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, and Jennifer L. Beneteau
Simon Fraser University

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 Given that research on these tools has yielded somewhat mixed results, we empirically synthesized 33 published and unpublished studies involving 6,196 male adolescents who had committed a sexual offense.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 areas under the receiver operating characteristic curve (AUCs). Total scores on each of the tools significantly predicted sexual reoffending, with aggregated correlations ranging from .12 to .20 and aggregated AUC scores ranging from .64 to .67. However, in many cases 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. Results are compared to other meta-analyses of risk tools used in the area of violence risk assessment and in other fields.

Keywords: sexual offending, juvenile offenders, risk assessment, Juvenile Sexual Offense Recidivism Risk Assessment Tool, Estimate of Risk of Adolescent Sexual Offense Recidivism, Static-99

Since the 1990s, numerous 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 many 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 policy-makers (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 several studies provide support for the use of the Static-99 with adult sexual offenders (e.g., Ducro & Pham, 2006; Hanson & Morton-Bourgon, 2009; Langton, et al., 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, several researchers have emphasized the need for tools designed specifically for adolescents, especially because 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 et al., 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 several 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 sexual and nonsexual reoffending among sexually abusive adolescents, the ERASOR's stated purpose is only to predict sexual reoffending. Second, although 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 in which 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). Although the J-SORRAT-II is a brief

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 cutoffs have not yet been developed.	Consists of 28 risk factors that are organized into four subscales (Sexual Drive and Preoccupation, Impulsive/Antisocial Behavior, Intervention, and 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 and is 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 five 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.

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 Behavior, 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, 2011), 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 several studies have found that J-SOAP-II total scores predict sexual reoffending (Martinez, Flores, & Rosenfeld, 2007; Prentky et al., 2010; Rajlic & Gretton, 2010a), other studies have reported nonsignificant 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, 2010a; Worling et al., 2011), whereas other research reported nonsignificant AUCs (Morton, 2003).

Several explanations may contribute to these equivocal findings. First, these results could suggest that it is somewhat challenging to predict adolescent sexual offending, possibly because of 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 (2010a) found that the ERASOR and J-SOAP-II had higher levels of predictive validity among adolescents who had committed only sexual offenses than among those who also had a history of nonsexual 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. Although 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 several 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

This meta-analysis captured 33 studies that were 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 2 for descriptive information on studies). Most studies focused exclusively on male adolescents, although a few did not explicitly mention gender. Most samples included predominantly Caucasian youth. Adolescents' index offenses ranged considerably and included penetrative and nonpenetrative offenses. The mean follow-up periods, during which reoffending was examined, ranged from 12 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 pseudoprospective 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, & The 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

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. (2011)	223	Switzerland	100	—	15.7	51.6	Charges[c]	Pseudoprospective	All youths convicted by courts	No	3.1[e] 44.8[f]
Beech et al. (2004)	77	United Kingdom	100	—	14.2	—	Convictions[c]	Unclear[d]	Community mental health services	No	9.0[e] —[f]
Caldwell (J-SORRAT-II) (2011a)	91 (same sample as Caldwell et al., 2008)	United States	100	52	15.4	71.6	Charges[c]	Pseudoprospective	Correctional treatment program	Yes	12.1[e] 69.0[f]
Caldwell (2011b) (Static-99)	172 (same sample as Caldwell & Dickinson, 2009)	United States	100	58.7	17.9	49.2	Charges[c]	Pseudoprospective	Correctional center	No	12.2[e] 59.3[f]
Caldwell & Dickinson (2009)	172	United States	100	58.7	17.9	49.2	Charges[c]	Pseudoprospective	Correctional center	No	12.2[e] 59.3[f]
Caldwell, Ziemke, & Vitacco (2008)	91	United States	100	52	15.4	71.6	Charges[c]	Pseudoprospective	Correctional treatment program	Yes	12.1[e] 69.0[f]
Chu et al. (2011)	104	Singapore	100	<2.0	15.2	54.6	Official records[c]	Pseudoprospective	Community mental health services	No	7.7[e] —[f]
Dahle et al. (2009)	273	Germany	100	—	—	99.6	Official records[c]	Pseudoprospective[d]	All youths adjudicated by courts	No	26.0[g] 17.0[e] —[f]
Epperson et al. (2006)	636	United States	100	76.4	15.2	—	Charges[b]	Pseudoprospective	All youths adjudicated by courts	No	19.8 (juvenile or adult)[e] —[f]
Epperson & Ralston (2009)	318	United States	—	—	—	—	Charges[a]	Pseudoprospective[d]	All youths adjudicated by courts	No	7.2[e] —[f]
Fanniff & Becker (2007)	277	United States	100	52	13.7	35.2	Official records[c]	Pseudoprospective[d]	All youths adjudicated by courts	No	9.6[e] 73.0[f]
Hersant (2007)	91	United States	100	63	15	—	Caught and sanctioned by adult[c]	Retrospective	Three residential treatment centers	Yes	—[e] —[f]
Lehman (2008)	97	United States	100	51.1	14.5	102.6	Arrest[b]	Pseudoprospective[d]	Community treatment program	Yes	17.5 60.5[f]
Martinez, Flores, & Rosenfeld (2007)	60	United States	100	16.7	14.9	—	Arrests, self-report, and other informant[c]	Pseudoprospective	Community treatment program	Yes	13.3[e] 20.0[f]
McCoy (2008)	128	United States	100	35.9	15.3	—	Arrests[b]	Pseudoprospective	Community treatment program	Yes	5.6[e] 57.8[f] 56.0[g]
Morton (2003)	80	Canada	100	—	15.2	68.1	Charges	Pseudoprospective	Community mental health services	No	16.9[e] 51.9[f]
Nelson et al. (2011)	93	United States	—	55	15.6	80.4	Official records[c]	Pseudoprospective	Referred for risk evaluation	No	10.8[e] —[f] 62.4[g]

