Sexual Abuse: A Journal of Research and Treatment
http://sax.sagepub.com/

Pedophilia: An Evaluation of Diagnostic and Risk Prediction Methods

Robin J. Wilson, Jeffrey Abracen, Jan Looman, Janice E. Picheca and Meaghan Ferguson

Sex Abuse published online 9 November 2010
DOI: 10.1177/1079063210384277

The online version of this article can be found at:
http://sax.sagepub.com/content/early/2010/11/06/1079063210384277

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
ATSA
Association for the Treatment of Sexual Abusers

Pedophilia: An Evaluation of Diagnostic and Risk Prediction Methods

Sexual Abuse: A Journal of Research and Treatment
XX(X) 1–15

DOI: 10.1177/1079063210384277
http://sax.sagepub.com

Robin J. Wilson[1], Jeffrey Abracen[2], Jan Looman[2], Janice E. Picheca[2], and Meaghan Ferguson[2]

Abstract

One hundred thirty child sexual abusers were diagnosed using each of following four methods: (a) phallometric testing, (b) strict application of *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision [DSM-IV-TR]) criteria, (c) Rapid Risk Assessment of Sex Offender Recidivism (RRASOR) scores, and (d) "expert" diagnoses rendered by a seasoned clinician. Comparative utility and intermethod consistency of these methods are reported, along with recidivism data indicating predictive validity for risk management. Results suggest that inconsistency exists in diagnosing pedophilia, leading to diminished accuracy in risk assessment. Although the RRASOR and *DSM-IV-TR* methods were significantly correlated with expert ratings, RRASOR and *DSM-IV-TR* were unrelated to each other. Deviant arousal was not associated with any of the other methods. Only the expert ratings and RRASOR scores were predictive of sexual recidivism. Logistic regression analyses showed that expert diagnosis did not add to prediction of sexual offence recidivism over and above RRASOR alone. Findings are discussed within a context of encouragement of clinical consistency and evidence-based practice regarding treatment and risk management of those who sexually abuse children.

Keywords

pedophilia, diagnosis, risk assessment, DSM-IV-TR, RRASOR

[1]Florida Civil Commitment Center, Arcadia
[2]Correctional Service of Canada, Ottawa, Ontario

Corresponding Author:
Robin J. Wilson, GEO Group Inc., 13613 SE Highway 70, Arcadia, FL 34266
Email: dr.wilsonrj@verizon.net

Introduction

As behavioral scientists, sexual offender specialists are tasked with identifying those individuals most at risk of committing sexual offenses against vulnerable persons. It is interesting to note that although treatment providers have apparently made great gains in demonstrating the efficacy of contemporary cognitive-behavioral treatment methods (see meta-analyses by Hall, 1995; Hanson et al., 2002; Lösel & Schmucker, 2005), diagnosticians have continuously failed to demonstrate consistency or precision in identifying that higher risk group of child sexual abusers (see Kingston, Firestone, Moulden, & Bradford, 2007). Given the public's current views regarding those who sexually abuse children, taken together with the prescriptions of the risk/needs/responsivity model (see Andrews & Bonta, 2007; Wilson & Yates, 2009), it would seem that identifying those offenders most at risk is equally as important as having good methods for treating them.

Meta-analytic research as to the predictors of sexual recidivism clearly indicates that deviant arousal to children is related to the increased risk of recidivism (e.g., Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). Perhaps the best means of objectively measuring deviant sexual interest is the phallometric test (see Blanchard, Klassen, Dickey, Kuban, & Blak, 2001; Freund & Blanchard, 1989; Freund & Watson, 1991). In the phallometric test, participants are presented audiovisual stimuli intended to evoke differential levels of sexual arousal to various categories of persons or behaviors, during which measurements of penile tumescence are taken. The assumption made is that greater erectile response is reflective of greater sexual interest in (or preference for) the particular category of the participant or activity. Although the psychometric properties of the method appear to be inconsistent from site to site (i.e., there appears to be little standardization—see Marshall, 2005), many practitioners continue to view phallometry as a reliable means of assessing deviant sexual interests.

