

Sexual Sadism Among Sex Offenders in Switzerland

Sexual Abuse

2020, Vol. 32(1) 79–100

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1079063218800473

journals.sagepub.com/home/sax



Leonel C. Gonçalves^{1,2}, Astrid Rossegger^{1,2},
Juliane Gerth^{1,2}, Jay P. Singh³,
and Jérôme Endrass^{1,2}

Abstract

The Sexual Sadism Scale (SeSaS) was developed to assist in the diagnosis of sexual sadism, and it revealed adequate psychometric properties in prior research. This study cross validated the SeSaS in Switzerland using a sample of 179 male sex offenders. Specifically, the SeSaS conformed to a Mokken model of double monotonicity (scalability coefficient [H] = .46, coefficient of reproducibility [CR] = .89), indicating that it measures a unidimensional construct of sexual sadism with hierarchically ordered items. The reliability of the scale was acceptable to high (ρ = .80, λ_2 = .75, κ = .88). In addition, the SeSaS was strongly associated with sexual sadism diagnoses based on mental health manuals (r_{pb} = .60, odds ratio [OR] = 13.02, area under the curve [AUC] = 1) but not with recidivism. The results suggest that the use of the SeSaS may improve the validity and reliability of sexual sadism diagnoses, therefore playing a role in the assessment and management of sex offenders.

Keywords

Sexual Sadism Scale, sexual sadism, assessment, diagnosis, recidivism

The clinical diagnosis of sexual sadism has been criticized for lacking validity and reliability (Eher et al., 2016; Kingston, Seto, Firestone, & Bradford, 2010; Marshall & Hucker, 2006; Marshall, Kennedy, & Yates, 2002; Mokros, Schilling, Weiss, Nitschke, & Eher, 2014; Nitschke, Mokros, Osterheider, & Marshall, 2013; Richards & Jackson,

¹Office of Corrections, Canton of Zurich, Switzerland

²University of Konstanz, Germany

³University of Pennsylvania, Great Falls, VA, USA

Corresponding Author:

Jérôme Endrass, Directorate of Justice and the Interior, Office of Corrections, Canton of Zurich, Hohlstrasse 552, 8090 Zurich, Switzerland.

Email: jerome.endrass@ji.zh.ch

2011). For this reason, researchers have begun developing psychometric assessment instruments for sexual sadism composed of behavioral indicators that can be rated based on crime-scene information (e.g., Nitschke, Osterheider, & Mokros, 2009). However, research on the psychometric properties of such tools is still scarce and, therefore, little is known about their utility in the forensic context (Longpré, Proulx, & Brouillette-Alarie, 2018). The present study cross validates the Sexual Sadism Scale (SeSaS; Nitschke et al., 2009) within a sample of male sex offenders in Switzerland. The SeSaS may be useful to assess sexual sadism severity as it is rated based on the information from official records, it does not require the rater to have a particular professional qualification or training, and prior research suggested that it has adequate psychometric properties (Mokros, Schilling, Eher, & Nitschke, 2012; Mokros & Stefanska, 2017; Nitschke et al., 2009). The results are important for practice as they shed light on the usefulness of the SeSaS as a complement or alternative to the clinical diagnosis of sexual sadism and as a predictor of recidivism in crime.

Sexual Sadism Diagnosis

There are several methodological problems associated with the study and diagnosis of sexual sadism based on diagnostic manuals. First, there are no pathognomonic symptoms characteristic of sexual sadism, which complicates the establishment of a definite diagnosis (Longpré, Guay, & Knight, 2017). Second, in addition to the variation in the definitions of sexual sadism (e.g., sexual sadism in the *Diagnostic and Statistical Manual of Mental Disorders* [5th ed.; *DSM-5*; American Psychiatric Association, 2013] and sadomasochism in the International Classification of Diseases [ICD-10; World Health Organization, 1993]), the operationalization of diagnostic criteria and their relevance is lacking, which is the reason why evaluators frequently choose an idiosyncratic list of criteria for the diagnoses (Fedoroff, 2008; Longpré, Proulx, Brouillette, & Alarie, 2018; Marshall & Hucker, 2006; Marshall & Kennedy, 2003; Richards & Jackson, 2011). Third, evaluators frequently assume that all violence and infliction of pain are diagnosable as sexual sadism (Frances & Wollert, 2012). However, with the exception of cruelty and torture, the remaining features are common of all rapes, making it difficult to distinguish sexual sadists from other sex offenders (Frances & Wollert, 2012; Marshall & Yates, 2004). Fourth, the diagnosis requires the offender to be sexually aroused by sadistic fantasies or victims' responses, and because offenders tend to be reluctant in admitting their sadistic fantasies, the diagnostician often has to infer a sexual desire from the individual's behavior (Eher et al., 2016; Marshall & Kennedy, 2003; Marshall & Yates, 2004; Richards & Jackson, 2011). Finally, because the criteria for sexual sadism differ between diagnostic manuals, it is difficult to compare results from various studies (Marshall & Yates, 2004).

These limitations are likely associated with the low level of agreement on sexual sadism diagnostics between evaluators. A systematic review (Nitschke et al., 2013) demonstrated that the estimates of clinician agreement in sexual sadism diagnoses differ widely, ranging from weak to substantial across seven studies, with Cohen's kappa

(κ) values varying between .14 and .93. According to Cicchetti (1994), reliability coefficients (in terms of kappa, weighted kappa, and intraclass correlation [ICC]) below .40, between .40 and .59, between .60 and .74, and above .75 represent poor, fair, good, and excellent levels of interclinician reliability, respectively. Some evidence exists that sexual sadism can be reliably diagnosed using diagnostic manuals when the evaluators are provided enough relevant information, when they anticipate having to defend their diagnoses in court, and when the prevalence of sexual sadism is not too low (Doren & Elwood, 2009). However, due to the methodological problems associated with the sadism criteria and the variability in clinicians' ratings, the validity and reliability of sadism diagnoses have been questioned.

Several authors have suggested that a dimensional approach that indicates the severity of sexual sadism, rather than the presence or absence of the disorder, could be a viable alternative to the current categorical system provided by diagnostic manuals (Berner, Berger, & Hill, 2003; Bradley, Shedler, & Westen, 2006; Longpré et al., 2018; Marshall & Kennedy, 2003; Mokros et al., 2014; Nitschke et al., 2013). Using indices from crime scenes, sexual sadism could be more behaviorally operationalized (Kingston et al., 2010; Marshall & Kennedy, 2003). Thus, clinicians could improve the accuracy of sadism diagnoses by describing sex offenders in terms of sadistic acts including the degree of aggression, control, and humiliation of the victim (Kingston et al., 2010; Marshall et al., 2002; Marshall & Yates, 2004). These behavioral indices could then be combined into sadism assessment tools that produce varying scores regarding offenders' sexual sadism level.

The SeSaS

The SeSaS (Nitschke et al., 2009) is a file-based observer rating of crime-scene characteristics indicative of sexual sadism that was developed based on the Marshall and Hucker's (2006) scale. The Marshall and Hucker's (2006) scale was constructed following a literature review on sexual sadism that revealed 35 characteristics used in previous studies to diagnose sex offenders (Marshall & Kennedy, 2003). Marshall, Kennedy, Yates, and Serran (2002) then conducted a study where they asked diagnosticians to rate the criteria identified in the literature review in terms of their relevance for the diagnosis of sexual sadism. Based on this study, the authors selected 17 criteria that professionals judged the most relevant to the evaluation of sexual sadism to form the Marshall and Hucker's (2006) scale.

