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

2007-01

R. Karl Hanson

Public Safety and Emergency Preparedness Canada

Kelly E. Morton-Bourgon

Department of Justice Canada

Cat. No.: PS4-36/2007E
ISBN No.: 978-0-662-45166-2

Table of Contents

Abstract

This paper reviewed the accuracy of different approaches to risk assessment for sexual offender recidivism. The risk assessments were classified according to the source of the items (empirical or conceptual) and the method for combining the items into an overall evaluation of risk (professional judgement or actuarial). Based on 577 findings from 79 distinct samples, the actuarial measures designed for specific types of outcome (sexual, violent, or general recidivism) were the best predictors of that specific outcome (i.e., sexual recidivism was best predicted by the measures designed to predict sexual recidivism). The results for evaluations based on structured professional judgement were variable, with average results intermediate between the findings for the actuarial assessments and unstructured professional judgements (which were consistently the least accurate). There were no significant differences in the predictive accuracy of empirically derived actuarial measures and the conceptually derived actuarial measures. These results suggest that it is possible to conduct psychologically informed risk assessments that have the dual advantages of high predictive accuracy and clinically useful understanding of specific cases.

The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis

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

The individual characteristics associated with recidivism among sexual offenders have been previous reviewed (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). In general, the two broad domains most strongly associated with sexual recidivism are sexual deviancy and lifestyle instability/criminality. The criminal lifestyle characteristics (e.g., history of rule violation, substance abuse) are also those most strongly related to violent and general (any) recidivism among sexual offenders (Hanson & Morton-Bourgon, 2004), general offenders (Gendreau, Little, & Goggin, 1995) and mentally-disordered offenders (Bonta, Law, & Hanson, 1998).

Although a number of recidivism risk factors have been identified, the relationships between any single risk factor and recidivism is small. Consequently, competent evaluations needs to consider a range of risk factors. The question addressed by the current review is the relative accuracy of different methods of combining risk factors into an overall evaluation of risk.

Much of the discussion of methods for combining information for applied decision making in psychology has been shaped by Meehl's (1954) review of clinical versus statistical prediction. For Meehl, the statistical (or actuarial) approach involved explicit procedures for grouping individuals into classes, and statistical frequencies for linking class membership to outcome. The clinical approach, in contrast, involves predicting outcome based on understanding the psychological structures and dynamics of specific individuals. The basic question posed by Meehl was the following: which approach is more accurate in predicting future behaviour? Meehl (1954) concluded in favour of the actuarial approach, as have most subsequent reviewers (Ægisdóttir et al., 2006; Andrews, Bonta, & Wormith, 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Quinsey, Harris, Rice, & Cormier, 2006).

Despite the evidence supporting statistical prediction, clinical prediction is still ubiquitous in judicial settings (Janus & Prentky, 2003) and has its advocates among forensic evaluators (Litwack, 2001). Some of the clinical-actuarial debate concerns technical issues that are in principle tractable through further research; the debate is also driven, however, by fundamentally different visions concerning the purpose of forensic evaluations of risk.

For some clinicians, the accuracy of the prediction is all that matters. To clinicians valuing this "pure prediction" approach, the ideal scale would include all non-redundant risk factors, the

factors would be optimally weighted, and the measure would be applied to individuals from the same population for which the measure was developed and validated. The content of a scale would not matter as long as it contributes to predictive accuracy. Examples of scales developed from this approach include the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 2006) and the iterative classification tree models developed by the Banks et al. (2004).

For other clinicians, the goal of risk assessment is to understand the case. From the perspective of maximizing understanding, the ideal risk assessment would be based on a valid theoretical model of risk. This theory would be generally true, and true for the specific individual being assessed. Rather than selecting variables based on their empirical relationships with recidivism (e.g., age at first offence), variables would be selected because they help explain why recidivism is likely or not (e.g., current procriminal attitudes). Understanding the case would not only provide an estimate of risk of recidivism, but it would also provide direction as to how the case can be managed to reduce that risk. Examples of risk assessment procedures developed from the "understanding" perspective include the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) and the Sexual Violence Risk – 20 (Boer, Hart, Kropp, & Webster, 1997).

Critics of the "understanding" approach argue that interpretations require evidence in order to be credible (Meehl, 1954, pp. 136-138). Critics of "pure prediction" approach complain that a score on a prediction instrument is of limited value for understanding and managing the case (Hart, 1998).

A close look at the approaches to risk assessment currently used demonstrates that the above distinction between prediction and understanding is not synonymous with Meehl's distinction between clinical and actuarial prediction. The items for the LS/CMI, for example, were based on a social psychological models of crime, and are organized into conceptually meaningful content areas (e.g., substance abuse; negative family relationships). There are, however, explicit rules for scoring the items, mechanical rules for combining the items into subscores, and mechanical rules for calculating the overall risk score. Meehl would consider the LS/CMI an actuarial measure, but it is quite different from the pure prediction approach of Banks' iterative classification trees (ICT). The ICT approach is explicitly atheoretical and, given the computational complexity, it "would clearly be impossible for a clinician to commit the multiple models and their scoring to memory" (Banks et al., 2004). Evaluators are also unlikely to memorize the LS/CMI scoring guide; nevertheless, after completing an assessment, they should be able to explain why the offender was placed into a particular risk category.

Given that the global distinction between clinical and actuarial prediction does not describe important differences in risk assessment procedures, the current review organized risk assessments into four categories (see Table 1). These categories were based on Sawyer's (1966) distinction between the factors examined in the risk assessment, and the method used to combine the factors into an overall evaluation of risk. In the field of sexual offence research, most of the existing risk assessments can be grouped into one of four categories based on

whether the factors are empirically or conceptually derived, and whether the final judgement is determined by professional judgement or an explicit algorithm.

Table 1. Types of risk assessments

	Items	Overall Judgement	Examples
Empirical Actuarial	Empirical	Mechanical	VRAG, Static-99
Conceptual Actuarial	Conceptual	Mechanical	Vermont's Sex Offender Treatment Needs and Progress Scale
			Adding items from SVR-20
Structured Professional Judgement	Conceptual	Unstructured	SVR-20
Unstructured	Case specific/ unknown	Unstructured	

In the empirical actuarial approach, the items are selected based on observed relationships with outcome, and explicit rules are provided for combining the items into an overall evaluation of risk (e.g., VRAG, Quinsey et al., 2006; Static-99, Hanson & Thornton, 2000). In the conceptual actuarial approach, the final judgement is determined by explicit rules, but the items are selected based on theory. Examples of sexual offender risk tools developed by this approach include Thornton's Structure Risk Assessment (Thornton, 2002) and the Sex Offender Treatment Needs and Progress Scale (McGrath & Cumming, 2003).

