Psychological Assessment
2009, Vol. 21, No. 1, 1–21

1040-3590/09/$12.00 DOI: 10.1037/a0014421

The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies

R. Karl Hanson
Public Safety Canada

Kelly E. Morton-Bourgon
Department of Justice, Canada

This review compared the accuracy of various approaches to the prediction of recidivism among sexual offenders. On the basis of a meta-analysis of 536 findings drawn from 118 distinct samples (45,398 sexual offenders, 16 countries), empirically derived actuarial measures were more accurate than unstructured professional judgment for all outcomes (sexual, violent, or any recidivism). The accuracy of structured professional judgment was intermediate between the accuracy found for the actuarial measures and for unstructured professional judgment. The effect sizes for the actuarial measures were moderate to large by conventional standards (average *d* values of 0.67–0.97); however, the utility of the actuarial measures will vary according to the referral question and samples assessed. Further research should identify the psychologically meaningfully factors that contribute to risk for reoffending.

Keywords: risk assessment, sexual offenders, meta-analysis

All societies must respond to individuals who commit serious offenses. One important determinant of these responses (e.g., punishment, detention, supervision) is the perceived risk of recidivism. Sexual offenders, in particular, are often the subject of special policies that are meant to improve community safety by managing sexual offenders' risk of future offending (e.g., postsentence detention, long-term community supervision). The effectiveness of these policies rests on the ability of evaluators to accurately differentiate offenders according to risk level.

The individual characteristics associated with recidivism among sexual offenders have been previously reviewed (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). In general, the two broad domains most strongly associated with sexual recidivism are sexual deviancy and lifestyle instability/criminality. The criminal lifestyle characteristics (e.g., history of rule violation, substance abuse) are also those most strongly related to violent and general (any) recidivism among sexual offenders (Hanson & Morton-Bourgon, 2004), general offenders (Gendreau, Little, & Goggin, 1996), and mentally disordered offenders (Bonta, Law, & Hanson, 1998).

Although a number of recidivism risk factors have been identified, the relationships between any single risk factor and recidivism are small. Consequently, a range of risk factors should be considered in competent evaluations. The question addressed in the current review is the relative accuracy of different methods of combining risk factors into a overall evaluation of risk.

It is widely accepted that evaluations based on unstructured professional judgment are less accurate than structured risk assessments (Andrews, Bonta, & Wormith, 2006; Janus & Prentky, 2003; Monahan, 2007; Quinsey, Harris, Rice, & Cormier, 2006). The general pattern has been documented for at least 50 years (Meehl, 1954), but only recently have forensic psychologists in the United States routinely used structured risk tools for evaluations of adult sexual offenders (Archer, Buffington-Vollum, Stredny, & Handel, 2006). In high-stakes evaluations, such as civil commitment procedures, most evaluators now consider structured risk tools to be essential (Jackson & Hess, 2007).

Static–99 (Hanson & Thornton, 2000) is by far the most commonly used risk tool with adult sexual offenders (Archer et al., 2006; Interstate Commission for Adult Offender Supervision, 2007; Jackson & Hess, 2007; McGrath, Cumming, & Burchard, 2003). It contains 10 items covering static, historical factors, such as the number of prior offenses, victim characteristics (unrelated, strangers, males), and the offender's age. None of the items were intended to measure psychologically meaningful constructs; they were selected purely on the basis of empirical relationships with recidivism and ease of administration. The scores on each of the items are summed to create a total score, and the total score is associated with the observed recidivism rates pooled from three development samples (n = 1,086).

Among forensic psychologists (Archer et al., 2006), the second most widely used measure is the Sexual Violence Risk—20 (SVR–20; Boer, Hart, Kropp, & Webster, 1997). The SVR–20 covers 20

R. Karl Hanson, Corrections Research, Public Safety Canada, Ottawa, Ontario, Canada; Kelly E. Morton-Bourgon, Department of Justice, Ottawa.

The views expressed are those of the authors and are not necessarily those of Public Safety Canada, the Department of Justice, Canada, or our many colleagues who made this work possible. We thank the following researchers who provided unpublished studies, raw data, or findings not included in other reports: Tony Beech, Susanne Bengtson, Jim Bonta, Liam Craig, Jerome Endrass, Doug Epperson, Grant Harris, Mark Hanlon, Raymond Knight, Calvin Langton, Roxanne Lieb, Jan Looman, Larry Motiuk, Thierry Pham, Jean Proulx, John Reddon, Marnie Rice, Sue Righthand, Steve Saum, Greg Schiller, David Thornton, Daniel Wilcox, Robin Wilson, and Annie Yessine. Thanks are also due to Caleb Lloyd and Leslie Helmus for locating articles and formatting tables and to Leticia Gutierrez and Kuan Li for translation of the Spanish and Chinese articles, respectively.

Correspondence concerning this article should be addressed to R. Karl Hanson, Corrections Research, Public Safety Canada, 340 Laurier Avenue West, Ottawa, Ontario K1A 0P8, Canada. E-mail: karl.hanson@ps-sp.gc.ca

features of the offender's criminal history and psychological functioning. The items are meant to assess risk of recidivism and to assist with case management. Example items include past supervision failure, sexual deviation, psychopathy, major mental illness, employment problems, and lack of realistic plans. No explicit procedure is provided for translating the ratings on the items into an overall evaluation of risk. Boer et al. (1997) stated that risk will typically increase monotonically with each additional risk factor, but they recommended against simply adding the items; instead, they directed evaluators to use their own professional judgment to rate the risk as low, moderate, or high. Scores on the are SVR–20 not linked to expected recidivism rates.

How Should Risk Tools Be Classified?

The differences between the above tools illustrate ongoing debates about how risk assessments should be "structured" (Monahan, 2007). The debate is not only about the relative predictive accuracy of specific tools; it also concerns the purpose of the risk assessment and the role of professional judgment. To appreciate the different positions, one can consider the different ways in which structured risk assessments have been conducted and classified.

Much of the discussion of methods for combining information for applied decision making in psychology has been shaped by Meehl's (1954) review of clinical versus statistical prediction. Meehl defined statistical (or actuarial) predictions as having two features: They use an explicit method of combining the information, and that information is linked to a probability figure on the basis of empirically determined relative frequencies (see also Dawes, Faust, & Meehl, 1989; Hilton & Harris, 2005). According to Meehl's (1954) description, the clinical approach to prediction was based on "hypotheses about the structure and dynamic of the particular individual" (p. 4). The accumulated research record clearly shows that it is possible for clinicians to predict without understanding a particular individual and that the subjective feeling of "understanding" need not lead to accurate predictions about future behavior (Quinsey et al., 2006).

The ideal risk assessment procedure, however, would provide more than simply accurate prediction: It would also provide information useful for case management (Douglas, Cox, & Webster, 1999; Grubin & Wingate, 1996; Hart, 1998; Heilburn, 1997). Bonta (1996) distinguished between three generations of risk assessments for offenders. The first involved unstructured professional opinion. The second generation involved actuarial risk scales composed of static, historical factors (e.g., criminal history). These static risk scales were widely implemented in the United States in the 1970s, and similar scales were subsequently adopted by the correctional systems of Canada and the United Kingdom (Bonta & Wormith, 2007). The third generation of risk assessment tools involved assessments of "criminogenic needs" or dynamic risk factors. Dynamic risk factors are defined as characteristics that are capable of change, and changes on these factors are associated with increased or decreased recidivism risk (Andrews, Bonta, & Hoge, 1990). Risk tools with dynamic risk factors are now widely used for general offenders (e.g., Level of Service/Case Management Inventory; Andrews, Bonta, & Wormith, 2004).

Establishment of a factor as a criminogenic need requires both conceptual and empirical justification. In addition to a plausible theory linking the factor to criminal persistence, there must be evidence that (a) the factor predicts recidivism, (b) the factor is capable of changing, and (c) changes in the factor are related to changes in recidivism risk, preferably through experimental manipulation of the risk factor. The causal role of risk factors becomes increasingly plausible as alternative hypotheses for the dynamic relationship with recidivism are eliminated.

There is strong evidence in support of dynamic risk factors for general offenders. A wide range of theoretically plausible factors predicts criminal recidivism (Gendreau et al., 1996). Criminogenic factors, such as irresponsibility and low self-control, decline over the lifespan (Roberts, Walton, & Viechtbauer, 2006). Almost all offenders eventually stop committing crimes (Sampson & Laub, 2003), and desistance can be predicted by pro-social changes on criminogenic needs (Andrews, 1980; Andrews & Wormith, 1984; Raynor & Vanstone, 2001). Programs that deliberately target criminogenic needs, such as criminal attitudes (Andrews, 1980) and substance abuse (Gottfredson, Najaka, Kearley, & Rocha, 2006), reduce general recidivism rates (see meta-analytic review by Andrews & Bonta, 2006, resource note 10.1, pp. 333–336).

The research base for dynamic (causal) risk factors among sexual offenders is, however, much less developed. A number of potentially dynamic risk factors for sexual offenders have been proposed, and many have been shown to predict sexual recidivism (Hanson & Morton-Bourgon, 2004, 2005). What has yet to be convincingly demonstrated is the extent to which changes on these factors are associated with reductions in recidivism risk. In many cases, changes on factors presumed to be criminogenic needs have not led to improved prediction (Langton, 2003; Quinsey, Khanna, & Malcolm, 1998). A recent exception is the study by Olver, Wong, Nicholaichuk, and Gordon (2007), in which treatment-induced progression through the stages of the transtheoretical model of change (Prochaska, DiClemente, & Norcross, 1992) was associated with decreased sexual recidivism. Although further empirical work is required for identification of the central causal factors in the persistence of sexual offending, there are no conceptual obstacles to creation of empirically grounded actuarial risk tools that are useful to intervention and case management.

Bonta's (1996) classification assumes that second- and third-generation measures are actuarial instruments à la Meehl (1954). Some widely used structured risk tools, however, lack a table linking scores to recidivism rates and thus fail to meet the second of Meehl's criteria. For example, the Structured Assessment of Risk and Need, widely used in the United Kingdom (Her Majesty's Prison Service, 2005; also called Structured Risk Assessment [SRA], Thornton, 2002), specifies the items to consider as well as the method for combining the items into an overall score; however, there are no tables linking the summary scores to recidivism rates. Such assessments do not fit into any of Bonta's "generations" and are perhaps best labeled "mechanical," following the language of Grove, Zald, Lebow, Snitz, and Nelson (2000). The distinction between an actuarial measure and a mechanical measure is not permanent; additional research can transform mechanical procedures into fully actuarial measures.

Another widely used, nonactuarial approach to violence risk assessment involves structured professional judgment. In this approach, the items to consider are specified in advance, but the overall evaluation of risk is left to the professional judgment of the evaluator (e.g., the SVR–20). Andrews et al. (2006) classified

these measures as first generation, but they are substantially different from the typical forms of unstructured professional judgment studied by Meehl (1954) or Grove et al. (2000). Promoters of structured professional judgment view it as providing clinically meaningful case formulations and as avoiding the dismal predictive accuracy associated with the unstructured clinical approach (Douglas et al., 1999; Hart, 1998).

Measures such as the SVR–20 are meant to structure professional judgments of risk, but it is not uncommon for researchers to omit the professional judgment and simply add the items from the checklist. In this case, the instrument becomes a mechanical measure. An important research question is the effectiveness of adding the items compared with using professional judgment to form the overall evaluation.

Hanson's (1998) review concluded that there were three plausible approaches to risk assessment with sexual offenders: structured professional judgment, pure actuarial prediction, and clinically adjusted actuarial prediction. The clinically adjusted actuarial approach starts with an established actuarial score and then considers factors external to the actuarial scheme (i.e., the evaluator is allowed to judge the extent to which the predicted recidivism rates are a fair evaluation of the offender's risk). Hanson (1998) noted that research often identifies new risk factors before they are included in actuarial scales. Consequently, evaluators are continually faced with the dilemma of knowing that a factor external to the actuarial scheme incrementally contributes to risk but of lacking an empirically validated method of including that factor in an overall evaluation. Examples of such external factors would be advanced age (60 plus) for the Static–99 (Hanson, 2006) or the interaction of psychopathy and sexual deviance (Harris et al., 2003).

Overview of the Current Study

Our purpose in the current review was to summarize the research on the predictive accuracy of various approaches to risk assessment with sexual offenders. We examined the predictive accuracy of specific measures (e.g., Static–99, SVR–20) as well as the average predictive accuracy for the different approaches to risk assessment. On the basis of the above review, we examined four ways of structuring risk assessments: empirical actuarial, clinically adjusted actuarial, mechanical, and structured professional judgment. All structured approaches to risk assessment were expected to be more accurate than was unstructured professional judgment. Another question addressed was the evidence of predictive accuracy for measures that include a substantial number of factors intended to guide case management or understanding.

We also considered a number of secondary questions concerning factors that influence the predictive accuracy of risk assessments. Empirical actuarial risk tools have been developed with different outcome criteria, and it is important to know how effective they are for assessing related, but not identical, forms of recidivism. For example, how accurate are measures designed to predict violent (including sexual) recidivism (e.g., the Sex Offender Risk Appraisal Guide [SORAG]; Quinsey et al., 2006) for prediction of the more narrow category of sexual recidivism or the broader category of any recidivism? The predictive accuracy also should vary with the amount of random variation introduced by design features (Harris & Rice, 2003). In particular, we expected that the relationships with recidivism would increase, given (a) fixed and equal follow-up times, (b) high levels of rater reliability, and (c) small amounts of missing data.

