

THE PREDICTION OF VIOLENCE IN ADULT OFFENDERS



A Meta-Analytic Comparison of Instruments and Methods of Assessment

MARY ANN CAMPBELL
SHEILA FRENCH
PAUL GENDREAU

University of New Brunswick–Saint John, Canada

Using 88 studies from 1980 to 2006, a meta-analysis compares risk instruments and other psychological measures on their ability to predict general (primarily nonsexual) violence in adults. Little variation was found amongst the mean effect sizes of common actuarial or structured risk instruments (i.e., Historical, Clinical, and Risk Management Violence Risk Assessment Scheme; Level of Supervision Inventory–Revised; Violence Risk Assessment Guide; Statistical Information on Recidivism scale; and Psychopathy Checklist–Revised). Third-generation instruments, dynamic risk factors, and file review plus interview methods had the advantage in predicting violent recidivism. Second-generation instruments, static risk factors, and use of file review were the strongest predictors of institutional violence. Measures derived from criminological-related theories or research produced larger effect sizes than did those of less content relevance. Additional research on existing risk instruments is required to provide more precise point estimates, especially regarding the outcome of institutional violence.

Keywords: risk assessment; violence; meta-analysis; adult offenders; forensic patients; recidivism; misconduct

Assessments of violence risk should play a central role in decision making pertaining to sentencing, release, case management, and the selection of rehabilitation methods to achieve risk reduction (Andrews & Bonta, 2006; Heilbrun, 1997). The ability to assess risk is facilitated by the use of structured, empirically derived, and theoretically driven instruments (Andrews, Bonta, & Wormith, 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000). Despite the availability of violence-specific risk tools and other measures associated with criminality and aggression, there are relatively few meta-analytic comparisons of their predictive validity for predicting risk, identifying risk-reduction targets, and monitoring changes in risk level. Comparisons of this nature are necessary to adequately inform professional practice parameters concerning the selection of instruments for inclusion in

AUTHORS' NOTE: *A version of this article was prepared under contract to Public Safety Canada. Views expressed herein are those of the authors and do not necessarily reflect those of Public Safety Canada. We recognize the assistance of Laurie Green, who conducted the interrater coding for this project. We are grateful to Dr. Karl Hanson for suggestions with respect to base rate consideration, to Dr. Paula Smith for assistance with database construction, and to Dr. Jim Bonta for his comments regarding an earlier version of this article. Finally, we appreciate the assistance of researchers who offered references and/or unpublished data for analysis. Correspondence should be directed to Mary Ann Campbell, PhD, Centre for Criminal Justice Studies, University of New Brunswick, PO Box 5050 (100 Tucker Park Road), Saint John, New Brunswick, Canada, E2L 4L5; phone: (506) 648-5969; fax: (506) 648-5780; e-mail: mcampbel@unb.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 36 No. 6, June 2009 567-590
DOI: 10.1177/0093854809333610
© 2009 International Association for Correctional and Forensic Psychology

violence risk assessments. Thus, this study provided a meta-analytic appraisal of a wide range of instruments and methods used in the literature to inform estimates of violence risk in adult offenders and forensic patients.

Tools for assessing risk have undergone various modifications in the past 50 years. First-generation risk assessment, arising in the mid-20th century, was based on unstructured clinical judgments of risk that were prone to error and bias (Grove et al., 2000; Monahan & Steadman, 1994; Rice, 1997). In light of these limitations, second-generation risk instruments offered a standardized assessment that was based on constructs statistically predictive of recidivism (e.g., criminal history, Diagnostic and Statistical Manual of Mental Disorders diagnoses). Examples of second-generation tools are the Violence Risk Assessment Guide (VRAG; Harris, Rice, & Quinsey, 1993) and the Statistical Information on Recidivism (SIR; Bonta, Harman, Hann, & Cormier, 1996). Some of these measures were criticized because their items were selected with little regard for their theoretical or rehabilitative value (Bonta, 2002). In addition, despite the fact that some second-generation instruments demonstrate fairly good predictive validity (e.g., $r = .30$ to $.35$; Bonta & Yessine, 2005; Gendreau, Little, & Goggin, 1996; Glover, Nicholson, Hemmati, Berfeld, & Quinsey, 2002; Loza & Green, 2003; Polvi, 2001), they are mainly composed of static risk items (Andrews & Bonta, 2006). Static risk factors are unchangeable (e.g., criminal history, age, gender). Sole reliance on static factors for risk assessment has been criticized because these factors do not capture the complexity of recidivism, do not permit measurement of changes in risk over time, and fail to identify areas for intervention (Andrews, Bonta, & Hoge, 1990; Hoge & Andrews, 1996; Wong & Gordon, 2006).

In contrast to second-generation measures, third-generation risk instruments emphasized the need for prediction models to not only predict risk but to also inform the identification of criminogenic needs that could be targeted for change as a means of reducing risk (Andrews et al., 2006; Bonta, 2002). Common examples are the Level of Supervision Inventory–Revised (LSI-R; Andrews & Bonta, 1995); Historical, Clinical, and Risk Management Violence Risk Assessment Scheme (HCR-20; Webster, Douglas, Eaves, & Hart, 1997); and Self-Appraisal Questionnaire (SAQ; Loza, 2005). These instruments included empirically supported risk factors, but item selection was more deliberately driven by theoretical understandings of persistent criminality and violence (i.e., social learning and cognition theories, the principles of risk-need-responsivity) than were second-generation measures (Andrews & Bonta, 2006; Gendreau, Goggin, French, & Smith, 2006).

Third-generation measures also included dynamic risk factors, which are risk factors that are variable in nature and can change with time or with the influence of social, psychological, biological, or contextual factors (e.g., intervention; Douglas & Skeem, 2005). Examples of such malleable risk factors (i.e., criminogenic needs) are substance use, interpersonal conflict, and antisocial attitudes. A meta-analysis by Gendreau et al. (1996) demonstrated that dynamic risk factors were as useful as static factors were in predicting risk. This finding encouraged the view held by some researchers that dynamic risk factors may even be more relevant than static factors when the focus is on risk reduction (Andrews, 1989; Douglas & Skeem, 2005; Heilbrun, 1997). The advantage of using instruments that assess dynamic risk factors is that they are sensitive to changes in risk that might occur with time and/or as a result of rehabilitation (Andrews & Bonta, 2006; Heilbrun, 1997). However, it should be noted that some dynamic risk factors are best described as “potential” dynamic risk factors (e.g., accommodation problems) until

additional research confirms their individual links to fluctuations in recidivism risk level (see Brown, St. Amand, & Zamble, 2009).

The latest evolution in risk instruments (i.e., fourth generation) are those specifically designed to be integrated into (a) the process of risk management, (b) the selection of intervention modes and targets for treatment, and (c) the assessment of rehabilitation progress (Andrews & Bonta, 2006; Andrews et al., 2006). These instruments are administered on multiple occasions and are particularly informative because they document changes in specific criminogenic needs that might occur between an offender's entrance into the criminal justice system through his or her exit from the criminal justice system. Fourth-generation instruments are intended to identify areas of success within a case management plan as well as areas in which intervention strategies need to be modified to maximize their potential for risk reduction. Examples are the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004), Violence Risk Scale (VRS; Wong & Gordon, 2006), Correctional Offender Management Profile for Alternative Sanctions (COMPAS; Brennan & Oliver, 2000), and Correctional Assessment and Intervention System (CAIS; National Council on Crime and Delinquency, 2004). Notably, few of these measures have sufficiently available prospective validity data.

With significant growth in risk research, many instruments have been advocated for use in the assessment of violence risk. These measures range from tools specifically designed to predict violence (e.g., Violence Prediction Scheme [VPS; Webster, Harris, Rice, Cormier, & Quinsey, 1994] and HCR-20 [Webster et al., 1997]), to measures that predict specific types of violence (e.g., Spousal Assault Risk Assessment Guide [SARA; Kropp, Hart, Webster, & Eaves, 1995]), to measures that assess personality constructs related to violence (e.g., Psychopathy Checklist-Revised [PCL-R; Hare, 2003]), and to measures designed to assess general recidivism (e.g., LSI-R [Andrews & Bonta, 1995]). Thus, professionals have access to a variety of tools to inform their predictions of violence risk.

With this variety in tools, the issue confronting a professional is "Which of these instruments should I use?" Complicating the answer to this question is the fact that there are differing administration methods of assessment within this array of instruments (e.g., paper-and-pencil vs. professional-rated forms; file review vs. interview) and they can vary in content (e.g., measure a single risk-related construct vs. multiple constructs). The diversity in both the administration format and content may lead some assessors to use multiple measures to generate a consensus estimation of risk (Doren, 2002). This practice can be problematic. Mills and Kroner (2006) used the PCL-R, LSI-R, VRAG, and the General Statistical Information on Recidivism (GSIR) to predict postrelease violence and general recidivism. For most offenders, there was agreement in the standardized risk scores generated for each of these instruments, but predictive accuracy was substantially reduced for cases with a high level of disagreement between instruments on their standardized risk scores. The challenges associated with formulating risk judgments based on several risk instruments highlight the need for research that identifies the most appropriate risk instrument(s) for a given offender population, forensic setting, and assessment purpose (see also Seto, 2005).

PREVIOUS META-ANALYTIC COMPARISONS OF RISK INSTRUMENTS

A number of individual studies have compared the relative utility of risk instruments for the prediction of violence in adults (e.g., see Dahle, 2006; Douglas, Yeomans, & Boer,

2005; Glover et al., 2002; Grann, Belfrage, & Tengström, 2000; Kroner & Loza, 2001; Mills & Kroner, 2006; Rice & Harris, 1995). Given the variation across these prediction and comparison studies in terms of such factors as sample characteristics, setting, and definitions of outcome, it is not surprising that it has been virtually impossible to identify a dominant violence risk measure. In fact, much of the variation across prediction studies may be due to sampling error, which, according to Hunter and Schmidt (2004), is the major source of variation in prediction studies. One means of addressing this issue is to conduct meta-analyses, which statistically culminate primary study data to better estimate true population parameters.