However, in addition to phallometry, there are other implicit means of assessing sexual deviance. The field of sexual offender risk management has learned much from the introduction of actuarial risk prediction tools (e.g., Sex Offender Risk Assessment Guide [SORAG—Quinsey, Harris, Rice, & Cormier, 2006]; Static-99 [Hanson & Thornton, 1999]). As noted by Hanson, Morton, and Harris (2003), actuarially based assessments are currently the most accurate means available of assessing risk. Those actuarial instruments, which have received the most empirical validation (e.g., SORAG, Static-99), include items that have a demonstrated relation to the criterion measure of interest (i.e., sexual offence recidivism). Furthermore, clear rules for scoring the items exist, as well as procedures whereby individual item scores can be combined to provide the assessor with an overall estimate of risk for a particular type of offending. In the present investigation, the Rapid Risk Assessment of Sex Offender Recidivism (RRASOR; Hanson, 1997; all four items of which also appear in the Static-99) was used to actuarially assess the participants included in the study. We used this index specifically because its items all have high face validity with commonly held features of pedophilia (e.g., male victims, high-density offending, etc.). For a discussion of the

relationship between deviant sexual interests and the RRASOR, the reader is referred to Doren (2004).

It is generally accepted that paraphilic individuals are more likely to recidivate sexually than nonparaphilic individuals, and most will further agree that persons so-diagnosed are requiring of special attention in regard to efficacious treatment programming. That said, in other research conducted by a subset of the current authors, no difference was found among treated sexual offenders in regard to rates of sexual recidivism for those with or without a diagnosis of a paraphilia (Abracen & Looman, 2006). In that study, it was observed that those sexual offenders who evidenced both a personality disorder and a paraphilic diagnosis were significantly more likely to recidivate sexually than those without such diagnoses (i.e., 20.6% vs. 9.6%, respectively). However, if recidivism rates were relatively low even among those with comorbid conditions, this has implications for current practices regarding long-term commitment (e.g., civil commitment of "sexually violent predators" [SVP]). It should be noted that the sample on which that study (i.e., Abracen & Looman, 2006) was based consisted exclusively of sexual offenders who scored at or above "5" on the Static-99, that is, they presented at the high end of the moderate range or higher with reference to risk of violent or sexual recidivism. As well, the mean number of prior sexual convictions for that sample was 4, with the mean number of violent nonsexual convictions being 1.9.

Another, more recent study has also raised concerns about the utility of diagnoses in risk assessment (Kingston et al., 2007). In that study, four potential methods for diagnosing pedophilia were compared, including *Diagnostic and Statistical Manual of Mental Disorders* (DSM) diagnosis by a psychiatrist, phallometric assessment, combined *DSM*/phallometry, and the Screening Scale for Pedophilic Interests (Seto & Lalumiére, 2001). Kingston et al. (2007) found that pedophiles and nonpedophiles were largely indistinguishable from each other on ancillary psychological measures and, more pertinent to the current study, that the four methods used to define pedophilia were not significantly related to one another. Their conclusion was that a diagnosis of pedophilia held limited utility for practitioners assessing and treating persons who sexually abuse children. Furthermore, they suggested that a sexual preference for children, as measured by phallometric testing, might provide more assistance in treatment and risk management.

Diagnosis of Pedophilia

The *DSM* (4th ed., text revision [DSM-IV-TR]; American Psychiatric Association, 2000) is regarded by many in the mental health sciences as a veritable diagnostic "Bible". The criteria set out in the *DSM-IV-TR* are used widely by various practitioners and are often entered into evidence in Court proceedings. However, there are ongoing concerns regarding the validity of diagnoses in applied settings (Levenson, 2004), and several authors have offered suggestions as to how a variety of diagnostic conditions, including the paraphilias, should be diagnosed in real world settings

(e.g., Doren, 2002). This often leads to difficulties in regard to reliability and validity, which would seem to be particularly true of the paraphilias, including pedophilia (see O'Donohue, Regev, & Hagstrom, 2000). The *DSM-IV-TR* defines paraphilia as "recurrent, intense sexually arousing fantasies, sexual urges, or behaviors generally involving 1) nonhuman objects, 2) the suffering or humiliation of oneself or one's partner, or 3) children or other nonconsenting persons that occur over a period of at least 6 months" (American Psychiatric Association, 2000, p. 566). The *DSM-IV-TR* diagnostic criteria for Pedophilia are as follows:

> Over a period of at least six months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger).
> The person has acted on these sexual urges, or the sexual urges or fantasies caused marked distress or interpersonal difficulty.
> The person is at least age 16 years and at least 5 years older than the child or children in Criterion A (American Psychiatric Association, 2000, p. 572).