Structured professional judgement requires evaluators to rate a list of pre-determined items, but the final evaluation is left to professional judgement (e.g., see SVR-20; Boer et al., 1997). It is not uncommon, however, for researchers to omit the professional judgement and simply add the items from the checklist. In this case, the instrument becomes a conceptual actuarial measure. An interesting research question is the effectiveness of adding the items compared to using professional judgement to form the overall evaluation. Structured professional judgement has been promoted as providing clinically meaningful case formulations while avoiding the dismal predictive accuracy associated with the unstructured clinical approach (Douglas, Cox, & Webster, 1999; Hart, 1998).

In unstructured professional judgement the risk factors are not specified in advance, nor is the method of combining the risk factors into an overall evaluation of risk. Unstructured professional judgement has been the traditional loser in the actuarial-clinical debates. It should be noted, however, that it is not one approach; unstructured professional judgement would be conducted differently by different evaluators.

The major question addressed in this review is the relative accuracy of these four approaches to risk assessment (empirical actuarial, conceptual actuarial, structured judgement, unstructured judgement). Other approaches to risk assessment have been described in the literature, such as the adjusted actuarial approach (Hanson, 1998: called "clinical synthesis" by Sawyer, 1966), and fully idiographic approaches (called "anamnestic approaches" by Doren, 2002). Although these approaches are commonly used by evaluators, they have received insufficient empirical study to merit inclusion in the current review.

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

Method

Sample

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

Studies were included if they examined the ability of risk assessments to predict sexual, violent (including sexual) or any recidivism among offenders released following an index sexual offence. Risk assessments were defined as global assessments of the risk for recidivism (e.g., dangerousness, likelihood of recidivism) made with or without the aid of guidelines or actuarial tools. Studies that only examined specific attributes related to risk (e.g., psychopathy, benefit from treatment, sexual deviancy) are reported elsewhere (Hanson & Morton-Bourgon, 2004; 2005).

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

As of May, 2006, our search yielded 100 usable documents (e.g., published articles, books, government reports, conference presentations). In 13 cases, the analyses were based on raw data or analyses provided by the original researchers. When the same data set was reported in several articles, all the results from these articles were considered to come from the same study. Consequently, the 100 documents represented 79 different studies (country of origin: 27 United States, 30 Canada, 14 United Kingdom, 2 Sweden, and one each from France, Netherlands, Germany, New Zealand, Belgium and Taiwan; 46 (58.2%) unpublished; produced between 1972 and 2006, with a median of 2002; average sample size of 276, median of 172, range of 10 to 2557).

Most of the studies examined mixed groups of adult sexual offenders (70 mixed offence types, 8 child molesters, 1 rapists; 68 predominantly adults, 11 adolescents; all male). All the offenders had committed offences that meet contemporary definitions of sexual crimes (i.e., old

studies containing homosexuals were excluded). Most of the offenders were released from institutions (48 institution only, 17 community only, 13 institution and community, and 1 unknown). The offenders in 30 studies came from treatment programs. When demographic information was presented, the offenders were predominantly Caucasian (34 of 38 studies).

Effect sizes were recorded for three outcome criteria: a) any sexual recidivism (versus no recidivism or only non-sexual recidivism - 281 effect sizes); b) any violent recidivism (sexual or non-sexual) (versus no recidivism or only non-violent recidivism - 149 effect sizes); and e) any recidivism (versus no recidivism - 147 effect sizes).

The most common sources of recidivism information were national criminal justice records (k = 44; 55.7%) followed by state or provincial records (k = 29; 36.7%). The source of the recidivism information was unknown for 8 studies. The average follow-up period ranged from 6 months to 276 months, with a mean of 67 months (SD = 46.7).

Coding procedure

Each study was coded using a standard list of variables and explicit coding rules (available upon request). The studies were coded by either one of two raters (the authors) and most were independently coded by both raters. All coding was reviewed by the first author. Only one finding per individual variable was coded per sample based on a) sample size, and b) completeness of information. If the sample sizes and descriptive detail were equivalent, the median value was used.

Inter-rater reliability was calculated for approximately 10% of the sample (n=10). Given that two of the reliability studies contained overlapping samples (Friendship, Mann & Beech, 2003; Thornton et al., 2003), one rater combined these two studies whereas the other rater considered them as separate studies. Consequently, the number of studies available for comparison was nine. Agreement for dichotomous and categorical variables was indexed by Kappa, and two-way random effects model intraclass correlation coefficients (type absolute agreement) (ICC) were used for ordinal and interval variables (Design 2 in Orwin, 1994). The agreement for the sample characteristics was perfect for 7 variables, and good to excellent for 8 others (Kappa > .70; ICC > .80). Agreement was fair for four variables: published or not (Kappa = .55 – disagreed on “in press” articles); victims (Kappa = .60; disagreed on whether hebephiles were child molesters or rapists); the total, combined sample size (ICC = .72, a result of one rater combining two studies and the other rater leaving them separate); and whether national criminal records were used to detect recidivism (Kappa = .59).

In the reliability studies, Rater 1 identified 56 findings, and Rater 2 identified 63 findings, with agreement on 104 of the 119 findings (87.4%). The inter-rater reliability of the effect sizes calculation was .68 for a single rater and .81 for the average of two raters. The agreement increased to .79 (single rater) and .88 (two raters) if one outlier was removed. Most differences involved simple omissions or clerical errors.

Index of predictive accuracy

The effect size indicator was the standardized mean difference, *d*, defined as follows: $d = (M_1 - M_2)/S_w$, where M_1 is the mean of the deviant group, M_2 is the mean of the non-deviant group, and S_w is the pooled-within standard deviation (Hasselblad & Hedges, 1995). In other words, *d* measures the average difference between the recidivists and the non-recidivists, and compares this difference to how much recidivists differ from each other, and how much non-recidivists differ from each other. Formula for calculating *d* can be found in Hanson and Morton-Bourgon (2004).

The *d* statistic was selected because it is less influenced by recidivism base rates than correlation coefficients – the other statistic commonly used in meta-analyses. According to Cohen (1988), *d* values of .20 are considered "small", .50 "medium", and .80 "large". The value of *d* is approximately twice as large as the correlation coefficient calculated from the same data. When the 95% confidence interval for *d* does not contain zero, it can be considered statistically significant at $p < .05$. When the confidence intervals for two predictors do not overlap, they can be considered significantly different from each other ($p < .05$).

