The Psychological Assessment of Risk for Crime and Violence

Article *in* Canadian Psychology/Psychologie canadienne · August 2009

DOI: 10.1037/a0015726

R. Karl Hanson
Carleton University

Canadian Psychology
2009, Vol. 50, No. 3, 172–182

0708-5591/09/$12.00 DOI: 10.1037/a0015726

The Psychological Assessment of Risk for Crime and Violence

R. Karl Hanson
Public Safety Canada

This article provides an overview of the current practices and challenges in psychological risk assessment for crime and violence. Risk assessments have improved considerably during the past 20 years. The dismal predictive accuracy of unstructured professional opinion has largely been replaced by more accurate, structured risk assessment methods. Consensus has not been achieved, however, on the constructs assessed by the various risk tools, nor the best method of combining factors into an overall evaluation of risk. Advancing risk assessment for crime and violence requires psychometrically sound evaluations of psychologically meaningful causal risk factors described using nonarbitrary metrics.

Keywords: risk assessment, violence, recidivism

Risk assessments for crime and violence are different from other forms of psychological assessment because the presenting problem is not directly observed. Therapists working with depression and anxiety routinely see clients who are visibly depressed and anxious. In contrast, therapists working with criminals may never see a crime. Risk assessment involves estimating the probability of a future event based on secondary, indicator variables. The general principles of assessment for the indicator variables are the same as for other areas of psychology (e.g., reliability, validity, observed, and latent constructs); the special concerns of violence risk assessment are the selection of factors to assess, and the methods for combining the factors into an overall evaluation of risk.

Psychological risk assessments are frequently requested for individuals who have violated social norms or displayed bizarre behavior, particularly when they appear menacing or unpredictable. Crucial decisions in the criminal justice, mental health, and child protection systems are predicated on the risk of harm. Individuals can even be detained involuntarily for risk assessment should they be suspected of being dangerous (see the Criminal Code of Canada, Section 752.1(1), and the mental health legislation for all the provinces and territories).

Some knowledge of risk assessment is required for all mental health professionals. The famous 1976 Tarasoff decision established the duty to protect third parties from the risk presented by mental health clients (Tarasoff v. Regents of the University of California, 1974, 1976). Although the Tarasoff case was based in California, almost every jurisdiction in the United States and Canada have followed its legal reasoning (Truscott, 1993; Truscott & Crook, 2004), and the basic recommendations are enshrined in the Code of Ethics of the Canadian Psychological Association (CPA, 2000, Section § II.39).

Many of the (better) ideas in this article were developed during ongoing conversations with David Thornton and Ruth Mann. I thank Kelly Babchishin for help with references and library searches. The views expressed are those of the author and not necessarily those of Public Safety Canada.

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

The Canadian case most analogous to Tarasoff (Smith v. Jones, 1999), involved a mental health professional (a psychiatrist) who took steps to inform the police that a private client was likely to kill prostitutes on Vancouver's lower east side. No one was killed, and the client subsequently attempted to sue the psychiatrist for breach of confidentiality. The Supreme Court of Canada eventually dismissed the suit against the psychiatrist on the grounds that the duty to protect outweighed confidentially. The Supreme Court specified that "[t]hree factors should be taken into consideration in determining whether public safety outweighs solicitor-client privilege: (a) Is there a clear risk to an identifiable person or group of persons? (b) Is there a risk of serious bodily harm or death? (c) Is the danger imminent?" (Smith v. Jones, 1999, ¶77).

Risk assessments involve judgments about uncertainty. They are formal methods of giving shape to our fears of future harm. Risk assessments can limit the range of plausible speculation, but they are never certain. The causal mechanisms for human behavior are sufficiently complex that it is impossible to reason through all the possibilities. Consequently, risk assessments are inherently stochastic. Risk assessors must consider the uncertainty intrinsic to the phenomenon being assessed in addition to the measurement error common to all psychological assessment.

Reporting Risk Assessments

An important consideration is the language with which the "likelihood" is communicated (Grann & Pallvik, 2002; Monahan, Heilbrun, & Silver, 2002). It is not uncommon that risk is presented as "low," "moderate," or "high" (Heilbrun et al., 2004). The problem with such descriptors is that they have no inherent scientific meaning, and are prone to divergent interpretations (Hilton, Carter, Harris, & Sharpe, 2008). One numeric approach to reporting risk is to use probabilities (e.g., Mr. Jones has a 30% chance of being convicted for a violent offence within 2 years). This approach has merit, but requires considerable qualification. Absolute rates vary based on the time periods specified, the outcome criteria used, and the local base rates for that outcome. In many cases, there is insufficient data to support precise estimates of absolute probabilities.

Absolute recidivism rates are not needed, however, for certain decisions. For example, police may only have the resources to verify the addresses of 15% of registered sexual offenders, and want to target the riskiest cases. For such decisions, risk could be presented as percentile ranks. It is also possible to express relative risk in terms of risk ratios (hazard ratios) when the absolute recidivism rates are unknown (see discussion of Cox regression below). For example, Mr. Jones could be described as being 2.5 times more likely to reoffend violently than the typical offender. In general, relative risk estimates would be expected to be more stable across settings than estimates of absolute risk because variance resulting from base rates is removed. Risk ratios are difficult to interpret, however, in the absence of base rate information. Most decision-makers care whether the risk for violence increases from 3.0% to 7.5%, or from 30% to 75%.

Typically, the consumer of the risk assessment report wants more than a number. Not only do decision-makers want an estimate of the likelihood of failure, but they also want an estimate of the potential consequences, and what can be done to mitigate the risk. The ideal risk assessment would have other desirable features as well. The following list summarizes some features to which risk assessment procedures should aspire:

- Assess risk factors whose nature, origins, and effects can be understood
- Enable reliable and valid assessment of clinically useful causal factors
- Provide precise estimates of recidivism risk
- Allow all relevant factors to be considered
- Inform the development of treatment targets and risk management strategies
- Allow the assessment of both long term and short term changes in risk
- Incorporate protective factors as well as risk factors
- Facilitate engaging the patient/offender in the assessment process
- Be easy to implement in a broad range of settings

How close are current practices to the ideal? There has been considerable progress in psychological risk for crime and violence during the past 25 years (Hanson, 2005). There is still, however, considerable room for improvement.

