Running Head: LONG-TERM FOLLOW-UP OF SEXUAL SADISTS

Comparing Indicators of Sexual Sadism as Predictors of Recidivism Among Adult Male Sexual Offenders

Drew A. Kingston

University of Ottawa

Michael C. Seto

Royal Ottawa Health Care Group

Philip Firestone

University of Ottawa

John M. Bradford

Royal Ottawa Health Care Group

Authors' Note

Address correspondence to Drew Kingston or Michael Seto, Integrated Forensic Program, Brockville Mental Health Center, 1804 Highway 2, Brockville, Ontario, Canada, K6V 5W7. Emails: drew.kingston@rohcg.on.ca or michael.seto@rohcg.on.ca.

We would like to thank Susan Curry for her assistance with data extraction from the clinical database. We would also like to thank Grant Harris for his comments on an earlier version of this paper. Preparation of this paper was supported in part by a grant from the Social Sciences and Humanities Research Council awarded to a team that included the second author.

Abstract

Objective: This longitudinal study compared the predictive validity of a psychiatric diagnosis of sexual sadism with three behavioral indicators of sadism: index sexual offense violence, sexual intrusiveness, and phallometrically-assessed sexual arousal to depictions of sexual or nonsexual violence.

Method: 586 adult male sexual offenders were assessed between 1982 and 1992 and were followed for up to 20-years post-release using official criminal records. Assessment information included DSM diagnosis, offense characteristics, phallometric assessment results, and an actuarial risk measure (the Sex Offender Risk Appraisal Guide).

Results: Predictive validity was demonstrated in univariate analyses for the behavioral indicators of sexual sadism (AUCs from .58 to .62) but not psychiatric diagnosis (AUC = .54). Cox regression analyses revealed that phallometrically-assessed sexual arousal to violence was still significantly associated with violent (including sexual) recidivism after controlling for actuarially-estimated risk to reoffend. A psychiatric diagnosis of sexual sadism, in contrast, was unrelated to recidivism.

Conclusions: The results support the use of more behaviorally operationalized indicators of sexual sadism, especially phallometric assessment of sexual arousal, and suggest the DSM criteria for sexual sadism require further work.

Keywords: sexual sadism; paraphilia; sexual offenders; diagnosis; recidivism

Comparing Indicators of Sexual Sadism as Predictors of Recidivism Among Adult Male Sexual Offenders

Sexual sadism is a paraphilia describing individuals who derive sexual pleasure from inflicting pain or humiliation on others. Richard von Krafft-Ebing (1886/1999) initially defined sadism as "the experience of sexual, pleasurable sensations (including orgasm) produced by acts of cruelty, bodily punishment afflicted on one's person or when witnessed by others, be they animals or human beings. It may also consist of an innate desire to humiliate, hurt, wound or even destroy others in order, thereby, to create sexual pleasure in oneself" (p. 109). The concept of sexual sadism was subsequently expanded to include the role of psychological pain in the form of humiliation (Eulenberg, 1911) and the notion of domination and control over another person (Karpmann, 1954). These aspects of sexual sadism are incorporated in the definitions used by the *International Classification of Diseases* (ICD-10) (World Health Organization, 1992) and the *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV-TR; American Psychiatric Association (APA), 2000].

The ICD-10 defines sexual sadism as "a preference for sexual activity which involves the infliction of pain or humiliation, or bondage. If the subject prefers to be the recipient of such stimulation this is called masochism; if the provider, sadism. Often an individual obtains sexual excitement from both sadistic and masochistic activities" (p. 220). The DSM-IV-TR specifies two criteria: "(A) Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving acts (real, not simulated) in which the psychological or physical suffering (including humiliation) of the victim is sexually exciting to the person; (B) The person must have acted on these sexual urges with a non-consenting person, or the sexual urges or fantasies cause marked distress or interpersonal difficulty" (p. 574).

Kendell (1975), Mezzich (2002) and others have argued that a major purpose of nosological diagnosis is to provide a common taxonomy to facilitate communication and categorize individuals into homogeneous groups that convey useful information about prognosis. Meehl (1973) suggested that a reliable and useful diagnosis contains a sufficient "etiological and prognostic homogeneity among patients…..so that the assignment of a patient to this group has probability implications which it is clinically unsound to ignore" (p. 92). More recently, both Spitzer (2001) and Kendell and Jablensky (2003) have suggested that a diagnostic concept should provide information regarding etiology, essential features, and future course of the disorder. Though a nosology such as DSM can consist entirely of highly reliable descriptions of different disorders, its clinical utility is greater if persons assigned to the same diagnostic category have a common course.

The relationship between diagnosis and prognosis among sexual offenders is particularly salient, given the potential for significant harm to others and its relevance to civil commitment laws in the United States and to sentencing laws in other countries (e.g., the Dangerous Offender designation in Canada, resulting in an indeterminate sentence). Twenty-one U.S. jurisdictions have enacted sexual offender civil commitment laws that were designed for the post-sentence commitment of sexual offenders who pose a legally defined risk to reoffend as a result of a mental condition (for a review, see Doren & Elwood, 2009). Consistent with the U.S. Supreme Court decision in *Kansas v. Hendricks* (1997), the criteria for civil commitment are that an individual has a prior history of criminal sexual activity, a legally defined likelihood of reoffense (often but not always set as "more likely than not"), and a mental condition such as sexual sadism or pedophilia that contributes to the likelihood for recidivism (see Doren & Levenson, 2009). Elwood (2009) defined predisposition, as it relates to civil commitment statutes, as: "the

effect of a mental disorder to increase the incidence of sexual recidivism" (pg. 401). He further stipulated that a predisposition that is associated with a particular mental disorder is equivalent to a risk factor that is testable via a statistical association.

All major explanations of sexual offending suggest paraphilias are an important motivation for some sexual offenders (for a review, see Ward, Polaschek, & Beech, 2006). Indeed, there is considerable evidence linking paraphilias, especially pedophilia, with recidivism among sexual offenders (for reviews, see Hanson & Morton-Bourgon, 2005; Lalumière, Harris, Quinsey, & Rice, 2005; Seto, 2008). The relationship between sexual sadism and risk to reoffend, however, is less clear (see Kingston & Yates, 2008). There is evidence that some correlates of sexual sadism predict sexual and violent recidivism, such as sexual arousal to depictions of sexual violence and the seriousness of the sexual offense (Hanson & Morton-Bourgon, 2004; Lalumière et al., 2005). But the degree of force and the degree of sexual intrusiveness during the sexual offense have been differentially associated with recidivism based on the type of recidivism outcome investigated. In a meta-analysis, Hanson and Morton-Bourgon (2004, 2005) found that the degree of force exhibited during the sexual offense was significantly associated with sexual recidivism, although the size of the effect was trivial (d = .09; 95% CI = .02 to .16). The degree of force was also significantly associated, and produced larger effects, for violent, non-sexual recidivism (d = .35; 95% CI = .22 to .47) and any violent recidivism (d = .22; 95% CI = .15 to .29). The degree of sexual intrusiveness was positively associated with violent, non-sexual recidivism (d = .36; 95% CI = .17 to .55) and produced a very small effect for any violent recidivism (d = .05; 95% CI = -.07 to .18). Sexual intrusiveness was negatively associated with sexual recidivism (d = -.17; 95% CI = -.29 to -.05), although this was likely due to the inclusion of studies with hands-off offenders (e.g., exhibitionists), as these individuals are

typically more likely to re-offend when compared to sexual offenders with solely hands-on offenses (Firestone, Kingston, Wexler, & Bradford, 2006; Hanson & Morton-Bourgon, 2004).

