Psychological Assessment
2017, Vol. 29, No. 6, 624–638

1040-3590/17/$12.00 http://dx.doi.org/10.1037/pas0000390

Actuarial Risk Assessment of Sexual Offenders: The Psychometric Properties of the Sex Offender Risk Appraisal Guide (SORAG)

Martin Rettenberger
Centre for Criminology, Wiesbaden, Germany, Johannes Gutenberg-University, and Federal Evaluation Centre for Violent and Sexual Offenders, Austrian Prison System, Vienna, Austria

Marnie E. Rice and Grant T. Harris
Waypoint Centre for Mental Health Care, Penetanguishene, Ontario, Canada

Reinhard Eher
Federal Evaluation Centre for Violent and Sexual Offenders, Austrian Prison System, Vienna, Austria, and Ulm University Hospital

The Sex Offender Risk Appraisal Guide (SORAG) is one of the most commonly used actuarial risk assessment instruments for sexual offenders. The aims of the present field study were to examine the predictive validity of the German version of the SORAG and its individual items for different offender subgroups and recidivism criteria in sexual offenders released from the Austrian Prison System (N = 1,104; average follow-up period M = 6.48 years) within a prospective-longitudinal research design. For the prediction of violent recidivism the German version of the SORAG yielded an effect size of AUC = .74 (p < .001, 95% CI = .70–.78). The predictive accuracy for general and violent recidivism was slightly higher than for general sexual and sexual hands-on recidivism. The effect sizes were found to be higher for the child molester sample than for rapists. However, the differences were significant only for general recidivism (z = 2.48, p = .001). Further analyses exhibited the SORAG to have incremental predictive validity beyond the VRAG and the PCL-R, and to remain the only significant predictor for violent recidivism once all 3 instruments were forced into a combined regression model. Twelve out of the 14 SORAG items were found to have a significant positive relationship with violent recidivism. The comparison of the relative and absolute risk indices between the Austrian and the Canadian samples showed that the normative data distribution yielded more (absolute risk indices) or less (relative risk indices) meaningful differences between the 2 countries.

Keywords: SORAG, actuarial risk assessment, sexual offender, recidivism, validity, reliability

Supplemental materials: http://dx.doi.org/10.1037/pas0000390.supp

The development and distribution of standardized psychological instruments for the assessment of recidivism risk in sexual and violent offenders is one of the most pioneering and influential improvements in the last decades of risk prediction research. In modern forensic psychology, there basically exist two different standardized methodological approaches to risk assessment (e.g., Hanson, 2009; Singh, Grann, & Fazel, 2011): actuarial risk assessment, and structured professional judgment. Fundamental characteristics of an actuarial risk assessment instrument (ARAI) are, first, that all risk factors included are empirically validated and combined into a total score based on an a priori fixed algorithm for combining the individual risk factors. Second, the total score can be linked to different kinds of empirically derived risk communication procedures: absolute (i.e., observed and/or calculated probabilities of recidivism risk for different risk categories or total scores) and relative risk indices (i.e., risk ratios, odds ratios, or percentile ranks). Structured professional judgment (SPJ) instruments like the Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997), on the other hand, also consist of an empirically derived list of risk and protective factors, but the final

Martin Rettenberger, Centre for Criminology, Wiesbaden, Germany, Department of Psychology, Johannes Gutenberg-University (JGU), and Federal Evaluation Centre for Violent and Sexual Offenders, Austrian Prison System, Vienna, Austria; Marnie E. Rice and Grant T. Harris, Research Department, Waypoint Centre for Mental Health Care, Penetanguishene, Ontario, Canada; Reinhard Eher, Federal Evaluation Centre for Violent and Sexual Offenders, Austrian Prison System, Vienna, Austria, and Department of Forensic Psychotherapy, Ulm University Hospital.

The current research project was conducted in accordance with the legal and ethical demands of the Austrian Department of Justice and in accordance with the national Data Protection Act. The views expressed are those of the authors and not necessarily those of the Austrian Prison System.

Martin Rettenberger and Reinhard Eher would like to express their appreciation and admiration of the tremendous scientific contribution made by Marnie E. Rice and Grant T. Harris, who passed away during the publication process of the present manuscript. Our deep condolences to their families, friends, and colleagues.

Correspondence concerning this article should be addressed to Reinhard Eher, Federal Evaluation Centre for Violent and Sexual Offenders, Austrian Prison System, Gerichtsgasse 6, A-1210 Vienna, Austria. E-mail: reinhard.eher@justiz.gv.at

risk allocation is typically based on a professional's estimate about which items apply best to an individual case backgrounded by the theoretical and empirical knowledge about (re)offending behavior.

Since Paul E. Meehl's seminal work about the superiority of actuarial assessment instruments compared to unstructured clinical judgment, experience, and intuition (Meehl, 1954), a number of research articles were published basically replicating his results (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996). These findings were transferred into the field of forensic psychology and psychiatry, where the difference between standardized and structured risk assessment methods like ARAIs and SPJ instruments and alternative approaches could be shown to be even more pronounced than in other areas of psychological prediction research and practice (Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009). In the last few years a couple of meta-analyses about the predictive accuracy of risk assessment instruments were published basically confirming that standardized and structured risk assessment usually yields moderate to large effect sizes in the prediction of recidivism in different offender samples (Singh et al., 2011; Tully, Chou, & Browne, 2013; Yang, Wong, & Coid, 2010). Furthermore, no meaningful differences between ARAIs and SPJ instruments could be found (Fazel, Singh, Doll, & Grann, 2012). But both structured approaches—ARAIs and SPJ-instruments—could be found to be more reliable and to have better predictive validity than unstructured or intuitive strategies (Hanson & Morton-Bourgon, 2009).

For the actuarial assessment of violent recidivism risk in sexual offenders the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 2006) is one of the most commonly used and best validated actuarial risk assessment instruments (Hanson & Morton-Bourgon, 2009; Harris, Rice, & Quinsey, 2010). The SORAG is a modification of the Violence Risk Appraisal Guide (VRAG), an instrument which was developed to predict violent and sexual recidivism among adult male offenders (Harris, Rice, & Quinsey, 1993). The SORAG is conceptualized for sexual offenders to assess violent recidivism risk that includes sexual offenses involving physical contact with the victim (sexual hands-on recidivism). Two aspects of the outcome variable "violent recidivism" are of particular relevance: First, violent recidivism was operationalized by using criminal charges rather than new convictions. The authors claim that even though not every charge may result in a conviction, previous research had shown that charges entailed less measurement error than convictions in relation to the "true" rate of recidivistic behavior (Quinsey et al., 2006; Rice, Harris, & Hilton, 2010). They further argue that even though every official recidivism source inevitably underestimates the true recidivism rate because of unknown numbers of unreported and unrecorded offenses, it can be assumed that charges are a more sensitive measure of recidivism than convictions.

Second, the outcome for which the SORAG is validated is violent (including sexual hands-on) recidivism, but not for sexual recidivism alone, which is usually the preferred and more commonly used outcome measure in sexual offender risk assessment instruments (Hanson, 1997; Hanson & Morton-Bourgon, 2009; Hanson & Thornton, 2000). One reason for the fact that the majority of (actuarial) risk assessment instruments measure the risk of sexual rather than general violent recidivism is that sexually violent predator civil commitment statutes in the United States generally require an assessment of the risk of a sexually motivated reoffense (Doren, 2002). However, the authors of the VRAG and SORAG argue that the more common outcome measure "violent reoffense" is a more accurate index of severe sexual recidivism than "sexual reoffense" alone (Quinsey et al., 2006). Rice, Harris, Lang, and Cormier (2006) conducted an empirical investigation of this assumption by analyzing comprehensively the current and previous delinquency of sexual offenders and by comparing the results of these more detailed analyses with the officially recorded offense categories obtained from the police "rapsheets". The results indicated that a substantial number of actually sexually motivated violent offenses were not recorded as sexual on police rapsheets. Even if more research is generally needed in this area, the authors argue that the preliminary conclusion is warrantable that general violent recidivism is a more accurate outcome measure for severe sexual reoffense than the narrower category "sexual reoffense" alone (Rice et al., 2006).

Despite the existence of many empirical studies providing strong support for the predictive usefulness of the VRAG and the SORAG (e.g., Quinsey et al., 2006; Rice et al., 2010), there still remain some aspects that are criticized or controversially discussed. For example, some authors have called the cross-national stability of the absolute and relative risk indices into question (Rossegger, Gerth, Singh, & Endrass, 2013). Another, but not less important, issue is the question of whether individual risk factors included in a particular risk assessment instrument are actually all related to the outcome and whether these associations might be replicated in independent studies with different samples (Coid et al., 2011; Rettenberger, Boer, & Eher, 2011). Also, a controversial research topic is the varying performance of risk assessment instruments for different offender (sub-)groups (Bartosh, Garby, Lewis, & Gray, 2003; Harris & Rice, 2003; Rettenberger, Matthes, Boer, & Eher, 2010). For example, previous studies reported that the predictive validity of the SORAG was higher for extrafamilial child molesters compared to other sexual offender subgroups, and that the predictive accuracy for intrafamilial child molesters was relatively low in general (Bartosh et al., 2003; Rettenberger et al., 2010). Finally, the question whether risk assessment instruments (including the SORAG) would provide incremental predictive validity to each other is—scientifically and clinically—also of high relevance (Babchishin, Hanson, & Helmus, 2012).

