Prediction of Adolescent Sexual Reoffending: A Meta-Analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99

Jodi L. Viljoen Sarah Mordell Jennifer Beneteau
Simon Fraser University

Law and Human Behavior

DOI: https://doi.org /10.1037/h0093938

Author Note

The authors thank our colleagues, the authors of these tools, and the international associations who graciously responded to our requests regarding conference papers and unpublished data. We also thank Drs. Constantine Gatsonis, David B. Wilson, Michael Borenstein, and Laura Guy for their assistance regarding meta-analyses of AUC scores, Dr. Karl Hanson for his comments regarding the Static-99, and our anonymous reviewers for their helpful suggestions. The first author gratefully acknowledges support from the Michael Smith Foundation for Health Research, Career Scholar Award.

Abstract

Several risk assessment tools, including the Juvenile Sex Offender Assessment Protocol-II (Prentky & Righthand, 2003), the Estimate of Risk of Adolescent Sexual Offense Recidivism (Worling & Curwen, 2001), the Juvenile Sexual Offense Recidivism Risk Assessment Tool-II (Epperson, Ralston, Fowers, DeWitt, & Gore, 2006), and the Static-99 (Hanson & Thornton, 1999), have been used to assess reoffense risk among adolescents who have committed sexual offenses. Given that research on these tools has yielded somewhat mixed results, we empirically synthesized 33 published and unpublished studies involving 6,196 male adolescents who had committed a sexual offense. We conducted two separate meta-analyses, first with correlations and then with AUCs. Total scores on each of the tools significantly predicted sexual reoffending, with aggregated correlations ranging from .12 – .20 and aggregated area under the receiver operating characteristic curve (AUC) scores ranging from .64 – .67. In many cases, however, heterogeneity across studies was moderate to high. There were no significant differences between tools, and although the Static-99 was developed for adults, it achieved similar results as the adolescent tools. To help interpret these findings, results are compared to other meta-analyses of risk tools used in the area of violence risk assessment and in other fields.

Prediction of Sexual Reoffending in Adolescents: A Meta-Analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99

Since the 1990s, a number of risk assessment tools have been developed to aid in the prediction and prevention of sexual violence (Hanson & Morton-Bourgon, 2009). These tools compile empirically-supported risk factors so that clinicians and other professionals who work with sexually abusive populations can systematically determine risk levels and manage risk, through means such as treatment and supervision. Most risk assessment tools for sexual offending were developed for use with adult sex offenders, and over the past several decades, a large body of research has examined the ability of these tools to predict reoffending (Hanson & Morton-Bourgon, 2009). In meta-analyses, the overall effect sizes for the most common of these tools, such as the Static-99 (Hanson & Thornton, 1999) and the Sexual Violence Risk—20 (Boer, Hart, Kropp, & Webster, 1997), have fallen in the moderate range (Guy, 2008; Hanson & Morton-Bourgon, 2009). In addition, these measures have gained widespread use (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Jackson & Hess, 2007; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010).

In contrast to the large number of studies on adult sex offending, knowledge regarding risk assessment approaches for sexually abusive adolescents is much less advanced. Whereas risk assessment tools for adult sex offenders were first developed in the 1990s (Borum 1996; Quinsey, Rice, & Harris, 1995), risk assessment tools for adolescents did not develop until the early 2000s (e.g., Prentky, Harris, Frizzell, & Righthand, 2000). In addition, although over 100 studies have examined risk assessment tools for adult sexual offenders (Hanson & Morton-Bourgon, 2009), only approximately 10 published studies had examined the predictive validity of risk assessment tools with sexually abusive adolescents as of 2010.

Several factors may contribute to this lag. First, the literature on risk factors for adolescent sexual offending is relatively scarce (McCann & Lussier, 2008), thus leading to challenges in the development of risk assessment tools for this population. In addition, in the past, adolescent sexual behavior problems were sometimes overlooked or dismissed as sexual experimentation (Chaffin et al., 2008; Trivits & Reppucci, 2002). As such, the risks and needs of this population were ignored. However, with the shift to more restrictive sanctions for sexually abusive adolescents, including lifetime placement on sexual offender registries, this population has increasingly come to the attention of researchers, clinicians, and policymakers (Vitacco, Viljoen, & Petrila, 2009). Clinical sites have increasingly adopted these risk assessment tools (McGrath et al., 2010), and a recent surge of research has examined their predictive validity.

As a starting point for the development of adolescent risk assessment approaches, some researchers have tested the predictive validity of common *adult* tools, such as the Static-99, among adolescents. Although a number of studies provide support for the use of the Static-99 with adult sexual offenders (e.g., Ducro & Pham, 2006; Hanson & Morton-Bourgon, 2009; Langton, Barbaree, Seto, Peacock, Harkins, & Hansen, 2007), efforts to apply this tool to adolescents have met with mixed success, with some studies reporting significant findings (Beech, Thornton, Tudway, Parish, & Print, 2004; Poole, Liedecke, & Marbibi, 2000) and other studies reporting null results (Morton, 2003; Viljoen, Elkovitch, Scalora, & Ullman, 2009). In addition, the use of the Static-99 with individuals who offended as juveniles has generated

controversy and legal challenges in court settings (*In re Anderson*, 2006; *In re Fox, Jones, & Jacka*, 2007; *In re J.P.*, 2001; *In re Sandry*, 2006; *R. v. R. (M.L.)*, 2002).

In contrast to applying adult measures to adolescents, a number of researchers have emphasized the need for tools designed specifically for adolescents, especially as risk factors for adolescent and adult sexual offending may differ somewhat (Caldwell, 2002; Miner, 2002; Prescott, 2004; Worling & Långström, 2006). This has led to the creation of several tools designed specifically for adolescents. The Juvenile Sex Offender Assessment Protocol II (J-SOAP-II; Prentky & Righthand, 2003), the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2001), and the Juvenile Sexual Offense Recidivism Risk Assessment Tool—II (J-SORRAT-II; Epperson, Ralston, Fowers, DeWitt, & Gore, 2006) are the most common of these measures (McGrath et al., 2010; Viljoen, McLachlan, & Vincent, 2010).

The J-SOAP-II, ERASOR, and J-SORRAT-II differ in a number of ways (see Table 1). First, they differ with respect to their intended purpose (Epperson et al., 2006; Prentky & Righthand, 2003; Worling & Curwen, 2001). Although the J-SOAP-II aims to predict both sexual and non-sexual reoffending among sexually abusive adolescents, the ERASOR's stated purpose is to predict sexual reoffending only. Second, while the J-SOAP-II and the J-SORRAT-II focus on numerical summary scores, the ERASOR was developed based on a structured professional judgment model, whereby raters can make their own structured judgment of low, moderate, and high risk. This structured professional judgment allows raters to consider additional factors that may not be captured by the items (e.g., stated plans to reoffend). Third, these tools differ with respect to the number and types of risk factors they include (Epperson et al., 2006; Prentky & Righthand, 2003; Worling & Curwen, 2001). While the J-SORRAT-II is a brief 12-item actuarial tool that focuses on static or historical risk factors (e.g., number of adjudications as a sex offender and history of special education), the J-SOAP-II and the ERASOR are longer and include dynamic or potentially modifiable risk factors. Finally, the tools differ in terms of their structure. The items on the J-SOAP-II are arranged into four subscales (Sexual Drive and Preoccupation, Impulsive/Antisocial Behaviour, Intervention, and Community Stability and Adjustment). In contrast, the ERASOR is broken down into separate sections (History of Sexual Assaults, Sexual Interests and Behaviors, Psychosocial Functioning, Family/Environment, and Treatment), but these sections were not developed as scales per se (Worling, Bookalam, & Litteljohn, in press), and the J-SORRAT-II does not have subscales.

Studies on the predictive validity of the J-SOAP-II, ERASOR, and J-SORRAT-II have varied in their conclusions. Although a number of studies have found that J-SOAP-II total scores predict sexual reoffending (Martinez, Flores, & Rosenfeld, 2007; Prentky, Li, Righthand, Schuler, Cavanaugh, & Lee, 2010; Rajlic & Gretton, 2010), other studies have reported non-significant results (Caldwell, Ziemke, & Vitacco, 2008; McCoy, 2008; Parks & Bard, 2006; Viljoen et al., 2008). The initial development study on the J-SORRAT-II reported very strong findings (Epperson et al., 2006), but the areas under the receiver operating characteristic curve (AUCs) for total scores were not as strong in several subsequent studies (Viljoen et al., 2008; Ralston, 2008). Similarly, several studies reported that the ERASOR significantly predicted sexual reoffending (Rajlic & Gretton, 2010; Worling et al., in press), whereas other research reported non-significant AUCs (Morton, 2003).

A number of explanations may contribute to these equivocal findings. First, these results could suggest that it is somewhat challenging to predict adolescent sexual offending, possibly due to limitations in our knowledge about risk factors for adolescent sexual offending. For instance, in their meta-analysis of risk factors for adolescent sexual reoffending, McCann and Lussier (2008) found that even the strongest risk factors had effect sizes that were relatively small, emphasizing that much remains unknown regarding predictors of adolescent sexual reoffending.

Second, some studies may have insufficient sample sizes to detect significant effects, particularly if the effects are small. Third, moderators may contribute to these mixed findings. For instance, these tools might be more effective with certain populations or in certain settings. Viljoen et al. (2008) reported that AUCs on the J-SOAP-II were higher for older adolescents than for younger adolescents. Also, Rajlic and Gretton (2010) found that the ERASOR and J-SOAP-II had higher levels of predictive validity amongst adolescents who had committed only sexual offenses than amongst those who also had a history of non-sexual offending. Methodological factors, such as study design, publication bias, and allegiance, might also moderate predictive validity (Blair, Marcus, & Boccaccini, 2008; Hanson & Morton-Bourgon, 2009). Finally, these mixed results may reflect simple random variation in study findings.