Problems associated with subjective judgments such as those used in making psychiatric diagnoses were identified decades ago by Meehl (1954), and different authors have identified problems in the definition, operationalization, and assessment of sexual sadism (see Kingston & Yates, 2008; Marshall, 2006; Marshall & Kennedy, 2003). For example, the terms "recurrent" and "intense" in Criterion A have been interpreted differently across evaluators, the motivation for behavior involving the infliction of pain must be inferred, and whether an act is intended or experienced as humiliating is subjective (see Doren, 2002; O'Donohue, Regev, & Hagstrom, 2000). There is accumulating evidence of specific problems with the reliability of paraphilia diagnoses, including sexual sadism (Kingston, Firestone, Moulden, & Bradford, 2007; Levenson, 2004; Moulden, Firestone, Kingston, & Bradford, 2009).

Marshall, Kennedy, and Yates (2002) found that sexual offenders diagnosed as sexually sadistic did not differ from those who did not receive the diagnosis on features identified as important aspects of sexual sadism among sexual offenders in the clinical literature (e.g., torturing victims, sexually violent fantasies, sexual arousal to depictions of rape); in fact, the nominally nonsadistic offenders scored higher on some of these variables. This study also found a lack of agreement across experienced diagnosticians on the features they considered to be important to the diagnosis of sexual sadism, as well as inconsistency in decisions made by the same diagnostician. In another study, Marshall, Kennedy, Yates, and Serran (2002) provided forensic psychiatrists with information on random samples of offenders who were diagnosed as sadists in the previously cited study. The reliability of diagnosis among 15 respondents was low (kappa = .14). Inadequate reliability was also reported by Levenson (2004) for sexual sadism

(kappa = .30). These findings, however, may underestimate the reliability of diagnosis as a result of the statistical procedures used (Packard & Levenson, 2006) or because of the quality of information available to the evaluators and increased scrutiny in adversarial situations (Doren & Elwood, 2009).

Poor reliability constrains validity and thus constrains the clinical utility of these diagnoses. Given the problems identified above, attempts have been made to identify sexual sadism among sexual offenders using more behaviorally operationalized features (see Marshall & Hucker, 2006; McLawsen, Jackson, & Vannoy, 2008; Nitschke, Osterheider, & Mokros, 2009). Candidate variables include use of a weapon, victim confinement, bondage, object insertion, torture, and post-death mutilation (Marshall & Kennedy, 2003). Proulx, Blais, and Beauregard (2006) compared 43 sadistic and 98 nonsadistic sex offenders, classified using the Massachusetts Treatment Center rapist typology (Knight & Prentky, 1990). Sadistic offenders were more likely to use weapons, engage in excessive violence, engage in bondage, confine victims, insert objects into the victim's vagina, torture, and mutilate. These findings have been replicated in subsequent investigations (Beauregard & Proulx, 2002; Gratzer & Bradford, 1995).

There is a potential tautology in these studies, however, as these same offense features may have been used to classify offenders as sexual sadists or nonsadists. Nonetheless, these results do suggest that offense features can be reliably identified, which is a first step towards demonstrating their validity. Showing that these offense features have discriminative validity by distinguishing sexually sadistic and nonsadistic offenders (classified according to other criteria), concurrent validity by being significantly correlated with other indicators of sexual sadism, and predictive validity (by predicting whether an offender commits another act of sexual violence in the future) would support their construct validity (Cronbach & Meehl, 1955). Similarly, showing

that the DSM diagnosis of sexual sadism has discriminative, concurrent, and/or predictive validity would support its clinical utility and suitability for consideration in high-stakes legal proceedings.

Another approach to identifying sexual sadists involves the assessment of sexual arousal to depictions of sexual or nonsexual violence using measures of penile response. Some studies have found that sexual arousal to depictions of sexual or nonsexual violence differentiates sadistic and nonsadistic rapists (Barbaree, Seto, Serin, Amos, & Preston, 1994; Fedora et al., 1992; Proulx, Aubut, McKibben, & Cote, 1994), whereas others have found no difference between groups (Langevin et al., 1985; Seto & Kuban, 1996). One study found that nonsadistic sexual offenders actually displayed greater arousal to stimuli depicting sexual violence than did sadistic offenders (Marshall, Kennedy, et al., 2002). Seto and Kuban (1996) found that self-identified sadists recruited from the community could be distinguished from sadistic or nonsadistic rapists, suggesting that prior results may be partially explained by the ability and motivation of sexual offenders to disguise their sexual arousal patterns. Other possible explanations for the inconsistent results are differences in how sadistic sexual offenders were identified and differences in the stimulus sets that were used. Stimulus sets that are more brutally violent produce greater discriminant validity, yet stimulus sets in use today vary greatly in their level of violence (see Lalumière et al., 2005).

The Present Study

The current investigation examined the predictive validity of sexual sadism, as indicated by psychiatric diagnosis, level of violence during the most recent offense, the intrusiveness of the sexual activity, and phallometrically-assessed sexual arousal to depictions of sexual or nonsexual violence. As noted earlier, demonstrating concurrent validity and predictive validity (such as

predicting further sexual violence) would support the construct validity of these different approaches to identifying sexual sadists. We hypothesized that: (a) all sadism indicators would predict recidivism among sex offenders, in line with both theories and empirical findings about the role of atypical sexual interests in sexual offending (Hanson & Morton-Bourgon, 2005; Lalumière et al., 2005; Seto, 2008); (b) the three behaviorally operationalized indicators (level of violence, sexual intrusiveness and phallometrically-assessed sexual arousal) would be better predictors than psychiatric diagnosis, in line with decades of research on the advantages of specific criteria over subjective, global judgments, and more recent work identifying problems with the reliability of the diagnosis of sexual sadism; and, (c) the indicators would be positively and significantly correlated with each other.

Method

Participants

Participants were 586 adult men who had been convicted of a contact sexual offense and were assessed between 1982 and 1992 at a predominantly outpatient sexology clinic that conducts assessments on men and women with problematic sexual behaviors or interests. All participants were assessed just prior to or just after their court appearance or sentencing. This sample has been previously examined in other studies (Firestone, Bradford, Greenberg, Larose, & Curry, 1998; Firestone et al., 1999). The follow-up period has been extended recently and, as such, some participants were lost to our follow-up as a result of death or deportation from Canada (McCoy, 1997; Wexler, 2005).

The sample consisted of 295 (50%) intrafamilial offenders against a child, 205 (35%) extrafamilial offenders against a child, and 86 (15%) rapists. Offenders with mixed victim types were not available in the database. The average age of the sample was 38.1 years (*SD* = 12.0,

range: 18 – 78 years) and approximately 67% of the participants reported that they had, at one time, been married or lived in a common-law relationship. The average education level was 10.8 years (*SD* = 3.6 years). Twenty-three percent had previous charges or convictions for sexual offenses, 37% had previous violent (including sexual) offenses, and 53% had prior criminal offenses. The sample was average in risk on the Sex Offender Risk Appraisal Guide (*M* = 3.35; *SD* = 2.21; Quinsey, Harris, Rice, & Cormier, 2006). Additional details about this sample are available from McCoy (1997) and Wexler (2005).