Table 2 (*continued*)

Study	Country	n	Male %	White %	Mean age	Mean follow-up Months	Definition of sexual reoffense	Design	Setting	Treatment	Base rate of recidivism %
Parks & Bard (2006)	United States	156	100	62.8	16.1	—	Convictions[b]	Pseudoprospective	Correctional center	Yes	6.4[e] / —[f] / 30.1[g]
Petersen (2010)	United States	129	100	—	13.7	12	Charges[c]	Pseudoprospective	Treatment program	Yes	8.8[e] / 23.2[f] / 17.6[g]
Poole, Liedecke, & Marbibi (2000)	United States	49	100	—	18.5	48	Arrests[c]	Pseudoprospective	Correctional center	No	8.2[e] / —[f]
Powers-Sawyer & Miner (2009)	United States	96	100	72	17.2	51.5	Arrests and parole violations[b]	Pseudoprospective	Correctional treatment program	Yes	7.3[e] / —[f]
Prentky et al. (2010)	United States	223 (adolescent sample)	100	62.3	14.3	84	Any sexually abusive "hands-on" behavior[c]	Pseudoprospective[d]	Referred for sexual risk evaluation by social services	No	13.9[e] / —[f]
Rajlic, Clift, & Gretton (2010)	Canada	198	—	65.3	15.8	84	Charges[c]	Pseudoprospective[d]	Community treatment program	Yes	9.6[e] / 42.9[f] / 33.3[g]
Rajlic & Gretton (2010a)	Canada	268	100	66	15.8	79.2	Charges[c]	Pseudoprospective	Community treatment program	Yes	9.4[e] / 43.4[f] / 33.9[g]
Ralston (2008)	United States	566	100	76	15	120	Charges[a]	Pseudoprospective	All youths adjudicated by courts	No	12.4[e] / —[f]
Skowron (2004)	Canada	110	100	—	14.2	47.3	Arrests[b]	Pseudoprospective	Community mental health services	No	35.0[e] / 62.0[f]
Viljoen, Scalora, Cuadra, et al. (2008)	United States	169	100	83.4	15.4	78.9	Charges[b]	Pseudoprospective	Residential treatment program	Yes	8.3[e] / 42.8[f]
Viljoen, Elkovitch, Scalora, & Ullman (2009)	United States	193	100	82.9	15.3	86.9	Charges[a]	Pseudoprospective	Residential treatment program	Yes	8.3[e] / 42.0[f]
Waite et al. (2005)	United States	256	100	43.5	16.8	61.9	Arrests[b]	Pseudoprospective	Correctional treatment program	Yes	4.7[e] / —[f]
Williams (2007)	United States	661	100	43.6	15	—	Charges[c]	Unclear[d]	Correctional treatment program	Yes	5.3[e] / —[f]
Worling (2004)	United States, Canada	136	100	—	14.9	—	Caught and sanctioned by any adult[a]	Retrospective	Community mental health and residential treatment	Yes	—[e] / —[f]
Worling (2011) (J-SORRAT-II, Static-99)	Canada	75	—	—	—	90	Charges and reports to agencies[b]	Pseudoprospective	Community mental health services	No	8.0[e] / —[f]
Worling et al. (2011)	Canada	191	100	—	15.3	43.9	Charges and reports to agencies[b]	Prospective	Community mental health services	No	9.4[e] / —[f]

[a] Included juvenile records only. [b] Included juvenile and adult records. [c] Did not specify whether juvenile or adult records were included. [d] Did not specify whether coders were blind to recidivism status. [e] Any recidivism. [f] Any sexual recidivism. [g] Nonsexual recidivism.

