

USING MULTIPLE ACTUARIAL INSTRUMENTS TO ASSESS SEXUAL RECIDIVISM RISK

DAVID THORNTON, PH.D..

JUNE 30TH 2022



SAARNA CURRENTLY SUPPORTS THREE STATIC ACTUARIAL INSTRUMENTS RELEVANT TO SEXUAL RECIDIVISM

- Static-99R
- Static-2002R
- Risk Matrix 2000/s

I AM GOING TO LOOK AT...

- How do the instruments compare?
- When should you pick one and just use that?
- What are the advantages of using multiple instruments?
- If you use multiple instruments how do you integrate information from them to form an overall estimate of static risk?
- Additional considerations when combining static actuarial scales with dynamic scales

ARE THEY INCREMENTAL TO EACH OTHER?

- We don't have large amounts of data on this for all the scales. Many samples are only scored for one scale and then items on other scales approximated
- Babchishin et al. (2012) reanalyzed available data from the STATIC normative samples and found that Static-99R and Static-2002R were each incremental to the other in Cox Regression
 - $P < .05$ for Static-2002R
 - $P < .01$ for Static-99R

WHAT ABOUT RISK MATRIX 2000/S

- In the STATIC normative data bases Risk Matrix 2000 can only be approximated because of differences in how the victim items are scored

WHAT ABOUT RISK MATRIX 2000/S

- Dynamic Supervision Project (Canadian data set)
- $N = 710$; most had 5-year follow-up
- Risk Matrix approximated from Static-99R items

COX REGRESSION

In this sample

- Static-99R incremental to RM2000/S
- RM2000/S had no predictive value after Static-99R controlled

AUCs for 5-year sex recidivism

- RM2000/S = .69
- Static-99R = .73

	b	P	Exp(B)
Static-99R	.331	<.001	1.393
RM2000/S	.014	.948	1.014

WHAT ABOUT WHEN RISK MATRIX 2000 IS SCORED CORRECTLY?

- New Austrian data set where Static-99R and Risk Matrix 2000 were both scored from the same files using their own instructions
- N= 332; almost all cases had 5 year follow up; many had up to 10 years

COX REGRESSION

In this sample

- RM2000/S incremental to Static-99R
- Static-99R had no predictive value after RM2000/S controlled

AUCs for 5-year sex recidivism

- RM2000/S = 0.75
- Static-99R = 0.71

	b	P	Exp(B)
Static-99R	-0.070	0.467	0.932
RM2000/S	+0.719	0.004	2.053

COMMENT

- How incremental scales are to each other is likely to vary from sample to sample
- We should not assume that the scales are redundant with each other, but we don't have enough data to empirically estimate a formula for combining them
- There may not be an optimum formula as how predictive different risk indicators are varies across samples (Helmus & Thornton, 2015)

HOW BIG A DEAL IS ALL THIS?

- Why do the scales sometimes give different risk estimates and how big can the differences be?
- If risk estimates from different scales can only be small then we don't need to worry about this. You could pick whichever scale you like and ignore the others
- If results from the scales can be very different then we need a strategy for handling this
- Let's look at this

HOW DIFFERENT ARE STATIC-99R, STATIC-2002R, AND RISK MATRIX 2000/S?

SIMILARITIES

- All use easily available static items
- All assess three underlying factors: sexual criminality, general criminality, and age

DIFFERENCES

- Weight given to uncertain “facts”
 - Unproven charges
 - Prison misconducts
 - Multiple charges for the same acts
- Relative weight given to the underlying factors
- Different definitions (e.g., non-sexual violence)
- Some variation in item content (e.g., young/unrelated victims)

HOW DOES THIS WORK OUT IN PRACTICE?

EXAMPLE 1

- Age 17 sentenced for stealing a car and joy riding
- Age 19 sentenced for possession of drugs, placed on probation
- Age 20 sentenced for street robbery and violation of probation
- Age 27 sentenced for three rapes of female strangers
- In prison conduct reports for masturbating in front of staff
- Now aged 37, Never married or in a live-in relationship

SEXUAL RECIDIVISM IN ROUTINE SAMPLES

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	31.0%	50.3%	99.1
Static-2002R	13.8%	25.1%	78.0
Risk Matrix 2000/S	10%	18.5%	51.2

EXAMPLE 2

- Age 16 sentenced for 4th degree sexual assault (known female) → Probation completed successfully
- Age 21 sentenced for exposing himself to female stranger – 12 months custody, 2 years supervision (which he completed)
- Age 25 sentenced for a stranger rape → 7 years prison
- Now aged 32, Never married or had a live in relationship

SEXUAL RECIDIVISM IN ROUTINE SAMPLES

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	23.7%	40.2%	97.2
Static-2002R	43.7%	67.8%	98.3
Risk Matrix 2000/S	40.0%	63.4%	96.5

