

Examining the Predictive Validity of the SORAG in Switzerland

Astrid Rossegger^{1,2}, Juliane Gerth^{1,2}, Jay P. Singh^{1,3}, Jérôme Endrass^{1,2}

¹Psychiatric/Psychological Service, Department of Justice, Canton of Zurich

²Department of Psychology, University of Konstanz

³Faculty of Health Sciences, Molde University College

[Sexual Offender Treatment, Volume 8 (2013), Issue 2]

Abstract

Background: The SORAG is one of the most commonly used actuarial risk assessment tools for sex offender evaluations. Although many studies have investigated the predictive validity of this instrument, few have examined whether its published expected rates of recidivism are useful in practice. The aim of the present investigation was to investigate predictive validity of the SORAG in Switzerland, considering both the discrimination and calibration components of predictive validity.

Material and Methods: The instrument was administered to two total cohorts of offenders (N = 137) in the Canton of Zurich, Switzerland that were followed for a fixed period of seven years after discharge. Recidivism was defined as new charges and/or convictions for violent (including sexual) offenses. Discrimination was measured using receiver operating characteristic (ROC) curve analysis and calibration using Sanders' decomposition of the Brier score. A dependent t-test was used to examine the difference in expected and observed percentiles, and a Kolmogorov-Smirnov Test was conducted to compare the distribution of offenders in risk bins.

Results: ROC analyses revealed an acceptable level of discrimination for both SORAG total risk scores (AUC = 0.69, 95% CI = 0.56-0.82) and risk bins (AUC = 0.67, 95% CI = 0.54-0.80). The seven year recidivism rates both overall as well as for each risk bin were considerably lower than the published SORAG norms. An average forecast error of 20.9% for each risk bin of the SORAG suggests a large difference between expected and observed recidivism rates.

Conclusion: Should further studies in Switzerland replicate the current findings, a re-calibration of the SORAG may be needed before the instrument can be considered a valid method to assess recidivism risk in Swiss sex offender populations.

Keywords: SORAG, sex offender, discrimination, calibration, distribution

Introduction

There is strong empirical evidence suggesting that evaluations conducted using mechanical instruments result in superior assessments of risk compared to clinical judgments (Ægisdóttir et al., 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Quinsey, Harris, Rice, & Cormier, 2006). These mechanical (or "actuarial") risk assessment instruments are composed of closed-ended questions with pre-determined responses. These responses are combined using objective, transparent, and inflexible rules proscribed by the tool authors, resulting in either a total risk score or risk classification.

Over the past 30 years, many actuarial risk assessment tools have been developed for the specific purpose of assessing the likelihood of recidivism in criminal offenders. Among these are a number of schemes designed for assessing the risk of sex offender recidivism. According to recent surveys (Jackson & Hess, 2007; Viljoen, McLachlan, & Vincent, 2010), the Sex Offender Risk Appraisal

Guide (SORAG; Quinsey et al., 2006) is one of the most popular of the current crop of sex offender recidivism risk assessment instruments. The SORAG is a Canadian tool composed of 14 items with a statistical association with new charges and/or convictions for violent (including sexual) offenses. The SORAG produces a total risk score that can be used to classify sex offenders into one of nine risk categories (i.e., "bins"), each of which has affiliated expected rate of recidivism for 7 and 10 years follow-up. In addition, total risk scores can be used to compare a given offender to the population of sex offenders using published percentiles.

There is a considerable empirical base concerning the ability of the SORAG to discriminate between recidivists and non-recidivists which includes both North American (AUC = 0.64-0.78) and European (AUC = 0.71-0.75) investigations (Bartosh, Garby, Lewis, & Gray, 2003); (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012); (Pham & Ducro, 2008); (Rice & Harris, 2002). However, study authors routinely discount the guidance of the SORAG's authors concerning how the instrument should be administered and its findings interpreted (Rossegger et al., 2013; Singh, Desmarais, & Van Dorn, 2013). Specifically, previous studies have not examined statistically whether the predicted recidivism rates for each risk bin are an accurate reflection of observed rates. And only four studies have descriptively reported the number of recidivists per risk bin (Table 1). In addition, the accuracy of the percentile ranks offered in the SORAG manual has not been replicated. Thus, further research is necessary to clarify the usefulness of the SORAG for practitioners, especially those working in jurisdictions other than Canada, where expected recidivism rates for risk bins and score distributions may be different.

