Examining the scope of questionable diagnostic reliability in Sexually Violent Predator (SVP) evaluations

Anthony D. Perillo [a,*], Ashley H. Spada [a], Cynthia Calkins [b], Elizabeth L. Jeglic [b]

[a] John Jay College and the Graduate Center, CUNY, New York, NY, United States
[b] John Jay College of Criminal Justice, New York, NY, United States

ABSTRACT

Research has suggested questionable reliability of diagnosing mental abnormality during Sexually Violent Predator (SVP) evaluations, despite this being a necessary requirement for SVP commitment. Findings have been inconsistent across studies, and little is known about the extent of such trends across diagnoses and clinicians. The current study includes data from 375 sex offenders referred for evaluation for SVP commitment in New Jersey. Clinicians ($n = 128$) rendered a variety of diagnoses, most commonly Pedophilia. Results suggested questionable agreement across paraphilic and non-paraphilic diagnoses, although agreement was fair for diagnoses of Pedophilia. Further analyses of cases ($n = 49$) involving clinicians receiving a large number of referrals ($n = 14$) were generally consistent with these findings, with no outlier effect apparent. Findings suggest questionable diagnostic reliability to be a widespread issue in SVP evaluations, present across a variety of diagnoses and across the general body of clinicians involved in evaluations.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Decisions pertaining to Sexually Violent Predator (SVP) commitment, the indefinite civil detainment of at-risk sex offenders, determine which sex offenders will be released into the outside community and which sex offenders will face indefinite detainment following completion of their prison sentence. Some of the struggles incurred by offenders as a result of SVP legislation include the indefinite loss of freedom as well as the inevitable challenge—should they eventually be released—of reintegrating into society (Schiavone & Jeglic, 2009). For states, SVP commitment has also become a heavy financial burden, with an average cost of $100,000 annually to civilly commit a single sex offender (La Fond, 2005; Rappleye, 2012). Further, these costs are expected to increase exponentially because of low release rates and increased expenditures associated with an aging population (James, Thomas, & Foley, 2007). Given the extraordinarily high costs (financial and beyond) of SVP commitment, it is critical that commitment decisions be based upon sound evaluation practices and accurate diagnosis. Overall, proper evaluation of those sex offenders at highest risk to commit future acts of sexual violence after release will maximize the impact of civil commitment on public safety, efficiency, and justice.

2. Criteria for SVP commitment

At the present time, 21 U.S. jurisdictions (20 states and the federal government; see Table 1) have active SVP commitment statutes, all of which are similar in nature to the original civil commitment statute enacted by Washington state. To be civilly committed as a Sexually Violent Predator, most statutes require that an offender be deemed too "sexually dangerous" to be released upon completion of their prison sentence (e.g., Commitment of Sexually Violent Predators, 2012[1]). The United States Supreme Court has upheld the constitutionality of SVP statutes (*Kansas v. Crane*, 534 U.S. 407, 2002[2]; *Kansas v. Hendricks*, 117 S. Ct. 2072, 1997[3]; *United States v. Comstock*, 551 F. 3d 274, 2010[4]) and, by result, upheld the legal criteria for defining sexual dangerousness (Doren & Elwood, 2009; First & Halon, 2008; Levenson, 2004b). Statutory criteria typically set forth that individuals eligible for SVP commitment must meet the following three criteria: 1) have a history of criminal sexual behavior, 2) suffer a mental abnormality or personality disorder that includes some form of volitional impairment, and 3) have an identifiable link between the mental abnormality and a high risk of future sexual violence (Janus, 2000).

---

[1] Commitment of Sexually Violent Predators, 2 Kan. Stat. Ann. § § 59-29a01-29a22 (2012).
[2] Kansas v. Crane, 534 U.S. 407 (U.S. Supreme Court 2002).
[3] Kansas v. Hendricks, 117 S. Ct. 2072 (U.S. Supreme Court 1997).
[4] United States v. Comstock, 551 F. 3d 274 (2010).

* Corresponding author. Tel.: +1 212 237 8252.
  E-mail address: aperillo@jjay.cuny.edu (A.D. Perillo).

Table 1
Civil commitment statutes by state (as of 2012).

State	Year	Title
Arizona	1996	Arizona Revised Statutes 13
California	1996	Welfare and Justice Code, sec. 6600 et seq.
Florida	1998	Jimmy Ryce Act
Illinois	1997	Sexually Violent Persons Commitment Act
Iowa	1998	Iowa Code, Chapter 229A
Kansas	1994	Sexually Violent Predator Act
Massachusetts	1998	Part I Title XVII, Chap. 123A et seq.
Minnesota	1994	Sexually Dangerous Persons Stat. § 253B.02
Missouri	1999	Sexually Violent Predator Act
Nebraska	2006	Neb. Rev. St. § 29-4005
New Hampshire	2007	House Bill 138
New Jersey	1994	Sexually Violent Predators Act (SVPA)
New York	2007	Sex Offender Management and Treatment Act
North Dakota	1997	House Bill 138
Pennsylvania	2003	42 Pa.C.S. § 6358
South Carolina	1998	General Bill 4360
Texas	1999	Health and Safety Code 841.001 et seq.
Virginia	2004	37.1-70 et seq.
Washington	1990	Community Protection Act
Wisconsin	1994	Wisconsin Statutes, Chapter 980
Federal	2006	Adam Walsh Child Protection and Safety Act

Identifying sex offenders who meet the first criterion is a fairly straightforward process. Under most SVP statutes, the offender must, at a minimum, have been convicted of one or more statutorily defined sexual offenses against a child or an adult victim (Jumper, Babula, & Casbon, 2011; Lieb & Matson, 1998). Components from the final two criteria (suffering a mental abnormality or personality disorder and being at high-risk for future sexual violence) are more complex and may require subjective judgment by way of a psychological evaluation conducted by a clinician. Of these final two criteria, far more research has been dedicated to evaluating risk for future sexual violence (Lieb & Matson, 1998). Actuarial risk assessment tools are a widely accepted and commonly used method for assessing the likelihood of future sexually violent behavior (Douglas, Cox, & Webster, 1999; Levenson, 2004b), and dynamic risk factors are also regularly integrated into the assessment of future dangerousness (Levenson & Morin, 2006). Analysis of civil commitment decisions in New Jersey and Florida indicates that sex offenders recommended for commitment score higher on actuarial risk scales such as the Static-99 (Hanson & Thornton, 1999) and MnSOST—R (Epperson, Kaul, Huot, Goldman, & Alexander, 2003), suggesting that determinations regarding who meets the criteria for high risk of future sexual violence generally follow empirically supported processes (Calkins, Jeglic, & Perillo, submitted for publication; Levenson, 2004b). Notably, however, high levels of variance have been observed among clinician scores on sex offender risk instruments (Murrie et al., 2009), bringing into question the interrater reliability on more standardized practices of sex offender risk assessment.