The aim of the present study was to contribute to answering these unresolved research questions concerning the SORAG by examining its predictive validity in a large sample of prison-released sexual offenders within a prospective-longitudinal field study research design by using different statistical analysis methods. First, the predictive validity of the SORAG and its individual items were tested. In the next step of our data analyses we investigated the predictive accuracy for different subgroups of sexual offenders (child molesters vs. rapists) and for different recidivism criteria (e.g., violent vs. sexual recidivism). Then we calculated absolute and relative risk indices and compared the results with the absolute and relative risk indices reported by the Canadian researcher group who had developed the SORAG (Quinsey et al., 2006), with the aim of analyzing the stability of the SORAG risk communication procedures. Finally, the incremental predictive validity of the SORAG beyond the VRAG and the Psychopathy Checklist-Revised (PCL-R; Hare, 2003) was examined.

Method

Measures

The VRAG was constructed by using a nonpreselected sample of $N = 618$ individuals having committed at least one "serious antisocial act" in the past (Rice, Harris, & Cormier, 1992), meaning, in fact, that a violent offense was not required to be included into the construction sample. The mean follow-up period of the developmental sample was $M = 81.5$ ($SD = 60.6$) months with time institutionalized for nonviolent offenses subtracted from follow-up periods (Quinsey et al., 2006; Rice et al., 2010). For identifying risk factors an item pool was generated with about 50 potential predictor variables showing empirical and/or theoretical support for their relationship with violent reoffense. All those independent variables were assigned to either one of the following predictor categories: childhood history (e.g., elementary school maladjustment), adult adjustment (e.g., socioeconomic and marital status), index offense characteristics (e.g., number and sex of victims), and previous psychological assessment results (e.g., IQ score). Any new criminal charge for a "serious violent offense" was defined as the outcome variable. Variables from the predictor pool without a significant bivariate relationship with the outcome were not considered for further analyses. Also, those variables of highly collinear pairs were removed that had the lower correlation with the outcome. Multiple regression analyses identified 12 VRAG items showing an independent and incremental contribution to the prediction of violent recidivism: lived with both biological parents up to age 16, elementary school maladjustment, history of alcohol problems, marital status, criminal history for nonviolent offenses, failure on prior conditional release, age at index offense, victim injury, any female victim in index offense, personality disorder, schizophrenia, and the score on the PCL-R. Because the use of individual item weights led to a small but significant improvement in predictive accuracy, every single risk factor was weighted in accordance to its deviation from the general base rate using a method described by Nuffield (1982). Recently, a revised version of the VRAG, the VRAG-R, was introduced (Rice, Harris, & Lang, 2013).

Until now a number of empirical studies on the psychometric properties of the VRAG have been conducted (for current overviews see, e.g., Quinsey et al., 2006, or Rice et al., 2010). Even if the VRAG was also successfully applied to sexual offender samples (Barbaree, Seto, Langton, & Peacock, 2001; Harris & Rice, 2003), further research indicated that there might exist some additional predictors for violent recidivism particularly relevant for sexual offenders (Quinsey, Rice, & Harris, 1995). Also, since sexual offenders have base rates for violent reoffense different from those of violent offenders, sexual offender specific norms were generated (Harris & Rice, 2007). The authors slightly modified the VRAG by eliminating two items (female index victim and index victim injury) and by adding four items (prior history of violent offending, prior convictions for sexual offenses, adult female or male child victims, and sexual deviance), leading to the 14-item version of the SORAG. Table 1 gives an overview of the items, the individual item score ranges, and the bivariate correlations between the items and violent recidivism in the developmental sample (Quinsey et al., 2006). The total score of the SORAG ranges between −26 (if no risk factor is present) and + 51 (if every risk factor is present). As with the VRAG, based on the SORAG total score the evaluator can allocate the offender to one of nine risk categories. By means of these risk categories, it is possible to infer empirically calculated probabilities of violent (including sexual) recidivism after 7 and 10 years at risk (absolute risk indices). Furthermore, SORAG scores are allocated to percentile ranks, which allow users to classify the offender's individual risk within a large offender sample (Rice et al., 2010). Up to four missing items can be prorated or substituted (Quinsey et al., 2006; see below for details regarding the prorating procedure).

In the construction sample the SORAG yielded an excellent interrater reliability of .90 (Quinsey et al., 2006), which was replicated in a number of subsequent studies (Barbaree et al., 2001;

Table 1
Number, Title, Score Range, and Correlations With Violent Recidivism of the 14 SORAG Items (Quinsey et al., 2006)

Number and item title	Score range	r[b] (Quinsey et al., 2006)	r[b] (Present study)
1. Lived with both parents to age 16	5	.19	.09*
2. Elementary school maladjustment	6	.18	.23***
3. Alcohol problems	3	.07	.23***
4. Never married	3	.18	.18***
5. Nonviolent criminal history	5	.10	.26***
6. Violent criminal history	7	.05	.29***
7. Convictions for prior sex offenses	6	.17	.10**
8. History of sex offenses against girls only[a]	4	.13	.16***
9. Failure on prior conditional release	3	.13	.28***
10. Age at index offense[a]	7	.18	.12**
11. DSM-III personality disorder	5	.25	.16***
12. DSM-III schizophrenia	4	.10	.02[n.s.]
13. Phallometric test results	2	.14	−.11**
14. Psychopathy Checklist-Revised (PCL-R) score	17	.26	.21***

Note. Third column: original values reported by Quinsey et al. (2006); fourth column: values of the present study.
[a] Inversely scored item. [b] Point-biserial correlations with violent recidivism as recidivism criterion.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Bartosh et al., 2003; Ducro & Pham, 2006; Rettenberger et al., 2010). Most of the studies about the predictive validity were conducted in Anglo-American countries (e.g., Barbaree et al., 2001; Bartosh et al., 2003; Harris & Rice, 2003) and predominantly yielded indices that could be classified as a good predictive validity (i.e., area under curve [AUC] ≥ .71 derived from receiver operating characteristics [ROC]; for details see, e.g., Rice & Harris, 2005). With regard to European samples, Ducro and Pham (2006) evaluated the predictive accuracy of the SORAG in Belgian sexual offenders committed to a forensic psychiatric hospital. They reported a good predictive validity for violent recidivism (AUC = .72) and a moderate predictive validity for sexual recidivism (AUC = .64). Rettenberger and Eher (2007b) reported good predictive validity values of the German version of the SORAG for general (AUC = .73), violent (AUC = .76), and sexual (AUC = .73) recidivism. A Swiss study of the SORAG yielded moderate predictive accuracy of AUC = .69 (Rossegger et al., 2013) by using new charges and/or convictions for violent (including sexual) offenses as outcome criterion after a fixed seven years' follow-up period. Furthermore, the predictive validity of the SORAG could also be demonstrated in studies using long-term follow-up periods (Rice et al., 2013) and for different sexual offender subgroups (Bartosh et al., 2003; Ducro & Pham, 2006; Harris & Rice, 2003; Rettenberger et al., 2010).

The PCL-R score is the item with the strongest individual item weight in the VRAG and SORAG (Quinsey et al., 2006). The PCL-R consists of 20 items and is a standard tool for the forensic and clinical assessment of psychopathy (Hare, 2003). Although originally not designed for risk assessment purposes in particular, research has shown that the PCL-R performs reasonably in predicting general, sexual, and violent recidivism in both prison and forensic psychiatric populations (e.g., Quinsey et al., 1995; Salekin, Rogers, & Sewell, 1996). The PCL-R is based on a semi-structured interview and a thorough review of file information. The PCL-R total score ranges between 0 and 40 and is obtained by summing up the individual item scores. Even if psychopathy as measured by the PCL-R is conceptualized as a dimensional construct (Hare, 2003), a conventional cut-off for the diagnosis of psychopathy often used in North America is 30, whereas in Europe the cut-off commonly used is 25 (Mokros et al., 2011). A recently published meta-analysis indicated that the PCL-R is a viable instrument for the prediction of violent recidivism also in the German-speaking countries (Mokros, Vohs, & Habermeyer, 2014).

Procedure

All participants included (N = 1,104) were prison sentenced sexual offenders registered between 2001 and 2013 at the Federal Evaluation Centre for Violent and Sexual Offenders (FECVSO), a department within the Austrian Ministry of Justice (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012). In brief, since 2001 every sexual offender sentenced to an unconditional prison term because of a sexually motivated offense has to be reported to the FECVSO. The report has to be done by the correctional facility where the offender is detained. Therefore, each of the 27 correctional facilities in the Austrian Prison System has to transfer file-based information about every sexual offender to the FECVSO as soon as they receive the data from the court. The obligation of the FECVSO is to assess the risk of every reported offender, to collect data about all offenders reported, and to continuously evaluate the accuracy of its forensic, diagnostic, and risk assessments tools. A substantial proportion of all reported offenders (about 60%) is clinically interviewed for risk assessment purposes by experienced forensic psychologists and psychiatrists. These selected offenders formed the sample of the present study. The selection for clinical risk assessment follows a screening process with a file-based risk assessment comprising the following five criteria: a Static-99 (Hanson & Thornton, 2000) total score of more than five, age under 25, a prison sentence of more than four years, a conviction for a child abuse offense with a non-related victim, and a reconviction for a sexual crime (Eher et al., 2012). If a reported offender meets at least one of these criteria, he is usually selected for the clinical assessment process.