EXAMPLE 3

- Age 16 sentenced for 4th degree sexual assault (known female) → 2 yrs Probation
- Age 17 Assault → Probation continued
- Age 21 Sentenced for 4th Degree Sexual Assault (known female)– 3 years custody, 5 years supervision
- Age 25 sentenced for 1st Degree Sexual Assault (Known female) → 12 years prison
- Now aged 37, Never married or had a live in relationship

SEXUAL RECIDIVISM IN ROUTINE SAMPLES

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	12.8%	23.4%	88.7
Static-2002R	34.3%	56.1%	95.9
Risk Matrix 2000/S	10.0%	18.5%	51.2

SUMMARY OF EXAMPLES

- Large difference between highest risk estimate and lowest risk estimate (around 30 points)
- For each case there is one scale which yields a risk estimate that exceeds 50% while other scales give a quite different picture
- **This means it is unsafe to score a single static actuarial instrument and assume that that represents the level of risk implied by static risk factors**

OPTIONS: FIRST DECISION – ONE VS MANY?

- Just score one (but then you don't know whether the others would give quite different results)
- Score all (but then how do you make sense of cases where the three scales give different results)
- Worth noting that if you have gathered the information to score Static-99R, laid out a criminal history table etc., then scoring Risk Matrix 2000 takes about 5 minutes and Static-2002R about another 10 minutes

SECOND DECISION: HOW INTEGRATE?

- Give a range of results
 - Works if all results have the same practical implications
- Pick within the range
 - Pick highest risk; Pick lowest risk estimate; average risk estimates
- Place within the range based on credibility of scale

THERE ARE SOME GROUNDS FOR AVERAGING

- Lehmann et al. (2013), for example, found that this worked better than other algorithms like choose the highest or choose the lowest. But they did not average recidivism estimates; their method equated base rates for all the scales and just averaged relative risk
- A risk estimate from a scale reflects
 - the base rate in the scale's norms
 - the relative risk assigned by the scale
 - averaging recidivism estimates averages both these components
- The main argument in favor of averaging is that it weights scales equally without requiring you to make difficult judgements as to which scale is better
- Also consistent with general psychological measurement; each estimate has error and can be integrated (averaging sensible in absence of evidence to privilege one measure)
- **Averaging is more objective**

BUT AN UNWEIGHTED MEAN COULD BE MISLEADING IF

- Scales differed substantially in how well they measured relative risk
- Scales differed substantially in how well they estimated the sexual recidivism base rate
- We need some principles for averaging which will warn us when averaging is likely to work less well and suggest what to do in that circumstances

SOME PRINCIPLES FOR COMBINING SCALES

- Don't include results from scales that are clearly biased
 - E.G., scale's risk estimates are for violent recidivism rather than sexual recidivism
 - E.G., scale does not take into account the protective effect of older age and the person being assessed is in their 60s
 - E.G., evidence or scale authors no longer support the scale (e.g., the RRASOR)
- Averaging makes most sense when the scales provide similar discrimination and their norms have similar base rates
- If base rates or discrimination differ across scales then consider a range and give more weight to scale with better discrimination or better estimated base rates

CONSIDERING RANGE

- Best estimate lies within the range, lies closer to the result from the scale that discriminates better or has better estimated base rates
- But don't disregard results from the other scales

SOMETIMES ONE SCALE MAY BE MORE CREDIBLE THAN THE OTHERS

- For example, Static-99R has been validated for individuals with MMI or ID while the other scales have not
- In this case consider the available validation evidence for these types of individuals when weighting scales

The slide features a teal-to-blue gradient background. In the corners, there are decorative white line-art elements resembling circuit traces or data paths, with small circles at the end of the lines.

WITH THAT FRAMEWORK, LET'S LOOK AT THE PROPERTIES
OF THE DIFFERENT SCALES AND THEIR NORMS

DISCRIMINATION IS SIMILAR FOR ALL THREE SCALES

- Static-99R
 - AUC =.69 based on 56 samples with a cumulative N of 71,515 (Helmus et al., 2022)
- Static-2002R
 - AUC =.69 based on 7 samples and a cumulative N of 2,609 (Babchishin et al., 2012)
- Risk Matrix 2000/S
 - $d = 0.74$ based on 15 samples and a cumulative N of 10,644 (Helmus et al., 2013)
 - This is equivalent to AUC= .70

NORMS FOR 5-YEAR SEX RECIDIVISM: SIZE

	Routine N	HRHN N
Static-99R	7,244 (483 recidivists)	860 (164 recidivists)
Static-2002R	1,964 (217 recidivists)	497 (97 recidivists)
Risk Matrix 2000	2,973 (386 recidivists)	-

COMPARING STATIC-99R VS. STATIC-2002R HRHN NORMS

STATIC-99R

- N=860 (5 yr; 164 recidivists), 350 (10 yr; 98 recidivists)
- 5 samples
- 1 Danish and 4 Canadian
- All use national data

STATIC-2002R

- N=497 (5 yr; 97 recidivists)
- 2 samples
- 2 Canadian
- All use national data

WHAT ABOUT AFTER THEY HAVE BEEN COMBINED WITH DYNAMIC INSTRUMENTS?