Method

Sample

Computer searches of PsycINFO, Web of Science, the National Criminal Justice Reference Service (USA), and the library of Public Safety Canada were conducted with the following key terms: *child molester, exhibitionism, exhibitionist, failure, frotteur, incest, indecent exposure, paraphilias, paraphiliac, pedophile, pedophilia, predict(ion), rape, rapist, recidivate, recidivism, recidivist, relapse, reoffend, reoffense, sex(ual) offender, sexual assault, sexual deviant.* As well, Proquest Digital Dissertations was searched with the names of known risk tools (e.g., Static–99, SORAG). Additional sources included the reference lists of empirical studies and previous reviews, conference programs, recent issues of 14 relevant journals (e.g., *Criminal Justice and Behavior, Sexual Abuse: A Journal of Research and Treatment*), and letters sent to 38 established researchers in the field of sexual offender recidivism.

Studies were included if they examined the ability of risk assessments to predict sexual, violent (including sexual), or any recidivism among offenders released following an index sexual offense. Risk assessments were defined as global assessments of the risk for recidivism (e.g., dangerousness, likelihood of recidivism) made with or without the aid of guidelines or actuarial tools. Studies that examined only specific attributes related to risk (e.g., psychopathy, benefit from treatment, sexual deviancy) were excluded.

To be included, structured risk procedures must have been developed with different samples than those reported in the study (i.e., all were replications on new samples). All risk assessments were conducted blind to recidivism status. Studies had to include sufficient statistical information for us to calculate *d,* the effect size, and the recidivism rate (sexual, violent, or any). For dichotomous variables, at least 5 subjects were needed for all marginal totals.

As of June 2008, our search had yielded 147 usable documents (e.g., published articles, books, government reports, conference presentations). In 18 cases, the analyses were based on raw data or analyses provided by the original researchers. When the same data set was reported in several articles, all the results from these articles were considered to come from the same study. Consequently, the 147 documents represented 110 studies with 118 distinct samples (number of studies per country of origin = 41 from the United States, 34 from Canada, 15 from the United Kingdom, 3 from Australia; 2 each from Germany, Sweden, Austria, New Zealand, and Belgium; and 1 each from France, Netherlands, Denmark, Taiwan, Japan, Switzerland, and Spain). Most of the studies identified by our search procedures were studies in English (104 studies; 4 French, 1 Chinese, 1 Spanish). Most were unpublished (62.7%) and were produced between 1972 and 2008, with a median year of 2004. In total, our study included data from 45,398 sexual offenders (average sample size = 417, *Mdn* = 180, range = 27–10,256).

Most of the studies examined mixed groups of adult sexual offenders (91 mixed offense types, 11 child molesters, 3 rapists, 1

sexual murderers; 91 predominantly adults, 18 adolescents, 1 mixed; all samples were male only). All the offenders had committed offenses that meet contemporary definitions of sexual crimes (i.e., old studies that contained homosexuals were excluded). Most of the offenders had been released from prisons or secure psychiatric institutions (60 prison/secure institutions, 24 community only, 24 prison/secure institutions and community, and 2 unknown). The offenders in 40 studies came from treatment programs. When demographic information was presented, the offenders were predominantly Caucasian (50 of 59 studies).

Effect sizes were recorded for three outcome criteria: (a) any sexual recidivism (vs. no recidivism or only nonsexual recidivism; 253 findings); (b) any sexual or violent recidivism (vs. no recidivism or only nonviolent recidivism; 137 findings); and (c) any recidivism (vs. no recidivism; 146 findings). Sexual recidivism included offenses with a sexual motivation, such as child molestation, sexual assault of adults, and exhibitionism. Offenses against public morals without an identifiable victim (e.g., prostitution) were typically excluded. Precise classification of the recidivism types was not recorded, but similar samples typically find that approximately 40%–70% of both the index and recidivism events involve child molestation, 20%–50% involve sexual assault of adults, and less than 10% involve noncontact sexual offenses (Hanson & Thornton, 2003).

Our category of sexual or violent recidivism included all sexual offenses as well as all offenses involving direct confrontation with the victim, such as robbery, assault, kidnapping, and murder. If known, violent offenses that had a sexual motivation were counted as sexual offenses (e.g., sexual homicide). The category of any recidivism included sexual and violent recidivism as well as nonviolent offenses, such as theft, fraud, and drug offenses. Technical violations (e.g., breach of curfew) were excluded if possible. Although the above rules were used in classifying the effect sizes, the precise definitions were restricted to the findings that most closely approximated these definitions in the individual studies.

The most common sources of recidivism information were national criminal justice records (58 studies), such as the Canadian Police Information Centre records (Royal Canadian Mounted Police, 2008) and the Offender Index in the United Kingdom. These national databases were designed for law enforcement purposes; officially recorded convictions can be extracted from them for research purposes. The second most common source of recidivism information involved state or provincial records (44 studies), such as those held by state police or provincial correctional systems. The source of the recidivism information was unknown for 11 studies. The average follow-up period ranged from 6 months to 276 months, with a mean of 70 months ($SD = 46.6$).

Coding Procedure

Each study was coded with a standard list of variables and explicit coding rules (available from R. Karl Hanson upon request). The studies were coded by one of two raters (R. Karl Hanson and Kelly E. Morton-Bourgon), and most were independently coded by both raters. Only one finding per individual variable was coded per sample on the basis of (a) sample size and (b) completeness of information. If the sample sizes and descriptive detail were equivalent, the median value was used.

Classification of Risk Tools

The risk assessment procedures were placed in one of the following five categories.

Empirical actuarial. These tools had explicit items determined in advance and explicit methods for combining the items into total scores. Both the items and the combination rules were selected on the basis of empirical evidence that linked them to recidivism. As well, these measures had tables that linked scores to expected recidivism rates.

Mechanical. These tools had explicit items determined in advance and explicit methods for combining the items into a total score. They did not have a table that linked the total score to recidivism probabilities. The method of selecting and combining the items was based primarily on theory or literature reviews instead of direct analysis of specific data sets.

Adjusted actuarial. These evaluations were based on the total scores of an actuarial or mechanical tool except when the evaluator determined that there were factors external to the actuarial or mechanical scheme that justified overriding the obtained rating. The external factors were not specified in advance, and neither was the method of combining the external factors with the results of the actuarial or mechanical tool.

Structured professional judgment. Evaluators were given a structured list of risk factors determined in advance. The method of combining the factors into a total score was not specified in advance, and the overall evaluation of risk was left to the professional judgment of the evaluator.

Unstructured. Neither the risk factors nor the method of combining the risk factors was specified in advance. Risk assessments were based on individual case analysis, case conferences, or professional experience.

In addition, we classified each measure according to the outcome that it is primarily intended to address. Outcomes included (a) sexual recidivism, for offenses that have a sexual motivation; (b) sexual or violent recidivism, for violent offenses (including sexual offenses) that involve victim confrontation or intent to harm; or (c) any recidivism (for these measures, only whether or not the offender commits any new offense was important and not the type of recidivism).

As well, we made a dichotomous judgment (yes/no) as to whether the risk assessment instrument was intended as a guide to intervention by specifying the factors that could be addressed in order to minimize the risk for recidivism (i.e., dynamic risk factors or needs). This determination was primarily based on the stated intentions of the authors of the instruments; however, we also considered the extent to which the measure included variables that were potentially amenable to intervention. Such variables included those that have been empirically associated with sexual, violent, or general recidivism risk (e.g., sexual preoccupations, impulsivity, intimacy deficits) as well as potentially changeable variables for which the association with recidivism has not been established (e.g., psychosis, low self-esteem, denial). Our classification of a variable as a "treatment target," therefore, did not require that there was convincing evidence that the factor played a causal role in criminal persistence; instead, we used the lower criteria of whether these variables were considered to be worthy of intervention by the developers of the measures. Variables that were defined as unchanging over the life course (e.g., psychopathy) were not included

as treatment targets. The most common examples of structured risk tools intended to guide intervention were the SVR–20 (Boer et al., 1997; 11 studies), the Juvenile Sex Offender Assessment Protocol (J-SOAP; Prentky, Harris, Frizzell, & Righthand, 2000; 6 studies), the Level of Service Inventory—Revised (LSI–R; Andrews et al., 2004; 6 studies) and variants, the Structured Assessment of Risk and Need/SRA (Thornton, 2002; 3 studies), and the HCR–20 (Webster, Douglas, Eaves, & Hart, 1997; 3 studies).

Rater Reliability

Two tests of rater reliability were conducted. The first entailed independent classifications by two raters (R. Karl Hanson and Kelly E. Morton-Bourgon) of 41 different risk assessment procedures. Overall, agreement was acceptable to high for the type of risk assessment ($\kappa = .66$, 75.6% agreement), the type of recidivism predicted ($\kappa = .76$, 87.8% agreement), and whether the measure meaningfully assessed case management variables ($\kappa = .83$, 90.2% agreement).

The second reliability test involved two independent ratings of the design features and effect sizes for 10 studies. Two of the reliability studies contained overlapping samples (Friendship, Mann, & Beech, 2003; Thornton et al., 2003); one rater combined these studies, whereas the other rater considered them as separate studies. Consequently, the number of studies available for comparison was 9. Agreement for dichotomous and categorical variables was indexed by kappa, and two-way random effects model intraclass correlation coefficients (ICC; type absolute agreement) were used for ordinal and interval variables (Design 2 in Orwin, 1994). The agreement for the sample characteristics was perfect for seven variables and was good to excellent for eight others ($\kappa > .70$, ICC $> .80$). Agreement was fair for four variables: published or not ($\kappa = .55$; disagreed on in press articles); victims ($\kappa = .60$; disagreed on whether hebephiles were child molesters or rapists); the total, combined sample size (ICC = .72; a result of one rater combining 2 studies and the other rater leaving them separate); and whether national criminal records were used to detect recidivism ($\kappa = .59$).

In the reliability studies, Rater 1 identified 56 findings and Rater 2 identified 63 findings; there was agreement on 104 of the 119 findings (87.4%). The interrater reliability of the effect sizes calculation was .68. The agreement increased to .79 if one outlier was removed. Most differences involved simple omissions or clerical errors.

Index of Predictive Accuracy

The effect size indicator was the standardized mean difference, *d*, defined as follows: $d = (M_1 - M_2)/S_w$, where M_1 is the mean of the deviant group, M_2 is the mean of the nondeviant group, and S_w is the pooled-within standard deviation (Hasselblad & Hedges, 1995). In other words, *d* measures the average difference between the recidivists and the nonrecidivists and compares this difference to how much recidivists differ from each other and how much nonrecidivists differ from each other. Formulas for calculating *d* were based on those presented by Cohen (1988), Hasselblad and Hedges (1995), Rosenthal (1991), and Swets (1986).

Aggregation of Findings

Two methods were used to summarize the findings: median values (Slavin, 1995) and weighted mean values (Hedges & Olkin, 1985). The averaged *d* value, *d*., was calculated by weighing each d_i by the inverse of its variance: $d. = \left(\sum_{i=1}^{k} w_i d_i\right) \Big/ \left(\sum_{i=1}^{k} w_i\right)$, where *k* is the number of findings ($w_i = 1/v_i$), and v_i is the variance of the individual d_i (fixed effect model). The variance of the weighted mean was used to calculate 95% confidence intervals: $Var(d.) = 1 \Big/ \left(\sum_{i=1}^{k} w_i\right)$; 95% confidence interval $= d. \pm 1.96(\text{var}[d.])^{1/2}$. Weighting *d* values by the inverse of their variance means that findings from small samples are given less weight than findings from large samples.

When d_i was calculated from 2×2 tables, the variance of d_i was estimated using Formula 6 from Hasselblad and Hedges (1995):

$$Var(d_i) = \frac{3}{\pi^2}\left(\frac{1}{a + .5} + \frac{1}{b + .5} + \frac{1}{c + .5} + \frac{1}{d + .5}\right).$$

When d_i was calculated from other statistics (e.g., *t*, receiver operating characteristic areas, means), the variance of d_i was estimated with Formula 3 from Hasselblad and Hedges (1995):

$$Var(d_i) = \left[\frac{N_1 + N_2}{N_1 N_2} + \frac{d_i^2}{2(N_1 + N_2)}\right].$$

To test the generalizability of effects across studies, we used Hedges and Olkin's (1985) *Q* statistic: $Q = \sum_{i=1}^{k} w_i(d_i - d.)^2$. The *Q* statistic is distributed as a chi-square with *k*–1 degrees of freedom (*k* is the number of studies). A significant *Q* statistic indicates that there is more variability across studies than would be expected by chance. Given only random variation across studies, the expected value of *Q* is similar to the degrees of freedom. Outliers were excluded from each category if the single extreme value accounted for more than 50% of the total variance (*Q*).