Aggregation of findings

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

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

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

When d_i was calculated from other statistics (*t*, ROC areas, means, etc.), the variance of d_i was estimated using Formula 3 from Hasselblad and Hedges (1995):

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

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

Results

The observed sexual recidivism rate was 12.4% (2,109/17,038; 72 studies), the violent recidivism rate (including sexual and non-sexual violence) was 17.5% (2,330/13,279; 36 studies) and the general (any) recidivism rate was 30.1% (3,237/10,755; 40 studies). Studies that specified in advance the number of recidivists and non-recidivists were excluded from the rate calculations (e.g., Dempster, 1998). The average follow-up time was 68 months. These figures should be considered underestimates because not all offences are detected.

The average predictive accuracy of the various approaches to risk assessment are summarized in Table 2. For the prediction of sexual recidivism, the empirical actuarial measures designed for sexual recidivism (*d.* = .70, 95% C.I. of .64 to .75) were more accurate than unstructured professional judgement (*d.* = .43, 95% C.I. of .28 to .58). There was no significant difference between the accuracy of the sexual actuarial measures constructed empirically (*d.* = .70) and those created conceptually (*d.* = .66, 95% C.I. of .56 to .75). The actuarial measures created empirically for the prediction of violent recidivism were somewhat less accurate (*d.* = .54, 95% C.I. of .43 to .65) than the empirical actuarial measures designed for sexual recidivism (*d.* = .70), but this difference should be interpreted cautiously given that their confidence intervals overlapped and there was considerable variability within each category. On average, the risk assessments based on structured professional judgement showed predictive accuracy intermediate between that observed for the actuarial measures and for unstructured professional judgement, and were not significantly different from the other approaches.

For the prediction of violent (including sexual) recidivism, the empirical actuarial measures designed for violent recidivism were more accurate (*d.* = .85, 95% C.I. of .76 to .94) than the actuarial measures designed for sexual recidivism (*d.* = .52 and *d.* = .55), structured professional judgement (*d.* = .31) or unstructured professional judgement (*d.* = .30). The same pattern was shown when the outcome criteria was any recidivism. The empirical actuarial measures designed for general (any) recidivism were more accurate (*d.* = 1.13, 95% C.I. of .93 to 1.32) than the empirical actuarial measures designed for violence (*d.* = .79, 95% C.I. of .67 to .92), the actuarial measures designed for the prediction of sexual recidivism (*d.* = .56 and *d.* = .53), or professional judgement – structured (*d.* = .24) or unstructured (*d.* =.22). The actuarial measures designed conceptually for the prediction of any recidivism also showed strong predictive accuracy (*d.* = .88, 95% C.I. of .64 to 1.11) and were significantly more accurate than all the other approaches, except for the empirical actuarial measures designed for violence (*d.* = .79).

Appendix I presents further details of the analyses, as well as the findings for individual risk tools. Within each category, the commonly used risk tools all showed moderate to large predictive accuracy, and, with rare exceptions, their confidence intervals overlapped. Readers interested in specific measure can peruse the tables at their leisure. The measure with the largest association with sexual recidivism was the SVR-20 professional judgment, but this finding was based on only three studies (n = 245) and showed significant variability (Q = 7.96, df = 2, p < .05).

Table 2. The average predictive accuracy of various forms of risk assessment for sexual offenders

Form of Risk Assessment	Recidivism Outcome Criteria					
	Sex		Any violence		Any	
	d. (95% C.I.)	k	d. (95% C.I.)	k	d. (95% C.I.)	k
Measures designed for sexual recidivism						
Empirical Actuarial	.70 (.64 - .75)	55	.52 (.47 - .58)	30	.56 (.51 - .62)	25
Conceptual Actuarial	.66 (.56 - .75)	22	.55 (.44 - .65)	9	.53 (.45 - .62)	13
Structured professional judgement	.42 (.25 - .60)[a]	5	.31 (.13 - .49)	3	.24 (.07 - .40)	8
Measures designed for violent recidivism						
Empirical Actuarial	.54 (.43 - .65)	12	.85 (.76 - .94)	10	.79 (.67 - .92)	6
Conceptual Actuarial	.22 (.01 - .42)	4	I.D.		.36 (.21 - .52)	3
Measures designed for any recidivism						
Empirical Actuarial	.52 (.34 - .71)	4	I.D.		1.13 (.93 –1.32)	3
Conceptual Actuarial	I.D.		I.D.		1.08 (.81 – 1.35)[c]	3
Unstructured professional judgement	.43 (.28 - .58)	9	.18 (.01 - .36)[b]	4	.22 (.09 - .35)	8

[a] Outlier excluded; with outlier *d.* = .57 (.41 - .73). [b] With outlier *d.* = .30 (.14 - .46). [c] With outlier *d.* = .88 (.64 – 1.11).

Note: k is the number of studies; I.D. = Insufficient data (less than three studies found).

Table 3. Comparing professional judgement and the simple addition of items for the prediction of sexual recidivism.

Study	Measure	Judgement *d*	Addition *d*	Recidivists/Total
Sjöstedt & Långström (2002)	SVR-20	.21	-.04	10/ 51
Dempster (1998)	SVR-20	1.23	1.27	24/ 73
de Vogel et al. (2004)	SVR-20	1.35	1.19	47/121
Kropp (2000)	RSVP	.97	.53	15/ 53
Morton (2003)	ERASOR	.14	.31	13/ 77
d. (95% C.I.)		.93 (.69 – 1.17)	.82 (.58 – 1.06)	109/375
Q		16.2 (p = .0028)	15.6 (p = .0036)	

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

To further examine the potential contribution of professional judgement, five studies were identified in which the evaluators rated a predetermined set of items and then formed an overall evaluation of risk based on either a) professional judgement or b) summing the items. As can be seen from Table 4, the results of both procedures were similar. Three studies favour professional judgement and two favoured the simple sums. In most cases, the differences between the approaches were not large enough to be meaningful. It is worth noting, however, that the two instances where summing did better than professional judgement both involved Master's thesis projects where the professional judgement was conducted by students who, at that time, had little experience in applied assessment (Dempster, 1998; Morton, 2003).