Later, using Mokken scaling with an iterative scaling algorithm on a sample of 100 male sex offenders in a German forensic psychiatric hospital, Nitschke et al. (2009) found that, out of the 17 criteria suggested by Marshall and Hucker (2006) plus one additional item (insertion of objects into victim's body orifices), 11 items conformed to a hierarchically ordered unidimensional scale. Overall, the 11 items had strong discrimination ability ($H = .83$) and conformed to a scale of the Guttman type (coefficient of reproducibility [CR] = .97). The total score of the SeSaS was highly reliable ($\rho = .93$) and the interrater reliability was excellent ($\kappa = .86$). However, in this study, the base rate of sexual sadism was inflated (50%) because the authors used an

oversampling of individuals diagnosed as sexual sadists for the derivation of the scale. The tool was named the SeSaS.

Next, the SeSaS was tested by Mokros et al. (2012) on a sample of 105 male sex offenders in Austria. Using a confirmatory factor analysis (CFA), the study replicated the one-dimensional structure of the SeSaS but failed to confirm its deterministic nature, as indicated by the level of Guttman errors and resulting CR of .79. However, the SeSaS was commensurate with a one-parameter logistic model (Rasch model), thus conforming to an ordinal scale. The reliability of the SeSaS total score was good ($\rho = .86$), the internal consistency of the scale was acceptable (Cronbach's $\alpha = .75$ and Guttman's $\lambda_2 [\lambda_2] = .78$), and the interrater reliability was moderate ($\kappa = .58$). Subsequently, based on a sample of 1,020 adult male sex offenders in Austria (which included the 105 offenders from Mokros et al., 2012), Mokros et al. (2014) observed that three of the 11 items (i.e., sexual arousal, inflicting humiliation, and keeping trophies) correlated negatively with some of the other items, therefore not representing a unidimensional latent construct. In this study, the interrater reliability of the SeSaS total was excellent, with an ICC of .91.

More recently, Mokros and Stefanska (2017) explored the psychometric properties of the SeSaS in the United Kingdom with a sample of 350 male sexual murderers. A CFA evidenced that a two-factorial structure corresponding to parts one (i.e., the 11 crime-scene items) and two (i.e., three biographical items: planful conduct, sadistic acts beyond listed offenses, and arousal through sadistic fantasies) of the SeSaS fitted the data. In this study, the data were in accordance with a two-parameter logistic model (Birnbaum model), indicating that the items vary not only in their difficulty but also in their discrimination, therefore not conforming to a cumulative scale. The reliability of the total score was adequate ($\rho = .76$) but the internal consistency of the items was questionable ($\alpha = .61$, $\lambda_2 = .63$). The interrater reliability of the SeSaS total was good (ICC = .80). In addition, in a study with 90 female sex offenders in the United States, Pflugradt and Allen (2013) reported a joint probability of agreement in the SeSaS total of 94%.

Criterion Validity of the SeSaS

A fundamental attribute of a sexual sadism assessment tool for mental health service practitioners is its ability to identify individuals diagnosed as sadists based on diagnostic manuals. In the SeSaS development study, Nitschke et al. (2009) found that the scale discriminated perfectly between diagnosed sadists and nonsadists when a cut score ≥ 4 was used (sensitivity and specificity = 100%). Mokros et al. (2012), adopting a cut score ≥ 7 , found that the ability of the tool to correctly classify individuals as sadists or nonsadists was 56% (sensitivity) and 90% (specificity), respectively. The cutoff of 4 points originally recommended by Nitschke et al. (2009) yielded a sensitivity of 83% and a specificity of 58%. The probability that a randomly selected sexual sadist had a higher score on the SeSaS than a randomly chosen nonsadist was 81% (area under the curve [AUC]). In a meta-analysis combining these and two North American studies, Nitschke et al. (2013) found that the SeSaS had an overall sensitivity estimate of 95% (95% confidence interval [CI] = [66, 99]) and an overall specificity estimate of 99% (95% CI = [64, 100]) using a cut score ≥ 4 .

Other studies lent support to the association between the SeSaS and the diagnosis of sexual sadism. Mokros and Stefanska (2017) reported a strong point-biserial correlation between the SeSaS total and clinicians' assessment of sexual sadism of ($r_{pb} = .57, p < .001$). Similarly, with a sample of 72 rapists incarcerated in Canada, Longpré et al. (2018) reported a moderate correlation between the SeSaS and the diagnosis of sexual sadism ($r_{pb} = .46, p < .001$). It must be noted that point-biserial correlation coefficients tend to be smaller than Pearson correlation coefficients and, therefore, the magnitude of the associations may be biased downward (Rice & Harris, 2005).

In addition, the ability of an assessment tool to estimate the probability of recidivism in crime is important for the criminal justice system. It is commonly assumed that sexual sadists are more dangerous and, thus, have a higher risk to reoffend than nonsadistic sex offenders (Eher et al., 2016; Fedora et al., 1992; Fedoroff, 2008; Marshall & Yates, 2004; Richards & Jackson, 2011). However, at present, results concerning the association between sexual sadism and recidivism are inconclusive. In a meta-analysis that combined seven studies from four countries, including Anglo Saxon and German-speaking ones, Eher et al. (2016; Study 1) found that the risk of violent and sexual reoffending among sexual sadists (classified either by diagnostic manuals or behavioral indicators) was not significantly higher compared with nonsadistic offenders. However, when combining only the five studies that relied on *DSM* criteria, sexual sadism was associated with both violent and sexual recidivism.

In a subsequent empirical study, Eher et al. (2016; Study 2) investigated whether different measures of sexual sadism were associated with reoffending in sex offenders released from prison, and whether these measures added incremental validity after controlling for violence risk. Using a sample of 768 adult male sex offenders in Austria, the authors observed that the SeSaS had a small effect in predicting violent but not sexual recidivism, whereas the *DSM* diagnosis of sexual sadism was not associated with either outcome. The strongest associations with reoffending were observed for instruments designed to assess (sexual) violence risk (Sex Offender Risk Appraisal Guide [SORAG; Quinsey, Harris, Rice, & Cormier, 2006] and Static-99; Hanson & Thornton, 1999) and psychopathy (Psychopathy Checklist-Revised [PCL-R; Hare, 2003]). When these instruments and sadism measures were combined in a Cox regression model, neither a *DSM* diagnosis of sexual sadism nor the SeSaS scores were related to violent recidivism. Similar results were reported by Kingston et al. (2010). Based on a sample of 586 adult male sex offenders in Canada, the authors observed that behavioral (level of violence) and physiological (phallometric testing) indices of sexual sadism were related to violent and sexual recidivism, but not to the *DSM* diagnosis. When combined in a Cox regression, only phallometric indices of sexual sadism added incremental validity over the SORAG.

In addition, Brouillette-Alarie, Proulx, and Hanson, (2018) observed that the SeSaS was not significantly associated with sexual, nonsexual violent, and nonsexual nonviolent recidivism, with AUC of .48, .55, and .54 for the different outcomes, respectively. However, using Harrell's *C* statistic (Harrell's *C* is similar to the AUC but takes time into account; Harrell, Califf, Pryor, Lee, & Rosati, 1982), the SeSaS had a small but significant effect on nonsexual violent (.57) and nonsexual nonviolent recidivism (.56).