The major findings are presented as average *d* values along with their 95% confidence intervals. The confidence intervals represent the range of plausible values for the population parameters. Substantive interpretations can be based on values anywhere in the interval (Cumming & Finch, 2005). When the observations are uncorrelated, two values can be considered to differ at the $p < .01$ level if their confidence intervals do not overlap (Cumming & Finch). Given that the different methods of evaluating recidivism risk would be expected to be substantially positively correlated, confidence intervals provide a test of differences with extremely low statistical power. Nevertheless, nonoverlapping confidence intervals were used as the primary indicator of statistically significant differences between findings. This decision was a reasonable compromise, given the large sample sizes (which decrease Type II error) and the large number of possible comparisons (which increases Type I error). The confidence interval "rule of eye" was supplemented by fixed-effects tests for specific hypotheses (Hedges, 1994). Specifically, the statistical significance of dichotomous or categorical moderator variables was indexed by the reduction in the overall *Q* statistic (Q_Δ) when the variable was included in the analysis. The statistical significance of linear moder-

ator variables was indexed by dividing the meta-regression coefficients by their standard errors, yielding *Z* statistics (Hedges, 1994).

Results

The observed sexual recidivism rate was 11.5% (n = 28,757, 100 samples), the sexual or violent recidivism rate was 19.5% (n = 17,421, 50 samples), and the general (any) recidivism rate was 33.2% (n = 23,343, 65 samples). Studies that specified in advance the number of recidivists and nonrecidivists were excluded from the rate calculations (e.g., Dempster, 1998). The average follow-up time was 70 months. Readers should note that the observed rates should be considered underestimates of the true rates, because not all offenses are detected. Although some of the "recidivists" would have been falsely accused, the proportion of falsely accused individuals would be smaller than the proportion of undetected offenders.

The average predictive accuracies of the various approaches to risk assessment are summarized in Table 1. For the prediction of sexual recidivism, the approaches with the strongest predictive accuracy were the actuarial measures designed for sexual recidivism ($d.$ = 0.67), the mechanical measures designed for sexual recidivism ($d.$ = 0.66), and the actuarial measures designed for general recidivism ($d.$ = 0.62). Unstructured professional judgment ($d.$ = 0.42) and actuarial measures for violent recidivism ($d.$ = 0.39) were less accurate in the prediction of sexual recidivism than were the actuarial or mechanical measures designed for sexual recidivism. The mechanical methods designed for violent recidivism were less accurate ($d.$ = 0.33) than were the actuarial measures designed for sexual recidivism ($d.$ = 0.67). The findings for structured professional judgment were intermediate between the results for unstructured professional judgment and actuarial measures. For structured professional judgment, however, there was large variability across the results of a small number of studies.

For the prediction of sexual or violent recidivism, the actuarial measures designed for violent recidivism ($d.$ = 0.78) and the actuarial measures designed for any recidivism ($d.$ = 0.79) were superior to any of the other methods. For the prediction of any recidivism, the most accurate approach involved actuarial measures designed for any recidivism ($d.$ = 0.97), followed by actuarial measures designed for violent recidivism ($d.$ = 0.74).

Actuarial measures designed for any recidivism were more accurate at predicting any recidivism ($d.$ = 0.97) than were actuarial measures designed for predicting violent recidivism at predicting violent recidivism ($d.$ = 0.78) or the actuarial measures designed for sexual recidivism at predicting sexual recidivism ($d.$ = 0.67). This does not mean that sexual and violent recidivism is inherently less predictable than general recidivism. Averaged across all findings, the average ability of all risk assessments to predict any recidivism ($d.$ = 0.57, 95% confidence interval = .55–.59, k = 146) was similar to the average ability of all risk assessments to predict sexual recidivism ($d.$ = 0.58, 95% confidence interval = .56–.60, k = 253). The risk assessments for violent recidivism were, on average, somewhat less accurate ($d.$ = 0.51, 95% confidence interval = .49–.53, k = 137) than were the risk assessments for sexual or any recidivism.

Unstructured professional judgment showed weak relationships with violent ($d.$ = 0.22) or any ($d.$ = 0.11) recidivism. The unstructured evaluations were conducted between 1970 and 1998 (*Mdn* = 1993). The accuracy of the unstructured evaluations did not change on the basis of the year in which the risk assessment was conducted (all *r*s were nonsignificant for the prediction of sexual, violent, or any recidivism).

The Appendix presents further details of the analyses, as well as the findings for individual risk tools. Within each category, the commonly used risk tools all showed moderate-to-large predictive accuracy; with rare exceptions, all their confidence intervals overlapped. Readers interested in specific measures can peruse the tables at their leisure. Measures with strong empirical support would have high mean values based on large sample sizes, narrow

Table 1
Average Predictive Accuracy of Various Forms of Risk Assessment for Sexual Offenders

	Recidivism outcome criteria					
Form of risk assessment	Sex *d* (95% CI)	*k*	Any violence *d* (95% CI)	*k*	Any *d* (95% CI)	*k*
Measures designed for sexual recidivism						
Empirical actuarial	0.67 (.63, .72)	81	0.51 (.47, .56)	42	0.52 (.48, .56)	43
Mechanical	0.66 (.58, .74)	29	0.40 (.31, .49)	10	0.37 (.30, .43)	19
Structured professional judgment	0.46 (.29, .62)[a]	6	0.31 (.13, .49)	3	0.26 (.11, .41)	4
Measures designed for violent recidivism						
Empirical actuarial	0.39 (.31, .46)	20	0.78 (.70, .86)	15	0.74 (.66, .82)	14
Mechanical	0.33 (.07, .60)	4	0.31 (.07, .56)	3	—	
Measures designed for any recidivism						
Empirical actuarial	0.62 (.49, .75)	9	0.79 (.63, .95)	5	0.97 (.93,1.01)	10
Unstructured professional judgment	0.42 (.32,0.51)	11	0.22 (.15, .29)[b]	7	0.11 (.06, .17)	9

Note. A dash indicates insufficient data. k is the number of studies; CI = confidence interval.
[a] Outlier excluded; with outlier, d = 0.59 (.43, .74). [b] With outlier, d = 0.24 (.17, .31).

confidence intervals, and consistent results across samples ($Q < df$, median equal to mean).

To further examine the potential contribution of professional judgment, we identified five studies in which the evaluators rated a predetermined set of items and then formed an overall evaluation of risk based on either (a) professional judgment or (b) summing the items. As can be seen from Table 2, the results of the procedures were similar. Three studies favored professional judgment, and two studies favored the simple sums. In most cases, the differences between the approaches was not large enough to be meaningful.

Three studies examined the difference between actuarial scores and adjusted actuarial risk ratings (Gore, 2007; Hanson, 2007; Vrana, Sroga, & Guzzo, 2008). In these studies, evaluators were required to complete an actuarial risk tool and then were allowed to adjust the final risk rating on the basis of factors external to the actuarial tool. All three studies were prospective, and evaluators completed the ratings as part of their routine procedures. In two studies, the raters were probation officers (Hanson, 2007; Vrana et al., 2008), and in the other study, the raters were either psychologists or correctional staff (Gore, 2007). For all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores (see Table 3).

On average, measures that contained factors presumed to be treatment targets were more accurate for the prediction of sexual recidivism than were measures based primarily on static, historical risk factors (.74 vs. .64, $Q_\Delta = 5.55$, $df = 1$, $p < .025$). The same pattern was not observed, however, for the other outcomes. For the prediction of violent or general recidivism, the measures with only static factors showed a slight (nonsignificant) advantage over measures with potential treatment targets (.55 vs. .53 for violent recidivism, $Q_\Delta = 0.13$, $df = 1$, $p > .50$; .62 vs. .57 for any recidivism, $Q_\Delta = 1.51$, $df = 1$, $p > .25$).

The possible effects of design features were examined in two ways. The first approach used the complete set of findings, including all measures and all outcome criteria (536 findings). This approach is comprehensive but involves considerable error, because consistency would not be expected and the lack of independence of findings within samples was ignored. Consequently, the second approach examined the largest category of findings that examined one specific measure and one specific outcome (i.e., Static–99 predicting sexual recidivism; 63 findings from 63 distinct samples). The statistical significance of the continuous moderator variables was tested with the fixed effects weighted least squares formula described by Hedges (1994).

Across all findings, published findings yielded larger effects ($d. = 0.61$, $k = 194$) than did the unpublished findings ($d. = 0.50$, $k = 342$, $Q_\Delta = 47.4$, $df = 1$, $p < .001$). The same pattern was observed for Static–99 findings: $d. = 0.80$ for the 21 published findings compared with $d. = 0.60$ for the 42 unpublished studies ($Q_\Delta = 15.7$, $df = 1$, $p < .001$). Neither analysis found significant relationships with the total sample size (unweighted $r = -.032$, $k = 536$; unweighted $r = -.088$, $k = 63$, respectively) or with the year of publication ($r < .001$, $k = 536$; $r = .12$, $k = 63$, respectively).

The effect sizes were stronger in the United Kingdom than in the other countries. For all findings, the average d in the United Kingdom was 0.66 ($k = 74$), 0.57 in Canada ($k = 218$), 0.52 in the United States ($k = 155$), and 0.56 in the other countries ($k = 89$, $Q_\Delta = 30.4$, $df = 3$, $p < .001$). The average d for Static–99 predicting sexual recidivism was 0.90 in the United Kingdom ($k = 12$), 0.58 in Canada ($k = 21$), 0.60 in the United States ($k = 14$), and 0.74 for other countries ($k = 16$, $Q_\Delta = 19.1$, $df = 3$, $p < .001$).

For the overall set of studies, there was no difference in the average effect size for the studies that used fixed follow-up periods ($d = 0.56$, $k = 87$) compared with those that used variable follow-up periods ($d = 0.55$, $k = 430$). However, Static–99 studies that used fixed follow-up times had larger effects than did studies that used variable follow-up times ($d. = 0.83$ [12] vs. $d. = 0.62$ [48], $Q_\Delta = 12.2$, $df = 1$, $p < .001$).

The effect of rater reliability was examined for findings from studies that reported the ICC (overall, $k = 125$, $Mdn = .90$, range = 0.60–1.00; for Static–99 findings, $k = 12$, $Mdn = .90$, range = 0.63–0.97). In the overall set of findings, assessments with high rater reliability showed stronger effect sizes than did assessments with low rater reliability ($r = .214$, $k = 125$, $Z = 4.12$, $p < .001$). For the Static–99 findings, the relationship between rater reliability and effect size was not significant ($r = -.040$, $k = 12$, $Z = 0.18$, $p > .50$).

The amount of missing data was measured as the percentage of the total items that was not rated for the average participant in the

Table 2
Comparison of Professional Judgment and the Simple Addition of Items for the Prediction of Sexual Recidivism

Study	Measure	Judgment, d	Addition, d	Recidivists/ total
Sjöstedt & Långström (2002)	SVR–20	0.21	−0.04	10/51
Dempster (1998)	SVR–20	1.23	1.27	24/73
De Vogel et al. (2004)	SVR–20	1.35	1.19	47/121
Kropp (2000)	RSVP	0.97	0.53	15/53
Morton (2003)	ERASOR	0.14	0.31	13/77
d (95% CI)		0.93 (0.69, 1.17)	0.82 (0.58, 1.06)	109/375
Q		16.2 ($p = .0028$)	15.6 ($p = .0036$)	

Note. The offenders in the Kropp (2000) study were a subset of offenders from Dempster (1998). SVR–20 = Sexual Violence Risk—20 (Boer et al., 1997); RSVP = Risk for Sexual Violence Protocol (Hart et al., 2003); ERASOR = Estimate of Risk of Adolescent Sexual Offense Recidivism (Worling & Curwen, 2000).

Table 3
Comparison of Empirical Actuarial With Clinically Adjusted Actuarial Risk Assessments

Study	Measure	Actuarial risk score	Adjusted risk category	Difference	Type of recidivism	Recidivists/ total
Hanson (2007)	Static–99	1.04	.95	−.09	Sexual	34/507
	Static–99	0.91	.78	−.13	Sexual or violent	65/507
Gore (2007)	MnSOST–R	0.50	.31	−.19	Sexual	19/381
Vrana et al. (2008)	LSI–OR	0.90	.54	−.36	Sexual or violent	25/198
	LSI–OR	0.98	.64	−.34	Any	52/198
Unweighted *M*		0.87	.64	−.22		

Note. MnSOST–R = Minnesota Sex Offender Screening Tool—Revised (Epperson et al., 1995); LSI–OR = Level of Service Inventory—Ontario Revision (Girard & Wormith, 2004).

study ($M = 5.7\%$, $Mdn = 1\%$, range = 0%–30%). The studies that reported the amount of missing data ($k = 149$) showed larger effect sizes ($d. = 0.60$) than did the studies that did not mention the amount of missing data ($d. = 0.52$, $k = 387$, $Q_{\Delta} = 26.0$, $df = 1$, $p < .001$). Contrary to expectation, larger amounts of missing data were associated with larger effect sizes ($r = .11$, $k = 149$, $Z = 2.39$, $p < .01$). The pattern was in the same direction for the Static–99 studies, although the effect was not significant ($r = .18$, $k = 22$, $Z = 1.32$, $p > .10$).