The possible effects of design features were examined in two ways. The first approach used the complete set of findings, including all measures and all outcome criteria (577 findings). This approach is comprehensive, but involves considerable error because consistency would not be expected and the lack of independence of findings within samples was ignored. Consequently, the second approach examined the largest category of findings that examined one specific measure and one specific outcome, which in this case was Static-99 predicting sexual recidivism (42 findings from 42 distinct samples). The statistical significance of the continuous moderator variables was tested using the weighted least squares formula described by Hedges (1994).

Across all findings, published findings yielded larger effects ($d.$ = .66; k = 240) than the unpublished findings ($d.$ = .52; k = 337; χ^2 = 75.8, df = 1, p < .001). This was not the case for the Static-99 findings, however, where the effects were similar in the published ($d.$ = .72, k = 16) and unpublished studies ($d.$ =.68, k = 26; χ^2 = 0.53, df = 1, p > .25). Neither analysis found significant relationships with the total sample size (r = -.033, k = 577; r = -.064, k = 42) nor the year of publication (r = .001, k = 577; r = .11, k = 42). For those studies that reported the date at which the assessments were conducted, the more recent risk assessment were more accurate than the older assessments (r = .20, k = 251). This finding can be mainly attributed to the introduction of structured risk tools for sexual offenders beginning in the mid 1990s. The clinical assessments of sexual recidivism risk examined in this study were conducted between 1970 and 1998 (median of 1984) and had not significantly improved during that time period (r = -.11, t = 0.26, df = 7, p > .50).

Across all findings, the effect sizes were larger in the UK ($d.$ = .72, k = 101) than in the US ($d.$ = .56, k = 129) or Canada ($d.$ = .52, k = 272; ; χ^2 = 77.5, df = 2, p < .001). The same pattern was found for Static-99 studies (UK $d.$ = .94, k = 10; US $d.$ = .69, k = 10; Canada $d.$ = .58, k = 16; ; χ^2 = 14.93, df = 2, p < .001).

Studies that used fixed follow-up times had larger effects than studies that used variable follow-up times (overall, $d.$ = .68 [110] versus $d.$ = .54 [451], χ^2 = 75.8, df = 1, p < .001; Static-99 studies, $d.$ = .81 [8] versus $d.$ = .65 [32], χ^2 = 5.5, df = 1, p < .05). The effect of rater reliability was examined in those studies that reported the intraclass correlation coefficient (k = 133;

range of .60 to 1.0, mean = .87, *SD* = .081). Rater reliability was unrelated to the effect size for both the full group of findings (r = .084, k = 133, p > .05) and the Static-99 findings (r = .039, k = 8, p > .05).

The amount of missing data was measured as the percentage of the total items that were not rated for the average participant in the study. This measure was available for 200 findings and ranged from zero to 30%, with a mean of 5.3% (*SD* = 9.0). The amount of missing data was positively related to the effect size for the full set of findings (r = .15, k = 200, p < .001), and was not related to the Static-99 findings (r = -.19, k = 18, t = 1.09, df = 16, p > .05), or the larger group of the findings using any actuarial-empirical measure (r = .095, k = 56, p > .05). Readers should note that a negative correlation would be expected if missing data degrades the accuracy of the risk tool.

Discussion

The results of the current review provide considerable support for actuarial approaches to risk assessment of sexual offenders. For all outcome measures, the most accurate approaches used empirical actuarial measures designed for that specific outcome (sex, violence, any). On average, the predictive accuracy of both the empirical actuarial measures and the conceptual actuarial measures was moderate to large, and there were no significant differences between these two categories of actuarial measures. The accuracy of structured professional judged varied from low (for any recidivism) to moderate (for sexual recidivism), and there was considerable variability across studies. The accuracy of unguided professional judgement was consistently low.

Readers who are sympathetic to professional judgement will counter that the strongest single predictor of sexual recidivism was a measure of structured professional judgement (the SVR-20). Furthermore, in those studies that directly compared structured professional judgement to simply adding the same items, clinical judgements showed a slight advantage (although the difference was not significant). Given the small sample size and significant variability for the SVR-20 findings, it is easy to dismiss these findings as a statistical fluke. Nevertheless, these findings do suggest it may be possible to create forms of structured professional judgement that have acceptable levels of accuracy.

Another important observation is that there was surprisingly little research on sex offender risk assessments as they are currently practiced. Almost all the studies of professional judgement examined evaluators who did not have access to actuarial risk tools and had little knowledge of empirically established recidivism risk predictors. The use of actuarial risk tools is now expected in many applied contexts (Association for the Treatment of Sexual Abusers, 2005; Doren, 2002), which raises the question of how the advances in research knowledge have influenced the judgements of today's professionals.

In agreement with Doren (2002), we believe that all forms of applied risk assessment involve some degree of professional judgement. There is more to risk assessment than simply scoring a risk tool. Evaluators must decide which actuarial measure to use, and must decide the extent to which the group data associated with an actuarial score applies to a specific individual. The central question is the relative weight given to the actuarial results in the overall risk assessment. Advocates of empirical actuarial measures believe that weight given to actuarial scores should be close to 100% (Quinsey et al., 2006). Critics of actuarial measures believe that the weight should be considerably less than 100% (Litwack, 2001). Given that evaluators have a responsibility to be informed by current research, both empirical actuarial measures and professional judgement need to be considered in applied assessments.

Deciding which actuarial measure to use is a non-trivial issue because no single measures has established itself as clearly more accurate than other similar measures. Kroner, Mills and Reddon (2005), for example, found that randomly selected items from the most commonly used

risk scales predicted general recidivism with the same accuracy as any of the established scales. In any specific data set, one measure may be superior to all others (Seto, 2005), but clinicians would not be expected to know a priori which measure applies best to a particular offender in a particular setting. As the current findings show, the relative accuracy of measures varies across studies and settings. Furthermore, different risk measures are not completely correlated with each other, different tools provide different risk rankings for the same offender (Barbaree, Langton, & Peacock, 2006).

One approach to resolving the divergent results of different empirical actuarial scales is to interpret them based on differences in content (e.g., Doren, 2004b). The problem with this approach is that it assumes construct validity for items that were selected on a purely empirical basis. It is possible to interpret empirically derived items (consider the MMPI), but interpretation is much easier with items that were originally intended to assess specific constructs (i.e., substance abuse, attitudes tolerant of crime). One advantage of the conceptual actuarial measures is that the subcomponents can be defined so that evaluators can identify the reasons for specific scores.