Four meta-analyses of the risk-prediction literature have provided a comparison of various instruments used in the assessment of risk in adults. In the first analysis, Gendreau et al. (1996) compared the LSI-R, PCL-R, Salient Factor Score (Hoffman, 1983), Wisconsin Classification System (Baird, 1981; Baird, Heinz, & Bemus, 1979), and Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1967) in the prediction of general recidivism across 131 primary studies. Although each instrument was moderately predictive, the LSI-R produced the strongest effect size. Subsequently, Gendreau, Goggin, and Law (1997) compared the LSI-R, MMPI, "other" risk measures, and non-MMPI measures of antisocial personality as predictors of an aggregate criterion of violent and nonviolent institutional misconducts. The LSI-R produced the highest predictive validity and outperformed the other measures. A third meta-analytic comparison by Gendreau, Goggin, and Smith (2002) focused specifically on the prediction of violent recidivism and found that the LSI-R had a slight advantage over the PCL-R. Other risk measures were not assessed by Gendreau et al., thus, it is unclear whether other instruments would perform on par with the LSI-R and PCL-R for violent risk prediction. Moreover, all of the above meta-analyses were concerned with a limited number of risk-related instruments and are in need of updating given the availability of additional primary studies since their publication.

More recently, Walters (2006) meta-analytically compared an aggregate category of selected structured/actuarial risk instruments (i.e., HCR-20, LSI-R, PCL-R, VRAG, and the Lifestyle Criminality Screening Form created by Walters, White, & Denney, 1991) with a number of self-report measures often used in risk judgments for institutional misconduct, general recidivism, and violence. Some of these self-report measures were specific to risk prediction (e.g., Psychological Inventory of Criminal Thinking Styles [PICTS; Walters, 1995, 1996] and Self-Appraisal Questionnaire [SAQ; Loza, 2005]), whereas others reflected general clinical constructs relevant to an individual's general personality and emotional functioning (e.g., NEO Personality Inventory-Revised, Multidimensional Anger Inventory, Beck Hopelessness Scale). Walters's findings supported the predictive validity of self-report measures in risk assessment but only if these instruments were based on constructs that were empirically tied to risk (e.g., antisocial attitudes). Walters suggested that the integration of content-relevant self-report measures with actuarial/structured risk instruments could add to the validity of risk assessment. Unfortunately, only a select number of structured/actuarial risk instruments were coded in Walters's meta-analysis, and their individual predictive validities were not reported. In addition, only nine effect sizes were available to compare the aggregate category of structured/actuarial methods with self-report measures in terms of their ability to predict violent recidivism. Across these nine effect sizes, the mean effect was larger for the structured/actuarial measures than for the

general category of self-report measures. A larger database that encompasses a greater range of measures is required to replicate Walters's findings.

In summary, many advances have been made in the assessment of risk in adults. Nonetheless, uncertainty remains concerning the most appropriate instruments for the prediction of violence given variations in item content, purpose and format, and administration method. Only a few meta-analyses (i.e., Gendreau et al., 1997; Gendreau et al., 1996; Gendreau et al., 2002; Walters, 2006) have been conducted to synthesize this literature for professionals, and none of these has been sufficiently comprehensive in its estimation of violence risk. A synthesis of this nature is timely given that very few correctional psychologists report using instruments empirically supported as relevant to the task of risk estimation (see Boothby & Clements, 2000). Thus, the primary objective of the current meta-analysis was to determine which instruments function most effectively as valid predictors of future violence (primarily nonsexual) within prison settings and in the community. Four secondary objectives of this meta-analysis were to compare the predictive utility of risk measures depending on which generation they represented, the type of items (static vs. dynamic), their method of administration, and their content relevance to corrections. With this information, guidelines can be generated to assist with the selection of risk instruments.

METHOD AND PROCEDURE

DESCRIPTION OF DATABASE

An electronic literature search was conducted via EBSCO databases (Academic Search Elite, PsycARTICLES, and PsycINFO). Key search terms included (a) assessment-related terms (e.g., *actuarial, clinical, prediction, LSI-R, PCL-R*), (b) terms related to the offender population (e.g., *adult offender, prisoner, parolee*), and (c) terms related to violent outcomes (e.g., *recidivism, misconduct*). Unpublished data were requested via an e-mail sent to 33 researchers and 23 research centers known to conduct risk research. Additional studies were added via reviews of article reference sections. The search was restricted to studies conducted between 1980 and 2006. Inclusion criteria required that primary prediction studies (a) were truly prospective in nature (i.e., assessment preceded the measurement of outcome), (b) involved adult general offender or forensic patient (i.e., sample mean of 18+ years at time of assessment), and (c) reported sufficient data to calculate an effect size (e.g., Pearson r ; Phi coefficient Φ) between the prediction measure and violent misconduct or recidivism outcomes; prison or probation studies were included regardless of length of follow-up but we included only postrelease recidivism studies that had least a 6-month follow-up period. For each study, data from the largest sample, longest follow-up period, and most specific type of criterion (i.e., conviction vs. arrest) were recorded. To avoid redundancy with Hanson and Morton-Bourgon's (2007) recent meta-analysis on the predictive validity of risk instruments for sexual and violent recidivism in sex offenders, this analysis excluded studies using samples that were exclusively of sex offenders. Likewise, instruments designed specifically to assess sexual recidivism were not included. Thus, studies included in the meta-analysis pertained almost exclusively to nonsexual offenders and forensic patients with nonsexual violent outcomes. It is estimated that sex offenders contributed only 2% of the total sample size for the predictors in the current meta-analysis.

The data set contained 88 coded studies reporting on various risk measures predicting institutional violence ($k = 76$) and violent recidivism ($k = 185$) in adults. Most of the data set was based on studies published in books, journals, or government reports (63.1%) that were primarily conducted in North America (60% Canadian and 24.8% American). Authors were largely academically affiliated (51.3%) and from the discipline of psychology (85.4%). The sample sizes for the predictors were 232,790 for institutional violence and 40,944 for violent recidivism. The majority of the data set (81.3% collapsed across outcomes) represented male-dominated samples. Samples representing general offender populations produced 63.9% of effect sizes, and the remainder were based on forensic psychiatric (30.7%) and mixed (5%) samples. Institutional violence effect sizes were based on an equivalent percentage of general offender (50.7%) and forensic samples (49.3%), and most violent recidivism effect sizes were from general offender samples (70.0%).

Overall, almost half of all effect sizes were drawn from samples coded as being of a low or moderate risk level (43.6% for institutional violence and 44.0% for violent recidivism), whereas only 7.5% came from high risk samples. Fewer than 3% were from mixed risk samples and 2.1% could not be coded on risk level because of insufficient information. Predisposition for violence among offenders could not be assessed with any degree of certainty across studies because information about previous/index violent offenses was not reported for more than half of the obtained effect sizes (67.6% of institutional violence outcomes and 56.0% of violent recidivism outcomes). The mean base rate for major institutional violence (excluding verbal threats) was 25.84% ($SD = 13.61$) and was 21.73% ($SD = 12.99$) for violent recidivism. Only 39.4% of institutional violence effect sizes were based on follow-up periods of greater than 1 year and most community-released offenders were followed from 2 to 5 years (41.7%). The most common index of institutional misconduct was official prison records (74.7%); rearrest, reconviction, and reincarceration were the most common violent recidivism indices (72.2% of effect sizes). In nearly all studies (97.0%), violent recidivists were compared to an aggregate group of offenders (i.e., offenders who did not reoffend at all combined with those who may have nonviolently recidivated).

Although studies examined more than 70 different risk measures in total, only instruments with ≥ 10 effect size estimates per outcome of interest will be reported to emphasize individual instruments for which the greatest amount of data were available. These instruments included the HCR-20 ($k = 11$ for misconduct; $k = 11$ for recidivism), LSI/LSI-R ($k = 19$ for recidivism), PCL/PCL-R ($k = 24$ for recidivism), SIR scale ($k = 17$ for recidivism), and VRAG ($k = 14$ for recidivism). Some instruments with ≤ 10 effect sizes are reported, but their predictive validities should be interpreted cautiously. Most effect sizes were based on risk assessment methods that involved only the use of file extraction methods (52.2% of effect sizes), followed by self-report questionnaires (17.4%), a combination of interview and file review (16.5%), only an interview (11.2%), or staff behavioral observations (1.8%). Most effects were based on measures containing potentially dynamic (51.9%) or static (34.9%) risk items, whereas 8% were derived from measures using relatively equal numbers of static and dynamic items. For almost 5% of effect sizes, their static or dynamic item composition could not be determined. The vast majority of effect sizes (85%) were based on measures rooted in a theory of criminal behavior and/or created specifically for use as a criminal risk instrument. Fewer than 3% were coded as first-generation methods of risk assessment, and the sample was relatively split between second- (52.3%)

and third-generation measures (42.3%). Only 2.5% of the data were from fourth-generation instruments.

CODING OF STUDIES

The descriptor, predictor, and outcome data were gathered from studies using a coding guide created for this current analysis. Major coded categories included (a) study and author characteristics (e.g., type of publication, author affiliation, publication year), (b) sample variables (e.g., ethnicity, gender, offender type), (c) risk assessment descriptors (e.g., name of measure, administration method, item content), and (d) effect size descriptors (e.g., type of outcome, calculated effect size). In accordance with previous definitions used in the literature (see Andrews et al., 2006; Bonta, 2002), each identified instrument was coded as to its most relevant generation of risk category. A list of instruments coded under each generation can be obtained from the first author, as can a copy of the entire coding manual. All studies were coded by S. French. Interrater reliability was established using a randomly selected sample of 15 studies, blindly coded by a second experienced coder. Using the Yeaton and Wortman (1993) formula, $\frac{\sum(\text{agreements})}{\sum(\text{agreements} + \text{disagreements})}$, the index for agreement was .82. The source of disagreements concerned less obvious sample characteristics (i.e., determination of sample risk level) and aspects of the nature of a particular risk instrument (i.e., type of item content, generation of risk instrument). Disagreements most often resulted from a misunderstanding or a clerical error when entering item codes. The two raters discussed disagreements and a consensus coding was achieved for those items prior to analysis.