The final set of analyses examined the incremental contribution of type of risk assessment after controlling for study features (published/unpublished, country of origin [Canada, United Kingdom, United States, other], fixed/variable follow-up, and whether or not the amount of missing data was reported). These features of the studies were selected because they were statistically significant in at least one analysis and were available for the majority of the findings. The type of risk assessment was indexed by two sets of dummy variables: (a) empirical actuarial, mechanical, or structured judgment (with unstructured judgment as the reference category) and (b) designed for sexual or violent recidivism (with general recidivism as the reference category). The seven findings concerning adjusted actuarial assessments were excluded.

For prediction of sexual recidivism (248 findings), sexual or violent recidivism (133 findings), or any recidivism (141 findings), the type of risk assessment significantly predicted the effect size after we controlled for the study variables (see Table 4). Even though the type of risk assessment (two variables) explained more variance than did the study variables (four variables), there was a considerable amount of unexplained variance (Birge ratios = 2.47–3.34) for all three outcome variables.

The multivariate analyses had the same pattern of results as that shown in Table 1. After we had controlled for the study variables, empirical actuarial and mechanical approaches were more accurate than was unstructured professional judgment for the prediction of sexual recidivism. As well, measures designed for sexual recidivism were more accurate than were the other measures for the prediction of sexual recidivism. For the prediction of violent recidivism, empirical actuarial, mechanical, and structured professional judgment were more accurate than unstructured professional judgment. Measures designed for sexual recidivism were less accurate than were the other measures for the prediction of sexual or violent recidivism. The results for the prediction of any recidivism were the same as those for the prediction of violent recidivism, with the exception that measures designed for general recidivism were more accurate than the measures designed for sexual and violent recidivism.

Discussion

Our purpose in this review was to compare the accuracy of various approaches to the prediction of recidivism among sexual offenders. For all outcome measures, unstructured professional judgment was significantly less accurate than were the empirically derived actuarial measures. For the prediction of any recidivism, the empirical actuarial measures designed for any recidivism were the most accurate and were significantly more accurate than all other approaches. For the prediction of violent (including sexual) recidivism, the most accurate approaches were actuarial measures designed for any recidivism and the actuarial measures designed for violent recidivism. There was no overall difference between these two categories of measures for the prediction of violent recidivism, and both were significantly more accurate than all other approaches.

Table 4
The Relative Contribution of Study Features (Four Variables) and Type of Risk Assessment (Two Variables) on the Observed Effect Sizes

	Source of variance				
Type of recidivism	Study features, *Q* (*df*)	Type of risk assessment, *Q* (*df*)	Unexplained, *Q* (*df*)	Birge ratio[a]	Total, *Q* (*k*)
Sexual	44.1 (6)	64.3 (5)	583.8 (236)	2.47	692.1 (248)
Sexual or violent	69.0 (6)	111.8 (5)	404.0 (121)	3.34	584.9 (133)
Any	114.2 (6)	685.0 (5)	418.2 (129)	3.24	1,217.3 (141)

Note. All *Q* values significant at $p < .001$.
[a] Birge ratios are defined as Q_{error}/df and can be interpreted as the ratio of unexplained between-study variability divided by within-study variability (random error). The expected value of the Birge ratio is 1.0 when the effect parameters are determined exactly by the linear model (Hedges, 1994).

For the prediction of sexual recidivism, the most accurate approaches were actuarial measures designed for sexual recidivism ($d. = 0.67$), mechanical measures designed for sexual recidivism ($d. = 0.66$), and actuarial measures designed for any recidivism ($d. = 0.62$). The accuracy of structured professional judgment was intermediate between the accuracy found for actuarial measures and for unstructured professional judgment. Given the small number of studies and significant variability in the results, further research is needed before strong statements can be made about the relative accuracy of structured professional judgment compared with that of other approaches. We expect, however, that advocates of structured professional judgment will be quick to point out that the measure with the largest average association with sexual recidivism was the SVR–20 professional judgment, even though this finding was based on only three studies ($n = 245$) and showed significant variability ($Q = 7.96$, $df = 2$, $p < .05$).

The clinically adjusted actuarial approach has been frequently proposed (Hanson, 1998; Webster, Harris, Rice, Cormier, & Quinsey, 1994) but rarely evaluated. There were only three direct tests of clinical adjustments to actuarial measures identified in the current review (Gore, 2007; Hanson, 2007; Vrana et al., 2008), but the results were clear. In each study, the clinical adjustments decreased the predictive accuracy over that observed for the pure actuarial measures. These studies are important because they were all prospective studies that examined current, applied risk assessment with sexual offenders. Defenders of clinical overrides could argue that the evaluations resulted in interventions that changed the expected recidivism risk of the offenders. Until evidence is collected in support of this assertion, the simplest interpretation is that the overrides simply added noise. In the Hanson (2007) study, the rater reliability for the override was not above chance levels (intraclass correlation = .14, $n = 74$).

Currently, the mechanical approaches to the prediction of sexual recidivism do as well as the empirical actuarial approaches. This will change as ongoing research transforms the most promising mechanical scales into fully actuarial measures and as new actuarial scales are created using the best empirical predictors. Comparison of the items of the most popular actuarial scale (Static–99) with the most popular mechanical scale (by adding the items from the SVR–20) indicates that the SVR–20 includes much more complex items (e.g., sexual deviant, psychopathy, major mental illness) than does the Static–99 (e.g., age, number of prior offenses). The future of sexual offender risk assessment lies in the development of fully actuarial measures that contain clinically relevant, causal risk factors (i.e., Bonta's third-generation measures).

Currently, the closest examples of third-generation risk assessments for sexual offenders are Vermont's Sex Offender Treatment Needs and Progress Scale (McGrath, Cumming, & Livingston, 2005; $d = 0.82$, one study) and the Violence Risk Scale—Sex Offender Version (VRS–SO; Olver et al., 2007; $d. = 0.89$, based on two studies, one by Olver et al. and one independent replication (Beggs & Grace, in press). Both measures assess theoretically meaningful constructs and present probability tables linking total scores to recidivism rates. As well, both studies of the VRS–SO found that posttreatment scores on VRS–SO Dynamic Variables added incrementally to the prediction provided by pretreatment scores. Further research is required, however, to help us establish the generalizability of these results and understand what is changing during treatment.

Readers should be aware that the methods used in the current review allowed for broad comparisons only. Lack of statistically significant differences between specific measures does not mean that equivalence has been established. Direct comparisons of measures that used the same sample would provide comparisons substantially more precise than those provided in the current review.

The analysis of design features explained only a small percentage of the variability across studies. Studies that used fixed follow-up periods found stronger effects among the Static–99 studies, but the same pattern was not observed for the full set of findings. Among the full set of findings, studies with high rater reliability had larger effects sizes than did studies with low rater reliability; however, the same pattern was not observed for the Static–99 studies. In the complete data set, studies that reported the amount of missing data had stronger effect sizes than did studies that did not report the amount of missing data; perplexingly enough, the amount of missing data was not associated with the accuracy of the Static–99 findings and was positively related to effect sizes in the full data set.

Although the psychometric quality of the assessment data must influence the magnitude of the observed effects, it appears that the design features examined in the current review had relatively little influence compared with the unmeasured features. One interpretation is that the commonly used actuarial instruments are relatively robust, given minor variations in scoring procedures and missing data. Possible design features worth considering in future reviews include (a) variability in the predictor measures (Hanson, 2008), (b) the quality of the information used in completing the assessment, and (c) whether the assessments were completed by researchers or by applied evaluators as part of their routine duties.

In the current review, published studies yielded stronger results than did unpublished studies. Although this may be another example of publication bias, it is also possible that the published studies had better control over extraneous variables. Publication bias would be expected to have made relatively little contribution to the overall findings of the current review, given that most of the studies (63%) were unpublished.

The observation of relatively high predictive accuracy in the U.K. studies is consistent with the United Kingdom's long-term commitment to maintaining comprehensive national criminal history records. Increased reliability of the criminal history records should translate into larger findings in the validity studies. In comparison with the United Kingdom, Canada and the United States have higher rates of recent immigration, and this fact increases the probability of failing to detect offenses committed in other countries. Långström (2004) found that the Static–99 predicted recidivism among Nordic and non-Nordic Europeans but failed to significantly predict recidivism among the African/Asian offenders in a Swedish national sample. The extent to which the differences in predictive accuracy are due to missing records or to real social–cultural differences remains to be explored. Regardless of the source, the observation of jurisdictional differences in predictive accuracy suggests that all countries should conduct local validity studies prior to routine implementation of measures developed in other jurisdictions (e.g., Rettenberger & Eher, 2006).

One limitation of the current review is that it examined the accuracy to which the assessments ranked the relative risk of

offenders but did not address the ability of the assessments to predict absolute recidivism rates. In many applied contexts, however, it is important to know whether the offenders' probability of recidivism exceeds some preestablished threshold (see Doren, 2004a; Mossman, 2006). Establishing reliable estimates of the absolute ("real") recidivism rates is difficult because many offenses are undetected, and it is plausible that the recidivism rates will change across cohorts and jurisdictions. It is possible, for example, that the changes in sexual values from the 1970s to 2010 could influence both the probability of recidivism (Minnesota Department of Corrections, 2007) and the likelihood that abusive sexual behavior would be reported as a crime (Todd, 2006). Given its genesis in data, the empirical actuarial approach will ultimately provide the best estimates of absolute risk. To date, those involved in the development of structured professional guidelines have not even attempted to specify absolute risk levels; instead, they have restricted themselves to relative risk defined in broad categories (e.g., "low," "moderate," "high"; Hart et al., 2003).

Although the effect sizes for most of the measures were moderate to large by conventional standards (Cohen, 1988; Rice & Harris, 2005), there are no direct procedures for translating the observed effect sizes into statements concerning the normative value or utility of specific measures (e.g., is a *d* value of 0.67 for the Static–99 poor, acceptable, good, or excellent?). The utility of a measure with any specific *d* value would vary depending on the purpose for which it is used. For decisions concerning allocation of scarce resources to priority cases (e.g., duration of treatment, intensity of community supervision), a rational policy would involve use of measures that balance the cost of administration with the measures' ability to discriminate between recidivists and nonrecidivists (i.e., large *d* values). However, even with a large *d* value, a measure may have limited utility for decisions concerning whether or not offenders meet some predetermined threshold of absolute risk. When the base rate is substantially different from the decision threshold, it is quite possible for the recidivists to have much higher scores than the nonrecidivists but for none of the offenders with high scores to exceed the decision threshold.

Implications for Applied Assessment

Given the weight of evidence in support of actuarial risk tools, we believe that they should be a major consideration in the evaluation of recidivism risk potential for sexual offenders. At this point, it is hard for evaluators to justify the use of unguided clinical opinion except for cases that fall outside the sampling frame of the existing actuarial measures (e.g., predicting sexual recidivism among adolescent female sexual offenders).

Evaluators have a number of measures to choose from depending on the offender, the goal of the assessments, and the information and resources available. The ideal measure would have the largest relationship to recidivism, the smallest confidence interval (with the low end of the confidence interval higher than all others), and the least variability across samples (smallest Birge ratio; Hedges, 1994). By these criteria, no single measure has yet to establish itself as clearly more accurate than other, similar measures.

For the prediction of sexual recidivism, the best supported measures were Static–99 (Hanson & Thornton, 2000), Static–2002 (Hanson & Thornton, 2003), MnSOST–R (Epperson et al., 2000), Risk Matrix—2000 Sex (Thornton et al., 2003), and adding the items from the SVR–20 (Boer et al., 1997). For the prediction of violent (including sexual) recidivism, the current findings supported the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 2006), the SORAG (Quinsey et al., 2006), the Risk Matrix—Combined (Thornton et al., 2003), the Statistical Index of Recidivism (SIR; Bonta, Harman, Hann, & Cormier, 1996), and the LSI–R and variants (Andrews et al., 2004). For the prediction of general recidivism, the best measures were the VRAG, the SORAG, the SIR, and the LSI–R.

An important unresolved question is how to interpret divergent findings from different risk tools. The use of multiple risk tools is common; 79.5% of civil commitment evaluators reported using more than one tool for the same case (Jackson & Hess, 2007). When the results of the different tools converge, evaluators are justified in having increased confidence in their conclusions (Mills & Kroner, 2006). However, the results of different risk tools often diverge. Barbaree, Langton, and Peacock (2006), for example, found that less than 5% of their sample was consistently identified as high risk or as low risk across five actuarial risk tools for sexual offenders (VRAG, SORAG, RRASOR, Static–99, MnSOST–R). As yet, no empirically justified method has been established for resolving such divergent results.

One proposed approach is to use only the "best" measure (Seto, 2005). Until a best measure is empirically established, however, this approach is equivalent to choosing a favorite. Another approach to resolving the divergent results of different empirical actuarial scales is to interpret them on the basis of differences in content (e.g., Doren, 2004b). The problem with this approach is that it assumes construct validity for items that may have been selected on a purely empirical basis.

The need to resolve the results of conflicting actuarial risk tools should motivate the development of new and better actuarial tools. In these new tools, variables associated with recidivism would be incorporated into psychometrically sound measures of psychologically meaningful constructs, and the tools would include a comprehensive set of factors responsible for the persistence of sexual crime. Until such a measure has been established, evaluators will need to rely on their professional judgment when they consider which measures to use and how to interpret the results for a particular case.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Allam, J. (1999). *Effective practice in work with sex offenders: A reconviction study comparing treated and untreated offenders.* Birmingham, West Midlands, England: West Midlands Probation Service Sex Offender Unit.

