Static-99(R) and Static-2002(R): How to Interpret and Report in Light of Recent Research

R. Karl Hanson, Ph.D.
Amy Phenix, Ph.D.
Leslie Helmus, M.A.
Pre-Conference Workshop at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, September 30, 2009

Why Assess Risk?

- Understand Threat
- Promote Public Safety
- Promote Effective Treatment
  - Risk/Need/Responsivity
- Risk-Based Decisions in Corrections and Mental Health
  - Family re-integration, Parole, Civil Commitment, Sentencing
- Allocation of Scarce Resources

Risk Assessment

- Source of the risk (explanation)
- Nature of potential harm
- Likelihood of harm
  - Relative risk (Karl is twice as risky as David)
  - Absolute risk (34% after 5 years)

Empirical Probabilities

- Life is too complicated to think through all the possibilities
- Estimate probabilities by observing the outcome in groups of offenders “like him”.

Types of Risk Assessment

<table>
<thead>
<tr>
<th>Type of Evaluation</th>
<th>Factors</th>
<th>Overall Evaluation</th>
<th>Recidivism Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured Clinical Judgement</td>
<td>?</td>
<td>Professional Judgement</td>
<td>No</td>
</tr>
<tr>
<td>Empirical-Actuarial</td>
<td>Empirically Derived</td>
<td>Mechanical Actuarial</td>
<td>Yes</td>
</tr>
<tr>
<td>Structured Professional Judgement</td>
<td>Theory</td>
<td>Professional Judgement</td>
<td>No</td>
</tr>
<tr>
<td>Mechanical</td>
<td>Theory</td>
<td>Mechanical</td>
<td>No</td>
</tr>
</tbody>
</table>

Hanson & Morton-Bourgon (2009) Meta-analysis

- 1972-2008 (median 2004)
- 151 documents; 110 studies; 118 samples
- 37% published
- Total n = 45,398 sexual offenders
- 16 countries
  - Canada, US, UK, France, Netherlands, Germany, Denmark, Australia, Sweden, Austria, New Zealand, Belgium, Taiwan, Japan, Switzerland, Spain
- Four languages
  - English, French, Chinese, Spanish
d "standardized mean difference"

- How much are the recidivists different from the non-recidivists, in comparison to how much the recidivists and non-recidivists are different from each other.

  \[
  \begin{align*}
  &.20 \text{ small} \\
  &.50 \text{ medium} \\
  &.80 \text{ large}
  \end{align*}
  \]

Prediction of sexual recidivism

<table>
<thead>
<tr>
<th>Measures</th>
<th>d (95% CI)</th>
<th>N (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designed for Sexual Recidivism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical Actuarial</td>
<td>.67 (.63-.72)</td>
<td>24,089 (81)</td>
</tr>
<tr>
<td>Mechanical</td>
<td>.66 (.58-.74)</td>
<td>5,838 (29)</td>
</tr>
<tr>
<td>Structured Judgement</td>
<td>.46 (.29-.62)</td>
<td>1,131 (6)</td>
</tr>
<tr>
<td>Unstructured</td>
<td>.42 (.32-.51)</td>
<td>6,456 (11)</td>
</tr>
</tbody>
</table>

Accuracy and Error

- Inter-Rater Reliability
- Relative Risk (rank order; rate ratios)
- Absolute Recidivism Rates
- Confidence Intervals for Group Estimates
- Extent of Unmeasured, External Risk Factors
  - Incremental validity studies
  - "unexplained" variability across studies

Inter-Rater Reliability – STATIC-99

<table>
<thead>
<tr>
<th>Study</th>
<th>Size</th>
<th>Statistic</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbaree et al. (2001)</td>
<td>30</td>
<td>Pearson r – total scores</td>
<td>.90</td>
</tr>
<tr>
<td>Hanson (2001b)</td>
<td>55</td>
<td>% agreement-items</td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>Kappa- items</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>Intra-class r – total score</td>
<td>.87</td>
</tr>
<tr>
<td>Sjöstedt &amp; Långström (2001)</td>
<td>20</td>
<td>Kappa – items</td>
<td>.90</td>
</tr>
<tr>
<td>Harris et al. (2003)</td>
<td>10</td>
<td>Intra-class r - total scores</td>
<td>.87</td>
</tr>
</tbody>
</table>