The SORAG data used in the present study was collected within this routine assessment procedure, that is, the application of the SORAG was part of the regular assessment procedure implemented in the Austrian Prison System. Results of the individual assessments are routinely used to provide therapists and decision-makers relevant case information. Therefore, the present study can be classified as a field study about the psychometric properties of risk assessment data in the context of the Austrian Prison System contributing to the current discussion about the utility of psychological assessment data in applied legal settings (e.g., Miller, Kimonis, Otto, Kline, & Wasserman, 2012; Murrie et al., 2009).

First, a German manual was written as a scoring guide for the SORAG adhering to the scoring rules of the original instrument (Quinsey et al., 2006; Rettenberger & Eher, 2007a). This guide was backtranslated and checked by one of the authors of the SORAG (M. E. R.). All clinicians (psychologists, psychiatrists, and psychotherapists) scoring the SORAG had attended at least one training workshop about how to correctly apply the German version of the SORAG. Furthermore, there had been regular peer consulting sessions at the FECVSO where difficult cases were discussed with experienced colleagues. In case of missing SORAG data at the time of data analyses, the officially proposed prorating procedure for the VRAG and SORAG was used in accordance with the manual (Quinsey et al., 2006). Prorating of missing SORAG items was performed in less than 1% of the cases and applied predominantly to item 1 or item 2 in cases where information was not available or not credible. In general, the prorating procedure for the VRAG and SORAG consists of the following steps: First, the highest possible score that could have been obtained on all available (i.e., non-missing) items was determined. Then the proportion of this highest possible score and the score an offender obtained on those items was calculated (proportion value). In the next step, the highest possible score an offender could have obtained on the missing items was determined. This score was multiplied by the proportion value. In the last step, the resulting number was added to the previously obtained total score for having the final prorated score. The items 11 (diagnosis of a personality disorder according to the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* [*DSM–III*; American Psychiatric Association, 1980]), 12 (*DSM–III* diagnosis of schizophrenia), and 13 (phallometric test results) were substituted in accordance with the official recommendations of the instrument developers (Quinsey et al., 2006): For items 11 and 12, diagnostic data derived from the *DSM–IV–TR* (American Psychiatric Association, 2000; for the German version of the *DSM–IV–TR* see Saß, Wittchen, Zaudig, & Houben, 2003) were used because this *DSM* edition was used during the time of data collection for the present study. The clinical diagnoses based on the *DSM–IV–TR* were adopted by consensus of two to three assessors at the end of the diagnostic procedure after conducting a structured clinical interview (Eher, Olver, Heurix, Schilling, & Rettenberger, 2015). Because in most European countries phallometric data are not available (Gazan, 2002; Marshall, 2006), *DSM–IV–TR* information about the diagnosis of pedophilia or sexual sadism was used to score item 13.

For calculating the incremental predictive validity of the SORAG beyond the VRAG and the PCL-R, the VRAG total score was calculated by using the relevant SORAG items which are identical to the VRAG and by retrospectively coding those items which are unique for the VRAG if the necessary information was provided by the database. Missing items were prorated (Quinsey et al., 2006). The application of the PCL-R is also an integral part of the FECVSO routine assessment procedure, and the PCL-R scores could therefore be retrieved from the FECVSO database as well (Mokros et al., 2014). As for the SORAG, all clinicians scoring the PCL-R had attended at least one official training workshop about how to correctly apply the PCL-R, and all clinicians regularly had the opportunity to attend peer consulting sessions where difficult cases could be discussed with experienced colleagues.

Data on recidivism (defined as a legally valid reconviction) were retrieved from the Federal Central Register of the Austrian Ministry of Internal Affairs. The evaluators of the reconviction data were blind for all other variables including the SORAG scores. Each new conviction listed in the rapsheets was counted as a reoffense. We used four different recidivism criteria: general criminal reconviction (i.e., each new conviction for any kind of reoffense), violent reconviction as commonly defined in the international risk assessment literature (i.e., each new conviction because of sexual and nonsexual violent reoffenses including threat and coercion; Quinsey et al., 2006), sexual reconviction (i.e., each new conviction because of a sexual hands-on or hands-off reoffense), and sexual violent reconviction (i.e., each new conviction because of a sexual offense involving physical contact or, in other words, each new conviction because of a sexual hands-on reoffense). Because Quinsey et al. (2006) used charges instead of convictions as the central outcome criterion (Rice et al., 2006), the follow-up periods in the Austrian sample were corrected by the time between charge and conviction, in order to be able to compare the absolute recidivism rates in both countries. Because the rapsheets of the Austrian Ministry of Internal Affairs do not provide any charge information but only dates of legally binding convictions, we analyzed the average time period between charge and conviction for a violent offense in a random sample of $n = 118$ sexual offenders with a prior violent offense where the relevant information about charges were available. This average time period between the date of charge and the date of the legally binding conviction was $M = 212.74$ days. Therefore, for comparing Canadian and Austrian fixed 7-year relapse rates, in the Austrian sample 7-year recidivism rates for violent recidivism were calculated after extending the 7-year follow-up period for another 213 days, because a charge would have needed another 213 days on average to lead to a final conviction.

Participants

The present study included data from $N = 1{,}104$ male sexual offenders including $n = 394$ individuals already described in the risk assessment study by Rettenberger et al. (2010) and a new sample of $n = 710$ participants. The sample of the study represented about 60% of all imprisoned sexual offenders in Austria since the implementation of the FECVSO in 2001. At the time of the collection of the recidivism data (December 17, 2012), $n = 749$ of the total sample had been released for a minimum period of at least 24 months, which was defined as the minimum follow-up period with reference to previously conducted risk assessment research (Rettenberger et al., 2011). The average follow-up period for this sample was $M = 6.48$ years ($SD = 2.26$, range = 2–14, $Mdn = 6.59$). In this latter subsample we analyzed the predictive accuracy of the SORAG and its individual items. Absolute risk indices were also calculated for the follow-up subsample ($n = 749$), whereas for identifying percentile ranks the total sample was used ($N = 1{,}104$). Further variables pertaining to the age at the time of release, duration of imprisonment, and criminal history for the follow-up subsample are presented in Table 2.

For subgroup analyses, participants were allocated either to the child molester subgroup (victims aged under 14, 48.6%, $n = 536$; in the follow-up subsample 51.5%, $n = 386$) or to the rapist subgroup (44.9%, $n = 496$; in the follow-up subsample 45.0%, $n = 337$) according to the age of all documented victims in the index offense as well as in all previous sexually motivated offenses. Allocation to the subgroups was based on an internal FECVSO coding manual that provided clear definitions of all relevant variables for the data collection process. To warrant high data quality, FECVSO staff members involved in data entry were trained in the application of the FECVSO coding manual. Mixed offender types were allocated to the subgroup that best matched their predominant victim type. Seventy-two (6.5%) participants were sexual hands-off offenders (e.g., exhibitionists or child pornography offenders), sexual murderers, or sexually motivated offenders not convicted for a sexual crime (e.g., sexual burglary, which is not an official statutory offense in the Austrian penal law; therefore, sexual burglary is judicially regarded as a conventional, i.e., nonsexual, burglary). These offenders were not allocated to either of the subgroups and were therefore excluded from subsample analyses. A substantial part of the sample but not all offenders had been assigned to at least one treatment regime during their prison sentence, after (conditional) release, or both. It can be estimated from previous experience and unstructured data collection that at least two thirds of the total sample were treated or supervised in some form.

Table 2
Sample Characteristics of the Follow-Up Subsample ($n = 749$)

	M	*SD*	Range
Age at time of release (years)	41.04	12.88	16.13–74.70
Duration of imprisonment (years)	2.67	2.62	.06–19.62
Number of prior offenses	3.55	5.88	0–38
Number of prior violent offenses	1.25	2.71	0–23
Number of prior sexual offenses (before the index offense)	.37	1.32	0–17

Statistical Analysis

The predictive accuracy of the SORAG was measured by using different statistical analysis methods: AUC values were derived from ROC curves (Hanley & McNeil, 1982), odds ratios (ORs) were calculated using logistic regression analyses (Hanson, Helmus, & Thornton, 2010; Hosmer & Lemeshow, 2000), and rate ratios (RRs) by Cox regression analyses (Allison, 1984). The incremental predictive validity was examined by using sequential Cox regression models (Hunsley & Meyer, 2003). For testing the relationship between the individual SORAG items and recidivism, point-biserial correlations were calculated between each item and violent recidivism (Quinsey et al., 2006). To examine the interrater reliability, intraclass correlation coefficients (ICC) were calculated.