The divergent concerns of those interested only in prediction and those interested in case interpretation are not inherently irresolvable. It would be easy to imagine a synthesis of the two visions through the creation of a theoretically coherent risk tool that predicts recidivism better than other existing methods. Meehl (1954) was correct when he argued that clinical opinions must be defended by evidence; however, the mere documentation of an empirical relationship between a variable and an outcome does not fulfill the needs of those interested in understanding their cases. Given that conceptual actuarial measures in the current study showed similar levels of predictive accuracy to empirical actuarial measures, the goal of theoretically sound and empirically valid risk assessment is well within reach.

One limitation of the current review is that it only examined the accuracy to which the assessments ranked the relative risk of offenders; it did not address their ability to predict absolute recidivism rates. In many applied contexts, however, it is important to know whether the offenders probability of recidivism exceeds some pre-established threshold (see Doren, 2004a; Mossman, 2006). Establishing reliable estimates of the absolute (“real”) recidivism rates is difficult because many offences are undetected, and it is plausible that the recidivism rates will changes across cohorts and jurisdictions. It is possible, for example, that the changes in sexual values from 1970 to 2010 could influence both the probability of recidivism and the likelihood that abusive sexual behaviour would be reported as a crime (Todd, 2006). Given its genesis in data, the empirical actuarial approach will ultimately provide the best estimates of absolute risk. To date, those involved in the development of structured professional guidelines have not even attempted to specify absolute risk levels; instead, they have restricted themselves to relative risk defined in broad categories (e.g., Low, Moderate, High; Hart et al., 2003).

The analysis of design features yielded some findings that should be of interest to researchers. As expected, studies that used fixed and equivalent follow-up periods found larger effects than

studies with variable follow-up times. Contrary to expectations, the size of the findings was unrelated to the rater reliability or the amount of missing data. In the full set of findings, the studies with more missing data actually were larger than those with less missing data. Although there are good reasons to believe that low rater reliability and missing data should decrease the level of predictive accuracy, the results suggest that the amount of variability typically found for these factors has relatively little influence on the predictive accuracy of the measures. What was not evaluated in the current study was the quality of the data used to make the ratings, which would be expected to vary across settings and jurisdictions.

The observation of relatively high predictive accuracy in the UK studies is consistent with the UK's long-term commitment to maintaining comprehensive national criminal history records. Increased reliability of the criminal history records should translate into larger findings in the validity studies. In comparison to the UK, Canada and US have higher rates of recent immigration, which increases the probability of failing to detect offences committed in other countries. Långström (2004) found that the Static-99 predicted recidivism among Nordic and Non-Nordic Europeans, but failed to significantly predict recidivism among the African/Asian offenders in a Swedish national sample. The extent to which the differences in predictive accuracy are due to missing records or due to real social-cultural differences remains to be explored. Regardless of the source, the observation of jurisdictional difference in predictive accuracy make its difficult to directly compare measures tested only in the UK (e.g., Risk Matrix-2000V, Thornton et al., 2003) with measures tested in other jurisdictions.

Implications for applied assessment

Given the weight of evidence supporting them, we believe actuarial risk tools should be a major consideration in the evaluation of recidivism risk potential. Those concerned purely with prediction have a number of different empirical actuarial measures to choose from depending on the offender, the goal of the assessments, and the information and resources available. For the prediction of sexual recidivism, there is strong evidence supporting the reliability and validity of the Static-99, MnSOST-R, and the Risk Matrix – 2000 Sex. The VRAG and SORAG both have demonstrated strong associations with violent (including sexual) recidivism, and in the UK, so has the Risk Matrix – 2000 Violence. The measures developed to predict general recidivism in the general population seem to work equally well at predicting general recidivism among sexual offenders (e.g., SIR, LSI-R).

For those wishing to understand their cases, there are also a number of risk assessment tools available, although the research on these measures is much less developed than the research on empirical actuarial measures. Currently, the most well established measures for understanding cases are the SVR-20 (professional judgment or simply adding the items) and Thornton's (2002) Structured Risk Assessment. At this point, it is hard for evaluators to justify the use of unguided clinical opinion except for cases that fall outside the sampling frame of the existing actuarial measures (e.g., predicting sexual recidivism among adolescent female sexual offenders).

Authors Note

We would like to thank the following researchers who provided unpublished studies, raw data or findings not included in other reports: A. Beech, L. Craig, D. Epperson, G. Harris, M. Hanlon, R. Knight, C. Langton, R. Lieb, J. Looman, L. Motiuk, J. Proulx, J. Reddon, M. Rice, S. Righthand, S. Saum, G. Schiller, and D. Thornton. Thanks are also due to Leslie Helmus for locating articles and formatting tables, and Kuan Li for translating the Chinese study. The views expressed are those of the authors and are not necessarily those of Public Safety and Emergency Preparedness Canada, or those of our many colleagues who made this work possible.

Correspondence should be addressed to R. Karl Hanson, Corrections Research, Public Safety and Emergency Preparedness Canada, 269 Laurier Ave., West, Ottawa, K1A 0P8. Email: Karl.Hanson@psepc-sppcc.gc.ca

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341-382.

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

Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The level of service/case management inventory (LS/CMI).* Toronto, Ontario: Multi-Health Systems.

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

Association for the Treatment of Sexual Abusers. (2005). *Practice standards for the evaluation, treatment and management of adult sexual abusers*. Beaverton, OR: Author.

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

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

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

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

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

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

*Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment, 14,* 155-167.

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

*Beggs, S. M., & Grace, R. C. (2005, November). *Improving sex offender risk assessments: The validity of incorporating measures of treatment outcome*. Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

*Bélanger, N. (1994). *La prédiction de la récidive chez un groupe d'agresseurs sexuels [The prediction of recidivism among a group of sexual offenders]*. Unpublished doctoral dissertation, University of Montreal, Montreal, Quebec, Canada.

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

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

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional guidelines for assessing risk of sexual violence*. Vancouver, BC: The British Columbia Institute Against Family Violence.

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

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

*Brown, J. A. (2004). *A comparison of actuarial methods of predicting sexual dangerousness*. Unpublished doctoral dissertation, Fielding Graduate Institute, Santa Barbara, CA.

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

*Cohen, M. L., Groth, A. N., & Siegel, R. (1978). The clinical prediction of dangerousness. *Crime and Delinquency, 24,* 28-39.

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

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

*Craissati, J., & Beech, A. (2005). Risk prediction and failure in a complete urban sample of sex offenders. *Journal of Forensic Psychiatry & Psychology, 16,* 24-40.

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

*Craissati, J., Webb, L., & Keen, S. (2005). *Personality disordered sex offenders.* Unpublished report. London, UK: Bracton Centre (Oxleas NHS Trust), London Probation Area and Home Office.

