

# Reliability of Sex-offense Risk Assessments at Civil Commitment: Threshold Levels in New Jersey SVP Proceedings

Author: James Krivacska, Psy.D.  
Avenel, New Jersey

with

Author: Howard Margolis, Ed.D. [Corresponding author]  
Professor Emeritus of Education  
Queens College of CUNY  
Flushing, NY 11367  
[hm08043@yahoo.com](mailto:hm08043@yahoo.com)

## Abstract

Forensic experts rely on various methods to assess sex-offense risk in sexually violent predator (SVP) proceedings in New Jersey. Forensic experts assume that such risk assessments can be conducted reliably and validly. The validity of this assumption was assessed by analyzing professional practice and legal standards for the admissibility of and reliance upon expert testimony in New Jersey SVP proceedings. A preliminary analysis, based on a limited set of Appellate Court decisions, suggested that current risk-assessment practice in New Jersey appears to rely upon a combination of normative risk estimates grounded in actuarial assessment and unstructured clinical judgment that fail to meet accepted professional/legal standards. Current practices most likely result in significant overestimates of recidivism risk and overclassification of SVPs. Recommendations for improving the reliability and validity of SVP determinations in New Jersey are provided.

**Keywords:** sexually violent predators, risk assessment, actuarial prediction, civil commitment

---

Despite more than a decade of conceptual debate and empirical research, questions remain about the ability of forensic psychologists and psychiatrists (Forensic Mental Health Professionals: FMHPs) to assist courts in reliably identifying a discrete subpopulation of those with sex-offending histories who pose the highest risk of reoffense (Janus & Meehl, 1997; Prentky, Janus, Barbaree, Schwartz, & Kafka, 2006; Schopp, Scalora, & Pearce, 1999). With twenty jurisdictions adopting, in one form or another, SVP statutes,<sup>1</sup> this remains more than an academic concern. At the crux of the debate remains the unsettled nature of the legal and psychological constructs at the heart of SVP stat-

---

<sup>1</sup> As of January, 2010, the twenty jurisdictions with SVP civil commitment laws were Arizona, California, Florida, Illinois, Iowa, Kansas, Massachusetts, Minnesota, Missouri, Nebraska, New Jersey, New York, North Dakota, Pennsylvania, South Carolina, Texas, Virginia, Washington, and Wisconsin as well as the federal government.

utes and conflict over what constitutes reliable and valid measurement of these constructs. Much of this conceptual debate occurs in the rarefied atmosphere of academic circles and professional discourse, with infrequent attempts at tying those analyses to real-world practices in SVP proceedings (such as Boccaccini, Murrie, Caperton, & Hawes, 2009; Jackson & Hess, 2007; Levenson, 2004). We attempt here to bridge those two worlds by defining the parameters of sex-offense risk assessment (SORA), a key component of SVP determinations, and then by applying the conceptual framework we describe to forensic psychological assessments as encountered in one state's (New Jersey's) implementation of its SVP statute.

### SVP Statute Constructs and Definitions

The New Jersey Sexually Violent Predators Act (SVPA, 1998: *N.J.S.A. §30:4-27.24 et seq.*) typifies the type of legislation that the United States Supreme Court held passed constitutional muster in *Kansas v. Hendricks* (1997). With over 400 hearings in New Jersey since the inception of the law a decade ago (Gallagher, 2009), a body of evidence exists as to how the SVPA is implemented and how FMHPs conduct their SORAs.

Classification as an SVP under the SVPA requires satisfaction of a three-prong test common to nearly all SVP legislation (see Miller, Amenta, & Conroy, 2005): (a) a judicial finding of a prior sex offense; (b) presence of “a mental condition that affects a person’s emotional, cognitive or volitional capacity in a manner that predisposes that person to commit acts of sexual violence” (*mental abnormality standard*); and (c) a likelihood “to engage in acts of sexual violence” (*likelihood standard*),<sup>2</sup> which is defined as “the propensity of a person to commit acts of sexual violence is of such a degree as to pose a threat to the health and safety of others.” (*N.J.S.A. §30:4-27.26*).

Although in most SVP cases, FMHPs play no role regarding the first prong, occasionally they may be called upon to assist in a court's determination as to whether the conduct underlying a criminal offense that is not specifically enumerated as a qualifying sex offense, was sexually motivated (See *N.J.S.A. 30:4-27.26*: “Sexually violent offense’ means: . . . (b) any offense for which the court makes a specific finding on the record that, based on the circumstances of the case, the person's offense should be considered a sexually violent offense . . . ”; see also *IMO the Civil Commitment of J.M.B.*, 2009).

With regard to the likelihood standard, typically assessed via SORA, the target behavior of concern is discrete: the future commission of a sexual offense in violation of the local

---

<sup>2</sup> In *IMO the Civil Commitment of W.Z.* (2002), the New Jersey Supreme Court, drawing on the U.S. Supreme Court's holding in *Kansas v. Crane* (2002) limiting SVP commitments to those with a “serious difficulty” controlling their sex-offending behavior, rejected a likelihood standard of “more likely than not” to recidivate. Rather, the New Jersey Supreme Court imputed into the statutory phrase “likely to engage in acts of sexual violence” both (a) a degree of risk which exceeds the more-likely-than-not threshold to an unquantified degree (described by the Court as “highly likely”), and (b) the constraint that the risk must be realizable within the “reasonable foreseeable future.” (*IMO the Civil Commitment of W.Z.*, 2002, p. 132).

jurisdiction's criminal code. FMHPs in New Jersey have access to an extensive body of evidence as to sex-offense recidivism rates on both a national (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005, 2007, 2009; Langan, Schmitt, & Durose, 2003) and statewide scale (Sager, 2000, 2001, 2002; Zgoba, Sager, & Witt, 2003; Zgoba & Simon, 2005; Zgoba, Witt, Dalessandro, & Veysey, 2008).