3. Diagnosing mental abnormalities in sex offenders

The final two criteria of SVP commitment require that the individual exhibit a mental abnormality or personality disorder that predisposes him or her to commit future "predatory acts of sexual violence" (*Kansas v. Hendricks*, 1997). That is, a risk of sexual dangerousness is not enough to conclude that an individual meets the criteria for civil commitment; the individual must also demonstrate mental dysfunction that exacerbates this risk. These criteria are thought to specify the pool of offenders eligible for civil commitment and have directly contributed to the Court's ruling that civil commitment statutes are indeed constitutional (First & Halon, 2008). As the term *mental abnormality* defined by the legal system is not commonly used in psychiatric circles, clinicians typically evaluate offenders for the presence of a qualifying impairment, specifically, a mental disorder or personality disorder as defined by the *DSM-IV* (American Psychiatric Association [APA], 1994) to determine

whether a mental abnormality is present (Levenson, 2004b; Wollert, 2007). Although this general consensus promotes consistency for approaching determinations of mental abnormality during SVP evaluations, the *DSM-IV* is a helpful guide in assigning diagnoses only to the extent that the clinicians who utilize the manual agree on what the criteria mean and who qualifies for the diagnoses (Doren & Elwood, 2009).

The reliability of the *DSM-IV* in forming diagnoses has been questioned through several empirical studies. Examinations of interrater reliability with *DSM-IV* diagnoses in general have revealed a kappa of .64 across seven data sets, indicating "fair" agreement (Levenson, 2004b; Meyer, 2002; Packard & Levenson, 2006). The use of semi-structured interviews has served to improve the reliability of the *DSM-IV* Axis I diagnoses (Meyer, 2002; Packard & Levenson, 2006). Fully structured joint interviews produce higher interrater reliability; however, this process relies heavily on patient self-report and removes much clinical judgment from the diagnostic formulation, which may prove to be more consistent but less accurate (Meyer, 2002).

4. Challenges in diagnosing paraphilic disorders

With regard to diagnoses in the sex offender civil commitment context, paraphilic disorders are the most commonly cited mental conditions as those that predispose offenders to future sexually violent behavior. The most commonly observed paraphilic disorders among civilly committed sex offenders are Pedophilia and Paraphilia Not Otherwise Specified (NOS), with base rates of 49% and 47%, respectively (Doren & Elwood, 2009; Jumper et al., 2011; Levenson & Morin, 2006). Other, less commonly cited diagnoses among civilly committed sex offenders include Sexual Sadism (6.4%), Voyeurism (8.6%), Exhibitionism (7.8%), Frotteurism (2.6%), and Fetishism (3.1%; Jumper et al., 2011).

Reliability of diagnoses as outlined in the *DSM-IV* has been criticized because of the potential to misapply or use paraphilic diagnoses in an overly broad manner (First & Halon, 2008; Levenson, 2004b). Diagnosis of a paraphilic disorder, in general, requires deviant "sexually arousing fantasies, sexual urges, or behaviors that involve nonhuman objects, suffering/humiliation, or children or other nonconsenting persons, with symptoms present for a period of at least six months" (APA, 2000). One potential concern is that a paraphilic diagnosis might be prematurely applied when there is simply evidence of criminal sexual behavior without conclusive knowledge of the underlying state of mind or motivation that led to the behavior. Convictions for child molestation or rape are not evidence enough, in and of themselves, that the individual has a deviant sexual arousal pattern (First & Halon, 2008). That is, not all sexual offenders will inevitably meet criteria for a paraphilic diagnosis. Among a sample of 450 adult male sex offenders evaluated for civil commitment in Florida, 277 (62%) were diagnosed with a paraphilia (Levenson, 2004b), while 38% of offenders had no paraphilic diagnosis on record. Similarly, among another sample of 113 male sex offenders, a slight majority (58%) were given a paraphilic diagnosis (Dunsieth et al., 2004). As these trends suggest, the motivation behind sexually deviant behavior cannot be determined solely from the behavior on record. Impulsivity, aggression, negative/hostile masculinity, entitlement, impersonal sex, and disregard for others are just a few motivations that may explain a criminal sexual act and thus not meet criteria for a paraphilic diagnosis (First & Halon, 2008).

Another issue with diagnosing a paraphilia is that the sexual deviance in question must be evident for at least six months (Levenson, 2004a). Several challenges emerge when attempting to establish a timeframe for sexually deviant behavior. For example, many sexual offenses go unreported, which can impede on identifying the duration of behavioral trends potentially important to assessing for a paraphilic diagnosis. At the other end of the spectrum, an offender with no apparent sexually deviant arousal pattern but who sexually abuses, for example, a child more than once over a time period greater than six months,

would meet the six month time period criteria for a paraphilia diagnosis (Levenson, 2004a).

The subjectivity required for a clinician to evaluate whether the deviant fantasies and behaviors are "recurrent" and "intense," as well as whether the disorder results in "distress" or "impairment" for the offender, presents yet another challenge that can impact the reliability of the paraphilia diagnoses (First & Halon, 2008; Levenson, 2004a). The ambiguity of the descriptors "recurrent" and "intense" requires clinicians to make subjective inferences about deviant sexual fantasies and behaviors that may not be on record and may not be disclosed by the sex offender being evaluated (Levenson, 2004a). Further, in addition to this overall subjectivity, "distress" and "impairment" due to arrest and criminal justice involvement must be differentiated from that of the paraphilia itself, as only distress and impairment that stem directly from the disorder qualify as meeting the diagnostic criteria for a paraphilia (Levenson, 2004a).