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

*Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments.* Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

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

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond.* Thousands Oaks: Sage.

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

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

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

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

*Epperson, D. L. (2003). Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota Prison and Probation Samples. Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation.

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

*Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W. & Goldman, R. (2000, November). *Cross-validation of the Minnesota Sex Offender Screening Tool-Revised*. Paper presented at the 19th annual conference of the Association for the Treatment of Sexual Abusers, San Diego, CA. Retrieved from http://psych-server.iastate.edu/faculty/epperson/atsa2000/index.htm

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

*Florida Department of Health and Rehabilitative Services. (1985). *Status of the sex offender treatment programs, fiscal year 1983-84 (Annual Report to the Florida Legislature).* Tampa, FL: Author.

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

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

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

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.

*Haag, A. M. (2005). *Do psychological interventions impact on actuarial measures: An analysis of the predictive validity of the Static-99 and Static-2002 on a re-conviction measure of sexual recidivism.* Unpublished doctoral dissertation, University of Calgary, Calgary, Alberta, Canada.

*Hanlon, M. J., Bender, S., Larson, S., & Schoepp, J. (2004). *The MnSOST and MnSOST-r in North Dakota: Validation and utility*. Unpublished manuscript. North Dakota State Penitentiary, PO Box 5521, Bismarck, ND 58506-5521, USA.

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

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

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

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

Hanson, R. K., & Morton-Bourgon, K. (2005). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

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

*Hanson, R. K., Harris, A. J. R., & Thomas, T. L. (2006). [Recidivism data for the Dynamic Supervision Project, March 8, 2006]. Unpublished raw data.

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

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

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

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

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

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

*Hartwell, L. L. (2001). *Sex Offender Risk Appraisal Guide: Validity and Utility for Hawaii Sex Offender Risk Assessments.* Unpublished clinical research project, American School of Professional Psychology, Hawaii Campus.

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

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

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

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

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

*Hills, A. L. (2002). *Assessment, treatment, and recidivism of Aboriginal and non-Aboriginal offenders: A comparison of intra-familial and extra-familial male sexual offenders in Saskatchewan.* Unpublished master's thesis, University of Regina, Regina, Saskatchewan, Canada.

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

*Hudson, S. A. (2003). *The Static-99 and community notification in Tarrant County*. Unpublished Master's thesis, University of Texas at Arlington, Arlington, TX.

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

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

*Kolko, D. J. (2005, November). Court-mandated juvenile sexual offenders in the community: Characteristics, collaborative treatment, and recidivism. Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Salt Lake City, UT.

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

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

*Kropp, P. R. (2000, November). *The Risk for Sexual Violence Protocol (RSVP).* Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

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

*Langton, C. M. (2003a). *Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy.* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

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

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7,* 409-443.

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

*Looman, J., Hammond, E., & Abracen, J. (2005). *A test of Doren (2004) dimensions of risk model in a sample of high risk sexual offenders.* Unpublished manuscript, Regional Treatment Centre, Kingston, Ontario, Canada.

*Martinez, R., Rosenfeld, B., & Flores, J. (2004, March). *Predictive validity of the J-SOAP II for sexually offending minority youth*. Paper presented at the Annual Meeting of the American Psychology-Law Society, Scottsdale, AZ.

McGrath, R. J., & Cumming, G. F. (2003). *Sex Offender Treatment Needs and Progress Scale Manual*. Downloaded from for the Center for Sex Offender Management's (CSOM) website: www.csom.org.

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

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

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

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

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

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

*Morton, K. E. (2003). *Psychometric properties of four risk assessment measures with male adolescent sexual offenders*. Unpublished master's thesis, Carleton University, Ottawa, Ontario, Canada.

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

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

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

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

*Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M. & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment, 14*, 253-269.

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

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

*Olver, M. E. (2003). *The development and validation of the Violence Risk Scale: Sexual Offender Version (VRS: SO) and its relationship to psychopathy and treatment attrition.* Unpublished doctoral dissertation, University of Saskatchewan, Saskatoon, Saskatchewan.

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

*Parks, G. A. (2004). *Juvenile sex offender recidivism: Typological differences in risk assessment.* Unpublished doctoral dissertation, Walden University, Minneapolis, MN.

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

*Prentky, R., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 12*, 71-93.

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

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

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

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

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

*Saum, S. (2006). *A comparison of an actuarial risk prediction measures (Static-99) and a stable dynamic risk prediction measure (Stable-2000) in making risk predictions for a group of sexual offenders.* Unpublished doctoral dissertation, Fielding Institute, Santa Barbara, CA.

Sawyer, J. (1966). Measurement and prediction: Clinical and statistical. *Psychological Bulletin, 66*, 178-200.

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

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

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

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

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

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

*Skowron, C. (2004). *Differentiation and predictive factors in adolescent sexual offending*. Unpublished Ph.D. Thesis, Carleton University, Ottawa, Ontario, Canada.

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

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

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

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

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

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

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

*Ternowski, D. R. (2004). *Sex offender treatment: An evaluation of the Stave Lake Correctional Centre Program*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, B.C.

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

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

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

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

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

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

*Tough, S. E. (2001). Validation of two standardized risk assessments (RRASOR, 1997; *Static-99, 1999) on a sample of adult males who are developmentally disabled with significant cognitive deficits.* Unpublished master's thesis, University of Toronto, Toronto, Ontario, Canada.

*Wieand, L. A. (1983). Judgments of dangerousness: An analysis of mental hospital decisions to release mentally disordered sex offenders. Doctoral dissertation, University of California, Riverside. (UMI No. 8323449).

*Williams, S. M., & Nicholaichuk, T. (2001, November). *Assessing static risk factors in adult female sex offenders under federal jurisdiction.* Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.

*Wilson, R. J., Picheca, J. E., & Prinzo, M. (2005). *Circles of support and accountability: An evaluation of the pilot project in South-Central Ontario* (No. R-168). Research Branch: Correctional Services of Canada Report.

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

*Witte, T., Di Placido, C., Gu, D., & Wong, S. (2002, May). *Criminal Sentiments Scale: Predictive validity in a sample of sexual offenders primarily composed of rapists*. Poster session presented at the Canadian Psychological Association's Annual Convention, Vancouver, British Columbia, Canada.