AUC values are statistical indices commonly used to examine the predictive accuracy of binary decisions (Mossman, 2013). Referring to Cohen (1992), Rice and Harris (2005) proposed the following interpretation criteria for AUC values: AUC $\geq$ .72 can be regarded as "good", AUC = .64–.71 are classified as "moderate", and significant AUC $\leq$ .63 are classified as "small." Differences between AUC values were tested in a pairwise manner for significance by using a z-statistic described in DeLong, DeLong, and Clarke-Pearson (1988). Given the limitations of AUC values for the measurement of predictive accuracy, some authors propose to additionally use odds ratios (OR) for reporting predictive accuracy (Hanson et al., 2010). ORs are calculated by logistic regression analyses (Hosmer & Lemeshow, 2000) by using equal follow-up periods. In order to allow the integration of all cases (which means to use unequal follow-up periods), Cox regression analyses were additionally calculated because this method controls for unequal follow-up periods (Allison, 1984). The hazard ratio (exp[B]) resulting from Cox regression analyses can be equaled with the risk ratio (RR). Contrary to the OR, it is a measure for the relationship between the probabilities of two groups rather than the odds of two groups. The calibration of the SORAG was examined using the *E/O* index (Gail & Pfeiffer, 2005) by contrasting the observed number of events (the corrected 7-years recidivism rates) with the number of events predicted by the risk assessment tool after 7 years reported in the developmental study (Quinsey et al., 2006; Viallon, Ragusa, Clavel-Chapelon, & Benichou, 2009). The *E/O* index is defined as the ratio of the predicted recidivists divided by the observed recidivists (Helmus, Thornton, Hanson, & Babchishin, 2012) and is an appropriate calibration measure only when fixed follow-up data are available. We also calculated 95% confidence intervals for the *E/O* index and an overall significance test by using the traditional χ^2-goodness of fit statistic as reported in Helmus, Hanson, Thornton, Babchishin, and Harris (2012). For comparing the relative risk indices derived from the present study with the normative data distribution of the original SORAG development study (Quinsey et al., 2006), a two-sample Kolmogorov–Smirnov test was conducted. The statistical analyses were calculated using the formulae in the above-mentioned publications (e.g., Gail & Pfeiffer, 2005) and by using IBM Statistical Package for Social Sciences (SPSS) version 21.0.0.1. For analyzing the difference between independent AUC values MedCalc Version 9.3.8.0 was used.

Results

Reliability, Scores, and Recidivism Rates

The interrater reliability of the SORAG was examined by using the SORAG scores of two independent clinicians for $n = 70$ cases derived from the total sample of the present study. Both independent raters had access to the files and both had conducted an independent clinical interview with the offender. Both raters were blind to the results of the clinical assessment and the SORAG ratings of each other. The clinicians who rated the cases had previously attended a rigorous training in the application of the SORAG. In accordance with the classification system proposed by McGraw and Wong (1996), the intraclass correlation coefficient (ICC[A,1]; random effects, single measure, absolute agreement) was ICC = .96 ($p < .001$, 95% CI = .94–.98), which can be regarded as an excellent value (Fleiss, 1981; Hart & Boer, 2010).

The average SORAG score for the total sample was $M = 10.39$ ($SD = 13.97$, range = −20–43). Rapists ($M = 14.98$, $SD = 13.22$, range = −13–43) exhibited significantly higher SORAG scores than child molesters ($M = 5.46$, $SD = 13.17$, range = −20–38), $t(1030) = 11.58$, $p < .001$, Cohen's $d = .72$. The difference of the SORAG total scores between the follow-up ($n = 749$, $M = 9.53$, $SD = 13.93$, range = −20–43) and the non-follow-up sample ($n = 355$, $M = 12.22$, $SD = 13.90$, range = −18–42) was small but still significant, $t(1102) = 3.00$, $p < .01$, Cohen's $d = .19$. The reconviction rates for the follow-up subsample ($n = 749$) as well as for child molesters ($n = 386$) and rapists ($n = 337$) separately for unequal and fixed 5-year follow-up periods are presented in Table 3. The average VRAG score for the total sample ($N = 1{,}104$) was $M = 5.33$ ($SD = 12.28$, range = −22–38), the average total score of the PCL-R ($n = 1{,}082$) was $M = 20.94$ ($SD = 7.60$, range = 1–40).

Predictive Validity

The predictive validity indices of the SORAG total score examined by using the follow-up subsample ($n = 749$) are presented in Table 4 and Table 5. For the prediction of violent recidivism—the outcome criterion the SORAG was originally designed for—the instrument yielded an effect size of AUC = .74 ($p < .001$, 95% CI = .70–.78) which can be regarded as "good" according to the internationally used conventional rules for the interpretation of AUC values (Rice & Harris, 2005). The predictive accuracy for general and violent recidivism was significantly higher than for sexual ($z = 2.36$ and $z = 2.29$, respectively) and sexual hands-on ($z = 2.51$ and $z = 2.45$) recidivism (all $p < .05$), whereas the AUC values for general and violent as well as for sexual and sexual hands-on recidivism did not differ significantly. Also, the effect sizes in general were somewhat higher for child molesters than for rapists. However, the differences between the AUC values of both subgroups were only significant for general recidivism ($z = 2.48$, $p = .001$) and failed statistical significance for violent ($z = .41$, $p = .681$), sexual ($z = .67$, $p = .501$), and sexual hands-on recidivism ($z = .44$, $p = .660$). Table 5 shows the predictive accuracy of the SORAG total scores for the total sample as well as for both subsamples by using logistic and Cox regression analyses. Both methods for calculating effect sizes yielded a comparable result pattern: For the total sample as well as for both subgroups

Table 3
Reconviction Rates for Unequal and Fixed 5-Year Follow-Up Periods

	Recidivism criterion			
	General criminal recidivism, % (*n*)	Violent recidivism, % (*n*)	Sexual recidivism, % (*n*)	Sexual hands-on recidivism, % (*n*)
Total sample				
Unequal follow-up ($N = 749$)	36.3 (272)	23.5 (176)	11.3 (85)	8.0 (60)
Fixed 5-year follow-up ($n = 586$)	40.6 (238)	24.1 (141)	10.8 (63)	7.5 (44)
Child molesters				
Unequal follow-up ($n = 386$)	28.0 (108)	15.3 (59)	11.7 (45)	8.0 (31)
Fixed 5-year follow-up ($n = 299$)	31.1 (93)	15.7 (47)	11.7 (35)	7.7 (23)
Rapists				
Unequal follow-up ($n = 337$)	45.4 (153)	33.8 (114)	9.5 (32)	8.6 (29)
Fixed 5-year follow-up ($n = 271$)	50.2 (136)	33.9 (92)	8.1 (22)	7.7 (21)

the SORAG total scores significantly predicted all four recidivism outcomes. The contribution of each individual item contained in the SORAG for predicting violent recidivism is presented in Table 1. With the exception of item 12 (schizophrenia) and item 13 (sexual deviance) every item showed a significant positive relationship with the main outcome variable (violent recidivism).

Normative Data: Relative and Absolute Risk Indices

In the next step of our analyses, relative (percentile ranks and relative risk ratios) and absolute (probabilities of recidivism risk for the different SORAG risk categories) risk indices were calculated and compared with the risk indices provided by Quinsey et al. (2006) in the original study. A Kolmogorov–Smirnov test revealed that both percentile distributions, one derived from the Canadian original dataset (Quinsey et al., 2006) and the other one from the sample of the present study, did not differ significantly from each other ($D_{max} = 0.133$, $p = .819$; for the percentiles of both samples see Table 1 of the online supplemental material).

In Table 6, comparisons of absolute risk indices for a 7-year follow-up period as a function of the nine SORAG risk categories between the Canadian original study (Quinsey et al., 2006) and the Austrian dataset are presented. For this purpose we used the corrected 7-year follow-up period for our sample as described above. The corrected 7-year recidivism rate for violent recidivism was 28.7% ($n = 77$). For the Austrian dataset, observed and estimated (predicted) recidivism rates are presented, the latter calculated by logistic regression analysis. The Hosmer-Lemeshow goodness of fit test indicated that the prediction model provides an adequate data fit, $\chi^2 = 8.034$, $df = 6$, $p = .236$.

In Table 7, results are presented for the comparison between absolute risk indices of the Canadian and Austrian sample by calculating *E/O* indices and corresponding 95% confidence intervals for each SORAG risk category. The overall goodness of fit statistic showed no overall significant difference between the original recidivism rates per SORAG risk category reported by Quinsey et al. (2006; expected values) and the recidivism rates derived from the dataset of the present study (observed values; $\chi^2 = 13.86$, $df = 8$, $p = .086$). On close inspection, the results nevertheless revealed that recidivism rates derived in the present study were lower in general (*E/O* indices > 1) than the rates of the Canadian dataset. However, all of the 95% confidence intervals include 1.0 indicating no significant differences between the observed (Austrian) and expected (Canadian) recidivism rates. For applied assessment purposes, Relative Risk Ratios (RRR; Hanson, Babchishin, Helmus, & Thornton, 2013) and absolute risk indices for violent recidivism (defined as reconvictions) for 3- and 5-year follow periods are provided in the online supplemental material (see Tables 2, 3, and 4 of the online supplemental material).