Inter-rater Reliability – STATIC-2002

<table>
<thead>
<tr>
<th>Study</th>
<th>Size</th>
<th>Statistic</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Langton et al. (2007)</td>
<td>25</td>
<td>Pearson r</td>
<td>.90</td>
</tr>
<tr>
<td>Haag (2005)</td>
<td>66</td>
<td>Pearson r</td>
<td>.84</td>
</tr>
<tr>
<td>Knight &amp; Thornton (2007)</td>
<td>258</td>
<td>Pearson r</td>
<td>.89</td>
</tr>
<tr>
<td>Bengtson (2008)</td>
<td>20</td>
<td>Intra-class Correlation</td>
<td>.96</td>
</tr>
</tbody>
</table>
**Standard Error of Measurement**

\[ SEM = SD \sqrt{1 - r_{xx}} \]

**SEM: STATIC-99**

\[ SEM = 1.97 \sqrt{1 - .87} = .71 \]

95% C.I. = 1.96 x .71 = 1.39

**Result:** 19 times out of 20, the offender’s true score will be within ± 1.4 points of the observed score.

**SEM: STATIC-2002**

\[ SEM = 2.6 \sqrt{1 - .90} = .82 \]

95% C.I. = 1.96 x .82 = 1.61

**Result:** 19 times out of 20, the offender’s true score will be within ± 1.6 points of the observed score.

**Variability of Group Estimates**

- **Confidence Intervals**
  - Get narrower as sample size increases
  - Intervals derived from logistic regression uses information from full sample (not just specific score)

**Example: 10 year Sexual Recidivism 95% Confidence Intervals for Logistic Regression Recidivism Estimates**

**Assumptions for Group Confidence Intervals**

- All individuals in each category have the same probability of recidivism
- All relevant risk factors have been measured
  - BUT neither Static-99 nor STATIC-2002 claim to measure all relevant risk factors (heterogeneity within groups is expected)
- Requires assumptions about the similarity between the individual and the group data
**Incremental Validity Studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Static</th>
<th>Other measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thornton (2002)</td>
<td>99</td>
<td>SRA</td>
</tr>
<tr>
<td>Hanson et al. (2007)</td>
<td>99</td>
<td>STABLE-2007</td>
</tr>
<tr>
<td>Hanson &amp; Helmus (2009)</td>
<td>02</td>
<td>STABLE-2007</td>
</tr>
<tr>
<td>Knight &amp; Thornton (2007)</td>
<td>99, 02</td>
<td>SRA</td>
</tr>
</tbody>
</table>

**Accuracy and Error: Strengths**

- **High** inter-rater reliability
- **Consistent** relative risk (rank order; rate ratios)
- **Narrow** confidence intervals for group estimates

**Accuracy and Error: Absolute Recidivism Rates**

- Unmeasured, external risk factors
- Shown by
  - Incremental validity studies
  - "unexplained" variability across studies

**10 Year Sexual Recidivism Rates**

(from logistic regression estimates)

![Graph showing 10 Year Sexual Recidivism Rates](image-url)
## Purpose of our research

- Are new norms needed for Static-99?
- What should the recidivism estimates for Static-99 and Static-2002 look like?
- What should we do with base rates variability across samples?

## Obtaining Samples

- Sought all Static-99/Static-2002 replications
- Required
  - Appropriate population (e.g., adult male sex offenders)
  - Complete data for Static scores (Ever Lived with Lover – only permissible missing item)
  - Recidivism rates based on fixed follow-up periods (5 years and 10 years)

### Static-2002: 8 samples

- Bengtson (2008)
- Bigras (2007)
- Boer (2003)
- Haag (2005)
- Hanson et al. (2007)
- Harkins & Beech (2007)
- Knight & Thornton (2007)
- Langton et al. (2007)

