 through December 2010. The total sample size for the study was 1,024.

Study Design

◆ Step 1-Identifying Predictors

The collection of STRONG-R and sex offender trailer make up the item pool used to develop the current instrument. Many of the static, and risk, needs and responsivity (RNR)\(^9\) factors overlap considerably with the content collected as part of the STRONG-R interview\(^{10}\). Therefore, the large pool of general recidivism predictors is taken from the STRONG-R tool. Another 11 sex offender specific measures selected from the Static-99R and the MnSOST-R. The final 36 measures identified for the model through a series of multiple regression approach.

◆ Step 2-Weighted Items by Logit

We created an algorithm (programmed in R), to compute a backward stepwise regression procedure that selected items for model inclusion only if they met two criteria – a) the item improved the AIC prediction value and b) the item’s logit was positively weighted. To increase fit and model stability we repeat model selection steps within 1,000 bootstrap samples. Items assessed to meet described criteria in greater than 51% of the samples were retained and represent the “training” model.

◆ Step 3-Internal Validation

Next, validation procedures were used identifying stability of estimates over another 1,000 bootstrapped samples. Exhibit 1 displays internal validity model based on the AUC statistics. Findings revealed strong effects for all three STRONG-R-SO models (OC-AUC range = .72-.74).

---


**Exhibit 1. The STRONG-R-SO Internal Validity Model**

<table>
<thead>
<tr>
<th>Model</th>
<th>A-AUC</th>
<th>AOD</th>
<th>OC-AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex Offense</td>
<td>0.83</td>
<td>0.11</td>
<td>0.72</td>
</tr>
<tr>
<td>FTR</td>
<td>0.89</td>
<td>0.15</td>
<td>0.74</td>
</tr>
<tr>
<td>Any Sex Related</td>
<td>0.84</td>
<td>0.11</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Note: 
- a Computed the AUC using the entire sample called Apparent AUC.
- b This statistic is created by taking another 1,000 bootstrap samples and computing the model AUC for each. The difference between the apparent AUC and each of the 1,000 samples is computed and averaged called the Average Optimism Distance.
- c Optimism Corrected AUC = A-AUC – AOD.

Additional calibration examination was conducted and the Brier score of STRONG-R-SO demonstrated slightly improved outcomes when compared to other instruments. Exhibit 2 showed internal validity comparisons results.

**Exhibit 2. Internal Validity Comparisons of Three Instruments**

<table>
<thead>
<tr>
<th>Model</th>
<th>Discrimination</th>
<th>Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OC-AUC</td>
<td>H</td>
</tr>
<tr>
<td>STRONG-R-SO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Offense</td>
<td>0.72</td>
<td>0.33</td>
</tr>
<tr>
<td>FTR</td>
<td>0.73</td>
<td>0.53</td>
</tr>
<tr>
<td>Any Sex Related</td>
<td>0.73</td>
<td>0.39</td>
</tr>
<tr>
<td>STATIC-99R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Offense</td>
<td>0.63***</td>
<td>0.16</td>
</tr>
<tr>
<td>FTR</td>
<td>0.52***</td>
<td>0.05</td>
</tr>
<tr>
<td>Any Sex Related</td>
<td>0.62***</td>
<td>0.09</td>
</tr>
<tr>
<td>MnSOST-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Offense</td>
<td>0.59***</td>
<td>0.10</td>
</tr>
<tr>
<td>FTR</td>
<td>0.52***</td>
<td>0.05</td>
</tr>
<tr>
<td>Any Sex Related</td>
<td>0.56***</td>
<td>0.07</td>
</tr>
</tbody>
</table>

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

<sup>a</sup> A widely-recognized statistic for assessing risk score calibration. Brier scores range from 0 (perfect) to 0.25 (worthless).

**Summarized Findings**

- The STRONG-R-SO is a unified system, in which a sex offender assessment is combined with a general risk assessment.
- The STRONG-R-SO incorporated RNR factors as well as sex offender specific items.
- The STRONG-R-SO is an actuarial risk assessment model, used analytic item weights and bootstrap strategy, and boosting its measurement validation.
- The STRONG-R-SO tailored to the needs of the local offender population and demonstrated strong effect size in terms of predictive accuracy.

**Conclusions**

The STRONG-R-SO tailored its design to sex offenders in Washington State, and their demographic characteristics. In the interests of fair and accurate risk classification for sex offenders as well as increased public safety, updating weights and recalibrating the STRONG-R-SO will be routinely necessary to retain predicative validity. As discussed, the preliminary results in the current study showed the improved accuracy, due in part to the norming of item weights for the Washington State population and the addition of general offender risk assessment items. This also allowed for additional specificity of outcomes, examining models differentiating predictions for both sex offenses and FTR.

The levels of predictive validity have reached substantial and satisfactory levels.
with use of the current item pool, however, the STRONG-R-SO is a work in progress and version updates are occurring regularly following initial implementation of the current instrument. With the assistance of the WADOC, the STRONG-R-SO software will be developed to compute individual risk scores as a partially automated computer-based software application for evaluators.

★What is the AUC?
The Area Under Curve (AUC) is widely used measure of predictive utility between recidivism and risk classification. AUCs range can vary between .5 and 1.0 and the larger AUCs represent higher accuracy. According to Rice and Harris (2005), the AUC values of small, medium, and large effects, are .556, .639, .714, respectively.

---


---


March 2014