The Incremental Predictive Validity

As expected, the SORAG score was highly intercorrelated with the VRAG and PCL-R total scores: The correlation was $r = .95$ ($p < .001$) between the SORAG and the VRAG scores and $r = .78$ ($p < .001$) between the SORAG and the PCL-R scores. The VRAG scores correlated also highly with the PCL-R scores, $r =$

Table 4
The Predictive Accuracy of the Sex Offender Risk Appraisal Guide (SORAG) Total Score for the Total Sample, Both Subsamples, and Different Recidivism Criteria Using ROC-Analyses

	General recidivism AUC [95% CI]	Violent recidivism AUC [95% CI]	Sexual recidivism AUC [95% CI]	Sexual hands-on recidivism AUC [95% CI]
Total sample ($n = 749$)	$.74^{***}$ [.70–.77]	$.74^{***}$ [.70–.78]	$.66^{***}$ [.61–.72]	$.64^{***}$ [.57–.71]
Child molesters ($n = 386$)	$.77^{***}$ [.71–.82]	$.72^{***}$ [.65–.80]	$.69^{***}$ [.61–.77]	$.66^{***}$ [.56–.76]
Rapists ($n = 337$)	$.67^{***}$ [.61–.72]	$.70^{***}$ [.65–.76]	$.65^{**}$ [.56–.73]	$.63^{*}$ [.54–.71]

Note. AUC = area under the receiver operating characteristic (ROC) curves; CI = confidence interval.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
The Predictive Accuracy of the Sex Offender Risk Appraisal Guide (SORAG) Total Score for the Total Sample, Both Subsamples, and Different Recidivism Criteria Using Logistic and Cox Regression Analyses

	Recidivism criterion			
	General criminal recidivism *RR/OR* [95% CI]	Violent recidivism *RR/OR* [95% CI]	General sexual recidivism *RR/OR* [95% CI]	Sexual hands-on recidivism *RR/OR* [95% CI]
Cox regression[a]				
Total sample (n = 749)	1.06*** [1.04–1.08]	1.06*** [1.05–1.07]	1.05*** [1.03–1.06]	1.04*** [1.02–1.06]
Child molesters (n = 386)	1.06*** [1.05–1.07]	1.07*** [1.05–1.09]	1.05*** [1.03–1.08]	1.05*** [1.02–1.08]
Rapists (n = 337)	1.07*** [1.05–1.09]	1.06*** [1.04–1.07]	1.05*** [1.02–1.08]	1.04** [1.01–1.07]
Logistic regression[b]				
Total sample (n = 586)	1.09*** [1.07–1.10]	1.08*** [1.06–1.10]	1.05** [1.03–1.08]	1.05** [1.03–1.08]
Child molesters (n = 299)	1.10*** [1.07–1.12]	1.08*** [1.05–1.11]	1.07*** [1.04–1.11]	1.07*** [1.04–1.11]
Rapists (n = 271)	1.07*** [1.05–1.10]	1.08*** [1.05–1.10]	1.05** [1.01–1.09]	1.05* [1.01–1.β0]

Note. *RR* = rate ratios derived from Cox regression analyses, *OR* = odds ratios derived from logistic regression analyses, CI = confidence interval.
[a] Cox regression analyses consider unequal follow-up periods. [b] Logistic regression analyses use fixed 5-year follow-up periods.
* $p < .05$. ** $p < .01$. *** $p < .001$.

.77 ($p < .001$). Effect sizes of the predictive validity of the VRAG (N = 1,104) and the PCL-R (n = 1,082; including validity indices for both factor scores and the four facet scores separately) are shown in the online supplemental material (see Table 5 of the online supplemental material). The differences between the AUC values of the SORAG and the VRAG were not significant. The predictive validity indices of the VRAG and the PCL-R did not differ significantly either. Only the difference between the SORAG and the PCL-R total scores significantly differed for general ($z = 2.47$, $p < .05$), violent ($z = 2.51$, $p < .05$), and sexual recidivism ($z = 2.14$, $p < .05$).

In order to examine the incremental predictive validity of the SORAG beyond the PCL-R and the VRAG, series of sequential Cox regression analyses were conducted with the SORAG, VRAG, and PCL-R scores as independent variables, and the four dichotomous recidivism criteria (general criminal, violent, sexual, and sexual hands-on recidivism) as the dependent variable. Following the preassumption that the SORAG—as it was specifically designed for predicting violent reoffense in sexual offenders—will add to the predictive accuracy beyond psychopathy (measured by the PCL-R) and beyond a risk assessment instrument not specifically developed for sexual offenders (VRAG), the PCL-R was first entered into the model followed by the VRAG. In the last step, the SORAG was included. As demonstrated in Table 8, the result pattern was comparable for all four recidivism criteria: Adding the VRAG significantly improved the prediction compared to the performance of the PCL-R alone. Including the SORAG further improved the model. However, once the SORAG was entered, it remained the only significant predictor.[1]

Discussion

The main aims of the present study were, first, to examine the psychometric properties of the German version of the SORAG and its individual items and, second, to compare absolute and relative risk indices between the original Canadian dataset (Quinsey et al., 2006) and a large sample of sexual offenders derived from the Austrian Prison System. In yielding an AUC value of .74 for the prediction of violent recidivism, which could be classified as good predictive validity for the outcome that the instrument was originally designed for (Rice & Harris, 2005), the SORAG exhibited an effect size comparable with that found for the developmental study (Quinsey et al., 2006). Also comparable with former studies, indices for the prediction of overall sexual and sexual hands-on recidivism were lower but still significant (Rettenberger et al., 2010). The finding that the German version of the SORAG exhibited lower predictive accuracy for sexuality-related outcome criteria could be interpreted as an indicator for its outcome specificity, because the SORAG was designed for predicting violent rather than sexual reoffense (Rice et al., 2006, 2013).

Predictive validity results of the SORAG scores separately for rapists and child molesters provided further support for its outcome specificity: in both subsamples the SORAG yielded larger effect sizes for the prediction of violent reoffense than for sexual reoffense alone. In contrast to previous studies on the differential effects of risk assessment instruments (Bartosh et al., 2003; Rettenberger et al., 2010), the results of the present study might be interpreted as an argument for the stability of the SORAG across different sexual offender subgroups. This result is in line with a recently published study by Harris and Rice (2013) investigating whether Bayesian adjustments due to subgroup differences would lead to a difference in the predictive accuracy of the VRAG. They found that the allocation to different offender subgroups (e.g., older vs. younger offenders or offenders with serious vs. less severe index offense) did not provide additional information relevant to the predictive performance of the instrument. In other words, the VRAG yielded high stability of predictive validity across different subgroups. Compared with the above-mentioned previous studies (Bartosh et al., 2003; Rettenberger et al., 2010), the present study and the study of Harris and Rice (2013) used substantially larger and more representative samples as well as

[1] This result pattern remains the same even when the order of entering the variables was changed and the VRAG was included after the SORAG, which was entered into the model in the second step after the PCL-R. Also in this case, the SORAG remained the only significant predictor. For interested readers, more detailed results are available on request from the first author.

Table 6
Absolute Violent 7-Year Recidivism Rates as a Function of the Nine Sex Offender Risk Appraisal Guide (SORAG) Risk Categories Derived From the Canadian Developmental Study (Quinsey et al., 2006) and the Austrian Dataset of the Present Study (Corrected 7-Years Relapse Rates; See Procedure)

SORAG Risk	SORAG Total score range	Original violent recidivism rate (Quinsey et al., 2006)	Observed violent recidivism rate (Present study)	Estimated[a] violent recidivism rate (Present study)
1	≤ −10	7%	7.4%	9.2%
2	−9 to −4	15%	11.4%	12.9%
3	−3 to + 2	23%	18.8%	17.8%
4	+ 3 to + 8	39%	19.4%	24.0%
5	+ 9 to + 14	45%	35.7%	31.6%
6	+ 15 to + 19	58%	43.8%	40.2%
7	+ 20 to + 24	58%	33.3%	49.6%
8	+ 25 to + 30	75%	53.3%	59.0%
9	≥ +31	100%	66.7%	67.7%

[a] Estimated rates were derived by logistic regression analysis.

longer follow-up periods. Therefore, one could hypothesize that the previously observed subgroup differences might be the result of random effects caused by relatively short follow-up periods and small sample sizes and therefore would decrease with increasing follow-up periods and (sub-)sample sizes.

Applying different methods for identifying effect sizes demonstrated a stable result pattern: Independently of which effect size (AUC values, odds ratios, or hazard ratios) was considered, the results remained the same. These findings support those of previous studies and provide further support for the stability of AUC values as an index for the predictive accuracy of risk assessment instruments even in smaller samples (Hanczar et al., 2010). The lack of stability of AUC values reported in previous investigations (e.g., Eher, Rettenberger, Schilling, & Pfäfflin, 2008) might be the result of small samples with comparably short follow-up periods. Hence, the main source of variability might not be attributed to the AUC index per se but rather to the study design and/or the data collection process.

In consideration of the high correlation between the SORAG and the VRAG, it might be prima facie a surprising finding that the SORAG provide nonetheless an incremental predictive contribution beyond the VRAG. However, Babchishin et al. (2012) showed that even highly correlated risk assessment instruments can add incremental validity to each other for the prediction of recidivism in sexual offenders. They conducted a meta-analysis of 20 samples stemming from different countries and jurisdictions ($N = 7{,}491$) and compared the predictive and incremental validity of the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997), the Static-99R (Hanson & Thornton, 2000), and the Static-2002R (Hanson et al., 2010)—three actuarial risk assessment instruments, which were very closely related. Despite a substantial item overlap and large intercorrelations up to $r = .92$, these instruments consistently added incremental predictive accuracy to one another. Given the results of the study published by Babchishin et al. (2012) and because of the fact that psychopathy as measured by the PCL-R might be quite as good (or even better) as customary risk assessment instruments in predicting violence (Salekin et al., 1996), we were interested in the (incremental) predictive validity of the SORAG and VRAG beyond the PCL-R. As mentioned above, the VRAG is an actuarial risk assessment instrument originally designed for the same outcome as the SORAG (violent recidivism) but not specifically for sexual offend-

Table 7
Comparison of the Absolute Violent Recidivism Risk Indices Between the Original Study (Quinsey et al., 2006) and the Present Study (Corrected 7-Years Relapse Rates; See Procedure)

SORAG Category	Original violent recidivism data (Quinsey et al., 2006)		Observed violent recidivism data (Present study)		Expected number of recidivists	Fit between original and observed rates	
	N	Recidivists	*N*	Recidivists	*N*	*E/O* Index	95% CI
1	14 (.07)	.98	27 (7.4%)	2	1.89	.95	[.24–3.80]
2	23 (.15)	3.45	44 (11.4%)	5	6.60	1.32	[.55–3.17]
3	40 (.23)	9.20	32 (18.8%)	6	7.36	1.23	[.55–2.73]
4	58 (.39)	22.62	36 (19.4%)	7	14.04	2.01	[.96–4.21]
5	52 (.45)	23.40	28 (35.7%)	10	12.60	1.26	[.68–2.34]
6	46 (.58)	26.68	32 (43.8%)	14	18.56	1.33	[.79–2.24]
7	32 (.58)	18.56	33 (33.3%)	11	19.14	1.69	[.93–3.05]
8	18 (.75)	13.50	15 (53.3%)	8	11.25	1.74	[.96–3.14]
9	5 (1.00)	5.00	21 (66.7%)	14	21.00	1.50	[.89–2.53]

Note. In brackets the violent recidivism rates, see Table 6.