Procedure

All participants signed a consent form at the time of their assessment permitting use of their data for research; this research was approved by the institutional ethics board. Each participant was first interviewed by a psychiatrist who specialized in working with sexual offenders and who provided a DSM diagnosis, if applicable. The psychiatrist had access to police and clinical reports, which included information about criminal history, psychosocial history, and any previous assessment results. Participants were then assessed in the phallometric laboratory.

Measures

Actuarial risk score. The Sex Offender Risk Appraisal Guide (SORAG) was developed as a modification of the Violence Risk Appraisal Guide (Quinsey et al., 2006), with the addition or revision of items that predict violent recidivism by sexual offenders. The measure consists of 14 items assessing child and adolescent adjustment, criminal history, psychopathy, and atypical sexual interests: Living with both biological parents until age 16, elementary school maladjustment, history of alcohol problems, marital history, extent of nonviolent offense history, extent of violent offense history, previous sexual offense history, sex and age of index sexual victim, failure on prior conditional release, age at index offense, met DSM criteria for any

personality disorder, met DSM criteria for schizophrenia, phallometrically-measured atypical sexual arousal, and Psychopathy Checklist – Revised score. SORAG scores can range from - 27 to + 51, which can be divided into nine equal-sized risk "bins". The predictive validity of the SORAG for both sexual and violent recidivism has been supported in a variety of studies (e.g., Hanson & Morton-Bourgon, 2009; Harris & Rice, 2003; Kingston, Yates, Firestone, Babchishin, & Bradford, 2008).

Diagnosis. The version of DSM used for diagnosis depended on the year of assessment; criteria from both the DSM-III (APA, 1980, *n* = 187) and the DSM-III-R (APA, 1987, *n* = 397) were used. These earlier versions of DSM are similar to the current edition (DSM-IV-TR) in that sadism is characterized by the infliction of psychological or physical suffering of another person in order to achieve sexual excitement. One notable difference was that DSM-III criteria included psychological or physical suffering toward a consenting partner with or without associated distress, whereas later versions only applied to nonconsenting persons or behavior that caused clinical distress or impairment. Diagnoses were coded as present or not present. Inter-rater reliability for these diagnoses was not available as diagnosis was made only by the evaluating psychiatrist.

Offense features. Offense features included a measure of the violence of the sexual offense and the intrusiveness of the sexual act in the index sexual offenses that brought the individual to the sexology clinic. The level of violence and the intrusiveness of the sexual act were rated by the psychiatrist using behaviorally anchored scales and were intended to measure violence that was excessive; that is, beyond what might have been used to overcome victim resistance. Level of violence was rated on a 10-point scale: no force or violence (0), threat of assault with no weapon (1), threat of assault with weapon (2), minor injury with no weapon (3),

minor injury with weapon (4), severe beating with no weapon (5), severe beating with weapon (6), potential homicide because the injuries could have caused death without successful medical intervention (7), homicide (8), and homicide with post-death mutilation (9). The intrusiveness of the sexual act was rated on a 6-point scale: no sexual intrusiveness (0), verbal threat (1), attempt (2), touching (3), penetration (4), and sexual assault with excessive violence (5).

Sexual arousal. Participants were tested in a private sound-proof room. Voice communication was made using an intercom system from the adjoining control room. Penile responses were recorded at one second intervals using an indium-gallium penile strain gauge from Farrell Instruments (Behavioural Technology Inc., Salt Lake City, UT, 84103). The testing procedure and stimulus sets used have been described in other peer-reviewed publications (e.g., Bradford & Pawlak, 1993; Firestone, Bradford, Greenberg, & Nunes, 2000). The order of stimulus presentation, held constant for all participants, was computer-controlled. Participants were presented with one or more of three series of audiotapes, consistent with their hypothesized sexual preferences. Thus, the length of sessions and the number of stimuli presented varied across participants.

The audiotaped battery consisted of vignettes of approximately two minutes' duration describing sexual activity with a person varying with respect to age, sex, and degree of coercion and violence portrayed (Abel, Blanchard, & Barlow, 1981). The female child series consisted of descriptions of sexual activity with a female person for eight categories. The male child series consisted of eight corresponding vignettes involving a male person, but also included one scenario involving an adult female partner. Each vignette was followed by 20 seconds where sexual arousal continued to be measured prior to the start of the next audiotaped segment. If the individual did not return to the baseline measure of arousal during this interval, the next stimuli

would not be started until the patient was within approximately 2 mm of the baseline measure.

For each of the female child and male child series, two scenarios for each category were included. Categories were as follows: (a) non-physical coercion of child, (b) physical coercion of child, (c) violent sex with child, (d) nonsexual assault of child, (e) consenting sex with female adult, and (f) sex with female child relative, (g) child initiates, and (h) child mutual. The latter two categories described the sexual offense taking place with no overt form of coercion. The audiotape series used to assess sexual arousal to coercive sexual activity toward an adult included two scenarios of two-minute duration for each of three categories: (a) consenting sex with adult female, (b) rape of adult female, (c) nonsexual assault of adult female.

The *Pedophilia Assault Index* was calculated by dividing the highest response to an assault stimulus involving a child victim (non-physical coercion of child, physical coercion of child, sadistic sex with child, or nonsexual assault of child) by the highest response to a child stimulus with no overt from of coercion. The *Rape Index* was calculated by dividing the response to the rape stimulus by the response to the adult-consenting stimulus. The *Adult Assault Index* was calculated by dividing the highest response to a nonsexual assault stimulus (against an adult) by the highest response to an adult consenting stimulus. For these indices, scores greater than one indicate greater arousal to the deviant category (assault of child, rape, or nonsexual assault). Because the database we used was archival and included only these indices, without raw scores, we were not able to calculate indices based on the differences between standardized *z* scores, even though differential indices tend to produce higher validity coefficients (for a review, see Lalumière & Harris, 1998). Still, indices based on ratios of raw responses have demonstrated good validity in previous studies (e.g., Firestone et al., 2000).

In the present sample, average scores on the Pedophilia Assault Index were 0.70 (*SD* =

0.81) for stimuli depicting a female victim and 0.85 (*SD* = 0.94) for stimuli depicting a male victim. The average score in the present sample for the Rape Index was 0.60 (*SD* = 0.97) and the average score for the Adult Assault Index was 0.29 (*SD* = 0.53). Among the entire sample, 44.7% (n = 262) demonstrated equal or greater arousal to coercive sex than to mutually consenting sex. Nonetheless, the average index values suggest that, as a group, the offenders responded more to adult consenting stimuli than to stimuli depicting coercive sex or nonsexual violence. For the analyses reported below, we created a new index of sexual arousal to (sexual or nonsexual) violence, irrespective of age, that was simply the highest score from any of the three indices noted above.