Overview of Risk Assessment Practices

In the 1980s, there was widespread pessimism concerning the ability of mental health experts to conduct assessments of dangerousness. Prominent reviews concluded that violence risk assessment was "doomed" (Monahan, 1976). Forensic evaluators were wrong twice as often as they were right (Monahan, 1981), and consistently overpredicted violence (Steadman & Cocozza, 1974). The risk prediction methods used at that time largely involved clinical judgment, in which the probability of violence was inferred from psychological structures and dynamics. Using current language, these risk assessments would be considered to be "unstructured professional opinion," meaning that neither the risk factors nor the method of combining the risk factors were specified in advance. Unstructured professional judgment evaluations have some ability to distinguish between the most and less "dangerous" offenders (Hanson & Morton-Bourgon, in press; Mossman, 1994), but they are not particularly "professional": the judgments of mental health experts were no different from those of otherwise intelligent lay people (Quinsey & Ambtman, 1979).

There are, however, better ways of conducting risk assessments. When risk evaluations focus on empirically based risk factors, their accuracy is substantially better than that provided by unstructured professional opinion (see reviews by Andrews, Bonta, & Wormith, 2006; Janus & Prentky, 2003; Monahan, 2007a; Quinsey, Harris, Rice, & Cormier, 2006). Empirically based risk assessments are not new. In 1928, Burgess published an empirically derived actuarial method for predicting recidivism amongst parolees (Burgess, 1928). Although Burgess's scale was not widely used, scales derived using similar methods have been widely adopted for risk assessment of general offenders. Variations on The Salient Factors Score (SFS; Hoffman, 1994; Hoffman & Beck, 1974) have been used since 1973 to assess parole releases from U.S. federal prisons. During the 1980s, the Correctional Service of Canada adopted the Statistical Information on Recidivism scale (SIR; Nuffield, 1982) to assess the risk of male offenders. These scales include static, historical factors (e.g., prior offences, young age) combined into a total score based on empirically derived weighting systems. Both the SFS and SIR scales predict general and violent recidivism with respectable levels of accuracy (Bonta, Harman, Hann, & Cormier, 1996; Hoffman, 1994; Mills, Kroner, & Hemmati, 2004).

These "second generation" risk assessment procedures (Bonta, 1996) were a clear advance over the "first generation" (i.e., unstructured professional judgment), but had little influence on psychological risk assessments. The second generation scales were not marketed to psychologists. They were not published in psychology journals nor was the content of these scales particularly "psychological." Instead, they were developed for, and used by, decision-makers within the criminal justice system. The risk factors were based on empirical associations with recidivism for the types of data routinely collected in corrections (e.g., offence history, demographics). These scales were not linked to, nor guided by, psychometric or behavioral theory.

The modern era of psychological risk assessment for crime and violence was led by Robert Hare's careful and committed development of the concept of psychopathy (Hare, 1980, 2003). The Hare Psychopathy Checklist–Revised (PCL-R) provided a psychometrically sound assessment of a psychologically interesting construct that was empirically related to crime and violence (e.g., Walters, 2003). Although the PCL-R predicted violent recidivism no better than the SIR scale (Campbell, French, & Gendreau, in press; Mills & Kroner, 2006), the PCL-R has received considerable attention amongst mental health professionals, compared to an almost complete neglect of the SIR.

The 1990s saw the rapid introduction of structured violence risk assessment tools, many of which include the PCL-R as an item (Violence Risk Appraisal Guide [VRAG], G. Harris, Rice, & Quinsey, 1993; Historical Clinical Risk-20 [HCR-20], Webster, Douglas, Eaves, & Hart, 1997; Sexual Violence Risk-20 [SVR-20], Boer, Hart, Kropp, & Webster, 1997). The superiority of structured, actuarial methods over clinical judgment has been documented for at least 50 years (Meehl, 1954). It was not until recently, however, that structured risk tools have been routinely used in forensic risk assessment (Archer, Buffington-Vollum, Stredny, & Handel, 2006). In high stakes evaluations, such as civil commitment procedures for sexual offenders, most evaluators now consider structured risk tools to be essential (Jackson & Hess, 2007).

For the assessment of violence and general recidivism, Archer et al. (2004) found that the measures most commonly used by forensic psychologists are the HCR-20, the VRAG, and the Level of Service Inventory–Revised (LSI-R; Andrews & Bonta, 1995; Andrews, Bonta, & Wormith, 2004). Static-99 (Hanson & Thornton, 2000; see www.static99.org) is by far the most commonly used risk tool for sexual offenders, followed by the SVR-20, the Minnesota Sex Offender Screening Tool–Revised (MnSOST-R; Epperson, Kaul, Huot, Goldman, & Alexander, 2003) and the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, et al., 2006). Specialized risk assessment tools have been developed for wife assaulters (Hilton et al., 2004; Kropp, Hart, Webster, & Eaves, 1999), young offenders (Borum, Bartel, & Forth, 2006; Hoge & Andrews, 2002), and young sexual offenders (Prentky, Harris, Frizzell, & Righthand, 2000; Worling & Curwen, 2001). For brief descriptions of the various risk tools currently available, see Craig, Browne, Stringer, and Beech (2005); Daffern (2006); Doyle and Dolan (2007); Loza (2003); and Otto and Douglas (in press). The association with recidivism of the most common risk tools is presented in Table 1.

How Should Risk Assessments Be Structured?

Although there is consensus that structured risk assessments are more accurate than unstructured professional opinion, there is considerable debate about how risk assessments should be structured (Monahan, 2007a). The three leading candidates are (a) structured professional judgment (Douglas, Cox, & Webster, 1999; Hart, 1998), (b) fully actuarial risk assessment (Quinsey, Harris, et al., 2006), and (c) clinical interpretation of actuarial results (Hanson, 1998; Monahan, 2008).

Structured professional judgment is the most widely used approach to risk assessment amongst forensic psychologist (Archer et al., 2006). In this approach, the risk factors are specified in advance, but the overall assessment of risk is left to professional judgment. No explicit rules are provided for combining risk factors into a total score. Examples of structured professional judgment tools include the HCR-20 (Webster et al., 1997) and the SVR-20 (Boer et al., 1997). Although the authors of the SVR-20 suggest that risk should increase monotonically with the number of risk factors present, evaluators are discouraged from simply adding the items. Instead, evaluators are directed to use their professional experience to form an overall judgment of risk. This risk judgment is expressed as "low," "moderate," or "high." No attempt is made to link these risk categorizations to estimated probabilities of outcomes.

In contrast, the developers of the VRAG argue for a fully actuarial approach to risk assessment (Quinsey, Harris, et al., 2006). In this approach, the items are specified in advanced and a total score is computed based on empirically derived rules. The total score is then linked to the outcome based on a table of probabilities. The role of professional judgment is limited to selecting the appropriate actuarial tool, and to scoring of the items. Quinsey and colleagues (2006) believe that "[a]ctuarial measures are too good and clinical judgement is too poor to risk contaminating the former with the latter" (p. 197).