Table 8
The Incremental Contribution of the SORAG Scores Beyond the VRAG and PCL-R Scores for the Prediction of General Criminal, Violent, Sexual, and Sexual Hands-On Recidivism Using Cox Regression Analyses

	χ^2 change			Regression coefficient				Rate ratio	
	change	*df*	*p*	b	*SE*	Wald	*p*	Exp(B)	95% CI
General criminal recidivism									
Step 1									
PCL-R	71.12	1	.000	.068	.008	69.11	.000	1.07	[1.05–1.09]
Step 2									
PCL-R				−.007	.012	.29	.588	.99	[.97–1.02]
VRAG	138.57	2	.000	.063	.008	61.42	.000	1.07	[1.05–1.08]
Step 3									
PCL-R				−.022	.013	2.87	.090	.98	[.95–1.00]
VRAG				.005	.017	.09	.762	1.01	[.97–1.04]
SORAG	153.07	3	.000	.061	.015	15.36	.000	1.06	[1.03–1.10]
Violent recidivism									
Step 1									
PCL-R	54.63	1	.000	.074	.010	52.74	.000	1.08	[1.06–1.10]
Step 2									
PCL-R				−.007	.016	.20	.653	.99	[.96–1.02]
VRAG	103.61	2	.000	.067	.010	44.17	.000	1.07	[1.05–1.09]
Step 3									
PCL-R				−.025	.017	2.26	.133	.98	[.94–1.01]
VRAG				−.001	.021	.00	.959	1.00	[.96–1.04]
SORAG	116.02	3	.000	.072	.019	13.92	.000	1.07	[1.04–1.12]
Sexual recidivism									
Step 1									
PCL-R	6.62	1	.010	.036	.014	6.55	.010	1.04	[1.01–1.07]
Step 2									
PCL-R				−.028	.022	1.64	.201	.97	[.93–1.02]
VRAG	21.17	2	.000	.052	.014	13.79	.000	1.05	[1.03–1.08]
Step 3									
PCL-R				−.069	.023	8.62	.003	.93	[.89–.98]
VRAG				−.098	.029	11.21	.001	.91	[.86–.96]
SORAG	54.29	3	.000	.160	.028	33.61	.000	1.17	[1.11–1.24]
Sexual hands-on recidivism									
Step 1									
PCL-R	4.97	1	.026	.037	.017	4.92	.027	1.04	[1.00–1.07]
Step 2									
PCL-R				−.016	.026	.38	.536	.98	[.94–1.04]
VRAG	12.24	2	.002	.044	.017	6.90	.009	1.05	[1.01–1.08]
Step 3									
PCL-R				−.047	.028	2.97	.085	.95	[.90–1.01]
VRAG				−.081	.035	5.57	.018	.92	[.86–.99]
SORAG	29.35	3	.000	.131	.032	17.24	.000	1.14	[1.07–1.21]

Note. N = 1,104. Values show incremental contribution of the VRAG scores beyond the PCL-R scores and the incremental contribution of the SORAG scores after controlling for the previously entered PCL-R and VRAG scores.

ers. Our assumption was that even if the SORAG was closely related to the VRAG, the SORAG would add to the predictive accuracy beyond the PCL-R and the VRAG because it was specifically designed for the prediction of recidivism in sexual offenders and consisted therefore of a few additional items that were particularly relevant for the assessment of reoffense risk in sexual offenders. This assumption was confirmed by the results of the incremental validity analyses: For all recidivism criteria investigated, the inclusion of the SORAG led to an improvement of the prediction model. These results, however, emphasize that risk assessment instruments including psychopathic traits as one item—at least in sexual offenders—are superior in the prediction of recidivism compared to a psychopathy measure alone. From the perspective of the general psychological assessment methodology this finding is not surprising: Since risk assessment instruments are measures specifically designed to predict recidivism, it seems to be self-evident that they yield better results than the PCL-R, which was originally not developed as a risk assessment measure but rather as a checklist for characterizing a personality trait (disorder). Also, current empirical results about the predictive validity of the PCL-R confirm that only a subset of its items is actually related to recidivism (Rice et al., 2013; Yang et al., 2010) and that the predictive accuracy of the instrument is generally lower than the predictive accuracy of conventional risk assessment instruments (Singh et al., 2011).

A more detailed inspection of the predictive accuracy of the individual items of the German version of the SORAG revealed that 12 out of 14 items were found to be significantly related to violent recidivism. This result is in line with the findings of previous studies which have usually exhibited comparable significant bivariate relationships between items and recidivism in empirically derived risk assessment tools even when replicated in independent cross-validation studies (Coid et al., 2011; Helmus & Thornton, 2015). However, in the present study two items were either not predictive or even negatively related to the outcome: item 12 (schizophrenia) and item 13 (sexual deviance). In the original versions of both the SORAG and the VRAG, item 12 (schizophrenia) was inversely coded, that is, that in the developmental study the diagnosis of schizophrenia was associated with a lower recidivism rate compared to the group of offenders without the diagnosis of schizophrenia (Quinsey et al., 2006). Item 13 (sexual deviance), however, was conceptualized like the other items as a risk factor, which means that the presence of sexual deviance was associated with a higher recidivism risk compared to the group of offenders without sexual deviant interests. Interestingly, item 12 (schizophrenia) was found not to predict violent recidivism above chance level also in a previous study (Coid et al., 2011). However, the meta-analytic literature on the relationship between psychosis and violence indicated that psychosis was significantly associated with an increase in the odds of violence but there was a substantial dispersion among effect sizes between different studies which was attributable in part to methodological factors (Douglas, Guy, & Hart, 2009). One reason for the failure of predictive accuracy of item 12 (schizophrenia) might be that both samples—Coid et al. (2011) and the present one—were prison samples with offenders suffering from psychotic disorders much less prevalent than in forensic psychiatric units (Eher, Rettenberger, & Schilling, 2010; Harsch, Bergk, Steinert, Keller, & Jockusch, 2006). As a consequence, the failure of an association between item and outcome might be related to an extremely low prevalence. In an overlapping sample psychotic disorders were found only in about 1% (Eher et al., 2010), which was substantially lower than the prevalence rate in the Canadian developmental study of the SORAG, which was about 16% (Rice et al., 2013).

A small but significantly negative relationship between item 13 (sexual deviance) and violent recidivism seemed to be unexpected given the well-known relationship between sexual deviance and sexual recidivism risk in sexual offenders (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). However, taking a closer look at the results of these meta-analyses, the SORAG development study (Quinsey et al., 2006), and the data collection process of the present study, there exist possible explanations: First, because in most European countries the application of phallometry is either not common or legally not allowed (Gazan, 2002; Marshall, 2006), *DSM–IV–TR* information was used to score item 13 instead of phallometric test results. However, Hanson and Bussière (1998) and Quinsey et al. (2006) found a strong relationship with recidivism explicitly by using phallometry rather than other diagnostic sources for sexual deviance. Even if nowadays a number of different approaches exist to measure sexual deviant interests, the research question about their reliability and validity remains still unanswered (Kalmus & Beech, 2005). Item 13 explicitly pertains to sexual sadism or to pedophilia. However, *DSM* diagnosed sadism recently was found not to predict sexual or violent recidivism (Eher et al., 2016). Also, *DSM* diagnosed pedophilia was not only reported to have limited clinical utility but was also found to lack predictive validity (Eher et al., 2015; First & Frances, 2008; Kingston, Firestone, Moulden, & Bradford, 2007). Of course, most of these findings might also be true for phallometric testing as well (Marshall, 2006), but its predictive relevance seems to be better grounded compared to other diagnostic approaches (Quinsey & Lalumière, 1996). A further reason for the lack of predictive accuracy of this item might be the fact that risk factors capturing sexual deviant interests are more relevant for the prediction of sexual recidivism rather than violent or general criminal recidivism (Eher et al., 2016). Support for this explanation can also be derived from a study about the predictive accuracy of the individual risk factors of the SVR-20 (Rettenberger et al., 2011): Using a mixed sample of adult sexual offenders ($N = 493$), item 1 (sexual deviance) of the SVR-20 was highly predictive for sexual recidivism (AUC = .67) but had no relevance for the prediction of violent (AUC = .42) and general criminal (AUC = .44) recidivism.[2] Interestingly, in the course of the revision process of the VRAG and SORAG, items specifically designed for capturing sexual deviance—like the original item 13 of the SORAG—were no longer included in the revised version (Rice et al., 2013).