An additional diagnostic issue is the potentially inconsistent and overly broad use of the Paraphilia NOS diagnosis. There is no specific diagnosis in the *DSM-IV* for offenders who commit rape, and this has resulted in the increased use of the Paraphilia NOS, nonconsent, diagnosis for sex offenders who commit rape but do not appear to meet criteria for Sexual Sadism (Levenson, 2004a). This can lead to misapplication of the Paraphilia NOS diagnosis to any offender convicted of rape, when the diagnosis was intended to apply to offenders whose sexual arousal is thought to be directly related to the nonconsent of the victim (Prentky, Coward, & Gabriel, 2008). Further, evaluators often rely on the Paraphilia NOS, Hebephilia, diagnosis for offenders with victims aged 13–15, as such offenders would not meet the victim age criteria for Pedophilia (Levenson, 2004a). Because the Paraphilia NOS diagnosis is not as clearly defined as other paraphilic disorders, clinicians must utilize a large degree of subjectivity, thus adding further ambiguity to the reliability of this diagnosis.

5. Diagnostic issues in SVP evaluations

Given these issues in rendering paraphilic diagnoses, diagnosing sex offenders with a mental abnormality that may increase their risk for sexually reoffending (as required by the civil commitment statutes) may not be as reliable a process as is desired. Such an issue is problematic, as paraphilic diagnoses are significantly associated with the clinician recommendations for civil commitment recommendations (Levenson & Morin, 2006). Research aiming to empirically examine the reliability of *DSM-IV* diagnoses during civil commitment evaluation has not been very positive, as agreement among clinicians has generally been considered questionable at best (Doren & Elwood, 2009; Levenson, 2004a; Marshall, Kennedy, Yates, & Serran, 2002).

In 2004, Levenson evaluated the interrater reliability of Sexually Violent Predator civil commitment criteria in the state of Florida among 25 licensed psychologists and psychiatrists. Levenson's results indicated fair reliability for diagnosing Pedophilia and poor reliability for diagnosing Sexual Sadism, Exhibitionism, and Paraphilia NOS, with kappa values for paraphilic diagnoses ranging from .23 to .70 (Levenson, 2004a). Further, the interrater reliability of civil commitment recommendation was poor, revealing that clinicians often disagreed not only on the identified mental abnormality, but also on the ultimate evaluation recommendations (Levenson, 2004a).

While the kappa coefficient is a commonly used and accepted measure, research has highlighted several concerns about reliance on kappa as a measure of diagnostic reliability (see Packard & Levenson, 2006). As such, Packard and Levenson reexamined the civil commitment data from Florida calculating odds ratios, relative risk ratios, and levels of uncertainty to provide a clearer examination of diagnostic reliability. The results of their reexamination revealed higher levels of agreement on diagnoses than was suggested by the kappa coefficients, as well as high agreement in final recommendations for civil commitment as SVPs. Namely, 85% agreement rates

were revealed for the presence of Pedophilia and 68% for the presence of Paraphilia NOS. The authors thus concluded that SVP evaluations appear to be sufficiently reliable. Further analyses of the Florida civil commitment data, however, supported Levenson's (2004a) initial findings of diagnostic unreliability among clinicians during the SVP commitment process (Wollert, 2007).

Beyond the aforementioned studies, there is little research detailing the reliability of diagnoses for sexual offenders, within or outside the SVP commitment context. In an examination of diagnostic agreement among twelve sex offender cases, clinicians agreed approximately 92% of the time regarding the presence of a paraphilic diagnosis in general (Doren & Elwood, 2009). In an examination of the relatively uncommon diagnosis of Sexual Sadism, Marshall et al. (2002) found rates of diagnostic agreement ranging from 60% to 100%.

In addition to the lack of clear and consistent trends of diagnostic reliability for disorders most commonly rendered to sex offenders, it is unclear how strongly questions about diagnostic reliability in SVP evaluations concern the general body of clinicians conducting the evaluations. Comparisons across clinicians will often include those with different degrees of experience. Clinical experience has a small but significant association with increased precision of clinical diagnoses (Spengler et al., 2009), lending support to the possibility that diagnostic reliability among high-referral clinicians is stronger than that of the general body of clinicians. On the other hand, agreement across clinicians can be significantly (and, arguably, artificially) compromised should a clinician frequently involved in evaluations offer judgments that consistently exist in contrast to those of other clinicians. Indeed, in an examination of the predictive validity of PCL-R scores during sex offender risk assessments, initially substandard results improved when focusing on a subset of clinicians who had conducted a proportionally higher number of the evaluations and removing a high-referral clinician whose findings appeared to differ from others (Murrie, Boccaccini, Caperton, & Rufino, 2012). In the context of SVP evaluations, it is possible that experienced evaluators generally agree on the diagnoses of offenders, with inconsistencies attributable to an outlier clinician rendering diagnoses in conflict with others' diagnoses. Such trends would impact the scope of any recommendations for addressing diagnostic reliability in SVP evaluations.

6. Current study

The current study examines the reliability of diagnoses among clinicians making diagnostic decisions concerning a sample of New Jersey sex offenders. Given the research of Levenson and colleagues and the overall subjective nature of paraphilic diagnoses, we hypothesized that interrater reliability would be questionable across most, if not all, diagnostic categories rendered by the clinicians in our study. With this hypothesis in mind, the current study expands upon this global assessment of diagnostic reliability to examine whether any questionable findings can be attributed to outlier clinicians that disagree with a body of clinicians that otherwise display high agreement. In the current study, follow-up analyses focusing on clinicians who conducted a proportionally higher number of evaluations attempt to address the extent that any questionable diagnostic reliability presents as a broad issue across clinicians. In doing so, the current study clarifies whether any questionable findings pertaining to diagnostic reliability can be attributed to a few outlier clinicians or can be considered a widespread issue, which carries significant implications for how to address any concerns with diagnostic reliability.