Another common approach to risk assessment involves using one or more empirical-actuarial risk tools as part of an overall risk evaluation (e.g., Hanson, 1998; Monahan, 2008). In sexual offender civil commitment evaluations, for example, 79.5% of evaluators report using more than one structured risk tool (Jackson & Hess, 2007). Given that there are no empirical rules for combining the results of different instruments, evaluators using more than one instrument must interpret the actuarial results. The extent to which these clinical interpretations add information or noise to the risk assessment is not known. The few studies that have examined professional overrides of empirically derived actuarial tools have consistently found overrides to degrade predictive accuracy (Gore, 2007; Hanson, 2007; Hilton & Simmons, 2001; Krauss, 2004; Vrana, Sroga, & Guzzo, 2008).

Clinical interpretations of actuarial results, however, are often unavoidable. The results of the various actuarial tools often differ (Barbaree, Langton, & Peacock, 2006a), the same tool can produce different estimates for different samples (Hanson, Helmus, & Thornton, in press), and offenders often display risk factors known to interact with actuarial scores (e.g., G. Harris et al., 2003; Thornton, 2002). Furthermore, scoring an actuarial risk tool is not a risk assessment. Evaluators will always need to make a separate judgment as to whether the risk scale score fairly represents the risk posed by the individual being assessed.

The current evidence does not clearly support any single approach to risk assessment. Respectable levels of predictive accuracy have been found for purely actuarial measures, structured professional judgment, and the mechanical combination of items from structured risk schemes (Campbell et al., in press; Daffern, 2006; Douglas, Guy, & Weir, 2006; Hanson & Morton-Bourgon, 2009). Although it is possible to find specific studies favoring a particular measure or approach, the results have not been consistent. Importantly, the evidence has not identified the conditions under which one approach would be expected to be superior to others.

Table 1
Correlation With Recidivism of Selected Risk Scales (Summary of Meta-Analyses)

Scales	Outcome criteria: General recidivism[a]	Violent recidivism[b]	Sexual recidivism sexual offenders[c]
HCR-20 (totals)	—	.22	.19
LSI/LSI-R	.40	.27	.22
PCL/PCL-R	.27	.26	.14[d]
SIR Scale	.42[e]	.22	.24
VRAG	—	.31	.25
Measures for sexual offenders[c]			
Static-99	.27	.25	.31
SVR-20 (totals)	.19	.28	.32
MnSOST-R	.21	.16	.36
SORAG	.40	.34	.30

Note. Base rate of 50% was assumed for transforming *d* values into *r*. [a] From Gendreau, Goggin, & Smith (2002). [b] From Campbell, French, & Gendreau (in press). [c] From Hanson & Morton-Bourgon (2009). [d] From Hanson & Morton-Bourgon (2005). [e] From Bonta, Harman, Hann, & Cormier (1996). Based on single sample of 3,267 (not meta-analysis).

There are strong arguments that pure empirical-actuarial approaches should ultimately prove the most accurate–and recent meta-analyses tend to favor the pure actuarial measures (Campbell et al., in press; Hanson & Morton-Bourgon, 2009). If a particular actuarial tool is found to be less accurate than any other measures, actuarial test developers can simply change the tool based on the new evidence. The need for actuarial risk tools to be replicated on large samples, however, limits the speed at which actuarial risk schemes can assimilate new information. Consequently, evaluators will frequently be faced with the dilemma of knowing that a risk factor contributes incrementally to a risk assessment scheme, but not knowing how to include the external risk factor in the evaluation (Hanson, 1998).

In many decision-making contexts, pure actuarial prediction has been found to be the most accurate approach (Dawes, Faust, & Meehl, 1989). There is, however, one important exception: meteorology. Weather prediction is a useful area to study prediction because the outcome is rapid, easily measured, and not influenced by the risk decision (unlike violence risk assessments). The most accurate approach to weather prediction involves expert interpretation of actuarial indicators (see Swets, Dawes, & Monahan, 2000, p. 18). The adjusted actuarial predictions of temperature and precipitation are consistently more accurate than the unadjusted actuarial predictions. Consequently, despite the arguments and evidence currently favoring pure actuarial prediction for crime and violence, it is possible that future research may find even better approaches.

Given valid risk factors, the ways that they are combined makes relatively little difference. Considerable statistical work has been devoted to optimizing the prediction for specific samples. In new samples, however, simple "1 = present, 0 = absent" rating schemes typically do as well or better than more complex alternatives at predicting crime and violence (Grann & Långström, 2007; Silver, Smith, & Banks, 2000). Kroner, Mills, and Reddon (2005) even found that randomly selected items from the PCL-R, LSI-R, VRAG, and SIR scales were as accurate at predicting recidivism as any of the original scales. Consequently, the primary concern of risk evaluators should be to ensure that they focus on relevant risk factors, and ignore irrelevant factors.

Identifying Risk Factors

Common sense is not a reliable guide to what matters in risk assessment. Research has found, for example, that the seriousness of the index offence and major mental illness have minimal relationship to new offences for mentally disordered offenders (Bonta, Law, & Hanson, 1998). Amongst sexual offenders, denial, poor social skills, and low victim empathy have no relationship to sexual recidivism (Hanson & Morton-Bourgon, 2005). Evidence is needed to identify valid risk factors.

Although it is possible to identify individual risk factors through case analysis (idiographic or "anamnestic" risk assessment), it is difficult to generalize the factors identified to other individuals. For example, Jack may have a repeated pattern of losing his job, playing violent video games, then assaulting family members. Knowing Jack's patterns can have considerable utility in case management, as Jack's family is no doubt aware. Group data are required, however, to determine whether Jack's personal risk factors (job loss, video games) apply to other individuals.

The simplest method for identifying risk factors are case-control studies. In these studies, individuals with the outcome of interest (e.g., recent violence) are compared to other individuals. These studies provide only weak evidence concerning risk factors because the temporal ordering of the features is unknown (e.g., did he hate the police before he was caught?). Furthermore, there are inherent ambiguities concerning who the index cases should be compared to (nonviolent patients? men? young, unemployed men?). Matching on certain variables virtually guarantees that they will be mismatched on other variables (Meehl, 1970). Nevertheless, case-control studies are a useful starting point for identifying potential risk factors (Whitaker et al., 2008), and it may be the only practical method for low frequency outcomes (e.g., violent extremism, homicide in schools).

Compared to case-control studies, follow-up studies provide much more convincing evidence. Some authorities even reserve the use of the term "risk factor" to characteristics that have been empirically demonstrated to precede (and predict) a future outcome (Kraemer et al., 1997).