Therefore, to empirically synthesize findings on the predictive validity of these tools and test potential moderators, we conducted a meta-analysis. While some useful qualitative reviews have been conducted (Vitacco, Caldwell, Ryba, Malesky, & Kurus, 2009), meta-analyses offer greater rigor and statistical power, and an opportunity to empirically test potential moderators (Egger & Smith, 1997; Ioannidis & Lau, 1999). We focused on the J-SOAP-II, ERASOR, and the J-SORRAT-II because they are the most widely used risk tools for sexually abusive adolescents (McGrath et al., 2010; Viljoen et al., 2010). We also examined the predictive validity of the Static-99 with adolescents because understanding how this tool performs with adolescents may advance knowledge regarding developmentally-appropriate approaches.

Although a number of meta-analyses have examined the predictive validity of adult sex offender risk assessment tools (Hanson & Morton-Bourgon, 2009) or more general adolescent risk assessment tools (Olver, Stockdale, & Wormith, 2009; Schwalbe, 2007), such as the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006), to our knowledge no prior meta-analyses have examined the predictive validity of tools designed for sexually abusive youth.

Method

Sample

The current meta-analysis captured 33 studies based on 31 separate samples, including 13 published studies and 20 unpublished or in-press studies. Included studies are marked with an asterisk in the reference list. In total, 15 studies were conducted on the J-SOAP-II, 11 on the ERASOR, 7 on the J-SORRAT-II, and 8 on the Static-99. Approximately one-quarter of studies (n = 8, 24.2%) examined multiple tools.

Overall, these studies involved 6,196 adolescents who had committed sexual offenses. The mean age of adolescents in these studies was approximately 16 years old (see Table 1 for

descriptive information on studies). Most studies focused exclusively on male adolescents, although a few did not explicitly mention gender. The majority of the samples included predominantly Caucasian youth. Adolescents' index offenses ranged considerably and included penetrative and non-penetrative offenses. The mean follow-up periods, during which reoffending was examined, ranged from 12 months to 120 months, with a median of approximately 71 months (6 years). In most cases, reoffending was measured through justice records and almost all studies relied on a pseudo-prospective design (also called a retrospective follow-up design) in which historical file information was coded and reoffense records were then obtained. During the follow-up periods, an average of 10.9% of youth sexually reoffended, whereas an average of 49.4% committed general offenses.

Procedures

In conducting this meta-analysis, we followed the PRISMA Statement for meta-analyses in health care (Liberati et al., 2009; Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009). The PRISMA statement was developed to provide enhanced reporting and rigor in meta-analyses, and includes a set of 27 items to include in a meta-analysis (e.g., "number of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage"; Liberati et al., 2009, p. 3).

Identification of Studies: A comprehensive search procedure was used to identify published and unpublished studies (see Figure 1). First, we searched names of each of the risk assessment tools and its abbreviations in 10 academic databases (PsycINFO, PsycARTICLES, PsycBOOKS, National Criminal Justice Reference Service, MEDLINE, Criminal Justice Abstracts, Sociological Abstracts, Social Services Abstracts, Health Source: Nursing/Academic Edition, and Web of Science).[1] Second, we searched three additional databases to identify unpublished dissertations (ProQuest Dissertations & Theses database, which was previously called the Digital Dissertation database, the Universal Index of Doctoral Dissertations in progress, and the Networked Digital Library of Theses and Dissertations). Third, to capture additional unpublished studies, we conducted a Google Scholar search.

Fourth, we requested conference programs from 10 organizations which include a focus on juvenile offender populations (American Academy of Forensic Sciences; American Psychology-Law Society Conference; American Association for the Treatment of Sexual Abusers; Australian and New Zealand Association of Psychiatry, Psychology, and Law; European Association of Psychology and Law; International Association of Forensic Mental Health Services; International Congress of Law and Mental Health; National Adolescent Perpetrator Network; Nordic Network for Research on Psychology and Law). We requested conference programs from the year 2000 (at which time these tools were first being developed) to 2011. Overall, 75% of the conference programs that we requested were available for review. Fifth, we contacted 34 experts, including the authors of these tools and other researchers in this area. Finally, we examined studies that were included in the meta-analysis in order to determine if they mentioned any other studies that could meet inclusion criteria.

Eligibility Screening: Based on the above procedures, we identified 1,164 relevant abstracts, which we reviewed to determine whether they met inclusion criteria. First, to be included, studies had to comprise a sample of adolescents who had committed sexual offenses. Sexual offenses were defined as illegal sexual acts committed by adolescents that resulted in

arrests, charges, convictions, and/or referral to a treatment program. Second, the majority of study participants had to fall in the age range of 12 to 18 years old. Where the range was not reported, studies were required to have a mean age falling in this range. Third, studies had to include empirical data for at least one of the risk assessment tools. Based on these criteria, 52 studies were deemed eligible to be included in the full-text review (see Figure 1). To examine interrater agreement for determinations of whether a study met eligibility criteria (Yeaton & Wortman, 1993[2]), each of the three co-authors blindly coded 25 cases. The interrater agreement rate for eligibility screening was 100%.

Full-Text Review for Inclusion in Meta-Analysis: We next reviewed the full text articles of studies that were screened in to determine whether they could be included in the meta-analysis. Of those screened in, 33 were included. Nine cases were excluded because they were identical or overlapping with other studies, such as when a conference presentation was later published as an article (Aebi & Bessler, 2010; Dahle, Janka, Gallasch-Nemitz, & Lehmann, 2008; Elkovitch, Viljoen, Scalora, & Ullman, 2008; Martinez, Rosenfeld, & Flores, 2004; Parks, 2004; Prentky, 2006; Prentky, Pimental, Cavanaugh, & Righthand, 2009; Rajlic & Gretton, 2010; Ralston & Epperson, 2007; Viljoen, Elkovitch, Bader, Scalora, & Ullman, 2008). In these instances, only the most recent and/or comprehensive version of the study was included. Six studies were excluded because they did not examine whether the tool predicted reoffending (Chavez, 2010; Costin, 2005; Fanniff & Letourneau, 2011; Rombouts, 2006; Schoenfeld, 2008; White, Cruise, & Frick, 2009). Two studies were excluded because they examined single items or unique combinations of items rather than total or subscale scores (Edwards, Beech, Bishopp, Erikson, Friendship, & Charlesworth, 2005; Young, 2006). One study was excluded because it used a case study methodology and relied primarily on qualitative data for 3 participants (Fleming, 2004), and one study was excluded because it focused on adults (Soothill, Harman, & Kirby, 2005).[3]

Data Extraction: Once studies were selected for inclusion, we coded effect sizes and potential moderator (e.g., setting of study, country, sample size, base rate of reoffending). Interrater agreement for the coding of *all* effect sizes and moderators was calculated by comparing ratings of the first author to those of the second and third authors (Yeaton & Wortman, 1993). The agreement rate was 97.5% for effect sizes and 97.1% for moderators. Discrepancies were resolved through a consensus procedure prior to proceeding with analyses.

Data Analyses

AUCs have come to be widely accepted in the risk assessment literature, as they are not as affected by base rates as correlations (Mossman, 1994; Rice & Harris, 1995, 2005). Despite this, most meta-analyses of risk assessment tools have relied on correlations or *d*-scores (Guy, 2008). Procedures for conducting meta-analyses of AUCs are not as well-established (Giles & Rothwell, 2010; Kester & Buntinx, 2000) and some questions have been raised regarding the viability of meta-analyses of AUCs (Singh et al., 2011). However, meta-analyses involving AUCs are becoming increasingly common in medicine (e.g., Giles & Rothwell, 2010; Parolari et al., 2010; Timmermans et al., 2010) and are similar to meta-analysis of other summary statistics (C. Gatsonis, Cochrane Center, personal communication, May 17, 2011; Kester & Buntinx, 2000; McClish, 1992).

As such, to compensate for the strengths and limitations of both approaches (i.e., meta-

analyses of correlations and AUCs), we conducted two separate meta-analyses, one using correlations and the other using AUC scores. In our analyses, we examined the ability of tools to predict both sexual reoffending and general reoffending (defined as any or nonsexual reoffending). In several studies, predictive validity for two separate time periods was presented (e.g., Caldwell, 2011a; Caldwell & Dickinson, 2009; Epperson et al., 2006; Worling et al., in press). In these cases, the longer time period was selected for the primary analyses as it was more comprehensive, but the multiple time periods were compared in moderator analyses.

Aggregated Correlations: We aggregated correlations using the statistical package Comprehensive Meta-Analysis Version 2 (CMA2; Borenstein, Hedges, Higgins, & Rothstein, 2005). Given that the goal was to generalize findings beyond the set of observed studies to the population of studies, we used a random-effects model rather than a fixed-effect model (Hedges & Vevea, 1998). Random-effects models provide more conservative estimates of effect sizes because they encompass both between and within study variability (Bornstein, Hedges, Higgins, & Rothstein, 2009). In studies in which the results were presented only as AUCs rather than correlations (n = 17), we converted data to correlations using the transformation formulas provided in Rice and Harris (2005) so that these studies could be included in the analyses.[4] Several studies presented the data in forms other than correlations or AUCs (i.e., mean scores, proportion of offenders who reoffended, or chi-square). In these cases, results were converted to correlation coefficients using CMA2. Orwin's (1983) fail-safe N was calculated (using CMA2) to evaluate how robust the findings were to possible missing studies. This procedure estimates how many missing studies with a null effect would reduce the estimate of the aggregate effect size to a specified level.

To test heterogeneity of findings, within-group Q statistics (Q_w) were calculated (using CMA2). The Q statistic is distributed as a chi-square test (Hedges & Olkin, 1985). Given that the Q statistic only tests the presence or absence of heterogeneity, CMA2 also generates an I^2 index, which provides an estimate of the amount of heterogeneity (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Heterogeneity can be classified as low, medium, or high according to I^2 values of 25%, 50%, and 75%, respectively (Huedo-Medina et al., 2006). The aggregated correlations of tools were compared using between-group Q statistics (Q_B).