The comparison of the relative (i.e., percentile ranks) and absolute (i.e., recidivism rates) risk indices between the original Canadian development study (Quinsey et al., 2006) and the present investigation yielded mixed results: On the one hand, the results indicated that both risk indices of the SORAG do not differ significantly between both countries, contradicting a former study with a substantially smaller sample from Switzerland (Rossegger et al., 2013). On the other hand, a more “clinical” inspection of the normative data distribution yielded more (with regard to the absolute risk indices) or less (with regard to the relative risk indices) meaningful differences between both countries. In detail, even if the 95% confidence interval of the *E/O* indices of the nine SORAG risk category indicated no significant differences between the observed (the Austrian) and the expected (the Canadian) recidivism rates, the actual difference between 75% and 53% (see SORAG risk category 8 as an example) is certainly of clinical and juridical relevance. Therefore, we would recommend—in accordance with recently published meta-analytic investigations of the variability of absolute risk indices derived from actuarial risk assessment instruments (Helmus et al., 2012)—to specifically collect normative data for each jurisdiction and country where the instrument should be used in applied risk assessment settings.

Findings of previous studies about less cross-cultural variability of relative risk indices (percentile ranks, relative risk ratios; Hanson et al., 2013; Hanson, Lloyd, Helmus, & Thornton, 2012) compared with absolute risk indices were supported by the results of the present study: we found no statistical difference between the Canadian and Austrian normative distributions, which was con-

[2] In the present sample item 13 (sexual deviance) showed also a significant positive correlation with sexual recidivism ($r = .10$, $p < .01$), which provides further support for the above-mentioned assumption.

firmed by a closer inspection of the two percentile ranking distribution. Nevertheless, the relative risk indices are not identical, so we would recommend using relative normative data tables based on an offender population for the jurisdiction and country where the offender lives.

What do the results of the present study mean for clinical practice and for applied risk assessment settings? First, the results show once again that the SORAG yielded large effect sizes for the ability to discriminate between recidivists and nonrecidivists. Therefore, the use of the SORAG in clinical practice and applied risk assessment settings is clearly supported by the results of the present study. However, the results showed also a few opportunities to improve the predictive accuracy of the instrument by revising (or deleting) items that were only weakly related to the outcome in their current form. Therefore, evaluators should be informed about revision efforts of the instrument that they have decided to use regularly in clinical practice. Second, evaluators from countries and jurisdictions outside North America must be aware that the absolute and relative risk indices originally published in the SORAG developmental study (Quinsey et al., 2006) should be very cautiously used in other countries and jurisdictions. The present results indicate that there might be relevant differences between the Canadian data and normative data from other countries. This result underscores the clinical and scientific importance of supporting every effort to gather own normative recidivism data for each country and jurisdiction where the instrument is used.

There are at least two major limitations of the present study that have to be addressed: First, there was only one recidivism data source available, and the conviction data gathered by using the computerized database of the Austrian Ministry of Internal Affairs could not be treated as equivalent to the comprehensive outcome information that was used by Quinsey et al. (2006; see also Rice et al., 2006, 2013). Also, when comparing absolute recidivism rates across studies, we had to adjust the follow-up periods because of the different outcome criteria (charges vs. convictions). This adjustment approach is only an approximation and should not blind us that the studies used different definitions for relapse. Second, the follow-up periods in our study were substantially shorter than the timeframes used by Rice et al. (2013) investigating the long-term predictive accuracy of the VRAG and SORAG. However, the prospective-longitudinal field study research design of the present study precluded the opportunity to extend the follow-up periods. At the same time the research design can be regarded as a clear strength of the present study given that prospective designs are usually deemed to be the best available frames for prediction and risk assessment studies (Hanson & Bussière, 1998).

Relevant research issues for future studies would be, for example, the question of whether the findings of the recently introduced revised versions of the VRAG and SORAG (Rice et al., 2013) could be replicated by using international samples. Another direction for future research would be to provide additional statistical evidence for the cross-cultural transferability of the SORAG by assessing measurement invariance by means of item response theory (Bolt, 2007) or multiple-group confirmatory factor analysis (Mokros et al., 2011). A further and still controversially discussed research question refers to the question whether treatment-related and/or maturational information contributes incrementally to the prediction of recidivism beyond what is captured by actuarial risk instruments like the SORAG alone (e.g., Hanson, Harris, Scott, & Helmus, 2007; Olver & Wong, 2011). Previous findings of our research group indicated that, on the one hand, different age-related items provide no significant incremental predictive accuracy beyond ARAIs (Rettenberger, Haubner-MacLean, & Eher, 2013). On the other hand, actuarial risk assessment instruments capturing dynamic (i.e., changeable) risk factors like the Stable-2007 (Hanson et al., 2007) are able to improve the predictive validity of instruments that consist mainly of static (i.e., biographical or historical) risk factors (Eher et al., 2012).

References

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage. http://dx.doi.org/10.4135/9781412984195

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment, 19,* 442–461. http://dx.doi.org/10.1177/1073191112458312

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,* 490–521. http://dx.doi.org/10.1177/009385480102800406

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,* 422–438. http://dx.doi.org/10.1177/0306624X03253850

Boer, D. P., 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: Mental Health, Law, and Policy Institute.

Bolt, D. M. (2007). Analyzing the Psychopathy Checklist-Revised (PCL-R) using item response theory and factor analysis: Overview and recent advances. In H. Hervé & J. C. Yuille (Eds.), *The psychopath: Theory, research, and practice* (pp. 105–139). Mahwah, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Coid, J. W., Yang, M., Ullrich, S., Zhang, T. Q., Sizmur, S., Farrington, D., & Rogers, R. (2011). Most items in structured risk assessment instruments do not predict violence. *Journal of Forensic Psychiatry & Psychology, 22,* 3–21. http://dx.doi.org/10.1080/14789949.2010.495990

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674. http://dx.doi.org/10.1126/science.2648573

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44,* 837–845. http://dx.doi.org/10.2307/2531595

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412990059

Douglas, K. S., Guy, L. S., & Hart, S. D. (2009). Psychosis as a risk factor for violence to others: A meta-analysis. *Psychological Bulletin, 135,* 679–706. http://dx.doi.org/10.1037/a0016311

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

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. *Sexual Abuse, 24,* 5–28.

Eher, R., Olver, M. E., Heurix, I., Schilling, F., & Rettenberger, M. (2015). Predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment. *Law and Human Behavior, 39,* 571–580. http://dx.doi.org/10.1037/lhb0000144

Eher, R., Rettenberger, M., & Schilling, F. (2010). Psychiatrische Diagnosen von Sexualstraftätern: Eine empirische Untersuchung von 807 inhaftierten Kindesmissbrauchstätern und Vergewaltigern [Psychiatric diagnoses in sexual offenders: An investigation of 807 incarcerated child molesters and rapists]. *Zeitschrift für Sexualforschung, 23,* 23–35. http://dx.doi.org/10.1055/s-0030-1247274

Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study. *Sexual Offender Treatment, 3,* 1–14. Retrieved from http://www.sexual-offender-treatment.org/1-2008_02.html

Eher, R., Schilling, F., Hansmann, B., Pumberger, T., Nitschke, J., Habermeyer, E., & Mokros, A. (2016). Sadism and violent reoffending in sexual offenders. *Sexual Abuse, 28,* 46–72. http://dx.doi.org/10.1177/1079063214566715

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24,827 people: Systematic review and meta-analysis. *BMJ: British Medical Journal, 345,* e4692. http://dx.doi.org/10.1136/bmj.e4692

First, M. B., & Frances, A. (2008). Issues for DSM-V: Unintended consequences of small changes: The case of paraphilias. *The American Journal of Psychiatry, 165,* 1240–1241. http://dx.doi.org/10.1176/appi.ajp.2008.08030361

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: Wiley.

Gail, M. H., & Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics, 6,* 227–239. http://dx.doi.org/10.1093/biostatistics/kxi005

Gazan, F. (2002). Penile plethysmography before the European Court of Human Rights. *Sexual Abuse, 14,* 89–93.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2,* 293–323. http://dx.doi.org/10.1037/1076-8971.2.2.293

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,* 19–30. http://dx.doi.org/10.1037/1040-3590.12.1.19

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics, 26,* 822–830. http://dx.doi.org/10.1093/bioinformatics/btq037

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29–36. http://dx.doi.org/10.1148/radiology.143.1.7063747

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (Corrections Research User Report No. 1997–04). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie canadienne, 50,* 172–182. http://dx.doi.org/10.1037/a0015726

Hanson, R. K., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for static-99R. *Sexual Abuse, 25,* 482–515. http://dx.doi.org/10.1177/1079063212469060

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66,* 348–362. http://dx.doi.org/10.1037/0022-006X.66.2.348

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. M. D. (2007). *Assessing the risk of sexual offenders on community supervision: The dynamic supervision project* (Corrections Research User Report No. 2007–05). Ottawa, Ontario, Canada: Public Safety Canada. Retrieved from http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-en.pdf

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism amongst sexual offenders: A multi-site study of static-2002. *Law and Human Behavior, 34,* 198–211. http://dx.doi.org/10.1007/s10979-009-9180-1

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *International Journal of Forensic Mental Health, 11,* 9–23. http://dx.doi.org/10.1080/14999013.2012.667511

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154–1163. http://dx.doi.org/10.1037/0022-006X.73.6.1154

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1–21. http://dx.doi.org/10.1037/a0014421

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24,* 119–136. http://dx.doi.org/10.1023/A:1005482921333

Hare, R. D. (2003). *Manual for the Psychopathy Checklist–Revised* (2nd ed.). Toronto, Otario, Canada: Multi Health Systems.