	Quinsey et al. (2006) ^a	Nunes et al. (2002)	Harris et al. (2003)	Looman (2006)	Eher et al. (2008b), (2008a) ^b ; subgroup of rapists	Eher et al. (2008b), (2008a) ^b ; subgroup of child molesters
Replication match ^c		7	7	6	3	3
No item approximations		No	Yes	Yes	No	No
No systematic item omission		No	Yes	Yes	Yes	Yes
Reliable scoring	Yes	No	Yes	Yes	No	No
Controlling for attrition	Yes	No	No	No	No	No
File information used for scoring	Yes	Yes	Yes	Yes	Yes	Yes
LoFU (years)	9.9 (Mean)	7.3 (Mean)	5.2 (Mean)	4.6 (Mean)	3.6 (Mean)	3.6 (Mean)
Legal status of recidivism	Charge + conviction	Charge + conviction	Charge + conviction	Conviction	Conviction	Conviction
Risk bin	Recidivism Rate (n)					

1	7% (14)	8% (87)	19% (NR)	0% (6)	0% (9)	5% (22)
2	15% (23)	8% (52)	18% (NR)	17% (6)	0% (13)	3% (33)
3	23% (40)	15% (46)	29% (NR)	11% (19)	15% (20)	0% (22)
4	39% (58)	30% (37)	50% (NR)	13% (28)	25% (16)	7% (15)
5	45% (52)	39% (18)	55% (NR)	32% (38)	33% (15)	15% (13)
6	58% (46)	0% (12)	63% (NR)	36% (42)	6% (15)	9% (11)
7	58% (32)	25% (4)	63% (NR)	33% (33)	19% (16)	33% (9)
8	75% (18)	50% (2)	71% (NR)	57% (44)	45% (16)	40% (5)
9	100% (5)	0% (0)	76% (NR)	57% (26)	43% (11)	0% (0)
Recidivism rate ^d (N)	42.0% (288)	15% (258)	48% (205)	34% (242)	24% (123)	6% (130)
Mean total risk score (SD)	8.9 (SD =11.3)	-3.2 (SD =10.3)	10.0 (SD =10.8)	16.6 (SD =12.3)	NR	NR

Note. LoFU = length of follow-up; N = size of total study sample; NR = not reported.

^a SORAG development sample.

^b This publication is the corresponding German version of Eher et al. (2008b).

^c Out of 12 matching criteria established by Rossegger and colleagues (Rossegger et al., 2013).

^d Base rate of violent (including sexual) recidivism for offenders with a SORAG score.

The Present Study

The aim of the present study was to conduct the first replication of the SORAG in the country of Switzerland. Two total forensic cohorts in the Canton of Zurich were followed for up to seven years, with novel charges and/or convictions for a violent (including sexual) offense used as criteria for recidivism. Both the discrimination and calibration of the SORAG were measured, and the distribution of total scores was compared to the percentile ranks published by the tool authors. As attention was paid to matching those participant and study design characteristics of the SORAG development study, high rates of predictive validity were hypothesized.

Material and Methods

Sample

The study sample consisted of two total cohorts of violent and sex offenders from the Canton of Zurich, Switzerland (N = 861). The first cohort was taken from the Zurich Forensic Study (Urbanik et al., 2007), which longitudinally followed all offenders with either a sentence of at least 10 months or court-ordered therapy who were supervised by the criminal justice system of the Canton as of August 2000 (N = 465). The second cohort consisted of all forensic patients receiving treatment in the Psychiatric/Psychological Service who began treatment between January 1, 1997 and December 31, 2009 (N = 296). The Psychiatric/Psychological Service is the largest provider of both outpatient and inpatient offender treatment, with approximately 250 violent and sexual offenders receiving services at any one time. To make the cohorts comparable to the SORAG development sample, only adult male offenders (n = 740) with contact sex offenses being the index offense (n =

267) and who were released into the community with a potential follow-up of seven years ($n = 168$) were included. Upon excluding participants who died ($n = 5$), were deported before recidivating ($n = 18$), or had more than four missing items on the SORAG ($n = 8$), a final study sample of 137 offenders was obtained.¹