EFFECT SIZE CALCULATION

For the rationale behind this study's approach to meta-analysis, the reader is referred to Gendreau and Smith (2007). Correlation coefficients were recorded for each measure's predictive validity with institutional violence and recidivism outcomes. Where statistics other than r were reported (i.e., F , t , χ^2 , p , AUC), we employed the appropriate formula for conversion to r (see Rosenthal, 1991; Swets, 1986). In light of generally low base rates for violent institutional misconduct and recidivism, it was necessary to consider this potential influence on effect sizes.¹ Correlation coefficients were adjusted using Ley's (1972) formula: $r' = [(r_{xy})(\delta_x' / \delta_x)] / [1 - r_{xy}^2 + (r_{xy}^2)(\delta_x'^2 / \delta_x^2)]^{1/2}$, where r_{xy} was the observed correlation, δ_x was the observed standard deviation of the base rate, δ_x' was the average standard deviation based on the average base rate for studies in the analysis, and r_{xy}' was the corrected correlation. The standard deviation of the base rate was calculated using the formula $\delta = [pq / (N)(N - 1)]^{1/2}$, where p was the number of participants who were institutional or community recidivists, q was the number of participants who were institutional or community nonrecidivists, and N was the total sample size.

The primary metric used to estimate and interpret the magnitude of the relationships between each risk predictor category and institutional violence and recidivism outcomes was the mean r value (M_r) weighted by sample size (Z^+ ; see Hedges & Olkin, 1985), along with its associated 95% confidence interval (CI_{Z^+}). Although M_r is reported in Tables 1 through 5, interpretation of relationships was based on the mean Z^+ values and their associated CI s. The CI s were used to interpret whether mean effect sizes from different variables (e.g., different risk measures) were likely drawn from the same population parameter. If

there was no overlap at all between the *CI*s for any two mean effect sizes, or the *CI*s just touched, then these two effects would be interpreted as representing different population parameters. This criterion is equivalent to statistical significance of $p < .006$ as long as the sample size was ≥ 10 and the width of the *CI*s did not vary by more than a factor of two (Cumming & Finch, 2005). When two *CI*s overlapped by no more than one quarter of the average length of the two intervals, then these mean effects were also interpreted as representing two different population parameters and were statistically different at approximately $p \leq .05$ (Cumming & Finch, 2005). Overlap in *CI*s exceeding the above criteria meant that the mean effect sizes likely represented the same population parameter and, therefore, were not statistically different from each other. A second use of *CI*s was to reflect the precision of effect size estimates, which was judged by noting the width of the *CI* (i.e., narrower intervals indicate a more precise estimate of a population parameter than do wider intervals; Cumming & Finch, 2001; Gendreau, Goggin, & Smith, 2000; Schmidt, 1996).

EFFECT SIZE HETEROGENEITY

Effect size variability was assessed using the Q statistic (Rosenthal, 1991). For each effect size, a q value was calculated using the following formula: $(n - 3)(z_r - Z^+)$, where n was the total sample size per effect size; z_r was the standardized r' value per effect size; and Z^+ was the sample-weighted M_r' value for each predictor category. These q values were then summed for each predictor category, yielding Q , which is an estimate of the heterogeneity of the effect sizes within that category. To test its significance, the Q was evaluated using the critical value of χ^2 with $(k - 1)$ degrees of freedom. A significant Q statistic indicates that there is more variability than would be expected by chance. In such cases, outlying effect sizes were inspected and only eliminated if there was a logical reason for exclusion (e.g., a coding error or a unique study characteristic, such as a restrictive sample).

FAIL-SAFE ESTIMATION

A fail-safe estimate was employed to provide an index of how many additional effect sizes would be required to alter an obtained effect size estimate. An index of the number of effect sizes ($Z^+ = .00$) needed for a given risk measure of greater accuracy in the prediction of misconduct or recidivism to approach an effect size equal to one of lesser accuracy was calculated using the following formula: $[(k_B(Z_B^+ - Z_A^+))]/(Z_A^+ - Z_{B=0}^+)$, where $Z_{B=0}^+$ indicates a null effect for the more accurate risk measure (see Gendreau et al., 2002). As applied to this meta-analysis, assume that the mean effect size was .30 ($k = 50$) for Measure A and .35 ($k = 40$) for Measure B. Using the above formula, an estimate of seven B predictions with a $Z^+ = 0$ would be necessary to negate Measure B's supremacy over A. That is, seven additional Measure B effect sizes, each with a magnitude of $Z^+ = .00$, would have to be located to conclude that the two measures were at predictive parity.

RESULTS

RISK MEASURES: PREDICTIVE VALIDITIES FOR INSTITUTIONAL VIOLENCE

Throughout the results section we focus on the Z^+ values, which produced similar results to the r values with the exception of four cases. These exceptions pertained to institutional

TABLE 1: Effect Size Comparisons of Risk Measures for the Prediction of Institutional Violence^a and Recidivism^b

Measure	k	N	M _r (SD)	CI _r	Z ⁺	CI _{Z⁺}	Q
Institutional violence							
HCR-20	11	758	.31 (.14)	.21 to .40	.28	.10 to .24	12.26
LSI/LSI-R	6	650	.24 (.08)	.16 to .33	.24	.09 to .25	5.91
PCL/PCL-R	5	626	.15 (.12)	.01 to .30	.14	.00 to .16	2.90
PCL:SV	7	504	.25 (.10)	.16 to .34	.22	.07 to .25	5.59
SIR scale ^c	1	215	.08	-.05 to .21	—	—	—
VRAG	2	222	.17 (.13)	-.98 to 1.00	.15	-.08 to .18	1.54
Violent recidivism							
HCR-20	11	1395	.25 (.15)	.14 to .35	.22	.17 to .27	6.68
LSI/LSI-R	19	4361	.25 (.08)	.21 to .28	.28	.25 to .31	57.15*
PCL/PCL-R	24	4757	.24 (.10)	.19 to .28	.27	.24 to .30	48.04*
SIR Scale	17	5618	.24 (.13)	.18 to .31	.22	.19 to .25	32.54*
VRAG	14	2082	.27 (.13)	.20 to .35	.32	.28 to .36	47.06*

Note. *k* = effect sizes per risk measure; *N* = offenders per risk measure; *CI* = confidence interval; *Z*⁺ = *r*' value weighted by sample size; *CI*_{*Z*⁺} = 95% confidence interval about *Z*⁺. HCR-20 = Historical, Clinical, and Risk Management Violence Risk Assessment Scheme; LSI = Level of Supervision Inventory; PCL = Psychopathy Checklist; PCL:SV = Psychopathy Checklist: Screening Version; SIR = Statistical Information on Recidivism; VRAG = Violence Risk Assessment Guide.

a. Although the total number of effect size estimates for risk measures with institutional violence was 76, there was only one category with *k* > 10. The other measures reported above are included to facilitate tentative comparisons of the predictive validity for those measures with misconduct and recidivism outcomes.

b. Although the total number of effect size estimates for risk measures with recidivism was 185, only those measures with more than 10 predictive validities were included in Table 1.

c. Only one effect size was available for the SIR scale (*r* = .08). Therefore, *Z*⁺ was not calculated for this instrument. **p* < .05, indicates that the level of variability is greater than would be expected by chance.

violence and related to second-generation tools (see Table 2), static-based instruments (see Table 3), file-extraction methods (see Table 4), and content-relevant instruments (see Table 5). In each of these cases, base rate and sample size adjustments to *r* resulted in a higher *Z*⁺ value and more precise *CI*s around that value. Table 1 contains the *Z*⁺ values and associated 95% *CI*s for risk measures and institutional violence. Only one measure was represented by more than 10 effect sizes (i.e., the HCR-20); however, preliminary data for some instruments are reported despite a *k* ≤ 10 to create consistency with instruments reported for violent recidivism. The HCR-20 and LSI-R had the largest mean weighted effect sizes for predicting institutional violence (*Z*⁺ = .28 and .24, respectively) and their *CI*s were wide and overlapping. The PCL: Screening Version (PCL:SV; *k* = 7) produced the third largest mean effect size (*Z*⁺ = .22), whereas the PCL-R and VRAG produced the weakest associations with institutional violence (*Z*⁺ = .14 and .15, respectively). However, 95% *CI*s for each of the above risk measures overlapped considerably, suggesting that they were all sampling from the same population parameter. Furthermore, the width of the *CI*s (all greater than .10) and the small number of effect sizes foreshadow a lack of precision for each instruments' effect size estimate. As a result, interpretations based on these estimates should be viewed as tentative until more studies have been conducted with institutional violence as the criterion. Given that a minimum of 10 effect sizes per instrument was set for calculation of fail-safe analyses, these metrics were not calculated for institutional violence.

RISK MEASURES: PREDICTIVE VALIDITIES FOR VIOLENT RECIDIVISM

The *Z*⁺ values and associated 95% *CI*s for predicting violent recidivism are displayed in the latter part of Table 1. The largest *Z*⁺ value was recorded for the VRAG. There was

TABLE 2: Comparison of Risk Assessment Generations for the Prediction of Institutional Violence^a and Recidivism^b

Measure	k	N	M_r (SD)	CI_r	Z^+	CI_{Z^+}	Q
Institutional violence							
Second generation	48	229397	.23 (.15)	.19 to .27	.34	.33 to .35	410.12*
Third generation	27	3349	.21 (.12)	.17 to .25	.20	.17 to .23	26.94*
Violent recidivism							
Second generation	92	19874	.20 (.14)	.17 to .23	.18	.17 to .19	328.13*
Third generation	81	15233	.22 (.12)	.19 to .25	.23	.21 to .25	247.38*

Note. k = effect sizes per risk measure; N = offenders per risk measure; CI = confidence interval; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .

a. Only 75 of 76 institutional violence effect sizes are represented. One effect size, produced by a fourth-generation measure, was not included in the table.

b. Only 173 of 185 recidivism effect sizes are represented. Seven effect sizes produced by a first-generation measure and 5 effect sizes produced by a fourth-generation measure were not included in the table.

* $p < .05$, indicates that the level of variability is greater than would be expected by chance.