Recidivism. Recidivism information was obtained in 2002 from a national database of criminal arrests and convictions maintained by the Royal Canadian Mounted Police. The dependent variables (recidivism outcomes) in this study were organized in a nested hierarchical manner, as follows: Sexual recidivism was defined as any charge or conviction for a sexual offense; violent (including sexual) recidivism was defined as any charge or conviction for a nonsexually violent or contact sexual offense; and any criminal recidivism, which included any new criminal charge or conviction. This method of coding recidivism is consistent with past research (Kingston, Firestone, Wexler, & Bradford, 2008), and coding violent recidivism this way captures sexual offenses that were "pled down" to nonsexually violent charges and sexually motivated offenses that resulted in nonsexually violent charges, such as a sexually motivated homicide that was charged as first degree murder (Rice, Harris, Lang, & Cormier, 2006).

Statistical Analyses

Independent-sample t-tests and chi-square analyses were used to compare sexual offenders diagnosed as sexually sadistic from those who were not so diagnosed on demographic

and offense history variables. Different effect size indicators were reported, chosen based on the underlying properties of the statistic and the utility over other commonly used indicators when used with a particular statistical test (e.g., Howell, 2002; Rice & Harris, 2005). The simplest indicator representing the magnitude of effect in the present study was Pearson's correlation coefficient; it is commonly considered that values of approximately .1, .3, and .5, represent small, medium, and large effects, respectively. Cohen's *d* was used for the continuous comparisons and it is generally accepted that corresponding values of .20 are small, .50 are medium, and .80 are large (Cohen, 1988; 1992). Cramér's V, which is equivalent to phi in 2 X 2 contingency tables, was used for categorical comparisons; values of .1, .3 and .5, represent small, medium, and large effects (Cohen, 1992).

Predictive accuracy was evaluated in two ways. First, the area under the curve (AUC) of the receiver operating characteristic (ROC) was calculated. AUC values have an advantage over other common indices of predictive accuracy (e.g., Pearson's correlation coefficient), as they are less affected by recidivism base rates or selection ratios (Rice & Harris, 1995, 2005; Swets, 1986), although they are also influenced by the amount of variability in the putative predictor (Humphreys & Swets, 1991). AUC values, which can range from 0 to 1, can be interpreted as the probability that a randomly selected recidivist has a higher score on a particular measure than a randomly selected non-recidivist. A value of 1 represents perfect prediction, while a value of .5 indicates chance prediction. For descriptive purposes, AUC values of .56, .64 and .71 can be described as small, medium, and large, respectively (Rice & Harris, 2005).

In order to detect significant differences between two AUC values, we determined the critical ratio z, the ratio of a difference score to the standard error of the difference score using

the formula shown below. This method is particularly appropriate when both areas are correlated (Hanley & McNeil, 1983).

$$Z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2r\,SE_1 SE_2}}$$

where A_1 and SE_1 refer to the AUC value and standard error for the first measure, A_2 and SE_2 refer to the AUC value and standard error for the second measure, and *r* refers to the estimated correlation between A_1 and A_2.

Additionally, Cox regression analyses were conducted to evaluate the unique contribution of sadism indicators as predictor variables on recidivism. Cox regression estimates relative risk ratios (hazard rates) and controls for time-at-risk and other important covariates (e.g., actuarially-estimated risk to re-offend). Cox regression analyses produces an exponent, reported as Exp(β), that can be interpreted as a rate ratio, defined as the change in recidivism rate for each unit change in the predictor variable. For categorical predictors such as DSM diagnosis, Exp(β), is the ratio of the estimated hazard for a case with the characteristic to that of a case without the characteristic (i.e., relative risk).

Results

Recidivism

The follow-up period began upon release to the community and ranged up to 20 years, with an average time-at-risk of 10.6 years (*SD* = 4.3 years). The overall rates of recidivism in this study were 16.7% for sexual recidivism, 27.5% for violent (including sexual) recidivism, and 37.4% for any criminal recidivism.

Sexually Sadistic Offenders

This study examined four possible indicators of sexual sadism: DSM diagnosis (coded as

present or not present), level of violence during the index offense, the intrusiveness of the sexual acts, and phallometrically-assessed sexual arousal to violence. Overall, 8.5 % (n = 50) of the sample were diagnosed with sexual sadism. Table 1 displays differences between sexually sadistic and non-sadistic offenders classified on the basis of DSM diagnosis with regard to age at time of assessment, offense characteristics, offense history, highest score on the phallometric index, type of index offense, and actuarial risk score.

Sexually sadistic offenders were significantly younger (d = .86; 95% CI = 0.56 – 1.16) had exhibited significantly greater violence (d = 1.15; 95% CI = 0.82 – 1.49) during the index offense and obtained higher scores on phallometrically-assessed sexual arousal to violence (d = .60; 95% CI = 0.30 – 0.89) when compared to non-sadistic sexual offenders. Sexually sadistic offenders also received higher scores on the SORAG (d = 1.51; 95% CI = 1.15 – 1.86) when compared to non-sadistic offenders. There was also a significant difference with regard to offender type, χ^2 (df = 2, n = 584) = 117.1, p < .001, Cramér's V = .45. Specifically, individuals who received a diagnosis of sexual sadism predominantly offended against an unrelated adult victim (66%), followed by an unrelated child victim (24%), and were least likely to have offended against a related child victim (10%).

In order to further test the relationship between type of sexual offender and a diagnosis of sexual sadism, differences between observed and expected values within each cell (i.e., the adjusted standardized residuals) were examined. Statistical departures from independence were noted among rapists and intra-familial child molesters: Individuals who had offended against an adult female victim were more likely to have received a diagnosis of sexual sadism (z = 10.7) while individuals who offended against a related child victim were less likely to receive a diagnosis of sexual sadism (z = -5.9) than would be expected by chance.

Insert Table 1 about here

Intercorrelations among the Indicators of Sexual Sadism and Recidivism

Pearson correlation coefficients were calculated to examine the relationships among the four indicators of sexual sadism, actuarially-estimated risk as measured by the SORAG, and recidivism outcomes. As shown in Table 2, the DSM was positively and significantly associated with the behaviorally operationalized indicators of sexual sadism. The indicators were also positively and mostly significantly correlated with each other. The association between diagnosis and recidivism was small (r = .09); the binomial effect size display (see Rosenthal, 1990) indicated that the effect between a diagnosis of sexual sadism and recidivism corresponded to a 9% difference in the base rate of recidivism between diagnosed sadists and non-sadists.

The behavioral indicators of sexual sadism were generally positively and significantly correlated with sexual, violent (including sexual), and any criminal recidivism. Actuarial risk score (SORAG) was positively and significantly associated with a diagnosis of sexual sadism, the behaviorally-operationalized measures of sexual sadism, as well as sexual, violent (including sexual) and any criminal recidivism.

Insert Table 2 about here

Predictive Validity

Table 3 presents the AUC values and 95% confidence intervals for the four indicators of sexual sadism in relation to sexual, violent (including sexual), and criminal recidivism. Results

indicated that the level of violence during the index offense, sexual intrusiveness, and phallometrically-assessed sexual arousal to violence predicted sexual, violent, and any criminal recidivism. Although statistically significant, the AUCs were lower than those typically obtained for actuarial risk measures comprising multiple variables (Hanson & Morton-Bourgon, 2009). A DSM diagnosis of sexual sadism was not significantly predictive of any type of recidivism. When comparing ROC indices using the formula noted earlier, the level of violence during the index offense was significantly more accurate than the DSM in predicting violent (including sexual) recidivism ($z = 2.18$) and any criminal recidivism ($z = 2.24$).