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 (N = 199)

Unpublished Sources
- Dissertation databases: ProQuest Dissertations & Theses, Universal Index of Doctoral Dissertations In Progress, and Networked Digital Library of Theses & Dissertations (N = 36)
- Google scholar (N = 887)
- Conference proceedings (N = 38)
- Experts in field (N = 7)

Total records screened using abstract (N = 1164)

Excluded for not meeting screening criteria for age or inclusion of tool (N = 1112)

Studies eligible for full-text review (N = 52)

Excluded for not meeting eligibility criteria, multiple disseminations of same study, other (N = 19)
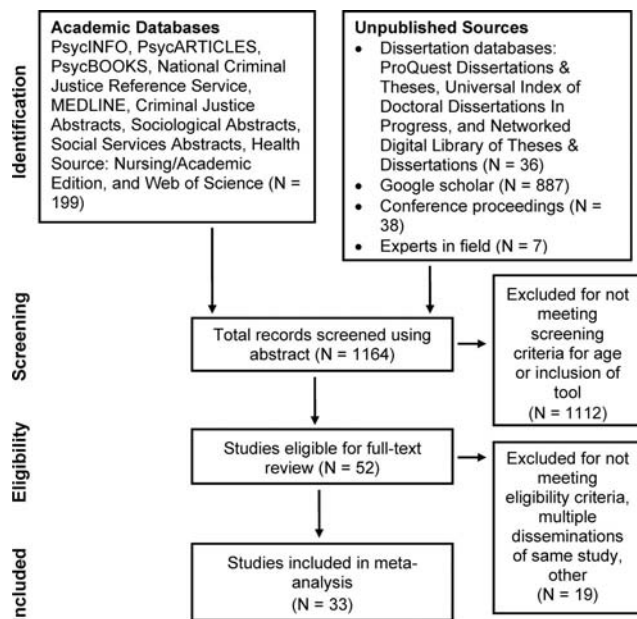
Studies included in meta-analysis (N = 33)

Figure 1. Search strategy and phases of review.

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 that 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; and 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 to determine if they mentioned any other studies that could meet inclusion criteria.

Eligibility Screening

On the basis of 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, most study participants had to fall in the age range of 12-18 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. On the basis of 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), each of the three coauthors 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, 2010b; 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 et al., 2005; Young, 2006). One study was excluded because it used a case study methodology and relied primarily on qualitative data for three participants (Fleming, 2004), and one study was excluded because it focused on adults (Soothill, Harman, & Kirby, 2005).

Data Extraction

Once studies were selected for inclusion, we coded effect sizes and potential moderators (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 re-

---

[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."

solved through a consensus procedure before proceeding with analyses.

Data Analyses

AUCs have come to be widely accepted in risk assessment literature because 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, Grann, & Fazel, 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-analyses of other summary statistics (C. Gatsonis, 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 sexual reoffending and general reoffending (defined as any or nonsexual reoffending). In several studies, the predictive validity for two separate time periods was presented (e.g., Caldwell, 2011a; Caldwell & Dickinson, 2009; Epperson et al., 2006; Worling et al., 2011). In these cases, the longer time period was selected for the primary analyses because 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 between- and within-study variability (Borenstein, Hedges, Higgins, & Rothstein, 2005). 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. Several studies presented the data in forms other than correlations or AUCs (i.e., mean scores, proportion of offenders who reoffended, or χ^2). 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 the heterogeneity of findings, within-group *Q* statistics (Q_w) were calculated (using CMA2). The *Q* statistic is distributed as a χ^2 test (Hedges & Olkin, 1985). Given that the *Q* statistic only tests the presence or absence of heterogeneity, CMA2 also generates an I^2 index that 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 and Wilson's MetaES macro (random-effects model). This macro can be used with AUC scores (David B. Wilson, personal communication, October 6, 2011), and weights studies by inverse variance. 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).

Because meta-analytic procedures for AUCs are less well established than those for correlations, we reran our analyses using the Hunter and Schmidt (1990, 2004; consistent with Schwalbe, 2007) method, which weights by sample size rather than by inverse variance. We obtained very similar patterns of results, and as such we present only the results based on Lipsey and Wilson's MetaES macro because weighting by inverse variance typically yields more accurate estimates than weighting by sample size (Marín-Martínez & Sánchez-Meca, 2010).

Moderators (Metaregression)

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). Several studies presented numerous effect sizes and used multiple tools. Thus, to avoid erroneously treating each effect size as independent, we ran these analyses in CMA2 (using aggregated correlations) because this enabled us to collapse findings across studies. For instance, if a study included two 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 structured professional judgment (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. Heterogeneity 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). On the basis of 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* =

Table 3

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

			Correlations					AUCs			
			95% confidence interval		Heterogeneity				95% confidence interval		Heterogeneity
	k	r_w			Q	I^2	AUC_w				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, * $p < .05$. ** $p < .01$. *** $p < .001$.