The Present Study

To summarize, although the SeSaS measures a one-dimensional construct and has adequate psychometric properties apparently, the results have been somewhat mixed (Longpré et al., 2018). Particularly, findings regarding the invariant ordering of the items was only observed in the development study of the scale (Nitschke et al., 2009). Similarly, more research is necessary to draw conclusions regarding the criterion validity of the SeSaS, especially its ability in predicting recidivism. Therefore, the objectives of this study were twofold: (a) to examine the psychometric properties of the SeSaS in Switzerland and (b) to examine the validity of the SeSaS in identifying individuals diagnosed as sexual sadists and in assessing recidivism in crime. The following were hypothesized:

Hypothesis 1: The 11 crime-scene items of the SeSaS represent a one-dimensional and cumulative scale.

Hypothesis 2: The SeSaS is associated with the clinical diagnosis of sadism and reoffending.

Method

Sampling and Procedure

The sample consisted of violent and sex offenders from three cohorts. The first cohort was comprised of individuals enrolled in the Zurich Forensic Study (Urbaniook et al., 2007). This study followed all offenders with either a minimum sentence of 10 months or court-mandated therapy who were supervised by the canton of Zurich criminal justice system as of August 2000 ($N = 465$). The second cohort included all patients who began treatment with the canton's Office of Corrections Department of Mental Health Services (DMH) between January 1, 1997, and December 31, 2009 ($N = 296$). The third cohort comprised of all offenders who were admitted to a correctional facility for young adults in the canton of Zurich and who were released between January 1, 2000, and December 31, 2009 ($N = 212$). From these cohorts, only adult male offenders who were convicted of a sexual offense and had already been released into the community, and for whom there was enough information available to score the SeSaS were included, resulting in a final study sample of 179 sex offenders.

Information used to code the SeSaS was obtained exclusively from official records, which includes judicial verdicts, criminal records, correctional files, clinical files, and forensic expert opinions (when available). Recidivism data were last reviewed in September 2013, which corresponds to the end of the follow-up period. All data were coded by master-level psychologists working at the DMH. The psychologists completed departmental training at the DMH for the coding of the assessment tools utilized in the current study. To avoid bias in the ratings, the coders were blinded to the individual sex offenders' outcomes.

This study exclusively uses data from criminal justice files that belong to the Office of Corrections of the Canton of Zurich and are not protected under medical

privacy laws. All analyses were performed on anonymized data. Thus, there was no need to submit the study to an external ethics committee. We report how we determined our sample size, all data exclusions, manipulations, and measures in the study.

Measures

Sexual sadism. Sexual sadism was measured in two different ways: (a) with the SeSaS and (b) using the clinical diagnosis made by the criminal justice practitioners. The SeSaS is composed of 11 items, including (1) offender is sexually aroused by the act, (2) offender exercises power/control/domination over the victim, (3) offender humiliates or degrades the victim, (4) offender tortures the victim or engages in acts of cruelty, (5) offender mutilates sexual parts of the victim's body, (6) offender engages in gratuitous violence toward the victim, (7) offender keeps trophies of the victim, (8) offender mutilates nonsexual parts of the victim's body, (9) victim is abducted or confined, (10) evidence of ritualism in the offense, and (11) insertion of objects into victim's body orifices. The items were coded dichotomously (0 = absent, 1 = present). Therefore, the total scores can range from 0 to a maximum of 11, with higher scores indicating a higher sadistic propensity. The clinical diagnoses of sadism were based on expert psychiatric opinions ordered by prosecutors prior to trial or on a formal psychological assessment by the DMH posttrial. The diagnoses followed diagnostic manuals, either the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychiatric Association, 2000) or the ICD-10. This variable was coded dichotomously (0 = nonsadist, 1 = sadist).

Assessment tools. In addition to the SeSaS, a measure of psychopathy, a measure of violence risk, and a measure of risk of sexual offenses were included. These constructs were assessed using the PCL-R (Hare, 2003), including the total score plus Factor 1 (affective deficits) and Factor 2 (antisocial deviance), the SORAG (Quinsey et al., 2006), and the Static-99 (Hanson & Thornton, 1999), respectively. All total scores were (nonnormal) continuous variables, in which higher values indicated an increased level of risk. These tools were utilized in prior studies on sexual sadism (see Eher et al., 2016; Kingston et al., 2010; Richards & Jackson, 2011).

Recidivism. Recidivism in crime was defined as new charges or convictions after the offenders were released into the community. Recidivism outcomes were categorized as (a) violent (severe offenses, nonsexual), (b) sexual (contact offenses), and (c) general recidivism (any offense). The three recidivism outcomes were coded dichotomously (0 = nonrecidivist, 1 = recidivist). The average time at risk of the sample in the community was 10.73 years (median = 10.36 years, $SD = 5.47$ years), ranging from 1 day to almost 37 years. In this study, time at risk did not refer to time of opportunity ("street time") but to time from release until the first new offense or until the date of the last follow-up information.

Analyses

Replicating the procedure used to develop the scale, the psychometric properties of the 11 SeSaS items were assessed using Mokken scaling (Mokken, 1971), which is a data reduction method for dichotomous items among nonparametric item response theory (IRT). Specifically, Mokken models serve to validate an ordinal measure of a latent variable (Hardouin, Bonnaud-Antignac, & Sébille, 2011), which was sadism in this case. There are two types of Mokken models: monotone homogeneity (MHM) and double monotonicity (DMM). MHMs are based on the assumptions of unidimensionality (the items are explained by a common latent trait), local independence (conditionally to the latent trait, the responses to the items are independent), and monotonicity (the probability of scoring positive in an item increases monotonically with increasing values of the latent trait). DMM are built onto MHM and, in addition to these three assumptions, assume nonintersection (the ordering of the items is the same across different values of the latent trait). An MHM allows ranking a set of persons in the same order on the latent continuum. However, it is not sufficient to establish a uniform rank ordering of items, which is the additional property of the DMMs (Van Schuur, 2003). Evidence that the model meets the MHM assumptions justifies the use of a scale in clinical practice to rank individuals based on their symptom severity. When DMM assumptions are met, items form a consistent hierarchy, which means that behaviors located at higher levels of the latent trait tend to only be observed if behaviors located at lower levels of the latent trait have also been observed (Murray, McKenzie, Booth, & Murray, 2013).

Detailed information on the methods used to analyze the assumptions and results of the Mokken models, as well as the other statistics utilized in the study, are presented in the online supplemental material. In summary, the unidimensionality and local independence assumptions of the Mokken model were assessed through indices from exploratory factor analysis, the dimensionality evaluation to enumerate contributing traits (DETECT; Zhang & Stout, 1999), and Loevinger's (1948) scalability coefficients. Monotonicity was assessed using indices based on the Guttman errors (Guttman, 1944) and nonintersection with indices based on the P matrix, as well as the CR proposed by Guttman (1944).

Regarding reliability, the reliability of the SeSaS total was assessed with the Sijtsma & Molenaar (1987) statistic (ρ [p]). The internal consistency of the scale was calculated using Cronbach's (1951) alpha and Guttman's (1945) lambda-2 coefficient, which are lower bound estimates for the true reliability values. The interrater reliability of the SeSaS total was calculated with Fleiss' (1971) kappa. The interrater reliability calculations were based on a subsample of 12 cases coded by six psychologists.