TABLE 3: Comparison of Static and Dynamic-Based Instruments for Institutional Violence^a and Recidivism^b

Measure	k	N	M_r (SD)	CI_r	Z^+	CI_{Z^+}	Q
Institutional violence							
Static	26	226026	.22 (.12)	.17 to .27	.32	.316 to .324	210.48*
Dynamic	37	5616	.20 (.14)	.15 to .25	.21	.18 to .24	165.50*
Combination ^c	12	1029	.27 (.14)	.18 to .36	.23	.17 to .29	14.36
Violent recidivism							
Static	64	13409	.22 (.13)	.19 to .26	.22	.20 to .24	152.17*
Dynamic	96	21913	.22 (.13)	.19 to .24	.25	.24 to .26	512.45*
Combination ^c	13	1697	.23 (.15)	.14 to .32	.20	.15 to .25	28.53*

Note. k = effect sizes per risk measure; N = offenders per risk measure; CI = confidence interval; Z^+ = r' value weighted by sample size; CI_{Z^+} = 95% confidence interval about Z^+ .

a. Only 75 of the 76 institutional violence outcomes are represented because the nature of predictors could not be determined for 1 effect size.

b. Only 173 of 185 recidivism effect sizes are represented because the nature of predictors could not be determined for 12 effect sizes.

c. Only measures based on an equivalent number of static and dynamic risk factors were included in this coded category. Thus, this coded category does not reflect a statistical combination of the primarily static and primarily dynamic risk measure categories.

* $p < .05$, indicates that the level of variability is greater than would be expected by chance.

CI overlap between this measure and the LSI-R and PCL-R, but its CI did not overlap with the HCR-20 or SIR scale. Based on the widths of the Z^+ CI s shown in Table 1, the LSI-R, PCL-R, and SIR scale each generated slightly more precise point estimates than the HCR-20 and VRAG. Fail-safe analyses indicated that only six additional null VRAG effect sizes would be needed to reduce its predictive ability to that of the HCR-20 or SIR scale. Only another two null VRAG effect sizes would be needed for the VRAG to perform at par with the LSI-R or PCL-R.

In terms of notable measures for violent recidivism with ≤ 10 effect sizes (not in Table 1), the LS/CMI ($k = 3$, $N = 841$) yielded relatively strong predictive validity ($Z^+ = .47$, $CI_{Z^+} = .40$ to $.54$), followed closely by the SAQ ($k = 8$, $N = 1094$, $Z^+ = .37$, $CI_{Z^+} = .31$ to $.43$). The CI s for these two measures only slightly overlapped and may be estimating distinct population parameters. Note, however, that any conclusions about these two measures must be made

TABLE 4: Comparison of Administration Methods for the Prediction of Institutional Violence^a and Recidivism^b

Measure	k	N	M _r (SD)	CI _r	Z ⁺	CI _{Z⁺}	Q
Institutional violence							
File review	32	223071	.23 (.14)	.19 to .29	.34	.336 to .344	209.14*
Interview only	6	635	.17 (.09)	.08 to .27	.14	.06 to .22	2.52
Self-report	13	2505	.18 (.11)	.11 to .25	.16	.12 to .20	21.63*
File/interview ^c	13	1352	.26 (.14)	.18 to .35	.22	.17 to .27	20.18
Violent recidivism							
File review	97	24648	.24 (.13)	.21 to .26	.26	.25 to .27	591.04*
Interview only	21	2921	.14 (.11)	.09 to .19	.11	.07 to .15	24.58
Self-report	29	5029	.16 (.13)	.11 to .21	.12	.09 to .15	53.31*
File/interview ^c	27	5741	.26 (.09)	.22 to .29	.30	.27 to .33	100.13*

Note. *k* = effect sizes per risk measure; *N* = offenders per risk measure; *CI* = confidence interval; *Z*⁺ = *r*' value weighted by sample size; *CI*_{*Z*⁺} = 95% confidence interval about *Z*⁺.

a. Only 64 of the 76 institutional violence outcomes are represented because the administration method could not be determined for 12 effect sizes.

b. Only 174 of 185 recidivism effect sizes are represented because the administration method could not be determined for 11 effect sizes.

c. Only measures scored from information gathered by means of using an interview with the offender and a file review are included in this category. It does not reflect the statistical combination of the interview-only and file-only effect size categories.

**p* < .05, indicates that the level of variability is greater than would be expected by chance.

TABLE 5: Comparison of Relevant Versus Less Relevant Measures for the Prediction of Institutional Violence and Recidivism^a

Measure	k	N	M _r (SD)	CI _r	Z ⁺	CI _{Z⁺}	Q
Institutional violence							
Relevant	63	214444	.22 (.12)	.19 to .25	.35	.346 to .354	286.10*
Less relevant	13	18346	.21 (.20)	.09 to .33	.27	.26 to .28	144.25*
Violent recidivism							
Relevant	153	33031	.23 (.13)	.21 to .25	.26	.25 to .27	647.86*
Less relevant	25	5835	.09 (.12)	.05 to .14	.07	.04 to .10	56.93*

Note. *k* = effect sizes per risk measure; *N* = offenders per risk measure; *CI* = confidence interval; *Z*⁺ = *r*' value weighted by sample size; *CI*_{*Z*⁺} = 95% confidence interval about *Z*⁺.

a. Only 178 of 185 recidivism effect sizes are represented because the relevance of the measures could not be determined for 7 effect sizes.

**p* < .05, indicates that the level of variability is greater than would be expected by chance.

in light of the few effect sizes available on their predictive validity, especially for the LS/CMI. Other notable measures were the Psychopathy Checklist: Screening Version (PCL:SV) (*k* = 5, *N* = 641, *Z*⁺ = .20, *CI*_{*Z*⁺} = .12 to .28), the Salient Factor Score (SFS; *k* = 5, *N* = 989, *Z*⁺ = .15, *CI*_{*Z*⁺} = .09 to .21), and measures comprised solely of criminal history variables (*k* = 9, *N* = 2230, *Z*⁺ = .23, *CI*_{*Z*⁺} = .19 to .27). The MMPI (using the Megargee Typology and the Prison Adjustment Scale) did not predict violent recidivism (*k* = 3, *Z*⁺ = .00).

COMPARISON OF EFFECT SIZES BY GENERATION OF RISK INSTRUMENT

Table 2 displays the mean effect sizes across generations of risk measures.² First- and fourth-generation measures were excluded from the table because each had ≤10 effect sizes, but tentative data on these methods are described below. As shown in Table 2,

second-generation instruments outperformed those of the third generation as predictors of institutional violence. This was because of the substantial weight given to three particularly large second-generation studies with $ns > 10,000$ offenders. Fail-safe calculations estimated that another 34 second-generation effect sizes of zero would be required before its mean effect would lower and become equivalent to third-generation measures in the prediction of institutional violence. The benefit of second- versus third-generation instruments was reversed when the outcome was violent recidivism. Third-generation measures had a slight advantage over those of the second generation, with no overlap of their *CI*s. According to the fail-safe index, another 23 null effect sizes for third-generation measures would be needed to reduce this category's mean effect to that of the second-generation instruments for the outcome of violent recidivism. For the generations not referenced in Table 2, first-generation methods produced a Z^+ of .18 ($k = 7, N = 1461, CI_{Z^+} = .13$ to .23) for violent recidivism. Of all the generations, fourth-generation measures ($k = 5, N = 3759$) resulted in the largest predictive estimate ($Z^+ = .52, CI_{Z^+} = .49$ to .55) for violent recidivism and shared no overlap with first-, second-, and third-generation effect size estimates.

COMPARISONS BASED ON THE CONTENT OF THE INSTRUMENT: STATIC VERSUS DYNAMIC

Table 3 summarizes the predictive validity for instruments containing primarily static or dynamic risk items and those with an equal combination of static and potentially dynamic items.³ For institutional violence, static instruments had a significantly larger mean effect ($Z^+ = .32$) than did dynamic ($Z^+ = .21$) and combined ($Z^+ = .23$) instruments. According to the fail-safe index, an additional 14 static effect sizes of zero would be needed to reduce its predictive magnitude to the level of dynamic instruments. Further, 10 additional nil effect sizes would be necessary to reduce the predictive estimate of static instruments to that of the combination instruments. In terms of violent recidivism, it was the dynamic instruments that had a slight advantage over static instruments as evidenced by very little *CI* overlap between these factors (i.e., $p < .05$). The mean effect for dynamic instruments was marginally larger than that for combination instruments, with very slight overlap of the two *CI*s as well. Fail-safe calculations indicated that another 13 dynamic effect sizes of zero would be needed to reduce this category's predictive validity to that of static measures. An additional 24 nil effect sizes would be required to lower the predictive power of dynamic measures to that of the combination measures.

COMPARISONS BASED ON MEASURE ADMINISTRATION METHOD

Comparisons of mean predictive validities between different administration methods are presented in Table 4. Beginning with institutional violence, the largest Z^+ value (.34) was attributed to the file review only. The *CI* associated with this effect shared no overlap with self-report, interview-only, or file-and-interview methods. Fail-safe calculation revealed that an additional 36 null file extraction effect sizes would be needed to reduce its mean effect to that of the self-report category; a further 17 effect sizes of zero would be needed for parity with the file-and-interview method; and 46 nil effect sizes for equality with the interview-only method. With regard to violent recidivism, the second part of Table 4 shows that the file-and-interview method had the largest predictive validity ($Z^+ = .30$). The *CI* for this category only touched that of file extraction methods, and shared no overlap with the other two methods. To reduce the predictive accuracy of file-plus-interview to that of file

review only, interview only, or self-report methods, an additional 4, 47, and 41 nil file review, respectively, would be needed.

COMPARISONS BASED ON INSTRUMENT RELEVANCE TO CORRECTIONS

The final comparison of interest was the relevance of an instrument to corrections. Each effect size was coded as to whether the measure was derived from a criminological theory and/or whether it was created specifically for use as a risk instrument.⁴ For instance, a measure like the LSI-R was coded as relevant to corrections because it was both derived from theories of criminality and created for use as a risk instrument. The VRAG also was coded as relevant because, although not created from theory, it was specifically created for risk evaluation. Less relevant instruments were those assessing constructs found to be unrelated or weakly related to correctional outcomes (e.g., literacy, self-esteem). Table 5 lists results for relevant versus less relevant instruments. Relevant instruments were better predictors of both institutional violence and recidivism, with no overlap in *CI*s with less relevant instruments. Fail-safe analyses indicate that, for institutional violence, an additional 19 effect sizes of zero would be needed to reduce their predictive validity of relevant measures to that found for less relevant measures. For violent recidivism, as many as 415 new null effects for the relevant instrument category would be needed to equate its validity to that found for less relevant measures.