Quantitative summaries (meta-analyses) are a useful guide concerning the extent to which risk factors generalize across samples (Cooper & Hedges, 1994; Hanson & Broom, 2005; Hunt, 1997). Several large-scale meta-analyses of risk factors for crime and violence have been conducted (Bonta, Law, & Hanson, 1998; Gendreau, Little, & Goggin, 1996; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2004, 2005), and the results are remarkably consistent (see Table 2). All the effects in Table 2 are

Table 2
Correlation With Recidivism of Selected Demographic, Criminal History, and Psychosocial Characteristics

	Outcome criteria		
Variables	General recidivism adult offenders[a]	Violent recidivism mentally disordered offenders[b]	Sexual recidivism sexual offenders[c]
Demographic			
Age	−.11	−.16	−.13
Minority race	.17	.12	.00
Criminal history			
Juvenile	.13	.27	.12
Adult	.17	.19	.13
Violence in index offence	—	.00	.04
Prior violence	—	.16	.09
Psychosocial characteristics			
Negative/criminal companions	.21	—	.13
Substance abuse	.10	.08	.06
Antisocial personality disorder/ psychopathy	.18	.18	.14
Personal distress/mood disorder	.05	.01	−.01[d]
Psychosis	.00	−.04	−.01
Low intelligence	.07	−.02	.09
Deviant sexual interests	—	—	.15

[a] From Gendreau, Little, & Goggin (1996). [b] From Bonta, Law, & Hanson (1998). [c] From Hanson & Bussière (1998), Hanson & Morton-Bourgon (2004, 2005), and Mann, Hanson, & Thornton (2009). A base rate of 50% was assumed for transforming d values into r. [d] Average of the findings for anxiety (.03) and depression (−.06).

based on large samples (mostly 1,000+) and have been replicated in at least three studies. The largest predictors of criminal and violent recidivism are a cluster of externalizing behaviors, which include antisocial personality disorder, substance abuse, criminal companions, and a history of rule violation. Major mental illness and internalizing psychological problems, such as anxiety and depression, are largely unrelated to recidivism risk.

The risk factors are similar for general, violent, and sexual recidivism—with certain exceptions. Sexual deviance has particular importance for sexual recidivism. Minority race appears more important for general or violent recidivism than for sexual recidivism. The following sources provide more detailed reviews of the risk factors for general recidivism (Andrews & Bonta, 2006), sexual recidivism (Craig et al., 2005; Hanson & Morton-Bourgon, 2004, 2005) and violent recidivism (Quinsey, Harris, et al., 2006).

Risk Statistics

As evidenced-based risk assessment gains popularity, evaluators need increasing statistical literacy to critically examine the research evidence (see Monahan, 2007b). Consequently, a general guide to commonly used statistics may help readers as they delve into the risk prediction literature.

The most familiar of the statistics linking a predictor variable to an outcome is the Pearson product-moment correlation coefficient – *r*. When the outcome criteria is dichotomous, *r* becomes the point-biserial correlation; when both variables are dichotomous, *r* becomes *phi.* The value of phi coefficients can be interpreted as (approximately) the difference in the probability of failure for the two groups. For example, the relationship between substance abuse and general criminal recidivism is .10 (see Table 2). Given a recidivism base rate of 40%, substance abusing offenders would be expected to have a recidivism rate of 45% compared to a rate of 35% for offenders without a history of substance abuse ($45 - 35 = 10$). Expressed as correlations, the magnitude of individual factors tends to be in the .10 to .20 range (see Table 2), and in the .20 to .35 range for risk scales (see Table 1). The major problem with *r* is that it varies widely with restriction of range in the predictor variables and changes in the base rate (e.g., *r* changes from .55 to .28 as the base rate changes from 50% to 5%). Consequently, most statistically minded commentators recommend other indices to report predictive accuracy.

The statistic most commonly recommended is the area under the receiver operating characteristic (ROC) curve (AUC; Mossman, 1994; Rice & Harris, 1995; Swets et al., 2000). The AUC is a robust, rank order measure (Erceg-Hurn & Mirosevich, 2008), designed for an ordinal predictor and a dichotomous outcome. The ROC curve is the plot of the hits (correctly identified failures) and false alarms (predicted failures who succeeded) for each point in the ordinal prediction scale. AUC values range from zero to one, with values of .50 indicating no information (chance levels of prediction). AUC values can be interpreted as the probability that a randomly selected failure would have a more deviant score than a randomly selected success. An important feature of the AUC is that it is expected to be invariant across changes in outcome's base rate. AUC values would be expected to change, however, given restrictions in the range of the predictor variable (Hanson, 2008; Humphreys & Swets, 1991).

The average AUC for the commonly used risk assessment tools are typically in the .65 to .75 range (Daffern, 2007; Douglas et al., 2006). ROC curves and standardized mean differences (Cohen's *d*) are based on similar statistical models and share many of the same strengths and weaknesses. For the prediction of sexual recidivism, Hanson and Morton-Bourgon (2009) found average *d* values of .67 for the Static-99, .76 for the MnSOST-R, and .62 for the SORAG. The *d* values for individual risk factors tend to be in the .20 to .40 range (Hanson & Morton-Bourgon, 2004, 2005).

Whether specific values of *d* or AUC are "small" or "large" depend on context. For the broad field of behavioral research, Cohen (1988) recommended labeling *d* values of .20, .50, and .80 as "small," "medium," and "large," respectively. Rice and Harris (2005) adopted Cohen's thresholds in the context of violence risk assessment and transformed them into AUC values of .56, .64, and .71. In the context of medical diagnosis, much larger AUC values are desired and expected: Swets (1988) refers to values of .50 to .70 as "rather low accuracy," .70 to .90 as "useful for some purposes," and .90 to 1.0 as "rather high accuracy." Even with AUC values greater than .90, an assessment procedure could have limited utility. For example, all violent patients may have high scores on Dr. Wise's Dangerousness Scale, but the proportion of high scoring patients who are violent may still be below a decision threshold (e.g., none of the patients would be predicted to be violent within the next 72 hours). Similarly, "tiny" effect sizes may have considerable consequences in some contexts (e.g., performance differences amongst elite athletes). Consequently, evaluators need to be mindful that the adequacy of an assessment for a specific purpose cannot be directly inferred from single effect size indicators.

When both the predictor and outcome variables are dichotomous, Fleiss (1994) recommends that the association be presented in terms of odds ratios. Odds ratios have the important feature that they are not directly influenced by the variance of the predictor or the recidivism base rate (Fleiss, 1994; Hanson, 2008). Furthermore, well-established statistical methods are available for describing the relationship between sets of predictor variables and dichotomous outcomes expressed as (log) odds ratios (i.e., logistic regression, see Hosmer & Lemeshow, 2000; Peng, Lee, & Ingersoll, 2002).

One advantage of logistic regression is that it can be used to predict absolute recidivism rates associated with scores on risk scales (and their confidence intervals). These estimates would be expected to be more stable than the raw observed rates when the number of individuals with a particular score is small (<100). Rather than basing the estimate on a limited set of the individuals with a particular score, logistic regression uses information from the whole distribution.