Regarding criterion validity, the relationship between the SeSaS and the clinical diagnosis of sexual sadism (concurrent validity) was assessed with a point-biserial correlation (r_{pb}) and the odds ratio (OR) from a logistic regression. In addition, an AUC analysis was performed, and the sensitivity and specificity, the proportion of participants correctly classified (ACC), the positive likelihood ratio (LR+) and

negative likelihood ratios (LR⁻), and the Youden's index (Youden, 1950) across all possible SeSaS cut scores were calculated.

The ability of the SeSaS to assess recidivism (predictive validity) was tested with Cox regressions to take into account the time-to-event component of the data and the different time at risk of the offenders. The analyses were restricted to a 20-year time-at-risk period ($n = 170$) because the cumulative hazard function for the different recidivism outcomes stagnated after this time point, and it is known that offenders become similar to nonoffenders in their risk of offending around this period (Hargreaves & Francis, 2014).

There was no variable with more than 16% of missing information. In addition, the SeSaS score was linearly related to the log odds of the sexual sadism diagnosis and all Cox regressions met the proportional hazards assumption. Robust standard errors were included in all regressions to deal with observations that exhibit large residuals, leverage, or influence. R 3.1.2 (R Core Team, 2013) was used to calculate the Rho statistic. G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) was used to perform post hoc power analyses. All other statistical analyses were conducted using the software Stata 13.1 (StataCorp, 2013).

Results

Sample Characteristics

The mean age of the 179 participants at time of the index offense was 39 years ($SD = 12.46$ years), ranging from 18 to 77 years. Most were Swiss citizens (84%, $n = 151$) and had a diagnosed mental health disorder (93%, $n = 161$), including mostly drug and alcohol addiction (28%, $n = 48$), personality disorder (51%, $n = 89$), and other mental disorder due to a known physiological condition (60%, $n = 104$). Half of the sample had prior convictions for violent and/or sexual offenses (50%, $n = 88$). During the follow-up period, 34% ($n = 61$) of the offenders reoffended (general recidivism), 12% ($n = 21$) reoffended with a sexual offense, and 3% ($n = 6$) reoffended with a violent offense. The average SeSaS total score in the current sample was 2.09 (median = 2, $SD = 1.80$), ranging from 0 to 10. Additional descriptive characteristics of the sample are presented in Table 1.

Psychometric Properties of the SeSaS

The principal axis factoring (PAF) revealed two initial eigenvalues higher than 1.0, suggesting two factors in the data. However, the initial eigenvalue of the first factor was 3.7 times higher than the value of the second factor (6.63 vs. 1.79) and, after rotation, accounted for 23% of the variance in the data. Furthermore, the DETECT value was equal to 0.04 (see the online supplemental material for details on the interpretation of the results).

Table 2 presents the difficulty (M) and scalability (H_i) of the 11 SeSaS items. Examining the difficulty (relative frequency) of the items revealed that most of the

Table 1. Descriptive Characteristics of the Sample.

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Sociodemographic					
Age at conviction	179	39.45	12.46	18.06	76.51
Age at release	179	41.50	12.45	19.01	76.51
Swiss nationality	179	0.84	0.36	0	1
Married	179	0.22	0.41	0	1
Unemployed	179	0.21	0.41	0	1
Secondary school or more	172	0.58	0.40	0	1
Clinical					
Mental health disorder (general)	173	0.93	0.25	0	1
Personality disorder	173	0.51	0.50	0	1
Substance abuse disorder	173	0.28	0.45	0	1
Mental health treatment history	173	0.39	0.49	0	1
Prior hospitalization	173	0.24	0.43	0	1
Criminological					
Prior convictions: general	177	0.64	0.48	0	1
Prior convictions: violent/sexual	176	0.50	0.50	0	1
Prior conviction: other	179	0.15	0.35	0	1
Time incarcerated (year) ^a	179	2.05	3.33	0	18.13
Time at risk (years)	179	10.73	5.47	0.003	36.98
Offense related					
Alcohol/drug at index offense	175	0.35	0.48	0	1
Female victim	167	0.72	0.45	0	1
Male victim	167	0.38	0.49	0	1
Weapon use	176	0.23	0.42	0	1
Risk assessment tools					
SeSaS	179	2.09	1.80	0.00	10.00
PCL-R total score	165	14.53	7.63	0.00	33.30
PCL-R F1	162	6.25	3.58	0.00	14.90
PCL-R F2	160	6.01	4.29	0.00	17.00
Static-99	158	4.03	2.11	0.00	10.00
SORAG	150	7.37	12.44	-16.00	37.00
Recidivism					
Violent	179	0.03	0.18	0	1
Sexual	179	0.12	0.32	0	1
General	179	0.34	0.48	0	1

Note. SeSaS = Sexual Sadism Scale; PCL-R = Psychopathy Checklist-Revised; F1 = PCL-R Factor 1; F2 = PCL-R Factor 2; SORAG = Sex Offender Risk Appraisal Guide.

^aTime incarcerated after index conviction (does not include remand imprisonment).

elements were observed in less than 15% of the sample, except Item 1 (offender is sexually aroused by the act) and Item 2 (offender exercises power/control/domination over the victim), which were much more frequent (75% and 59%, respectively). All

Table 2. Difficulty and Scalability of the 11 SeSaS Items.

Item	Current study N = 179		Nitschke, Osterheider, and Mokros (2009) N = 100	Mokros, Schilling, Eher, and Nitschke (2012) N = 105
	M	H_i	M	M
1 Offender is sexually aroused by the act	.75	.75	.50	.61
2 Offender exercises power/control/domination over the victim	.59	.66	.57	.59
3 Offender humiliates or degrades the victim	.15	.47	.53	.30
4 Offender tortures the victim or engages in acts of cruelty	.06	.55	.44	.25
5 Offender mutilates sexual parts of the victim's body	.01	.81	.08	.05
6 Offender engages in gratuitous violence toward the victim	.13	.42	.76	.79
7 Offender keeps trophies of the victim	.07	.15	.03	.10
8 Offender mutilates nonsexual parts of the victim's body	.02	.65	.08	.04
9 Victim is abducted or confined	.14	.36	.37	.60
10 Evidence of ritualism in the offense	.11	.45	.38	.35
11 Insertion of objects into victim's body orifices	.07	.31	.12	.10

Note. The sexual sadism base rate was equal to 3% ($n = 6$), 50% ($n = 50$), and 17% ($n = 18$) in the current study, Nitschke et al. (2009), and Mokros et al. (2012), respectively. SeSaS = Sexual Sadism Scale; M = difficulty (*relative frequency*, higher values indicate sadistic behaviors observed more often); H_i = scalability (Loevinger items scalability coefficient, higher values indicate more homogeneous items).

items had acceptable to good scalability with H_i values between .31 (for Item 11: insertion of objects into victim's body orifices) and .81 (for Item 5: offender mutilates sexual parts of the victim's body) except Item 7 ($H_i = .15$; offender keeps trophies of the victim). Item 7 also had an item-rest correlation (IRC) considerably lower than the other items ($r_{pb} = .15$). The other items had IRCs ranging from .31 (for Item 11) to .52 (for Item 4: offender tortures the victim or engages in acts of cruelty). The SeSaS as a whole had medium scalability with an H value of .46.

Table 3 presents the results on the tests of the monotonicity and nonintersection assumptions. There was one violation of the monotonicity assumption for Item 7

Table 3. Monotonicity and Nonintersection of the 11 SeSaS Items.