Aggregated AUC Scores: We aggregated AUCs using Lipsey's and Wilson's MetaES macro (random effects model). This macro is appropriate for AUC scores (Dr. David B. Wilson, personal communication, October 6, 2011), and weights studies by inverse variance, a procedure which typically yields more accurate estimates than weighting by sample size (Marín-Martínez & Sánchez-Meca, 2010). If a study did not include an estimate of standard error (SE) from which inverse variance could be calculated, we estimated SE using the formulas outlined in Hanley and McNeil (1982). In studies that did not provide AUC scores (n = 10), correlations were converted to AUCs using formulas provided in Rice and Harris (2005). We used z-tests to test whether AUCs for tools differed significantly (Hanley & McNeil, 1982, 1983).[5]

Moderators (Meta-Regression): Given that there were insufficient studies to test potential moderators by instrument, moderators were tested at an aggregate level, aggregating total scores for the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99, similar to the approach used by Hanson and Morton-Bourgon (2009). A number of studies presented numerous effect sizes and used multiple tools. Thus, to avoid erroneously treating each effect size as independent,

we ran these analyses in CMA (using aggregated correlations) because this enabled us to collapse findings across studies. For instance, if a study included 2 or more effect sizes, these effect sizes were averaged so that the same moderators (from the same study) were not counted twice.

Results

Sexual Reoffending

Aggregated Correlations: At an aggregate level, total scores on the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99 predicted sexual reoffending, as did SPJ ratings on the ERASOR (see Tables 3-5). Aggregated correlations for total scores ranged from .12 to .20, and did not significantly differ across tools. Heterogenity for J-SOAP-II, ERASOR, and Static-99 total scores was medium to high (I^2 values > 50%), as illustrated in the forest plots (Figures 2-4). Based on Orwin's (1983) fail-safe N, 11 studies with an effect size of 0 would need to be added to the analyses to drive the estimate of aggregate effect for ERASOR total score down to $r = .10$ (i.e., a small effect). There would need to be 8 such studies on the J-SOAP-II total score, 4 on the Static-99 total score, and 2 on the J-SORRAT-II total score to drive the pooled estimate to $r = .10$.

Aggregated AUCs: Aggregated AUCs for total scores ranged from .64 to .70, but there was significant heterogeneity for J-SOAP-II and J-SORRAT-II total scores. Although the rank ordering of tools from highest to lowest AUC differed slightly than that of the aggregated correlations, there were no significant differences between AUCs of the total scores for any of the tools.

Further Examination of Potentially Biasing Studies: Several additional analyses were conducted to examine if studies that were potentially biasing impacted the results. First, although the Parks and Bard (2006) study on the J-SOAP-II was excluded in the initial analyses because no effect size was reported (it was simply described as "non-significant" in the article), the results were re-run with the correlation conservatively coded as 0 (and then again, with the AUC coded as .50). The aggregated correlations and AUCs for the J-SOAP-II total score decreased somewhat but remained significant ($r = .19$ to .17, p's < .01; AUC = .67 to .66), indicating that the inclusion of this study would not have changed the findings to non-significant. Second, because the study conducted by Prentky et al. (2009) was classified as an outlier (z-score > 5.00) the results were re-run excluding this study. The aggregated correlations and AUCs for the J-SOAP-II total score decreased somewhat but the aggregated correlation remain significant ($r = .19$ to .15, p's < .001; AUC = .67 to .65). Finally, although the initial development sample for the J-SORRAT-II was included in the initial analyses described above (Epperson et al., 2006), the test development sample can lead to inflated estimates of predictive validity (Silver, Smith, & Banks, 2000). When the results were re-run with this study excluded, aggregated correlations and AUCs for the J-SORRAT-II total score decreased somewhat but the aggregated correlation remained significant ($r = .12$ both before and after, p's < .01; AUC = .64 to .61).

Moderators: We separately tested the following categorical moderators: a) setting (mental health setting or other type of setting), b) treatment sample (i.e., whether or not the sample consisted of youth in a treatment program), c) publication bias (published or

unpublished), d) allegiance effects (whether any of the study's authors was also an author of the tool under investigation, or in the case of dissertations, whether the student was supervised by an author of the tool), e) country where the study was conducted (USA or other), f) sample size (greater than or less than the median of 150 participants), g) base rate of sexual reoffending (greater than or less than 10%), h) length of follow-up (greater than or less than 5 years), i) interrater reliability of tool (strong interrater reliability vs. no information or inadequate interrater reliability).[6] After correcting for multiple comparisons using a Boneferoni correction, none of the moderators reached significance (p = .05/9 comparisons = .005).

General Reoffending

We examined effect sizes for general reoffending for the J-SOAP-II and ERASOR but did not include the J-SORRAT-II and Static-99, given the limited number of studies that have examined general reoffending with these tools (n = 2 and 3 respectively). Notably, some studies defined it to include sexual offending (n = 9) whereas other studies excluded sexual offending from the definition (n = 5).[7]

Aggregated Correlations: At an aggregate level, total scores on the J-SOAP-II and ERASOR predicted general reoffending, as did a number of scales, particularly the Impulsive/Antisocial Scale of the J-SOAP-II and the Psychosocial Functioning section of the ERASOR (see Tables 2-3). While the aggregated r for the J-SOAP-II was higher than the ERASOR for the prediction of general reoffending, this difference was not significant. Based on Orwin's (1983) fail-safe N, 11 studies with a mean effect size of 0 would need to be added to the analyses to drive the estimate of aggregate effect for J-SOAP-II total score down to r = .10, and 5 such studies on the ERASOR. Heterogeneity was high for aggregated correlations on the J-SOAP-II and ERASOR total scores (I^2 >75%).

Aggregated AUCs: Similar to the pattern of results for aggregated correlations, the aggregated AUC for the J-SOAP-II total score was higher than that of the ERASOR total score. However, this difference was not significant. There was significant heterogeneity for aggregated AUCs on the J-SOAP-II and ERASOR total scores.

Further Examination of Potentially Biasing Studies: Although the Parks and Bard (2006) study was excluded in the initial analyses because no effect size was reported (it was simply described as "non-significant"), the results were re-run with the correlation conservatively coded as 0 (and then again, with the AUC coded as .50). The aggregated correlations and AUCs for the J-SOAP-II total score decreased somewhat but the aggregated correlation remain significant (r = .25 to .17, p's < .01; AUC = .66 to .64).

Moderators: After correcting for multiple comparisons using a Boneferoni correction, none the potential moderators examined (described earlier) were significant.

Further Examination of Studies that Directly Compared Two Tools: We examined studies which compared multiple tools in greater detail because they enable a more controlled comparison of the tools under similar conditions and methodology (see Table 6). Again, the tools did not differ significantly in these analyses. These results should be interpreted with caution however given the small n.

Discussion

Concerns have been raised that it may be challenging to predict sexual reoffending in adolescents due to the tremendous developmental changes that occur during this period (Caldwell et al., 2008; Viljoen et al., 2008). Several risk assessment tools, including the J-SOAP-II, ERASOR, and J-SORRAT-II, have been developed to assist in these assessments. However, research on the predictive validity of these tools has been mixed. As such, the purpose of this meta-analysis was to empirically synthesize research on these tools.

Primary Findings

Despite a range of findings across individual studies, aggregated correlations for total scores on the ERASOR, J-SOAP-II, J-SORRAT-II, and Static-99 significantly predicted sexual reoffending. Aggregated correlations for total scores fell in the range of .12 to .20, which is typically considered to be a fairly small correlation (Cohen, 1988). Correlations may, however, underestimate predictive validity for low base rate events (Rice & Harris, 2003). As such, we also performed a separate meta-analysis using AUC scores. This yielded comparable result (though perhaps slightly more favorable) in that AUCs for total scores ranged from .64 to .67, which is generally considered to reflect moderate AUCs (Douglas, Guy, Reeves, & Weir, 2008; Rice & Harris, 2005). There were no significant differences between tools regardless of whether analyses were conducted with correlations or AUCs.

Our findings are positive in many respects. When compared to other meta-analyses, the sex offense-specific tools examined in the present meta-analysis generally appear to outperform more general tools such as the PCL:YV in the prediction of sexual reoffending, indicating that sex offense-specific tools are preferable to general tools for the assessment of sexual reoffense risk (see Table 7; Edens et al., 2007; Olver et al., 2009). Furthermore, the effect sizes for adolescent sexual risk assessment tools appear to be higher than the effect sizes for individual risk factors (e.g., stranger victim, child victims, adult victims, use of threats/weapons, and prior sexual and nonsexual offending; McCann & Lussier, 2008), suggesting the value of combining factors in making judgments.

In addition, when compared to tools that are used to assess risk in other domains, the effect sizes found in the present meta-analysis appear to be fairly similar to the effect sizes found for the prediction of self-harm risk with the Beck Hopelessness Scale (see Table 7; McMillan, Gilbody, Beresford, & Neilly, 2007). They are higher than those found for the prediction of risk of driving accidents with a popular self-report questionnaire (i.e., Driving Behaviour Questionnaire, de Winter & Dodou, 2010), although somewhat lower than those for some tools in other fields (e.g., assessment of risk of driving problems with certain cognitive tests, assessment of early stroke risk with the ABCD and ABCD2, assessment risk of mortality from pneumonia with various tools; see Table 7).

Although our results generally provide support for the tools, they also suggest that the prediction of adolescent sexual reoffending may bring some challenges. For instance, although the tools significantly predicted reoffending, they explained only a modest amount of variance in sexual reoffending (i.e., 1.4% to 4% for total scores). Also, our aggregated correlations were not as high as those reported for tools such as the SAVRY and YLS/CMI in the prediction of general reoffending (Edens et al., 2007; Olver et al., 2009; see Table 7), suggesting that it may be more

challenging to predict sexual than general reoffending.

In addition, despite the overall significant effect sizes for the risk assessment tools at an aggregate level, in many cases, there were high levels of heterogeneity across studies, meaning that studies did not find uniformly positive results. Although concerns regarding heterogeneity have arisen in some other risk assessment contexts as well (Edens et al., 2007; Olver et al., 2009), risk assessment of adolescent sexual reoffending appears to be an area that is plagued by particularly high levels of inconsistency. At the present time, it is largely unknown what might explain the mixed findings across studies, as none of the moderators we tested reached statistical significance.

We predicted that the tools that were designed specifically for adolescents would outperform the Static-99, a tool that was developed for adult sex offenders. However, this hypothesis was not supported. While it is possible that certain risk factors, such as dynamic factors, may be particularly important during adolescence, recent research indicates that a number of risk factors for *adolescent* sexual reoffending overlap with those for *adult* sex reoffending (e.g., stranger victim; McCann & Lussier, 2008; Hanson & Morton-Bourgon, 2009); this overlap in risk factors likely contributes to our finding. That said, the effect size for the Static-99 in the present meta-analysis was lower than the effect size reported for adult samples (Dahle et al., 2008, 2009; Hanson & Morton-Bourgon, 2009), which suggests that the Static-99 is better suited to adults than adolescents.