**NOTE:** These 8 samples were also included in the Static-99 research

### Static-99: 28 samples with sexual recidivism data

- Allan et al. (2007)
- Bartosh et al. (2003)
- Bengtson (2008)
- Bigras (2007)
- Boer (2003)
- Bonta & Yessine (2005)
- Brouillette-Alarie & Proulx (2008)
- Cortoni & Nunes (2007)
- Craig et al. (2006)
- Craissati et al. (2008)
- de Vogel et al. (2004)
- Eher et al. (2008)
- Endrass et al. (2009)
- Epperson (2003)
- Haag (2005)
Static-99: 28 samples with sexual recidivism data
• Hanson et al. (2007)
• Harkins & Beech (2007)
• Hill et al. (2008)
• Johansen (2007)
• Knight & Thornton (2007)
• Långström (2004)

Preparing the datasets
• Corrected coding errors or inconsistencies where possible
• Deleted cases if:
  – No follow-up information
  – Any item other than “ever lived with a lover for two years” is missing
  – Inconsistencies not resolvable

Basic descriptive information
• Most offenders released 1990 or later
  – Static-99: >80%; Static-2002: ~70%
• Samples from Canada, US, UK, Europe
• Samples primarily treated
  – Static-99: only one untreated sample
  – All other samples: either mostly treated or mixed
• Mean age was 39 (Static-2002) to 40 (Static-99)

Basic descriptive information
• Roughly half used charges as recidivism criteria
  – Static-99: 13 samples used charges, 15 used convictions
  – Static-2002: 4 used charges, 4 used convictions
• Approximately half the offenders were child molesters
  – Static-99 (k = 15, n = 6,335): 53% child molesters, 37% rapists
  – Static-2002 (k = 5, n = 1,860): 55% child molesters, 45% rapists

Analyses
• Logistic regression
  – Absolute and relative risk ($B_0$ and $B_1$)
• Cox regression analyses
  – Does not provide a base rate estimate
• Meta-analysis of logistic regression coefficients
  – Fixed effect for moderator analyses
  – Random effect for recidivism estimates

Overview of Analyses

Note: Harris et al. (2003) also obtained but with violent recidivism data only
Logistic regression

- Requires standard (fixed) follow-up time
- $B_0$ (intercept) – predicted value of DV when IV equals zero
  - Logistic regression: expressed as log odds
  - Proxy for base rate – absolute risk
  - Can center on different scores to examine base rates at different points on Static
- $B_1$ (slope) – amount of change in DV associated with one-unit increase in IV
  - Logistic regression: expressed as average log odds ratio
  - Measure of predictive accuracy – relative risk

Cox Regression

- Purpose: examine incremental contribution of predictor variable(s) in survival data
- Has the advantages of survival analysis (correct for unequal follow-up – allows for more cases to be included)
- No intercept (looks at relative risks): can’t be directly used to produce recidivism estimates

Meta-analysis

- Fixed effect model
  - Conceptually, results restricted to the studies used
  - Variability across samples measured separately (in the Q statistic)
- Random effect model
  - Conceptually, estimates population from which the studies are a part
  - Incorporates variability across samples into the error term (confidence intervals are larger)
- If variability across studies is low ($Q < df$), both models provide identical results
- Aggregating logistic regression results – uses fixed follow-up periods

Warning: Fluctuating sample sizes

- Overall data on sexual recidivism
  - Static-2002 ($k = 8, n = 2,959$)
  - Static-99 ($k = 28, n = 8,893$)
- 5 year logistic regression: approximately 2/3 of total sample
  - Static-2002 ($k = 8, n = 1,865$)
  - Static-99 ($k = 27, n = 6,285$)
- 10 year logistic regression: approximately 1/3 of total sample
  - Static-2002 ($k = 5, n = 1,104$)
  - Static-99 ($k = 18, n = 2,528$)
- Moderator analyses: n’s fluctuate depending on which samples have info

Are new Static-99 norms needed?

![Sexual Recidivism at 5 years (Survival Analysis)](image)
Exploring Static-99 & Static-2002 risk properties (relative and absolute) across samples