Insert Table 3 about here

A series of Cox regression survival analyses were used to examine the unique contribution of the sadism indicators, after controlling for time-at-risk and risk to re-offend. Estimated hazard rate ratios (Exp[β]), confidence intervals, and regression coefficients are presented in Table 4. SORAG score was entered as a covariate in the first block of each analysis, to control for risk. Following this, the sadism indicators were entered as a block: that is, the behaviorally operationalized indicators of sexual sadism (i.e., level of violence, sexual intrusiveness and phallometrically-assessed sexual arousal to violence) and the dichotomously coded DSM diagnosis of sexual sadism. The dependent variables were sexual and violent (including sexual) recidivism, given the importance of these outcomes for dispositional purposes such as civil commitment or indeterminate sentencing. With regard to sexual recidivism, the SORAG was significantly related to outcome, χ^2 (1, $N = 368$) = 19.02, $p < .001$. Specifically, a one unit increase on the SORAG increased the hazard rate by 24% ($e^{.22}$). In the second block,

the additional predictor variables (i.e., indicators of sexual sadism) did not significantly add to the predictive equation, after controlling for risk to re-offend, χ^2 (4, N = 368) = 4.46, *ns*.

With regard to violent (including sexual) recidivism, the SORAG was significantly related to outcome, χ^2 (1, N = 369) = 35.30, $p < .001$. Specifically, a one unit increase on the SORAG increased the hazard rate by 26% ($e^{.23}$). In the second block, the indicators of sexual sadism significantly added to the predictive model, after controlling for risk, χ^2 (4, N = 369) = 12.08, $p < .05$. Among the sadistic indicators, only sexual arousal to violence was significantly associated with the outcome, such that a one unit increase in phallometrically-assessed sexual arousal increased the hazard rate by 32% ($e^{.28}$).

Insert Table 4 here

Discussion

Consistent with our hypotheses, indicators of sexual sadism predicted sexual and violent recidivism and the strength of these associations were stronger for more behaviorally operationalized indicators than for DSM diagnosis. Phallometrically-assessed sexual arousal to violence, in particular, added to the prediction of violent (including sexual) recidivism after actuarially-estimated risk to reoffend was taken into account in multivariate analysis. This is perhaps a surprising result, because phallometrically-assessed atypical sexual arousal is already included as an item on the SORAG. Unless standardized differential index scores are available, this item is scored dichotomously as +1 if any phallometric test indicates deviant sexual arousal and -1 if no test indicates deviant sexual arousal. The indices we used in this analysis contain more information about deviant sexual arousal than the dichotomously scored SORAG item and

thus could add to the prediction provided by the SORAG as a whole.

These results replicate and extend past research suggesting validity problems with the DSM diagnosis of sexual sadism. In conjunction with studies showing problems with inter-rater reliability and debate about the core aspects of sexual sadism and how it relates to sexual assault (see Krueger, in press; Thornton, in press), our results raise questions about the clinical utility of the DSM diagnosis of sexual sadism. This appears to be a problem for other paraphilia diagnoses as well, as both Moulden et al. (2009) and Wilson, Abracen, Picheca, Malcolm, and Prinzo (2003) have found that a diagnosis of pedophilia was unrelated to sexual recidivism, even though multiple studies have shown that correlates of pedophilia, such as phallometrically-assessed sexual arousal to children or victimizing male children, are significant predictors of sexual recidivism (Hanson & Morton-Bourgon, 2005).

Some might wonder whether it is necessary for DSM diagnosis to show predictive validity. It is not necessary for the stated primary purpose of the DSM, which is clinical communication using agreed upon definitions. Moreover, groups distinguished on the basis of DSM diagnosis of sexual sadism were significantly different from each other in level of index offense violence and phallometric responding, and the comparison approached statistical significance for sexual intrusiveness. Thus, our results did provide some evidence for the concurrent validity of DSM diagnosis. However, our results do challenge the idea that DSM diagnosis is important when considering predisposition or risk to sexually offend (see Elwood, 2009).

The most likely explanation for the poor prediction provided by DSM diagnosis is that the subjective judgment involved in applying the diagnostic criteria incorporates information that is unrelated to risk for recidivism, and this in turn constrains the predictive validity that can be

obtained. To illustrate, we believe that level of violence in the index offense is a relevant indicator of sexual sadism (see Gratzer & Bradford, 1995; Proulx et al., 2006). The DSM criteria, however, require an inference to be made as to whether the violence is intended to cause physical suffering that is sexually arousing to the offender, as opposed to other possible motivations (e.g., gratuitous violence because the offender is angry at the victim). In contrast, simply rating the level of violence in the index offense had a significant relationship with violent or specifically sexual recidivism in the univariate analyses.

Our findings do not support the idea that behavioral indicators should be relied upon in making determinations of risk to reoffend. Neither level of violence nor sexual intrusiveness added to the prediction provided by the SORAG, and the SORAG was positively and significantly correlated with all of the sadism indicators, especially DSM diagnosis and level of violence, suggesting the association between sadism indicators and the recidivism outcomes is already captured by this actuarial risk measure. Nonetheless, because of its relationship with violent recidivism and relevance for treatment and management decisions, as shown here and in previous studies, sexual sadism is an important construct to consider when evaluating adult male sexual offenders. The current study suggests that behaviorally operationalized measures are preferred over psychiatric diagnosis, which has implications for how assessments of sexual sadism are conducted in sex offender evaluations. For example, adding behavioral criteria akin to those used for nonsexual disorders that are commonly encountered in forensic settings, such as conduct disorder or antisocial personality disorder, could improve the reliability and thus the validity of paraphilia diagnoses. We discuss one approach at the end of the Discussion.

Study Limitations

The first limitation to the present study reflected the fact that DSM diagnoses were made

only by the evaluating psychiatrist, and thus varied in reliability across clinicians. We did not examine the predictive validity of diagnosis under optimal conditions. In addition, we did not subject these diagnoses to scrutiny by other evaluators, unlike the diagnoses presented in high-stake arenas such as civil commitment hearings (see Doren & Elwood, 2009). At the same time, many clinicians conduct evaluations alone and thus our analysis may be more ecologically valid than an alternative study design that used multiple evaluators and strived to attain high inter-rater reliability. It is also worth noting that the behavioral ratings of offense characteristics – level of violence and sexual intrusiveness—were also made by the same clinician, and these variables were significantly associated with recidivism outcomes in the univariate analyses, whereas DSM diagnosis was not.

A related limitation is the low base rate of sexual sadism diagnosis (8.5%). This low base rate limited our statistical power to detect significant effects for DSM diagnosis. It is possible that we could have detected a unique contribution of DSM diagnosis to the prediction of recidivism outcomes if the base rate was higher. This limitation is less relevant to our univariate analysis because ROC analysis is less sensitive to low base rates than correlational analyses.