.10 (i.e., a small effect). There would need to be eight such studies on the J-SOAP-II total score, four on the Static-99 total score, and two 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 or-

dering 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 affected the results. First, although the Parks and Bard (2006) study on the J-SOAP-II was excluded in the initial

Table 4

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

			Correlations					AUCs			
			95% confidence interval		Heterogeneity				95% confidence interval		Heterogeneity
	k	r_w			Q	I^2	AUC_w				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 professional judgment rating											
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, and 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, * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
J-SORRAT-II and Static-99 Total Scores: Relationship to Sexual Reoffending

| | | Correlations | | | | | AUCs | | | | |
| | | | 95% confidence interval | | Heterogeneity | | | 95% confidence interval | | Heterogeneity | |
	k	r_w			Q	I^2	AUC_w			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, * $p < .05$. ** $p < .01$. *** $p < .001$.

analyses because no effect size was reported (it was simply described as "nonsignificant" in the article), the results were rerun 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 $.17$, $p < .01$; AUC $= .67$ to $.66$), indicating that the inclusion of this study would not have changed the findings to nonsignificant. Second, because the study conducted by Prentky et al. (2009) was classified as an outlier (z-score >5.00) the results were rerun 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 $.15$, $p < .001$; 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 rerun with this study excluded, AUCs for the J-SORRAT-II total score decreased somewhat but the aggregated correlation remained significant ($r = .12$ before and after, $p < .01$; 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 (United States 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), and (i) interrater reliability of tool (strong interrater reliability vs. no information or inadequate interrater reliability). 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. After correcting for multiple comparisons using a Bonferroni 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 we 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 ($n = 5$).

Aggregated correlations. At an aggregate level, total scores on the J-SOAP-II and ERASOR predicted general reoffending, as did several scales, particularly the Impulsive/Antisocial Scale of the J-SOAP-II and the Psychosocial Functioning section of the ERASOR (see Tables 3 and 4). Although the aggregated r for the J-SOAP-II was higher than the ERASOR for the prediction of general reoffending, this difference was not significant. On the
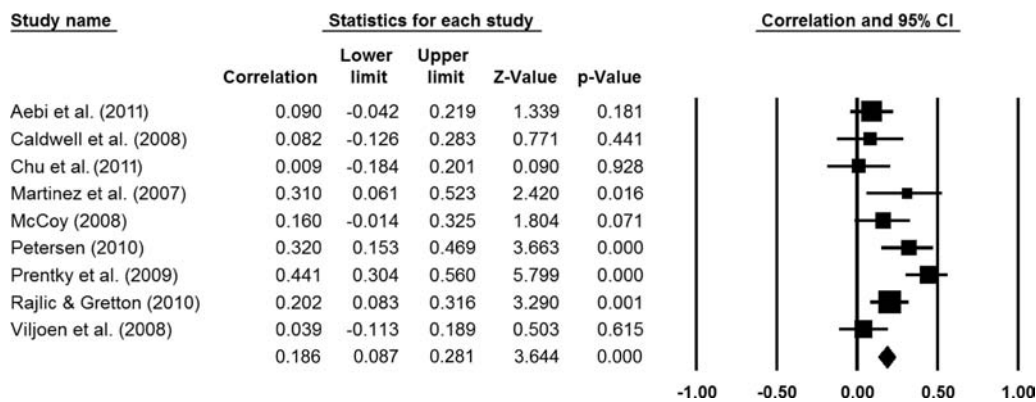
Study name	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Aebi et al. (2011)	0.090	-0.042	0.219	1.339	0.181	
Caldwell et al. (2008)	0.082	-0.126	0.283	0.771	0.441	
Chu et al. (2011)	0.009	-0.184	0.201	0.090	0.928	
Martinez et al. (2007)	0.310	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	
	0.186	0.087	0.281	3.644	0.000	

Figure 2. Forest plot: Correlations between total scores on the J-SOAP-II and sexual reoffending.

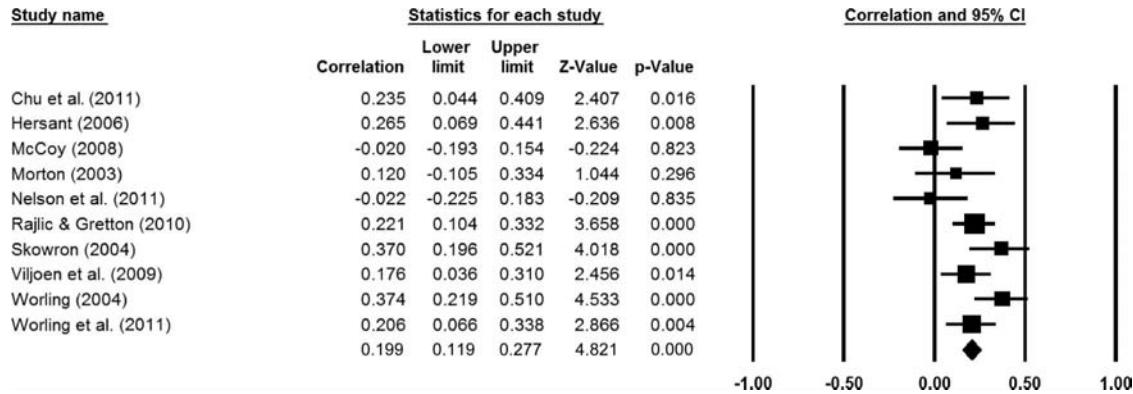| Study name	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Chu et al. (2011)	0.235	0.044	0.409	2.407	0.016	
Hersant (2006)	0.265	0.069	0.441	2.636	0.008	
McCoy (2008)	-0.020	-0.193	0.154	-0.224	0.823	
Morton (2003)	0.120	-0.105	0.334	1.044	0.296	
Nelson et al. (2011)	-0.022	-0.225	0.183	-0.209	0.835	
Rajlic & Gretton (2010)	0.221	0.104	0.332	3.658	0.000	
Skowron (2004)	0.370	0.196	0.521	4.018	0.000	
Viljoen et al. (2009)	0.176	0.036	0.310	2.456	0.014	
Worling (2004)	0.374	0.219	0.510	4.533	0.000	
Worling et al. (2011)	0.206	0.066	0.338	2.866	0.004	
	0.199	0.119	0.277	4.821	0.000	