NOTE: Analyses will focus on sexual recidivism from here onwards

<table>
<thead>
<tr>
<th>Static-2002 logistic regression meta-analysis</th>
<th>$M$</th>
<th>95% CI</th>
<th>$Q$</th>
<th>$k$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Five Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_j$</td>
<td>0.26</td>
<td>0.20–0.31</td>
<td>5.69</td>
<td>7</td>
<td>1,892</td>
</tr>
<tr>
<td>$B_0$ Centered 0</td>
<td>4.1%</td>
<td>2.8–6.1</td>
<td>17.62**</td>
<td>7</td>
<td>1,892</td>
</tr>
<tr>
<td><strong>Ten Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_j$</td>
<td>0.24</td>
<td>0.17–0.30</td>
<td>4.45</td>
<td>4</td>
<td>1,085</td>
</tr>
<tr>
<td>$B_0$ Centered 0</td>
<td>7.0%</td>
<td>4.6–10.4</td>
<td>12.62**</td>
<td>4</td>
<td>1,085</td>
</tr>
</tbody>
</table>
Static-99 logistic regression meta-analysis

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>95 % CI</th>
<th>Q</th>
<th>k</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Five Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_0$</td>
<td>0.31</td>
<td>0.27 – 0.35</td>
<td>24.91</td>
<td>25</td>
<td>6,233</td>
</tr>
<tr>
<td>$B_0$ Centered 0</td>
<td>4.3%</td>
<td>3.6 – 5.1</td>
<td>59.73***</td>
<td>27</td>
<td>6,281</td>
</tr>
<tr>
<td>$B_0$ Centered 2</td>
<td>8.4%</td>
<td>7.5 – 9.3</td>
<td>149.09***</td>
<td>27</td>
<td>6,281</td>
</tr>
<tr>
<td>$B_0$ Centered 5</td>
<td>17.6%</td>
<td>16.3 – 19.0</td>
<td>144.71***</td>
<td>27</td>
<td>6,281</td>
</tr>
</tbody>
</table>

| **Ten Years** |       |         |        |     |    |
| $B_0$    | 0.29  | 0.24 – 0.34 | 23.92  | 17  | 2,628 |
| $B_0$ Centered 0 | 7.1%  | 5.7 – 8.8 | 42.74*** | 17  | 2,628 |
| $B_0$ Centered 2 | 11.8% | 10.3 – 13.4 | 56.10*** | 17  | 2,628 |
| $B_0$ Centered 5 | 24.6% | 22.4 – 27.0 | 57.73*** | 17  | 2,628 |

Variability Across Studies: Static-99

Trying to explain variability across samples….

Moderator Analyses using Static-99

What Factors Might Affect Absolute Recidivism Estimates?

<table>
<thead>
<tr>
<th>Methodological Factors</th>
<th>Individual-level Factors</th>
<th>Systems-level Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of recidivism sources</td>
<td>Treatment</td>
<td>Country</td>
</tr>
<tr>
<td>Street time</td>
<td>Dynamic risk factors</td>
<td>Sample type</td>
</tr>
<tr>
<td>Recidivism definition</td>
<td>Age at release</td>
<td>Time period</td>
</tr>
<tr>
<td>Length of follow-up</td>
<td>Race</td>
<td>Detection rates</td>
</tr>
<tr>
<td>Quality of assessment</td>
<td>Rapist vs. child molester</td>
<td>Correctional philosophy</td>
</tr>
<tr>
<td>Use of national criminal records</td>
<td></td>
<td>Community supervision</td>
</tr>
</tbody>
</table>
Methodological Factors: Non-significant

<table>
<thead>
<tr>
<th></th>
<th>Q due to moderator</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used 2+ recid sources</td>
<td>1.84</td>
<td>1</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>Used national records</td>
<td>0.43</td>
<td>1</td>
<td>&gt;.50</td>
</tr>
<tr>
<td>Cited coding rules</td>
<td>1.14</td>
<td>1</td>
<td>&gt;.25</td>
</tr>
</tbody>
</table>

Methodological Factors: Closer look

- Use of street time
  - Significantly higher recidivism rates when street time used ($Q = 14.74, df = 1, p < .001$)
  - More than 80% of cases from one unusually high risk sample (Bridgewater; Knight & Thornton, 2007)
  - Cox regression with larger sample of non-Bridgewater cases using street time
    - No effect ($x^2$ change = 1.8, df = 1, $p = .179$)

Methodological Factors: Closer look

- Recidivism Criteria (charges vs convictions)
  - Significant ($Q = 16.51, df = 1, p < .001$), but interacted with Static-99 scores ($x^2$ change = 10.5, $df = 1, p < .001$)
  - Pattern of results not logical