In addition, although some researchers have raised questions about whether it may be particularly challenging to assess risk of sexual reoffending in adolescents compared to adults, our results are somewhat mixed in this regard. The overall effect sizes for the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99 to be fairly comparable to the averages found for tools used in adult sex offender risk assessment (Hanson & Morton-Bourgon, 2009). However, based on the meta-analysis by Hanson and Morton-Bourgon (2009), a number of risk assessment tools have achieved higher mean effect sizes in predicting adult sexual reoffending than any of the tools examined in the present meta-analysis (e.g., SVR-20, MnSost-R, Static-2002, Static-99)[8], which could suggest that the assessment of sexual reoffense risk in adolescents may currently be more difficult.

Finally, our results may have the potential to offer some information about risk factors for adolescent sexual reoffending. Consistent with other studies (e.g., McCann & Lussier, 2008; Seto et al., 2010), our results provide some support for the hypothesis that sexual reoffending is predicted by both unique factors (e.g., sexually deviant interests) and general factors reflecting an antisocial orientation. In particular, we found that the Sexual Drive/Preoccupation scale of the J-SOAP-II predicted sexual reoffending but not general reoffending, whereas the Impulsive/Antisocial scale predicted both sexual and general reoffending.

Clinical Implications

Overall, this meta-analysis provides support for the use of the ERASOR, J-SOAP-II, J-SORRAT-II, and Static-99 in assessing sexual reoffense risk in adolescents. Risk assessment tools such as these offer clear benefits over unstructured clinical judgments (Hanson & Morton-Bourgon, 2009). However, given that the effect sizes were quite modest, these tools may be insufficient to make predictions that require a high degree of precision, such as preventative

detention (Yang, Wong, & Coid, 2010), civil commitment of adolescent sex offenders, or the placement of adolescent offenders on sex offender registries for life (see Caldwell et al., 2008).

Despite common desires to identify which tool is "best," no single tool emerged as significantly stronger than the others. Also, while Hanson and Morton-Bourgon (2009) concluded that actuarial tools outperformed structured professional judgment tools in the prediction of sexual reoffending, we did not replicate this finding. Instead, the effect size for the ERASOR SPJ did not differ significantly from that of total scores. Although the tools appear to be fairly interchangeable strictly in terms of predictive validity, they differ in other ways that affect their suitability for particular purposes. For instance, the ERASOR and J-SOAP-II measure putatively dynamic factors that can serve as treatment targets, and therefore may offer benefits when the goal is to manage or prevent reoffending.

While the Static-99 achieved a similar degree of predictive validity as adolescent tools, several factors may limit its applicability to adolescents. First, the effect size for the Static-99 in the present meta-analysis was lower than the effect size reported for adult samples (Dahle et al., 2008, 2009; Hanson & Morton-Bourgon, 2009). Second, given that the juvenile justice system places a greater focus on treatment than the adult criminal justice system, the Static-99's focus on static factors may make it less relevant to juvenile settings. Third, although the Static-99 may be able to discriminate adolescents who recidivate from those who do not, it may not be properly calibrated for juveniles (R. Karl Hanson, personal communication, July 20, 2011). In particular, because adolescents automatically receive points on two Static-99 items (i.e., young age, unmarried), the Static-99 might overestimate adolescents' risk level. Thus, if this tool were to be used with adolescents, adolescent norms would first be needed (R. Karl Hanson, personal communication, July 20, 2011).

Limitations and Future Directions

This meta-analysis used comprehensive search procedures (i.e., 14 databases and search engines, conference programs, contacts with experts, etc.) and captured a large number of unpublished studies (n = 19), although it is possible that some relevant studies were not captured by search procedures. Also, while most meta-analyses of risk tools have used correlations, in this meta-analysis, analyses were conducted with AUC scores as well correlations, in order to address the limitations of correlations.

Like any meta-analysis, however, this meta-analysis is limited by the quality of research. Nearly all of the studies included in this meta-analysis were pseudo-prospective studies in which tools were coded from file information by research assistants and reoffending was measured through official records alone (see Table 2). Also, some studies made adaptations to the tools (e.g., coding items as present or absent or coding only a single scale). Thus, more rigorous designs and methodologies are needed.

Based on our findings, several areas may be important for future research to address. First, although examinations of sex offender risk assessment tools have focused on the ability of these tools to predict reoffending, tools such as the J-SOAP-II and ERASOR are also intended to help manage risk and plan treatment in order to prevent reoffending. Increased attention to the utility of tools for these purposes will enable us to move beyond simply the *prediction* of reoffending towards the *prevention* of reoffending.

Second, given the significant heterogeneity in the effect sizes, subsequent research should carefully test potential moderators of predictive validity. Length of follow-up may be particularly important to test as a moderator, as some authors have suggested that risk assessments of youth may have a relatively quick expiration period due to developmental change (Prescott, 2004; Worling et al., in press). In the present study, it was not possible to precisely test this because studies ranged considerably in their follow-up periods rather than presenting fixed periods.

Finally, our finding that the Static-99 functioned fairly similarly to adolescent-specific tools may be viewed as surprising or potentially disappointing given the enormous investments in the development of adolescent tools. Rather than suggesting we abandon efforts to develop developmentally-appropriate approaches, our findings instead emphasize the need for further knowledge in this area. Thus far, efforts to develop risk assessment tools for adolescents who have sexually offended have proceeded in the virtual absence of research on how risk and protective factors overlap and differ across various age groups. This basic underlying knowledge may help to advance adolescent risk assessment.

References

*Aebi, M., Plattner, B., Steinhausen, H. C., & Bessler, C. (in press). Predicting sexual and nonsexual recidivism in a consecutive sample of juveniles convicted of sexual offences. *Sexual Abuse: A Journal of Research and Treatment.*

Aebi, M., & Bessler, R. (2010, February). J*uvenile sex offenders in Switzerland: Demographic factors, psychopathology, offence characteristics, and recidivism*. Poster presented at 18th European Congress of Psychiatry, Munich, Germany.

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

*Beech, A. R., Thornton, D., Tudway, J. A., Parish, R. J. H., & Print, B. (2004). *Sexual recidivism in juvenile sexual offenders*. Unpublished manuscript, University of Birmingham, UK.

Blair, P., Marcus, D., & Boccaccini, M. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice*, *15*, 346-360.

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

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis Version 2.* Engelwood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chicester, UK: Wiley.

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist*, *51*, 945-956.

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY).* Odessa FL: Psychological Assessment Resources.

*Caldwell, M. (2011a). *Unpublished data on the Static-99*.

*Caldwell, M. (2011b). *Unpublished data on the J-SORRAT-II*.

Caldwell, M. F. (2002). What we do not know about youth sexual reoffense risk. *Child Maltreatment, 7*, 291-302.

Caldwell, M. (2010). Study characteristics and recidivism base rates in juvenile sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology*, *54*, 197-212.

*Caldwell, M., & Dickinson, C. (2009). Sex offender registration and recidivism risk in juvenile sexual offenders. *Behavioral Sciences & the Law*, *27*, 941-956.

*Caldwell, M., Ziemke, M., & Vitacco, M. (2008). An examination of the Sex Offender Registration and Notification Act as applied to juveniles: Evaluating the ability to predict sexual recidivism. *Psychology, Public Policy, and the Law, 14*, 89-114.

Chaffin, M., Berliner, L., Block, R., Johnson, T.C., Friedrich, W.N., Louis, D.G., …Silovsky, J.F. (2008). Report of the ATSA Task Force on Children With Sexual Behavior Problems. *Child Maltreatment, 13*, 199–218.

Chavez, V. (2010). Is everyone rated equal? An examination of factors related to sexual risk in ethnically diverse male adolescents who have sexually offended. Unpublished doctoral dissertation, University of Nebraska – Lincoln, NE.

*Chu, C. M., Ng, K., Fong, J., & Teoh, J. (in press). Assessing youth who sexually offended: The predictive validity of the ERASOR, J-SOAP-II, and YLS/CMI in a non-Western context. *Sexual Abuse: Research and Treatment.*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Costin, D. (2005). The relationship between the victimization history and sexual offences of male adolescent sexual offenders: A focus on sexual victimization, and sexual intrusiveness and force in offences. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*, 5393.

*Dahle, K. P., Janka, C., Gallasch-Nemitz, F., & Lehmann, R. (2009). Tatcharakteristika, Rückfallrisiko und Rückfallprognose bei Sexualstraftätern vom Jugend- bis ins Seniorenalter. (Offending characteristics, risk of recidivism and prediction of reoffence in sex offenders, from adolescence to later adulthood). *Forensische Psychiatrie, Psychologie, Kriminologie, 3*, 210-220.

Dahle, K. P., Janka, C., Gallasch-Nemitz, F., & Lehmann, R. (2008). Jugendliche Sexualstraftäter: Spezialfälle der Prognoseerstellung? Ein Vergleich jugendlicher, heranwachsender und erwachsener Sexualstraftäter bezüglich Tatcharakteristika, Rückfallhäufigkeit und Prognose. *Forensiche Psychiatrie, Psychologie, Kriminologie,* 2, 213-221.

Devos, H., Akinwuntan, A., Nieuwboer, A., Truijen, S., Tant, M., & De Weerdt, W. (2011). Screening for fitness to drive after stroke: a systematic review and meta-analysis. *Neurology*, 76(8), 747-756.

de Winter, J. F., & Dodou, D. D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *Journal of Safety Research*, *41*, 463-470.

Douglas, K. S., Guy, L. S., Reeves, K. A., & Weir, J. (2008). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography* [On-line]. Available: http://kdouglas.wordpress.com/

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

Edens, J., Campbell, J., & Weir, J. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior*, *31*, 53-75.