Figure 3. Forest plot: Correlations between total scores on the ERASOR and sexual reoffending.

basis of 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 the 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 "nonsignificant"), the results were rerun 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, $p < .01$; AUC = .66 to .64).

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

Further Examination of Studies that Directly Compared Two Tools

We examined studies that 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. However, these results should be interpreted with caution given the small *n*.

Discussion

Concerns have been raised that it may be challenging to predict sexual reoffending in adolescents because of the tremendous developmental changes that occur during this period (Caldwell et al., 2008; Viljoen, Scalora, Caudra, Bader, Chavez, Ullman, & Lawrence, 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,
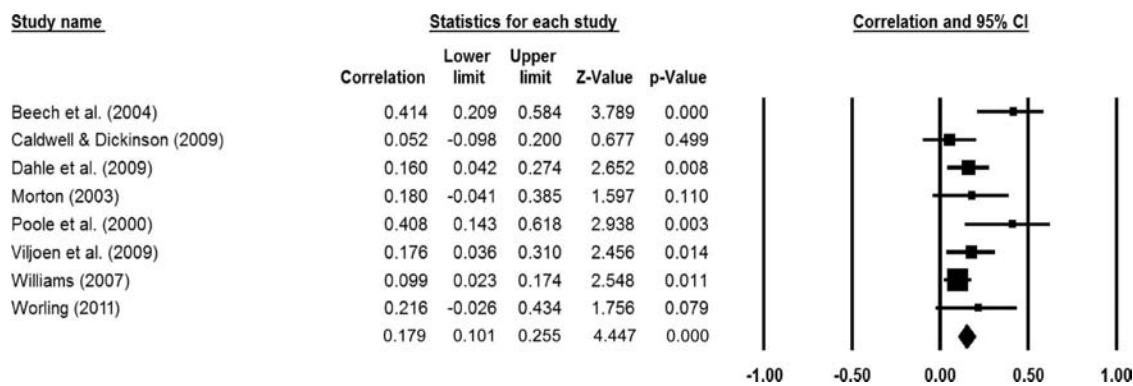

Study name	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Beech et al. (2004)	0.414	0.209	0.584	3.789	0.000	
Caldwell & Dickinson (2009)	0.052	-0.098	0.200	0.677	0.499	
Dahle et al. (2009)	0.160	0.042	0.274	2.652	0.008	
Morton (2003)	0.180	-0.041	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 (2011)	0.216	-0.026	0.434	1.756	0.079	
	0.179	0.101	0.255	4.447	0.000	

Figure 4. Forest plot: Correlations between total scores on the Static-99 and sexual reoffending.

Table 6
Studies That Directly Compared Risk Assessment Tools

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

Note. n.s. = not significant.

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). However, correlations may underestimate predictive validity for low base rate events (Rice & Harris, 2005). As such, we also performed a separate meta-analysis using AUC scores. This yielded a comparable result; AUCs for total scores ranged from .64 to .67, which is generally considered to reflect moderate AUCs (Douglas, Blanchard, Guy, Reeves, & Weir, 2010; Rice & Harris, 2005). There were no significant differences between tools regardless of whether analyses were conducted with correlations or AUCs. A benefit of conducting meta-analyses with AUCs, however, is that AUCs are more commonly used in risk assessment research, and thus perhaps more easily understood.

Our findings are positive in many respects. The sex-offense-specific tools examined in this meta-analysis appear to outperform more general tools such as the Hare Psychopathy Checklist: Youth Version in the prediction of sexual reoffending when compared with findings from other meta-analyses, indicating that sex-offense-specific tools are preferable to general tools for the assessment of sexual reoffense risk (Edens et al., 2007; Olver et al., 2009; see Table 7). 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 victims, 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.