Sexual Recidivism (%)

5 year Fixed follow-up

![Sexual Recidivism Graph]

- Charges vs convictions
- Better test: compare charges vs convictions within same study
- 5 studies available ($n = 1,318$)
- 181 charged; 159 subsequently convicted
- Rate ratio of 1.14
- Insufficient to explain variability in base rates

Methodological Factors: Closer look

- Recidivism Criteria (charges vs convictions)
- Better test: compare charges vs convictions within same study
  - 5 studies available ($n = 1,318$)
  - 181 charged; 159 subsequently convicted
- Rate ratio of 1.14
- Insufficient to explain variability in base rates

Treatment & Race: Non-significant

<table>
<thead>
<tr>
<th></th>
<th>Q due to moderator</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Started treatment</td>
<td>0.64</td>
<td>1</td>
<td>&gt;.25</td>
</tr>
<tr>
<td>Completed treatment</td>
<td>2.76</td>
<td>1</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Non White</td>
<td>2.40</td>
<td>1</td>
<td>&gt;.10</td>
</tr>
</tbody>
</table>
**Rapists vs Child Molesters**

- Confusing!
- **Static-2002**
  - Significant ($\chi^2$ change = 4.5, df = 1, $p < .05$), with higher recidivism among *rapists*
- **Static-99**
  - Non-significant in Cox regression ($\chi^2$ change = 0.1, df = 1, $p = .800$)
  - Significant in meta-analysis ($Q = 5.05$, df = 1, $p < .05$), with higher recidivism among *child molesters*
- No effect that is consistent and large enough to be of substantive value

**Year of Release**

- Some pattern discernible, with and without controlling for sample type
- Insufficient evidence to justify including it in recidivism estimates

**Country**

- Predicted recidivism rates for Static-99 score of 2 ($Q = 27.39$, df = 4, $p < .001$)
  - United States: 8.9% ($k = 5$, $n = 1516$)
  - Canada: 6.8% ($k = 11$, $n = 1,793$)
  - United Kingdom: 5.4% ($k = 3$, $n = 491$)
  - Europe: 3.8% with outlier removed ($k = 5$, $n = 1,697$)
- Europe significantly lower than US and Canada
Country

- Effects largely disappear when control for sample type (to be discussed)
  - Logistic regression with country, age at release, sample type
    - 5 years: Canada significantly lower than US
    - 10 years: Country not significant
  - Meta-regression: after sample type entered, country does not contribute to prediction of base rates (Q = 6.98, df = 3, p > .05)

What moderators are we left with?

- Age at release
- Sample type

Age at Release

- Rate ratio of .98 (95% C.I. of .98 to .99)
  - Expected recidivism rate of 32-year old offenders is 98% of the recidivism rate of 31-year old offenders, which is 98% of 30-year old offenders, etc.…
  - Tested with Cox regression using sample as strata (x2 change = 28.7, df = 1, p < .001)
- Non-linear (adding age², x2 change = 10.7, df = 1, p = .001)
- Adding age³ non-significant

Age at release: Static-99

Static-99 without the age item

Developing new age item

- Cases with age at release info and age-free Static-99 scores (k = 23, n = 8,128)
  - Development sample (k = 23, n = 5,736) – all cases with < 10 years follow-up
  - Validation sample (k = 15, n = 2,392) – all cases with 10+ years follow-up
Selecting new age weights

- Principles guiding selection
  - Similar odds ratio for Static-99 (one-point Static increase associated with ~1.35 increase in odds of recidivism)
  - Median age (39) get score of 0
  - Higher predictive accuracy than original Static
  - Age should no longer contribute

Selecting new age weights: Process

Result: Virtually the same weights

New age item

<table>
<thead>
<tr>
<th>Age at release</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-34.999</td>
<td>+1</td>
</tr>
<tr>
<td>35-39.999</td>
<td>0</td>
</tr>
<tr>
<td>40-59.999</td>
<td>-1</td>
</tr>
<tr>
<td>60+</td>
<td>-3</td>
</tr>
</tbody>
</table>

Comparing Static-99 to Static-99R: Validation sample (n = 2,392)