DISCUSSION

Although professionals are presented with a range of tools for use in risk assessment, a challenge arises when trying to decide which of these instruments is most suitable. To assist with the decision-making process, the current meta-analysis synthesized research focusing on the predictive validities of various instruments used to assess violence risk. From a pool of 88 studies, a total of 185 effect sizes were produced for violent recidivism and 76 were obtained for institutional violence. Collapsed across instruments, their moderate ability to predict risk outcomes was consistent with estimates reported in other risk prediction meta-analyses (e.g., Gendreau et al., 1997; Gendreau et al., 1996; Schwalbe, 2007; Walters, 2006). Although most of the common risk instruments analyzed produced relatively equivalent predictive estimates, variations related to specific instruments are discussed below.

The following discussion should be considered with a mind to the limitations of the current meta-analysis. The first set of limitations related to serious deficits in the primary studies that prohibited the coding of important variables as potential moderators (e.g., 68% of effect sizes were based on studies in which there was insufficient information to code or define the violence history within a particular sample). In addition, none of the institutional violence studies provided details about their sample's pre-existing level of institutional violence. Omission of violence history data precluded an examination of this variable as a moderator of effect size. Furthermore, information was lacking for 56% of effect sizes about the nature of the sample's index offenses (violent vs. nonviolent), which prevented examination of the moderating effects of index offense severity on predictive validity. It was also noted that 21 effect sizes were derived from studies that did not report the gender

composition of their samples. When gender was noted, it was clear that most of the effect sizes were generated from male samples. Thus, generalization of the results to female offenders, and other poorly represented offender subgroups (e.g., native offenders), is limited. One final methodological issue was that over 88% of effect sizes were generated from samples defined as low or moderate risk to reoffend. Thus, it is difficult to generalize the current findings to high-risk samples with any degree of certainty until additional data with this population has been accrued.

PREDICTION OF VIOLENT RECIDIVISM

Instruments comprised primarily of dynamic risk items generated the strongest effect size for violent recidivism ($Z^+ = .25$). The *CI* for this category shared minimal overlap with the *CI* for instruments derived primarily of static-based risk items ($Z^+ = .22$). When considered with the fail-safe index, this finding suggests a small advantage for potentially dynamic over static risk instruments when it comes to predicting violent reoffending and replicates Gendreau et al. (1996). However, it is interesting to note the performance of the primarily static VRAG, which produced an effect size equivalent to that generated by dynamic risk measures. In addition, third-generation instruments produced a better estimate of violent recidivism risk than did second-generation measures. This finding is consistent with Schwalbe (2007), whose meta-analysis of adolescent risk measures also found a slight predictive advantage for third-generation measures over those of the second generation. While noting the limitations associated with only five effect sizes for fourth-generation measures, this category produced the strongest predictive estimate of the different generations ($Z^+ = .52$). Thus, further evaluation of this newer generation of risk measures is crucial.

In examining the mean effect size magnitudes for individual instruments with ≥ 10 effect sizes, it was clear that each predicted violent recidivism with at least a moderate degree of success. The mean effect sizes ranged from .22 for the HCR-20 and SIR scales to .32 for the VRAG. The LSI-R, PCL-R, and SIR scales provided the most precise point estimates (i.e., the narrowest *CI*s), but no one measure stood out as the most effective for predicting violent recidivism. The VRAG performed well, but its *CI* overlapped with those of the LSI-R and the PCL-R. According to fail-safe indexes, only two null VRAG effect sizes would be required to reduce its mean effect to that of the LSI-R and PCL-R. Thus, they are all likely sampling the same population parameter. Our analysis also found that the LSI-R was equivalent in its predictive validity to that of the PCL-R, and to a lesser degree with the HCR-20 and SIR scale. The current meta-analysis updated Gendreau et al. (2002), who had originally found a slight advantage of the LSI-R over the PCL-R in predicting violent recidivism. With the inclusion of additional effect sizes published since Gendreau et al., these results suggest that the PCL-R and LSI-R are actually more comparable than not as predictors of violent reoffending.

In summary, most of the measures reported in Table 1 appear to be similar in their predictive power. The one exception was that the VRAG had a predictive advantage over both the HCR-20 and the SIR scale. Collectively, these data suggest that the variation across primary studies in the predictive validity estimates of most risk instruments is a reflection of sampling error (Hunter & Schmidt, 2004). Only more primary studies will offer a definite conclusion on this matter. At present, our results are congruent with findings that suggest that many of the commonly used risk instruments are moderately to highly

intercorrelated (e.g., Dahle, 2006; Glover et al., 2002). The similarity between instruments was further reflected in Kroner, Mills, and Reddon (2005), who randomly generated four hybrid risk measures based on the item content of the PCL-R, LSI-R, VRAG, and GSIR. When they tested each of these measures on their ability to predict general recidivism, the hybrid instruments performed as well as each of their respective parent instruments. Thus, there is a significant degree of overlap between the common risk measures.

PREDICTION OF INSTITUTIONAL VIOLENCE

Unlike the prediction of violent recidivism, there was much more variability within the individual risk instruments in their ability to predict institutional violence. An aggregate category of criminal history indexes ($Z^+ = .26$) produced the most precise mean effect size, as noted by its very narrow *CI*. This was a catch-all category of measures related to past criminality, which makes its value difficult to interpret. In terms of standardized risk measures, the HCR-20 had the greatest number of effect sizes and produced the largest mean effect size for institutional violence ($Z^+ = .28$). Despite its strong performance, the HCR-20 has challenges related to its use in that the current data were derived from the numerical risk score of the HCR-20 and not the structured clinical prediction judgments advocated for use in its clinical application by the test developers (Webster et al., 1997). In addition, data for the HCR-20 were primarily generated from forensic psychiatric samples; this limits its generalizability to institutional violence in nonpsychiatric correctional facilities. In terms of other measures, the PCL:SV ($Z^+ = .22$) and LSI-R ($Z^+ = .24$) were moderately predictive of institutional violence, while the VRAG ($k = 2$) and PCL-R ($k = 5$) each recorded small associations with this outcome ($Z^+ = .15$ and $.14$, respectively). A few primary VRAG studies predicting institutional violence have come to light since the completion of our analysis (e.g., McDermott, Edens, Quanbeck, Busse, & Scott, 2008; Nadeau, Nadeau, Smiley, & McHattie, 1999). The inclusion of these data in future meta-analyses may clarify the role of the VRAG in predicting this outcome. Consistent with the current results regarding the PCL-R, Guy, Edens, Anthony, and Douglas (2005) found that the PCL-R produced a mean weighted effect size of $.17$ for physical aggression in the institution. Guy et al.'s analysis suggested that the PCL-R was better used as a predictor of verbal aggression than physical aggression within institutional settings. Thus, caution is warranted in the choice of instrument to predict risk within an institutional setting until data has been sufficiently compiled (i.e., at least 10 effect sizes per instrument).

In contrast to our violent recidivism data, second-generation instruments had an advantage over third-generation measures ($Z^+ = .34$ vs. $.20$, respectively) in predicting institutional violence. More specifically, instruments based on criminal history and other static variables were more informative than other types of measures when estimating the risk of institutional violence. It is possible that static factors were more valuable as risk items when assessing institutional violence because of the short-term follow-up duration of these assessments. Most of these effect sizes were based on studies with follow-up periods of less than 1 year. Arguably, the effect of dynamic factors on behavior may have had little time to emerge over such short periods. The inclusion of dynamic risk factors may be more relevant to longer-term predictions of institutional violence (as was the case for recidivism, which had longer follow-up periods).

Despite justifiable concern about the accuracy of predicting future violence, and the ongoing debate as to which measure is best to achieve this goal, there are still remarkably

few studies available to address these issues (i.e., the largest number was obtained for the PCL-R at $k = 24$). We caution that there is likely little value in the generation of new risk measures at this point. The last thing the risk assessment field needs is to imitate the wasted efforts found in the psychiatric rehospitalization prediction literature, in which 419 scales have been produced with only 3 reporting more than 10 predictive validity estimates (Smith, Gendreau, & Goggin, 2007). Instead, research should focus on further validation of existing risk measures within different forensic contexts and offender subgroups. Such information will likely better showcase an individual measure's strengths and weaknesses. Additional data will also provide much more precise estimates (i.e., width of *CIs*) of these point estimates.

CONTENT RELEVANCE AND ADMINISTRATION METHOD

Similar to Walters's (2006) meta-analysis, measures with content relevant to criminal behavior and risk constructs yielded more accurate predictive validities for violent recidivism than instruments containing unrelated and/or less relevant content (e.g., anxiety). Although less relevant instruments performed substantially better in predicting cases of institutional violence, content-relevant instruments were still superior for this outcome as well. A notable finding within this data set was the little attention received by the MMPI-2 as a predictor of future violence in recent prospective research. This is of concern because the MMPI has been one of the most commonly used assessment instruments by psychologists working in correctional settings in the United States (Boothby & Clements, 2000). Only one study reported on the predictive validity of the MMPI (Megargee Typology) as an index of future violence (L. L. Motiuk, 1991). This study found that it was a poor predictor of violent recidivism and only performed slightly better as a predictor of institutional violence. Thus, assessors must be cautious when using this instrument to inform decisions about violence risk in light of the lack of recent data. Its use should be directed more to understanding potential personality dynamics and mental health problems that may be relevant to responsivity concerns.

These data support the inclusion of self-report measures in the assessment of violence risk but not as the sole means of prediction. Specifically, the mean effect size for the general category of self-report measures was small for both violent recidivism and institutional violence ($Z^+ = .12$ and $.16$, respectively). The file review only and file-and-interview approaches to assessment produced the largest predictive validities for both outcomes. Given that we did not separate relevant and less relevant self-report measures, it is possible that the less relevant self-report measures in this category detracted from its overall mean effect size. This possibility is supported by Walters (2006). Of the self-report measures that are content relevant in nature, the SAQ stands out. The mean SAQ effect size for violent recidivism in the current data was very promising ($Z^+ = .37$). The SAQ's prediction of institutional violence was based on only one effect size (Loza & Loza-Fanous, 2002) but might have some utility in that domain as well. The advantage of the SAQ is that it yields valuable information about risk-need factors that have been empirically associated with risk outcomes (see Bonta, Law, & Hanson, 1998; Gendreau et al., 1996), including anti-social attitudes, characteristics of antisocial personality disorder, early behavior problems, past criminal behavior, substance abuse, and antisocial associates. This instrument also contains a validity index and an anger subscale (Loza & Loza-Fanous, 2003). Nevertheless,

the SAQ and other content-relevant self-report measures (e.g., PICTS) require additional study to document their ability to predict future violence and require testing within different subgroups of offender populations. Only one prospective study has tested the validity of the SAQ with female offenders from different ethnic groups (Loza, Neo, Shahinfar, & Loza-Fanous, 2005).

