Accuracy of the Static-99 in Predicting Recidivism in Switzerland

Jerome Endrass
Frank Urbaniok
Psychiatric/Psychological Service, Zurich, Switzerland
Leonhard Held
Institute for Social and Preventive Medicine, Zurich, Switzerland
Stefan Vetter
Center for Disaster and Military Psychiatry, Zurich, Switzerland
Astrid Rossegger
Psychiatric/Psychological Service, Zurich, Switzerland

The Static-99 is a widely used actuarial risk assessment instrument. Various international validation studies have found satisfactory to good predictive validity for the Static-99, with the area under the curve (AUC) between 59% and 95%. This study is the first evaluation of the predictive accuracy of the Static-99 among sex offenders in Switzerland. The Static-99 scores of 69 violent/sex offenders in Switzerland were assessed using data from their psychiatric assessments. Recidivism was operationalized as reconviction assessed from penal records. The Static-99 risk levels were predictive for recidivism (AUC = .758) among our population. The results are discussed on the basis of the literature.

Keywords: *Static-99; calibration; sex offender; validity; recidivism; Switzerland*

R isk assessment—specifically, that of the recidivism of offenders—is a key aspect of forensic psychiatry. In North America, the application of actuarial prognosis instruments is considered state of the art (Andrews & Bonta, 2003; Quinsey, Harris, Rice, & Cormier, 2003) and is seen to be superior to clinical and especially intuitive judgment (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hilton & Simmons, 2001; Mossman, 1994; Quinsey et al., 2003; Rice & Harris, 1997; Swets, Dawes, & Monahan, 2000).

Actuarial instruments for the prognosis of recidivism of sex offenders include the Sex Offender Risk Appraisal Guide (Quinsey et al., 2003), the Static-99 (Hanson &

Thornton, 2000), the Minnesota Sex Offender Screening Tool (Epperson, Kaul, & Hesselton, 1998), the Rapid Risk Assessment for Sex Offence Recidivism (Hanson, 1997), and the Structured Anchored Clinical Judgment (Grubin, 1998).

The Static-99 was developed by Hanson and Thornton in 1999 on a sample of 1,301 offenders from Canada and Great Britain using four sets of data (Hanson & Thornton, 1999) and was topic of many validation studies in multiple countries since its development. In most of the studies, the area under the curve (AUC) ranged from .70 to .80 (Barbaree, Seto, Langton, & Peacock, 2001; Beech, Friendship, Erikson, & Hanson, 2002; Hood, Shute, Feilzer, & Wilcox, 2002; McGrath, Cumming, Livingstone, & Hoke, 2000; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002; Sjöstedt & Langström, 2001), which reflects a satisfactory to good prognostic accuracy.

Bartosh, Garby, Lewis, and Gray (2003) compared the results of four actuarial prognosis instruments—the Static-99, the Rapid Risk Assessment for Sex Offence Recidivism, the Minnesota Sex Offender Screening Tool, and the Sex Offender Risk Appraisal Guide—on groups of sex offenders. The prognostic accuracy of the instruments depended on the kind of sex offense committed. Only the Static-99 (the AUC for sexual recidivism was .65 and for violent or sexually violent recidivism, .87) and the Sex Offender Risk Appraisal Guide were significantly predictive for sex and violent offenses by offenders who had perpetrated sexual abuse against children outside their families. All four risk assessment instruments showed at least satisfactory predictive validity for sexual recidivism of incest offenders (Static-99: AUC = .74). The Static-99 also performed well for violent recidivism (AUC = .92). For the categories of *rapist* and *hands-off* sex offenders, none of the four instruments showed significant validity—with the exception being that the Static-99 and Sex Offender Risk Appraisal Guide achieved sufficient predictive validity for sexual recidivism of rapists (Static-99: receiver operating characteristic [ROC] = .71).