Another potential limitation is that we did not directly examine the utility of DSM-IV-TR criteria in this study, as the length of follow-up time required to examine recidivism meant diagnoses were made using DSM-III or DSM-III-R criteria. Although diagnostic criteria for sexual sadism have been similar across versions of DSM, it was noted earlier that DSM-III criteria included psychological or physical suffering toward a consenting partner with or without associated distress, whereas later versions only applied to nonconsenting persons or behavior that caused clinically significant distress or impairment. In our study, individuals diagnosed with DSM-III were more violent and sexually intrusive during their index offense when compared to

individuals diagnosed with the DSM-III-R; these differences were small to moderate (d = 0.50 and d = 0.32 for level of violence and sexual intrusiveness, respectively). There was no difference in degree of sexual arousal to violence between individuals diagnosed with the two different versions of DSM. It is interesting to note that the proposed DSM-V criteria for Sexual Sadism are very similar to post DSM-III versions of the criteria, suggesting our findings will continue to have relevance (Krueger, in press).

Additionally, level of violence and sexual intrusiveness were rated based on the most recent (i.e., index) offense. This means offenders who had engaged in higher levels of violence or intrusiveness in prior sexual offenses might not have received higher scores on these ratings, depending on their index offense. The classification of sadistic offenders as nonsadists on the basis of our decision rule – which reflected the fact that there was usually less details available about prior offenses, compared to the most recent offense – would be expected to attenuate the strength of the relationships we found between the two offense features (level of violence and sexual intrusiveness) and the recidivism outcomes.

Finally, we did not have access to the raw scores from phallometric testing in the data set we used for this study, and so could only report the indices that were available. The Pedophilia Assault Index included sexual arousal to depictions of nonphysical coercion of children in its calculation. Arousal to nonphysical coercion of a child is relevant to pedophilia but is less relevant to the question of whether someone has sexually sadistic interests in coercive sex or violence. Inclusion of nonphysical coercion of children likely attenuated the strength of the relationships that we observed for the phallometric data.

Future Directions

Our results support increasing the availability and use of phallometric assessments of

sexual arousal to violence, given the predictive value of phallometrically-assessed sexual arousal. Developing stimulus sets that disentangle sexual arousal to violence from sexual arousal to nonconsent could lead to better discrimination and identification of sexual sadism. Sexual violence stimuli that are currently used in forensic assessments usually include both kinds of cues (e.g., in the description of a violent rape). Recently, Seto, Lalumière, Harris, and Chivers (2009) developed stimulus sets designed to disentangle the dimensions of coercion, injury, and sexual activity and results have supported the ability of the stimuli to discriminate self-identified sadists from nonsadistic controls through their sexual arousal profiles. The results of this study suggest that the critical cues for sadistic arousal are violence and physical injury.

An option to improve the reliability and validity of diagnosing sexual sadism is to develop a scale based on objective indicators that are uniquely related to sexual and violent recidivism among sexual offenders. For pedophilia, Seto and Lalumière (2001) developed a brief four-item scale of offense characteristics (having a boy victim, having multiple child victims, having younger child victims, and having unrelated child victims) that is significantly correlated with phallometrically-assessed sexual arousal to children and to violent (including sexual) recidivism among sexual offenders with child victims (Seto, 2008; Seto, Harris, Rice, & Barbaree, 2004). There is also evidence for this scale's criterion-related validity among adolescent sexual offenders (Seto, Murphy, Page, & Ennis, 2003).

With regard to sexual sadism, Marshall & Hucker (2006) created a 17-item sadism scale using predominantly official criminal record data, thereby minimizing the emphasis on subjective, global inferences. There has been recent support for this scale's reliability: Nitschke et al. (2009) found that 11 of the 17 items formed a highly reliable Guttman-type severe sadism scale. Moreover, they achieved almost perfect discrimination using this scale between a group of

50 sexual sadists and 50 nonsadists, diagnosed on the basis of consensus decisions using DSM-IV-TR criteria. The ability of this severe sadism scale to predict recidivism has not yet been examined. The results of the present study support the use of more behaviorally operationalized indicators in a sadism scale intended to assist in determining prognosis.

References

Abel G. G., Blanchard E. B., & Barlow D. H. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. *Behavior Research and Therapy, 19*, 25-33. doi:10.1016/0005-7967(81)90109-1

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Barbaree H. E., Seto, M. C., Serin, R. C., Amos N. L., & Preston D. L. (1994). Comparisons between sexual and nonsexual rapists subtypes: Sexual arousal to rape, offense precursors, and offense characteristics. *Criminal Justice and Behavior, 21*, 95-114. doi:10.1177/0093854894021001007

Beauregard, E., & Proulx, J. (2002). Profiles in the offending process of nonserial sexual murderers. *International Journal of Offender Therapy and Comparative Criminology, 46*, 386-399. doi:10.1177/0306624X02464002

Bradford, J. M., & Pawlak, A. (1993). Double-blind placebo crossover study of cyproterone acetate in the treatment of the paraphilias. *Archives of Sexual Behavior, 22*, 383-402. doi:10.1007/BF01542555

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin,* 112, 155-159. doi:10.1037/0033-2909.112.1.155

Cronbach L. J., & Meehl P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. doi:10.1037/h0040957

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

Doren D. M & Elwood, R. W. (2009). The diagnostic reliability of sexual sadism. *Sexual Abuse: A Journal of Research and Treatment*, 21, 251-261. doi:10.1177/1079063209342072

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

Elwood, R. W. (2009). Mental disorder, predisposition, prediction, and ability to control: Evaluating sex offenders for civil commitment. *Sexual Abuse: A Journal of Research and Treatment, 21*, 395-411. doi:10.1177/1079063209347723

Eulenberg, A. (1911). *Sadism and masochism*. New York, NY: Bell.

Fedora, O., Reddon, J. R., Morrison, J. W., Fedora, S. K., Pascoe, H., & Yeudall, L. (1992). Sadism and other paraphilias in normal controls and aggressive and nonaggressive sex offenders. *Archives of Sexual Behavior, 21,* 1-15. doi:10.1007/BF01542713

Firestone, P., Bradford, J. M., Greenberg, D. M., Larose, M. R., & Curry, S. (1998). Homicidal and non-homicidal child molesters: Psychological, phallometric and criminal features. *Sexual Abuse: A Journal of Research and Treatment, 10*, 305-323. doi:10.1023/A:1022146105664

Firestone, P., Bradford, J. M., Greenberg, D. M., McCoy, M., Larose, M. R., & Curry, S. (1999). Prediction of recidivism in incest offenders. *Journal of Interpersonal Violence, 14*, 511-531. doi:10.1177/088626099014005004

Firestone, P., Bradford, J. M., Greenberg, D. M., & Nunes, K. L. (2000). Differentiation of homicidal child molesters, non-homicidal child molesters, and non-offenders by phallometry. *American Journal of Psychiatry, 157*, 1847-1850. Retrieved from http://ajp.psychiatryonline.org/archive/index.dtl

Firestone, P., Kingston, D. A., Wexler, A., & Bradford, J. M. (2006). Long term follow-up of exhibitionists: Psychological, phallometric and offense characteristics. *The Journal of the American Academy of Psychiatry and the Law, 34,* 349-359. Retrieved from http://www.jaapl.org/

Gratzer, T. & Bradford, J. M. (1995). Offender and offense characteristics of sexual sadists: A comparative study. *Journal of Forensic Sciences, 40*, 450-455. Retrieved from http://www3.interscience.wiley.com/journal/118519059/home?CRETRY=1&SRETRY=0

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*, 839-843. Retrieved from http://radiology.rsna.org/content/by/year

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

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

73, 1154-1163. doi:10.1037/0022-006X.73.6.1154

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. doi:10.1037/a0014421

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In. R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually Coercive Behavior: Understanding and Management* (pp. 198-210). New York, NY: The New York Academy of Sciences.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Vermont: Duxbury/Thompson Learning.