Additional specifications are then made as to victim gender choice, familial versus extrafamilial context, and exclusive versus nonexclusive choice of children as "partners."

Wollert (2007) has recently published an article noting that only two studies have been conducted on *DSM*-based diagnostic reliability in SVP cases (Levenson, 2004; Packard & Levenson, 2006). In reviewing these data, Wollert (2007) noted that the earlier study indicated poor to fair diagnostic reliability for diagnostic category and decision as to whether to refer a client for commitment. In a reanalysis of these data, Packard and Levenson (2006) obtained much higher levels of reliability. Wollert (2007) reported the results from two studies, questioning the reliability of paraphilic diagnoses, and suggesting that the results obtained by Packard and Levenson (2006) were based on faulty assumptions. Nonetheless, Doren and Levenson (2008) have sharply criticized the approach taken by Wollert (2007).

Marshall (2007), one of the pioneers in the contemporary treatment of sexual offenders, has also expressed reservations as to the diagnostic reliability of *DSM*-based diagnoses. As an alternative to *DSM*-based diagnoses, he has argued for the rating of the features of each type of sexual offender on a variety of dimensions ranging from normal to seriously problematic. Marshall noted further that for the paraphilas (including those that identify sexual offences), the most recent versions of *DSM* (including *DSM-IV-TR*) do not include evidence with respect to reliability. Reliability data for the paraphilias extend back to *DSM-III* in 1980, but the diagnostic criteria have changed in important ways since *DSM-III* was introduced. It is interesting to note in this regard that Jackson and Hess (2007), in an examination of the practices of professionals conducting SVP evaluations, observed that all the experts surveyed agreed that paraphilic diagnoses were critical, yet only 80% assessed for paraphilia.

The goal of the present investigation was to investigate the level of agreement—specifically, in regard to identification of pedophilia—between four methods or

Table 1. Demographic Data

Demographic variables	Value
Age, *M* (*SD*)	42.3 (12.7)
% Female victims only	51.5
% Familial victims only	9.8
% With prior sexual offenses	45.0
RRASOR, *M* (*SD*)	2.4 (1.5)
% Low risk	63.7
% Moderate risk	11.5
% High risk	24.8
LSI-R, *M* (*SD*)	21.6 (11.0)
Modal SORAG % risk rating	.35
PCL-R, *M* (*SD*)	16.0 (8.0)
Phallometric deviance index	–0.24 (1.19)
% Preference for children	56.6
% Nondiscriminating	12.3
% Adult preference	31.0
% Meet *DSM-IV-TR* pedophilia criteria	71.3
% Expert judgment	77.3

Note: *N*s vary by analysis because of missing data for some variables. *DSM-IV-TR* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision); RRASOR = Rapid Risk Assessment of Sex Offender Recidivism; SORAG = Sex Offender Risk Assessment Guide.

indices typically used in sexual offender assessments. In particular, level of agreement was examined between actuarial assessment of risk (i.e., the RRASOR), phallometric assessment, expert diagnosis (which included knowledge of all the available evidence), and *DSM-IV-TR*-based diagnoses. As well, the relationship between each of these four predictor variables and sexual offence recidivism was examined. It was expected that the different methods of assessment would agree in terms of who was classified as deviant (or high risk), and that the various methods would each be related to sexual recidivism.

Method

Participants

For the purposes of this investigation, 200 sexual offenders were selected at random from the intake assessment database at the Millhaven Institution Sexual Behaviour Clinic in Kingston, Ontario, Canada. Of those, 148 were convicted of sexual offenses against children. Ten participants were excluded because of low response levels on phallometric testing, with a further 8 participants being excluded because of excessive missing data—resulting in a final sample size of 130 offenders. Demographic data are presented in Table 1.

Also included in Table 1—for comparison purposes—are mean (and some modal) scores on risk assessment indices not specifically used in this study but presented for information purposes. Because this article also makes several comments regarding the use of diagnostic and actuarial assessment information in risk prediction, many readers may be interested in scores on other measures. Specifically, we have included scores on the Level of Service Inventory–Revised (LSI-R—see Andrews & Bonta, 2007), SORAG (see Quinsey et al., 2006), and the Psychopathy Checklist–Revised (PCL-R—see Hare, 1991, 2003). In each measure, the central tendency score reported is equivalent to a moderate risk rating, which is commensurate with the RRASOR average score relied on in this study.