<table>
<thead>
<tr>
<th></th>
<th>ROC 5 years</th>
<th>ROC 10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static-99</td>
<td>.713</td>
<td>.706</td>
</tr>
<tr>
<td>Static-99R</td>
<td>.720</td>
<td>.710</td>
</tr>
</tbody>
</table>

Static-99R

- Revised version of Static-99
  - Original age item removed
  - New age item added

- Total scores range from -3 to 12
### Age fully accounted for in Static-99R

<table>
<thead>
<tr>
<th>Static-99</th>
<th>Static-99R</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2 Δ</td>
<td>p</td>
</tr>
<tr>
<td>Log. Reg. 5yr</td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>128.92</td>
</tr>
<tr>
<td>Age</td>
<td>4.14</td>
</tr>
<tr>
<td>Log. Reg. 10yr</td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>157.91</td>
</tr>
<tr>
<td>Age</td>
<td>5.54</td>
</tr>
<tr>
<td>Cox Regression</td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>165.92</td>
</tr>
<tr>
<td>Age</td>
<td>7.87</td>
</tr>
</tbody>
</table>

### Static-99R: Statistical Shrinkage?

<table>
<thead>
<tr>
<th></th>
<th>ROC</th>
<th>Log. Reg. Odds Ratio</th>
<th>Cox Reg. Rate Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>.709</td>
<td>1.32</td>
<td>1.36</td>
</tr>
<tr>
<td>Validation</td>
<td>.720</td>
<td>1.34</td>
<td>1.28</td>
</tr>
</tbody>
</table>

### Static-99R for rapists and child molesters

<table>
<thead>
<tr>
<th></th>
<th>Rapists</th>
<th>Child Molesters</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>p</td>
<td>b</td>
</tr>
<tr>
<td>Log. Reg. 5yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static-99R</td>
<td>.317</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Age</td>
<td>-.006</td>
<td>.583</td>
</tr>
<tr>
<td>Log. Reg. 10yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static-99R</td>
<td>.325</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Age</td>
<td>.001</td>
<td>.966</td>
</tr>
<tr>
<td>Cox Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static-99R</td>
<td>.281</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Age</td>
<td>.002</td>
<td>.778</td>
</tr>
</tbody>
</table>

### What about Static-2002 and age?

<table>
<thead>
<tr>
<th>Age at release</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24.999</td>
<td>3</td>
</tr>
<tr>
<td>25-34.999</td>
<td>2</td>
</tr>
<tr>
<td>35-49.999</td>
<td>1</td>
</tr>
<tr>
<td>50+</td>
<td>0</td>
</tr>
</tbody>
</table>

### New Static-99 age item

<table>
<thead>
<tr>
<th>Age at release</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-34.999</td>
<td>+1</td>
</tr>
<tr>
<td>35-39.999</td>
<td>0</td>
</tr>
<tr>
<td>40-59.999</td>
<td>-1</td>
</tr>
<tr>
<td>60+</td>
<td>-3</td>
</tr>
</tbody>
</table>
Static-2002 and Age at release

• Static-2002 much better at accounting for age than Static-99
  – BUT, non-linear effect still significant
• Like Static-99, if we drop the age item from Static-2002, age shows a significant LINEAR effect
• Tested the same age item used in Static-99R
  – No reason to expect age effects would be different
  – Larger sample sizes for Static-99 analyses
  • All Static-2002 datasets included in Static-99 analyses

Age fully accounted for in Static-2002R

<table>
<thead>
<tr>
<th></th>
<th>Static-2002</th>
<th>Static-2002R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log. Reg. 5yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>86.56</td>
<td>91.00</td>
</tr>
<tr>
<td>Age</td>
<td>3.40</td>
<td>0.18</td>
</tr>
<tr>
<td>Age²</td>
<td>2.65</td>
<td>0.33</td>
</tr>
<tr>
<td>Log. Reg. 10yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>61.34</td>
<td>62.54</td>
</tr>
<tr>
<td>Age</td>
<td>0.31</td>
<td>1.80</td>
</tr>
<tr>
<td>Age²</td>
<td>4.69</td>
<td>1.56</td>
</tr>
<tr>
<td>Cox Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>129.78</td>
<td>131.99</td>
</tr>
<tr>
<td>Age</td>
<td>1.90</td>
<td>1.90</td>
</tr>
<tr>
<td>Age²</td>
<td>6.14</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Static-99R and Static-2002R