Moreover, researchers have identified many factors correlated with sex-offense recidivism (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004, 2005) and have combined factors into systematic protocols for assessing risk (e.g., structured professional judgment, Douglas & Skeem, 2005; Sexual Violence Risk-20: SVR-20, Boer, Hart, Knopp, & Webster, 1998). Some investigators have combined factors into standardized assessment instruments producing estimates of recidivism risk such as the Static-99 and its progeny (Hanson & Thornton, 1999; Helmus, Hanson & Thornton, 2009; Phenix, Helmus, & Hanson, 2009), the Minnesota Sex Offending Screening Tool-Revised (MnSOST-R; Epperson, Kaul, Hesselton, 1999), the RRASOR (Hanson, 1997), and the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998). Finally, FMHPs have access to statistical models and methods to increase the reliability and predictive validity of their risk estimates (e.g., Donaldson & Wollert, 2008; Mossman, 2006; Wollert, 2006; Woodworth & Kadane, 2004). Together, all those components establish an empirical framework within which FMHPs may operate when conducting SORA under the likelihood standard.

In contrast to the likelihood standard, the mental-abnormality standard represents a conceptual minefield, lacking definitional clarity or a consensus as to how to assess its presence (LaFond, 1999; Prentky et al., 2006; Wollert, 2007).

Thus, independent of mental-abnormality determinations, because improving SORA is likely to significantly reduce the number of erroneous SVP classifications, this article will focus on the type of expert evidence and analytic framework needed to inform likelihood standard determinations using the New Jersey SVPA as a legal framework. But first, we review constraints imposed on forensic practice, and the nature of classification.

### **Forensic Practice and Professional Standards in SVP Proceedings**

Several ethical codes of conduct and specialty guidelines inform the forensic practice of FMHPs. Psychological assessment is guided by the *Ethical Principles and Code of Conduct of Psychologists* (American Psychological Association, 2002, hereafter *APA Ethical Principles*) and the *Specialty Guidelines for Forensic Psychologists* (Committee on Ethical Guidelines for Forensic Psychologists, 1991, hereafter *Forensic Specialty Guidelines*).<sup>3</sup> Psychologists must also adhere to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychologi-

---

<sup>3</sup> The American Psychological Association is currently reviewing revisions to the *Forensic Specialty Guidelines* (Committee on the Revision of the Specialty Guidelines for Forensic Psychology, 2009). In anticipation of their approval, reference to applicable standards in the proposed revision will follow the 1991 standards in [brackets].

cal Association, & National Council on Measurement in Education, 1999, hereafter *Testing Standards*).

Forensic psychiatrists look to the ethical guidelines and practice standards for forensic psychiatry adopted by the American Academy of Psychiatry and Law (2005) as well as the 2006 edition of *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry* (American Psychiatric Association, 2006, hereafter *APA Psychiatric Ethics*).<sup>4</sup> Under these standards, psychological and psychiatric evaluations in SVP proceedings constitute forensic practice governed by the codes listed in this section.<sup>5</sup>

### The Nature of Prediction

FMHPs' ability to predict individual behavior is extraordinarily limited. Some argue that SORA, regardless of the method employed, entails the classification of the examinee—on the basis of some predetermined discriminative criteria—to a group about which some recidivism information is known (Berlin, Galbreath, Geary, & McGlone, 2003; Helmus et al., 2009; Mossman, 2008; Seto, 2005). How one describes the group depends on the type of classification desired, which can be differentiated into two broad categories: relative and absolute classifications of risk (Helmus et al., 2009; *Testing Standards*, p. 29).

Absolute classifications of risk, or Absolute Probability Estimates of Risk (APERs), estimate the percentage of individuals in a group into which an FMHP has classified the examinee who will manifest the target behavior (“John Doe belongs to a group in which 45% of the members are predicted to reoffend within five years of release”).

Relative classifications of risk, or Relative Probability Estimates of Risk (RPERs), represent rank order classifications of groups of individuals, from those posing the least risk to those at greatest risk. FMHPs typically present RPERs in three forms: degree of risk or risk ratios (“John Doe belongs to a group whose members are three times more likely to commit a future sex offense than a typical offending individual”); percentile rank (“John Doe obtained a score placing him in a group which scored higher than 60% of the study sample”); or normative risk descriptions (“John Doe is at high risk to reoffend”; Helmus et al., 2009).

---

<sup>4</sup> State statutes and regulations may provide additional guidance for professional practice, while the professional literature and other published practice guidelines by professional organizations may inform the application of the above principles and standards in specific contexts.

<sup>5</sup> While ethical codes are binding, the *Forensic Specialty Guidelines* and *Testing Standards* provide greater flexibility for psychologists' divergent practices. Nevertheless, substantial deviation from these standards should be documented and justified (*Testing Standards*, p. 3).

## Presumptions in Classification

The above-cited literature on the use of APERs and RPERs typically describes SORA as entailing the classification of an examinee into either a “likely-to-reoffend” or “not-likely-to-reoffend” group (using APERs, usually in the context of an SVP proceeding), lower versus higher risk groups (using RPERs) or to a series of discrete groups described by their associated risk probabilities (using either APERs or RPERs; see e.g., Hanson & Morton-Bourgon, 2009; Helmus et al., 2009; Mossman, 2006; Seto, 2005; Wollert, 2006). Prior to conducting an assessment and classifying the examinee into one of these groups, the FMHP holds a neutral position as to risk group assignment. SORA structured in this manner seeks to maximize the accurate placement of the individual into groups (Mossman, 2008).

In SVP proceedings, however, the task is not to start from a neutral position and then classify the examinee into likely-to-reoffend versus not-likely-to-reoffend groups or classification bins. Rather, from a legal perspective, the examinee is presumed to belong to the not-likely-to-reoffend bin and to not suffer from a mental abnormality. The burden of proof rests with the state, which must overcome these presumptions in order to classify the examinee as an SVP (*Kansas v. Hendricks*, 1997; Wollert, 2007). Nor does the U.S. Constitution permit an equal apportionment of risk of erroneous determinations between the state and the examinee; the state must bear the greater portion of risk and thus must establish each element of the SVP statute by, at minimum, clear and convincing evidence (*Kansas v. Hendricks*).