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

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

Appendix 1: Table 1. Prediction of sexual recidivism

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial – empirical/sex	.73	.70	.64	.75	102.83***	55	14,160	1, 2.99, 3, 4, 5, 6.1, 8, 10.99, 11, 12, 13, 14, 15.99, 17, 18a, 18b, 21, 22, 23, 25, 26, 27, 31.1b, 31.99, 32, 34.99, 37.99, 38.99, 41, 42, 43, 44.1, 45, 46, 47, 48.99, 49, 51, 53, 54, 57, 61, 63, 65, 66, 67, 69.1, 70, 71, 72, 73, 74, 75, 77, 78.1a
Static-99	.74	.70	.64	.76	78.09***	42	13,288	2.1, 5, 6.1, 8, 10.2, 11, 13, 14, 15.1, 17, 18a, 18a, 21, 23, 25, 27, 31, 32, 34.1, 37, 38, 41, 42, 44.2, 45, 46, 47, 48, 51, 53, 57, 65, 66, 67, 69.1, 70, 72, 73, 74, 75, 78.1a, 78.1b
Static-2002	.78	.78	.65	.91	2.30	5	2,290	8, 21, 37.1, 53, 74
RRASOR	.69	.59	.52	.65	80.66***	28	8,673	2.1, 10, 10.1, 11, 15, 18a, 18b, 23, 27, 34, 37, 38.1, 41, 44.2, 46, 47, 48.1, 49, 53, 61, 63, 65, 69.1, 71, 74, 75, 77, 78
MnSOST-R	.81	.72	.58	.86	9.19	8	1,684	2.1, 15.1, 18a, 18b, 22, 37, 44, 74
SACJ-Min	.57	.48	.29	.67	1.89	5	914	1, 11, 12, 26, 54
Risk Matrix –2000 sex	.72	.82	.68	.97	4.50	6	1,814	3, 4, 11, 31.1a, 31.1b, 74

Table continues

Appendix 1: Table 1 continued

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial –empirical/violence	.55	.54	.43	.65	12.47	12	2,634	2.1, 11, 15, 17, 27, 28, 37, 38, 48, 56, 61, 74
VRAG	.46	.51	.39	.63	8.23	7	1,699	15, 17, 27, 37, 56, 61, 74
SORAG	.60	.61	.49	.73	12.77	8	2,192	2.1, 15, 27, 28, 37, 38, 48, 74
Actuarial – empirical/any recidivism								
Statistical Index of Recidivism (SIR)	.79	.52	.34	.71	10.61*	4	736	9, 10.5, 24, 47
Actuarial – conceptual/sex	.62	.66	.56	.75	28.66	22	4,592	1, 4, 10.3, 11.99, 14, 15.99, 23, 25, 29, 37, 39, 40, 41, 46, 50, 52, 57, 61, 62, 67, 73, 74
SVR-20 adding the items	.60	.66	.52	.81	23.64**	8	1,346	10.4, 11, 14, 15, 37, 61, 67, 74
Structure Risk Assessment (SRA)	.78	.79	.57	1.02	0.53	3	637	11.1, 73, 74
JSOAP	.40	.31	-.11	.74	4.02	4	317	29, 39, 50, 52
Beech Deviance	.54	.71	.17	1.26	1.26	3	229	1, 4, 5
Actuarial – conceptual/violence	.14	.22	.01	.42	5.61	4	705	10.4, 37, 42, 67
HCR-20 adding the items	.24	.41	.12	.71	2.36	3	237	10.4, 42, 67

Table continues

Appendix 1: Table 1continued

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Structured Professional Judgement – sex	.36	.42	.25	.60	5.74	5	844	15.99, 37, 44, 46, 61
With outlier	.40	.57	.41	.73	22.84***	6	965	14
SVR-20	1.23	1.11	.82	1.40	7.96*	3	245	14, 15, 61
Unstructured professional judgement	.44	.43	.28	.58	5.71	9	1,723	16, 32, 37, 55, 58, 59, 64, 68, 76

Appendix 1: Table 2. Prediction of violent recidivism

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial – empirical/sex	.53	.52	.47	.58	74.92***	30	11,308	2.99, 6, 8, 10.2, 11, 12, 13, 15, 23, 25, 26, 27, 31, 32, 33, 37, 38.99, 41, 46, 47, 48.99, 53, 54, 60, 65, 67, 69, 70, 71, 78
Static-99	.65	.58	.53	.64	74.00***	25	10,166	2.1, 6, 8, 10.2, 11, 13, 23, 25, 27, 31, 32, 33, 37, 38, 41, 46, 47, 48, 53, 60, 65, 67, 69, 70, 78
Static-2002	.72	.71	.57	.85	1.26	3	1,211	8, 37.1, 53
RRASOR	.29	.32	.26	.38	63.23***	17	6,131	2.1, 10.2, 11, 15, 23, 27, 37, 38.1, 41, 46, 47, 48.1, 53, 65, 69, 71, 78
SACJ-Min	.45	.45	.27	.62	1.60	4	789	11, 12, 26, 54
Actuarial –empirical/violence	.89	.85	.76	.94	23.33**	10	2,571	2.1, 11, 15, 27, 31.1a, 31.1b, 37, 38, 48, 56
VRAG	.87	.80	.66	.93	5.42	3	1,023	27, 37, 56
With outlier	.97	.85	.72	.98	12.68**	4	1,118	15
SORAG	.70	.74	.63	.86	3.31	5	1,550	2.1, 27, 37, 38, 48
With outlier	.72	.79	.68	.91	18.01**	6	1,645	15
Risk Matrix –2000 violence	1.00	.98	.81	1.16	2.11	3	767	11, 31.1a, 31.1b

Table continues

Appendix 1: Table 2 continued

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial – conceptual/sex	.53	.55	.44	.65	22.80**	9	2,832	11, 15, 23, 25, 37, 40, 41, 46, 67
Structure Risk Assessment (SRA)	.47	.46	.29	.64	3.21	3	670	11, 37, 67
With outlier	.57	.57	.41	.73	13.88**	4	765	15
Structured Professional Judgement – sex	.39	.31	.13	.49	12.75**		579	15, 37, 46
Unstructured professional judgement	.31	.18	.01	.36	3.61	4	928	32, 37, 55, 68
With outlier	.31	.30	.14	.46	16.09**	5	1,263	36