7. Method

7.1. Sample

Archival case data were collected from the New Jersey Department of Corrections (NJDOC) files of male offenders incarcerated for a sexual

offense within the NJDOC and scheduled to be released from correctional custody between the years 1996 and 2007. Files were reviewed for all male sex offenders who during the years above were housed in a New Jersey state prison or the sex offender treatment facility of the NJDOC. In addition, file data were gathered for those offenders in custody of the Special Treatment Unit (STU), the facility that houses those committed as SVPs in New Jersey. The current study utilizes a sample of 375 offenders who were evaluated for SVP commitment by two independent mental health professionals during the timeframe of the investigation. Cases included for analysis were thus subjected to two independent forensic evaluations complete with independent diagnostic assessments.

The majority of offenders (55.9%) were classified as White, 34.8% as Black, 8.1% as Hispanic, and 1.2% as other ethnic minorities. Offenders had been incarcerated for a variety of sex offenses, with 93.5% of offenders incarcerated for contact sex offenses at the time of evaluation. A notable portion of offenders had a documented history of psychiatric treatment prior to incarceration (46.1%).

Data were analyzed from 128 mental health professionals who conducted SVP evaluations on the offenders in the current study. The majority (71.4%) of these clinicians were male. In terms of the level of training of evaluators across the sample, 40.3% of evaluations involved at least one Ph.D.-level clinician, 24.2% involved a PsyD.-level clinician, 88.8% a M.D.-level clinician, and 11.0% a Master's-level clinician.

7.2. Procedures

Implemented in 1998, the New Jersey Sexually Violent Predator Act (NJ SVP Act) calls for the indefinite civil commitment of persons with histories of sex offenses and who "suffer from mental abnormalities or personality disorders which make them likely to engage in repeat acts of predatory sexual violence if not treated for their mental conditions" (Sexually Violent Predator Act, 2004[5]). Near the completion of their sentence, sex offenders within the NJDOC may be petitioned by the district attorney to be subject to commitment as an SVP upon their scheduled release from incarceration. Those petitioned by the district attorney are referred for evaluation by mental health professionals. These evaluators are responsible for performing independent psychological evaluations that include clinical diagnosis, risk assessment, and recommendations regarding SVP commitment.

As part of a larger study with Institutional Review Board approval from the researchers' host institution and the NJDOC (Mercado, Jeglic, & Markus, 2011), a group of trained psychology graduate students reviewed the offender files and completed data collection tools. Prior to completing data collection tools, records were deidentified by replacing all identifying information with a case identification number. Data collection tools were developed for the multiple sites within the NJDOC to code for offender and offense characteristics (e.g., demographic information, offense history, index offense characteristics, victim information) and SVP evaluation characteristics available in the offender files. Information extracted pertaining to SVP evaluations included demographic information for each evaluating clinician and conclusions offered by each clinician (including diagnoses rendered, risk scores, and recommendations for commitment). The current study focused on the reliability of diagnostic categories offered during SVP evaluations. That is, the results focused on clinician agreement as to whether a given diagnosis was present or not present for a given offender, not total agreement on the entire set of diagnoses that clinicians offered when evaluating a given offender. All data were coded dichotomously (no–yes) for the presence of each diagnosis or diagnostic category included by each clinician.

---

[5] Sexually Violent Predator Act, N.J. Stat. Ann. § § 30:4–27.24–27.38 (2004).

Table 2
Prevalence of diagnoses during SVP evaluations.

Diagnosis	Prevalence (%) either clinician	Prevalence (%) both clinicians
Pedophilia	48.4	27.1
Substance Disorder	31.1	11.1
Paraphilia NOS	31.1	9.8
Antisocial Personality Disorder	28.0	12.9
Other Personality Disorder	13.3	2.2
Sexual Sadism	5.3	0.9
Other Sexual Disorder	3.1	0.4
Exhibitionism	2.7	0.0

8. Results

8.1. Prevalence of diagnostic categories

Clinicians offered a variety of paraphilic and non-paraphilic diagnoses for offenders evaluated for SVP commitment (see Table 2). Almost half of the sex offenders in our sample (48.4%) were diagnosed with Pedophilia by at least one clinician, with 27.1% of offenders diagnosed with Pedophilia by both evaluating clinicians. The other commonly diagnosed paraphilic disorder was Paraphilia NOS; almost one-third of the offenders (31.1%) were diagnosed with Paraphilia NOS by at least one clinician and 9.8% by both evaluating clinicians. Diagnoses of Sexual Sadism, Exhibitionism, and other sexual disorders were less commonly rendered.

Among non-paraphilic diagnoses, almost one-third of the offenders (31.1%) were rendered a substance disorder by at least one clinician, although a substantially lower rate was diagnosed as such by both evaluating clinicians (11.1%). The most commonly diagnosed personality disorder was Antisocial Personality Disorder, with more than one quarter of offenders (28.0%) rendered the diagnosis by at least one clinician. Other personality disorders were far less likely to be diagnosed in this sample.

8.2. Diagnostic reliability

8.2.1. Reliability in overall sample

Table 3 presents reliability statistics for each diagnostic category. Kappa coefficients were calculated to measure interrater agreement for each diagnostic category. All kappa coefficients were statistically significant ($p < 0.01$) except for that of Exhibitionism ($p = 0.84$). In fact, although six offenders were diagnosed with Exhibitionism, there were no cases in which both clinicians rendered the diagnosis to the same offender. Using the Bloom, Fischer, and Orme (1999) standard for kappa agreement ("poor" = below 0.60, "fair" = 0.60–0.74, and "good" = 0.75 and above), all diagnostic categories exhibited "poor" reliability. Diagnostic reliability approached "fair" for Pedophilia (kappa = 0.55) and Antisocial Personality Disorder (kappa = 0.54). Using the Landis and Koch (1977) standard, results suggest "moderate" reliability for Pedophilia and Antisocial Personality Disorder and "fair" reliability for all other diagnostic categories (except Exhibitionism).