- STABLE-2007 has combination tables for all three static instruments

STABLE-2007 Scores

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
-2	I	I	I	I	I	I	I	I	II	II	II	II	II	III	III	III	III	III	III	III	III
-1	I	I	I	I	I	I	I	I	II	II	II	II	II	III	III	III	III	III	III	III	III
0	I	I	II	II	II	II	II	II	II	II	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa
1	I	I	II	II	II	II	II	II	II	II	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa
2	II	II	II	III	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb
3	II	II	II	III	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb
4	II	II	II	III	III	III	III	III	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb
5	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb
6	III	III	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb
7	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
8	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
9	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
10	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb
11+	III	III	III	IVa	IVa	IVa	IVa	IVa	IVa	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb	IVb

Static-2002R Scores

RISK MATRIX & STABLE

- The table provided in the workbook uses old Risk Matrix category labels but you can convert them
 - Low = Below Average
 - Moderate = Average
 - High = Above Average
 - Very High = Well Above Average
- And uses different labels for the combined categories

STABLE-2007 Scores

Risk Matrix 2000.S/V/C Risk Levels

	Low	Moderate	High	Very High
0	Low	Low	Moderate	Moderate-High
1	Low	Low	Moderate	Moderate-High
2	Low	Low	Moderate	Moderate-High
3	Low	Low	Moderate	Moderate-High
4	Low	Moderate	Moderate-High	High
5	Low	Moderate	Moderate-High	High
6	Low	Moderate	Moderate-High	High
7	Low	Moderate	Moderate-High	High
8	Low	Moderate	Moderate-High	High
9	Low	Moderate	Moderate-High	High
10	Low	Moderate	Moderate-High	High
11	Low	Moderate	Moderate-High	High
12	Moderate	Moderate-High	High	Very High
13	Moderate	Moderate-High	High	Very High
14	Moderate	Moderate-High	High	Very High
15	Moderate	Moderate-High	High	Very High
16	Moderate	Moderate-High	High	Very High
17+	Moderate	Moderate-High	High	Very High

COMMENT ON STABLE INTEGRATIONS

- Averaging seems a good fit for combining results of the three scales when they have been integrated with STABLE

RECOMMENDATIONS #1

- Score at least two static instruments, preferably three
- The two most robustly empirically supported are
 - Static-99R
 - Risk Matrix 2000
- Adding Risk Matrix 2000 to your assessment will take only a few minute
- Preferably combine with STABLE-2007

RECOMMENDATIONS #2

- Present both the average sexual recidivism estimate and the range
- In most circumstances the average is the more comprehensive, more objective, way of estimating risk
- Consider the range of results as informing you about the potential margin of error

EXAMPLE 1: SEXUAL RECIDIVISM

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	31.0%	50.3%	99.1
Static-2002R	13.8%	25.1%	78.0
Risk Matrix 2000/S	10%	18.5%	51.2
Average of all Three	18.3%	31.3%	
Average of 99R & RM2000/s	20.5%	34.4%	

EXAMPLE 2: SEXUAL RECIDIVISM

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	23.7%	40.2%	97.2
Static-2002R	43.7%	67.8%	98.3
Risk Matrix 2000/S	40.0%	63.4%	96.5
Average of all Three	35.8%	57.1%	
Average of 99R & RM2000/s	31.9%	51.8%	

EXAMPLE 3: SEXUAL RECIDIVISM

Scale	5-Year Sex Recidivism	20-Year Projected Sexual Recidivism	Percentile
Static-99R	12.8%	23.4%	88.7
Static-2002R	34.3%	56.1%	95.9
Risk Matrix 2000/S	10.0%	18.5%	51.2
Average of all Three	19.0%	32.7%	
Average of 99R & RM2000/s	11.4%	21.0%	

TRAINING EVENTS FOR ALL THREE SCALES ARE AVAILABLE THROUGH SAARNA.ORG

- The next one is Risk Matrix 2000 training (two half days)
- July 18 & 19, 2022, 9am-1pm Eastern time
- \$350 Canadian (about \$272 US)
- \$280 Canadian for SAARNA affiliates (about \$217 USA)
- 16 seats currently available

<https://saarna.org/product/risk-matrix-2000-training>



LEVELS OF ACCESS

- **Lurker (no need to sign up)**

- Almost everything on old Static website is free on saarna.org, and more
- All resources needed to score Static-99R/2002R and RM2000
- Research papers we can legally post on website

- **Scale user or Affiliate (\$25/year CAD)**

- Warm fuzzy of supporting our mission, additional papers/presentations, newsletters, discussion/policy papers, scoring exercises, training discounts

- **Trainer (\$75/year CAD)**

- Everything above, sample training materials and exams, some free webinars, confirmation of certification, templates (e.g., certification for attendees)