Humphreys, L. G., & Swets, J. A. (1991). Comparison of predictive validities measured with biserial correlations and ROCs of signal detection theory. *Journal of Applied Psychology, 76*, 316-321. doi:10.1037/0021-9010.76.2.316

Kansas v. Hendricks, 521 U. S. (1997). Retrieved from http://www.oyez.org/cases/1990-1999/1996/1996_95_1649/

Karpmann, B. (1954). *The sex offender and his offenses*. New York: Julian Press.

Kendell, R. E. (1975). *The role of diagnosis in psychiatry*. Oxford, United Kingdom: Blackwell.

Kendell, R. & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry, 160*, 4-12. Retrieved from http://ajp.psychiatryonline.org/archive/index.dtl

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. doi:10.1007/s10508-006-9091-x

Kingston, D. A., Firestone, P., Wexler, A., & Bradford, J. M. (2008). Factors associated with

recidivism among intrafamilial child molesters. *Journal of Sexual Aggression, 14*, 3-8. doi:10.1080/13552600802074924

Kingston D. A., & Yates P. M. (2008). Sexual sadism: Assessment and treatment. In D. R. Laws & W. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (2nd ed, pp 231-249). New York, NY: The Guilford Press.

Kingston D. A., Yates P. M., Firestone P., Babchishin, K., & Bradford J. M. (2008). Long-term predictive validity of the Risk Matrix 2000: A Comparison with the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment, 20*, 466-484. doi:10.1177/1079063208325206

Knight, R. A. & Prentky, R. A. (1990). Classifying sexual offenders: The development and corroboration of taxonomic models. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 23-52). New York: Plenum Press.

Krafft-Ebing, R. von. (1999). *Psychopathia sexualis*. (12th ed., F. J. Rebman, Trans.). Burbank, CA: Bloat. (Original work published 1886).

Krueger, R. B. (in press). The DSM diagnostic criteria for sexual sadism. *Archives of Sexual Behavior.* doi:10.1007/s10508-009-9586-3

Lalumière, M. L., & Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10,* 227-237. doi:10.1023/A:1021369904984

Lalumière M. L, Harris G. T, Quinsey V. L, & Rice M. E. (2005). *The causes of rape: understanding individual differences in male propensity for sexual aggression*. Washington, DC: American Psychological Association.

Langevin, R., Bain, J., Ben-Aron, M. H., Coulthard, R., Day, D., Roper, V., et al. (1985). Sexual aggression: Constructing a predictive equation: A controlled pilot study. In R. Langevin (Ed.), *Erotic preference, gender identity, and aggression in men: New research studies* (pp. 39-76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levenson, J. S. (2004). Reliability of sexually violent predator civil commitment criteria. *Law and Human Behavior, 28,* 357-368. doi:10.1023/B:LAHU.0000039330.22347.ad

Marshall, W. L. (2006). Diagnostic problems with sexual offenders. In W. L., Marshall, Y. M. Fernandez, L. E., Marshall, & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp. 33-43). Chichester, UK: John Wiley & Sons.

Marshall W. L., & Hucker S. J. (2006). Issues in the diagnosis of sexual sadism. *Sexual Offender Treatment,* Available online at http://sexual-offender-treatment.org. Retrieved July 15, 2008.

Marshall, W. L., & Kennedy, P. (2003). Sexual sadism in sexual offenders: An elusive diagnosis. *Aggression and Violent Behavior: A Review Journal, 8,* 1-22. doi:10.1016/S1359-1789(01)00052-0

Marshall, W. L., Kennedy, P., & Yates, P. M. (2002). Issues concerning the reliability and validity of the diagnosis of sexual sadism applied in prison settings. *Sexual Abuse: A Journal of Research and Treatment, 14,* 310-311. doi:10.1023/A:1019917519457

Marshall, W. L., Kennedy, P., Yates, P. M., & Serran, G. A. (2002). Diagnosing sexual sadism in sexual offenders: Reliability across diagnosticians. *International Journal of Offender Therapy and Comparative Criminology, 46,* 668-676. doi:10.1177/0306624X02238161

McCoy, M. E. (1997). *Predictors of recidivism in a population of Canadian sex offenders: Psychological, physiological, and offense factors.* Unpublished doctoral dissertation.

University of Ottawa, Ottawa, Canada.

McLawsen, J. E., Jackson, R.L., & Vannoy, S.D. (2008). Professional perspectives on sexual sadism. *Sexual Abuse: A Journal of Research and Treatment, 20*, 272-304. doi:10.1177/1079063208320250

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

Meehl, P. E. (1973). *Psychodiagnosis: Selected papers*. Minneapolis: University of Minnesota Press.

Mezzich, J. E. (2002). Comprehensive diagnosis: A conceptual basis for future diagnostic systems. *Psychopathology, 35*, 162-165. doi:10.1159/000065138

Moulden, H. M., Firestone, P., Kingston, D. A., & Bradford, J. M. (2009). Recidivism in pedophiles: An investigation using different methods of defining pedophilia. *Journal of Forensic Psychiatry and Psychology, 20,* 680-701. doi:10.1080/14789940903174055

Nitschke, J., Osterheider, M, & Mokros, A. (2009). A cumulative scale for severe sexual sadism. *Sexual Abuse: A Journal of Research and Treatment, 21*, 262-278. doi:10.1177/1079063209342074

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. doi:10.1177/107906320001200202

Packard, R. L., & Levenson, J. S. (2006). Revisiting the reliability of diagnostic decisions in sex offender civil commitment. *Sexual Offender Treatment 1*. Retrieved from http://www.sexual-offender-treatment.org

Proulx J., Aubut J., McKibben A., & Cote M. (1994). Penile responses of rapists and nonrapists

to rape stimuli involving physical violence or humiliation. *Archives of Sexual Behavior, 23*, 295-310. doi:10.1007/BF01541565

Proulx, J., Blais, E., & Beauregard, E. (2006). Sadistic sexual aggressors. In W. L. Marshall, Y. M. Fernandez, L. E. Marshall & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp.61-77). Chichester, UK: John Wiley & Sons.

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.