The odds ratios demonstrate the likelihood that when a clinician rendered a given diagnosis, the other clinician would agree by issuing the same diagnosis. For example, when one clinician offered a diagnosis of Pedophilia, another clinician was 13.43 times more likely to agree (by also diagnosing Pedophilia) than to disagree (by not diagnosing Pedophilia). For all diagnostic categories with significant kappa coefficients, odds ratios were significantly above 1.00. This trend suggests that for all of these diagnostic categories, a second clinician was more likely to concur with a diagnosis offered by another clinician than to disagree with this diagnosis. The magnitude of this trend varied across diagnoses, with the most robust odds ratios for Sexual Sadism (OR = 17.75), Antisocial Personality Disorder

Table 3
Clinical diagnostic reliability (overall).

Diagnosis	Kappa (95% C.I.)	Odds ratio (95% C.I.)	Pos. predictive value (95% C.I.)	Neg. predictive value (95% C.I.)
Pedophilia	0.55[*,a1,b2] (0.44–0.66)	13.43 (6.89–26.18)	0.78 (0.67–0.86)	0.79 (0.71–0.85)
Paraphilia NOS	0.35[*,a1,b2] (0.20–0.50)	6.09 (2.94–12.60)	0.52 (0.37–0.68)	0.85 (0.78–0.89)
Antisocial Personality Disorder	0.54[*,a1,b2] (0.40–0.68)	16.48 (7.53–36.10)	0.60 (0.45–0.74)	0.92 (0.86–0.95)
Other Personality Disorder	0.23[*,a1,b2] (0.03–0.43)	7.74 (2.23–26.88)	0.42 (0.16–0.71)	0.92 (0.87–0.95)
Substance Disorder	0.40[*,a1,b2] (0.25–0.55)	7.66 (3.72–15.75)	0.53 (0.38–0.68)	0.87 (0.81–0.91)
Sexual Sadism	0.26[*,a1,b2] (−0.04–0.58)	17.75 (2.71–116.48)	0.33 (0.06–0.76)	0.97 (0.94–0.99)
Exhibitionism	−0.01[a1] (0.00–0.02)	–	0.00 (0.00–0.69)	0.99 (0.96–1.00)
Other Sexual Disorder	0.24[*,a1,b2] (−0.22–0.70)	27.25 (2.03–365.49)	0.33 (0.02–0.87)	0.98 (0.95–0.99)

* $p < 0.05$.
[a1] Denotes "poor" agreement per Bloom et al. (1999) standard.
[b1] Denotes "fair" agreement per Landis and Koch (1977) standard.
[b2] Denotes "moderate" agreement per Landis and Koch (1977) standard.

(OR = 16.48), and Pedophilia. An odds ratio could not be computed for Exhibitionism because there were no cases in which both clinicians agreed on the presence of the disorder.

Positive Predictive Values (PPV) illustrate the probability that both clinicians will agree on the presence of a given diagnosis, given that the first clinician offered that diagnosis. A PPV of 0.50 indicates chance agreement, with clinicians no more likely to agree or disagree on the presence of the diagnosis when the diagnosis is offered by the first clinician. The strongest PPV was for Pedophilia, with a 78% likelihood of agreement when the first clinician diagnosed an offender with Pedophilia. Only three other diagnostic categories had PPVs above 0.50: Antisocial Personality Disorder (PPV = 0.60), Substance Disorder (PPV = 0.53), and Paraphilia NOS (PPV = 0.52); however, 95% confidence intervals demonstrate that none of these PPVs were above chance levels of agreement. PPVs for Sexual Sadism, Exhibitionism, Other Personality Disorders, and Other Sexual Disorders were all below chance, suggesting that clinicians were more likely to *disagree* on the presence of that diagnostic category when the first clinician offered the diagnosis. Similarly, Negative Predictive Values (NPV) illustrate the probability that both clinicians will agree on the *absence* of a given diagnosis, given that the first clinician did not offer that diagnosis. NPV trends were consistent across all diagnostic categories: when the first clinician failed to render a diagnosis, the second clinician was significantly more likely to concur by also failing to render that diagnosis.

8.2.2. Reliability among high-referral clinicians

To examine the extent to which findings generalized across the sample of clinicians, the analyses above were conducted with a subset of clinicians who conducted a high proportion of the evaluations (high-referral clinicians). Our sample includes 14 clinicians who had each conducted at least 50 SVP evaluations for the NJDOC. We repeated all analyses for the subset of cases where both evaluating clinicians were high-referral clinicians ($n = 49$). Analysis for diagnoses not observed in this subset of cases could not be completed (Sexual Sadism

and Other Sexual Disorder). Table 4 includes the results of these repeated analyses. Kappa coefficients were significant for the following diagnostic categories: Pedophilia, Antisocial Personality Disorder, and Substance Disorders. Kappa for Exhibitionism remained nonsignificant, and kappa coefficients were no longer significant for Paraphilia NOS, Sexual Sadism, and Other Personality Disorder. Using the Bloom et al. (1999) standard, kappa coefficients remained in the poor range for all diagnostic categories except for one: Pedophilia, whose kappa was in the "fair" range (kappa = 0.62). Using the Landis and Koch (1977) standard, Pedophilia was elevated to the "substantial" range, while all other categories with significant kappa coefficients remained in the same range as in the whole sample.

Among high-referral clinicians, the PPV and NPV trends for Pedophilia were unlike those for other diagnostic categories. The PPV for Pedophilia (PPV = 0.96) suggested high agreement on the presence of Pedophilia; when the first clinician offered a diagnosis of Pedophilia, there was a 96% likelihood that the clinicians would agree by also offering this diagnosis. Conversely, the NPV for Pedophilia (NPV = 0.64) suggested lower agreement on the absence of Pedophilia; there was only a 64% likelihood of clinician agreement when the first clinician did not diagnose an offender with Pedophilia. For all other diagnostic categories for which analysis could be conducted, PPVs were nonsignificant, suggesting that levels of agreement when the first clinician offered a diagnosis were no better than chance. On the other hand, NPVs were significant and above 0.75 for all other diagnostic categories; when the first clinician did not offer a given diagnosis, clinicians were significantly likely to concur on the absence of this diagnosis. Such trends may be partly reflected by the overall low base rates of many of these diagnostic categories.