Measures

RRASOR. Karl Hanson's RRASOR (1997) is a brief actuarial scale designed to predict sexual offense recidivism. The RRASOR consists of four, easily scored items (age—greater or less than 25, number of previous sexual offences, unrelated victims, and male victims), and the test's author reports that it demonstrates moderate predictive accuracy (i.e., receiver operating characteristic [ROC] = .71).

Phallometric test. All offenders in the current study were phallometrically assessed using the age/gender slide protocol, as described in Looman and Marshall (2001). The age/gender assessment consisted of 21 colored slides of single, nude individuals: three adults, three pubescents, and three prepubescent individuals of each gender, as well as three neutral (scenery) slides. In this protocol, a "pedophilia index" is calculated by subtracting standard scores of arousal to prepubescent children from standard scores of arousal to adults. Thus, a score of less than 0 indicates a preference for children, whereas a score above 0 indicates a preference for adults.

DSM-IV-TR. Participants were evaluated regarding presence of pedophilia using a strict application of the *DSM-IV-TR* diagnostic criteria as described above. All diagnoses were made by the first author (RJW) with a subset diagnosed by the third author (JL) for interrater reliability purposes (see below).

"Expert" judgment. The risk assessment literature of the past 25 years has been very clear in stating that clinical judgment alone is often likely to result in inaccuracies with respect to risk prediction. Actuarial instruments such as the RRASOR have been constructed specifically to counter these difficulties and to add a much-needed degree of objectivity to the process. However, it is the authors' observation that despite the general acceptance of the importance of actuarial assessment many experts continue to adjust actuarial assessments using "risk factors outside the actuarials" (Static-99, 2010). This type of assessment is, perhaps, more similar to that performed by expert witnesses in SVP commitment proceedings. It reflects an overall comment as to the offender's level of deviancy, typically anchored in an actuarial assessment, and may be reflective of the assessor's confidence as to the level of sexual offender recidivism risk presented by the person being assessed.

Interrater Agreement

Ratings for the *DSM* diagnoses and expert rating were made for all participants by the first author (RJW). Interrater agreement for these variables was obtained by having the third author (JL) independently rate every fifth offender in the data set on these items ($n = 25$). Note that both raters have more than 20 years of experience in the assessment and treatment of sexual offenders. For *DSM-IV-TR* diagnoses $\kappa = .72$, $p < .000$ and for expert ratings $\kappa = .57$, $p < .004$. These are considered to be substantial and moderate levels of agreement, respectively (Viera & Garrett, 2005).

Procedure

All participants were diagnosed using each of the methods described above. Congruence between the various diagnostic methods was then measured using Cramer's *V* statistic. The ability of each method to predict recidivism (i.e., outcome) was evaluated using χ^2. Deviant arousal, via phallometric testing, was defined as a clear preference for children over adults. *DSM-IV-TR*-based diagnoses were achieved following strict adherence to the criteria outlined above. RRASOR scores were divided into three categories for the purpose of these analyses, such that individuals scoring 1 to 2 were identified as low, individuals with a score of 3 were classified as moderate, and individuals with scores of 4 to 6 were classified as high.

Results

Rater Agreement

Investigation of interrater agreement revealed substantial levels on the *DSM-IV-TR* diagnoses, with agreement on the diagnoses of 22 of 25 offenders. In all cases of disagreement, RJW diagnosed pedophilia whereas JL did not. For the expert ratings, agreement reached only moderate levels. This decrease in reliability is not surprising given that the task involves departing from a structured rating (i.e., RRASOR or *DSM* criteria) and incorporating other information in a less-structured fashion. This process has implications for predictive validity, to be discussed below.

Classification of Offenders

As noted above, offenders were classified as to whether they met the *DSM-IV-TR* diagnostic criteria for pedophilia, deviancy according to phallometric testing, and whether they were judged as deviant according to an "expert rating." For these analyses, phallometric deviance indices were recoded so that participants demonstrating a clear preference (i.e., deviance index of −.25 or less) for children over adults were evaluated as pedophilic; a deviance index between −.25 and +.25 were considered to

Table 2. Cross-tabulation of *DSM-IV-TR* Diagnosis With Phallometrics, RRASOR Category, and Overall Rating