Thus, in contrast to researchers examining the validity of classification schemes, SVP determinations *begin* with the assignment of the examinee to the not-likely-to-reoffend bin. Practically speaking, this means that the risk of reoffense for the examinee is presumed to not differ from the base rate of the reoffending population from which the examinee was drawn.<sup>6</sup> Fact finders must then assess if proffered evidence of risk is sufficiently elevated above the local base rate and above the threshold set by the likelihood standard, to reclassify the examinee to the likely-to-reoffend bin. Because APERs provide information useful for determining whether to reclassify examinees to the likely-to-reoffend bin, APERs are relevant to the likelihood standard.

RPERs, however, provide no useful information for overcoming the presumptive classification of not-likely-to-reoffend. Indeed, presenting RPER data may easily confuse fact finders into believing the likelihood standard has been met. Examinees may present with high RPERs, but if the overall recidivism base rate in the population from which they are drawn is low, their absolute recidivism risk may also be very low (e.g., Boccacini et al., 2009, reporting a group of Texas sex offenders who, despite scoring above the 91st percentile [RPER] on the Static-99, had a five-year sex-offense recidivism rate

---

<sup>6</sup> Of course, determination of the local base rate may also be a matter of disputed expert opinion, which the fact finder must resolve, particularly when local base-rate information isn't available from published, peer-reviewed studies.

[APER] of only 6%).<sup>7</sup> Consequently, RPERs are not relevant to SVP determinations (Hanson & Morton-Bourgon, 2009; Mossman, 2006; Seto, 2005).

FMHPs assist fact finders to properly analyze SORA results by reporting those results in a manner that accurately portrays group-related risk information. For example, FMHPs obtaining a score on an actuarial scale that is associated with X% risk estimate over Y years may express the result in the form of Statement A.

*Statement A:*

*Based on the selection criteria, the examinee may be classified as belonging to a group of individuals X% of whom are expected to reoffend within Y years of release.*

Statement A, however, still falls short of satisfying professional standards because it lacks critical information regarding measurement error and potential threats to validity (*APA Ethical Principles*, Standard 9.02(b), 9.06; *Forensic Specialty Guidelines*, Standard IV.A.4 [12.01]; *Testing Standards*, Standards 2.1, 2.2, 2.14, 2.15; *APA Psychiatric Ethics*, Section 2).

Donaldson and Wollert (2008), Mossman (2006), Hanson (2005) and Phenix et al. (2009) have argued that proffering a risk estimate in the form of a single percentage, risk ratio, or percentile, as above, fails to account for measurement error. Indeed all assessment entails some measurement error, which must be known or estimated in order for the consumer of the assessment results to interpret those results and determine how much weight to give them (*Testing Standards*, pp. 25, 27). In psychometrics, measurement error is often reported as the confidence interval (CI; *Testing Standards*, Standard 2.2). The CI assumes the true score or value associated with the subject is fixed and thus describes a range of scores or values which an FMHP can assert has an X% probability of including the true score or value (Aron & Aron, 1999, p. 201). The size of a CI for any instrument is likely to vary with the raw score obtained as a factor of the number of subjects obtaining that score. Generally speaking, as one obtains scores that deviate further from the mean, the amount of error associated with the score increases (as the *n* for those with the extreme scores declines; Aron & Aron, 1999, p. 242).

This has serious implications for SVP determinations, which, by definition, are seeking a small subgroup of individuals whose risk deviates significantly from the mean of the local sex-offending population. For these reasons, especially for scores yielding risk probabilities near the thresholds for making SVP classification decisions, reporting the CI is critical (*Testing Standards*, Standard 2.14).

---

<sup>7</sup> Throughout this article, recidivism rates cited obviously only reflect detected sex-offense recidivism. Actual recidivism rates are presumably higher due to unreported/unprosecuted offenses. However, as FMHPs have no reliable and valid method available to them by which to calculate an "actual" recidivism rate from the detected rate, FMHPs should base their risk estimates on the best available empirical data while acknowledging that any risk estimate proffered may be an indeterminate underestimation of risk.

Thus, we can modify Statement A to reflect the CI associated with the risk estimate for the group into which the examinee has been classified.

*Statement B:*

*Based on the selection criteria, there is a 95% probability that the subject belongs to a group of individuals between  $X_1\%$  and  $X_2\%$  of whom are expected to reoffend within  $Y$  years of release.*

In other words, the FMHP uses the risk estimate to classify the individual into a bin described by the upper and lower ranges of risk defined by the CI, while the probability associated with that CI (in the above example, 95%) expresses the degree of certainty with which the FMHP has made that classification.

In jurisdictions that distinguish members of the likely-to-reoffend group from the not-likely-to-reoffend group using some manner of a quantified likelihood standard (such as more likely than not or greater than 50% probability of reoffense), the CI associated with an APER will sometimes encompass that likelihood standard threshold or cut score. The fact finder must then determine whether the fact that the FMHP has classified the examinee into a group with a range of risk that includes the likelihood standard threshold, suffices to place the examinee into the likely-to-reoffend bin. Two points of view have emerged in the field on this point. Wollert (2007) has argued that an examinee can only be classified as a likely recidivist if the entire CI for the APER lies above the cut score. Mossman (2008), in contrast, has argued that despite a CI that includes the cut score, classifying examinees based on whether their APER falls above or below the cut score remains, statistically, the most correct decision.

In a research or academic context, where classifications are unbounded by prior presumptions and where the risk of error is equally apportioned among sorting bins, Mossman would be correct. However, as noted above, in SVP proceedings, fact finders are constrained both by a prior presumption that the examinee belongs in the not-likely-to-reoffend bin, and an apportionment of risk which favors the examinee's freedom over public safety (the clear and convincing standard of proof; *Addington v. Texas*, 1979).