Procedure

The SORAG was coded based on clinical and criminal justice files by Master's-level psychologists who had attended an accredited Psychopathy Checklist-Revised (Hare, 2003) workshop and were blind to individual offenders' outcomes. The integrity of the SORAG assessments was ensured through the use of a validated and peer-reviewed translation of the instrument (Rossegger, Gerth, Urbaniok, Laubacher, & Endrass, 2010). As recommended by the SORAG authors (Quinsey et al., 2006), item 13 (phallometric test results) was substituted with diagnoses of pedophilia or sadism according to the Diagnostic and Statistical Manual of the American Psychiatric Association or the Screening Scale for Pedophilic Interests (Seto, Harris, Rice, & Barbaree, 2004). This systematic substitution was necessary, as phallometric assessments are not legally admissible in Switzerland. Further, there is no validated penile plethysmography test available in Switzerland (Cantonal Court of Zurich, 2012). When items were missing, the prorating algorithm suggested by the SORAG authors was followed. Using this administration strategy, trained raters reached good interrater reliability of > 0.70 (Fleiss, Levin, & Paik, 2003).

Following the SORAG manual, recidivism was defined as a new charge and/or conviction for a violent (including sexual) offense within seven years of release. Data on recidivism was based on criminal records, which were last reviewed in 2011. Acts of pseudo-recidivism (e.g., a new charge and/or conviction after an index offense that was precipitated by an incident prior to the index offense) were identified by the construction of crime trajectories for each offender and were not considered acts of recidivism (cf. Quinsey et al., 2006).

Statistical Analysis

Both the discrimination and calibration components of predictive validity were investigated for the SORAG. Discrimination was measured using receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC) parameter. To measure calibration, the expected seven year recidivism rates according to the SORAG's published norms (Quinsey et al., 2006) were compared to those rates observed in the present sample both overall and for each of the nine risk bins, individually. This was evaluated using the χ^2 test to assess the goodness-of-fit between expected and observed recidivism rates per risk bin, as well as Sanders' (1963) decomposition of the Brier score (1950) to provide an index of variation in forecasting. To further investigate the current controversy surrounding the usefulness of base rate-adjusted actuarial models (G. T. Harris & Rice, 2013; Mossman, 2006), we calculated the likelihood ratio (LR) for each SORAG risk bin and compared it to that established for the SORAG development sample (Rice, personal communication, July 17, 2013). Following the guidance of Mossman (2006), we also tested a calibration model with estimated rates obtained by applying Bayes' theory.

In addition to exploring discrimination and calibration validity, a one-sample t-test was used to examine the difference in expected and observed percentile. Further, a Kolmogorov-Smirnov Test (K-S test) was conducted to compare the distribution of risk bins. The K-S test produces a D statistic and corresponding p-value which are not affected by scale changes but rather serve to capture information on the relative distribution of the SORAG data. All analyses were two-tailed, used a significance threshold of $\alpha = 0.05$, and were conducted using STATA/IC 12.1 for Windows and OSX (StataCorp, 2012).

Results

Sample Characteristics

The present study sample was composed of offenders convicted of either child molestation ($n = 83$, 60.6%) or rape ($n = 54$, 39.4%), with a mean age at conviction of 39.4 years ($SD = 11.8$). The majority of the sample ($n = 105$, 76.6%) was enrolled in a treatment program, with most of these offenders having had therapy mandated by the court ($n = 92$, 87.6%). Diagnostic criteria for a personality disorder were met by 60 (43.8%) offenders, while 16 (11.7%) met criteria for schizophrenia and 29 (21.2%) for substance abuse or dependency. The base rate for violent (including sexual) recidivism within seven years after release was 16.1% ($n = 22$).²