	RRASOR group			Phallometric deviance		Expert rating		
	N (%)			*N* (%)		*N* (%)		
DSM	Low	Moderate	High	Child preference	Nondiscriminating preference	Adult	Deviant	Not pedophilia
No	28 (22.8)	5 (3.3)	4 (3.3)	21 (17.1)	5 (4.1)	10 (8.1)	21 (17.1)	15 (12.2)
Yes	52 (42.3)	10 (8.1)	25 (20.3)	50 (40.7)	10 (8.1)	27 (22.0)	10 (8.1)	77 (62.6)

Note: *DSM-IV-TR* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision); RRASOR = Rapid Risk Assessment of Sex Offender Recidivism.

be indicative of no preference between children and adults (i.e., nondiscriminating); and indices of +.25 and greater were considered to be indicative of an adult preference. In terms of diagnosis, 70.7% of the sample met the *DSM-IV-TR* diagnostic criteria for pedophilia, whereas 57.7% demonstrated preferential deviant arousal to children on phallometric testing. Finally, 23.6% of the sample scored in the high-risk category on the RRASOR, whereas 11.4% scored in the moderate-risk category. In terms of the expert judgment, 74.8% were judged as being sexually deviant.

Measures of Association

In terms of agreement between ratings, 57.5% of those who were classified as pedophiles according to *DSM-IV-TR* criteria were also classified as deviant on phallometrics; however, 58.3% of those who were not classified as pedophiles also had deviant phallometric profiles. Of those offenders classified as pedophiles according to *DSM-IV-TR* criteria, 59.8% scored in the low range on the RRASOR, 11.5% in the moderate-risk range, and 28.7% scored in the high-risk range. For those who were not classified as pedophiles according to *DSM-IV-TR* criteria, the corresponding percentages were 77.8, 11.1, and 11.1. Of those offenders classified as pedophiles according to *DSM-IV-TR* criteria, 88.5% were also classified as sexually deviant according to the expert rating, compared to 41.7% of those not classified as pedophiles. These data are summarized in Table 2.

For the expert ratings, 56.5% of those who were rated as deviant scored in the low range on the RRASOR, whereas 29.3% scored in the high range. For those not rated as deviant, 90.3% scored in the low range, whereas 6.5% scored high. Regarding sexual deviance, 62.0% of those rated as deviant by the expert rating had a preference for children in phallometric testing, whereas 25.0% had an adult preference. Of those not rated as sexually deviant, 54.2% had a child preference in testing.

Analyses were performed using the Cramer's *V* statistic to ascertain whether the four diagnostic techniques used in the current investigation were significantly related.

Table 3. Relationship Among Measures

	RRASOR	*DSM-IV-TR* criteria	Deviant arousal	Expert rating
RRASOR		.19	.10	.31**
DSM-IV-TR criteria			.04	.49***
Deviant arousal				.19
Expert rating				

Note: RRASOR was coded categorically (1 = *Low*; 2 = *Moderate*; 3 = *High*). All measures of association are Cramer's *V* with **$p < .01$. ***$p < .001$. *DSM-IV-TR* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision); RRASOR = Rapid Risk Assessment of Sex Offender Recidivism.

As can be seen in Table 3, *DSM-IV-TR* diagnoses were not associated with any other diagnostic technique with the exception of the expert rating method ($V = .49$, $p < .001$). Given that the expert rating diagnoses were based on all data available (including the *DSM-IV-TR* diagnosis), this result was not unexpected. Expert rating diagnoses were also found to be significantly associated with RRASOR scores ($V = .31$, $p < .003$) but were not associated with deviant arousal to children via phallometric testing ($V = .19$, $p = ns$).

Outcome

In this investigation, outcome was defined as reconviction for any sexual offence according to finger print service records of the Royal Canadian Mounted Police. Of the 130 offenders in the study, 106 were released and eligible for follow-up. Fourteen (13.7%) sexually reoffended over an average follow-up period of 8.8 ($SD = 2.9$) years.