Within a group described by a range of risk ( $X_1\%$  to  $X_2\%$ ) will be members whose risk of reoffense will be  $X_1\%$  and others whose risk will be  $X_2\%$ . Because the FMHP cannot say whether the examinee's risk corresponds to the lower ( $X_1\%$ ) or upper ( $X_2\%$ ) limit, or somewhere in between, arguably, the FMHP cannot state with reasonable certainty (see discussion of *Legal Standards for Expert Testimony* below) that the examinee's true risk lies above the threshold.<sup>8</sup>

---

<sup>8</sup> The CI informs both a legal determination as to whether an expert is reasonably certain as to a risk estimate (see discussion below as to admissibility of expert opinion) and the standard of proof employed by the fact finder. Wollert (2007) equated a 90-95% level of confidence in a forensic diagnosis or conclusion with the legal standard of proof beyond a reasonable doubt. However, neither the United

Having discussed how FMHPs should present risk-estimate information, the question remains: What methods may FMHPs employ to obtain those risk estimates?

## Assessment Methods

Three assessment methods predominate in SORA: actuarial assessment, clinical judgment, and a hybrid mix of both actuarial and clinical assessment, which some have called the clinically adjusted actuarial method (Hanson & Morton-Bourgon, 2009). First, we'll explore the limitations of relying *solely* on either actuarial scales or clinical judgment when estimating reoffense risk. Then we'll examine the viability of combining these two methods in a multi-factor assessment model as typified by the clinically adjusted actuarial method.

The first method relies on actuarially developed scales (see Monahan, 2006 for a description of how these scales are developed). Rates of reoffense associated with groups of individuals obtaining each of the summed raw scores on the scale are reported in experience tables (thus constituting APERs). Two actuarial scales that provide APERs relevant to SVP determinations predominate in sex-offense risk assessment in New Jersey: the MnSOST-R (Epperson, et al., 1999), and the Static-99 and its progeny, the Static-99R and Static-2002R<sup>9</sup> (Phenix et al., 2009; see *IMO the Civil Commitment of R.S.*, 2001). These two scales are also widely used in SVP proceedings across the country (Jackson & Hess, 2007).

In addition to APERs, Phenix et al. (2009) report RPERs (risk ratios and percentile ranks) for the raw scores obtained on the Static-99R and Static-2002R. Both of these scales, as well as the MnSOST-R, assign normative labels to categories of raw scores, such as “high,” “moderate,” or “low” risk. As mentioned before, neither the RPERs nor the normative labels associated with raw scores on these scales are legally relevant in SVP proceedings because they provide no information about whether an examinee’s risk reaches the likelihood standard.

Regarding the second method, reliance on clinical judgment or clinical experience constitutes an assessment methodology—albeit a “less rigorous or quantitative version of categorization and prediction” than actuarial assessment (Harcourt, 2004, p. 35)—sub-

---

States Supreme Court nor the New Jersey SVPA requires such a strict standard of proof in SVP proceedings. So, arguably, requiring the whole 95% CI to lie above the likelihood standard may be more stringent than required by law. Unfortunately, if available at all, researchers usually only report the 95% CI for actuarially assessed risk (e.g., Hanson, 2005; Phenix et al., 2009). Unless FMHPs can calculate CIs for lower confidence levels (e.g., 65% or 75%), the alternative—ignoring measurement error altogether—opens the risk estimate to attack that recidivism risk above the threshold level hasn’t been established by clear and convincing evidence. Ultimately, best practice may be simply to inform the fact finder as to the role of the CI in estimating error around the SVP risk threshold, and leaving it to the fact finder to determine if the proofs offered satisfy the jurisdiction’s standard of proof.

<sup>9</sup> The Static-99R and Static-2002R account for age-related reductions of risk better than the original Static-99, and were also normed on much larger, more contemporary samples of offending individuals (Phenix et al., 2009). The Static-2002R also includes a broader range of items covering general criminality than the Static-99R, but has a smaller research base (Phenix et al., 2009).

ject to the same standards of practice applicable to other assessment methods.<sup>10</sup> FMHPs rarely define clinical judgment and clinical experience, yet such definitions are critical to evaluating the method's reliability and the validity of interpretations generated from its data.

Clinical experience refers to the data set of observations of client behavior that an individual clinician makes and retains in memory during one or more phases of the clinical cycle (e.g., referral, assessment, evaluation, discharge) across many clients. The process of transitioning from "novice" to "expert" status in any profession entails exposure to specific cases (perhaps thousands), and a testing of the boundaries of classification decisions (Johnston, 2003).

Clinical judgment represents the application of a set of heuristics to a particular set of clinical experiences (Johnston, 2003) by which decisions are made—including diagnosis, treatment plans, predictions of behavior, and other decisions regarding the client. Heuristics may derive from research-based findings or idiosyncratic clinical experiences. In exercising clinical judgment, clinicians attach meaning to their clinical experiences, even while remaining frequently unaware of the biases that mediate the attached meanings. In the literature regarding sex offending, there is wide consensus that FMHPs cannot justify, ethically or scientifically, reliance upon unstructured clinical judgment to conduct SORA (Campbell, 2000; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Prentky et al., 2006; Wollert, 2006; Zonanna, 2000).

Attempts have been made to structure clinical judgment by systematizing the analysis of factors. One approach which Hanson and Morton-Bourgon (2009) refer to as "structured professional judgment," entails limiting the factors considered to those which empirical research has documented correlate significantly with recidivism or nonrecidivism, such as the systematic evaluation of recidivism-correlated factors proposed by Douglas & Skeem (2005). FMHPs using either structured professional judgment or unstructured clinical judgment are limited to generating RPERs: neither method provides APERs relevant for SVP proceedings (Hanson & Morton-Bourgon, 2009).