Item		Monotonicity				Nonintersection			
		#ac	#vi	#zsig	crit	#ac	#vi	#zsig	crit
1	Offender is sexually aroused by the act	3	0	0	-23	90	0	0	-23
2	Offender exercises power/control/domination over the victim	6	0	0	-18	90	0	0	-18
3	Offender humiliates or degrades the victim	3	0	0	-9	90	0	0	-9
4	Offender tortures the victim or engages in acts of cruelty	1	0	0	-13	90	1	1	17
5	Offender mutilates sexual parts of the victim's body	0	0	0	—	90	0	0	-26
6	Offender engages in gratuitous violence toward the victim	6	0	0	-6	90	0	0	-6
7	Offender keeps trophies of the victim	3	1	0	66	90	1	1	37
8	Offender mutilates nonsexual parts of the victim's body	0	0	0	—	90	0	0	-17
9	Victim is abducted or confined	3	0	0	-3	90	0	0	-3
10	Evidence of ritualism in the offense	3	1	0	64	90	0	0	-8
11	Insertion of objects into victim's body orifices	3	0	0	-1	90	0	0	-1

Note. $N = 179$. SeSaS = Sexual Sadism Scale; #ac = number of active comparisons; #vi = number of assumption violations; #zsig = number of statistically significant assumption violations; crit = Molenaar and Sijtsma (2000) fit statistic; — = not computable.

and Item 10 (evidence of ritualism in the offense) but none reached statistical significance. However, this assumption could not be tested for Item 5 and Item 8 (offender mutilates nonsexual parts of the victim's body) because there were not enough positive responses at each possible rest-score group to draw comparisons (the remaining items had between one and six active comparisons). The trace lines of these two items did not decrease but remained flat, increasing only in the higher rest-score group.

One violation of the nonintersection assumption (90 comparisons for all items) was statistically significant for Item 4 and Item 7 but their *crit* values (17 and 37, respectively) suggested that this violation could be due to sampling variation. Furthermore, with 225 Guttman errors for 1,969 total responses, the CR was equal to .89.

The reliability of the SeSaS total score was high ($\rho = .80$) and the internal consistency of the scale was acceptable ($\alpha = .72$, $\lambda_2 = .75$). In addition, the SeSaS revealed good to excellent level of interrater reliability ($\kappa = .88$, 95% CI = [.73, 1]).

Table 4. Association Between the SeSaS and the Clinical Diagnosis of Sexual Sadism.

SeSaS cut score	Sensitivity	Specificity	ACC	LR+	LR-	<i>J</i>
≥0	1	.00	.03	1.00	—	.00
≥1	1	.20	.22	1.24	0.00	.20
≥2	1	.38	.40	1.62	0.00	.38
≥3	1	.73	.74	3.76	0.00	.73
≥4	1	.89	.89	9.11	0.00	.89
≥5	1	.94	.94	15.73	0.00	.94
≥6	1	.98	.98	43.25	0.00	.98
≥7	.83	.99	.99	144.17	0.17	.83
≥8	.50	1	.98	—	0.50	.50
≥9	.33	1	.98	—	0.67	.33
≥10	.17	1	.97	—	0.83	.17
>10	.00	1	.97	—	1.00	—

Note. *N* = 179. SeSaS = Sexual Sadism Scale; ACC = accuracy (proportion of correctly classified offenders); LR+ = likelihood ratio for a positive test result; LR- = likelihood ratio for a negative test result; *J* = Youden index; — = not computable.

Criterion Validity of the SeSaS

The mean SeSaS score was 1.89 (median = 2, *SD* = 1.45, range = 0-7) among nonsadists and 7.83 (median = 7.5, *SD* = 1.47, range = 6-10) among diagnosed sexual sadists. The point-biserial correlation between the SeSaS and the clinical diagnosis of sadism was strong ($r_{pb} = .60, p < .001$) and the logistic regression revealed that the odds of being a diagnosed sexual sadist was approximately 13 times higher for each one-point increase in the SeSaS total (OR = 13.02, 95% CI = [2.32, 73.06], *SE* = 11.46, $p = .004$).

Furthermore, the SeSaS discriminated diagnosed sadists from nonsadists perfectly (AUC = 1, 95% CI = [0.99, 1], *SE* = 0.003). An SeSaS score ≥6 was the point in which sensitivity (100%, 95% CI = [54, 100]) and specificity (98%, 95% CI = [.94, .99]) were maximized (*J* = .98). With this cut score, 98% of the offenders were correctly classified (ACC) and the likelihood ratio for a positive test result was 43.25 (LR+). It would correspond to a sadism base rate of 6% ($n = 10$). Although less efficient (*J* = .89), a cut score ≥4 also had high sensitivity (100%, 95% CI = [.54, .100]) and specificity (89%, 95% CI = [.83, .93]). It correctly classified 89% of the offenders (ACC), produced a likelihood ratio for a positive test result of 9.11 (LR+), and implied a base rate of 14% ($n = 25$). The results are presented in Table 4.

On the contrary, the SeSaS was not related to subsequent violent, sexual, or general recidivism. Among the other risk assessment tools, the PCL-R total score (hazard ratio [HR] = 1.10, $p = .022$) and Factor 1 (HR = 1.41, $p = .034$) were related to the hazard of violent recidivism, and the Static-99 was related to the hazard of sexual recidivism (HR = 1.23, $p = .044$). None of the tools was associated to general recidivism at the 5% level, although the SeSaS almost reached this threshold of statistical significance

Table 5. Association Between the SeSaS and Other Assessment Tools With Recidivism Outcomes.

Tool	N	Violent	Sexual	General
		HR (SE)	HR (SE)	HR (SE)
SeSaS	170	1.07 (0.14)	1.13 (0.10)	1.14 (0.08)
PCL-R total	157	1.10 (0.05)*	1.01 (0.03)	1.02 (0.02)
PCL-R F1	154	1.41 (0.23)*	1.05 (0.06)	1.04 (0.04)
PCL-R F2	152	1.06 (0.09)	1.00 (0.04)	1.01 (0.03)
Static-99	150	1.05 (0.24)	1.23 (0.13)*	0.99 (0.07)
SORAG	141	1.05 (0.03)	1.02 (0.02)	1.00 (0.01)

Note. Differences in the number of participants for the different analyses is due to missing data in the assessment tools. The analyses are limited to offenders with a maximum of 20-year time-at-risk period. HR = hazard ratio from Cox regression; SE = robust standard error (Huber–White sandwich estimator); SeSaS = Sexual Sadism Scale; PCL-R = Psychopathy Checklist–Revised; F1 = PCL-R Factor 1; F2 = PCL-R Factor 2; SORAG = Sex Offender Risk Appraisal Guide.

* $p < .05$, two-tailed.

(HR = 1.14, $p = .067$). Besides not having an independent effect on recidivism outcomes, the SeSaS added no incremental validity to the prediction of any of the outcomes when entered together with the other tools in Cox regressions. Spearman correlations evidenced that the SeSaS had a small but significant association with the PCL-R Factor 1 ($r_s = .19$, $p = .013$), but it was not significantly correlated with any of the other assessment tools (r_s ranging between .03 [with the Static-99] and .13 [with the PCL-R total]). The analyses are presented in Table 5.

Table 6 presents the frequency of offenders, including diagnosed sadists and recidivists, across SeSaS scores. It can be seen that none of the diagnosed sexual sadists reoffended with a violent or sexual offense, and only one reoffended with a general offense. Among individuals with an SeSaS score ≥ 4 , none reoffended with a violent offense, two reoffended with a sexual offense, and eight reoffended with a general offense. Among individuals with an SeSaS score ≥ 6 , only one reoffended (general recidivism). No individual with an SeSaS score ≥ 8 ($n = 3$) reoffended in any of the three crime categories.