CONCLUSION

A practical issue for professionals involved in risk prediction and treatment is the selection of the best instruments for their work with offenders (Bonta, 2002). Although this analysis found little difference among the predictive validities of actuarial and structured instruments for violent reoffending, this does not mean that they would be equally informative for case planning when the goal is risk reduction. The interested reader should refer to Bonta (2002) and Quinsey, Harris, Rice, and Cormier (1998) for useful professional practice parameters relevant to the selection of instruments for the purposes of violence risk assessment and reduction. These parameters stress the importance of considering the context and objective of the specific risk assessment as well as the content and structure of a particular risk instrument that is being considered for use. These parameters should be applied within the structure of the risk-need-responsivity principles (see Andrews & Bonta, 2006) for effective case management and risk reduction. Adherence to the risk-need-responsivity principles contributes to greater risk reduction than when these principles are ignored or minimally adopted (e.g., Dowden & Andrews, 2000; French & Gendreau, 2006; Gendreau et al., 2006).

In addition to ongoing prospective validation of existing risk measures, an area for future research is the identification of factors predictive of the nature and context of an offender's violent behavior. Such research aims to identify acute and/or transitory risk factors relevant to determining the imminence of violence or assist with judgments about the likely occurrence of various forms of aggressive behavior (e.g., reactive vs. instrumental or proactive; see Quinsey et al., 1998). The detailed aspects of violence risk and the conditions under which violence is most likely to occur are arguably more useful to case managers than a vague statement about the general estimate of violence risk. Research in this area is important given that assessors have difficulty accurately predicting the likelihood of various dimensions of violent behavior (e.g., severity of aggression, likely imminence of the violent event, weapon use; Douglas, Ogloff, & Hart, 2003).

Examination of incremental validity research may also assist professionals in the selection of the most appropriate measures for the assessment of violence risk. As a case in point, Edens, Skeem, and Douglas (2006) found that the PCL:SV had incremental validity over a modified version of the VRAG (with the PCL and two other items removed) in the prediction of violence in discharged civil psychiatric patients. The VRAG continued to predict violence but it did so more modestly than the PCL:SV itself. Edens et al. concluded that the personality traits assessed by the PCL:SV may be more useful than the VRAG for the assessment of violence risk in nonoffenders. Walters, Knight, Grann, and Dahle (2008) also reported on incremental validity variations within the four facet scales of the PCL-R and PCL:SV. Thus, the combined interpretation of data from both incremental validity studies and meta-analytic research provides useful information about the comparative contributions between measures and within various components of a single measure.

Another issue for future research is attention to the composition of the comparison group for violent recidivists in prediction studies. The majority of coded studies (97%) defined their outcome criterion for nonrecidivism in such a way that this group likely included recidivists of other types of crimes (e.g., nonviolent recidivists). As a result, little predictive data were available using a pure outcome criterion of no recidivism at all. This problem is compounded by the practice of plea bargaining and police discretion at the early stages of legal involvement, which may result in some violent offenses being reduced to nonviolent charges. These offenders would then be incorrectly classified as nonviolent recidivists if the coding of recidivism categories was based solely on the type of charge. Examination of offense descriptions may help minimize this classification difficulty. Effect size estimates for risk instruments may be larger when distinguishing between participants with no recidivism at all and violent recidivists. Future violence prediction studies should attempt to operationalize the outcome criteria in a way that reflects a pure violent–nonviolent recidivist dichotomy to determine whether there is an effect on predictive validity estimates.

A final issue relates to the practice of interpreting confidence intervals within meta-analyses. Examination of confidence intervals is one means by which researchers can determine the degree to which an effect size estimate represents population parameters. There are no consistent decision rules about the appropriate width of a *CI* required to create a precise estimate of the population parameters (Smithson, 2003). In response to the lack of decision rules, Snook, Eastwood, Gendreau, Goggin, and Cullen (2007) argued in a related criminal justice field that correlation coefficients effect sizes with *CI* widths of $>.10$ are imprecise estimates. The number of effect sizes required to achieve this criterion will depend on many factors (e.g., variables of interest, study quality, representativeness of the data) but the objective is to collect as much data as is necessary to efficiently narrow the confidence interval around a point estimate. Some variability is to be expected, but wide *CI*s ($>.10$; see Gendreau & Smith, 2007) only point to the need for additional research and offer little insight into the true state of population parameters.

NOTES

1. Base rate adjustments were required because correlations based on the binomial effect size display assume a 50% base rate (i.e., that half of the population would reoffend violently and half would not; Randolph & Edmondson, 2005). Given that the real-world base rate of violence is lower than 50%, effect size correlations need to be adjusted to account for this lower base rate (see Thompson & Schumacker, 1997).

2. Page constraints limit a detailed reporting of each measure included in the generation categories, but more information can be obtained from the first author.

3. Common examples of indexes coded as containing primarily static risk items included the Statistical Information on Recidivism scale; Violence Risk Assessment Guide; Historical, Clinical, and Risk Management Violence Risk Assessment Scheme, historical scale only; Salient Factor Score; Custody Rating Scale: Institutional Adjustment/Security Risk; Risk Assessment for Prison Scale; history of conduct disorder; and miscellaneous criminal history variables. Common examples of measures coded as containing primarily dynamic risk factors included the Level of Supervision Inventory (LSI)/LSI-Revised/LSI: Screening Version, Violence Risk Scale, Self-Appraisal Questionnaire, and the Psychopathy Checklist (PCL)/PCL-Revised/PCL: Screening Version.

4. Examples of measures coded as content relevant (see Andrews and Bonta, 2006, and Walters, 2006, for elaboration) included such instruments as the Psychopathy Checklist-Revised; Level of Supervision Inventory-Revised; Historical, Clinical, and Risk Management Violence Risk Assessment Scheme; Criminal Sentiments Scale-Modified; Self-Appraisal Questionnaire; Criminal Insensitivity and Irresponsibility Scale; Lifestyle Criminality Screening Form; Measures of Criminal Attitudes and Associates; Diagnostic and Statistical Manual of Mental Disorders-related measures of conduct problems and antisocial personality disorder; Violent Beliefs Inventory; Wisconsin Assessment of Client Risk Scale; and the Offender Group Reconviction Scale. Measures coded as “less relevant” include the Coping Situations Scale, Minnesota Multiphasic Personality Inventory-2, Positive Affect/Negative Affect Scale, Psychological Referral Screening Form, and the Perceived Stress Index.

REFERENCES

Entries marked with an asterisk (*) were included in the meta-analysis.

- Andrews, D. A. (1989). Recidivism is predictable and can be influenced: Using risk assessments to reduce recidivism. *Forum on Corrections Research, 1*, 11-17.
- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory-Revised*. Toronto: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: 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, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7-27.
- Baird, S. C. (1981). Classifying juveniles: Making the most of an important management tool. *Corrections Today, 47*, 32-38.
- Baird, S. C., Heinz, R. C., & Bemus, B. J. (1979). *The Wisconsin Case Classification/Staff Development project: A two-year follow-up report*. Madison, WI: Wisconsin Bureau of Community Corrections.
- *Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *Journal of Forensic Psychiatry, 11*, 167-175.
- *Blanchette, K. (2005). *Field-test of a gender-informed security re-classification scale for female offenders*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Blanchette, K., Verbrugge, P., & Wichmann, C. G. (2002). *The Custody Rating Scale, initial security level placement, and women offenders* [Research Report #127]. Ottawa, Ontario: Correctional Services of Canada.
- *Bonta, J. (1989). Native inmates: Institutional response, risk, and needs. *Canadian Journal of Criminology, 31*, 49-61.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29*, 355-379.
- *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, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*, 123-142.
- *Bonta, J., & Motiuk, L. L. (1986, August). *Use of the Level of Supervision Inventory for assessing incarcerates*. Paper presented at the annual convention of the American Psychological Association, Washington, DC.
- *Bonta, J., & Yessine, A. K. (2005). *The national flagging system: Identifying and responding to high-risk, violent offenders* [Report No. 2005-04]. Ottawa, Ontario: Department of Public Safety and Emergency Preparedness Canada.
- Boothby, J. L., & Clements, C. B. (2000). A national survey of correctional psychologists. *Criminal Justice and Behavior, 27*, 715-731.
- Brennan, T., & Oliver, W. L. (2000). *Evaluation of reliability and validity of COMPAS scales: National aggregate sample*. Traverse City, MI: Northpointe Institute for Public Management.
- Brown, S. L., St. Amand, M. D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law and Human Behavior, 33*, 25-45.
- *Collie, R. M., & Polaschek, D. L. L. (2003). Sorting women's risk: New Zealand women prisoners' misconducts and internal security risk. *New Zealand Journal of Psychology, 32*, 101-109.
- *Cooke, D. (1996). Predicting offending in prison: The predictive validity of the prison behavior rating scales. *Legal and Criminological Psychology, 1*, 65-82.
- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-574.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170-180.
- *Cunningham, M. D., & Sorensen, J. R. (2006). Actuarial models for assessing prison violence risk: Revisions and extensions of the Risk Assessment Scale for Prison (RASAP). *Assessment, 13*, 253-265.
- *Cunningham, M. D., Sorensen, J. R., & Reidy, T. J. (2005). An actuarial model for assessment of prison violence risk among maximum security inmates. *Assessment, 12*, 40-49.
- *Daffern, M., Ogloff, J. R. P., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory-Revised: Screening Version. *International Journal of Forensic Mental Health, 4*, 201-206.
- *Dahle, K.-P. (2006). Strengths and limitations of actuarial prediction of criminal reoffense in a German prison sample: A comparative study of LSI-R, HCR-20, and PCL-R. *International Journal of Law and Psychiatry, 29*, 431-442.
- *Dernevik, M., Grann, M., & Johannson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the HCR-20. *Psychology, Crime, & Law, 8*, 93-111.