8.2.3. Examination of a potential outlier effect

To examine whether any results would be attributed to an outlier effect (i.e., a high-referral clinician rendering diagnoses greatly inconsistent with others and thus artificially compromising reliability results among other clinicians), we again repeated the analyses 14 times,

Table 4
Clinical diagnostic reliability (high-referral clinicians).

Diagnosis	Kappa (95% C.I.)	Odds ratio (95% C.I.)	Pos. predictive value (95% C.I.)	Neg. predictive value (95% C.I.)
Pedophilia	0.62[*,a2,b3] (0.40–0.84)	45.50 (5.15–401.71)	0.96 (0.79–1.00)	0.64 (0.41–0.82)
Paraphilia NOS	0.01[a1] (−0.27–0.29)	1.04 (0.23–4.68)	0.25 (0.07–0.57)	0.76 (0.58–0.88)
Antisocial Personality Disorder	0.45[*,a2,b2] (0.11–0.79)	12.67 (2.06–77.96)	0.50 (0.17–0.83)	0.93 (0.79–0.98)
Other Personality Disorder	0.30[a1,b1] (−0.07–0.67)	13.33 (1.04–170.63)	0.67 (0.13–0.98)	0.87 (0.73–0.95)
Substance Disorder	0.31[*,a1,b1] (0.03–0.58)	4.36 (1.11–17.17)	0.44 (0.21–0.69)	0.85 (0.67–0.94)
Exhibitionism	−0.03[a1] (−0.08–0.02)	–	0.00 (0.00–0.69)	0.98 (0.87–1.00)

* $p < 0.05$.
[a1] Denotes "poor" agreement per Bloom et al. (1999) standard.
[a2] Denotes "fair" agreement per Bloom et al. (1999) standard.
[b1] Denotes "fair" agreement per Landis and Koch (1977) standard.
[b2] Denotes "moderate" agreement per Landis and Koch (1977) standard.
[b3] Denotes "substantial" agreement per Landis and Koch (1977) standard.

each time only excluding diagnoses from one high-referral clinician. In other words, we repeated analyses while excluding cases involving High Referral Clinician 1, then repeated analyses while excluding cases involving High Referral Clinician 2 (with High Referral Clinician 1 again included), then repeated analyses while excluding cases involving High Referral Clinician 3 (with High Referral Clinician 2 again included), and proceeded in this fashion. An outlier effect would be demonstrated if reliability results significantly improved in the absence of a particular high-referral clinician. The results demonstrated no outlier effect. All analyses involving the exclusion of one high-referral clinician followed the same pattern as those displayed in Table 4.

9. Discussion

When determining which sex offenders will be committed as SVPs, states such as New Jersey rely on clinicians to evaluate sex offenders and offer insight into their eligibility and suitability for SVP commitment. Among the tasks for these clinicians is assessment of the offender to determine whether he suffers a mental abnormality or personality disorder. SVP evaluations are thus a specialized responsibility for clinicians that carry serious implications for multiple parties: the offenders being evaluated, the courts responsible for commitment decisions, and the general community to which released offenders will attempt to reintegrate. These evaluations contribute toward decisions regarding the indefinite commitment of sex offenders, decisions that impact both the public and the individuals evaluated. As such, it is in the interest of all parties that these evaluations be conducted in manners that are sound and reliable.

The current study adds to the literature on the reliability of diagnoses offered during SVP evaluations by comparing the diagnostic agreement between clinicians evaluating the same offender. Given the context of these evaluations, it is reasonable to expect that clinicians would have access to the same information covering the same course of time, increasing the likelihood of diagnostic agreement. Nonetheless, the current study suggests questionable diagnostic agreement between clinicians assessing the same sexual offender during SVP evaluations, similar to those conclusions in the original Levenson (2004b) study. Different statistical analyses were conducted to offer multiple approaches for examining diagnostic reliability. Overall, clinicians had poor to fair agreement on the presence of different paraphilic and non-paraphilic disorders.

Reliability results were moderately better for the diagnosis of Pedophilia. The kappa coefficient was highest for Pedophilia and approached "mild" agreement range, while PPVs indicated Pedophilia to be the only disorder among those we examined whereby an initial diagnosis was likely to be "confirmed" by a second clinician at levels greater than chance. NPVs suggested higher likelihoods of clinicians agreeing and confirming the *absence* of a given diagnosis. It should be noted, however, that absence of a diagnosis could represent different things in different cases. In some cases a clinician may have actively determined that a diagnosis was not appropriate for a given offender, whereas in other cases a clinician may have not considered that diagnosis at all (thus not actively "rejecting" the diagnosis). For example, a clinician who feels that an offender's substance abuse history is not as relevant in an SVP evaluation would be less likely to devote time to gathering information about the offender's substance abuse and whether he meets criteria for a Substance Disorder. In contrast, more confidence can be placed in the presence of a diagnosis representing a proactive decision by the clinician. Courts are ultimately more concerned on the presence of diagnoses (i.e., whether the offender meets the statutory criteria of having a mental abnormality) than the absence of diagnoses (i.e., which mental abnormalities are *not* applicable to the offender).

Although it may initially appear that two clinicians who recommend commitment for an offender that they opine to suffer a mental abnormality or personality disorder have reached a consensus, findings

from the current study suggest that only a superficial agreement is apparent, as clinicians do not arrive at the same conclusions regarding the specific mental impairment in question as often as would likely be desired. Moreover, this questionable reliability appears to reflect a widespread issue. The current study examined reliability within a subset of clinicians who received a high number of SVP evaluation referrals. This follow-up analysis was based on the idea that diagnostic disagreement among clinicians could be artificially skewed if a clinician involved in a high number of cases arrived at different diagnoses than other clinicians who generally agreed with one another. The current findings do not support the notion of an "outlier clinician" disagreeing with a general consensus. Instead, the current findings point to questionable diagnostic reliability being a broader issue across the range of clinicians. As such, efforts to improve diagnostic reliability would appear to be more beneficial on a broader scale.