• Neither Static-99 nor Static-2002 fully accounted for age at release
• New age item created. Same item replaces age items in both Static-99 and Static-2002
• Using either Static-99R or Static-2002R, no further age adjustments would improve prediction

Nominal risk categories

• Static-99
  – 0-1: Low
  – 2-3: Moderate-Low
  – 4-5: Moderate-High
  – 6+: High
• Static-2002
  – 0-2: Low
  – 3-4: Moderate-Low
  – 5-6: Moderate
  – 7-8: Moderate-High
  – 9+: High
• Static-99R
  – -3 to 1: Low
  – 2-3: Moderate-Low
  – 4-5: Moderate-High
  – 6+: High
• Static-2002R
  – -3 to 2: Low
  – 3-4: Moderate-Low
  – 5-6: Moderate
  – 7-8: Moderate-High
  – 9+: High

Static-99R and Static-2002R Nominal risk categories

• Compared original and R versions
  – Proportion of offenders in each category (similar)
  – Recidivism rates per category (same or better)
• Same categories retained
  – Negative scores join lowest risk group

Distribution of Sex Offenders by Risk Category
**Routine Correctional Samples**

- Research ideal
- Large, unselected samples of sex offenders
  - Representative of general population of adjudicated sex offenders
- Does not describe most research studies
- May not describe the offender sitting in front of you
  - Possible he was sent to you because he is NOT representative of typical offenders
- How does this routine/non-routine distinction affect the data?
  - And what do I do with it?

---

**Non-Routine Samples**

- Preselected in some way
  - From a particular treatment setting
  - Referred to a particular setting for assessment/treatment
    - e.g. psych assessment
  - From a particular institution
    - e.g., max security
  - By some kind of condition
    - e.g., indefinite sentence, detained until warrant expiry, other special measures
- Do offenders preselected in some way vary in their recidivism rates from random, unselected samples (e.g., routine)?
Preselection

- Preselection processes
  - Likely consider factors already included in Static-99
  - Possibly consider factors unrelated to risk (e.g., offence severity, treatment availability, publicity for a case)
  - Likely consider risk factors external to Static-99 (e.g., treatment need, institutional behaviour)

- Do offenders preselected in some way vary in their recidivism rates from random, unselected samples (e.g., routine)?

**Routine samples vs all others (5 years)**

<table>
<thead>
<tr>
<th></th>
<th>B(95)</th>
<th>95% C.I.</th>
<th>Q</th>
<th>k</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>7.4%</td>
<td>5.3 – 10.2</td>
<td>154.82***</td>
<td>23</td>
<td>5,760</td>
</tr>
<tr>
<td>Routine</td>
<td>5.0%</td>
<td>3.2 – 7.8</td>
<td>19.57**</td>
<td>8</td>
<td>2,406</td>
</tr>
<tr>
<td>Non-Routine</td>
<td>9.1%</td>
<td>6.3 – 13.0</td>
<td>92.04***</td>
<td>15</td>
<td>3,354</td>
</tr>
</tbody>
</table>

Q due to Routine variable 43.21***

**5 year sexual recidivism: Static-99R**

![Graph showing 5 year sexual recidivism: Static-99R](image)

Can we do better than this?
Categorizing the non-routine samples

**2 Categories of Preselection**

1) Preselected based on treatment need
   - Through some formal or informal process, offenders judged as having treatment needs in need of intervention

2) Preselected as high risk/need
   - Offenders considered for rare (infrequent) measure/intervention/sanction reserved for highest risk cases
     - Detention until Warrant Expiry (in Canada)
     - Indefinite detention (civil commitment, Dangerous Offender, indefinite treatment order)
     - High-intensity treatment (if given to small subset and assigned for high risk/need
       - Civil commitment (U.S.), Regional Treatment Centres (Canada)
       - Does not include typical, moderate intensity treatment programs (or one-size-fits-all programs)
     - Offenders sent for specialized psychiatric services
       - E.g., Penetanguishene