Discussion

Utilizing a sample of 179 male sex offenders, the aims of the present study were to examine the psychometric properties of the SeSaS in Switzerland and its validity in identifying individuals diagnosed as sexual sadists and in assessing reoffending. Overall, the results suggest that the SeSaS has adequate to good psychometric properties, representing a unidimensional scale with hierarchically ordered items that measure the construct of sexual sadism reliably. In addition, the scale was strongly associated with the clinical diagnosis of sexual sadism, but not with recidivism in crime.

Table 6. Number of Diagnosed Sadists and Recidivists Across SeSaS Scores.

SeSaS score	n	%	Sadists	Recidivists (sadists)		
				Violent	Sexual	General
0	34	19.0	0	2 (0)	5 (0)	13 (0)
1	32	17.9	0	0 (0)	5 (0)	9 (0)
2	61	34.1	0	2 (0)	4 (0)	19 (0)
3	27	15.1	0	2 (0)	5 (0)	12 (0)
4	8	4.5	0	0 (0)	0 (0)	2 (0)
5	7	3.9	0	0 (0)	2 (0)	3 (0)
6	4	2.2	1	0 (0)	0 (0)	2 (1)
7	3	1.7	2	0 (0)	0 (0)	1 (0)
8	1	0.6	1	0 (0)	0 (0)	0 (0)
9	1	0.6	1	0 (0)	0 (0)	0 (0)
10	1	0.6	1	0 (0)	0 (0)	0 (0)
Total	179	100	6	6 (0)	21 (0)	61 (1)

Note. SeSaS = Sexual Sadism Scale. Sadists refer to the diagnosed sexual sadists.

Psychometric Properties of the SeSaS

In accordance with prior research from the United Kingdom (Mokros & Stefanska, 2017), Austria (Mokros et al., 2012), and Germany (Nitschke et al., 2009), our results suggest that the SeSaS measures a one-dimensional construct of sexual sadism. This is indicated by the dominant factor in the PAF (more than 3 times higher than the second eigenvalue and accounting for more than 20% of the variance), the scalability coefficient of most items and the scale as a whole (H_i and $H > .30$, except Item 7), and the DETECT value (< 0.2). However, Item 7 (offender keeps trophies of the victim) appeared to be too indiscriminate in the present sample ($H_i = .15$), and had a low correlation with the rest score ($r_{pb} = .15$). This may be due to the fact that keeping trophies from the victim is generally more frequent among sexual murderers than sex offenders (see Mokros & Stefanska, 2017) and, therefore, it may be more difficult to find in a lesser form of sexual sadism, as was the case in the current study, where there were only five (attempted) sexual murders. Because the low scalability of Item 7 is probably sample related, and because in this study the Mokken scaling of the SeSaS was made in a confirmatory way, this item was kept in the scale for further analyses. However, removing it would cause the SeSaS H value to increase to .53 and the CR to increase to .92, which would indicate a strong scale of the Guttman type, as reported by Nitschke et al. (2009).

Based on the trace lines and tests of monotonicity (no significant violations, although not calculated for Items 5 and 8), the SeSaS appears to conform to an ordinal person scale. Specifically, higher scores on the SeSaS are suggestive of a higher level of sexual sadism, as previously evidenced by Nitschke et al. (2009), Mokros et al. (2012), and Mokros and Stefanska (2017). However, we cannot draw definitive

conclusions regarding monotonicity because the relative frequencies of Item 5 and Item 8 were too low to test this assumption reliably. These two items were also less observed in the prior studies (see Table 2) of Mokros et al. (2012) and Nitschke et al. (2009; after Item 7). In fact, Items 5 and 8 measure the same construct (mutilation), and only differ regarding the body parts that are mutilated (sexual vs. nonsexual). Thus, combining them should be considered to reduce the difficulty of these items and improve the overall construct validity of the scale.

Examining the tests of nonintersection (no $crit > 80$) and the CR (.89; slightly below the .90 value), the 11 SeSaS items appear to form a consistent and reproducible hierarchy (i.e., a Guttman scale), which is in accord with the findings of Nitschke et al. (2009). However, although the items appeared to be hierarchically ordered, nine of them had a narrow range from 1% to 15%, falling in the higher level of the latent trait. This was not the case in previous studies where the difficulty of the items had a greater range (see Table 2). The lower prevalence of sexual sadists in our sample may explain why the majority of the offenders exhibited only a few sadistic behaviors. In addition, although the item “offender engages in gratuitous violence toward the victim” was the most frequent in the studies of Nitschke et al. (2009) and Mokros et al. (2012), this behavior was observed for only 13% of the current sample. Again, this might be related to the low prevalence of sexual sadists, who could arguably manifest behaviors that are more violent. Nevertheless, despite the lower sadism-based rate in the current sample, the ordering of the items has been different across studies and additional research is necessary to conclude about their hierarchy.

Finally, the reliability of the SeSaS total ($\rho = .80$), the internal consistency of the scale ($\alpha = .72$, $\lambda_2 = .75$), and the interrater agreement ($\kappa = .88$) were all acceptable to excellent, which is in line with prior research (Mokros et al., 2012; Mokros et al., 2014; Mokros & Stefanska, 2017; Nitschke et al., 2009; Pflugradt & Allen, 2013). Together, these studies cover research on different types of samples (forensic vs. correctional, sex offenders vs. sexual murderers, and female vs. male) and from different countries (United States, United Kingdom, Germany, Austria, and Switzerland), which suggests that the SeSaS is reliable.

Criterion Validity of the SeSaS

Like in prior studies (Longpré et al., 2018; Mokros et al., 2012; Mokros & Stefanska, 2017; Nitschke et al., 2009), the SeSaS was strongly related to the clinical diagnosis of sexual sadism ($r_{pb} = .60$, $OR = 13.02$) and discriminated sadists from nonsadists perfectly ($AUC = 1$). A score equal to or higher than 4 (Nitschke et al., 2009) and 7 (Mokros et al., 2012) was previously identified as the best cut scores. In this study, the cut score that maximized sensitivity (100%) and specificity (98%) was a score equal to or higher than 6, making diagnosed sexual sadists 43 times more likely to have a positive SeSaS test when compared with nonsadists, which represents a large probability ($LR+ > 10$). However, although this cut score fits the current data, it is sample related and may be biased. In fact, our sample is quite small, and the prevalence of sadists, as well as core components of sexual sadism, was low. Therefore, the cut score of 4 is more reasonable until better evidence is provided.

Nevertheless, future research should address the question of whether the cut score of 4 points is too low (Mokros et al., 2012). In the current sample, an SeSaS score equal to or higher than 6 would result in a sadism base rate of 6% ($n = 10$), whereas a cut score equal to or higher than 4 would result in a 14% ($n = 10$) base rate. In the meta-analysis of Nitschke et al. (2013), which included 491 offenders from three samples in the United States and Austria, the prevalence of diagnosed sadists was about 6%. In the Eher et al. (2016; Study 2) study, which included 768 sex offenders from Austria, less than 5% had a diagnosis of sexual sadism. A cut score equal to or higher than 6 would, therefore, correspond to a sadism base rate more in accordance with the one observed in prior comparable studies. Furthermore, it must be noted that, as a lower cut score in the SeSaS would result in more individuals identified as sadists, when using the tool for decision making, more persons could be subjected to harsher criminal justice measures and more intensive treatment programs, which would have repercussion for the offenders in terms of their human rights if they were not correctly diagnosed (false positives), as well as costs for the criminal justice system.