Edwards, R., Beech, A., Bishopp, D., Erikson, M., Friendship, C., & Charlesworth, L. (2005). Predicting dropout from a residential programme for adolescent sexual abusers using pre-treatment variables and implications for recidivism. *Journal of Sexual Aggression*, *11*,

139-155.

Egger, M., & Smith, G. (1997). Meta-analysis: Potentials and promise. *British Medical Journal (Clinical Research Ed.)*, *315*, 1371-1374.

Elkovitch, N., Viljoen, J. L., Scalora, M. J., & Ullman, D. (2008). Assessing risk of reoffending in adolescents who have committed a sexual offense: the accuracy of clinical judgments after completion of risk assessment instruments. *Behavioral Sciences & the Law, 26*, 511-528.

*Epperson & D. L., & Ralston, C. A. (2009). *Validation of the Juvenile Sexual Offender Recidivism Risk Assessment Tool – II (JSORRAT-II) in Iowa*. Paper presented at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

*Epperson, D. L., Ralston, C. A., Fowers, D., DeWitt, J., & Gore, K. S. (2006). Actuarial risk assessment with juveniles who sexually offend: Development of the Juvenile Sexual Offense Recidivism Risk Assessment Tool-II (JSORRAT-II). In D. S. Prescott (Ed.), *Risk assessment of youth who have sexually abused* (pp. 118–169). Oklahoma City, OK: Wood N' Barnes.

*Fanniff, A. M. & Becker, J.V. (2007, June). Predicting recidivism in a court-referred sample of juvenile sex offenders. In J.V. Becker (Chair), *Sex Offenders IV*. Symposium conducted at the 30th International Congress on Law and Mental Health, Padua, Italy.

Fanniff, A. M. & Letourneau, E. (2011). *Unpublished data on the J-SOAP-II.*

Fleming, J. S. (2004). A pragmatic case study evaluation of the process of juvenile sexual offender treatment. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 65*, 3156.

Giles, M., & Rothwell, P. (2010). Systematic review and pooled analysis of published and unpublished validations of the ABCD and ABCD2 transient ischemic attack risk scores. *Stroke; A Journal Of Cerebral Circulation*, *41*, 667-673.

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey*. Unpublished doctoral dissertation, Simon Fraser University, Burnaby, British Columbia, Canada.

Hanson, R., & Morton-Bourgon, K. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*, 1-21.

Hanson, R. K., & Thornton, D. (1999). *Static–99: Improving actuarial risk assessments for sex offenders* (User Rep. No. 1999–02). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Harris, A., Phenix, A., Hanson, R.K., & Thornton, D. (2003). *Static-99 coding rules revised – 2003*. Ottawa, ON: Corrections Directorate, Solicitor General Canada.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*, 486-504.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in*

research findings. Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2nd ed). Thousand Oaks, CA: Sage.

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

*Hersant, J. (2007). *Risk assessment of juvenile sex offender reoffense*. Unpublished doctoral dissertation, Southern Illinois University, Carbondale.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q* statistics or I^2 index? *Psychological Methods, 11*, 193 – 206.

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

In re Anderson, 134 Wn. App. 309 (Wash. App. 2006).

In re Fox, Jones, and Jacka, 138 Wn. App. 374 (Wash. App. 2007).

In re J.P., 339 N.J. Super. 443. (App. Div. 2001).

In re Sandry, 367 Ill. App. 3d 949 (Ill. App. 2006).

Ioannidis, J., & Lau, J. (1999). Pooling research results: benefits and limitations of meta-analysis. *The Joint Commission Journal On Quality Improvement*, *25*, 462-469.

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

Kester, A. D., & Buntinx, F. (2000). Meta-analysis of ROC curves. *Medical Decision Making, 20*, 430-439.

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

*Lehman, F. (2008). *Historical factors that predict re-offense in juvenile sexual offenders*. Unpublished doctoral dissertation, Alliant University, CA.

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., …Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Plos Medicine*, *6*, e1000100?.

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

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

Mathias, J. L., & Lucas, L. K. (2009). Cognitive predictors of unsafe driving in older drivers: A

meta-analysis. *International Psychogeriatrics*, *21*, 637-653.

McCann, K., & Lussier, P. (2008). Antisociality, sexual deviance, and sexual reoffending in juvenile sex offenders: A meta-analytical investigation. *Youth Violence and Juvenile Justice*, *6*, 363-385.

McClish, D. K. (1992). Combining and comparing area estimates across studies or strata. *Medical Decision-making, 12*, 274-279.

*McCoy, W. (2008). *Predicting treatment outcome and recidivism among juvenile sex offenders: The utility of the JSOAP-II and ERASOR in an outpatient treatment program*. Unpublished doctoral dissertation, Sam Houston State University, TX.

McGrath, R., J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.

McMillan, D., Gilbody, S., Beresford, E., & Neilly, L. (2007). Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences*, *37*, 769-778.

Miner, M. (2002). Factors associated with recidivism in juveniles: An analysis of serious juvenile sex offenders. *Journal of Research in Crime and Delinquency*, *39*, 421-436.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Plos Medicine, 6*, e1000097?.

*Morton, K. E. (2003). *Psychometric properties of four risk assessment measures with male adolescent sex offenders*. Unpublished Master's Thesis, Carleton University, Ottawa, Canada.

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

*Nelson, R., Owens, T., Fuson, A., Griffin, C., & DiCataldo, F. (2011, March). *Estimating risk among juvenile sex offenders: The utility of the ERASOR and the SAVRY*. Poster presented at the American Psychology-Law Society Conference, Miami, Florida.

Olver, M. E., Stockdale, K. C., & Wormith, J. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior*, *36*, 329-353.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159.

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

*Parks, G., & Bard, D. (2006). Risk factors for adolescent sex offender recidivism: Evaluation of predictive factors and comparison of three groups based upon victim type. *Sexual Abuse: Journal of Research and Treatment*, *18*, 319-342.

Parolari, A., Pesce, L., Trezzi, M., Cavallotti, L., Kassem, S., Loardi, C., & ... Alamanni, F. (2010). EuroSCORE performance in valve surgery: a meta-analysis. *The Annals of Thoracic Surgery, 89*, 787.

*Petersen, J. (2010). *Validation of the Juvenile Sex Offender Assessment Protocol (JSOAP-II) for predicting sexual and non-sexual crimes*. Unpublished Master's thesis, Michigan State University.

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

*Powers-Sawyer, A. B., & Miner, M. H. (2009). Actuarial prediction of juvenile recidivism: The static variables of the juvenile se assessment protocol-II (J-SOAP-II). *Sexual Offender Treatment, 4*, 1-11.

Prentky. R. A. (2006). Risk management of sexually abusive youth: A follow-up study. US Department of Justice Final Report.

Prentky, R. A., 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.

Prentky, R., Pimental, A., Cavanaugh, D., & Righthand, S. (2009). Predicting risk of sexual recidivism in juveniles: Predictive validity of the J-SOAP-II. *Assessment and treatment of sex offenders: A handbook* (pp. 265-290). New York, NY US: John Wiley & Sons Ltd.

*Prentky, R., Li, N., Righthand, S., Schuler, A., Cavanaugh, D., & Lee, A. (2010). Assessing risk of sexually abusive behavior among youth in a child welfare sample. *Behavioral Sciences & the Law*, *28*, 24-45.

Prentky, R., & Righthand, S. (2003). *Juvenile Sex Offender Assessment Protocol II (J-SOAP-II) manual.* Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Prescott, D. (2004). Emerging strategies for risk assessment of sexually abusive youth: Theory, controversy, and practice. *Journal of Child Sexual Abuse*, *13*, 83-105.

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of lnterpersonal Violence, 10,* 85-105.

*Rajlic, G., & Gretton, H. (2010). An examination of two sexual recidivism risk measures in adolescent offenders: The moderating effect of offender type. *Criminal Justice and Behavior*, *37*, 1066-1085.

*Rajlic, G., Clift, R. J. W., & Gretton, H. M. (2010, May). *The predictive validity of the Juvenile Sexual Offense Recidivism Risk Assessment Tool-II (JSORRAT-II)*. Poster presented at the 10th Annual International Association of Forensic Mental Health Services Conference, Vancouver, BC, Canada.

Rajlic, G., & Gretton, H. (2010). *Predictive Validity of the J-SOAP and ERASOR in Two Types of Adolescent Sexual Offenders*. Paper presented at the American Psychology-Law Society, Vancouver, BC, Canada.

*Ralston, C. A. (2008). *Validation of the Juvenile Sexual Offense Recidivism Risk Assessment Tool-II*. Unpublished doctoral dissertation, Iowa State University, Ames, Iowa.

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

Diego, CA.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63(5), 737-748.

Rice, M., & Harris, G. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*, 615-620.

Rombouts, S. (2006). *Development of a risk assessment checklist for juvenile sexual offenders: A meta-analytic approach*. Unpublished Honours thesis, Griffith University, Australia.

R. v. R. (M.L.), WL 35377 (Ont. C. J. 2002).

Schoenfeld, T. M. (2008). *Risky business: a regional comparison of the levels of risk and service needs of sexually offending youth*. Unpublished Master's thesis, University of British Columbia – Okanagan, Kelowna, British Columbia, Canada.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior*, *31*, 449-462.

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

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior*, *37*, 965-988.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*, 499-513.

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *The International Journal of Forensic Mental Health*, *1*, 179-183.

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

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 & Psychology*, *16*, 225-247.

Timmermans, A., Opmeer, B., Khan, K., Bachmann, L., Epstein, E., Clark, T., ... Mol, B. (2010). Endometrial thickness measurement for detecting endometrial cancer in women with postmenopausal bleeding: a systematic review and meta-analysis. *Obstetrics and Gynecology, 116*, 160-167.

Trivits, L., & Reppucci, N. (2002). Application of Megan's law to juveniles. *American Psychologist*, *57*, 690-704.

Viljoen, J., McLachlan, K., & Vincent, G. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, *17*, 377-395.

*Viljoen, J., Elkovitch, N., Scalora, M., & Ullman, D. (2009). Assessment of reoffense risk in adolescents who have committed sexual offenses: Predictive validity of the ERASOR, PCL:YV, YLS/CMI, and Static-99. *Criminal Justice and Behavior*, *36*, 981-1000.