Given that most risk assessments are concerned about when failure is likely to occur, survival analysis is another useful class of statistics (Aalen, Borgan, & Gjessing, 2008; Allison, 1984; Greenhouse, Stangl, & Bromberg, 1989). Life table analysis is appropriate if the primary interest involves differences in "survival" time between a small number of groups. Cox regression (proportional hazard analysis) is the standard method for estimating the contribution of one or more linear predictor variables to different rates of failure. One useful property of Cox regression analyses is that it is possible to estimate relative risk that is independent of base rates (e.g., sexual offenders with a score of 0 on the Static-99 are half [.44 times] as likely to reoffend sexually as the typical Canadian sexual offender).

A number of statistical methods have been developed for extracting information from large, complex data sets (Banks et al., 2004; Dow, Jones, & Mott, 2005; Grann & Långström, 2007). Although prone to shrinkage when applied to new cases, these techniques can useful when the available data is regularly updated, and the associations between predictor variables and the outcome are not expected to be stable or consistent (e.g., identifying smugglers at the border).

Psychologically Meaningful Risk Factors

Even though it is possible to conduct a risk assessment based purely on empirically established predictors, evaluations respond better to the needs of decision-makers (and science) when the evaluation also *explains* the source of the risk. The distinction between simple correlates and clinically useful risk factors has been discussed by Andrews and Bonta (2006; Bonta, 1996) using the terms "static" and "dynamic" risk factors. Static risk factors are fixed aspects of offenders' histories that cannot be changed through deliberate intervention, such as age and the extent of previous offending. Dynamic risk factors are psychological or behavioral features of the offender that are potentially changeable, such as procriminal attitudes or aimless use of leisure time. Because Andrews and Bonta consider that dynamic risk factors should be the focus of correctional programming, these factors are also called "criminogenic needs." Bonta (1996) refers to actuarial risk assessment tools using primarily dynamic factors as "third generation" to distinguish them from the "second generation" risk assessment tools, which were based largely on static factors.

Establishing that a feature is a dynamic risk factor requires both theory and evidence. First, there must be some theoretical justify for why the factor is a potential *cause* of failure. Next, the feature must be shown to precede failure, and between-groups variation in the predictor must be associated with between-groups variation in the rate of failure (see Kraemer et al., 1997). The factor must be capable of changing, and intraindividual changes must be linked to change in the probability of failure, preferably through experimental manipulation of the risk factor.

Readers should note that the research agenda to establish dynamic, causal risk factors is unlikely to ever reach definitive conclusions. There will always be alternate explanations for any set of experimental results. As research advances, however, some interpretations become increasingly plausible and others lose favor in the scientific community (i.e., they become part of either advancing or deteriorating research programs; Lakatos, 1970).

Currently, there is strong evidence supporting certain risk factors as causes of general offending. Substance abuse, procriminal attitudes, and lifestyle impulsivity all predict criminal behavior (Gendreau et al., 1996; Gottfredson & Hirschi, 1990), are capable of changing, and changes on these factors are associated with changes in criminal behavior (Andrews, 1980; Andrews & Wormith, 1984; Raynor & Vanstone, 2001). Importantly, programs that deliberately target substance abuse (Gottfredson, Najaka, Kearley, & Rocha, 2006), criminal attitudes (Andrews, 1980), and impulsivity (Tong & Farrington, 2006) reduce recidivism rates. In general, programs that target criminogenic needs are more likely to be effective in reducing recidivism than programs that target other factors (see meta-analytic review by Andrews & Bonta, 2006, Resource note 10.1, pp. 333–336).

Psychological risk factors exist as propensities and as manifestations. When we described somebody as "aggressive," it does not mean that they are aggressive all the time. It means that they are more likely than other people to be aggressive in certain circumstances (Mischel, Shoda, & Mendoza-Denton, 2002). In the past, I have discussed a similar distinction between stable and acute risk factors (Hanson & Harris, 2000). Stable risk factors are relatively enduring psychological characteristics (e.g., sexual deviancy, impulsivity) whereas acute risk factors are rapidly changing features that signaled the timing of reoffence (e.g., emotional collapse, intoxication).

Although there is some evidence that there are identifiable features that precede offending (Hanson & Harris, 2000; Jones & Gondolf, 2001; Quinsey, Jones, Book, & Barr, 2006), these features may not be particularly "acute." In a large study of sexual offenders on community supervision, we found that problematic behavior averaged over the previous 6 months was a better predictor of recidivism than the behavior during the most recent session (Hanson, Harris, Scott, & Helmus, 2007). Consequently, it may be better to consider current behavior as evidence that the underlying propensities are currently active, rather than as short-term predictors of imminent relapse.

Many of the characteristics associated with crime and violence are found in other populations. Not all individuals who drink are also engaged in crime or violence. Consequently, psychologically minded evaluators need some theory linking the risk factors to problematic behavior. The LSI-R and its variants, for example, were developed based on a cognitive-social learning model of human behaviors (Andrews & Bonta, 1995, 2006; Andrews et al., 2004). Currently, my colleagues and I have been exploring the utility of the Theory of Planned Behavior (Ajzen, 2005) to explain how risk factors can lead to a decision to offend sexually (Mann, Hanson, & Thornton, 2008).

Broad social learning models assume that offending is multiply determined, and that it is possible to intervene with at least some of these factors. This perspective is not universally shared. An alternate vision of the risk assessment task presumes that there is a discrete group of crime-prone social predators ("hawks" vs. "doves"). With this assumption, the evaluation task is to identify whether or not the individual in question is a member of the predator class; there is little interest in dynamic risk factors because the people are not expected to change (G. Harris & Rice, 2003; G. Harris, Rice, & Quinsey, 1994). This vision has a certain popular appeal (Dodge, 2008), but relies on the questionable assumption that a discrete crime-prone taxon exists. The most plausible candidate for such a taxon is psychopathy, which Hare now considers to be a dimension (Guay, Ruscio, Knight, & Hare, 2007).

Strengths and Limits With Existing Risk Assessment Procedures

Most of the current risk assessment instruments show levels of predictive accuracy superior to that of unstructured professional opinion. The accuracy, however, is far from ideal. Most of the research has focused on relative risk (e.g., AUC, r). The stability of the recidivism rates associated with specific assessment procedures has received relatively little research attention (see G. Harris et al., 2003, for an exception). A recent meta-analysis of the Static-99 has found that the observed recidivism rates can vary more than 200% based on base rates and other factors external to the risk scheme (A. Harris, Helmus, Hanson, & Thornton, 2008).