Proportional analyses were conducted as to whether any of the diagnostic methods were associated with sexual offence recidivism. To facilitate analyses, RRASOR data were once again categorized as low, moderate, or high. Results are summarized in Table 3. Individuals scoring high on the RRASOR were significantly more likely to be reconvicted for a sexual offence, $\chi^2(2) = 5.76$, $p < .05$, than expected (Adjusted Standardized residual = 2.4) in comparison to those scoring low (Adjusted Standardized residual = −1.9). High-risk participants based on expert rating diagnosis (though these decisions were presumably based in part on RRASOR scores) were also more likely to recidivate, although not significantly so, $\chi^2(1) = 1.31$, $p = ns$. However, individuals who met *DSM-IV-TR*-based diagnoses of pedophilia were no more likely to be convicted of a new sexual offence than those who failed to meet the *DSM-IV-TR* diagnostic criteria for pedophilia, $\chi^2(1) = 0.69$, $p = ns$. Whether the offender who was phallometrically assessed as having deviant arousal on the age/gender assessment was also unrelated to sexual offence recidivism, $\chi^2(2) = 1.68$, $p = ns$, although a trend was evident with 17.9% of sexually deviant offenders reoffending and only 8.8% of nondeviant offenders committing a new sexual offense. Percentages of offenders reoffending by method of diagnosis are presented in Table 4.

Table 4. Sexual Recidivism and Assessment Type

	No reoffence	Reoffence
	N (%)	*N* (%)
RRASOR*		
Low	63 (90.0)	7 (10.0)
Moderate	8 (89.9)	1 (11.1)
High	13 (68.4)	6 (31.6)
DSM-IV-TR criteria		
No	27 (90.0)	3 (10.0)
Yes	56 (83.6)	11 (16.4)
Deviant arousal		
Adult preference	31 (91.2)	3 (8.8)
Nondiscriminating	10 (90.9)	1 (9.1)
Child preference	46 (82.1)	10 (17.9)
Expert rating		
Not deviant	24 (92.3)	2 (7.7)
Deviant	59 (83.1)	12 (16.9)

Note: *DSM-IV-TR* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision); RRASOR = Rapid Risk Assessment of Sex Offender Recidivism.
*$p < .05$, denoting a statistically significant relationship between score and reoffence.

Table 5. ROC AUC Values for Predictor Variables

		95% Confidence interval	
Test result variable(s)	Area	Lower bound	Upper bound
Age-gender phallometric results	.41	.235	.579
RRASOR	.67*	.498	.839
DSM-IV-TR pedophilia diagnosis	.54	.373	.704
Expert rating	.55	.389	.713

Note: AUC = area under the curve; *DSM-IV-TR* = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision); ROC = receiver operating characteristic; RRASOR = Rapid Risk Assessment of Sex Offender Recidivism.
*$p < .05$.

ROC area under the curve analyses were also conducted with each of the predictor variables, with RRASOR scores and the phallometric testing results treated as continuous variables. Results are displayed in Table 5. As can be seen, only the RRASOR was significantly associated with sexual recidivism.

Finally, incremental validity of the various methods of diagnosis in predicting sexual recidivism in addition to the RRASOR was evaluated by means of a Cox regression analysis. With sexual recidivism (yes/no) as the outcome, the RRASOR score was entered in the first block, and *DSM-IV-TR* pedophilia diagnosis (yes/no), expert rating (deviant vs. not), and phallometrically assessed sexual deviance as a continuous variable were entered in a stepwise fashion on the second block. The only significant predictor was the RRASOR score, $\chi^2(1.91) = 6.34$, $p < .01$; Wald = 5.78, $df = 1$, $p < .016$, E(B) = 1.53. None of the other predictors added significantly to the prediction of sexual recidivism, residual $\chi^2(3.91) = 1.41$, $p = ns$.

Discussion

In the investigation outlined above, we found a considerable degree of disparity in terms of how a child sexual abuser might be diagnosed as pedophilic depending on the method used. Although the expert rating diagnostic method was significantly associated with RRASOR and a strict application of *DSM-IV-TR* criteria, *DSM-IV-TR* was *not associated* with RRASOR or results of phallometric testing indicating preferential deviant arousal to children. The latter finding—no association between *DSM-IV-TR* and phallometry is puzzling, given that these would appear to be the two most common means of diagnosing this condition. Indeed, the fact that an offender who had a deviant profile in phallometric testing was just as likely to be diagnosed as nonpedophilic using *DSM-IV-TR* criteria as he was to be pedophilic suggests that we may need to rethink our conception of what arousal to children in phallometric testing indicates. It appears that the simple presence of "deviant" arousal is not necessarily indicative of the *DSM-IV-TR* constellation of traits associated with pedophilia.