Predictive Validity of the SORAG

The predictive validity of the SORAG was measured using both discrimination and calibration performance indicators. ROC analyses revealed an acceptable level of discrimination for both SORAG total risk scores ($AUC = 0.69$, 95% CI = 0.56-0.82, $p < 0.05$) and risk bins ($AUC = 0.67$, 95% CI = 0.54-0.80, $p < 0.05$). Five of the nine risk bin LRs fell outside of the 95% confidence intervals established for the SORAG development sample (Table 2). This is reflected by the irregular shape of the ROC graph for the ZSOP compared to the developmental sample (Figure 1 and 2). The seven year recidivism rates both overall as well as for each risk bin were considerably lower than for the SORAG norms (Table 2). This lack of goodness-of-fit was reflected in a significant χ^2 statistic and a resolved Sanders' decomposition score of 0.04, corresponding to an average forecast error of 20.9% per risk bin. The ratio of the excess forecast variance to the minimum forecast variance for the SORAG was 9.3, with ratios higher than 6.0 suggesting "considerably excess variation in forecasts" (Spiegelhalter, 1986, p. 427). When the norms were adjusted with respect to the base rate of the ZSOP sample by applying Bayes' theory, comparison between the observed and estimated expected recidivism rates showed reduced but still substantial differences. As seen in Figure 3, recidivism rates were under-estimated as well as over-estimated (Sanders' decomposition score = 0.007; average forecast error of 8.4% per risk bin; ratio of the excess forecast variance to the minimum forecast variance = 6.2).

Risk bin	Total risk score	Percentage of sample in each risk bin		Recidivism rate ^e		LR of risk bin (95% intervall)	
		Quinsey et al. ^{af}	ZSOP	Quinsey et al. ^a	ZSOP	Quinsey et al. ^{ag}	ZSOP
1	≤ -10	4.9%	13.9%	7%	5.2%	0.11 (0.01-0.80)	0.29
2	-9 to -4	8.0%	15.3%	15%	14.2%	0.21 (0.06-0.68)	0.87
3	-3 to +2	13.9%	22.6%	23%	6.4%	0.40 (0.20-0.81)	0.36
4	+3 to +8	20.1%	19.0%	39%	23.1%	0.87 (0.54-1.40)	1.57
5	+9 to +14	18.1%	12.4%	45%	11.8%	1.09 (0.67-1.80)	0.70
6	+15 to +19	16.0%	5.1%	58%	42.9%	1.79 (1.05-3.06)	3.92
7	+20 to +24	11.1%	5.8%	58%	12.5%	1.77 (0.92-3.43)	0.75

8	+25 to +30	6.3%	2.9%	75%	0.0%	4.83 (1.63-14.31)	0
9	≥ +31	1.7%	2.9%	100%	100.0%	NA	NA

Note. ZSOP = Zurich sex offender population. LR = Likelihood Ratio. NA = not applicable.

^aSORAG development sample.

^eRecidivism rate over seven years follow-up.

^fM.E. Rice, personal communication, July 17, 2013.

^gLRs of the development study were estimated on the basis of the distribution of SORAG risk bins (column two in table 1) and recidivism rate within risk bins in the development sample (column five in table 2).

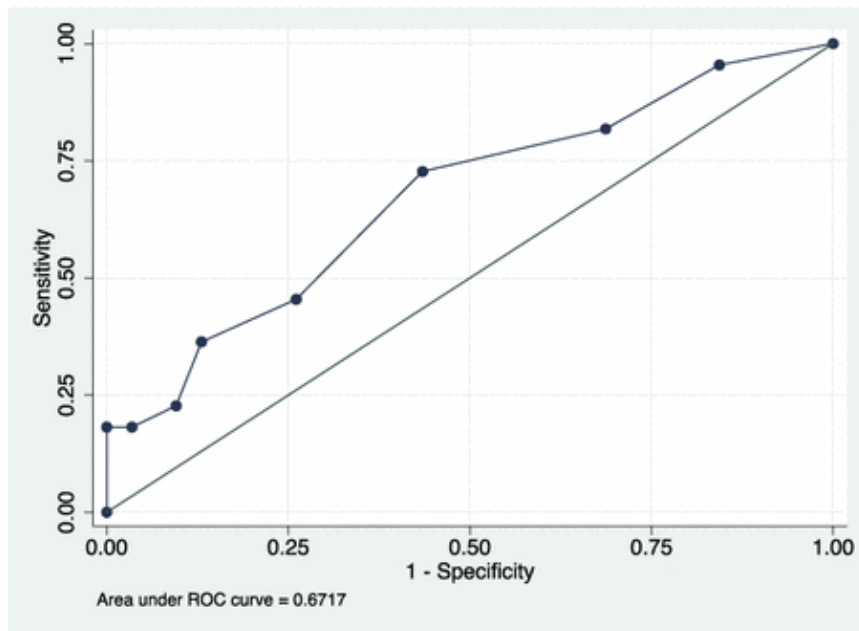


Figure 1: Receiver operating characteristic (ROC) graph displaying the discrimination of the SORAG risk bins in the ZSOP