Although there is increasing consensus concerning the risk factors worth considering in crime and violence risk assessments, there is less agreement on the meaning of these risk factors. This creates practical problems when the results of different actuarial risk tools disagree. Barbaree, Langton, and Peacock (2006a), for example, found that less than 5% of their sample were consistently identified as "high risk" or as "low risk" across five actuarial risk tools for sexual offenders (VRAG, SORAG, RRASOR, Static-99, MnSOST-R). One approach to resolving such divergent results is through content analysis (Doren, 2004). The problem with this approach is that it assumes construct validity for items that may have been selected on a purely empirical basis.

Another limitation of the existing risk assessment procedures is that little attention has been paid to strengths. Although it is possible to consider "strengths" as simply the opposite of risk factors, strengths and risk factors can co-occur (e.g., offenders may have criminal friends *and* prosocial friends). Furthermore, there is some evidence that strengths can moderate the influence of risk factors (Griffin, Beech, Print, Bradshaw, & Quayle, 2008).

Implementation Is Rarely Perfect

Although a risk scale may work well in research studies, it may not work in practice (Bonta, Bogue, Crowley, & Motiuk, 2001; Flores, Lowenkamp, Holsinger, & Latessa, 2006; Lowencamp, Latessa, & Holsinger, 2004; Lyle & Graham, 2000). Flores et al. (2006) found that the predictive accuracy of the LSI-R was $r = .21$ for formally trained evaluators ($n = 2{,}030$) compared to $r = .08$ for untrained evaluators ($n = 1{,}635$). Similarly, Bonta et al. (2001) found that the error rate for scoring the LSI-R was between 8% and 15% across sites during routine use in Colorado. These error rates decreased to 1% following feedback.

In a prospective study of risk assessment of sexual offenders on community supervision, we found large differences in the predictive accuracy of the risk tools depending on the "conscientiousness" of the parole officers completing the forms (Hanson et al., 2007). Conscientiousness was defined as completing a minimum number of the requested research forms, including a subjective "override" rating. All officers attended a formal 2-day training. The AUC for the prediction of sexual recidivism using the Static-99/Stable-2007 was .81 for the cases completed by the conscientious parole officers (502 offenders) compared to .67 for the cases rated by the less-than-conscientious officers (290 offenders).

Investment in training, monitoring, and motivating evaluators is needed to avoid degradation in the applied use of risk assessment procedures (Bonta et al., 2001). Currently, there is insufficient research to determine the relative costs of implementing the different risk assessment procedures. It is likely, however, that the most accurate applied assessments are made with simple, easy to use measures, even if more complex measures do better in research studies.

Directions for Future Development of Risk Assessment

The future of risk assessment for crime and violence involves psychometrically sound evaluation of psychologically meaningful causal risk factors described using nonarbitrary metrics. Although certain risk scales have been guided by conceptual models (e.g., LS/CMI, Andrews et al., 2004), these are rare. Most of the commonly used risk scales are simply lists (e.g., VRAG, HCR-20, Static-99). Even if lists can produce total scores with acceptable predictive accuracy, more work is needed to determine what is being measured. By understanding the constructs assessed by risk tools, it would be possible to resolve contradictions and combine the results of different assessment procedures. Furthermore, it would allow evaluators to easily identify how many relevant risk factors are addressed by their assessment procedures.

Distinguishing the latent constructs from their indicators is a perennial theme in psychological assessment. Recently, this distinction has been discussed in terms of linear versus dimensional theory (Loftus, Oberg, & Dillon, 2004).

> The general idea of dimensional theory is that independent variables combine at various stages into internal psychological *dimensions* that underlie performance. By determining (a) how many such dimensions are necessary to account for performance in a given situation along with (b) the nature of the mathematical functions that describe how the independent variables combine to produce the dimensional values, the underlying nature of the relevant structures can be unveiled (p. 836).

An example of risk research consistent with the dimensional approach is the study by Walsh and Kosson (2008) on the relationship of psychopathy and violent recidivism. In this study, they replicated the familiar finding that PCL-R Factors 2 scores (impulsive and antisocial lifestyle) have a greater relationship to future violence than do Factor 1 scores (callous, unemotional, manipulative; Walters, 2003). Walsh and Kosson (2008) found, however, that the interaction between Factor 1 and Factor 2 significantly improved prediction. Such findings suggest that there are at least two factors measured by the PCL-R relevant to risk assessment, and the relationship between them may be complex. Use of a wider range of risk indicators identifies many more dimensions associated with recidivism risk (e.g., Barbaree, Langton, & Peacock, 2006b; Kroner et al., 2005). Sorting out the number of "necessary" dimensions and their mathematical relationships remains an outstanding challenge to the field of risk assessment.

The risk assessment field would also advance by increased use of nonarbitrary metrics (Blanton & Jaccard, 2006a). Nonarbitrary metrics have been used to describe the risk of failure (e.g., percentages, relative risk), but nonarbitrary metrics are rarely used for describing the risk factors that contribute to that risk. This is an important gap. Nonarbitrary metrics are necessary if evaluators are to combine information from different measures into an overall evaluation of risk.

The use of nonarbitrary metrics first requires scientific consensus concerning the nature of the latent psychological construct, and consensus concerning the metrics to describe the variation on the latent constructs (Blanton & Jaccard, 2006b; Embretson, 2006). For dichotomous latent constructs, one possible metric is to report the probability that the condition is present given a score on the indicator measure (Blanton & Jaccard, 2006b). For example, David Thornton, Leslie Helmus, and I have been examining a 3-item "sexual deviancy" scale from the Static-2002 (e.g., any male victims, any noncontact sexual offences, two young victims). Combined across two samples (approximately 800 sexual offenders), we found that the chance of being identified as having "deviant sexual interests" based on a more comprehensive evaluation was as follows: "0" – 7%, "1" – 29%, "2" – 64%, and "3" – 74%. Such percentages depend, of course, on the quality of the "gold standard," but provide much more interpretable description of the offenders than the raw scores of 0 to 3.

Further work is also needed on offender/patient engagement in the risk assessment process. Just as evaluators need to be instructed and motivated to conduct high quality assessments, so do those being assessed. It is possible to conduct risk assessments without cooperation or consent; however, access to many psychologically meaningful factors can only be reliably achieved with a motivated and reliable informant.

Concluding Comment

This article has focused on the psychological assessment of risk for crime and violence, but many of the same issues apply to other areas of psychology (e.g., smoking relapse, recovery from depression). These issues include the utility of structured decision-making, the need for psychologically meaningful causal constructs, and the value of dimensional theory and nonarbitrary metrics. Quality risk assessments are grounded in, and contribute to, psychological knowledge as a whole.