- *Dolan, M., & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behaviour, 30*, 469-483.
- Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Thousand Oaks, CA: Sage.
- *Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Examination of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54*, 1372-1379.
- Douglas, K. S., & Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383.
- Douglas, K. S., Yeomans, N., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*, 479-510.
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent re-offending: A meta-analysis. *Canadian Journal of Criminology, 44*, 449-467.
- *Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology, 7*, 141-154.
- *Edens, J. F., & Ruiz, M. A. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment, 18*, 220-224.
- Edens, J. F., Skeem, J. L., & Douglas, K. S. (2006). Incremental validity analyses of the Violence Risk Appraisal Guide and the Psychopathy Checklist: Screening Version in a civil psychiatric sample. *Assessment, 13*, 368-374.
- French, S., & Gendreau, P. (2006). Reducing prison misconducts: What works! *Criminal Justice and Behavior, 33*, 185-218.
- *Fujii, D. E., Tokioka, A. B., Lichton, A. I., & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services, 56*, 711-716.
- *Gagliardi, G. J., Lovell, D., Peterson, P. D., & Jemelka, R. (2004). Forecasting recidivism in mentally ill offenders released from prison. *Law and Human Behavior, 28*, 133-155.
- Gendreau, P., Goggin, C., French, S., & Smith, P. (2006). Practicing psychology in correctional settings. In I. B. Weiner & A. K. Hess (Eds.), *The handbook of forensic psychology* (3rd ed., pp. 722-750). Hoboken, NJ: Wiley & Sons.
- Gendreau, P., Goggin, C., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior, 24*, 414-431.
- Gendreau, P., Goggin, C., & Smith, P. (2000). Generating rational correctional policies: An introduction to advances in cumulating knowledge. *Corrections Management Quarterly, 4*, 52-60.
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29*, 397-426.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult recidivism: What works! *Criminology, 34*, 401-433.
- Gendreau, P., & Smith, P. (2007). Influencing the "people who count": Some perspectives on the reporting of meta-analytic results for prediction and treatment outcomes with offenders. *Criminal Justice and Behavior, 34*, 1536-1559.
- *Girard, L. (1999). *The Level of Service Inventory-Ontario Revision: Risk/need assessment and recidivism*. Unpublished doctoral dissertation, University of Ottawa, Ontario, Canada.
- *Glover, A. J. J., Nicholson, D. E., Hemmati, T., Berfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high risk federal offenders. *Criminal Justice and Behavior, 29*, 235-249.
- Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27*, 97-114.
- *Grann, M., Långström, N., Tengström, A., & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior, 23*, 205-218.
- *Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of the HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71*, 443-451.
- *Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, T. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL:SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*, 523-530.
- *Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry and Psychology, 15*, 278-292.
- 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.
- Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting and Clinical Psychology, 73*, 1056-1064.
- *Halsall, S., & Van Dine, S. (2002). Security reclassification risk instrument for males: Development and methodology. *Ohio Corrections Research Compendium, 1*, 95-96.
- Hanson, R. K., & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessment for sexual offenders: A meta-analysis*. Ottawa, Ontario: Public Safety and Emergency Preparedness Canada.
- *Hanson, R. K., & Wallace-Capretta, S. (2000). *Predicting recidivism among male batterers* [User report 2000-2006]. Ottawa, Ontario: Solicitor General of Canada.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto: Multi-Health Systems.

- *Harer, M. D., & Langan, N. P. (2001). Gender differences in predictors of prison violence: Assessing the predictive validity of a risk classification system. *Crime and Delinquency, 47*, 513-536.
- *Harris, G. T., Rice, M. E., & Cormier, C. A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior, 15*, 626-637.
- *Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26*, 377-394.
- *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.
- Hathaway, S. R., & McKinley, J. C. (1967). *Minnesota Multiphasic Personality Inventory manual* (rev. ed.). New York: Psychological Corporation.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for analysis*. Toronto: Academic Press.
- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*, 347-359.
- *Heilbrun, K., Hart, S. D., Hare, R. D., Gustafson, D., Nunez, C., & White, A. J. (1998). Inpatient and post-discharge aggression in mentally disordered offenders: The role of psychopathy. *Journal of Interpersonal Violence, 13*, 514-527.
- *Hemphill, J. F. (1998). *Psychopathy, criminal history, and recidivism*. Unpublished doctoral dissertation, University of British Columbia, Canada.
- *Hemphill, J. F., Templeman, R., Wong, S., & Hare, R. D. (1998). Psychopathy and crime: Recidivism and criminal careers. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 375-399). Dordrecht, Netherlands: Kluwer Academic.
- *Hildebrand, M., De Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among male offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence, 19*, 13-29.
- Hoffman, P. B. (1983). Screening for risk: A revised salient factor score (SFS 81). *Journal of Criminal Justice, 11*, 539-547.
- Hoge, R. D., & Andrews, D. A. (1996). *Assessing the youthful offender: Issues and techniques*. New York: Plenum.
- *Holland, T. R., Holt, N., Levi, M., & Beckett, G. E. (1983). Comparison and combination of clinical and statistical predictions of recidivism among adult offenders. *Journal of Applied Psychology, 68*, 203-211.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- *Jayjohn, J., & Van Dine, S. (2002). Standard initial security classification risk instrument for males: Development and methodology. *Ohio Corrections Research Compendium, 1*, 93-94.
- *Jemelka, R. P., Wiegand, G. A., Walker, E. A., & Trupin, E. W. (1992). Computerized offender assessment: Validation study. *Psychological Assessment, 4*, 138-144.
- *Kroner, D. G., & Loza, W. (2001). Evidence for the efficacy of self-report in predicting nonviolent and violent criminal recidivism. *Journal of Interpersonal Violence, 16*, 168-178.
- *Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471-489.
- Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A coffee can, factor analysis, and prediction of antisocial behaviour: The structure of criminal risk. *International Journal of Law and Psychiatry, 28*, 360-374.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1995). *Manual for the Spousal Assault Risk Assessment Guide* (2nd ed.). Vancouver, British Columbia, Canada: British Columbia Institute of Family Violence.
- *Law, M. A. (2004). *A longitudinal follow-up of federally sentenced women in the community: Assessing the predictive validity of the dynamic characteristics of the community intervention scale*. Unpublished doctoral thesis, Carleton University, Ottawa, Ontario, Canada.
- Ley, B. A. (1972). *Quantitative aspects of psychological assessment: An introduction*. London: Duckworth.
- *Lovell, D., Gagliardi, G., & Phipps, P. (2005). *Washington's dangerous mentally-ill offender law: Was community safety increased?* [Report #05-03-10901]. Olympia: Washington State Institute for Public Policy.
- Loza, W. (2005). *The Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and non-violent recidivism*. Toronto: Mental Health Systems.
- *Loza, W., & Green, K. (2003). The Self-Appraisal Questionnaire: A self-report measure for predicting recidivism versus clinician-administered measures: A 5-year follow-up study. *Journal of Interpersonal Violence, 18*, 781-797.
- *Loza, W., & Loza-Fanous, A. (1999). *Predictive validity of the LSI: A tool for assessing violent and non-violent recidivism*. Unpublished manuscript.
- *Loza, W., & Loza-Fanous, A. (2000). Predictive validity of the Self-Appraisal Questionnaire (SAQ): A tool for assessing violent and nonviolent release failures. *Journal of Interpersonal Violence, 15*, 1183-1191.
- *Loza, W., & Loza-Fanous, A. (2001). The effectiveness of the Self-Appraisal Questionnaire in predicting offenders' post-release outcome: A comparison study. *Criminal Justice and Behavior, 28*, 105-121.
- *Loza, W., & Loza-Fanous, A. (2002). The effectiveness of the Self-Appraisal Questionnaire (SAQ) as an offenders' classification measure. *Journal of Interpersonal Violence, 17*, 3-14.
- *Loza, W., & Loza-Fanous, A. (2003). More evidence for the validity of the Self-Appraisal Questionnaire for predicting violent and nonviolent recidivism: A 5-year follow-up study. *Criminal Justice and Behavior, 30*, 709-721.

- Loza, W., Neo, L. H., Shahinfar, A., & Loza-Fanous, A. (2005). Cross-validation of the Self-Appraisal Questionnaire: A tool for assessing violent and non-violent recidivism with female offenders. *International Journal of Offender Therapy and Comparative Criminology*, 49, 547-560.
- *Loza, W., Villeneuve, D. B., & Loza-Fanous, A. (2002). Predictive validity of the Violence Risk Appraisal Guide: A tool for assessing violent offenders' recidivism. *International Journal of Law and Psychiatry*, 25, 85-92.
- *Luciani, F., Motiuk, L., & Nafekh, M. (1996). *An operational review of the Custody Rating Scale: Reliability, validity, and practical utility* [Research Report #R-47]. Ottawa, Ontario: Correctional Services of Canada.
- McDermott, B. E., Edens, F., Quanbeck, C. D., Busse, D., & Scott, C. L. (2008). Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable and person-focused analyses. *Law and Human Behavior*, 32, 325-338.
- *McHattie, L., Hills, A., Smiley, C., & MacKenzie, R. (1999). *Does the CATS predict recidivism as well as the PCL-R?* Abbotsford, British Columbia, Canada: Regional Health Centre.
- *Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice and Behavior*, 32, 565-585.
- *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.
- *Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence of the robustness of self-report among offenders. *Criminal Behaviour and Mental Health*, 13, 140-150.
- Monahan, J., & Steadman, H. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.
- *Motiuk, L. L. (1991). *Antecedents and consequences of prison adjustment: A systematic assessment and reassessment approach*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Motiuk, M. S., Motiuk, L. L., & Bonta, J. (1992). A comparison between self-report and interview-based inventories in offender classification. *Criminal Justice and Behavior*, 19, 143-159.
- *Müller-Isberner, R., Sommer, J., Özokuy, K., & Freese, R. (1999, November). *Clinical use of the HCR-20 for predicting violence in a German Forensic Psychiatric Hospital*. Paper presented at the International Conference on Risk Assessment and Risk Management: Implications for the Prevention of Violence, Vancouver, British Columbia, Canada.
- Nadeau, J., Nadeau, B., Smiley, W. C., & McHattie, L. (1999, November). *The PCL-R and VRAG as predictors of institutional behaviour*. Paper presented at the International Conference on Risk Assessment and Risk Management: Implications for the Prevention of Violence, Vancouver, British Columbia, Canada.
- National Council on Crime and Delinquency. (2004). *Correctional assessment and intervention system*. Oakland, CA: Author.
- *Nicholls, T. (2001). *Violence risk assessments with female NCRMD acquittees: Validity of the HCR-20 and PCL-SV*. Unpublished doctoral dissertation, Simon Fraser University, British Columbia, Canada.
- *Nicholls, T. L., Vincent, G. M., & Whitemore, K. E. (1999, November). *Assessing risk of inpatient violence in a sample of forensic psychiatric patients: Comparing the PCL:SV, HCR-20, and VRAG*. Paper presented at the International Conference on Risk Assessment and Risk Management: Implications for Prevention of Violence, Vancouver, British Columbia, Canada.
- *Nugent, P. M. (2000). *The use of detention legislation: Factors affecting detention decisions and recidivism among high-risk federal offenders in Ontario*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- *Polvi, N. H. (2001). *The prediction of violence in pre-trial forensic patients: The relative efficacy of statistical versus clinical predictions of dangerousness*. Unpublished doctoral dissertation, Simon Fraser University, British Columbia, Canada.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk* (1st ed.). Washington, DC: American Psychological Association.
- Randolph, J. J., & Edmondson, R. S. (2005). Using the binomial effect size display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical Assessment, Research, and Evaluation*, 14, 1-7.
- *Raynor, P. (1998). *Implementing the Level of Service Inventory-Revised (LSI-R) in Britain: Report 3, risk and need assessment in the five pilot areas*. Swansea, UK: Centre for Criminal Justice and Criminology, Department of Social Policy and Applied Social Studies, University of Wales & The Cognitive Centre Foundation, Dinas Powys Vale, Glamorgan, UK.
- *Rettinger, L. J. (1998). *A recidivism follow-up study investigating risk and need within a sample of provincially sentenced women*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- Rice, M. E. (1997). Violent offender research and implications for the criminal justice system. *American Psychologist*, 52, 414-423.
- *Rice, M. E., & Harris, G. T. (1992). A comparison of criminal recidivism among schizophrenic and nonschizophrenic offenders. *International Journal of Law and Psychiatry*, 15, 397-408.
- *Rice, M. E., & Harris, G. T. (1995). Psychopathy, schizophrenia, alcohol abuse, and violent recidivism. *International Journal of Law and Psychiatry*, 18, 333-342.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