With respect to the validity of the Static-99 in other-than-Anglo-Saxon countries, A. J. R. Harris, Phenix, Hanson, and Thornton (2003) stressed that there is evidence that the Static-99 may be culturally independent and a useful risk assessment tool in a range of countries. They referred to a study by Nicholaichuk (2001), who showed that the Static-99 is as predictive for Canadian Native Americans as it is for Caucasians. The few studies from continental Europe confirmed the assumption of cultural independence: Sjöstedt and Langström (2001) examined the predictive validity of the Static-99 and the Rapid Risk Assessment for Sex Offence Recidivism retrospective on a Swedish sample of 1,400 sex offenders. The authors were able to show that in addition to demonstrating a good predictive validity throughout, the Static-99 was superior to the Rapid Risk Assessment for Sex Offence Recidivism in all tested areas (rapists and child abusers) concerning recidivism of a sexual or violent (including sexually violent) nature, with the AUC ranging from .70 to .76. With regard to a Dutch sample of sex offenders (95 rapists and 27 child abusers), de Vogel, de Ruiter, van Beek, and Mead (2004) found a moderate predictive validity for the Static-99 for sexual recidivism (AUC = .71) but neither for violent (nonsexual) nor

for general recidivism. Ducro and Pham (2006) investigated the predictive validity of the Static-99 for 147 sexual offenders (rapists and child abusers) in the maximum-security section of a Belgium psychiatric hospital. The authors distinguished three types of recidivism: general, sexual, and sexually violent. With regard to general recidivism, they found a satisfactory predictive validity (AUC = .70). However, the predictive validity for sexual and sexually violent recidivism was moderate (sexual recidivism: AUC = .66; sexually violent offenses: AUC = .67; Ducro & Pham, 2006).

In a recent study, Stadtland et al. (2006) investigated a sample of 134 sex offenders in Germany for whom a court had ordered a forensic psychiatric expert assessment to be conducted. In this study, the Static-99 showed the highest predictive validity (AUC = .71) when compared to the Historical Clinical Risk–20 (Webster, Douglas, Eaves, & Hart, 1997), the Sexual Violence Risk–20 (Boer, Hart, Kropp, & Webster, 1997), and the Psychopathy Checklist–Revised (Hare, 1991).

There is strong empirical evidence for the discriminant validity of the Static-99 in predicting sexual recidivism among sex offenders. The evidence for German-speaking countries is weaker, and it is unclear whether the Static-99 is an accurate instrument for predicting recidivism among sex offenders in Switzerland. Furthermore, it is not clear whether the Static-99 is accurately calibrated for Swiss offenders. As such, the aim of this study was to conduct the first evaluation of the discriminant validity and calibration of the Static-99 among sex offenders in Switzerland.

Method

Sample Selection Criteria

The sample examined in this study is a subsample from the Zurich Forensic Study (Urbaniok et al., 2007), which comprises 451 violent and sex offenders. The Zurich Forensic Study included all sex and violent offenders of the canton of Zurich, Switzerland, who as of August 2000 were in jail, on parole, on probation, or in forensic therapy (inpatient and outpatient) and had been sentenced to a minimum prison sentence of 10 months.

Of the sample of all sex offenders (*N* = 159) of the Zurich Forensic Study, a sub-sample of sex offenders were selected for the current analysis who had a time at risk of at least 5 years, until March 2006. Seventy offenders were thus excluded because of this criterion, and the Static-99 was thus scored for the remaining 89 sex offenders. Furthermore, offenders who could not be scored because of a lack of file information (at least one missing value) were excluded as well (*n* = 20). The final sample analyzed therefore consisted of 69 persons.

Sample Characteristics

Sociodemographic variables. In sum, 65.2% (*n* = 45) were Swiss nationals; 29.4% (*n* = 20) were married at the time of offense; 47.1% (*n* = 32) had a child; and

58.7% ($n = 37$) had completed vocational training. At the time of the offense, 83.1% ($n = 54$) of the offenders were employed.

Offense related variables. The index offense was rape in 40.6% of the cases ($n = 28$), child molestation in 52.2% ($n = 36$), defilement in 1.5% ($n = 1$), and exhibitionism in 5.8% ($n = 4$). Overall, 60.8% ($n = 42$) of the offenders were previously convicted, with 37.7% ($n = 28$) having committed a sexual offense.