Although some researchers have raised questions about whether it may be particularly challenging to assess risk of sexual reoffending in adolescents compared with adults, our overall effect sizes were quite similar to the mean effect sizes found for tools used in adult sex offender risk assessment (Hanson & Morton-Bourgon, 2009; see Table 7). In addition, when compared with tools that are used to assess risk in other domains, the effect sizes found in this meta-analysis appear to be fairly similar to the effect sizes found for the prediction of self-harm risk with the Beck Hopelessness Scale (McMillan, Gilbody, Beresford, & Neilly, 2007; see Table 7). They are higher than those found for the prediction of risk of driving accidents with a popular self-report questionnaire (i.e., Driving Behavior Questionnaire, de Winter & Dodou, 2010), but they are 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., 1.4-4% for total scores). Also, our aggregated correlations were not as high as those recently reported for tools such as the Structured Assessment of Violence Risk in Youth and Youth Level of Service/Case Management Inventory in the prediction of general reoffending (Olver et al., 2009; see Table 7). This could suggest that it may be more challenging to predict sexual than general reoffending. However, an earlier meta-analysis on general reoffending reported AUCs that were quite similar to our results (Schwalbe, 2007).

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 also arisen in some other risk assessment contexts (Edens et al., 2007; Olver et al., 2009), risk assessment of adolescent sexual reoffending may 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 because 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. Although it is possible that certain risk factors, such as dynamic factors, may be particularly important during adolescence, recent research indicates that several risk factors for adolescent sexual reoffending overlap considerably 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.

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

Table 7

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

Meta-analyses	Aggregated findings for total scores		
	r	AUC	*d*
Adolescent Reoffense Risk			
Present study: J-SOAP-II, ERASOR, J-SORRAT-II, & Static-99	.12–.20 (sexual) .14–.25 (general)	.64–.67 (sexual) .59–.66 (general)	.51–.62 (sexual)[a] .33–.58 (general)
Psychopathy Checklist (Edens et al., 2007)	.07 (sexual) .24 (general)	—	—
SAVRY, YLS/CMI, & Psychopathy Checklist (Olver et al., 2009)	.06–.19 (sexual) .25–.33 (general)	—	—
Various risk tools (Schwalbe et al., 2007)	—	.59–.70 (general)	—
Sexual Reoffense Risk in Adults			
Actuarial, mechanical, & SPJ tools for sex offenders (Hanson & Morton-Bourgon, 2009)	—	—	.46–.67 (sexual) .26–.52 (general); specific tools ranged from .33–1.11 (sexual)
Suicide Risk			
Beck Hopelessness Scale (McMillan et al., 2007)	—	.70 (suicide) .63 (self-harm)	—
Risk for Driving Problems			
Various cognitive tests in stroke patients (Devos et al., 2011)	—	—	.15–1.22 (on-road driving)
Driving Behaviour Questionnaire (de Winter & Dodou, 2010)	.10–.13 (accidents)	—	—
Various cognitive tests in older adults (Mathias & Lucas, 2009)	—	—	.04–1.18 (driving problems)
Risk for Adverse Medical Outcomes			
Pneumonia severity assessment tools (Chalmers et al., 2010)	—	.79–.81 (mortality)	—
ABCD & ABCD2 (Giles & Rothwell, 2010)	—	.78 (stroke)	—
European System for Cardiac Operative Risk Evaluation (Parolani et al., 2010)	—	.72–.73 (cardiac events)	—

Note.   To derive this table, we conducted a 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. 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.
[a] Aggregated AUCs were converted to *d* scores using the formula in Rice and Harris (2005).

sexual reoffending is predicted by 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 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 moderate, these tools may be insufficient to make predictions that require a high degree of precision, such as civil commitment of adolescent sex offenders or the placement of adolescents on sex offender registries for life (see Caldwell et al., 2008). A recent meta-analysis of adult violence risk tools similarly cautioned that risk assessment tools should not be relied upon as a primary or sole source of information to make decisions that

"require a very high level of accuracy such as preventive detention" (Yang, Wong, & Coid, 2010, p. 740).

Despite common desires to identify which tool is "best," no single tool emerged as significantly stronger than the others. Also, whereas 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.

Although 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 this 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 focus of Static-99 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 many unpublished studies ($n = 19$), although it is possible that some relevant studies were not captured by search procedures. Also, although most meta-analyses of risk tools have used correlations, in this meta-analysis, analyses were conducted with AUC scores and correlations to address the limitations of correlations.

However, like any meta-analysis this meta-analysis is limited by the quality of research. Nearly all of the studies included in this meta-analysis were pseudoprospective 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.

On the basis of 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 ERASOR and J-SOAP-II are also intended to help manage risk and plan treatment to prevent reoffending. Increased attention to the utility of tools for these purposes will enable us to move beyond simply the prediction of reoffending toward 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 because some authors have suggested that risk assessments of youth may have a relatively quick expiration period because of developmental change (Prescott, 2004; Worling et al., 2011). In this 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 somewhat surprising or potentially disappointing given the enormous investments in the development of adolescent tools. Rather than suggesting we abandon efforts to develop adolescent 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 absence of research that directly tests how risk and protective factors overlap and differ across age groups. This basic underlying knowledge may help to advance adolescent risk assessment.