*Allan, A., & Dawson, D. (2002). *Developing a unique risk of violence tool for Australian indigenous offenders* (No. CRC 6/00–01). Canberra, Australia: Criminology Research Council.

*Allan, A., Dawson, D., & Allan, M. (2006). Prediction of the risk of male sexual reoffending in Australia. *Australian Psychologist, 41,* 60–68.

*Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19,* 347–367.

Andrews, D. A. (1980). Some experimental investigations of the principles

of differential association through deliberate manipulations of the structure of service systems. *American Sociological Review, 45,* 448–462.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed). Cincinnati, OH: LexisNexis/Anderson.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17,* 19–52.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The Level of Service/Case Management Inventory (LS/CMI).* Toronto, Ontario, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency, 52,* 7–27.

Andrews, D. A., & Wormith, S. J. (1984). *Criminal sentiments and criminal behaviour* (Corrections Research User Report No. 1984–30). Ottawa, Ontario, Canada: Solicitor General of Canada.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87,* 84–94.

*Austin, J., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the Static–99/RRASOR sex offender risk assessment instruments: Final report submitted to the Pennsylvania Board of Probation and Parole.* Washington, DC: Institute on Crime, Justice and Corrections, George Washington University.

Barbaree, H. E., Langton, C., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18,* 423–440.

*Barnoski, R. (2005). *Sex offender sentencing in Washington State: Notification levels and recidivism* (Document No. 05–12–1203). Olympia, WA: Washington State Institute for Public Policy.

*Barnoski, R. (2006a). *Sex offender sentencing in Washington State: Predicting recidivism based on demographic and criminal history* (Document No. 06–01–1207). Olympia, WA: Washington State Institute for Public Policy.

*Barnoski, R. (2006b). *Sex offender sentencing in Washington State: Predicting recidivism based on the LSI–R* (Document No. 06–02–1201). Olympia, WA: Washington State Institute for Public Policy.

*Barnoski, R. (2006c). *Sex offender sentencing in Washington State: Sex offender risk level classification tool and recidivism* (Document No. 06–01–1204). Olympia, WA: Washington State Institute for Public Policy.

*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 & Comparative Criminology, 47,* 422–438.

*Bates, A., Falshaw, L., Corbett, C., Patel, V., & Friendship, C. (2004). A follow-up study of sex offenders treated by Thames Valley Sex Offender Groupwork Programme, 1995–1999. *Journal of Sexual Aggression, 10,* 29–38.

*Beech, A., & Ford, H. (2006). The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending. *Psychology, Crime & Law, 12,* 685–701.

*Beech, A., 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,* 155–167.

*Beech, A. R. (2005, November). *An evaluation of the effectiveness of Static–99 with juveniles.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

*Beech, A. R., Tudway, J. A., Parish, R. J. H., Print, B., & Thornton, D. (2004). *An evaluation of the effectiveness of an actuarial risk assessment tool for people adjudicated as sex offenders as adolescents.* Unpublished manuscript, University of Birmingham, Birmingham, United Kingdom.

*Beggs, S. M., & Grace, R. C. (2008). Psychopathy, intelligence, and recidivism in child molesters: Evidence of an interaction effect. *Criminal Justice and Behavior, 35,* 683–695.

*Beggs, S. M., & Grace, R. C. (in press). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment.*

*Bélanger, N. (1995). La prédiction de la récidive chez un groupe d'agresseurs sexuels [The prediction of recidivism among a group of sexual offenders]. *Dissertation Abstracts International, 57*(06), 4021B. (UMI No. NN09572).

*Bélanger, N., & Earls, C. (1996). Sex offender recidivism prediction. *Forum on Corrections Research, 8,* 22–24.

*Bengtson, S. (2007). [Recidivism data collected in 2002 from 445 sexual offenders assessed at Aarhus University Hospital or Clinic of Forensic Psychiatry, Copenhagen, Denmark, between 1978 and 1992]. Unpublished raw data.

*Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: A Journal of Research and Treatment, 19,* 135–153.

*Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertation Abstracts International, 68*(09), B. (UMI No. NR30941).

*Boer, A. (2003). *Evaluating the Static–99 and Static–2002 risk scales using Canadian sexual offenders.* Unpublished master's thesis, University of Leicester, Leicester, United Kingdom.

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: British Columbia Institute Against Family Violence.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18–32). Thousand Oaks, CA: Sage.

*Bonta, J., & Hanson, R. K. (1995). [Ten-year recidivism data for offenders released from the Correctional Service of Canada in 1983/1984]. Unpublished raw data.

Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). Prediction of recidivism among federally sentenced offenders: A re-validation of the SIR (Statistical Information on Recidivism) Scale. *Canadian Journal of Criminology, 38,* 61–79.

Bonta, J., Law, M., & Hanson, R. K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123,* 123–142.

Bonta, J., & Wormith, S. (2007). Risk and need assessment. In G. McIvor & P. Raynor (Eds.), *Developments in social work with offenders* (pp. 131–152). London: Jessica Kingsley.

*Bonta, J., & Yessine, A. K. (2005). [Recidivism data for 124 released sexual offenders from the offenders identified in *The National Flagging System: Identifying and responding to high-risk, violent offenders*]. Unpublished raw data.

*Bright, D., Hoy, A., Murphy, K., Harle, K., Marshall, R., & Senior, J. (2007, June). *Treatment effectiveness: An appraisal of custody based sexual offender treatment programmes in New South Wales.* Paper presented at the Third Congress on Psychology, Psychiatry, and Law, Montreal, Quebec, Canada.

*Brown, J. A. (2003). A comparison of actuarial methods of predicting sexual dangerousness. *Dissertation Abstracts International, 55*(02), 1068B. (UMI No. 3120894).

*Caperton, J. D. (2005). Predicting recidivism among sex offenders: Utility of the Static–99, Minnesota Sex Offender Screening Tool—Revised, and

the Psychopathy Checklist—Revised. *Dissertation Abstracts International, 66*(09), 5080B. (UMI No. 3190110)

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Cortoni, F., & Nunes, K. L. (2007). *Assessing the effectiveness of the National Sexual Offender Program* (Research Report No. R-183). Unpublished manuscript.

*Cotton, D. J. (1991). *Final evaluation report: Pilot juvenile sex offender treatment program. Report to the California legislature in response to SB 890 (Seymour, 1985).* Pinole, CA: Author.

*Craig, L. A., Beech, A. R., & Browne, K. D. (2006). Evaluating the predictive accuracy of sex offender risk assessment measures on UK samples: A cross-validation of the Risk Matrix 2000 scales. *Sexual Offender Treatment, 1*(1). Retrieved June 1, 2008, from http://www.sexual-offender-treatment.org/19.98.html

*Craig, L. A., Thornton, D., Beech, A., & Browne, K. D. (2007). The relationship of statistical and psychological risk markers to sexual reconviction. *Criminal Justice and Behavior, 34,* 314–329.

*Craissati, J., Bierer, K., & South, R. (2008). *What do sex offenders really get up to? Risk prediction, community failure and "sexually risky behaviors" in a nine-year follow-up study.* Unpublished manuscript.

*Craissati, J., Falla, S., McClurg, G., & Beech, A. (2002). Risk, reconviction rates and pro-offending attitudes for child molesters in a complete geographical area of London. *Journal of Sexual Aggression, 8,* 22–38.

*Craissati, J., Webb, L., & Keen, S. (2005). *Personality disordered sex offenders.* Unpublished manuscript.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

*Dempsey, C. (2002). *A comparison of risk assessment instruments with community based sexual offenders.* Unpublished master's thesis, University of Leicester, Leicester, United Kingdom.

*Dempster, R. J. (1998). Prediction of sexually violent recidivism: A comparison of risk assessment instruments. *Masters Abstracts International, 77*(06), 1986. (UMI No. MQ37511).

*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,* 235–251.

*Dix, G. E. (1976). Differential processing of abnormal sex offenders: Utilization of California's mentally disordered sex offender program. *Journal of Criminal Law & Criminology, 67,* 233–243.

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

Doren, D. M. (2004b). Toward a multidimensional model for sexual recidivism risk. *Journal of Interpersonal Violence, 19,* 835–856.

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4,* 149–184.

*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,* 15–26.

*Dufresne, M. H. (2005). Impact de la réussite au traitement de la délinquance sexuelle sur la récidive, en lien avec les scores obtenus au PCL–R, au SORAG et à la Statique–99 [Impact of success in sexual offender treatment on recidivism in relation to scores obtained on the PCL–R, SORAG and Static–99]. *Dissertation Abstracts International, 67*(05), 4021B. (UMI No. NR14991).

*Endrass, J., Urbaniok, F., Held, L., Vetter, S., & Rossegger, A. (in press). The accuracy of the Static–99 in predicting recidivism in Switzerland. *International Journal of Offender Therapy and Comparative Criminology.*

*Epperson, D. L. (2003). *Validation of the MnSOST–R, Static–99, and RRASOR with North Dakota prison and probation samples.* Unpublished manuscript.

*Epperson, D. L., Kaul, J. D., & Huot, S. J. (1995, October). *Predicting risk of recidivism for incarcerated sex offenders: Updated development on the Sex Offender Screening Tool (SOST).* Poster session presented at the 14th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, New Orleans, LA.

*Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (2000, November). *Cross-validation of the Minnesota Sex Offender Screening Tool—Revised.* Paper presented at the 19th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

*Fischer, D. R. (2000). *Sex offender risk assessment validation study.* Phoenix, AZ: Arizona Department of Corrections.

*Florida Department of Health and Rehabilitative Services. (1985). *Status of the sex offender treatment programs, fiscal year 1983–1984* (*annual report to the Florida legislature).* Tampa, FL: Author.

*Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *Journal of Interpersonal Violence, 18,* 744–759.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology,* 575–607.

*Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory—Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior, 31,* 150–181.

*Gore, K. S. (2007). Adjusted actuarial assessment of sex offenders: The impact of clinical overrides on predictive accuracy. *Dissertation Abstracts International, 68*(07), 4824B. (UMI No. 3274898).

Gottfredson, D. E., Najaka, S. S., Kearley, B. W., & Rocha, C. M. (2006). Long-term effects of participation in the Baltimore City drug treatment court: Results from an experimental study. *Journal of Experimental Criminology, 2,* 67–98.

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.

Grubin, D., & Wingate, S. (1996). Sexual offence recidivism: Prediction versus understanding. *Criminal Behaviour and Mental Health, 6,* 349–359.

*Haag, A. M. (2005). Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static–99 and Static–2002 on a re-conviction measure of sexual recidivism. *Dissertation Abstracts International, 66*(08), 4531B. (UMI No. NR05662).

*Hanlon, M. J., Bender, S., Larson, S., & Schoepp, J. (2004). *The MnSOST and MnSOST–R in North Dakota: Validation and utility.* Unpublished manuscript.

*Hanlon, M. J., & Herlickson, A. (2007, October). *"Risk assessment" and recidivism of sex offenders released from NDSP: What the data tell us about sending letters to state's attorneys.* Paper presented at the fall meeting of the North Dakota Psychological Association, Fargo, ND.

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Report 97–04). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hanson, R. K. (1998). What do we know about sexual offender risk assessment? *Psychology, Public Policy and Law, 4,* 50–72.

*Hanson, R. K. (2002). [Evaluation of Manitoba's Secondary Risk Assessment]. Unpublished raw data.

Hanson, R. K. (2006). Does Static–99 predict recidivism among older sexual offenders? *Sexual Abuse: A Journal of Research and Treatment, 18,* 343–355.

*Hanson, R. K. (2007, March). *How should risk assessments for sexual offenders be conducted?* Paper presented at the Fourth Annual Forensic Psychiatry Conference, Victoria, British Columbia, Canada.

Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 15*(1), 15–17. Retrieved August 14, 2008, from http://www.cpa.ca/sections/criminaljustice/publications/

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.

*Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27,* 6–35.

*Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (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.

Hanson, R. K., & Morton-Bourgon, K. E. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Corrections Research User Report No. 2004–02). Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.

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.

*Hanson, R. K., Steffy, R. A., Scott, H., & Gauthier, R. (1993). [Long-term recidivism of child molesters released from Millbrook Institution between 1958 and 1974]. Unpublished raw data.

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.

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static–2002* (Corrections Research User Report 2003–01). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management* (pp. 193–210). New York: New York Academy of Sciences.

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

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3,* 121–137.

Hart, S. D., Kropp, P. R., Laws, D. R., Klaver, J., Logan, C., & Watt, K. A. (2003). *The Risk for Sexual Violence Protocol (RSVP): Structured professional guidelines for assessing risk of sexual violence.* Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

*Hartwell, L. L. (2001). *Sex offender risk appraisal guide: Validity and utility for Hawaii sex offender risk assessments.* Unpublished manuscript, American School of Professional Psychology, Hawaii Campus.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117,* 167–178.