Viljoen, J. L., Elkovitch, N., Bader, S., Scalora, M., & Ullman, D. (2008, March). *The Predictive Validity of the ERASOR, PCL:YV, and YLS/CMI Among Adolescents Who Have Sexually Offended*. Paper presented at the Annual Conference of the American Psychology-Law Society, Jacksonville, FL.

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

Vitacco, M., Caldwell, M., Ryba, N., Malesky, A., & Kurus, S. (2009). Assessing risk in adolescent sexual offenders: Recommendations for clinical practice. *Behavioral Sciences & the Law*, *27*, 929-940.

Vitacco, M., Viljoen, J., & Petrila, J. (2009). Introduction to this issue: Adolescent sexual offending. *Behavioral Sciences & the Law*, *27*, 857-861.

*Waite, D., Keller, A., McGarvey, E. L., Wieckowski, E., Pinkerton, R., & Brown, G. L. (2005). Juvenile sex offender re-arrest rates for sexual, violent nonsexual and property crimes: A 10-year follow-up. *Sexual Abuse: Journal of Research and Treatment*, *17*, 313-331.

White, S. F., Cruise, K. R., & Frick, P. J. (2009). Differential correlates to self-report and parent-report of callous-unemotional traits in a sample of juvenile sexual offenders. *Behavioral Sciences & the Law*, *27*, 910-928.

*Williams, J. D. (2007). Juvenile sex offenders: Predictors of recidivism. Unpublished doctoral dissertation, Auburn University, Auburn, Alabama.

*Worling, J. R. (2004). The Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR): Preliminary psychometric data. *Sexual Abuse: A Journal of Research and Treatment*, *16*, 235-254.

*Worling, J. R. (2011). *Unpublished data on the J-SORRAT-II and Static-99.*

*Worling, J. R., Bookalam, D., & Litteljohn, A. (in press). Prospective validity of the Estimate of Adolescent Sexual Offense Recidivism (ERASOR). *Sexual Abuse: Research and Treatment.*

Worling, J. R., & Curwen, T. (2001). *Estimate of Risk of Adolescent Sexual Offense Recidivism, Version 2.0.* Toronto, ON: Ontario Ministry of Community and Social Services.

Worling, J., & Långström, N. (2006). Risk of sexual recidivism in adolescents who offend sexually: Correlates and assessment. *The juvenile sex offender (2nd ed.)* (pp. 219-247). New York: Guilford.

Yang, M., Wong, S. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767.

Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review*, *17*, 292-309.

Young, S. A. (2006). Psychopathy as a predictor of recidivism among adolescent male sexual offenders treated in a prison setting. Unpublished doctoral dissertation, Chicago School of Professional Psychology.

Table 1

Characteristics of Risk Assessment Tools

Tool	**Purpose**	**Model and Scoring**	**Items and Organization**
J-SOAP-II (Prentky & Righthand, 2003)	Prediction of sexual and nonsexual reoffending in male adolescents (aged 12-18) with a history of sexual offenses or sexually coercive behavior.	Developed based on a systematic review of the literature. Scoring focuses on numerical scores but actuarial cut-offs have not yet been developed.	Consists of 28 risk factors which are organized into 4 subscales (sexual drive & preoccupation, impulsive/antisocial behaviour, intervention, & community stability & adjustment). The first two subscales focus on static factors whereas the other two focus on dynamic risk factors.
J-SORRAT-II (Epperson et al. 2006)	Prediction of sexual reoffending in juvenile sexual offenders, particularly sexual recidivism as juveniles.	Developed based on an actuarial approach.	Consists of 12 static or historical items.
ERASOR (Worling & Curwen, 2001)	Prediction of sexual reoffending in adolescents (aged 12-18) who have committed a previous sexual assault.	Developed based on a systematic review of the literature. Based on a structured professional judgment model whereby raters make a structured judgment of risk level after completing the tool .	Consists of 25 items that are organized into 5 sections (history of sexual assaults, sexual interests and behaviors, psychosocial functioning, family/environment, and treatment). All of the sections except for history of sexual assaults focus on factors that are dynamic.
Static-99 (Hanson & Thornton, 1999)	Prediction of sexual and nonsexual reoffending in adult sex offenders. Harris et al. (2003) note that it may be appropriate for youth in some instances.	Developed based on an actuarial approach with adult sex offenders.	Consists of 10 static items.

Table 2

Descriptive Characteristics of Included Studies

Study	*n*	Country	Male (%)	White (%)	Mean Age	Mean Follow Up (months)	Definition of Sexual Reoffense	Design	Setting	Treatment	Base rate of Recidivism (%)
Aebi et al, in press	223	Switzerland	100	-	15.7	51.6	Charges[3]	Pseudo-Prospective	All youths convicted by courts	No	a) 3.1 b) 44.8
Beech et al., 2004	77	UK	100	-	14.2	-	Convictions[3]	Unclear[4]	Community mental health services	No	a) 9.0 b) –
Caldwell, 2011a (Static-99)	172 (same sample as Caldwell & Dickinson, 2009)	USA	100	58.7	17.9	49.2	Charges[3]	Pseudo-Prospective	Correctional centre	No	a) 12.2 b) 59.3
Caldwell, 2011b (J-SORRAT-II)	91 (same sample as Caldwell et al., 2008)	USA	100	52	15.4	71.6	Charges[3]	Pseudo-Prospective	Correctional treatment program	Yes	a) 12.1 b) 69.0
Caldwell & Dickinson, 2009	172	USA	100	58.7	17.9	49.2	Charges[3]	Pseudo-Prospective	Correctional centre	No	a) 12.2 b) 59.3
Caldwell, Ziemke, & Vitacco, 2008	91	USA	100	52	15.4	71.6	Charges[3]	Pseudo-Prospective	Correctional treatment program	Yes	a) 12.1 b) 69.0
Chu et al., in press	104	Singapore	100	< 2.0	15.2	54.6	Official records[3]	Pseudo-Prospective	Community mental health services	No	a) 7.7 b) – c) 26.0

Dahle et al. 2009	273	Germany	100	-	-	99.6	Official records[3]	Pseudo-Prospective[4]	All youths adjudicated by courts	No	a) 17.0 b) -
Epperson et al., 2006	636	USA	100	76.4	15.2	-	Charges[2]	Pseudo-Prospective	All youths adjudicated by courts	No	a) 19.8 (juvenile or adult) b) -
Epperson & Ralston, 2009	318	USA	-	-	-	-	Charges[1]	Pseudo-Prospective[4]	All youths adjudicated by courts	No	a) 7.2 b) -
Fanniff & Becker, 2007	277	USA	100	52	13.7	35.2	Official records[3]	Pseudo-Prospective[4]	All youths adjudicated by courts	No	a) 9.6 b) 73.0
Hersant, 2007	91	USA	100	63	15	-	Caught and sanctioned by adult[3]	Retrospective	3 residential treatment centres	Yes	a) – b) –
Lehman, 2008	97	USA	100	51.1	14.5	102.6	Arrest[2]	Pseudo-Prospective[4]	Community treatment program	Yes	a) 17.5 b) 60.5
Martinez, Flores, & Rosenfeld, 2007	60	USA	100	16.7	14.9	-	Arrests, self report, other informant[3]	Pseudo-Prospective	Community treatment program	Yes	a) 13.3 b) 20.0
McCoy, 2008	128	USA	100	35.9	15.3	-	Arrests[2]	Pseudo-Prospective	Community treatment program	Yes	a) 5.6 b) 57.8 c) 56.0
Morton, 2003	80	CAN	100	-	15.2	68.1	Charges	Pseudo-Prospective	Community mental health services	No	a) 16.9 b) 51.9
Nelson et al., 2011	93	USA	-	55	15.6	80.4	Official records[3]	Pseudo-Prospective	Referred for risk evaluation	No	a) 10.8 b) – c) 62.4

Parks & Bard, 2006	156	USA	100	62.8	16.1	-	Convictions[2]	Pseudo-Prospective	Correctional centre	Yes	a) 6.4 b) – c) 30.1
Petersen, 2010	129	USA	100	-	13.7	12	Charges[3]	Pseudo-Prospective	Treatment program	Yes	a) 8.8 b) 23.2 c) 17.6
Poole, Liedecke, & Marbibi, 2000	49	USA	100	-	18.5	48	Arrests[3]	Pseudo-Prospective	Correctional centre	No	a) 8.2 b) –
Powers-Sawyer & Miner, 2009	96	USA	100	72	17.2	51.5	Arrests, Parole violations[2]	Pseudo-Prospective	Correctional treatment program	Yes	a) 7.3 b) –
Prentky et al., 2010	223 (adolescent sample)	USA	100	62.3	14.3	84	Any sexually abusive "hands-on" behaviour[3]	Pseudo-Prospective[4]	Referred for sexual risk evaluation by social services	No	a) 13.9 b) –
Rajlic, Clift, & Gretton, 2010	198	CAN	-	65.3	15.8	84	Charges[3]	Pseudo-Prospective[4]	Community treatment program	Yes	a) 9.6 b) 42.9 c) 33.3
Rajlic & Gretton, 2010	268	CAN	100	66	15.8	79.2	Charges[3]	Pseudo-Prospective	Community treatment program	Yes	a) 9.4 b) 43.4 c) 33.9
Ralston, 2008	566	USA	100	76	15	120	Charges[1]	Pseudo-Prospective	All youths adjudicated by courts	No	a) 12.4 b) –
Skowron, 2004	110	CAN	100	-	14.2	47.3	Arrests[2]	Pseudo-Prospective	Community mental health services	No	a) 35.0 b) 62.0

Viljoen et al., 2008	169	USA	100	83.4	15.4	78.9	Charges[2]	Pseudo-Prospective	Residential treatment program	Yes	a) 8.3 b) 42.8
Viljoen, Elkovitch, Scalora, & Ullman, 2009	193	USA	100	82.9	15.3	86.9	Charges[1]	Pseudo-Prospective	Residential treatment program	Yes	a) 8.3 b) 42.0
Waite et al., 2005	256	USA	100	43.5	16.8	61.9	Arrests[2]	Pseudo-Prospective	Correctional treatment program	Yes	a) 4.7 b) –
Williams, 2007	661	USA	100	43.6	15	-	Charges[3]	Unclear[4]	Correctional treatment program	Yes	a) 5.3 b) –
Worling, 2004	136	USA, CAN	100	-	14.9	-	Caught and sanctioned by any adult[1]	Retro-spective	Community mental health, and residential treatment	Yes	a) – b) –
Worling, 2011 (J-SORRAT-II, Static-99)	75	CAN	-	-	-	90	Charges, reports to agencies[2]	Pseudo-Prospective	Community mental health services	No	a) 8.0 b) –
Worling et al., in press	191	CAN	100	-	15.3	43.9	Charges, reports to agencies[2]	Prospective	Community mental health services	No	a) 9.4 b) –

Note. [1] Included juvenile records only, [2] Included juvenile and adult records, [3] Did not specify whether juvenile or adult records were included, [4] Did not specify whether coders were blind to recidivism status; a) Any Sexual Recidivism, b) Any Recidivism, c) Nonsexual Recidivism.