References

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

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

*Aebi, M., Plattner, B., Steinhausen, H. C., & Bessler, C. (2011). Predicting sexual and nonsexual recidivism in a consecutive sample of juveniles convicted of sexual offences. *Sexual Abuse: A Journal of Research and Treatment, 23,* 456–473. doi:10.1177/1079063210384634

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. doi:10.1207/s15327752jpa8701_07

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

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. doi:10.1111/j.1468-2850.2008.00147.x

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, Canada: Institute Against Family Violence.

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

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist, 51,* 945–956. doi:10.1037/0003-066X.51.9.945

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). [Static-99 and recidivism]. Unpublished raw data.

*Caldwell, M. (2011b). [J-SORRAT-II and recidivism]. Unpublished raw data.

*Caldwell, M., & Dickinson, C. (2009). Sex offender registration and recidivism risk in juvenile sexual offenders. *Behavioral Sciences & the Law, 27,* 941–956. doi:10.1002/bsl.907

*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. doi:10.1037/a0013241

Caldwell, M. F. (2002). What we do not know about juvenile sexual reoffense risk. *Child Maltreatment, 7,* 291–302. doi:10.1177/107755902237260

Chaffin, M., Berliner, L., Block, R., Johnson, T. C., Friedrich, W. N., Louis, D. G., . . . Madden, C. (2008). Report of the ATSA Task Force on Children With Sexual Behavior Problems. *Child Maltreatment, 13,* 199–218. doi:10.1177/1077559507306718

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. Lincoln: University of Nebraska–Lincoln.

*Chu, C. M., Ng, K., Fong, J., & Teoh, J. (2011). 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.* Advance online publication. doi:10.1177/1079063211404250

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

Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

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. (2008). Jugendliche sexualstraftäter: Spezialfälle der prognoseerstellung? Ein vergleich jugendlicher, heranwachsender und erwachsener sexual-straftäter bezüglich tatcharakteristika, rückfallhäufigkeit und prognose [Juvenile sexual offenders: Special cases for the prediction of recidivism? A comparison of juvenile, adolescent and adult sexual offenders regarding modus operandi, recidivism rates and prediction of reoffence]. *Forensiche Psychiatrie, Psychologie, Kriminologie, 2,* 213–221. doi: 10.1007/s11757-008-0091-7

*Dahle, K. P., Janka, C., Gallasch-Nemitz, F., & Lehmann, R. (2009). Tatcharakteristika, rückfallrisiko und rückfallprognose bei sexual-straftätern vom jugend bis ins seniorenalter [Offending characteristics, risk of recidivism and prediction of reoffence in sex offenders, from adolescence to later adulthood]. *Forensiche Psychiatrie, Psychologie, Kriminologie, 3,* 210–220. doi:10.1007/s11757-009-0007-1

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. doi: 10.1212/WNL.0b013e31820d

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. doi:10.1016/j.jsr.2010.10.007

Douglas, K. S., Blanchard, A. J., Guy, L. S., Reeves, K. A., & Weir, J. (2010). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography.* Retrieved from http://kdouglas.files.wordpress.com/2007/10/hcr-20-annotated-biblio-sept-2010.pdf

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. doi:10.1177/107906320601800102

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. doi:10.1007/s10979-006-9019-y

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. doi:10.1080/13552600500063641

Egger, M., & Smith, G. (1997). Meta-analysis: Potentials and promise. *British Medical Journal, 315,* 1371–1374. doi:10.1136/bmj.315.7119.1371

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. doi:10.1002/bsl.832

*Epperson, D. L., & Ralston, C. A. (2009, October). *Validation of the Juvenile Sexual Offender Recidivism Risk Assessment Tool—II (JSORRAT-II) in Iowa.* Paper presented at the meeting of the Association for the Treatment of Sexual Abusers, Dallas, TX.

*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 meeting of the International Congress on Law and Mental Health, Padua, Italy.

Fanniff, A. M., & Letourneau, E. (2011). [J-SOAP-II and revidivism]. Unpublished raw data.

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. doi:10.1161µSTROKEAHA.109.571174

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, Canada.

Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143,* 29–36.

Hanley, J., & McNeil, B. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

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. doi:10.1037/a0014421

Hanson, R. K., & Thornton, D. (1999). *Static–99: Improving actuarial risk assessments for sex offenders* (User Rep. No. 1999–02). Ottawa, Canada: Department of the Solicitor General of Canada. Retrieved from http://www.courtdiagnostic.com/Static%2099–02.pdf

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules revised – 2003.* Ottawa, Canada: Corrections Directorate, Solicitor General Canada. Retrieved from http://www.publicsafety.gc.ca/res/cor/rep/_fl/2003–03-stc-cde-eng.pdf

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

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. doi:10.1037/1082-989X.3.4.486

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

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. doi:10.1037/1082-989X.11.2.193

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.