Measure and Procedure

The Static-99 consists of 2 demographic items, 5 items concerning the criminal history, and 3 items concerning the victims. All 10 items can be scored without any direct contact with the offender, and the sum score of the 10 items permits one to assign the offender to one of four risk categories. In this study, all data used for the assessment of the Static-99 were collected from correctional and court files; there was no direct contact with the offenders. The files contained the psychiatric reports and all previous charges. Therefore, extensive historical details on the offender were available, including criminal history, exact type and circumstances of the offense, and any personality and psychiatric diagnoses. The Static-99 data, as well as data concerning sociodemographic background, psychiatric disorder, and criminal history, were collected by forensic psychiatrists. In a pilot study, the interrater agreement was computed for the variables, yielding an alpha value greater than .65. The interrater reliability was assessed with Krippendorff's alpha (2003). The advantage of this measure is that it can be used to analyze the agreement of multiple raters even if there are unequal sample sizes or missing data, and it can be computed when the variables are nominal, ordinal, or continuous.

The dependent variable, recidivism, was defined as any reconviction by a criminal court for a sex offense with a minimum prison sentence of 10 months. Information about recidivism (reconviction) was gathered from penal records in March 2006. The register contains the date of conviction, the type of offense, the length of sentence, reprieve, deportation, the date of conditional release, and the duration of probation. The date that the incarceration commenced was not noted. To prevent bias, the Static-99 was coded before examination of the penal records.

The recidivism rates of the Swiss sample were compared with the 5-year recidivism rates published in G. T. Harris et al. (2003) using the four risk categories: low = 5.7% (Static-99 sum score of 0–1), medium = 10.2% (Static-99 sum score of 2–3), medium-high = 28.9% (Static-99 sum score of 4–5), and high = 38.8% (Static-99 sum score of 6 or higher).

Statistical Analysis

To determine the predictive accuracy of the Static-99, exact logistic regression analysis was used to assess its discriminant validity. Calibration was assessed by

Table 1
Static-99 Risk Scores and Recidivism Rates (in Percentages)

Static-99 Risk Level	Five-Year Recidivism Rate[a]	Swiss Sex Offender Population Recidivism Rate
1	5.7	0.0 (0 of $n = 7$)
2	10.2	6.1 (2 of $n = 33$)
3	28.9	22.2 (4 of $n = 18$)
4	38.8	36.4 (4 of $n = 11$)

a. Based on data from G. T. Harris et al. (2003).

comparing the 5-year recidivism rates reported in G. T. Harris et al. (2003) with the recidivism rates in the Swiss sex offender population in each risk category. All models were estimated using Stata/SE 10.0.

Results

Recidivism Rates and Static-99 Scores

The rate of reconviction for a sex offense was 14.5% ($n = 10$). Two sex offenders were also convicted for a nonsexually motivated violent offense. There was no sex offender who recidivated with only a violent offense.

The mean sum score in the Static-99 was 3.6 ($SD = 1.8$, range $= 0$–8). Overall, 10.1% ($n = 7$) of the sex offenders were in the lowest risk category; 47.8% ($n = 33$) were medium risk; 26.1% ($n = 1$) were in the medium-high category; and 15.9% ($n = 11$) were in the highest risk category.

Predictive Validity of the Static-99

The AUC based on the four risk categories was .758. The exact logistic regression analysis yielded an odds ratio (OR) of 3.057 (95% confidence interval [CI] = 1.68, 8.406) per one risk category increase. Table 1 shows, as expected, that the higher the category, the more recidivists are present.

In a subsequent analysis the calibration of the Static-99 was assessed. If the model was perfectly calibrated, then the rates of recidivism within the risk categories reported in G. T. Harris et al. (2003) could be replicated in our sample. As such, the table reveals that the recidivism rates reported by G. T. Harris et al. and the Swiss sex offender population recidivism rates do not exactly concur. Sanders calibration score (Schmid & Griffith, 2005), which measures the squared error between the average forecast and the average outcome in each risk category, was 0.0024, which corresponds to an average error of 4.9%. However, all recidivism rates reported by G. T. Harris et al. are higher than the recidivism rates in the Swiss Sex offender population.