Contrary to our research hypothesis, the SeSaS was not significantly related to the hazard of any recidivism outcome. Prior research evidenced that the SeSaS (Brouillette-Alarie et al., 2018; Eher et al., 2016; Study 2) and other behavioral measures (Kingston et al., 2010) of sexual sadism were related to reoffending, but the effects were small and lost significance after controlling for other variables generally associated to recidivism. Overall, the findings indicate that sexual sadism (measured with the SeSaS, other behavioral measures, and diagnostic manuals) has a negligible effect on reoffending and may be the results of other correlates associated with both reoffending and sexual sadism, and not a direct result of sadism (Eher et al., 2016; Kingston et al., 2010).

Eher et al. (2016) argued that the failure of the SeSaS to predict recidivism beyond actuarial risk assessment tools might be because all the relevant violence risk items in the scale are subsumed by such instruments. In addition, sadistic sex offenders might be released into the community at a considerably older age than nonsadistic offenders and, because age is negatively related to reoffending, at the time of release, sexual sadists may have a lower recidivism risk (Eher et al., 2016). However, this did not appear to be the case in the current study because there was no significant difference regarding time incarcerated ($z = 0.06, p = .952$) or age at release ($z = 1.40, p = .162$) between sex offenders with an SeSaS score ≥ 4 ($n = 25$) and sex offenders with an SeSaS score < 4 ($n = 154$). Yet, another possible explanation is that sexual sadists, due to the nature of their crimes, might receive more intensive correctional treatment and be more controlled in the community, by both probation services and the surrounding population, which may also reduce their likelihood to reoffend. Additional research is necessary to test this hypothesis and potential others that may explain the link (or lack of it) between sexual sadism and reoffending.

Limitations

The major limitation of this study is the small size of the sample, and the small number of diagnosed sexual sadists and violent recidivists ($n = 6$, both), which might have

limited the precision of the analyses. Nevertheless, regarding the association between the SeSaS and the clinical diagnosis of sexual sadism, a post hoc power analysis indicated that the sample (six sadists vs. 173 nonsadists) was sufficient to detect group differences in the SeSaS score at a 5% level (one sided) with a power of 80% if the effect size was equal to or larger than Cohen's $d = 1.06$, using the Wilcoxon Mann–Whitney test for two groups as a proxy. A d score of 1.06 represents an AUC value of .77 (Rice & Harris, 2005). Consequently, the sample appears to be large enough for the concurrent validity analysis, given the effect sizes reported in the literature so far (i.e., AUCs > .80).

Similarly, the sample size should be sufficient for the nonmetric IRT scaling analyses as Mokken scaling on the SeSaS items was previously used in a sample of 100 participants (Nitschke et al., 2009). The minimum requirement for Mokken scale analysis to obtain stable results is still unclear, although Straat, van der Ark, and Sijtsma (2014) recommended at least 250 participants when item qualities are high. However, that study refers to the development of new scales from a pool of items using an automated item selection and, therefore, does not apply to the present context, where the analyses were made on preselected items. Nevertheless, replications with larger data sets would improve confidence in the results.

Another limitation pertains to the mixing of different types of sex offenders. The prevalence of sexual sadism is normally lower among child molesters than rapists (as cited in Eher et al., 2016), and the current study included a substantial number of child molesters ($n = 77$) as compared with rapists ($n = 60$) and mixed offenders ($n = 42$). Accordingly, in this sample, child molesters had significantly lower SeSaS scores than rapists ($z = -2.80, p = .005$). Had only rapists been selected, the base rate of sexual sadism would have been higher, providing higher statistical power. Unfortunately, due to the small size of the available sample, the analyses could not be stratified by sex offender type. Future research on the properties of the SeSaS may want to focus on more homogeneous samples (Eher et al., 2016).

In addition, only individuals released into the community were selected for the current study, which led to the exclusion of offenders judged as too high risk to be released from prison (perhaps, including a number of sexual sadists), potentially leading to the decreased accuracy of the SeSaS. The low sadism base rate in the present study might also be partially explained by this sampling criterion.

Implications and Future Directions

Despite limitations, the results of the current study indicate that the SeSaS is a psychometrically sound measure of sexual sadism. The items appear to measure a dimensional construct of sadism with some evidence of cumulative property, although the ordering of the items might vary across different samples. The results also indicate that the SeSaS is reliable, and a score equal to or higher than 4 might be used to identify individuals with a higher level of sexual sadism. However, additional research is necessary to establish whether this cut score is the most appropriate. Conversely, the SeSaS does not seem useful to assess recidivism in crime. Assessment tools such as

the PCL-R and the Static-99 are more appropriate for this purpose. Still, the SeSaS might be used for the dimensional diagnosis of sadism, which could increase the transparency of the criteria used in the assessments, as well as their validity and reliability. Furthermore, the use of the SeSaS might be more convenient, as coding 11 items using information from official records could take considerably less time and require less expertise than the clinical diagnosis of sexual sadism.

Although the SeSaS appears to have adequate psychometric properties, more research is necessary, especially to draw conclusions about the ordering of its items. Future research should also examine other sadism tools, such as the Massachusetts Treatment Center (MTC) Sadism Scale (Longpré et al., 2017), to ascertain which one is most useful. Importantly, the etiology of sexual sadism remains uncertain (DeLisi et al., 2017). Studies on the causation of this disorder, its association with other pathologies, and the characteristics of sexual sadists are necessary to develop more accurate theories, assessment tools, and treatment programs for this population (for a review on sadism theories, see Yates, Hucker, & Kingston, 2008). Finally, the present findings contribute to the accumulating literature evidencing sexual sadism as a dimensional construct and the usefulness of behavioral indicators from crime scenes to assess this disorder (Longpré et al., 2017; Mokros et al., 2014; Nitschke et al., 2009). Additional research would help supporting the inclusion of behaviorally based scales in diagnostic manuals aiming to improve the validity and reliability of sexual sadism diagnoses (Nitschke et al., 2013).

Authors' Note

The present article has not been published elsewhere and is not currently under consideration by any other journal. The authors take responsibility for the integrity of the data, the accuracy of the data analyses, and have made every effort to avoid inflating statistically significant results.

Acknowledgments

The authors thank Cornel Gmür for his collaboration on a prior version of this study, Madeleine Kirschstein for her comments on the manuscript and assistance in revising the text, and BioScience Writers LLC. for editing the final version of the article.