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

Rice, M. E., & Harris, G. T. (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., Lang, C., & Cormier, C. A. (2006). Violent sex offenses: How are they best measured from official records? *Law and Human Behavior, 30*, 525-541. doi:10.1007/s10979-006-9022-3

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775-777. doi:10.1037/0003-066X.45.6.775

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

Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic Interests and recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior, 33,* 455-466. doi:10.1023/B:ASEB.0000037426.55935.9c

Seto, M. C., & Kuban, M. (1996). Criterion-related validity of a phallometric test for paraphilic rape and sadism. *Behavior Research and Therapy, 34*, 175-183. doi:10.1016/0005-7967(95)00056-9

Seto, M. C., & Lalumière, 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. doi:10.1177/107906320101300103

Seto, M. C., & Lalumière, M. L., Harris, G. T., & Chivers, M. L. (2009, October). *Distinguishing biastophilia and sadism: The relative importance of violence and non-consent cues*. Paper presented at the annual research and treatment conference for the Association for the Treatment of Sexual Abusers. Dallas, TX.

Seto, M. C., Murphy, W.D., Page, J., & Ennis, L.E. (2003). Detecting anomalous interests in juvenile sex offenders. In R. A. Prentky, E. S. Janus, & M. C. Seto. (Eds.), *Sexually Coercive Behavior: Understanding and Management* (pp. 118-130). New York: New York Academy of Sciences.

Spitzer, R. L. (2001). Values and assumptions in the development of DSM-III and DSM-III-R: An insider's perspective and a belated response to Sadler, Hulgus, and Agich's "On values in recent American psychiatric classification". *The Journal of Nervous and Mental Disease, 189*, 351-359. Retrieved from http://journals.lww.com/jonmd/pages/default.aspx

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99*, 100-117. doi:10.1037/0033-2909.99.1.100

Thornton, D. (in press). Evidence regarding the need for a diagnostic category for a coercive paraphilia. *Archives of Sexual Behavior*. doi:10.1007/s10508-009-9583-6

Ward, T., Polaschek, D., & Beech, A. R. (2006). *Theories of sexual offending*. Chicester, UK: Wiley & Sons.

Wexler, A. (2005). *An examination of recidivism in a population of Canadian sex offenders*: Unpublished doctoral dissertation, University of Ottawa, Canada.

Wilson, R. J., Abracen, J., Picheca, J. E., Malcolm, P. B., & Prinzo, M. (October, 2003). *Pedophilia: An evaluation of diagnostic and risk management methods*. Paper presented at the 23rd Annual Research and Treatment Conference of the Association for the Treatment of Sexual Offenders. St. Louis, MI.

World Health Organization (1992). *International classification of diseases and related health problems* (10th rev.). Geneva: Author.

Table 1

Demographic and Offense-Related Data for Sadistic and Nonsadistic Sex Offenders as a Function of Diagnosis

Variable	**DSM-III or DSM-III-R Criteria**						
	Sexual Sadist		**Non-Sadist**				
	M/% *(SD)*	*N*	*M/%* *(SD)*	*N*	*t* or χ^2	*p*	*d or V*
Age at time of assessment (in years)	28.9 (6.9)	48	38.9 (11.9)	534	8.7	< .001	0.86
Type of Sexual Offender					117.1	< .001	.45
Rapist	66.0%	33	9.9%	53			

Extrafamilial against child	24%	12	36.1%	193			
Intrafamilial against child	10%	5	53.9%	288			
Level of Violence	2.40 (2.11)	40	0.69 (1.40)	376	- 4.9	< .001	1.15
Sexual Intrusiveness	4.07 (1.16)	45	3.73 (.68)	404	- 1.9	.058	.46
Phallometric Index	1.38 (1.05)	50	0.84 (.89)	526	-4.0	< .001	.60

SORAG	6.17	36	3.09	400	-8.6	<.001	1.51
	(1.83)		(2.06)				
Number of Prior Charges or Convictions							
Sexual	.40	50	.69	534	.84	.404	0.12
	(.93)		(2.45)				
Violent	1.68	50	1.13	534	-1.3	.197	0.19
	(2.12)		(2.91)				
Any criminal	6.94	50	3.50	534	-1.9	.067	0.43
	(12.76)		(7.46)				

Note. DSM = Diagnostic and Statistical Manual for Mental Disorders (DSM-III or DSM-III-R) diagnosis; SORAG = Sex Offender Risk Appraisal Guide; *d* = Cohen's effect size; *V* = Cramér's V effect size.

Table 2

Correlation Matrix for Indicators of Sexual Sadism, Actuarial Risk Assessment, and Recidivism

	DSM	Level of Violence	Sexual Intrusiveness	Phallometric Index	SORAG	Sexual Recidivism	Violent Recidivism	Any Recidivism
DSM	-	.32***	.14**	.17***	.38***	.09*	.09*	.09*
		(416)	(449)	(576)	(436)	(584)	(584)	(584)
Level of Violence		-	.51***	.14**	.47***	.18***	.22***	.22***
			(406)	(410)	(383)	(416)	(416)	(416)
Sexual			-	.07	.35***	.08	.10*	.14**
Intrusiveness				(443)	(414)	(449)	(449)	(449)
Phallometric				-	.21***	.13**	.17***	.17***
Index					(430)	(578)	(578)	(578)
SORAG					-	.28***	.38***	.43***
						(436)	(436)	(436)

Note. Sample sizes are in parentheses. DSM = Diagnostic and Statistical Manual for Mental Disorders (DSM-III or DSM-III-R) diagnosis; SORAG = Sex Offender Risk Appraisal Guide.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Predictive Accuracy of the Indicators of Sexual Sadism.

	Type of Recidivism					
	Sexual		Violent (including sexual)		Any Criminal	
Sadism Indicator	AUC	95% CI	AUC	95% CI	AUC	95% CI
DSM	.54	.47-.62	.54	.48-.60	.54	.48-.60
Level of Violence	.61**	.54-.69	.62***	.55-.68	.62***	.56-.67
Sexual Intrusiveness	.57	.49-.65	.58*	.52-.64	.59**	.53-.65
Phallometric Index	.60**	.53-.67	.58*	.52-.64	.58**	.52-.63

Note. DSM = Diagnostic and Statistical Manual for Mental Disorders (DSM-III or DSM-III-R) diagnosis; CI = confidence interval.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Cox regression survival analyses for risk and sexual sadism indicators predicting sexual and violent (including sexual) recidivism

	β	*SE* β	e^b	95% CI for e^b		χ^2 change from previous block
				Lower	Upper	
Sexual Recidivism						
Block 1						19.02***
SORAG	.22	.05	1.24	1.13	1.36	
Block 2						4.46
Violence of Act	-.03	.09	.97	.81	1.17	
Sexual Intrusiveness	-.09	.19	.91	.62	1.34	
Phallometric Index	.21	.11	1.24	1.01	1.52	
DSM	-.30	.44	.74	.31	1.74	
Violent Recidivism						
Block 1						35.30***
SORAG	.23	.04	1.26	1.17	1.35	

Block 2						12.08*
Violence of Act	-.08	.07	.92	.80	1.07	
Sexual Intrusiveness	-.05	.16	.95	.70	1.30	
Phallometric Index	.28	.08	1.32	1.13	1.54	
DSM	-.22	.34	.80	.41	1.55	

Note. SORAG = Sex Offender Risk Appraisal Guide; DSM = Diagnostic and Statistical Manual for Mental Disorders (DSM-III or DSM-III-R) diagnosis; CI = confidence interval.

* $p < .05$. *** $p < .001$.