Discussion

Actuarial prognosis instruments are considered to have an important relevance in risk assessment because they lead to satisfying, precise predictions (Grove et al., 2000; Hilton & Simmons, 2001; Mossman, 1994; Quinsey et al., 2003; Rice & Harris, 1997; Swets et al., 2000). However, empirically generated forensic risk assessment models cannot be easily generalized to different populations. Even slight changes in the composition of the calibration sample from the training sample can lead to different statistical models. In some cases, only the weights of the predictors differ; in other cases, some variables may even "lose" their statistical significance (Urbaniok et al., 2007). Actuarial risk assessment instruments therefore have to be validated for each population before they can be implemented in the decision-making process of the criminal justice system.

Given that the validity of the Static-99 has not yet been determined in Switzerland, we investigated the question of whether the instrument was valid in predicting recidivism of Swiss offenders. To measure validity, we compared the predictive validity found in our study with the results given in G. T. Harris et al. (2003).

Because ROC analyses are considered the gold standard for the validity of risk assessment instruments (Douglas, Ogloff, Nicholls, & Grant, 1999; Grann, Belfrage, & Tengstrom, 2000), we compared the AUC of our study with that of the empirical literature. The ROC analysis of the original sample of male prisoners in Canada and the United Kingdom found an AUC of .71 (G. T. Harris et al., 2003). Various validation studies have found a range of AUCs between 59% and 95% (Doren, 2004). In most of these studies, the ROC of the Static-99 ranged from .70 to .80 (Barbaree et al., 2001; Beech et al., 2002; Hood et al., 2002; McGrath et al., 2000; Nunes et al., 2002; Sjöstedt & Langström, 2001), which reflects a satisfactory to good prognostic validity.

The ROC analyses in the current study yielded a satisfactory AUC of .758, indicating that the Static-99 is suitable for predicting sexual recidivism among sex offenders. Our results thus replicate the findings of other authors concerning the validity of the Static-99 for predicting sexual recidivism. However, we considered that a measure of discriminant validity such as the ROC was not sufficient to decide whether the Static-99 was suitable for risk assessment in Switzerland. The problem of relying solely on ROC analyses to assess the validity of a forensic risk assessment instrument can be demonstrated with the following hypothetical example: If all the recidivists in our sample had scored 10 points on the Static-99 and all the nonrecidivists had scored 9 points, then the AUC would have been 1.00. In this case, ROC analysis would have demonstrated that the Static-99 score allows perfect discrimination between recidivists and nonrecidivists. However, the risk categories of the Static-99 are hypothesized to be incremental, meaning that the higher an offender scores on the instrument, the higher the risk is for reoffending. In the example given, this assumption is true only for the discrimination of high (9 points) and very high (10 points) scores, and there would be no evidence base for the use of the Static-99 for

low-risk offenders. Such a result could not be therefore interpreted as a validation of the Static-99, given that the distribution of the recidivism rates across the risk categories would have been substantially different from those of the original publications (e.g., G. T. Harris et al., 2003).

To decide whether an instrument is suitable across all risk categories—in the sense that the probabilities for reoffending within the risk categories can be replicated—a measure of calibration has to be used. Hence, aside from ROC analyses, the Sanders calibration score was computed (Schmid & Griffith, 2005) and was convenient to compute because the probabilities for recidivism for each Static-99 score are published for each risk category (G. T. Harris et al., 2003). If the instrument is calibrated for a Swiss offender population, similar recidivism rates for each Static-99 score can be expected.

The results indicate some degree of lack of fit, which is largely due to an overall difference between the mean predicted and the mean actual recidivism rate. This finding suggests that a recalibration (Schmid, D'Agostino, Griffith, Beshansky, & Selker, 1997) of the Static-99 to the Swiss mean recidivism rate would largely remove the lack of fit in each risk category while maintaining good discrimination.