Résumé

Cet article fait un survol des pratiques courantes et des défis liés à l'évaluation du risque pour le crime et la violence. L'évaluation du risque s'est grandement améliorée au cours des 20 dernières années. La capacité prédictive incertaine de l'opinion professionnelle non structurée a été largement remplacée par des méthodes d'évaluation du risque structurées et plus précises. Cependant, aucun consensus n'existe quant aux construits évalués par les différents outils, ni en ce qui a trait à la meilleure méthode pour combiner les facteurs en une évaluation globale du risque. Les progrès liés à l'évaluation du risque pour le crime et la violence reposent sur des évaluations psychométriques rigoureuses des facteurs de risques significatifs, en utilisant des mesures non arbitraires.

Mots-clés : évaluation du risque, violence, récidive

References

Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view.* New York: Springer.

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Milton-Keynes, England: Open University Press.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data.* Beverly Hills, CA: Sage.

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. (1995). *The Level of Service Inventory–Revised.* Toronto, Canada: Multi-Health Systems.

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

Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The Level of Service/Case Management Inventory (LS/CMI).* Toronto, Ontario: 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* (User Report No. 1984–30). Ottawa, Ontario: 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.

Banks, S., Robbins, P. C., Silver, E., Vesselinov, R., Steadman, H. J., Monahan, J., et al. (2004). A multiple-models approach to violence risk assessment among people with mental disorder. *Criminal Justice and Behavior, 31,* 324–340.

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

Barbaree, H. E., Langton, C., & F. E. Peacock, E. J. (2006b). The factor structure of static actuarial items: Its relation to prediction. *Sexual Abuse: A Journal of Research and Treatment, 18,* 207–226.

Blanton, H., & Jaccard, J. (2006a). Arbitrary metrics in psychology. *American Psychologist, 61,* 27–41.

Blanton, H., & Jaccard, J. (2006b). Arbitrary metrics redux. *American Psychologist, 61,* 62–71.

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: The 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., Bogue, B., Crowley, M., & Motiuk, L. (2001). Implementing offender classification systems: Lessons learned. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice* (pp. 227–245). Chichester, NY: Wiley.

Bonta, J., Harman, W. G., Hann, R. G., & Cormier, R. B. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the SIR 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.

Borum, R., Bartel, P., & Forth, A. (2006). *SAVRY: Structured Assessment of Violence Risk in Youth: Professional manual.* Lutz, FL: Psychological Assessment Resources.

Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce, A. J. Harno, E. W. Burgess, & J. Landesco (Eds.). *The workings of the indeterminate sentence law and the parole system in Illinois* (pp. 221–234). Springfield, IL: State Board of Parole.

Campbell, M. A., French, S., & Gendreau, P. (in press). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior.*

Canadian Psychological Association. (2000). *Canadian code of ethics for psychologists* (3rd ed.). Ottawa, Ontario: Canadian Psychological Association.

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

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression, 11,* 65–84.

Daffern, M. (2006). The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior, 12,* 116–130.

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

Dodge, K. A. (2008). Framing public policy and prevention of chronic violence in American youths. *American Psychologist, 63,* 573–590.

Doren, D. M. (2004). 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.

Douglas, K. S., Guy, L. S., & Weir, J. (2006). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography.* Retrieved January 15, 2009, from http://www.sfu.ca/psyc/faculty/hart/resources.htm

Dow, E., Jones, C., & Mott, J. (2005). An empirical modeling approach to recidivism classification. *Criminal Justice and Behavior, 32,* 223–247.

Doyle, M., & Dolan, M. (2007). Standardized risk assessment. *Psychiatry, 6,* 409–414.

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist, 61,* 50–55.

Epperson, D. L., Kaul, J. D., Huot, S. J., Goldman, R., & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool – Revised (MnSOST-R) technical paper: Development, validation and recommended risk level cut scores.* Retrieved January 23, 2009 from http://www.psychology.iastate.edu/faculty/epperson/TechUpdatePaper12–03.pdf

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63,* 591–601.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of quality assurance. *Journal of Criminal Justice, 34,* 523–529.

Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? *Criminal Justice and Behavior, 29,* 397–426.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34,* 575–607.

Gore, K. S. (2007). Adjusted actuarial assessment of sex offenders: The impact of clinical overrides on predictive accuracy. *Dissertations Abstracts International, 68,* B. (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.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford, CA: Stanford University Press.

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh. *Criminal Justice and Behavior, 34,* 22–36.

Grann, M., & Pallvik, A. (2002). An empirical investigation of written risk communication in forensic psychiatric evaluations. *Psychology, Crime & Law, 8,* 113–130.

Greenhouse, J. B., Stangl, D., & Bromberg, J. (1989). An introduction to survival analysis: Statistical methods for analysis of clinical trial data. *Journal of Consulting and Clinical Psychology, 57,* 536–544.

Griffin, H. L., Beech, A., Print, B., Bradshaw, H., & Quayle, J. (2008). The development and initial testing of the AIM2 framework to assess risk and strengths in young people who sexually offend. *Journal of Sexual Aggression, 14,* 211–225.

Guay, J. P., Ruscio, J., Knight, R. A., & Hare, R. D. (2007). A taxometric analysis of the latent structure of psychopathy: Evidence for dimensionality. *Journal of Abnormal Psychology, 116,* 701–716.

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

Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence, 20,* 212–217.

Hanson, R. K. (2007, March). *How should risk assessments for sexual offenders be conducted?* Paper presented at the 4th Annual Forensic Psychiatry Conference, Victoria, British Columbia.

Hanson, R. K. (2008). What statistics should we use to report predictive accuracy? *Crime Scene, 15,* 15–17. Retrieved January 9, 2009, from http://www.cpa.ca/cpasite/userfiles/Documents/ Criminal%20Justice/Crime%20Scene%202008–04.pdf

Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: A Journal of Research and Treatment, 17,* 357–373.

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 sex 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* (User Report No 2007–05). Ottawa, Ontario: Public Safety Canada.

Hanson, R. K., Helmus, L., & Thornton, D. (in press). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior*.

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (User Report No. 2004–02). Ottawa, Ontario: Public Safety 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., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1–21.

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.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences, 1,* 111–119.

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

Harris, A. J. R., Helmus, L., Hanson, R. K., & Thornton, D. (2008, October). *Are new norms needed for Static-99?* Paper presented at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA. Retrieved January 26, 2009, from http://www.static99.org/pdfdocs/arenewnormsneededforstatic-99october2008.pdf

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.). *Sexually coercive behavior: Understanding and management, Annals of the New York Academy of Sciences* (Vol. 989, pp. 193–210). New York: New York Academy of Sciences.

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

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology, 62,* 387–397.

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.

Heilbrun, K., O'Neill, M., Stevens, T. N., Strohman, L. K., Bowman, Q., & Lo, Y. W. (2004). Assessing normative approaches to communicating violence risk: A national survey of psychologists. *Behavioral Sciences and the Law, 22,* 187–196.