Table 3

J-SOAP-II Total Scores and Scales: Relationship to Sexual and General Reoffending

		Correlations					AUCs			
	k	r_w	95% Confidence Interval		Heterogeneity		AUC_w	95% Confidence Interval		Heterogeneity
					Q	I^2				*Q*
Total Score										
Sexual Reoffending	9	.19***	.09	.28	25.90**	69.11	.67	.59	.75	18.17*
General Reoffending	7	.25**	.11	.37	30.67***	80.44	.66	.57	.75	34.98***
Sex Drive/ Preoccupation										
Sexual Reoffending	13	.12**	.03	.20	43.13***	72.17	.61	.53	.69	39.99***
General Reoffending	10	-.01	-.08	.05	15.42	41.63	.49	.45	.53	15.78
Impulsive/ Antisocial										
Sexual Reoffending	11	.14***	.08	.19	13.80	27.51	.63	.58	.69	12.86
General Reoffending	9	.26***	.17	.35	25.09***	68.12	.66	.60	.72	29.30***
Intervention										
Sexual Reoffending	9	.09*	.02	.16	12.80	37.50	.60	.54	.66	8.99
General Reoffending	7	.17*	.04	.29	26.85***	77.66	.60	.52	.69	28.67***
Community Stability and Adjustment										
Sexual Reoffending	8	.19***	.07	.30	32.11***	78.20	.70	.60	.80	23.28**
General Reoffending	7	.21**	.05	.35	36.39***	83.51	.65	.57	.73	29.09***

Note. For r_w and *Q*, *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 4

ERASOR Total and Section Scores: Relationship to Sexual and General Reoffending

		Correlations					AUCs			
	k	r_w	95% Confidence Interval		Heterogeneity		AUC_w	95% Confidence Interval		Heterogeneity
					Q	I^2				*Q*
Total Score										
Sexual Reoffending	10	.20***	.12	.28	20.58*	56.28	.66	.61	.72	15.37
General Reoffending	7	.14*	.00	.27	25.74***	76.69	.59	.50	.67	32.61***
Structured Profl Judgment										
Sexual Reoffending	9	.21***	.13	.29	15.99*	49.97	.66	.60	.71	15.05
General Reoffending	6	.16*	.01	.31	22.74***	78.01	.59	.51	.68	23.56***
Sexual Interests, Drive & Preoccupation										
Sexual Reoffending	7	.08	-.04	.20	19.34**	68.98	.55	.45	.66	23.02***
General Reoffending	7	.13**	.06	.21	8.25	27.31	.58	.54	.63	10.18
Historical Sexual Assaults										
Sexual Reoffending	8	.11	-.01	.21	22.42**	68.77	.58	.48	.68	29.42***
General Reoffending	8	.00	-.07	.08	9.40	25.55	.50	.46	.55	13.12
Psychosocial Functioning										
Sexual Reoffending	7	.13***	.06	.19	3.63	0	.61	.56	.66	4.92
General Reoffending	7	.22**	.09	.34	22.27**	73.06	.62	.55	.69	23.86***
Family/Environmental Functioning										
Sexual Reoffending	7	.11**	.04	.18	7.65	21.51	.60	.55	.66	5.75
General Reoffending	7	.10	-.01	.21	17.36**	65.43	.56	.49	.63	18.88**
Treatment										
Sexual Reoffending	7	.05	-.03	.13	8.29	27.58	.52	.46	.58	4.96
General Reoffending	7	.07	-.07	.20	26.46***	77.32	.53	.45	.61	24.96***

Note. For r_w and *Q*, *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 5

J-SORRAT-II and Static-99 Total Scores: Relationship to Sexual Reoffending

		Correlations					AUCs			
	k	r_w	95% Confidence Interval		Heterogeneity		AUC_w	95% Confidence Interval		Heterogeneity
					Q	I^2				*Q*
J-SORRAT-II Total Score	7	.12***	.06	.18	10.08	40.45	.64	.54	.74	32.48***
Static 99 Total Score	8	.18***	.10	.26	13.82	49.33	.67	.59	.74	13.28
Static 99 Risk Category	4	.22***	.05	.38	10.24*	70.71	.70	.56	.85	8.93

Note. For r_w and *Q*, *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 6

Studies Which Directly Compared Risk Assessment Tools

		Correlations		AUCs	
	k	r_w	Q_B	AUC_w	
Sexual Reoffending					
J-SOAP-II vs. ERASOR	3	.14 vs. 15	.01 (ns)	.62 vs. .68	-.73 (ns)
ERASOR vs. Static-99	2	.16 vs. .18	.04 (ns)	.56 vs. .63	-. 73 (ns)
General Reoffending					
J-SOAP-II vs. ERASOR	3	.31 vs. .18	.49 (ns)	.69 vs. .61	.63 (ns)
ERASOR vs. Static-99	2	.07 vs. .09	.02 (ns)	.54 vs. .56	-.35 (ns)

Note. For r_w, Q, and z, *** $p < .001$, ** $p < .01$, * $p < .05$.

Table 7

Comparison to Recent Meta-Analyses on Other Tools to Assess Risk: Reoffending, Suicide, Driving Problems, and Medical Outcomes

Authors and Date	Topic	Findings for Total Scores		
		r	AUC	*d*
Present meta-analysis (authors names blinded, 2011)	Prediction of adolescent reoffense risk using J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99	.12-.20 (sexual) .14-.25 (general)	.64-.67 (sexual) .59-.66 (general)	.51-.62 (sexual)[9] .33-.58 (general)
Meta-Analyses on Tools to Assess General Reoffense Risk in Adolescents				
Edens et al. (2007)	Prediction of adolescent reoffending with Psychopathy Checklist measures	.07 (sexual) .24 (general)	-	-
Olver et al. (2009)	Prediction of adolescent reoffending the SAVRY, YLS/CMI, and Psychopathy Checklist	.06 - .19 (sexual) .25 - .33 (general)	-	-
Meta-Analyses on Tools to Assess Sexual Reoffending in Adults				
Hanson & Morton-Bourgon (2009)	Prediction of sexual reoffending with actuarial, mechanical, and SPJ measures (in primarily adult samples)	-	-	Averages for actuarial, mechanical, and SPJ sex-specific tools: 46 - .67 (sexual); .26 - .52 (general). However, individual tools ranged from .33 to 1.11 (sexual).
Meta-Analyses on Tools to Assess Suicide Risk				
McMillan, Gilbody, Beresford, & Neilly (2007)	Prediction of suicide and self-harm with the Beck Hopelessness Scale	-	.70 (suicide) . 63 (non-fatal self-harm)	-
Meta-Analyses on Tools to Assess Risk for Driving Problems				
Devos et al. (2011)	Prediction of fitness to drive after stroke using numerous cognitive tests	was not calculated for all tests	-	5 tests had $d > .80$ (large effects), $p < .05$ (on-road driving exam)
de Winter & Dodou (2010)	Prediction of driving accidents using Driving Behaviour Questionnaire (self-report questionnaire)	.10 – .13 (self-reported accidents)	-	-

Mathias & Lucas (2009)	Prediction of driving performance in older adults using numerous cognitive tests	-	-	6 tests had $d > .80$ (driving problems, i.e., accidents, violations)
Meta-Analyses on Tools to Assess Risk for Poor Medical Outcomes (i.e., Stroke, Cardiac Events, Mortality)				
Chalmers et al. (2010)	Prediction of mortality in hospitalised patients with community-acquired pneumonia using the PSI, CURB65, and CRB65	-	.79-.81 (mortality)	-
Giles & Rothwell (2010)	Prediction of early stroke risk with the ABCD and ABCD2	-	.78 (stroke)	-
Parolani et al. (2010)	Prediction of outcomes of cardiac valve operations with the European System for Cardiac Operative Risk Evaluation	-	.72 - .73 (cardiac events); authors interpreted this as having "low discrimination"	-

Note. To derive this table, we conducted a broad search in PsycInfo and MedLine for meta-analyses on tools used to assess risk for future negative outcomes (search terms: meta-analysis and risk and [tool or assess*]). We included meta-analyses that a) focused on tools rather than single risk factors, b) presented data in *r*s, AUCs, or *d* scores, and c) focused on risk of future negative outcomes rather than diagnostic tests. Based on this search, we identified several areas where relevant meta-analyses have been conducted (i.e., suicide risk, driving outcomes, and negative medical outcomes). To ensure we had identified relevant studies, we then proceeded to search each of these areas in greater detail, reviewing over 2,000 study titles and abstracts. For the broad category of negative medical outcomes, we focused only on meta-analyses published in the year 2010, as an unrestricted search generated over 5,000 hits. If several meta-analyses had examined the same topic, we presented data only on the most recent and comprehensive meta-analysis.