*Haynes, A. K., Yates, P. M., Nicholaichuk, T., Gu, D., & Bolton, R. (2000, June). *Sexual deviancy, risk and recidivism: The relationship between deviant arousal, the RRASOR and sexual recidivism.* Paper presented at the annual conference of the Canadian Psychological Association, Ottawa, Ontario, Canada.

*Hecker, J., Scoular, J., Righthand, S., & Nangle, D. (2002, October). *Predictive validity of the J–SOAP over 10-plus years: Implications for risk assessment.* Poster session presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Montreal, Quebec, Canada.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Heilburn, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21,* 347–359.

*Helmus, L. (2007). A multisite comparison of the validity and utility of the Static–99 and Static–2002 for risk assessment with sexual offenders. Unpublished manuscript, Carleton University, Ottawa, Ontario, Canada.

*Helmus, L. M. D., & Hanson, R. K. (2007). Predictive validity of the Static–99 and Static–2002 for sex offenders on community supervision. *Sexual Offender Treatment, 2*(2), 1–14. Retrieved June 1, 2008, from http://www.sexual-offender-treatment.org/60.0.html

Her Majesty's Prison Service. (2005). *Structured assessment of risk and need (sexual offenders): Manual version 2. Offending Behavior Programmes Unit.* London: Author.

*Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52,* 5–20.

*Hills, A. L. (2003). Assessment, treatment, and recidivism of Aboriginal and non-Aboriginal offenders: A comparison of intra-familial and extra-familial male sexual offenders in Saskatchewan. *Masters Abstracts International, 24*(02), 684B. (UMI No. MQ82620).

Hilton, N. Z., & Harris, G. T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence, & Abuse, 6,* 3–23.

*Hiscox, S. P., Witt, P. H., & Haran, S. J. (2007). Juvenile Risk Assessment Scale (JRAS): A predictive validity study. *Journal of Psychiatry and Law, 35,* 503–539.

*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' judgments of their risk. *British Journal of Criminology, 42*(2), 371–394.

*Hudson, S. A. (2003). The Static–99 and community notification in Tarrant County. *Masters Abstracts International, 42*(04), 1168B. (UMI No. 1418170).

*Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: The Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment, 14,* 103–119.

Interstate Commission for Adult Offender Supervision (2007, April). *SO assessment information survey 4–2007.* Retrieved October 30, 2008, from http://www.interstatecompact.org/LinkClick.aspx?fileticket=jLoQPVeviaQ%3d&tabid=105&mid=431

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment, 19,* 409–448.

Janus, E. S., & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility and accountability. *American Criminal Law Review, 40,* 1443–1499.

*Johansen, S. H. (2007). Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods. *Dissertation Abstracts International, 68*(03), 1929B. (UMI No. 3255527).

*Johnson, J. L. (2006). Sex offenders on federal community supervision: Factors that influence revocation. *Federal Probation, 70,* 18–32.

*Kingston, D. A., Firestone, P., Babchishin, K., & Yates, P. M. (2007, November). *The predictive validity of the Risk Matrix 2000 in a Canadian sample of sexual offenders.* Poster session presented at the 26th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

*Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Unpublished manuscript.

*Kolko, D. J. (2005, November). *Court-mandated juvenile sexual offenders in the community: Characteristics, collaborative treatment, and*

recidivism. Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

*Kozol, H. L., Boucher, R. J., & Garofalo, R. F. (1972). The diagnosis and treatment of dangerousness. *Crime and Delinquency, 18,* 371–392.

*Kropp, P. R. (2000, November). *The Risk for Sexual Violence Protocol (RSVP).* Paper presented at the 26th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

*Långström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment, 16,* 107–120.

*Langton, C. M. (2003). Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy. *Dissertation Abstracts International, 64*(04), 1907B. (UMI No. NQ78052).

*Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static–2002 and five other instruments. *Criminal Justice and Behavior, 34,* 37–59.

*Looman, J. (2006). Comparison of two risk assessment instruments for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18,* 193–206.

*Looman, J., Hammond, E., & Abracen, J. (2005). *A test of Doren (2004). Dimensions of risk model in a sample of high-risk sexual offenders.* Unpublished manuscript.

*Marghem, B. (2007, September). *Évaluation du risqué de récidive en Clinique Évaluative à la Prison de Tournai* [Evaluating the risk for recidivism at Tournai Prison's Evaluation Clinic]. Paper presented at Fourth Congrès International Francophone sur L'agression Sexuelle, Paris.

*Martinez, R., Flores, J., & Rosenfeld, B. (2007). Validity of the Juvenile Sex Offender Assessment Protocol—II (J-SOAP–II) in a sample of urban minority youth. *Criminal Justice and Behavior, 34,* 1284–1295.

*McCoy, W. K. (2007). Predicting treatment outcome and recidivism among juvenile sex offenders: The utility of the J-SOAP–II and ERASOR in an outpatient treatment sample. *Dissertation Abstracts International, 68*(11), 7673B. (UMI No. 3288394).

McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 nationwide survey.* Brandon, VT: Safer Society Press.

*McGrath, R. J., Cumming, G., & Livingston, J. A. (2005, November). *Predictive validity of the Sex Offender Treatment Needs and Progress Scale (TPS).* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

*McGrath, R. J., Hoke, S. E., Livingston, J. A., & Cumming, G. F. (2001, November). *The Vermont Assessment of Sex-Offender Risk (VASOR): An initial reliability and validity study.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Mills, J. F., & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health, 16,* 155–166.

*Milton, J. (2003). *Care pathways and outcome of mentally disordered sex offenders referred to high secure psychiatric hospitals.* Unpublished master's thesis, University of Nottingham, Nottingham, United Kingdom.

*Min-chieh, J. L., & Tzu-yi, D. (2005). The establishment of a Taiwan Sex Offender Static Risk Assessment Scale (TSOSRAS–2004): A validity study by various samples. *Asian Journal of Domestic Violence and Sexual Offense, 1,* 49–110.

Minnesota Department of Corrections. (2007). *Sex offender recidivism in Minnesota.* St. Paul, MN: Author.

Monahan, J. (2007). Clinical and actuarial predictions of violence. In D. Faigman, D. Kaye, M. Saks, J. Sanders, & E. Cheng (Eds.), *Modern scientific evidence: The law and science of expert testimony* (pp. 122–147). St. Paul, MN: West Publishing.

*Montana, S., & Thompson, G. (2005, November). *Risk assessment for priest sex offenders: Developing a risk assessment tool to measure relapse risk.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

*Morton, K. E. (2003). Psychometric properties of four risk assessment measures with male adolescent sexual offenders. *Dissertation Abstracts International, 41*(06), 1833. (UMI No. MQ79677).

Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment, 18,* 41–63.

*Motiuk, L. L., & Brown, S. L. (1995). [Survival time for sexual offenders released from the Correctional Service of Canada in 1991 and 1994]. Unpublished raw data.

*Nicholaichuk, T. (1997). [Relationship between a modified version of the RRASOR and sexual recidivism among 283 untreated sexual offenders]. Unpublished raw data.

*Nicholaichuk, T. (2001, November). *The comparison of two standardized risk assessment instruments in a sample of Canadian Aboriginal sexual offenders.* Paper presented at the annual Research and Treatment 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 (SORAG). *Sexual Abuse: A Journal of Research and Treatment, 14,* 253–269.

*Nunes, K. L., Wexler, A., Firestone, P., & Bradford, J. M. (2003, October). *Incarceration and recidivism in a sample of sexual offenders.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, St. Louis, MO.

*Ohio Department of Rehabilitation and Correction. (2001). *Ten-year recidivism follow-up of 1989 sex offender releases.* Columbus, OH: Author.

*Olver, M. E. (2003). The development and validation of the Violence Risk Scale: Sexual Offender Version (VRS: SO) and its relationship to psychopathy and treatment attrition. *Dissertation Abstracts International, 64*(09), 4628B. (UMI No. NQ83561).

*Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale—Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19,* 318–329.

Orwin, R. G. (1994). Evaluating coding decisions. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 139–162). New York: Russell Sage Foundation.

*Parks, G. A. (2004). Juvenile sex offender recidivism: Typological differences in risk assessment. *Dissertation Abstracts International, 65*(06), 3176B. (UMI No. 3135791).

*Pérez, M., Redondo, S., Martínez, M., Garcia, C., & Andrés, A. (2008). Predicción de riesgo de reincidencia en agresores sexuales [Prediction of risk for recidivism among sexual offenders]. *Psichotherma, 20,* 205–210.

*Poole, D., Liedecke, D., & Marbibi, M. (2000). *Risk assessment and recidivism in juvenile sexual offenders: A validation study of the Static–99.* Austin, TX: Texas Youth Commission.

*Prentky, R., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial

procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 12,* 71–93.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to the addictive behaviors. *American Psychologist, 47,* 1102–1114.

*Proulx, J. (2004). [Static predictors of recidivism among sex offenders assessed in the Quebec Region of the Correctional Service of Canada between 1995 and 2000]. Unpublished raw data.

*Proulx, J., Pellerin, B., McKibben, A., Aubut, J., & Ouimet, M. (1995). [Static and dynamic predictors in sexual aggressors]. Unpublished raw data.

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

Quinsey, V. L., Khanna, A., & Malcolm, P. B. (1998). A retrospective evaluation of the Regional Treatment Centre sex offender treatment program. *Journal of Interpersonal Violence, 13,* 621–644.

*Ralston, C. A., & Epperson, D. L. (2006, November). *Assessment of the validity of adult sexual offender risk assessment tools with juveniles.* Paper presented at the 25th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago.

*Ralston, C. A., & Epperson, D. L. (2007, November). *Validation of the Juvenile Sexual Offense Recidivism Risk Assessment Tool*—II (JSORRAT–II). Paper presented at the 26th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Raynor, P., & Vanstone, M. (2001). "Straight thinking on probation": Evidence-based practice and the culture of curiosity. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice* (pp. 189–203). Chichester, United Kingdom: Wiley.

*Reddon, J. R., Studer, L., & Estrada, L. (1996). [Recidivism data from the Phoenix Program for sex offender treatment]. Unpublished raw data.

*Rettenberger, M., & Eher, R. (2006). Actuarial assessment of sex offender recidivism risk: A validation of the German version of the Static–99. *Sexual Offender Treatment, 1*(3), 1–11. Retrieved June 1, 2008, from http://www.sexual-offender-treatment.org/51.0.html

*Rettenberger, M., & Eher, R. (2007). 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*(2), 1–12. Retrieved June 1, 2008, from http://www.sexual-offender-treatment.org/62.2html

*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.

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.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Royal Canadian Mounted Police. (2008). *Fact sheet: Canadian Police Information Service (CPIC*). Retrieved September 8, 2008, from http://www.rcmp-grc.gc.ca/factsheets/fact_cpic_e.htm

Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41,* 555–592.

*Saum, S. (2005). [Sexual recidivism among 175 sexual offenders in relationship to Static–99 and Stable–2000 scores]. Unpublished raw data.

*Saum, S. (2007). A comparison of an actuarial risk prediction measure (Static–99) and a stable dynamic risk prediction measure (Stable–2000) in making risk predictions for a group of sexual offenders. *Dissertation Abstracts International, 68*(03), 1943B. (UMI No. 3255539).

*Schiller, G. (2000). [Hospital and aftercare clinicians' estimates of recidivism probability compared with actual reoffense rates by type of outcome and type of prediction]. Unpublished raw data.

*Schram, D. D., Milloy, C. D., & Rowe, W. E. (1991). *Juvenile sex offenders: A follow-up study of reoffense behavior.* Olympia, WA: Washington State Institute for Public Policy.

*Seager, J. A., Jellicoe, D., & Dhaliwal, G. K. (2004). Refusers, dropouts, and completers: Measuring sex offender treatment efficacy. *International Journal of Offender Therapy and Comparative Criminology, 48,* 600–612.

Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17,* 156–167.

*Sjöstedt, G., & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static–99 in Sweden. *Law and Human Behaviour, 25,* 629–645.

*Sjöstedt, G., & Långström, N. (2002). Assessment of risk for criminal recidivism among rapists: A comparison of four different measures. *Psychology, Crime & the Law, 8,* 25–40.

*Skelton, A., Riley, D., Wales, D., & Vess, J. (2006). Assessing risk for sexual offenders in New Zealand: Development and validation of a computer-scored risk measure. *Journal of Sexual Aggression, 12,* 277–286.

*Skowron, C. (2005). Differentiation and predictive factors in adolescent sexual offending. *Masters Abstracts International, 66*(03), 1788B. (UMI No. NR00807).

Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48,* 9–18.

*Smiley, W. C., McHattie, L., & Mulloy, R. (1998, October). *Predicting sexual recidivism using a brief actuarial risk scale.* Poster session presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Vancouver, British Columbia, Canada.

*Smith, W. R., & Monastersky, C. (1986). Assessing juvenile sexual offenders' risk for reoffending. *Criminal Justice and Behavior, 13,* 115–140.

*Song, L., & Lieb, R. (1994). [Recidivism data for the study of Washington State's sentencing alternative for sex offenders]. Unpublished raw data.

*Soothill, K., Harman, J., Francis, B., & Kirby, S. (2005). Identifying future repeat danger from sexual offenders against children: A focus on those convicted and those strongly suspected of such crime. *Journal of Forensic Psychiatry and Psychology, 16,* 225–247.

*Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 16,* 92–108.

Stalans, L. J., Seng, M., & Yarnold, P. R. (2002). Long-term impact evaluation of specialized sex offender probation programs in Lake, DuPage and Winnebago Counties.* Chicago: Illinois Criminal Justice Information Authority.

*Sturgeon, V. H., & Taylor, J. (1980). Report of a five-year follow-up study of mentally disordered sex offenders released from Atascadero State Hospital in 1973. *Criminal Justice Journal, 4,* 31–63.

Swets, J. A. (1986). Indices of discrimination and diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

*Ternowski, D. R. (2004). Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program. *Dissertation Abstracts International, 66*(06), 3428B. (UMI No. NR03201).

*Thornton, D. (1997). [A 16-year follow-up of 563 sexual offenders from HM Prison Service in 1979]. Unpublished raw data.

*Thornton, D. (2000). [Recidivism information for 110 sexual offenders released from HM Prison Service during the 1990s]. Unpublished raw data.

*Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14,* 139–153.

*Thornton, D., & Knight, R. (2006). [Actuarial risk scales for sexual offenders referred to Bridgewater Treatment Center: 5-year follow-up]. Unpublished raw data.

*Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management* (pp. 225–235). New York: New York Academy of Sciences.

Todd, D. (2006, October 14). Quest trial: Was teacher–student sex acceptable in the 1970s? *Vancouver Sun,* pp. C1, C11.

*Tough, S. E. (2001). Validation of two standardized risk assessments (RRASOR, 1997; Static–99, 1999) on a sample of adult males who are developmentally disabled with significant cognitive deficits. *Masters Abstracts International, 39*(06), 1626B. (UMI No. MQ58817).

*Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chávez, V., Ullman, D., & Lawrence, L. (2008). Assessing risk for violence in adolescents who have sexually offended: A comparison of the J-SOAP–II, J-SORRAT–II, and SAVRY. *Criminal Justice and Behavior, 35,* 5–23.

*Vrana, G. C., Sroga, M., & Guzzo, L. (2008). *Predictive validity of the LSI–OR among a sample of adult male sexual assaulters.* Unpublished manuscript, Nipissing University, North Bay, Ontario, Canada.

*Watanabe, K., Yokota, K., Yoshimoto, K., Ihara, N., & Fujita, G. (2007). Recidivism in child rapists: Identifying high-risk factors. Unpublished manuscript.

*Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of Internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19,* 449–465.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR–20: Assessing the risk of violence* (2nd ed.). Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). Violence prediction scheme: Assessing dangerousness in high risk men. Toronto, Ontario, Canada: University of Toronto, Centre of Criminology.

*Wieand, L. A. (1983). Judgments of dangerousness: An analysis of mental hospital decisions to release mentally disordered sex offenders. *Dissertation Abstracts International, 44*(06), 2018B. (UMI No. 8323449).

*Wilcox, D., Beech, A. R., Markall, H. F., & Blacker, J. (2008). [Actuarial risk assessment and recidivism for 27 intellectually disabled sexual offenders assessed at the West Midlands Probation Service between 1994 and 2002]. Unpublished raw data.

*Williams, J. D. (2007). Juvenile sex offenders: Predictors of recidivism. *Dissertation Abstracts International, 68*(05), 2185A. (UMI No. 3265532).

*Wilson, R. J., Cortoni, F., & Vermani, M. (2007). *Circles of support and accountability: A national replication of outcome findings* (Report No. R-185). Ottawa, Ontario, Canada: Correctional Service of Canada.

*Wilson, R. J., Picheca, J. E., & Prinzo, M. (2007). Evaluating the effectiveness of professionally-facilitated volunteerism in the community-based management of high-risk sexual offenders: Pt. 2: A comparison of recidivism rates. *Howard Journal, 46,* 327–337.

*Wilson, R. J., & Prinzo, M. (2001, November). *The concurrent validity of actuarial measures of sexual and violent risk in high-risk sexual offenders detained until sentence completion.* Paper presented at the annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

*Witte, T. D., Di Placido, C., Gu, D., & Wong, S. C. P. (2006). An investigation of the validity and reliability of the Criminal Sentiments Scale in a sample of treated sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 18,* 249–258.

*Witte, T., Di Placido, C., & Wong, S. (2001). *How dangerous are dangerous sex offenders? An estimation of recidivism and level of risk using a matched control group.* Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre.

Worling, J. R., & Curwen, T. (2001). *The ERASOR: Estimate of risk of adolescent sexual offense recidivism (Version 2.0).* Toronto, Ontario, Canada: SAFE-T Program, Thistletown Regional Centre.

*Wormith, J. S., & Ruhl, M. (1986). Preventive detention in Canada. *Journal of Interpersonal Violence, 1,* 399–430.

Appendix

Table A1
Prediction of Sexual Recidivism

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
Actuarial empirical/sex	0.74	.67	0.63, 0.72	182.87***	81	24,089	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 20, 21, 22, 23, 25, 27, 28, 29, 30, 31a, 31b, 34, 35, 36, 37, 38, 40, 41, 44, 47, 48, 51, 54, 55, 56, 59, 60, 61, 62, 63, 64, 65, 66, 69, 71, 72, 73, 74, 76, 79, 80, 81, 85, 86a, 86b, 86c, 88, 90, 91, 92, 93, 95, 96, 97, 98, 99, 100, 101, 103, 104, 106, 107, 108a, 108b, 109a, 109b
Static–99	0.74	.67	0.62, 0.72	129.85***	63	20,010	3, 4.1, 8, 9.1, 11, 12, 14, 15, 16, 17, 18.2, 20, 21.1, 22.1, 23, 25.1, 27, 28, 29, 30, 31a, 31b, 35, 38, 41, 44, 47, 48, 51.2, 54, 55, 56, 60, 61, 63.2, 64, 65, 69, 72, 76, 79, 80.2, 81, 86a, 86b, 86c, 90, 91, 92, 93, 95.1, 96, 98, 99, 100, 103.4, 104, 106, 107, 108.1a, 108.1b, 109a, 109b
Static–2002	0.71	0.70	0.59, 0.81	8.90	8	3,330	11, 12, 15, 27.2, 35, 54, 72, 80.2
RRASOR	0.66	0.60	0.54, 0.65	85.08***	34	11,031	2, 3, 4.1, 11, 15, 20, 18, 18.1, 22, 25, 31a, 31b, 38, 41, 51, 54, 60, 63.2, 65, 66, 72, 76, 79.1, 80.1, 85, 88, 90, 95.1, 97, 100, 103.1, 104, 106, 108
MnSOST	1.61	0.59	0.35, 0.84	15.39***	3	316	36, 37, 63.1
MnSOST–R	0.80	0.76	0.65, 0.87	51.33***	12	4,672	4.1, 15, 17, 25.1, 31a, 31b, 34, 36, 37, 54, 63.1, 73
SACJ–Min	0.46	0.42	0.27, 0.57	2.96	6	1,392	1, 20, 21, 40, 71, 93
Risk Matrix 2000, sexual	0.66	0.67	0.56, 0.77	15.87	10	2,755	6, 7, 11, 15, 20, 21.1, 47.1a, 47.1b, 79.2, 106
Actuarial empirical/violence	0.48	0.39	0.31, 0.46	44.44***	20	4,539	2, 3, 4, 11, 14, 15, 20, 25, 28, 29, 41, 42, 54, 77, 78, 79, 80, 85, 103, 106
VRAG	0.47	0.52	0.40, 0.64	8.86	8	1,933	14, 15, 25, 41, 54, 78, 85, 103.4
SORAG	0.60	0.62	0.51, 0.73	17.93	12	3,058	4.1, 15, 25, 28, 29, 41, 42, 54, 77, 79, 80, 103.4
Risk Matrix 2000, violence only	0.14	0.24	0.10, 0.38	11.36*	5	1,082	11, 15.1, 20, 79.2, 106
Risk Matrix 2000, sexual/violence	0.38	0.42	0.28, 0.57	2.75	4	997	11, 15.1, 79.2, 106
Actuarial empirical/any recidivism	0.54	0.62	0.49, 0.75	37.96***	9	6,613	2, 13, 14, 18, 22, 39, 65, 87, 103
SIR	0.64	0.51	0.33, 0.70	10.08*	5	800	13, 14, 18.6, 22, 39
LSI–R (and variants)	0.44	0.45	0.22, 0.68	1.23	4	904	14, 65, 87, 103.3
Mechanical/sex	0.61	0.66	0.58, 0.74	83.23***	29	5,838	1, 3, 7, 8, 15, 18, 20, 23, 25, 27, 38, 43, 44, 46, 54, 57, 58, 61, 65, 67, 68, 70, 81, 85, 87, 92, 99, 101, 103
SVR–20, adding the items	0.60	0.68	0.55, 0.81	40.20***	10	1,699	15, 18.4, 20, 23, 25, 44, 54, 68, 85, 92
Structured Risk Assessment (SRA)	0.78	0.79	0.57, 1.02	0.53	3	637	15, 20, 99
J-SOAP	0.24	0.33	0.03, 0.63	7.14	6	629	43, 57, 58, 67, 70, 101
Beech deviance	0.54	0.71	0.17, 1.26	1.26	3	229	1, 7, 8
ERASOR, adding the items	0.31	0.60	0.29, 0.91	5.37	3	316	58, 65, 87

(*table continues*)

Table A1 (*continued*)

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
Mechanical/violence	0.14	0.33	0.07, 0.60	3.52	4	398	18, 44, 92, 101
HCR–20, adding the items	0.18	0.40	0.10, 0.70	2.68	3	232	18.4, 44, 92
Structured professional judgment/sex	0.40	0.46	0.29, 0.62	7.99	6	1,131	19, 25, 54, 63, 65, 85
With outlier	0.44	0.59	0.43, 0.74	24.06***	7	1,252	23
SVR–20	1.23	1.11	0.82, 1.40	7.96*	3	245	23, 25, 85
Unstructured professional judgment	0.44	0.42	0.32, 0.51	14.56	11	6,456	11, 26, 48, 54, 75, 82, 83, 94, 103a, 103b, 105

Note. RRASOR = Rapid Risk Assessment of Sexual Offense Recidivism (Hanson, 1997); MnSOST = Minnesota Sex Offender Screening Tool—Revised (Epperson et al., 1995); SACJ–Min = Structured Anchored Clinical Judgment Scale—Minimum criteria; VRAG = Violence Risk Appraisal Guide(Quinsey et al., 2006); SORAG = Sex Offender Risk Appraisal Guide (Quinsey et al., 2006); SIR = Statistical Index of Recidivism (Bonta et al., 1996); LSI–R = Level of Service Inventory—Revised (Andrews et al., 2004); SVR–20 = Sexual Violence Risk—20 (Boer et al., 1997); J-SOAP = Juvenile Sex Offender Assessment Protocol (Prentky, Harris, Frizzell, & Righthand, 2000); ERASOR = Estimate of Risk of Adolescent Sexual Offense Recidivism (Worling & Curwen, 2001); HCR–20 = Historical–Clinical Risk—20 (Webster et al., 1997).

Table A2
Prediction of Violent Recidivism

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
Actuarial empirical/ sex	.50	.51	.47, .56	103.44***	42	16,625	4, 9, 11, 12, 14, 16, 17, 18, 20, 21, 22, 25, 27, 29, 38, 40, 41, 44, 47, 48, 49, 54, 55, 59, 60, 65, 71, 72, 76, 79, 80, 84, 90, 92, 93, 95, 96, 97, 101, 103, 106, 108
Static–99	.51	.57	.52, .62	98.85***	35	14,197	4.1, 9, 11, 12, 14, 16, 17, 18.2, 20, 22.1, 27, 29, 38, 41, 44, 47, 48, 49a, 54, 55, 60, 65, 72.1, 76, 79, 80.2, 84, 90, 92, 93, 95, 96, 103.4, 106, 108
Static–2002	.61	.65	.54, .75	3.20	5	1,890	11, 12, 54.1, 72.1, 80.2
With outlier	.66	.72	.63, .82	12.48*	6	2,592	27.2
RRASOR	.30	.32	.26, .38	66.46***	21	7,269	4.1, 11, 18.2, 20, 22, 25, 38, 41, 54, 60, 65, 72.1, 76, 79.1, 80.1, 90, 95, 97, 103.1, 106, 108
MnSOST–R	.28	.33	.20, .47	3.35	3	2,481	4.1, 17, 54
SACJ–Min	.36	.36	.22, .49	3.91	5	1,267	20, 21, 40, 71, 93
Risk Matrix 2000, sexual	.50	.49	.34, .63	.85	4	809	11, 20, 79.2, 106
Actuarial empirical/ violence	.79	.78	.70, .86	41.74***	15	3,485	4, 11, 14, 20, 25, 29, 41, 47, 54, 77, 78, 79, 80, 103, 106
VRAG	.94	.87	.75, .98	13.47*	6	1,499	14, 25, 41, 54, 78, 103.4
SORAG	.72	.77	.67, .87	9.55	8	2,269	4.1, 29, 41, 54, 77, 79, 80, 103.4
With outlier	.74	.81	.71, .91	23.52**	9	2,364	25
Risk Matrix 2000, violence only	.74	.62	.47, .77	5.22	3	782	11, 20, 79.2
With outlier	.60	.58	.43, .73	13.74**	4	809	106
Risk Matrix 2000, sexual/violence	.70	.75	.63, .87	15.51**	5	1,406	11, 47.11, 47.12, 79.2, 106
Actuarial empirical/ any recidivism	.73	.79	.63, .95	1.88	5	860	13, 14, 22, 65, 102
SIR	.73	.80	.61, .99	1.36	3	566	13, 14, 22
LSI–R (and variants)	.73	.72	.47, .97	1.36	3	390	14, 65, 102
LSI–R (and variants), adjusted	.54	.54	.18, .91	0.00	2	242	33, 102