9.1. Limitations

Reliance on archival data limits the conclusions drawn from this study. Many files had incomplete data, a common challenge when conducting archival research. Many cases had to be excluded from the current study because of incomplete information regarding diagnoses offered and the corresponding clinician offering those diagnoses. The small sample size available for analyzing reliability among high-referral clinicians in particular resulted in elevated standard error, making conclusions more tenuous. In addition, several of the diagnostic categories examined in the current study had low base rates. Although these rates generally reflect patterns in prior research (Jumper et al., 2011), low base rates can overstate agreement on the absence of a diagnosis (resulting in inflated NPVs) and understate agreement on the presence of a diagnosis (resulting in deflated PPVs).

During the current study, the researchers did not have access to the actual evaluations done by clinicians. As such, no examination was conducted regarding the processes that led to clinicians' diagnoses. This prevented any evaluation of the validity of diagnoses or identification of any reasons for disagreement. As stated previously, the absence of a given diagnosis does not always indicate that the clinician proactively rejected this diagnosis; however, because the researchers did not have access to the actual evaluations, no inferences could be made for the reasons behind a diagnosis or absence thereof. Finally, the true independence of each evaluation is unknown. It is possible that some evaluations were interdependent, either conducted simultaneously or through shared materials. The data for the current study did not include information offering insight into the extent by which the two clinicians conducted their evaluations independent of each other. It should be noted that concerns about the independence of evaluations regard artificial inflation of agreement between clinicians; thus, diagnostic reliability would likely be elevated (not lowered) if evaluation interdependence were an issue.

9.2. Implications

Despite the limitations noted, the current study provides insight into the extent of SVP diagnostic reliability concerns observed in prior studies. The original Levenson (2004b) study examined SVP evaluations in Florida. The current study includes analysis of SVP evaluations conducted in New Jersey, expanding analysis to another state and another region of the U.S. Results from the current study run in accordance with initial concerns about questionable diagnostic agreement during SVP evaluations and suggest this to be a widespread issue. Although clinicians appear to agree on the presence of Pedophilia with some consistency, clinicians conducting SVP evaluations appear to largely disagree on the presence of other paraphilic and non-paraphilic disorders. The current study affirms prior findings on questionable diagnostic reliability during SVP evaluations and further clarifies the extent of this issue. Specifically, the current study fails to support the notion that

disagreement may be attributed to isolated incidences of outlier clinicians offering diagnoses divergent of an otherwise consensus. Instead, the current study suggests that questionable diagnostic reliability in SVP evaluations is a widespread issue apparent across different studies (in different states), across various diagnostic categories, and across the scope of clinicians involved.

Given the wide scope of inconsistencies across clinicians' diagnostic conclusions, it may be necessary to reexamine the processes involved in SVP decision making, including the role of clinicians and the criteria on which they base their recommendations to the courts. As it currently stands, the degree of inconsistency in diagnosis of mental abnormality suggests that clinical judgment in SVP evaluations is not meeting proper reliability standards as required by the *Daubert* standard of expert testimony admissibility (Jackson, Rogers, & Shuman, 2004), although these error rates pose less significant of a threat to the admissibility of such testimony for states following the *Frye* standard (such as New Jersey). In either case, for clinicians to properly assist the courts in SVP evaluations, better efforts must be made to improve reliability.

The specific tasks given to clinicians during SVP evaluations may contribute to these inconsistencies. The mental abnormality criterion in SVP laws has been described as ambiguous with language deviating from the standard criteria clinicians use when determining sexual recidivism risk (Fabian, 2011). Although it is common for clinicians to rely on *DSM* diagnoses to address the mental abnormality criterion, there are few standardized and established approaches for reliably assessing the disorders commonly associated with SVP commitment, namely paraphilic disorders (Jackson, 2008; Miller, Amenta, & Conroy, 2005). Further, the empirically supported risk assessment instruments currently used to offer insight into recidivism risk do not emphasize mental abnormality, bringing to question whether the current strengths in sex offender assessment are consistent with the current standards for SVP decision making (Jackson et al., 2004). Should states wish to use the best supported and most reliable practices in sex offender assessment during SVP evaluations, a review of the criteria for evaluating suitability for SVP commitment may be necessary.

Significant diagnostic disagreement risks confusing the courts with inconsistent information about an offender and makes an already serious process more complicated. Courts tend to concur with the conclusions of clinicians hired to assist the court (e.g., Zapf, Hubbard, Galloway, Cox, & Ronan, 2004), but less is known about judicial understanding and decision making when forensic clinicians present conflicting information. It is certainly possible that courts would disregard diagnostic agreement as long as clinicians agree more broadly that the offender suffers a "mental abnormality." Nonetheless, conflicting information makes it increasingly likely that the courts will need to resolve such inconsistencies in the information being received regarding an offender evaluated for SVP commitment. Efforts must be made to better assure that clinicians evaluating the same offender offer reliable diagnoses that truly assist courts responsible for making SVP commitment decisions.

9.3. Future directions

Before improvements in SVP diagnosis can be made, further clarity of the reasons that clinicians disagree with such frequency is necessary. Future research on processes related to diagnostic decision making may offer insight into diagnostic reliability and, in turn, areas for global improvement. Content analysis of clinicians' SVP evaluations would offer comparisons of clinicians' sources of information, areas emphasized during the evaluation, and stated bases for diagnostic decisions. Identifying specific factors that decrease diagnostic reliability could clarify the issue as well. For example, to what extent can diagnostic reliability during SVP evaluations be attributed to offender factors (e.g., number of undocumented victims), clinician factors (e.g., reliance on collateral information), procedural factors (e.g., time allotted between referral and submission of report), and diagnostic factors (e.g., explicitness of diagnostic criteria)? The latter is sure to be an area of particular concern

with the eventual shift toward *DSM-5* diagnoses. Discussion about the validity and reliability of paraphilic and non-paraphilic diagnostic categories has feverishly ensued with the impending release of the *DSM-5*. Paraphilic diagnostic studies such as the current study should be revisited with the evolution of paraphilic diagnoses and their continued importance to SVP evaluations.

Acknowledgments

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (Washington, DC).

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

Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating Practice: Guidelines for the Accountable Professional* (3rd ed.)Boston: Allyn and Bacon.

Calkins, C., Jeglic, E. L., & Perillo, A. D. (2013). *Sexually Violent Predator (SVP) Commitment: An Examination of the Characteristics and Recidivism Rates of Committed and Nearly Committed Offenders.* (submitted for publication).