Declaration of Conflicting Interests

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

Funding

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

Supplemental Material

Supplemental material for this article is available online.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Berner, W., Berger, P., & Hill, A. (2003). Sexual sadism. *International Journal of Offender Therapy and Comparative Criminology*, *47*, 383-395. doi:10.1177/0306624X03256131
- Bradley, R., Shedler, J., & Westen, D. (2006). Is the appendix a useful appendage? An empirical examination of depressive, passive-aggressive (negativistic), sadistic, and self-defeating personality disorders. *Journal of Personality Disorders*, *20*, 524-540. doi:10.1521/pedi.2006.20.5.524
- Brouillette-Alarie, S., Proulx, J., & Hanson, R. K. (2018). Three central dimensions of sexual recidivism risk: Understanding the latent constructs of Static-99R and Static-2002R. *Sexual Abuse: A Journal of Research and Treatment*, *30*, 676-704. doi:10.1177/1079063217691965.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284-290. doi:10.1037/1040-3590.6.4.284
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334. doi:10.1007/BF02310555
- DeLisi, M., Drury, A., Elbert, M., Tahja, K., Caropreso, D., & Heinrichs, T. (2017). Sexual sadism and criminal versatility: Does sexual sadism spillover into nonsexual crimes? *Journal of Aggression, Conflict and Peace Research*, *9*, 2-12. doi:10.1108/JACPR-05-2016-0229
- Doren, D., & Elwood, R. (2009). The diagnostic reliability of sexual sadism. *Sexual Abuse: A Journal of Research and Treatment*, *21*, 251-261. doi:10.1177/1079063209342072
- Eher, R., Schilling, F., Hansmann, B., Pumberger, T., Nitschke, J., Habermeyer, E., & Mokros, A. (2016). Sadism and violent reoffending in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *28*, 46-72. doi:10.1177/1079063214566715
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi:10.3758/BF03193146
- Fedora, O., Reddon, J. R., Morrison, J. W., Fedora, S. K., Pascoe, H., & Yeudall, L. T. (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
- Fedoroff, P. J. (2008). Sadism, sadomasochism, sex, and violence. *The Canadian Journal of Psychiatry*, *53*, 637-646. doi:10.1007/BF01542713
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378-382. doi:10.1037/h0031619
- Frances, A., & Wollert, R. (2012). Sexual sadism: Avoiding its misuse in sexually violent predator evaluations. *Journal of the American Academy of Psychiatry and the Law Online*, *40*, 409-416.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139-150. doi:10.2307/2086306
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255-282. doi:10.1007/BF02288892
- Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders (1999-02)*. Ottawa, Ontario: Solicitor General of Canada.

- Hardouin, J.-B., Bonnaud-Antignac, A., & Sébille, V. (2011). Nonparametric item response theory using Stata. *Stata Journal*, *11*, 30-51.
- Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised (PCL-R)*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hargreaves, C., & Francis, B. (2014). The long term recidivism risk of young sexual offenders in England and Wales: Enduring risk or redemption? *Journal of Criminal Justice*, *42*, 164-172. doi:10.1016/j.jcrimjus.2013.06.017
- Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *The Journal of the American Medical Association*, *247*, 2543-2546. doi:10.1001/jama.1982.03320430047030
- Kingston, D. A., Seto, M. C., Firestone, P., & Bradford, J. M. W. (2010). Comparing indicators of sexual sadism as predictors of recidivism among adult male sexual offenders. *Journal of Consulting and Clinical Psychology*, *78*, 574-584. doi:10.1037/a0019734
- Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, *45*, 507-529. doi:10.1037/h0055827
- Longpré, N., Guay, J.-P., & Knight, R. A. (2017). MTC Sadism Scale: Toward a dimensional assessment of severe sexual sadism with behavioral markers. *Assessment*. doi:10.1177/1073191117737377
- Longpré, N., Proulx, J., & Brouillette-Alarie, S. (2018). Convergent validity of three measures of sexual sadism: Value of a dimensional measure. *Sexual Abuse: A Journal of Research and Treatment*, *30*, 192-208. doi:10.1177/1079063216649592
- Marshall, W. L., & Hucker, S. J. (2006). Issues in the diagnosis of sexual sadism. *Sexual Offender Treatment*, *1*(2). Retrieved from <http://www.sexual-offender-treatment.org/40.0.html>
- Marshall, W. L., & Kennedy, P. (2003). Sexual sadism in sexual offenders: An elusive diagnosis. *Aggression and Violent Behavior*, *8*, 1-22. doi:10.1016/S1359-1789(01)00052-0
- Marshall, W. L., Kennedy, P., & Yates, P. (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*, 301-311. doi:10.1177/107906320201400402
- Marshall, W. L., Kennedy, P., Yates, P., & Serran, G. (2002). Diagnosing sexual sadism in sexual offenders: Reliability across diagnosticians. *International Journal of Offender Therapy and Comparative Criminology*, *46*, 668-677. doi:10.1177/107906320201400402
- Marshall, W. L., & Yates, P. (2004). Diagnostic issues in sexual sadism among sexual offenders. *Journal of Sexual Aggression*, *10*, 21-27. doi:10.1080/13552600410001470973
- Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. The Hague, The Netherlands: Mouton.
- Mokros, A., Schilling, F., Eher, R., & Nitschke, J. (2012). The Severe Sexual Sadism Scale: Cross-validation and scale properties. *Psychological Assessment*, *24*, 764-769. doi:10.1037/a0026419
- Mokros, A., Schilling, F., Weiss, K., Nitschke, J., & Eher, R. (2014). Sadism in sexual offenders: Evidence for dimensionality. *Psychological Assessment*, *26*, 138-147. doi:10.1037/a0034861
- Mokros, A., & Stefanska, E. (2017, March). *Sexual sadism among sexual murderers in the United Kingdom: An item response theory analysis (ICPS)*. Paper presented at the 2nd Biennial International Convention of Psychological Science, Vienna, Austria.
- Murray, A., McKenzie, K., Booth, T., & Murray, G. (2013). Estimating the level of functional ability of children identified as likely to have an intellectual disability. *Research in Developmental Disabilities*, *34*, 4009-4016. doi:10.1016/j.ridd.2013.08.008

- Nitschke, J., Mokros, A., Osterheider, M., & Marshall, W. L. (2013). Sexual Sadism: Current diagnostic vagueness and the benefit of behavioral definitions. *International Journal of Offender Therapy and Comparative Criminology*, *57*, 1441-1453. doi:10.1177/0306624X12465923
- Nitschke, J., Osterheider, M., & Mokros, A. (2009). A cumulative scale of severe sexual sadism. *Sexual Abuse: A Journal of Research and Treatment*, *21*, 262-278. doi:10.1177/1079063209342074
- Pflugradt, D. M., & Allen, B. P. (2013). Identifying sadists among female sexual offenders using the cumulative scale of Severe Sexual Sadism. *Sexual Offender Treatment*, *8*(1). Retrieved from <http://www.sexual-offendertreatment.org/116.html>
- 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.
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org/>
- 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
- Richards, H., & Jackson, R. L. (2011). Behavioral discriminators of sexual sadism and paraphilia nonconsent in a sample of civilly committed sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, *55*, 207-227. doi:10.1177/0306624X10377073
- Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika*, *52*, 79-97. doi:10.1007/BF02293957
- StataCorp. (2013). Stata statistical software: Release 13. College Station, TX: Author.
- Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, *74*, 809-822. doi:10.1177/0013164414529793
- Urbaniok, F., Endrass, J., Rossegger, A., Noll, T., Gallo, W. T., & Angst, J. (2007). The prediction of criminal recidivism: The implication of sampling in prognostic models. *European Archives of Psychiatry and Clinical Neurosciences*, *257*, 129-134. doi:10.1007/s00406-006-0678-y
- Van Schuur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis*, *11*, 139-163. doi:10.1093/pan/mpg002
- World Health Organization (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Belgium: World Health Organization.
- Yates, P. M., Hucker, S. J., & Kingston, D. A. (2008). Sexual sadism: Psychopathology and theory. In D. R. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (2nd ed., pp. 213-230). New York, NY: Guilford Press.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32-35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
- Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*, 213-249. doi:10.1007/BF02294536