(*table continues*)

Table A2 (*continued*)

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
Mechanical/sex	.34	.40	.31, .49	36.26***	10	3,376	20, 25, 27, 38, 44, 54, 65, 92, 101, 103
SVR–20, adding the items	.27	.39	.23, .55	7.65	4	760	20, 44, 54, 92
With outlier	.47	.50	.34, .65	20.54***	5	855	25
Mechanical/violence	.28	.32	.07, .56	2.25	3	375	44, 92, 101
Structured professional judgment/sex	.39	.31	.13, .49	12.75**	3	579	25, 54, 65
Unstructured professional judgment	.30	.22	.15, .29	6.37	7	5,705	11, 48, 54, 75, 94, 103a, 103b,
With outlier	.30	.24	.17, .31	19.33**	8	6,040	53

Note. RRASOR = Rapid Risk Assessment of Sexual Offense Recidivism (Hanson, 1997); MnSOST–R = Minnesota Sex Offender Screening Tool—Revised (Epperson, 1995); SACJ–Min = Structured Anchored Clinical Judgment—Minimum criteria; VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); SORAG = Sex Offender Risk Appraisal Guide (Quinsey et al., 2006); SIR = Statistical Index of Recidivism (Bonta et al, 1996); LSI–R = Level of Service Inventory—Revised (Andrews et al., 2004); SVR–20 = Sexual Violence Risk—20 (Boer et al., 1997).

Table A3
Prediction of Any Recidivism

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
Actuarial empirical/sex	0.50	0.52	0.48, 0.56	113.12***	43	15,015	4, 5a, 5b, 11, 14, 17, 18, 20, 21, 22, 23, 24, 27, 28, 29, 36, 37, 38, 40, 45, 47, 49a, 49b, 51, 54, 55, 56, 59, 61, 65, 71, 72, 76, 79, 80, 90, 92, 93, 96, 97, 101, 103, 106
Static–99	0.50	0.53	0.49, 0.57	87.30***	35	12,523	4.1, 5a, 5b, 11, 14, 17, 18.3, 20, 21.1, 22.1, 23, 24, 27, 28, 29, 38, 45, 47, 49a, 49b, 51.3, 54, 55, 56, 61, 65, 72, 76, 80.2, 90, 92, 93, 96, 103.4a, 106
Static–2002	0.61	0.63	0.53, 0.73	0.82	4	1,816	11, 54.1, 72, 80.2
With outlier	0.62	0.73	0.64, 0.81	13.49**	5	2,518	27.2
RRASOR	0.25	0.27	0.20, 0.34	8.44	13	4,000	4.1, 11, 20, 22, 38, 54, 65, 72, 76, 90, 97, 103.1, 106
MnSOST–R	0.60	0.43	0.34, 0.53	7.59	4	2,511	4.1, 17, 37, 54
With outlier	0.62	0.45	0.35, 0.54	16.36**	5	2,541	36
SACJ–Min	0.33	0.44	0.32, 0.56	16.02**	5	1,267	20, 21, 40, 71, 93
Risk Matrix 2000, sexual	0.46	0.50	0.38, 0.62	7.69	6	1,257	11, 20, 21.1, 55, 79.2, 106
Actuarial empirical/violence	0.80	0.74	0.66, 0.82	36.55***	14	2,700	4, 10, 11, 14, 20, 24, 28, 29, 42, 54, 77, 79, 103, 106
VRAG	0.74	0.76	0.62, 0.89	3.72	4	897	14, 24, 54, 103.4a
SORAG	0.82	0.82	0.72, 0.92	9.35	8	1,741	4.1, 10, 28, 29, 42, 54, 77, 103.1a
Risk Matrix 2000, violence only	0.53	0.55	0.40, 0.69	28.78***	4	809	11, 20, 79.2, 106
Risk Matrix 2000, sexual/violence	0.45	0.64	0.48, 0.79	8.95*	3	724	11, 79.2, 106
Actuarial empirical/any recidivism	1.00	0.97	0.93, 1.01	38.12***	10	14,816	10, 13, 14, 18, 22, 45, 50, 65, 102, 103
SIR	1.02	1.02	0.85, 1.20	2.28	4	626	13, 14, 18.6, 22
With outlier	1.10	1.09	0.92, 1.26	10.33**	5	679	10.1
LSI–R (and variants)	0.58	0.69	0.47, 0.91	5.83	3	390	14, 65, 102
LSI–R (and variants), adjusted	0.76	0.70	0.41, 0.98	0.44	2	242	33, 102
Mechanical/sex	0.51	0.37	0.30, 0.43	68.06***	19	4,319	20, 23, 24, 27, 38, 43, 45, 46, 54, 55, 57, 58, 61, 65, 70, 87, 92, 101, 103
SVR–20, adding the items	0.67	0.58	0.43, 0.72	7.82	5	839	20, 23, 24, 54, 92

(*table continues*)

Table A3 (*continued*)

Variable	*Mdn*	*M*	95% CI	*Q*	*k*	Total	Studies
J-SOAP	0.10	0.16	−0.05, 0.37	4.52	4	413	43, 58, 70, 101
With outlier	0.21	0.24	0.04, 0.43	10.18*	5	476	57
ERASOR, adding the items	0.29	0.28	0.05, 0.50	5.92	3	316	58, 65, 87
Structured professional judgment/sex	0.26	0.26	0.11, 0.41	6.12	4	882	19, 23, 54, 65
Unstructured professional judgment	0.08	0.11	0.06, 0.17	20.97**	9	5,801	26, 32, 52, 54, 75, 83, 103a, 103b, 110

Note. RRASOR = Rapid Risk Assessment of Sexual Offense Recidivism (Hanson, 1997); MnSOST = Minnesota Sex Offender Screening Tool—Revised (Epperson, 1995); SACJ–Min = Structured Anchored Clinical Judgment—Minimum criteria; VRAG = Violence Risk Appraisal Guide (Quinsey et al., 2006); SORAG = Sex Offender Risk Appraisal Guide (Quinsey et al., 2006); SIR = Statistical Index of Recidivism (Bonta et al., 1996); LSI–R = Level of Service Inventory—Revised (Andrews et al., 2004); SVR–20 = Sexual Violence Risk—20 (Boer et al., 1997); J-SOAP = Juvenile Sex Offender Assessment Protocol (Prentky et al., 2000); ERASOR = Estimate of Risk of Adolescent Sexual Offense Recidivism (Worling & Curwen, 2001).

Table A4
Key to Studies Used in the Meta-Analysis

No.	Study	No.	Study
1	Allam (1999)	54	Langton (2003)
2	A. Allan & Dawson (2002)	54.1	Langton et al. (2007)
3	A. Allan et al. (2006)	55	[London Probation] Craissati et al. (2005)
4	[Arizona] Fischer (2000)	55.1	[London Probation] Webb et al. (2007)
4.1	[Arizona] Bartosh et al. (2003)	56	Marghem (2007)
5(a, b)	Austin et al. (2003)	57	Martinez et al. (2007)
6	Bates et al. (2004)	58	McCoy (2007)
7	Beech & Ford (2006)	59	McGrath et al. (2005)
8	Beech et al. (2002)	60	McGrath et al. (2001)
9	Beech et al. (2004)	61	Milton (2003)
9.1	Beech (2005)	62	Min-chieh & Tzu-yi (2005)
10	Bélanger & Earls (1996)	63	[Minnesota] Epperson et al. (1995)
10.1	Bélanger (1995)	63.1	[Minnesota] Epperson et al. (2000)
11	Bengtson & Långström (2007)	63.2	[Minnesota] Brown (2003)
11.1	Bengtson (2007)	64	Montana & Thompson (2005)
12	Boer (2003)	65	Morton (2003)
13	Bonta & Hanson (1995)	66	Ohio (2001)
14	Bonta & Yessine (2005)	67	Parks (2004)
15	[Bridgewater] Thornton & Knight (2006)	68	Pérez et al. (2008)
15.1	[Bridgewater] Knight & Thornton (2007)	69	Poole et al. (2000)
16	Bright et al. (2007)	70	Prentky et al. (2000)
17	Caperton (2005)	71	Proulx et al. (1995)
18	[Clearwater] Haynes et al. (2000)	72	[Quebec] Bigras (2007)
18.1	[Clearwater] Nicholaichuk (1997)	72.1	[Quebec] Proulx (2004)
18.2	[Clearwater] Nicholaichuk (2001)	73	Ralston & Epperson (2006)
18.3	[Clearwater] Olver (2003)	74	Ralston & Epperson (2007)
18.4	[Clearwater] Witte et al. (2001)	75	Reddon et al. (1996)
18.5	[Clearwater] Olver et al. (2007)	76	Rettenberger & Eher (2006)
18.6	[Clearwater] Witte et al. (2006)	77	Rettenberger & Eher (2007)
19	Cotton (1991)	78	Rice & Harris (1997)
20	Craig et al. (2006)	79	[ROH] Nunes et al. (2002)
20.1	Craig et al. (2007)	79.1	[ROH] Nunes et al. (2003)
21	Craissati et al. (2002)	79.2	[ROH] Kingston et al. (2007)
21.1	Craissati et al. (2008)	80	[RTC] Looman (2006)
22	[CSC] Motiuk & Brown (1995)	80.2	[RTC] Looman et al. (2005)
22.1	[CSC] Cortoni & Nunes (2007)	80.3	[RTC] Helmus (2007)
23	de Vogel et al. (2004)	81	Saum (2007)
24	Dempsey (2002)	81.1	Saum (2005)
25	Dempster (1998)	82	Schiller (2000)
25.1	[Dempster] Kropp (2000)	83	Schram et al. (1991)
26	Dix (1976)	84	Seager et al. (2004)

(*table continues*)

Table A4 (*continued*)

No.	Study	No.	Study
27	[DSP] Hanson et al. (2007)	85	Sjöstedt & Långström (2002)
27.1	[DSP] Hanson (2007)	86 (a,b,c)	Skelton et al. (2006)
27.2	[DSP] Helmus & Hanson (2007)	87	Skowron (2005)
28	Ducro & Pham (2006)	88	Smiley et al. (1998)
29	Dufresne (2005)	89	Smith & Monastersky (1986)
30	Endrass et al. (in press)	90	Song & Lieb (1994)
31(a,b)	Epperson (2003)	91	Soothill et al. (2005)
32	Florida (1985)	92	Stadtland et al. (2005)
33	Girard & Wormith (2004)	93	Stalans et al. (2002)
34	Gore (2007)	94	Sturgeon & Taylor (1980)
35	Haag (2005)	95	[Sweden] Sjöstedt & Långström (2001)
36	Hanlon & Herlickson (2007)	95.1	[Sweden] Långström (2004)
37	Hanlon et al. (2004)	96	Ternowski (2004)
38	Hanson (2002)	97	Thornton (1997)
39	Hanson & Harris (2000)	98	Thornton (2000)
40	Hanson et al. (1993)	99	Thornton (2002)
41	Harris et al. (2003)	100	Tough (2001)
42	Hartwell (2001)	101	Viljoen et al. (2008)
43	Hecker et al. (2002)	102	Vrana et al. (2008)
44	Hill et al. (2008)	103	[Washington State] Barnoski (2005)
45	Hills (2003)	103.1	[Washington State] Barnoski (2006c)
46	Hiscox et al. (2007)	103.2	[Washington State] Barnoski (2006a)
47	[HM Prison] Friendship et al. (2003)	103.3	[Washington State] Barnoski (2006b)
47.1 (a, b)	[HM Prison] Thornton et al. (2003)	103.4 (a,b)	[Washington State] Johansen (2007)
48	Hood et al. (2002)	104	Watanabe et al. (2007)
49(a, b)	Hudson (2003)	105	Wieand (1983)
50	Johnson (2006)	106	Wilcox et al. (2008)
51	[Kia Marama] Hudson et al. (2002)	107	Williams (2007)
51.1	[Kia Marama] Beggs & Grace (in press)	108	Wilson & Prinzo (2001)
51.2	[Kia Marama] M. Allan et al. (2007)	108.1 (a,b)	Wilson, Picheca, & Prinzo (2007)
51.3	[Kia Marama] Beggs & Grace (2008)	109	Wilson, Cortoni, & Vermani (2007)
52	Kolko (2005)	110	Wormith & Ruhl (1986)
53	Kozol et al. (1972)		

Note. Studies with the same number but different decimal points are different reports from the same or overlapping samples. When a study number has letters in parentheses, distinct subsamples were drawn from the same report.

Received November 3, 2006
Revision received September 26, 2008
Accepted October 6, 2008 ■