Hilton, N. Z., Carter, A. M., Harris, G. T., & Sharpe, A. J. B. (2008). Does using nonnumerical terms to describe risk aid violence risk communication: Clinician agreement and decision making. *Journal of Interpersonal Violence, 23,* 171–188.

Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16,* 267–275.

Hilton, N. Z., & Simmons, J. L. (2001). Actuarial and clinical risk assessment in decisions to release mentally disordered offenders from maximum security. *Law and Human Behavior, 25,* 393–408.

Hoffman, P. B. (1994). Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's Salient Factor Score. *Journal of Criminal Justice, 22,* 477–494.

Hoffman, P. B., & Beck, J. L. (1974). Parole decision-making: A Salient Factor Score. *Journal of Criminal Justice, 2,* 195–206.

Hoge, R. D., & Andrews, D. A. (2002). *Youth Level of Service/Case Management Inventory (YLS/CMI).* Toronto, Ontario: Multi-Health Systems.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Humphreys, L. G., & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *Journal of Applied Psychology, 76,* 316–321.

Hunt, M. (1997). *How science takes stock: The story of meta-analysis.* New York: Russell Sage Foundation.

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.

Jones, A. S., & Gondolf, E. W. (2001). Time-varying risk factors for reassault among batterer program participants. *Journal of Family Violence, 16,* 345–359.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupler, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54,* 337–343.

Krauss, D. A. (2004). Adjusting risk of recidivism: Do judicial departures worsen or improve recidivism prediction under the Federal Sentencing Guidelines? *Behavioral Sciences and the Law, 22,* 731–750.

Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of criminal risk: The structure of "criminality". *International Journal of Law and Psychiatry, 28,* 360–374.

Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1999). *Spousal Assault Risk Assessment: User's guide.* Toronto, Ontario: Multi-Health Systems.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, UK: Cambridge University Press.

Loftus, G. R., Oberg, M. A., & Dillon, A. M. (2004). Linear theory, dimensional theory, and the face-inversion effect. *Psychological Review, 111,* 835–863.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2004). *Empirical evidence on the importance of training and experience in using the Level of Service Inventory-Revised. Topics in Communications* (pp. 49–53). Washington, DC: National Institute of Corrections.

Loza, W. (2003). Predicting violent and nonviolent recidivism of incarcerated male offenders. *Aggression and Violent Behavior, 8,* 175–203.

Lyle, G., & Graham, E. (2000). Looks can be deceiving: Using a risk assessment instrument to evaluate the outcomes of child protective services. *Children and Youth Services Review, 22,* 143–155.

Mann, R. E., Hanson, R. K., & Thornton, D. (2008, October). *What should be assessed in sexual offender risk assessments?* Paper presented at the 27th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Mann, R. E., Hanson, R. K., & Thornton, D. (2009). *What are the plausible treatment targets for sexual offenders?* Unpublished manuscript.

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

Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science* (Vol. IV, pp. 373–402). Minneapolis, MN: University of Minnesota.

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.

Mills, J. F., Kroner, D. G., & Hemmati, T. (2004). The Measures of Criminal Attitudes and Associates (MCAA): The prediction of general and violent recidivism. *Criminal Justice and Behavior, 31,* 717–733.

Mischel, W., Shoda, Y., & Mendoza-Denton, R. (2002). Situation-behavior profiles as a locus of consistency in personality. *Current Directions in Psychological Science, 11,* 50–54.

Monahan, J. (1976). The prevention of violence. In J. Monahan (Ed.), *Community mental health and the criminal justice system* (pp. 13–34). New York: Pergamon Press.

Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques.* Beverly Hills, CA: Sage.

Monahan, J. (2007a). 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 Company.

Monahan, J. (2007b). Statistical literacy: A prerequisite for evidence-based medicine. *Psychological Science in the Public Interest, 8,* i–ii.

Monahan, J. (2008). Structured risk assessment of violence. In R. Simon & K. Tardiff (Eds.), *Textbook of violence assessment and management* (pp. 17–33). Washington, DC: American Psychiatric Publishing.

Monahan, J., Heilbrun, K., & Silver, E. (2002). Communicating violence risk: Frequency formats, vivid outcomes, and forensic settings. *International Journal of Forensic Mental Health, 1,* 121–126.

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

Nuffield, J. (1982). *Parole decision-making in Canada.* Ottawa, Ontario: Solicitor General Canada.

Otto, R. K. & Douglas, K. (Eds) (2009). *Handbook of violence risk assessment tools.* Milton Park, UK: Routledge.

Peng, C. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research, 96,* 3–14.

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.

Quinsey, V. L., & Ambtman, R. (1979). Variables affecting psychiatrists' and teachers' assessments of the dangerousness of mentally ill offenders. *Journal of Consulting and Clinical Psychology, 47,* 353–362.

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., Jones, G. B., Book, A. S., & Barr, K. N. (2006). The dynamic prediction of antisocial behaviour among forensic psychiatric patients: A prospective field study. *Journal of Interpersonal Violence, 21,* 1539–1565.

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, UK: Wiley.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737–748.

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.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior, 27,* 733–764.

Smith v. Jones (1999), 1 SCC, 455. Retrieved January 28, 2009, from http://scc.lexum.umontreal.ca/en/1999/1999rcs1–455/1999rcs1–455.html

Steadman, H. J., & Cocozza, J. J. (1974). *Careers of the criminally insane: Excessive social control of deviance.* Lexington, MA: Lexington Books.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240,* 1285–1293.

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

Tarasoff v. (1974). Regents of the University of California, 118 Cal., 529 P. 2d 553, R. 119.

Tarasoff v. (1976). Regents of the University of California, 131 Cal., 551 P. 2d 334, R. 14.

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

Tong, L. S. J., & Farrington, D. P. (2006). How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries. *Psychology, Crime & Law, 12,* 3–24.

Truscott, D. (1993). The psychotherapist's duty to protect: An annotated bibliography. *Journal of Psychiatry & Law, 21,* 221–244.

Truscott, D., & Cook, K. H. (2004). *Ethics for the practice of psychology in Canada.* Edmonton, AB: University of Alberta Press.

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, Ontario.

Walsh, Z., & Kosson, D. S. (2008). Psychopathy and violence: The importance of factor level interactions. *Psychological Assessment, 20,* 114–120.

Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factors scores: A meta-analysis. *Law and Human Behavior, 27,* 541–558.

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

Whitaker, D. J., Le, B., Hanson, R. K., Baker, C. K., McMahon, P., Ryan, G., et al. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse & Neglect, 32,* 529–548.

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

Received January 28, 2006
Revision received February 20, 2006
Accepted February 23, 2008 ■