Appendix 1: Table 3. Prediction of any recidivism

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial – empirical/sex	.53	.56	.51	.62	44.19**	25	8,298	2.99, 10.3, 11, 12.1, 13, 14, 17, 22, 23, 25, 26, 30, 31, 33, 33, 37, 42, 46, 47, 53, 54, 65, 67, 70, 71
Static-99	.55	.61	.55	.66	25.77	21	7,411	2.1, 10.3, 11, 12.1, 13, 14, 17, 23, 25, 30, 31, 33, 33, 37, 42, 46, 47, 53, 65, 67, 70
RRASOR	.25	.27	.19	.35	3.76	9	2,754	2.1, 11, 23, 37, 46, 47, 53, 65, 71
MnSOST-R	.62	.62	.45	.80	0.87	3	570	2.1, 22, 37
SACJ-Min	.53	.57	.41	.74	9.75*	4	789	11, 12, 26, 54
Risk Matrix –2000 sex	.50	.56	.35	.78	1.21	3	530	11, 12.1, 13
Actuarial – empirical/violence	.82	.79	.67	.92	4.97	6	1,107	2.1, 7, 11, 17, 28, 37
SORAG	.85	.82	.68	.96	3.47	4	875	2.1, 7, 28, 37
Actuarial – empirical/any recidivism								
Statistical Index of Recidivism (SIR)	1.22	1.13	.93	1.32	9.05*	3	543	7.1, 9, 47

Table continues

Appendix 1: Table 3 continued

Variable	Median	Mean	95% C.I.		Q	k	Total	Studies
Actuarial - conceptual/sex	.56	.53	.45	.62	19.72	13	3,096	11, 13, 14, 23, 25, 29, 30, 37, 40, 46, 52, 62, 67
SVR-20 adding the items	.59	.54	.39	.69	3.34	4	791	11, 14, 37, 67
Actuarial – conceptual/violence	.61	.36	.21	.52	4.61	3	682	37, 42, 67
Actuarial – conceptual/any	1.02	1.08	.81	1.35	0.62	3	263	20, 30, 47
With outlier	.95	.88	.64	1.11	8.52*	4	340	46
Structured Professional Judgement – sex	.18	.24	.07	.40	5.77	3	605	14, 37, 46
Unstructured professional judgement	.10	.22	.09	.35	15.67*	8	1, 189	16, 19, 35, 37, 55, 59, 64, 79

Appendix 1: Table 4. Key to studies used in the meta-analysis

Number	Study	Number	Study
1	Allam (1999)	36	Kozol et al. (1972)
2.0	[Arizona] Fischer (2000)	37.0	Langton (2003)
2.1	[Arizona] Bartosh et al. (2003)	37.1	Langton et al. (2007)
2.99	[Arizona]	37.99	[Langton]
3	Bates et al. (2004)	38.0	Looman (2006)
4	Beech & Ford (2006)	38.1	Looman et al. (2005)
5	Beech et al. (2002)	38.99	[Looman et al.]
6.0	Beech et al. (2004)	39	Martinez et al. (2004)
6.1	Beech (2005)	40	McGrath et al. (2005)
7.0	Bélanger & Earls (1996)	41	McGrath et al. (2001)
7.1	Bélanger (1994)	42	Milton (2003)
8	Boer (2003)	43	Min-chieh (2005)
9	Bonta & Hanson (1995)	44.0	[Minnesota] Epperson et al. (1995)
10.0	[Clearwater] Haynes et al. (2000)	44.1	[Minnesota] Epperson et al. (2000)
10.1	[Clearwater] Nicholaichuk (1997)	44.2	[Minnesota] Brown (2004)
10.2	[Clearwater] Nicholaichuk (2001)	45	Montana & Thompson (2005)
10.3	[Clearwater] Olver (2003)	46	Morton (2003)
10.4	[Clearwater] Witte et al. (2001)	47	Motiuk & Brown (1995)
10.5	[Clearwater] Witte et al. (2002)	48.0	Nunes et al. (2002)
10.99	[Clearwater]	48.1	Nunes et al. (2003)
11.0	Craig et al. (2006)	48.99	[Nunes et al.]
11.1	Craig et al. (in press)	49	Ohio (2001)
11.99	[Craig et al.]	50	Parks (2004)
12.0	Craissati et al. (2002)	51	Poole et al. (2000)
12.1	Craissati & Beech (2005)	52	Prentky et al. (2000)
13	Craissati et al. (2005)	53	Proulx (2004)

Table continues

Appendix 1: Table 4 continues

Number	Study	Number	Study
14	de Vogel et al. (2004)	54	Proulx et al. (1995)
15.0	Dempster (1998)	55	Reddon et al. (1996)
15.1	[Dempster] Kropp (2000)	56	Rice & Harris (1997)
15.99	[Dempster]	57	Saum (2005)
16	Dix (1976)	58	Schiller (2000)
17	Ducro & Pham (2006)	59	Schram et al. (1991)
18 (a, b)	Epperson (2003)	60	Seager et al. (2004)
19	Florida (1985)	61	Sjöstedt & Långström (2002)
20	Girard & Wormith (2004)	62	Skowron (2004)
21	Haag (2005)	63	Smiley et al. (1998)
22	Hanlon et al. (2004)	64	Smith & Monastersky (1986)
23	Hanson (2002)	65	Song & Lieb (1994)
24	Hanson & Harris (2000)	66	Soothill et al. (2005)
25	Hanson et al. (2006)	67	Stadtland et al. (2005)
26	Hanson et al. (1993)	68	Sturgeon & Taylor (1980)
27	Harris et al. (2003)	69.0	[Sweden] Sjöstedt & Långström (2001)
28	Hartwell (2001)	69.1	[Sweden] Långström (2004)
29	Hecker et al. (2002)	70	Ternowski (2004)
30	Hills (2002)	71	Thornton (1997)
31.0	[HM Prison] Friendship et al. (2003)	72	Thornton (2000)
31.1 (a, b)	[HM Prison] Thornton et al. (2003)	73	Thornton (2002)
31.99	[HM Prison]	74	Thornton & Knight (2006)
32	Hood et al. (2002)	75	Tough (2001)
33	Hudson (2003)	76	Wieand (1983)
34.0	[Kia Marama] Hudson et al. (2002)	77	Williams & Nicholaichuk (2001)

Table continues

Appendix 1: Table 4 continues

Number	Study	Number	Study
34.1	[Kia Marama] Beggs & Grace (2005)	78.0	Wilson & Prinzo (2001)
34.99	[Kia Marama]	78.1 (a, b)	Wilson et al. (2005)
35	Kolko (2005)	79	Wormith & Ruhl (1986)

Notes: Studies with the same number but different decimal points are different reports from the same sample. When a study number has a & b in parentheses, two distinct subsamples were drawn from the same report. Study numbers ending in .99 are averaged results over more than one report on the same sample.