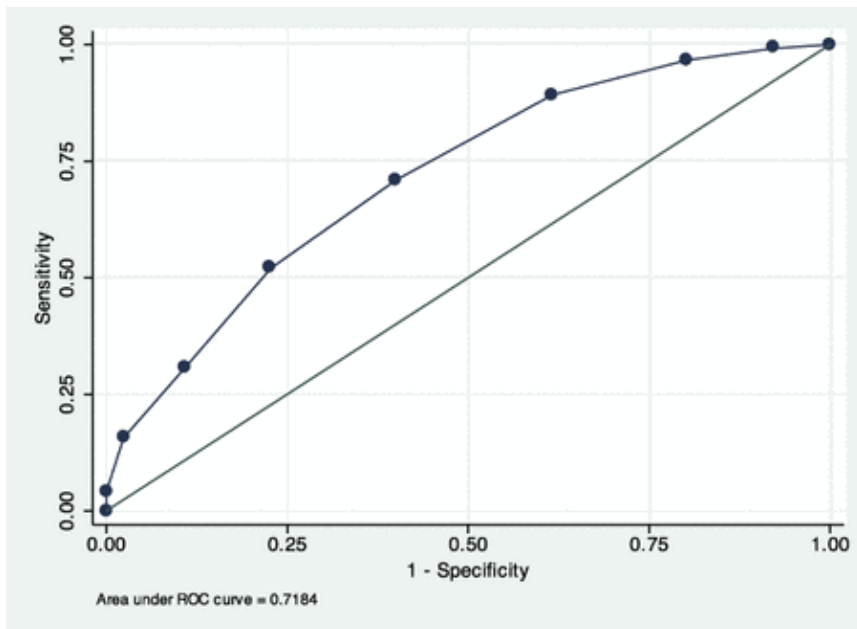


Figure 2: Estimated receiver operating characteristic (ROC) graph displaying the discrimination of the SORAG risk bins in the development sample

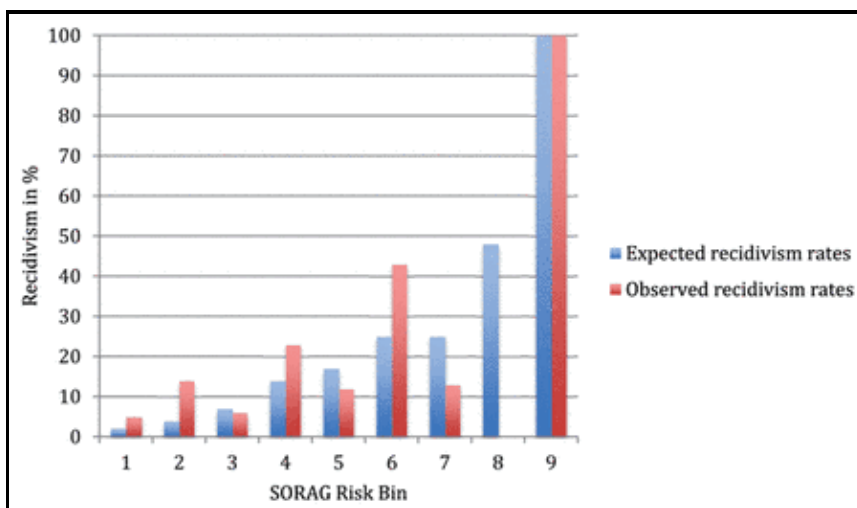


Figure 3: Comparing expected and observed recidivism rates by applying Bayes' theory

The mean SORAG total risk score was +3.4 (SD = 11.6), with scores ranging from -16 to +39. This mean score was significantly lower than that in the tool's development study (M = 8.9, SD = 11.3), $t(136) = -5.54$, $p < 0.001$. Further, the distribution of SORAG scores as exhibited by percentile ranks was significantly different than in the SORAG development study, $t(59) = 11.1$, $p < 0.001$ (Figures 4 and 5). The mean deviation from the expected percentile was 7.6 (SD = 5.3). When a K-S test was conducted to investigate the difference in the distribution of offenders in the SORAG risk bins between the development study and the present study, a significant effect was found with

approximately half (51.8%) of the offenders in the Zurich sex offender population classified into the first three risk bins ($D = 0.25$, $p < 0.001$).