Figure 1:

Search Strategy and Phases of Review

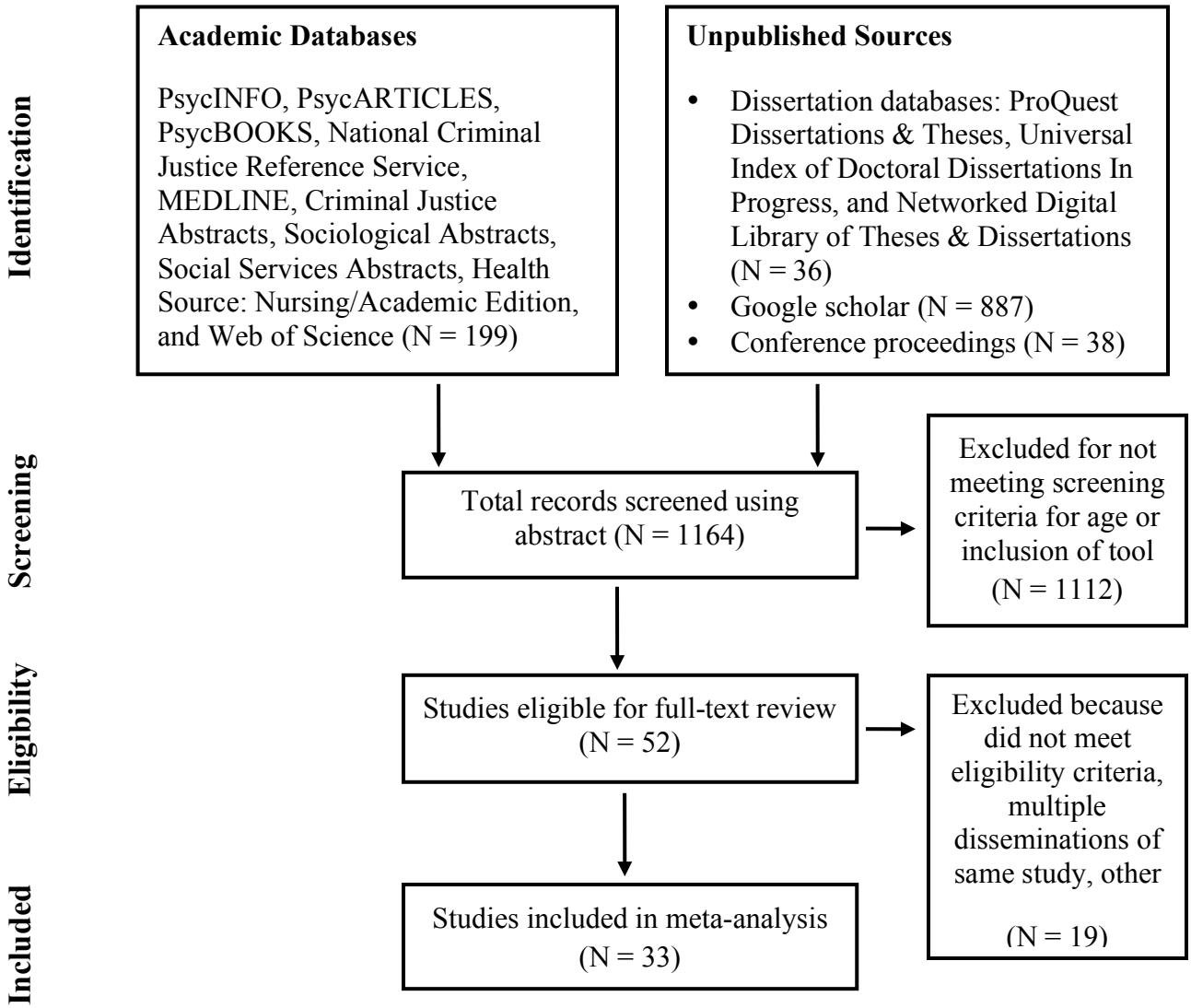


Figure 2:

Forest Plot: Correlations Between Total Scores on the J-SOAP-II and Sexual Reoffending

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Aebi et al. (in press)	0.090	-0.042	0.219	1.339	0.181
Caldwell et al. (2008)	0.082	-0.126	0.283	0.771	0.441
Chu et al (in press)	0.009	-0.184	0.201	0.090	0.928
Martinez et al. (2007)	0.310	0.061	0.523	2.420	0.016
McCoy (2008)	0.160	-0.014	0.325	1.804	0.071
Petersen (2010)	0.320	0.153	0.469	3.663	0.000
Prentky et al. (2009)	0.441	0.304	0.560	5.799	0.000
Rajlic & Gretton (2010)	0.202	0.083	0.316	3.290	0.001
Viljoen et al. (2008)	0.039	-0.113	0.189	0.503	0.615
	0.182	0.129	0.235	6.613	0.000

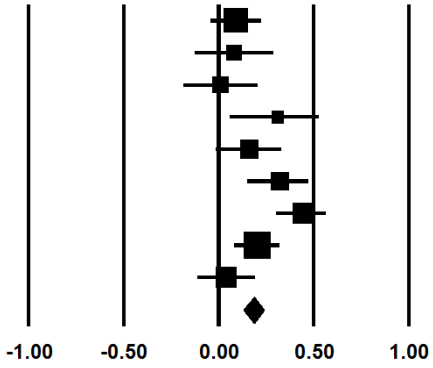


Figure 3:

Forest Plot: Correlations Between Total Scores on the ERASOR and Sexual Reoffending

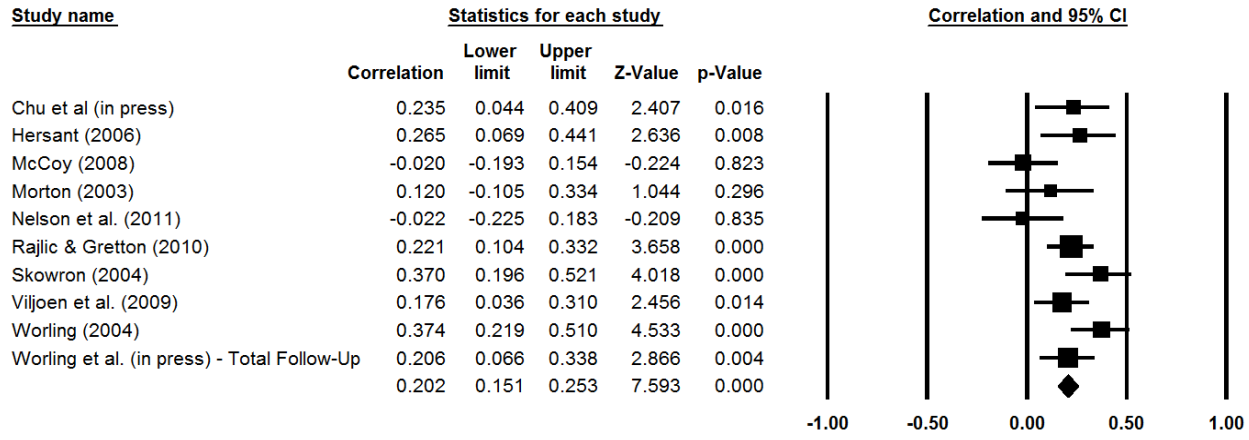
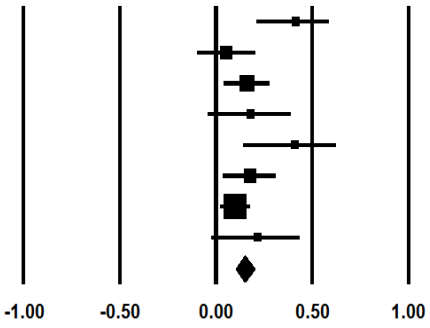


Figure 4:

Forest Plot: Correlations Between Total Scores on the Static-99and Sexual Reoffending

Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Beech et al. (2004)	0.414	0.209	0.584	3.789	0.000
Caldwell & Dickinson (2009) - Total Follow-Up	0.052	-0.098	0.200	0.677	0.499
Dahle et al. (2009)	0.160	0.042	0.274	2.652	0.008
Morton (2003)	0.180	-0.041	0.385	1.597	0.110
Poole et al. (2000)	0.408	0.143	0.618	2.938	0.003
Viljoen et al. (2009)	0.176	0.036	0.310	2.456	0.014
Williams (2007)	0.099	0.023	0.174	2.548	0.011
Worling (unpublished)	0.216	-0.026	0.434	1.756	0.079
	0.149	0.100	0.197	5.901	0.000



Endnotes

[1] Search terms for J-SOAP-II: "Juvenile Sex Offender Assessment Protocol" or "J-SOAP* or "JSOAP"; search terms for ERASOR: "Estimate of Risk of Adolescent Sexual*" or "Estimated Risk of Adolescent Sexual" or "ERASOR"; search terms for J-SORRAT-II: "Juvenile Sexual Offender Recidivism Risk*" and "JSORRAT*" or "J-SORRAT*"; search terms for Static-99: "Static-99" or "Static 99."

[2] The formula used was $\sum$ (agreements)/ $\sum$ (agreements + disagreements) (Yeaton & Wortman, 1993).

[3] Two studies (both on the J-SOAP-II) described the effect size as "non-significant (Fanniff & Becker, 2007; Parks & Bard, 2006). In these cases, the authors were contacted for information in order to code the effect sizes. In one case, the authors responded and provided the necessary information to proceed (Fanniff, personal communication, February 21, 2011).

[4] CMA2 does not handle AUCs.

[5] As meta-analytic procedures for AUCs are less well-established than those for correlations, we re-ran our analyses using several additional approaches: a) we weighted by sample size using the Hunter and Schmidt methods (1990, 2004 consistent with Schwalbe, 2007), b) we conducted McClish's (1992) fixed effect model which utilizes inverse variance, and c) we calculated simple mean AUCs (i.e., an unweighted average). In general, we obtained similar patterns of results. These results are available from the authors upon request.

[6] We considered testing several additional moderators, including sources of information used to code risk assessment tool (file information or interview/file information), definition of recidivism (reconviction or charges/arrest), and study design (retrospective examinations of repeaters vs. non-repeaters, pseudo-prospective studies based on historical file information, or truly prospective studies). However, these factors could not be tested as moderators due to the lack of variability across studies. For instance, there was only one truly prospective design (Table 1). Age was not feasible to test as a moderator because some studies measured age at admission whereas others measured it at discharge or at some other time.

[7] One study presented both "any" and "nonsexual" categories. For this study, the two categories were averaged using CMA.

[8] To compare our findings to Hanson and Morton-Bourgon (2009), aggregated AUCs were converted to *d* scores using the formula in Rice and Harris (2005).

[9] Aggregated AUCs were converted to *d* scores using the formula in Rice and Harris (2005).