**2 Categories of Preselection**

2) Preselected as high risk/need
   - Offenders considered for rare (infrequent) measure/intervention/sanction reserved for highest risk cases
     - Detention until Warrant Expiry (in Canada)
     - Indefinite detention (civil commitment, Dangerous Offender, indefinite treatment order)
     - High-intensity treatment (if given to small subset and assigned for high risk/need
       - Civil commitment (U.S.), Regional Treatment Centres (Canada)
       - Does not include typical, moderate intensity treatment programs (or one-size-fits-all programs)
     - Offenders sent for specialized psychiatric services
       - E.g., Penetanguishene
Non-Routine Samples

- Treatment Need
  - Allan et al. (2007)
  - Brouillette-Alarie & Proulx (2008)
  - Harkins & Beech (2007)
  - Johansen (2007)
  - Swinburne Romine et al. (2008)
  - Ternowski (2004)

- High Risk/Need
  - Bengtson (2008)
  - Bonta & Yessine (2005)
  - Haag (2005)
  - Knight & Thornton (2007)
  - Nicholaichuk (2001)
  - Wilson et al. (2007a,b)

What samples are gone?

- Don’t fit the 2 preselected groups
  - Cortoni & Nunes (2007)
  - Hill et al. (2008)
  - Saum (2007)

- No age info for Static-99R scores
  - Craig et al. (2006)
  - De Vogel et al. (2004)
  - Endrass et al. (2008)

Sample Type (5 years)

<table>
<thead>
<tr>
<th></th>
<th>$B_\text{Q}(2)$</th>
<th>95% C.I.</th>
<th>Q</th>
<th>k</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>7.4%</td>
<td>6.6 – 8.3</td>
<td>62.83***</td>
<td>20</td>
<td>5,501</td>
</tr>
<tr>
<td>Routine</td>
<td>5.0%</td>
<td>3.2 – 7.8</td>
<td>19.57**</td>
<td>8</td>
<td>2,406</td>
</tr>
<tr>
<td>Treatment Need</td>
<td>7.2%</td>
<td>6.0 – 8.8</td>
<td>4.47</td>
<td>6</td>
<td>1,782</td>
</tr>
<tr>
<td>High Risk/Need</td>
<td>12.2%</td>
<td>9.9 – 15.0</td>
<td>4.01</td>
<td>6</td>
<td>1,313</td>
</tr>
</tbody>
</table>

Q due to Routine variable 34.78***

5 year sexual recidivism: Static-99R

10 year sexual recidivism: Static-99R
Static-2002 Recidivism Estimates

<table>
<thead>
<tr>
<th></th>
<th>5 years</th>
<th>10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine</td>
<td>$k = 3, n = 526$</td>
<td></td>
</tr>
<tr>
<td>High Risk/Need</td>
<td>$k = 3, n = 931$</td>
<td>$k = 2, n = 642$</td>
</tr>
<tr>
<td>Non-Routine</td>
<td>$K = 4, n = 1,121$</td>
<td>$K = 3, n = 766$</td>
</tr>
</tbody>
</table>

Notes. Only 1 treatment need sample (insufficient for separate estimates)
Non-Routine group includes all cases in high risk/need group, plus the 1 treatment need sample

5 year sexual recidivism: Static-2002R

10 year sexual recidivism: Static-2002R

Summary

- Static-99 and Static-2002 provide consistent measures of relative risk
- Incremental effect of age
  - Static-99R; Static-2002R
- Variability in Base Rates
  - Routine/Non-Routine
  - Treatment Needs
  - High Risk/Need

What’s an evaluator to do?

- Focus on relative risk
  - Percentiles
  - Risk Ratios
- Any statements about absolute risk requires justification

Option #1: Ideal, but not often possible

- Use local norms
  - Recidivism studies
  - These can be estimated from the distribution of Static-99 or Static-2002 and overall recidivism rate (assuming B1 and distribution to be constant)
Option #2: Routine norms

- The estimates from routine samples are the default position
- Representative of general population of adjudicated sex offenders
- This option is sufficient in most circumstances

Option #3: Justify that routine norms do not apply

- Possible justifications
  - Sufficient criminogenic needs to recommend treatment: use treatment need norms
  - Member of small minority selected on risk/need factors external to Static-99R/Static-2002R: use high risk/need norms
  - Sufficient evidence that offender is non-routine, but insufficient information to differentiate between treatment need or high risk/need: use non-routine norms