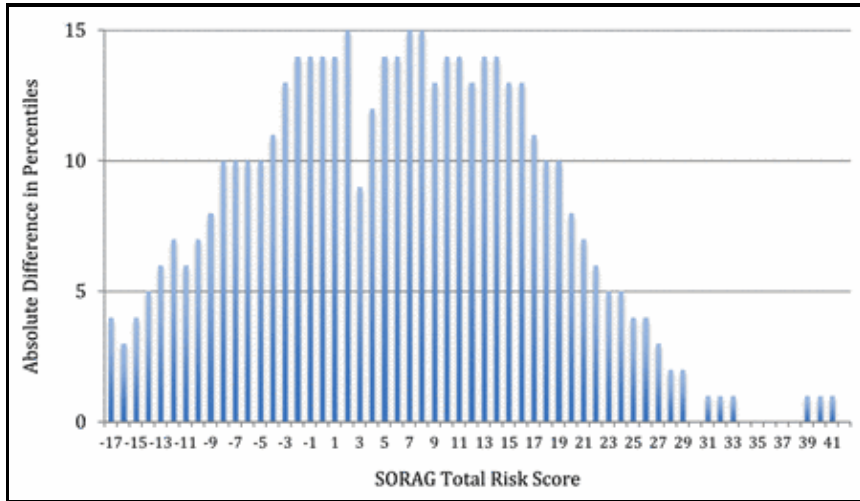


Figure 4: Absolute differences in percentiles between the SORAG development sample and the Zurich sex offender population

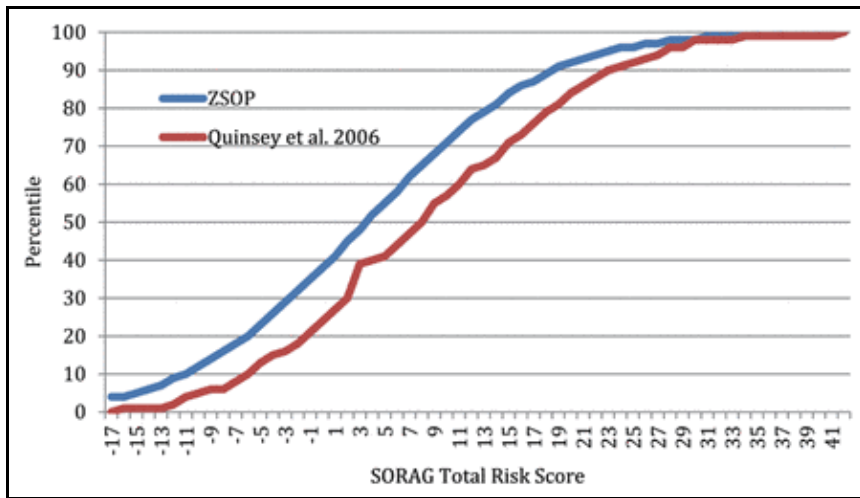


Figure 5: Percentiles corresponding to SORAG total risk scores for the tool's development sample and the Zurich sex offender population

Discussion

The aim of the present study was to provide the first examination of the predictive validity of the

SORAG in Switzerland. The sample consisted of two total forensic cohorts from the Canton of Zurich, Switzerland that were followed up for seven years in the community. The SORAG was coded using file information for all sex offenders with strict adherence to the manual-based protocol. Recidivism was defined as a new charge and/or conviction for a violent (including sexual) offense. Given the match between the participant and study design characteristics of the present investigation and those of the SORAG development study, rates of predictive validity and percentile distributions were expected to be comparable to those in the tool's development sample.

There were three main findings of the study: First, ROC analyses revealed that the ability of the SORAG to discriminate between recidivists and non-recidivists was lower for the Zurich sex offender population than for the Canadian sex offender population. With an AUC of 0.69 it was of moderate performance and lower than in the instrument's development sample in Canada (AUC=0.75; Quinsey et al., 2006). This finding supports previous research (e.g. Bartosh et al., 2003; Looman, 2006; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002) that has found lower AUCs for the SORAG compared to those found in the original study. According to the tool's authors, lower AUCs can be explained by lack of adherence to the administration protocol published in the SORAG's manual and/or substantial deviation from the study design used in the original study. As we used a total cohort of sex offenders and adopted a study design with a strong replication match, these explanations are not available for the present study. In addition, the comparatively lower base rate in Switzerland compared to Canada is not an explanation, as ROC analyses are robust to base rate variations. We also found that the likelihood ratios of five out of the nine SORAG risk bins differed substantially from those of the instrument's development sample, which led to considerable differences between the two ROC plots. Hence, there is preliminary evidence suggesting a difference in the SORAG's ability to discriminate between recidivisms and non-recidivists in Switzerland compared to Canada.