- *Rowe, R. (1995). *The utilization of an interview-based classification instrument for parole board decision making in Ontario*. Unpublished doctoral dissertation, Carleton University, Ottawa, Ontario, Canada.
- *Rowe, R. (1997). *Applying risk-needs assessments to parole decisions*. Unpublished manuscript.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training researchers. *Psychological Methods, 1*, 115-129.
- Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31*, 449-462.
- *Serin, R. C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior, 20*, 207-217.
- *Serin, R. C., & Brown, S. L. (1998). *Empirical guidelines for clinicians: Enhancing the practice of offender risk assessment*. Ottawa, Ontario: Research Branch, Correctional Service of Canada.
- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17*, 156-167.
- *Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31*, 306-323.
- *Simourd, D. J., & Van De Ven, J. (1999). Assessment of criminal attitudes: Criterion-related validity of the Criminal Sentiments Scale—Modified and the Pride in Delinquency Scale. *Criminal Justice and Behavior, 26*, 90-106.
- Smith, P., Gendreau, P., & Goggin, C. G. (2007). "What works" in predicting psychiatric hospitalization and relapse: The specific responsivity dimension of effective correctional treatment for mentally disordered offenders. In R. Ax & T. Fagan (Eds.), *Corrections, mental health, and social policy: International perspectives* (pp. 209-233). Springfield, IL: Charles C. Thomas.
- Smithson, M. (2003). *Confidence intervals: Quantitative applications in the social sciences series, No. 140*. Thousand Oaks, CA: Sage.
- Snook, B., Eastwood, J., Gendreau, P., Goggin, C., & Cullen, R. M. (2007). Taking stock of criminal profiling: A narrative review and meta-analysis. *Criminal Justice and Behavior, 34*, 437-453.
- *Stadtland, C., Kleindienst, N., Kröner, C., Eidt, M., & Nedopil, N. (2005). Psychopathic traits and risk of criminal recidivism in offenders with and without mental disorders. *International Journal of Forensic Mental Health, 4*, 89-97.
- *St. Amand, M. D. (2003). *The coping relapse model of recidivism and the prediction of failure after relapse*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- *Strand, S., Belfrage, H., Fransson, G., & Lavender, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders—More important than historical data? A retrospective study of 40 mentally disordered offenders assessed with the HCR-20 Violence Risk Assessment Scheme. *Legal and Criminological Psychology, 4*, 67-76.
- *Stribling, A. L. (2003). *Exploring differences in clinical judgments of dangerousness of North Carolina psychiatric inpatients: Utility of the HCR-20 risk assessment scheme*. Unpublished doctoral dissertation, University of Tennessee.
- Swets, J. A. (1986). Indices of discrimination and diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100-117.
- *Tengström, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry, 55*, 243-249.
- *Tengström, A., Grann, M., Långström, N., & Kullgren, G. (2000). Psychopathy (PCL-R) as a predictor of violent recidivism among criminal offenders with schizophrenia. *Law and Human Behavior, 24*, 45-58.
- Thompson, K. N., & Schumacker, R. E. (1997). An evaluation of Rosenthal and Rubin's binomial effect size display. *Journal of Educational and Behavioural Statistics, 22*, 109-117.
- *Urbaniok, F., Noll, T., Grunewald, S., Steinbach, J., & Endrass, J. (2006). Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. *Journal of Forensic Psychiatry and Psychology, 17*, 23-31.
- *Villeneuve, D. B., Oliver, N., & Loza, W. (2003). Cross-validation of the Self-Appraisal Questionnaire with a maximum-security psychiatric population. *Journal of Interpersonal Violence, 18*, 1325-1334.
- *Villeneuve, D. B., & Quinsey, V. L. (1995). Predictors of general and violent recidivism among mentally disordered inmates. *Criminal Justice and Behavior, 22*, 397-410.
- Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles (PICTS). Part I: Reliability and preliminary validity. *Criminal Justice and Behavior, 22*, 307-325.
- Walters, G. D. (1996). The Psychological Inventory of Criminal Thinking Styles (PICTS). Part III: Predictive validity. *International Journal of Offender Therapy and Comparative Criminology, 40*, 105-112.
- Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes. *Criminal Justice and Behavior, 33*, 179-304.
- *Walters, G. D., Duncan, S. A., & Geyer, M. D. (2003). Predicting disciplinary adjustment in inmates undergoing forensic evaluation: A direct comparison of the PCL-R and PAI. *Journal of Forensic Psychiatry and Psychology, 14*, 382-393.
- Walters, G. D., Knight, R. A., Grann, M., & Dahle, K. P. (2008). Incremental validity of the Psychopathy Checklist facet scores: Predicting release outcome in six samples. *Journal of Abnormal Psychology, 117*, 396-405.
- *Walters, G. D., & Mandell, W. (2007). Incremental validity of the Psychological Inventory of Criminal Thinking and Psychopathy Checklist: Screening Version in predicting disciplinary outcome. *Law and Human Behavior, 31*, 141-157.

- Walters, G. D., White, T. W., & Denney, D. (1991). The Lifestyle Criminality Screening Form: Preliminary data. *Criminal Justice and Behavior, 18*, 406-418.
- Webster, C. D., Douglas, K., Eaves, D., & Hart, D. (1997). *HCR-20: Assessing risk for violence* (Ver. 2). Vancouver, British Columbia, Canada: Mental Health, Law, & Policy Institute, Simon Fraser University.
- Webster, C. D., Harris, G. T., Rice, M. E., Cormier, C., & Quinsey, V. L. (1994). *The Violence Prediction Scheme: Assessing dangerousness in high risk men*. Toronto: University of Toronto, Centre of Criminology.
- *Wintrup, A. M. (1996). *Assessing risk of violence in mentally disordered offenders with the HCR-20*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.
- *Wong, S. C. P., & Gordon, A. (2001). The Violence Risk Scale. *Forensic Update, 67*, 19-23.
- *Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and the Law, 12*, 279-309.
- *Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2006). *The long-term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment*. Unpublished manuscript.
- Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review, 17*, 292-309.

Mary Ann Campbell, PhD, is an assistant professor in the Department of Psychology and director of the Centre for Criminal Justice Studies at the University of New Brunswick (Saint John Campus). She is also a private practice licensed psychologist in the provinces of New Brunswick and Nova Scotia, and she sits as a consultant with the Adult Forensic Team with Mental Health Services (Region B) in Saint John, New Brunswick, Canada.

Sheila French, MA, is a member of the Program Development and Therapeutic Services Unit within the Ministry of Corrections, Public Safety, and Policing, Government of Saskatchewan, Canada, and is the director of clinical services at the Prince Albert Youth Residence in Prince Albert, Saskatchewan. She is currently pursuing a doctoral degree through the University of New Brunswick.

Paul Gendreau, PhD, is a professor emeritus in the Department of Psychology at the University of New Brunswick (Saint John campus) and former director of the Centre for Criminal Justice Studies at the University of New Brunswick. He was inducted as an officer in the Order of Canada in 2007 for his national and international work related to the advancement of correctional theory, risk assessment, and offender rehabilitation.



Criminal Justice and Behavior (CJB) is pleased to offer the opportunity to earn continuing education credits online for this article. Please follow the access instructions below. The cost is \$9 per unit, and a credit card number is required for payment.

Access for IACFP members:

- Go to the IACFP homepage (www.ia4cfp.org) and log in to the site using your email address
- Click on the Publications heading in the sidebar on the left-hand side of the page or on the Publications icon at the bottom of the page
- Click on the *CJB* cover image located at the bottom of the page. You will be prompted for your user name and password to log in to the journal
- Once you have logged in to *CJB*, select the appropriate issue and article for which you wish to take the test, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

Access for institutional subscribers:

- Log in through your library or university portal to access *CJB*
- Select the appropriate issue and article, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

If you are not a member of the IACFP and your institution does not subscribe to *CJB*, individual articles (and corresponding test links within the articles) are available on a pay-per-view basis by visiting <http://cjb.sagepub.com>. However, membership in the IACFP provides you with unlimited access to *CJB* and many additional membership benefits! Learn more about becoming a member at www.ia4cfp.org.