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

Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology*, 4, 149–184. http://dx.doi.org/10.1348/135532599167824.

Dunsieth, N. W., Nelson, E. B., Brusman-Lovins, L. A., Holcomb, J. L., Beckman, D., Welge, J. A., et al. (2004). Psychiatric and legal features of 113 men convicted of sexual offenses. *Journal of Clinical Psychiatry*, 65, 293–300. http://dx.doi.org/10.4088/JCP.v65n030.

Epperson, D. L., Kaul, J.D., Huot, S., Goldman, R., & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool—Revised (MnSOST—R) Technical Paper: Development, Validation, and Recommended Risk Level Cut Scores.* (Retrieved from http://www.psychology.iastate.edu/faculty/epperson/TechUpdatePaper12-03.pdf)

Fabian, J. M. (2011). Paraphilias and predators: The ethical application of psychiatric diagnoses in partisan sexually violent predator civil commitment proceedings. *Journal of Forensic Psychology Practice*, 11, 82–98. http://dx.doi.org/10.1080/15228932.2011.521723.

First, M. B., & Halon, R. L. (2008). Use of DSM paraphilia diagnoses in sexually violent predator commitment cases. *The Journal of the American Academy of Psychiatry and the Law*, 36, 443–454 (Retrieved from http://www.jaapl.org)

Hanson, R. K., & Thornton, D. (1999). *STATIC-99: Improving Actuarial Risk Assessments for Sex Offenders (User Report No. 1999–02).* Ottawa: Department of the Solicitor General of Canada.

Jackson, R. L. (2008). Sex offender civil commitment: Recommendations for empirically guided evaluations. *Journal of Psychiatry & Law, 36*, 389–429.

Jackson, R. L., Rogers, R., & Shuman, D. W. (2004). The adequacy and accuracy of sexually violent predator evaluations: Contextualized risk assessment in clinical practice. *International Journal of Forensic Mental Health*, 3, 115–129. http://dx.doi.org/10.1080/14999013.2004.10471201.

James, N., Thomas, K. R., & Foley, C. (2007). *Civil commitment of sexually dangerous persons.* Congressional Research Service Report for Congress.

Janus, E. S. (2000). Sexual predator commitment laws: Lessons for law and the behavioral sciences. *Behavioral Sciences and the Law, 18*, 5–21. http://dx.doi.org/10.1002/(SICI)1099-0798(200001/02)18:1<5::AID-BSL374>3.0.CO;2-C.

Jumper, S., Babula, M., & Casbon, T. (2011). Diagnostic profiles of civilly committed sexual offenders in Illinois and other reporting jurisdictions: What we know so far. *International Journal of Offender Therapy and Comparative Criminology*, 56, 838–855. http://dx.doi.org/10.1177/0306624X11415509.

La Fond, J. Q. (2005). *Preventing Sexual Violence: How Society Should Cope With Sex Offenders.* Washington, DC: American Psychological Association.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Levenson, J. S. (2004a). Reliability of sexually violent predator civil commitment criteria in Florida. *Law & Human Behavior, 28*, 357–368 (doi: 0147-7307/04/0800-0357/1).

Levenson, J. S. (2004b). Sexual predator civil commitment: A comparison of selected and released offenders. *International Journal of Offender Therapy and Comparative Criminology, 48*, 638–648. http://dx.doi.org/10.1177/0306624X04265089.

Levenson, J. S., & Morin, J. W. (2006). Factors predicting selection of sexually violent predators for civil commitment. *International Journal of Offender Therapy and Comparative Criminology, 50*, 609–629. http://dx.doi.org/10.1177/0306624X06287644.

Lieb, R., & Matson, S. (1998). *Sexual Predator Commitment Laws in the United States [No. 38].* Olympia, WA: Washington State Institute for Public Policy.

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. http://dx.doi.org/10.1177/0306624X02238161.

Mercado, C. C., Jeglic, E. L., & Markus, K. A. (2011). *Sex Offender Management, Treatment, and Civil Commitment: An Evidence Based Analysis Aimed at Reducing Sexual Violence.* : National Institute of Justice.

Meyer, G. L. (2002). Implications of information-gathering methods for a refined taxonomy of psychopathology. In L. E. Beutler, & M. L. Malik (Eds.), *Rethinking the DSM: A Psychological Perspective* (pp. 69–106). Washington, DC: American Psychological Association.

Miller, H. A., Amenta, A. E., & Conroy, M.A. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior*, *29*, 29–54. http://dx.doi.org/10.1007/s10979-005-1398-y.

Murrie, D. C., Boccaccini, M. T., Caperton, J., & Rufino, K. (2012). Field validity of the Psychopathy Checklist—Revised in sex offender risk assessment. *Psychological Assessment*, *24*, 524–529. http://dx.doi.org/10.1037/a0026015.

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, *15*, 19–53. http://dx.doi.org/10.1037/a0014897.

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

Prentky, R. A., Coward, A. I., & Gabriel, A.M. (2008). Commentary: Muddy diagnostic waters in the SVP courtroom. *Journal of the American Academy of Psychiatry and Law*, *36*, 455–458 (Retrieved from http://www.jaapl.org)

Rappleye, H. (April 17). *Releasing sex offenders.* The Crime Report (Retrieved from http://www.thecrimereport.org)

Schiavone, S. K., & Jeglic, E. (2009). Public perception of sex offender social policies and the impact on sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, *53*, 679–695. http://dx.doi.org/10.1177/0306624X08323454.

Spengler, P.M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2009). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist*, *37*, 350–399. http://dx.doi.org/10.1177/0011000006295149.

Wollert, R. (2007). Poor diagnostic reliability, the Null-Bayes logic model, and their implications for sexually violent predator evaluations. *Psychology, Public Policy, and Law*, *13*, 167–203. http://dx.doi.org/10.1037/1076-8971.13.3.167.

Zapf, P. A., Hubbard, K. L., Galloway, V. A., Cox, M., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determinations of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice*, *4*, 27–44. http://dx.doi.org/10.1300/J158v04n01_02.