Second, calibration analyses suggested that published expected recidivism rates produced by the SORAG should perhaps not be used as official predictions of recidivism rates for sex offenders in Switzerland at this time. Inconsistent fit was found between the expected and observed recidivism rates in all nine SORAG risk bins, with over-estimation errors ranging from 1.4% to 81.8%. This is consistent with prior findings in other German-speaking countries (Eher, Rettenberger, Schilling, & Pfäfflin, 2008a; Eher, Rettenberger, Schilling, & Pfäfflin, 2008b). Additionally, our findings suggested that calibration problems could not be solved by simply adjusting expected recidivism rates. This concurs with recent research by Harris and Rice (2013) who tested a re-calibration of a related risk assessment tool (the Violence Risk Appraisal Guide) using Bayesian methods. Given these differences, simply adjusting expected rates by taking base rate measures into account may not be as useful as undergoing a jurisdiction-specific re-calibration process.

Third, the distribution of SORAG total risk scores (and hence percentiles) differed from the distribution for the tool's development sample. Thus, while the supposed advantage of actuarial instruments is the provision of population norms to aid in the interpretation of assessment results, this finding suggests that Swiss sex offenders may not be accurately ranked using published guidelines. This is important, as whether an offender is considered to be at the 25th or the 75th percentile regarding his risk to recidivate may influence decision makers in determining therapeutic resource allocation and release conditions. Albeit published percentile rankings appear to have less use in Switzerland, independent research suggests that such rankings may still be useful in Canada (Barbaree, Langton, & Peacock, 2006). At this time, published expected recidivism rates and percentiles are not advised to be provided in reports on Swiss sexual offenders.

Limitations

Although we used a total cohort, the investigated sample was nevertheless small. Given the high prevalence of sex offenders receiving treatment in the Swiss criminal justice system, approximately three out of four offenders (76.6%) participated in a court-ordered treatment program. This figure is substantially higher than in the development sample of the SORAG (36.1%). Since studies investigating the effectiveness of sex offender treatment programs have shown preliminary evidence that specific psychotherapeutic interventions - especially those adhering to Risk-Needs-Responsivity principles (see Hanson, Bourgon, Helmus, & Hodgson, 2009 for a review) - may be able to reduce recidivism rates, it could be that the lower recidivism rate in Switzerland is the result of the routine implementation of treatment programs (cf. Lipsey & Cullen, 2007). However, future research into the efficacy of sex offender-specific treatment programs is warranted (Dennis et al., 2012), and it should be kept in mind that the present sample appears to have been an overall lower risk sample than the SORAG development sample. The latter may have affected the base rate of recidivism, as may have the fact that we did not subtract time spent in a criminal justice or mental health institution for a non-violent crime from the total follow-up period for each offender. Hence, appropriate caution is warranted before further research by independent research can be conducted in Switzerland to replicate our findings.

Notes

¹Offenders who changed their names were still able to be followed.

²Stratified by offense type: child sexual abuse (n = 13, 9.5%), rape (n = 6, 4.4%), assault (n = 3, 2.2%), and homicide (n = 1, 0.7%).

References

1. Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist, 34*(3), 341-382. doi: 10.1177/0011000005285875
2. Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sex Abuse, 18*, 423-440.
3. Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to the sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*(4), 422-438. doi: 10.1177/0306624X03253850
4. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1-3.
5. Cantonal Court of Zurich. (2012). Decree of June 5th, 2012, from http://www.gerichte-zh.ch/fileadmin/user_upload/entscheide/oeffentlich/UG070045-O2.pdf
6. Dennis, J. A., Khan, O., Ferriter, M., Huband, N., Powney, M. J., & Duggan, C. (2012). Psychological interventions for adults who have sexually offended or are at risk of offending (Review). *Cochrane Database of Systematic Reviews: Issue 12*. Art. No.: CD007507. DOI: 10.1002/14651858.CD007507.pub2.
7. Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sex Abuse, 24*(5), 5-28.
8. Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008a). Validity or benefit? The prediction of relapses with Static-99 and SORAG. A prospective study with 275 sex