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In R. Prentky, E. Janus, & M. Seto (Eds.), *Understanding and managing sexually coercive behavior* (pp. 198–210). New York, NY: Annals of the New York Academy of Sciences. http://dx.doi.org/10.1111/j.1749-6632.2003.tb07306.x

Harris, G. T., & Rice, M. E. (2007). Characterizing the value of actuarial violence risk assessment. *Criminal Justice and Behavior, 34,* 1638–1658. http://dx.doi.org/10.1177/0093854807307029

Harris, G. T., & Rice, M. E. (2013). Bayes and base rates: What is an informative prior for actuarial violence risk assessment? *Behavioral Sciences & The Law, 31,* 103–124. http://dx.doi.org/10.1002/bsl.2048

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315–335. http://dx.doi.org/10.1177/0093854893020004001

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2010). Allegiance or fidelity? A clarifying reply. *Clinical Psychology: Science and Practice, 17,* 82–89. http://dx.doi.org/10.1111/j.1468-2850.2009.01197.x

Harsch, S., Bergk, J. E., Steinert, T., Keller, F., & Jockusch, U. (2006). Prevalence of mental disorders among sexual offenders in forensic psychiatry and prison. *International Journal of Law and Psychiatry, 29,* 443–449. http://dx.doi.org/10.1016/j.ijlp.2005.11.001

Hart, S. D., & Boer, D. P. (2010). Structured professional judgment guidelines for sexual violence risk assessment: The Sexual Violence Risk-20 (SVR-20) and Risk for Sexual Violence Protocol (RSVP). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 269–294). New York, NY: Routledge/Taylor & Francis.

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39,* 1148–1171. http://dx.doi.org/10.1177/0093854812443648

Helmus, L.-M., & Thornton, D. (2015). Stability and predictive and incremental accuracy of the individual items of Static-99R and Static-2002R in predicting sexual recidivism: A meta-analysis. *Criminal Justice and Behavior, 42,* 917–937. http://dx.doi.org/10.1177/0093854814568891

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse, 24,* 64–101.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York, NY: Wiley. http://dx.doi.org/10.1002/0471722146

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment, 15,* 446–455. http://dx.doi.org/10.1037/1040-3590.15.4.446

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10,* 193–217. http://dx.doi.org/10.1016/j.avb.2003.12.002

Kingston, D. A., Firestone, P., Moulden, H. M., & Bradford, J. M. (2007). The utility of the diagnosis of pedophilia: A comparison of various classification procedures. *Archives of Sexual Behavior, 36,* 423–436. http://dx.doi.org/10.1007/s10508-006-9091-x

Marshall, W. L. (2006). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment, 1,* 1–19. Retrieved from http://www.sexual-offender-treatment.org/index.php?id=18&type=123.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46. http://dx.doi.org/10.1037/1082-989X.1.1.30

Meehl, P. E. (1954). *Clinical vs. statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Miller, C. S., Kimonis, E. R., Otto, R. K., Kline, S. M., & Wasserman, A. L. (2012). Reliability of risk assessment measures used in sexually violent predator proceedings. *Psychological Assessment, 24,* 944–953. http://dx.doi.org/10.1037/a0028411

Mokros, A., Habermeyer, E., Neumann, C. S., Schilling, F., Hare, R. D., & Eher, R. (2014). Assessment of psychopathy in Austria. *European Journal of Psychological Assessment, 30,* 243–250. http://dx.doi.org/10.1027/1015-5759/a000177

Mokros, A., Neumann, C. S., Stadtland, C., Osterheider, M., Nedopil, N., & Hare, R. D. (2011). Assessing measurement invariance of PCL-R assessments from file reviews of North American and German offenders. *International Journal of Law and Psychiatry, 34,* 56–63. http://dx.doi.org/10.1016/j.ijlp.2010.11.009

Mokros, A., Vohs, K., & Habermeyer, E. (2014). Psychopathy and violent reoffending in German-speaking countries: A meta-analysis. *European Journal of Psychological Assessment, 30,* 117–129. http://dx.doi.org/10.1027/1015-5759/a000178

Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & The Law, 31,* 23–39. http://dx.doi.org/10.1002/bsl.2050

Murrie, D. C., Boccaccini, M. T., Turner, D., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15,* 19–53. http://dx.doi.org/10.1037/a0014897

Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, Ontario, Canada: Supply and Services Canada.

Olver, M. E., & Wong, S. C. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38,* 113–126. http://dx.doi.org/10.1177/0093854810389534

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders* (2nd ed.). Washington, DC: American Psychological Association.

Quinsey, V. L., & Lalumière, M. L. (1996). *The assessment of sexual aggressors against children*. Newbury Park, CA: Sage.

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence, 10,* 85–105. http://dx.doi.org/10.1177/088626095010001006

Rettenberger, M., Boer, D., & Eher, R. (2011). The predictive accuracy of risk factors in the Sexual Violence Risk-20 (SVR-20). *Criminal Justice and Behavior, 38,* 1009–1027. http://dx.doi.org/10.1177/0093854811416908

Rettenberger, M., & Eher, R. (2007a). *Aktuarische Kriminalprognose bei Sexualstraftätern: Die deutsche Version des Sex Offender Risk Appraisal Guide (SORAG)* [Actuarial risk prediction for sexual offenders: The German version of the Sex Offender Risk Appraisal Guide (SORAG)]. Vienna, Austria: Institut für Gewaltforschung und Prävention.

Rettenberger, M., & Eher, R. (2007b). Predicting reoffense in sexual offender subtypes: A prospective validation study of the German version of the sexual offender risk appraisal guide (SORAG). *Sexual Offender Treatment, 2,* 1–12. Retrieved from http://www.sexual-offender-treatment.org/62.html

Rettenberger, M., Haubner-MacLean, T., & Eher, R. (2013). The contribution of age to the Static-99 risk assessment in a population-based prison sample of sexual offenders. *Criminal Justice and Behavior, 40,* 1413–1433. http://dx.doi.org/10.1177/0093854813492518

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology, 54,* 169–186. http://dx.doi.org/10.1177/0306624X08328755

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d,* and *r*. *Law and Human Behavior, 29,* 615–620. http://dx.doi.org/10.1007/s10979-005-6832-7

Rice, M. E., Harris, G. T., & Cormier, C. (1992). Evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior, 16,* 399–412. http://dx.doi.org/10.1007/BF02352266

Rice, M. E., Harris, G. T., & Hilton, N. Z. (2010). The Violence Risk Appraisal Guide and Sex Offender Risk Appraisal Guide for violence risk assessment and the Ontario Domestic Assault Risk Assessment and Domestic Violence Risk Appraisal Guide for wife assault risk assessment. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment tools* (pp. 99–120). Oxford, England: Routledge/Taylor & Francis.

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychological Assessment, 25,* 951–965. http://dx.doi.org/10.1037/a0032878

Rice, M. E., Harris, G. T., Lang, C., & Cormier, C. (2006). Violent sex offenses: How are they best measured from official records? *Law and Human Behavior, 30,* 525–541. http://dx.doi.org/10.1007/s10979-006-9022-3

Rossegger, A., Gerth, J., Singh, J. P., & Endrass, J. (2013). Examining the predictive validity of the SORAG in Switzerland. *Sexual Offender Treatment, 8,* 1–12. Retrieved from http://www.sexual-offender-treatment.org/index.php?id=123&type=123

Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3,* 203–215. http://dx.doi.org/10.1111/j.1468-2850.1996.tb00071.x

Saß, H., Wittchen, H.-U., Zaudig, M., & Houben, I. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen—Textrevision—DSM–IV–TR* [Diagnostic and statistical manual of mental disorders, text revised fourth ed.]. Göttingen, Germany: Hogrefe.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31,* 499–513. http://dx.doi.org/10.1016/j.cpr.2010.11.009

Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33,* 287–316. http://dx.doi.org/10.1016/j.cpr.2012.12.002

Viallon, V., Ragusa, S., Clavel-Chapelon, F., & Bénichou, J. (2009). How to evaluate the calibration of a disease risk prediction tool. *Statistics in Medicine, 28,* 901–916. http://dx.doi.org/10.1002/sim.3517

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136,* 740–767. http://dx.doi.org/10.1037/a0020473

Received September 21, 2015
Revision received July 26, 2016
Accepted August 18, 2016 ■

Correction to Bovin et al. (2016)

In the article "Psychometric Properties of the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition* (PCL-5) in Veterans" by Michelle J. Bovin, Brian P. Marx, Frank W. Weathers, Matthew W. Gallagher, Paola Rodriguez, Paula P. Schnurr, and Terence M. Keane (*Psychological Assessment*, 2016. Vol. 28, No. 11, pp. 1379–1391. http://dx.doi.org/10.1037/pas0000254), the departments and affiliations were incorrectly listed for authors Michelle J. Bovin, Brian P. Marx, Matthew W. Gallagher, Paola Rodriguez, Paula P. Schnurr, and Terence M. Keane. The first department and affiliation for authors Michelle J. Bovin, Brian P. Marx, Matthew W. Gallagher, Paola Rodriguez, and Terence M. Keane and should have read "National Center for PTSD at VA Boston Healthcare System, Boston, Massachusetts". The first department and affiliation for author Paula P. Schnurr should have read "National Center for PTSD, White River Junction, Vermont."

The online version of this article has been corrected.

http://dx.doi.org/10.1037/pas0000504