Finally, Hanson and Morton-Bourgon (2009), and others (Doren, 2002; Fabian, 2005) have advocated reliance on a hybrid form of SORA, the clinically adjusted actuarial method, in which information obtained from actuarial assessment is combined with assessment of other factors not scored on the scales, but which are known to correlate with recidivism. For example, research documents age-related reductions in sex-offense risk over the entire lifespan, yet, until recently, actuarials failed to fully account for this effect (Hanson, 2005; Wollert, 2006). For a factor to be considered, there must

---

<sup>10</sup> While application of clinical judgment may appear to entail individualized estimates of risk, we argue that, like actuarial estimation, the clinician employing clinical judgment is, at best, stating: "This individual is similar in many respects to others with whom I have had professional experience, a high percentage of whom have reoffended. Therefore, I conclude based on those similarities that he belongs to that group of high-risk offenders."

be evidence it correlates with recidivism and can be reliably and validly measured (Hanson, 1998; Hanson & Morton-Bourgon, 2009).

Hanson (1998) has pointed out one means of weighing such factors. He notes that the correlation coefficient approximates the difference in the base rate between those who have a particular attribute/factor and those who do not. For example, with a base rate of 20% and  $r = .10$  for the factor "never married," 15% of those who were married and 25% of those never married will likely reoffend. Unfortunately, this method breaks down when one seeks to combine the impact of multiple factors because of unknown levels of covariance between factors. Moreover, many so-called "risk factors" occur more frequently among nonrecidivist versus recidivist samples (Campbell & DeClue, 2010a).

One method for combining factors to predict recidivism is to use regression analysis (*Testing Standards*, Standard 1.17) as Hanson & Harris (2000) did to obtain a nine-factor constellation of risk factors that correlated with recidivism ( $r = .60$ ). While one could conceivably combine an APER of 14% (Static-99R routine sample, score = 6, CI [10.2, 19.8]) with this nine-factor model to obtain an adjusted APER = 44% [ $.14 + (.60/2)$ ], the precision suggested by such calculations is illusory since considerable covariance is likely.

The example does, however, illustrate an important point: The inability to obtain risk estimates that exceed chance levels of reoffending, given nine factors empirically correlated with recidivism, should give pause to FMHPs confident in their ability to classify individuals into likely-to-reoffend versus not-likely-to-reoffend bins based on a few idiosyncratically identified factors. Any other approach to combining APERs generated by actuarial assessments with structured professional judgment, inevitably converts the APER to an RPER.

In any event, FMHPs must exercise caution when using mathematical models such as these to adjust APERs from actuarial assessments as they may convey a degree of precision that is unwarranted and because any mode of calculation and combination of results obtained across methods depends on unknown assumptions about the development sample, the scale, reliability and validity, variances, and individual subject differences.

### **Legal Standards Applicable to Proceedings under the SVPA**

Under the New Jersey SVPA, FMHPs' testimony is subject to analysis under three legal standards. First, the court must find the proffered expert testimony admissible under state law. Second, FMHPs must assert their conclusions to a reasonable certainty. Only after these two standards are met is the testimony presented to the fact finder<sup>11</sup> for consideration under the likelihood standard and mental-abnormality standard.

### **Legal Standards for Expert Testimony: Admissibility and Reasonable Certainty**

---

<sup>11</sup> In New Jersey, the Superior Court judge presiding over the civil commitment hearing also acts as the fact finder and makes the ultimate determination as to whether or not the individual is an SVP.

For over a half century, most expert-witness testimony in American courts was admissible under the *Frye* standard (*Frye v. U.S.*, 1923) if the expert's methods and reasoning were generally accepted (*general acceptance standard*) in the relevant professional community (Faigman, 1999, p. 62).<sup>12</sup> Under *Frye* as applied in New Jersey, the trial court must evaluate the evidentiary reliability<sup>13</sup> of the expert's testimony, the soundness of the expert's methodology, and its relevance to the issues before the court (*State of New Jersey v. Harvey*, 1997).

In most jurisdictions, using either the *Daubert* or *Frye* standard,<sup>14</sup> experts must prove the evidentiary reliability (and thus admissibility: *admissibility standard*) of their methodology and testimony by a preponderance of evidence<sup>15</sup> (e.g., *In re Paoli R. R. Yard PCB Litigation*, 1994; *Moore v. Ashland Chemical, Inc.*, 1998; *Rutigliano v. Valley*, 1996).

Most jurisdictions also require experts to assert a "reasonable degree of psychological or medical certainty" (Wecht, 2008, 1§2.02) in their conclusions (*reasonable certainty standard*) which has been described as, at minimum, constituting greater than chance (50%) certainty (*Dallas v. Burlington N., Inc.*, 1984; *Mauro v. Raymark Industries*, 1989; *State of New Jersey v. Fortin*, 2004; Wecht, 2008, at 1§2.02).

Thus, in New Jersey, FMHPs must establish both the reliability of the method (admissibility standard) based on the *Frye* general acceptance standard as well as the certainty of the opinion offered (reasonable certainty standard) by a preponderance of evidence.

### The Threshold Question: The Likelihood Standard

---

<sup>12</sup> An alternative standard was adopted by the United States Supreme Court in *Daubert v. Merrell Dow Pharmaceutical* (1993) which replaced the *Frye* standard in all federal jurisdictions. Under the *Daubert* standard, also adopted by a number of states, the trial court must evaluate the evidentiary reliability of the expert's testimony and its relevance to the issues before the court (p. 593). An expert can prove reliability by demonstrating the scientific validity of the reasoning or methodology underlying the testimony (p. 592-93). As a non-exhaustive list of factors to evaluate evidentiary reliability, the *Daubert* Court advised judges to seek evidence of a theory or method's: (a) falsifiability; (b) submission to peer review; (c) error rates; and (d) general acceptance (p. 593-594).