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

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 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. doi:10.1177/107906320701900407

Kester, A. D., & Buntinx, F. (2000). Meta-analysis of ROC curves. *Medical Decision Making, 20,* 430–439. doi:10.1177/0272989X0002000407

Langton, C., Barbaree, H., Seto, M., F. E. 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. doi:10.1177/0093854806291157

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

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., . . . Mohler, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *British Medical Journal, 339,* b2700. doi:10.1136/bmj.b2700

Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational And Psychological Measurement, 70,* 56–73. doi:10.1177/0013164409344534

*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. doi:10.1177/0093854807301791

Martinez, R., Rosenfeld, B., & Flores, J. (2004, March). *Predictive ability of the J-SOAP-II for sexually offending minority youth.* Paper presented at the 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. doi:10.1017/S1041610209009119

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. doi:10.1177/1541204008320260

McClish, D. K. (1992). Combining and comparing area estimates across studies or strata. *Medical Decision Making, 12,* 274–279. doi:10.1177/0272989X9201200405

*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. Huntsville, TX: Sam Houston State University.

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, 37,* 769–778. doi:10.1017/S0033291706009664

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. doi:10.1177/002242702237287

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6,* e1000097. doi:10.1371/journal.pmed.1000097

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

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783–792. doi:10.1037/0022-006X.62.4.783

*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 meeting of the American Psychology-Law Society, Miami, FL.

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. doi:10.1177/0093854809331457

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics, 8,* 157–159. doi:10.3102/10769986008002157

*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. doi:10.1177/107906320601800402

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

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. doi:10.1016/j.athoracsur.2009.11.032

*Petersen, J. (2010). *Validation of the Juvenile Sex Offender Assessment Protocol (JSOAP-II) for predicting sexual and non-sexual crimes.* Unpublished Master's thesis. East Lansing, MI: 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.* Austin, TX: Texas Youth Commission.

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

*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. doi:10.1002/bsl.920

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

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. Retrieved from https://www.ncjrs.gov/pdffiles1/ojjdp/202316.pdf

Prentky. R. A. (2006). Risk management of sexually abusive youth: A follow-up study. U.S. Department of Justice Final Report. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/214261.pdf

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. doi:10.1177/107906320001200201

Prescott, D. (2004). Emerging strategies for risk assessment of sexually abusive youth: Theory, controversy, and practice. *Journal of Child Sexual Abuse, 13,* 83–105. doi:10.1300/J070v13n03_05

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of lnterpersonal Violence, 10,* 85–105. doi:10.1177/088626095010001006

*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 meeting of the International Association of Forensic Mental Health Services Conference, Vancouver, Canada.

*Rajlic, G., & Gretton, H. (2010a). An examination of two sexual recidivism risk measures in adolescent offenders: The moderating effect of offender type. *Criminal Justice and Behavior, 37,* 1066–1085. doi:10.1177/0093854810376354

Rajlic, G., & Gretton, H. (2010b). *Predictive validity of the J-SOAP and ERASOR in two types of adolescent sexual offenders.* Paper presented at the meeting of the American Psychology-Law Society, Vancouver, Canada.

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

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

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. doi:10.1007/s10979-005-6832-7

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737–748. doi:10.1037/0022-006X.63.5.737

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

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. Kelowna, Canada: University of British Columbia–Okanagan.

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31,* 449–462. doi:10.1177/0093854808324377

Seto, M. C., & Lalumière, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin, 136,* 526–575. doi:10.1037/a0019700

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. doi:10.1177/0093854800027006004

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. doi:10.1016/j.cpr.2010.11.009

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

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. doi:10.1080/14789940500098277

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. doi:10.1097/AOG.0b013e3181e3e7e8

Trivits, L., & Reppucci, N. (2002). Application of Megan's law to juveniles. *American Psychologist, 57,* 690–704. doi:10.1037/0003-066X.57.9.690

*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. doi:10.1177/0093854809340991

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. doi:10.1177/1073191109359587

*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. doi:10.1177/0093854807307521

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 meeting of the American Psychology-Law Society, Jacksonville, FL.

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. doi:10.1002/bsl.909

Vitacco, M., Viljoen, J., & Petrila, J. (2009). Introduction to this issue: Adolescent sexual offending. *Behavioral Sciences & the Law, 27,* 857–861. doi:10.1002/bsl.915

*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. doi:10.1007/s11194-005-5061-4

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. doi:10.1002/bsl.911

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

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 Press.

*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. doi:10.1177/107906320401600305

*Worling, J. R. (2011). [J-SORRAT-II and Static-99 and recidivism]. Unpublished raw data.

*Worling, J. R., Bookalam, D., & Litteljohn, A. (2011). Prospective validity of the Estimate of Adolescent Sexual Offense Recidivism (ERASOR). *Sexual Abuse: Research and Treatment.* Advance online publication. doi:10.1177/1079063211407080

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

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. doi:10.1037/a0020473

Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review, 17,* 292–309. doi:10.1177/0193841X9301700303

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