In regard to risk-assessment utility, only the RRASOR scores were significantly predictive of recidivism. Given that expert rating diagnoses included consideration of the items leading to the RRASOR scores, it is somewhat surprising that this method was not also predictive of recidivism. It is possible that since the expert rating took into account each of the three variables explored in this study (RRASOR, *DSM-IV-TR* diagnosis, and phallometrics), the predictive validity was contaminated by consideration of irrelevant information. This provides further support to the contention (see Quinsey et al., 2006) that actuarial models of risk assessment are more reliable than solely clinical or clinically adjusted assessments and is consistent with the finding reported by Hanson and Morton-Bourgon (2009) indicating that the predictive validity of adjusted actuarial assessments is lower than that of actuarial assessments alone. To reiterate, results obtained in this study show that empirically derived composite risk estimates best predict risk for recidivism.

Although the *DSM-IV-TR* and phallometric testing methods are arguably the more prevalent pedo-diagnostic schemes, both were *not predictive* of recidivism. As noted above, we found it particularly curious that deviant sexual interest—as measured by phallometry—was not helpful in determining who would or would not go on to commit future sexual offenses, especially, given that both Hanson meta-analyses (Hanson

& Bussière, 1998; Hanson & Morton-Bourgon, 2005) found deviant sexual interests to be a robust predictor of recidivism. One potential explanation for this may be the small sample size in the current study in combination with the relatively low base rate of sexual reoffence.

We were not surprised to see that *DSM-IV-TR* was not predictive of recidivism, largely because of the difficulties implicit in using 1990s diagnostic criteria without employing additional knowledge derived from 21st-century science and specialized practice. Clearly, aspects of the *DSM-IV-TR* criteria are useful in diagnosing pedophilia, and these aspects were considered in rendering the expert rating diagnoses. One of the greatest difficulties in strictly adhering to the *DSM-IV-TR* criteria is that many persons are either diagnosed or not diagnosed based solely on the "6-month" criterion. Furthermore, the age limit (13) often fails to appreciate the body shape (degree of secondary sex characteristics development) of the offender's target population. We would suggest that evaluators use a physical development scale, like Tanner, which appears to be more useful in this regard (Seto, 2008). It would also appear that attendance to practical empiricism is precisely what the *DSM-V* subworkgroup has in mind regarding its attempts to redefine sexual interest in underage persons (Blanchard, 2009a, 2009b).

Conclusion

One aspect of the current study addressed the issue of whether consideration of information over and above an actuarial assessment of risk was useful in predicting recidivism in child molesters. In particular, we were interested to explore the relevance of the *DSM-IV-TR* diagnostic criteria for pedophilia, phallometric testing (another putative indicator of pedophilia), and expert ratings as both diagnostic indices and risk predictors. It was found that expert ratings, *DSM-IV-TR* diagnosis, and phallometric testing were unrelated to each other and unrelated to recidivism. In regression analyses, the inclusion of these items did not add to prediction of recidivism when the RRASOR scores were already accounted for. Overall, the implications of these findings for forensic practice (e.g., SVP hearings in the United States, Dangerous Offender hearings in Canada) are clear. Based on the results reported here, it would appear that an actuarial-only model of risk consideration results in the most reliable risk assessments, and that additional information potentially clouds the assessment.

Acknowledgments

The authors are thankful to Michele Prinzo and P. Bruce Malcolm for their assistance with an earlier version.

Authors' Notes

An earlier version of this article—without the recidivism follow-up data—was previously presented at a conference held by the Association for the Treatment of Sexual Abusers (see Wilson, Abracen, Picheca, Prinzo, & Malcolm, 2003). The views expressed in this study are those of the

authors and do not necessarily represent the views of the Florida Civil Commitment Center or the Correctional Service of Canada.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

Abracen, J., & Looman, J. (2006). Evaluation of civil commitment criteria in a high risk sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and the Law, 1*, 124-140.

American Psychiatric Association. (2000). *The Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision.* Washington, DC: Author.

Andrews, D. A., & Bonta, J. (2007). *The psychology of criminal conduct* (4th ed.). Cincinnati, OH: Anderson.

Blanchard, R. (2009, October). *Paraphilias vs. paraphilic disorders, pedophilia vs. pedo- and pedohebephilia, and gynephilia vs. fetishistic transvestism.* Paper presented at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

Blanchard, R. (2010). The DSM diagnostic criteria for Pedophilia. *Archives of Sexual Behavior, 39*, 304 -316.

Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*, 118-126.

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

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

Doren, D. M., & Levenson, J. S. (2008). Diagnostic reliability and sex offender civil commitment evaluations: A reply to Wollert (2007). *Sexual Offender Treatment, 4*. Available from http://www.sexual-offender-treatment.org/1-2009_01.html

Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology*, *57*, 1-6.

Freund, K., & Watson, R. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of "Phallometric diagnosis of pedophilia." *Psychological Assessment*, *3*, 254-260.

Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting & Clinical Psychology, 63*, 802-809.

Hanson, R. K. (1997). *The development of a brief actuarial scale for sexual offense recidivism* (User Report 1997-2004). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

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

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*, 169-194.

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163.

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

Hanson, R. K., Morton, K. E., & Harris, A. J. R. (2003). Sexual offender recidivism risk: What we know and what we need to know. *Annals of the New York Academy of Sciences, 989*, 154-166.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sexual offenders* (User Report 1999-2002). Ottawa, Ontario: Department of the Solicitor General of Canada.

Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist–Revised*. Toronto, Ontario, Canada: Multi-Health Systems.

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

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*, 425-448.

Kingston, D. A., Firestone, P., Moulden, H. M., & Bradford, J. M. (2007). The utility of the diagnosis of pedophilia: A comparison of various classification procedures. *Archives of Sexual Behavior, 36*, 423-436.

Levenson, J. S. (2004). Reliability of sexually violent predator civil commitment criteria in Florida. *Law and Human Behavior, 28*, 357-368.

Looman, J., & Marshall, W. L. (2001). Phallometric assessments designed to detect arousal to children: The responses of rapists and child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13*, 3-13.

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*, 117-146.

Marshall, W. L. (2005). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment, 1*, 14-41.

Marshall, W. L. (2007). Diagnostic issues, multiple paraphilias, and comorbid disorders in sexual offenders: Their incidence and treatment. *Aggression and Violent Behavior, 12*, 16-35.

O'Donohue, W., Regev, L. G., & Hagstrom, A. (2000). Problems with the *DSM-IV* diagnosis of Pedophilia. *Sexual Abuse: A Journal of Research and Treatment, 12*, 95-105.

Packard, R. L., & Levenson, J. S. (2006). Revisiting the reliability of diagnostic decisions in sex offender civil commitment. *Sexual Offender Treatment, 1*, 1-15.

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

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC: American Psychological Association.

Seto, M., & Lalumiére, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13*, 15-25.

Static-99. (2010). *Worksheet of risk factors to assess outside the Static-99*. Conference handout available from author RJW.

Viera, A. J. & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37, 360-3.

Wilson, R. J., Abracen, J., Picheca, J. E., Prinzo, M., & Malcolm, P. B. (2003, October). *Pedophilia: An evaluation of diagnostic and risk management methods*. Paper presented at the 22nd Annual Conference of the Association for the Treatment of Sexual Abusers, St. Louis, MO.

Wilson, R. J., & Yates, P. M. (2009). Effective interventions and the Good Lives Model: Maximizing treatment gains for sexual offenders. *Aggression & Violent Behavior, 14*, 157-161.

Wollert, R. (2007). Poor diagnostic reliability, the null-Bayes logic model, and their implications for sexually violent predator evaluations. *Psychology, Public Policy, and Law, 13*, 167-203.

Bios

Robin J. Wilson, Ph.D., ABPP is a board certified clinical psychologist with more than 25 years experience working with sexual offenders. He is presently the Clinical Director of the Florida Civil Commitment Center in Arcadia, FL.

Jeffrey Abracen, Ph.D., C.Psych. is the Chief, Community Correctional Research for the Research Branch of the Correctional Service of Canada. He has also worked as the Chief Psychologist, Central District Parole, Correctional Service of Canada.

Jan Looman, Ph.D., C.Psych. is the Clinical Director of the High Intensity Sex Offender Treatment program at the Regional Treatment Centre (Ontario). He has published research in the area of sexual offender treatment, psychopathy, and risk assessment.

Janice Picheca, Ph.D., C.Psych. is a forensic psychologist working for Correctional Service Canada (Parole) in Toronto, Ontario. She provides community-based assessment and counselling services for offenders on federal conditional release.

Meaghan Ferguson is a doctoral candidate in Health Psychology at York University in Toronto. She has participated in various research projects regarding risk management and sexual offenders.