<sup>13</sup> Evidentiary reliability, in a legal context, is not synonymous with scientific reliability, the latter referring to the consistency of a method's measurement (American Educational Research Association et al., 1999, p. 25). As Heilbrun, DeMatteo, and Marczyk (2004) note, in the law, reliability more closely aligns with the concept of scientific validity (implicit in which is its reliability) or the evidence supporting the interpretations generated by a method's data or results (American Educational Research Association et al., 1999, p. 9).

<sup>14</sup> Because application of the *Frye* standard, unlike *Daubert*, depends on state law precedent and local rules of evidence, how general acceptance under *Frye* is proved will vary from state to state. Nevertheless, "the lynchpin to expert testimony [under *Frye*] is that it be reliable." (*Rubanick v. Witco Chem. Corp.*, 1990, p. 66).

<sup>15</sup> *Preponderance of evidence* refers to the standard of proof required to make a factual finding and is defined as *more likely than not* or greater than fifty percent probability (*Bourjaily v. U.S.*, 1987, p. 175; see also *IMO the Civil Commitment of W.Z.*, 2002; Garner, 2004). Admission of expert testimony in New Jersey is dependent upon a showing, by a preponderance of evidence, that the methods employed by the expert are generally accepted under the *Frye* standard (*State of New Jersey v. Harvey*, 1997).

States differ as to how and whether they quantify the likelihood standard. For example, the New Jersey Supreme Court in *IMO the Civil Commitment of W.Z.* (2002) established a likelihood standard of “highly likely” to reoffend, a threshold the court determined exceeded the more-likely-than-not threshold to an unspecified degree (see footnote 2 above).<sup>16</sup> The risk threshold question has serious implications for SORA analysis and its legal application. As discussed below, the admissibility of SORA methods will depend on the local recidivism base rate accepted by the experts and fact finder as well as on the threshold of risk (expressed as a probability of reoffense over a fixed period of time) associated with the likelihood standard.

In jurisdictions in which neither the legislature nor the courts have quantified the likelihood standard, attorneys and experts should explain to trial courts the importance of agreeing on a threshold level of risk both to ensure consistency across examinees and fact finders, and to ensure that the methodologies employed are capable of yielding risk estimates in the vicinity of the likelihood standard’s risk threshold.

Because the New Jersey Supreme Court made no further attempt in *W.Z.* to quantify the likelihood standard of “highly likely” beyond specifying it exceeds the more-likely-than-not standard, for purposes of this analysis we will accept a minimal likelihood standard of more-likely-than-not. To the degree that New Jersey’s likelihood standard exceeds more-likely-than-not (or greater than 50% probability), any failure of an SORA method or combination of methods to achieve admissibility under the more-likely-than-not standard would also preclude its admissibility under the more stringent, unspecified “highly likely” standard. This analysis should also inform SVP proceedings in other jurisdictions employing a de jure likelihood standard equal to or greater than more likely than not. In jurisdictions lacking a de jure likelihood standard, the analysis may aid in identifying the de facto likelihood standard actually being used.

### **FMHP and Legal Standards**

Although FMHPs have an obligation to be familiar with legal standards in the jurisdictions in which they testify (*APA Ethical Principles*, Standard 2.01(f); *Forensic Specialty Guidelines*, Standard III.C [4.04]; *APA Psychiatric Ethics*, Section 4), final determinations regarding admissibility of evidence and presence of reasonable certainty lie with the court, whereas adjudication of whether the likelihood standard has been met rests with the fact finder. Nevertheless, FMHPs have an obligation to understand the parameters of these standards so that they may provide the information relevant to a court or fact finder’s evaluation of the FMHPs’ testimony.

Under the *Frye* admissibility standard, a method will be admissible if, at the local recidivism base rate, the method is capable of yielding an APER (expressed in the form of

---

<sup>16</sup> This standard is shared by Arizona, Illinois and Minnesota (Doren, 2006). Six other states employ the more-likely-than-not standard: Florida, Iowa, Missouri, Washington, Wisconsin (Sreenivasan, Weinberger, & Garrick, 2003) and Massachusetts (Woodworth & Kadane, 2004). California’s courts do not require a finding of more likely than not (Sreenivasan et al., 2003). The remaining nine jurisdictions have not yet addressed this issue.

Statement B) which exceeds the likelihood standard (Wollert & Waggoner, 2009). Thus, as the highest recidivism risk estimate yielded by the Static-99R routine sample is 29.5% over five years (raw score = 9), the Static-99R with the routine sample table would likely be inadmissible as *sole* proof that an examinee's risk level exceeds the likelihood standard in a jurisdiction using a more-likely-than-not likelihood standard.<sup>17</sup> On the other hand, a score of 9 or less on the same version of the Static-99R may be admissible by the defense as proof that the examinee's risk level does not approach the likelihood standard.

Additionally, significant differences between an actuarial scale's development sample and the local population of SVP candidates or the examinee may render a particular SORA method inadmissible.<sup>18</sup> This is especially true when the sex-offense recidivism base rate in the development sample of actuarial scales differs significantly from the local base rates of sex reoffending. The validity of estimates of risk is intrinsically and mathematically related to the reliability of an actuarial method *and* the base rate of the target behavior (Helmus et al., 2009; Vrieze & Grove, 2008; Wollert, 2006). The lower the base rate, the greater the risk of overestimating risk of recidivism (Wollert & Waggoner, 2009).

Because base-rate information may be critical to the court's determination of an SORA method's admissibility, FMHPs have an ethical obligation to attend to and report on base-rate effects on APERs. For example, professional and ethical standards require FMHPs to obtain validity evidence for interpretations provided, especially given differences between local and development samples (*Testing Standards*, pp. 27, 83, Standards 1.4, 1.5, 11.2; Mossman, 2008) and, where appropriate, obtain and consider information regarding local norms (*Testing Standards*, p. 30; Boccaccini et al., 2009).