- offenders. *Recht und Psychiatrie*, 26(2), 79-88.
9. Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2008b). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment*, 3(1), 1-14. Retrieved from http://www.sexual-offender-treatment.org/1-2008_02.html
 10. Fleiss, J. L., Levin, B., & Paik, M. C. (2003). The measurement of interrater agreement. In D. J. Balding, N. A. C. Cressie, N. I. Fisher, I. M. Johnstone, J. B. Kadane, L. M. Ryan, D. W. Scott, A. F. M. Smith & J. L. Teugels (Eds.), *Statistical methods for rates and proportions* (3 ed., pp. 598-626). New York: John Wiley & Sons.
 11. 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.
 12. Hanson, K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). *A meta-analysis of the effectiveness of treatment for sexual offenders: Risk, need, and responsivity*: Public Safety Canada.
 13. Hare, R. D. (2003). *Manual for the Revised Psychopathy Checklist* (2nd ed.). Toronto, ON: Multi-Health Systems.
 14. Harris, G. T., & Rice, M. E. (2013). Bayes and base rates: What is an informative prior for actuarial violence risk assessment? *Behavioral Sciences and the Law*, 31, 103-124.
 15. Harris, G. T., Rice, M. E., Lalumière, M. L., Quinsey, V. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15(3), 413-425. doi: 10.1037/1040-3590.15.3.413
 16. 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, 409-448. doi: 10.1007/s11194007-9062-3
 17. Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *The Annual Review of Law and Social Science*, 3, 297-320.
 18. Looman, J. (2006). Comparison of Two Risk Assessment Instruments for Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment*, 18(2). doi: 10.1007/s11194-006-9013-4
 19. Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment*, 18(1), 41-63.
 20. Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment*, 14, 253-269.
 21. Pham, T.-H., & Ducro, C. (2008). Risk assessment in social defence: Preliminary factorial analysis of the Sex Offender Recidivism Appraisal Guide (SORAG) and the Static-99. *Annales Médico-Psychologiques*, 166, 575-579.
 22. Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2. ed.). Washington, DC: American Psychological Association.
 23. Rice, M. E., & Harris, G. T. (2002). Men who molest their sexually immature daughters: Is a special explanation required? *Journal of Abnormal Psychology*, 111(2), 329-339. doi: 10.1037//0021-843X.111.2.329
 24. Rossegger, A., Gerth, J., Seewald, K., Urbaniok, F., Singh, J. P., & Endrass, J. (2013). Current obstacles in replicating risk assessment findings: A systematic review of commonly used actuarial instruments. *Behavioral Sciences and the Law*, Online First.
 25. Rossegger, A., Gerth, J., Urbaniok, F., Laubacher, A., & Endrass, J. (2010). The Sex Offender Risk Appraisal Guide (SORAG). Validity and authorised German translation. [Review]. *Fortschritte der Neurologie-Psychiatrie*, 78(11), 658-667. doi: 10.1055/s-0029-1245688

26. Sanders, F. (1963). On subjective probability forecasting. *Journal of Applied Meteorology*, 2(2), 191-201.
27. Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The screening scale for pedophilic interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior*, 33, 455-466.
28. Singh, J. P., Desmarais, S., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences and the Law*, 31, 55-73.
29. Spiegelhalter, D. (1986). Probabilistic prediction in patient management and clinical trials. *Statistics in Medicine*, 5, 421-433.
30. StataCorp. (2012). *Stata statistical software: Release 12*. College Station: TX: StataCorp LP.
31. Urbaniok, F., Endrass, J., Rossegger, A., Noll, T., Gallo, W. T., & Angst, J. (2007). The prediction of criminal recidivism: The implication of sampling in prognostic models. *European Archives of Psychiatry and Clinical Neuroscience*, 257(3), 129-134. doi: 10.1007/s00406-006-0678-y
32. Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17, 377-395.

Author address

Jérôme Endrass, PhD

Department of Justice

Psychiatric/Psychological Service

Canton of Zurich

Feldstrasse 42, P.O. Box 8090

Zurich, Switzerland

jerome.endrass@ji.zh.ch