A limitation of the study is the small sample size. Unfortunately, it was impossible to differentiate the predictive validity of the Static-99 for different offender groups (child molestation, rape, hands-off offenses).

This study represents the first evaluation of the predictive accuracy of the Static-99 among sex offenders in Switzerland. Regarding the value of early detection of potentially recidivating sex offenders, we recommend further investigation concerning violent recidivism—especially, sexual recidivism—focusing also on other risk assessment instruments.

References

Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati, OH: Anderson.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28*(4), 490-521.

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*(4), 422-438.

Beech, A. R., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 156-167.

Boer, D. R., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk–20. Professional guidelines for assessing risk of sexual violence.* Vancouver, British Columbia, Canada: British Columbia Institute on Family Violence, and Mental Health, Law, & Policy Institute, Simon Fraser University.

de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28*(3), 235-251.

Doren, D. M. (2004). Stability of the interpretative risk percentages for the RRASOR and Static-99. *Sexual Abuse: A Journal of Research and Treatment, 16*(1), 25-36.

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing the risk for violence among psychiatric patients. *Journal of Consulting and Clinical Psychology, 67,* 917-930.

Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: A Journal of Research and Treatment, 18*(1), 15-26.

Epperson, D. L., Kaul, J. D., & Hesselton, D. (1998, October). *Final report on the development of the Minnesota Sex Offender Screening Tool–Revised (MnSOST-R).* Paper presented at the 17th annual research and treatment convention of the Association for the Treatment of Sex Abusers, Vancouver, British Columbia, Canada.

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*(1), 97-114.

Grove, W., Zald, D., Lebow, B., Snitz, B., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19-30.

Grubin, D. (1998). *Sex offending against children: Understanding the risk.* London: Home Office.

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offence recidivism* (No. 1997-04). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders* (No. 1999-02). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised.* Toronto, Ontario, Canada: MultiHealth Systems.

Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules revised—2003.* Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15,* 413-425.

Hilton, N. Z., & Simmons, J. L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*(4), 393-408.

Hood, R., Shute, S., Feilzer, M., & Wilcox, A. (2002). Sex offenders emerging from long-term imprisonment: A study of their long-term reconviction rates and of parole board members' judgement of their risk. *British Journal of Criminology, 42*(2), 371-394.

Krippendorff, K. (2003). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

McGrath, R. J., Cumming, G., Livingstone, J. A., & Hoke, S. E. (2000, November). *The Vermont Treatment Program for Sexual Aggressors: An evaluation of a prison-based treatment program.* Paper presented at the 19th annual conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Mossman, D. (1994). Assessing predictions of violence. Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783-792.

Nicholaichuk, T. (2001, November). *The comparison of two standardized assessment instruments in a sample of Canadian Aboriginal sexual offenders.* Paper presented at the 20th annual conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

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*(3), 253-269.

Quinsey, V. L., Harris, G., Rice, M., & Cormier, C. A. (2003). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Rice, M. E., & Harris, G. T. (1997). Cross validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21,* 231-241.

Schmid, C. H., D'Agostino, R. B., Griffith, J. L., Beshansky, J. R., & Selker, H. P. (1997). A logistic regression model when some events precede treatment: The effect of thrombolytic therapy for acute myocardial infarction on the risk of cardiac arrest. *Journal of Clinical Epidemiology, 50,* 1219-1229.

Schmid, C. H., & Griffith, J. L. (Eds.). (2005) *Encyclopedia of biostatistics* (2nd ed.). Chichester, UK: Wiley.

Sjöstedt, G., & Langström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the PRASOR and the STATIC 99 in Sweden. *Law and Human Behavior, 25*(6), 629-645.

Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2006). Rückfallprognosen bei Sexualstraftätern—Vergleich der prädiktiven Validität von Prognoseinstrumenten [Risk assessment of sex offenders—Comparison of the predictive validity of different risk assessment instruments]. *Der Nervenarzt, 77,* 587-595.

Swets, J., Dawes, R., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1-26.

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.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence, Version 2.* Burnaby, British Columbia, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.