Reasonable certainty as to FMHPs' risk estimates corresponds to the level of precision selected for the CI reported. Although in the psychological sciences the 95% CI is commonly selected by convention, such a high level of precision may be greater than that required for a court to accept an expert's opinion (see above discussion noting that courts typically require certainty exceeding only 50%, or more likely than not). Unfortunately, when reported at all, actuarial scale authors typically only report the 95% CIs. Thus, unless FMHPs are able to obtain information necessary to calculate CIs at different levels of certainty, they must rely on the published ranges when available.

FMHPs seeking to base SORA on clinical judgment (structured or unstructured), or the clinically adjusted actuarial method, face significant hurdles in describing the level of confidence with which they report their risk estimate conclusions. In most cases,

---

<sup>17</sup> This is analogous to a breathalyzer device calibrated so as to yield as its highest blood alcohol level .07+ in a state where the threshold for DWI is .08. As the device is incapable of providing any value at or above the .08 threshold, the device yields no relevant evidence as to whether the defendant's blood alcohol level exceeds the legal limit and thus would be inadmissible.

<sup>18</sup> For example, in a New Jersey case involving an individual who committed his sex offenses as a juvenile and who spent his entire adult life in prison, the state's Appellate Court overturned an SVP determination which was grounded, in part, on an actuarial scale for which the development sample included no juveniles who had grown up in prison (*IMO the Civil Commitment of J.P.*, 2001).

FMHPs would have to admit that their clinical judgment, and risk estimates based on that judgment, lack any demonstrable or knowable reliability (unless FMHPs scrupulously maintained data over the years as to their personal rates of accurate versus inaccurate estimates of recidivism).

Indeed, to satisfy professional and ethical standards (*APA Ethical Principles*, Standards 1.01; 9.06; *Forensic Specialty Guidelines*, Standards II.B, III.B, IV.A.4, VII.A & VII.D [3.01, 4.05, 4.09, 13.01]; *APA Psychiatric Ethics*, Sections 2 and 4), FMHPs would have to acknowledge the limitations of clinically based APERs using language similar to that proposed below.

*Statement C:*

*Based on a review of factors of indeterminate validity as predictors of recidivism, assessed using a nonsystematic, unspecified process of unknown reliability, and combined according to an unspecified method of weighing those factors, I have classified the examinee as belonging to a group whose risk of reoffense over an indeterminate future period I have described as "high" based on my personal tolerances for risk.*

The fact that presentation of an expert conclusion phrased in this manner would likely lead to a bench ruling that such testimony failed to satisfy the local admissibility standard does not justify failure to disclose the limitations of such conclusions.<sup>19</sup>

Finally, assuming that the FMHPs' methods and conclusions survive analysis under the above two standards, the fact finder must evaluate whether the group to which the FMHP has assigned the examinee fits *within* the group of individuals defined by the likelihood standard (SVPs). For example, assume the FMHP reports, with 95% confidence, that the examinee belongs to a group whose members have between a 45% and 55% probability of reoffending in five years. As a practical matter, the fact finder must treat the individual as an interchangeable member of that group (since no further discrimination of the members of that group into smaller subsets based on reoffense risk is possible). The fact finder, in essence, then must determine if there is clear and convincing evidence that the group defined by law (SVPs) contains, as a subset, the group into which the examinee has been classified by the FMHP.<sup>20</sup> When the CI does not lie

---

<sup>19</sup> Indeed, at least two courts have ruled that normative predictions of risk, presented alone, are not admissible as expert scientific testimony due to their lack of reliability (*State of New Hampshire v. Ploof*, 2009; *U.S. v. Shields*, 2008).

<sup>20</sup> To most fact finders who consider themselves to be evaluating an individual's likelihood to reoffend, this will seem counterintuitive. To better illustrate this point, consider a scenario where likelihood determinations were based *solely* on an actuarial score (such as a score of 13 on the MnSOST-R). In that scenario, it becomes obvious that to find that one member of the group of individuals who scores a 13 is thus likely to reoffend means that all individuals who score a 13 meet that standard. If we apply a confidence interval to that score, the fact finder must likewise conclude (if we rely solely on the actuarial score) that if the SVP candidate's range of risk probabilities satisfies the likely-to-reoffend standard, then

entirely above the likelihood standard threshold, a fact finder may not be able to conclude with the requisite certainty that the examinee's true risk level, though bounded by the CI, falls above the risk threshold.

As fact finders are unlikely to perceive the classification process they are implicitly employing, FMHPs may play an important role in explaining the ramifications of conducting a probability determination. At the very least, FMHPs are obligated by ethical codes of conduct to provide sufficient contextual information so as to ensure that the risk information conveyed is not misused or misunderstood. FMHPs who, when reporting SORA results, report percentile ranks and subjective normative labels of risk, and who downplay, dismiss, or withhold APERs for an examinee, may fail to satisfy these ethical obligations (*APA Ethical Principles*, Standards 1.01, 3.04, 9.02, 9.06; *Forensic Specialty Guidelines*, Standards II.B, VII.A and VII.D [4.09, 13.01]; *APA Psychiatric Ethics*, Section 2).

Having established the ethical, professional, and legal framework for SORA, the manner in which SORA is conducted in one state, New Jersey, can be evaluated for conformity to these standards.

### **SORA in New Jersey**

#### **Expert Testimony Regarding the Likelihood Standard in New Jersey SVP Cases**

To what extent do actual professional and legal practices in New Jersey adhere to the conceptual framework described above? And is there evidence that fact finders in New Jersey are employing a likelihood standard that exceeds more likely than not to an appreciable degree? Because the records in SVP proceedings in New Jersey are sealed to protect the identities of examinees, the only public access to the expert testimony presented in SVP cases comes from reported decisions of the New Jersey Superior Court, Appellate Division, and the New Jersey Supreme Court hearing appeals of lower court decisions. Since not all lower court decisions are appealed and only a small fraction of these appeals may result in a published opinion, review of law reporters provides a somewhat limited overview of what may transpire during actual SVP proceedings. Nevertheless, as controlling precedent, these cases establish, even if somewhat indirectly, what currently constitutes acceptable expert opinion. Moreover, those decisions frequently contain extensive summaries of, or quotations from, the most probative expert testimony heard at the proceedings. As a sample of the type of testimony prof-

---

anyone whose score yields the same range of probabilities must also satisfy this standard. In an alternative scenario in which we now add other risk-correlated factors (as in the clinically adjusted actuarial method), we have only potentially narrowed the class of eligible individuals; the fundamental analysis hasn't changed. If a fact finder determines that an individual with a score of 9 on the Static-99R (APER between 22.8% and 37.2% on the routine sample) who also manifests Factors A, B, and C, satisfies the likelihood standard, then, to avoid a claim of arbitrary and capricious decision making, any individual with a Static-99R score of 9 who manifests these three factors must also be found to satisfy this standard.

ferred, at the very least, a review of these cases may provide insight into what courts in New Jersey consider probative in the reports and testimony of FMHPs.

Toward this end, we searched the New Jersey edition of the Lexis Nexus legal database of reported trial, appellate, and Supreme Court cases dating back to 1999, the year the SVPA was implemented. A search for the term “sexually violent predator” in the text of any published decision and the word “commitment” in the caption of the case<sup>21</sup> yielded 31 hits. Each case was reviewed to locate a description of recidivism risk provided by an expert witness and reported in the lower court record. Two cases were duplicates; the same case heard at two different levels of appeals. Of the unduplicated cases, thirteen dealt with procedural issues or various constitutional issues unrelated to risk assessment per se. The remaining sixteen cases included summaries and quotations from the risk assessments conducted by forensic experts at the trial level, most of whom scored multiple actuarial scales.

Of these, only two cases cited forensic evidence containing quantified APERs/RPERs (in both cases reporting the risk percentages associated with actuarial scores; *IMO the Civil Commitment of A.E.F.*, 2005; *IMO the Civil Commitment of E.S.T.*, 2004). For the remaining fourteen cases, 87.5% of those surveyed, only FMHPs’ reports of normative risk estimates appeared. The terms used to describe risk included: “high,” “high to moderate,” “high danger,” “highly likely,” “moderate,” “extraordinary risk,” “significant risk,” “significant level of risk,” and “very substantial risk.” Typical was the form of report quoted in *IMO the Civil Commitment of G.G.N.* (2004, p. 48). “[Dr. Schaupp’s] report at one point states: . . . [G.G.N.] was administered the MnSOST-R and Static-99, which both yielded results that place [G.G.N.] in the high risk category for sexual reoffense.” Despite the ready availability of APERs from the actuarials’ experience tables, it appears, at least among the published decisions, that the courts view these normative risk estimates as most salient.

This finding is especially disconcerting given the emphasis placed on the reliability of actuarial assessment by state experts in the seminal case of *IMO the Civil Commitment of R.S.* (2001), which upheld the admissibility of actuarial instruments under the *Frye* standard. Testifying for the state, Dr. Dennis Doren averred that a clinically adjusted actuarial method, such as that purportedly used in New Jersey SVP proceedings by the state experts, was the most accurate method for risk assessments in that it starts with an actuarial assessment and then makes clinical adjustments to the obtained risk estimate based on the specific details of each case (p. 521).

---

<sup>21</sup> In New Jersey SVP cases the word “commitment,” as in “*IMO the Civil Commitment of W.Z.*” appears in the case caption. A review of SVP civil commitment docket numbers indicates that as of January 2011, approximately 600 petitions for orders of temporary civil commitment were filed in New Jersey Superior Courts since 1999. As the 16 reported cases show, only a very small percentage of these cases generated a published appellate decision describing findings as to the likelihood standard in the lower courts.

To examine the extent to which actual practice in New Jersey courts, as represented by these cases, coincides with the idealized practice and mode of analysis recommended by state experts in *R.S.* and the standards and analytic model propounded above, each of the sixteen reported cases was again reviewed to discern the extent to which clinicians started with the actuarial risk estimate and then adjusted their final rating of risk by other empirically validated factors. In no reported case did this pattern emerge, at least on the appellate record. Indeed, as exemplified in the excerpts below, almost universally, state experts described a litany of “factors” that pointed to the examinee’s high reoffense risk, which they then “validated” with a normative risk rating associated with the examinee’s actuarial score.

For example, in *IMO the Civil Commitment of T.J.N.* (2007, p. 223), the state expert described the examinee’s repeated sexual institutional infractions, poor impulse control, and lack of remorse concluding, based on those factors, that he was “at extraordinary risk to sexually reoffend if released. *Actuarials are consistent with this analysis*” (*emphasis added*). Thus, at least in this case, the state expert appeared to validate the clinical findings with the actuarial scale results, just the opposite of the procedure endorsed in *R.S.*

In an effort to obtain more information about how SORA is conducted by experts testifying for the State in SVP civil commitment proceedings, the authors contacted two attorneys with extensive experience defending individuals against SVP. Based on their personal experiences both generally confirmed this description of the SORA process in New Jersey (J. S. Furlong personal communication, May 23, 2011; J. Van Pelt, personal communication, June 21, 2011).<sup>22</sup> Their perceptions of the frequency with which state experts reported risk estimates expressed as percentages for actuarial raw scores differed, however.

Furlong noted that in none of the cases he defended did state experts report the percentage of reoffenders associated with each actuarial raw score, which observation comports with the experience of the first author who had the opportunity to review reports filed in approximately two dozen SVP cases between 2000 and 2009. In no case he reviewed did the evaluating FMHP report risk estimates as percentages.

In contrast, Van Pelt reported that, in her recent experience, state experts had been reporting the percentages of reoffenders associated with actuarial raw scores, if not in their written reports, then in oral testimony at the SVP hearing. Notwithstanding reporting these risk percentages, since most estimates obtained were relatively low, Van Pelt noted that the state experts had to rely on other factors to justify elevating an individual’s risk to a level approaching or exceeding the threshold of “highly likely” to reoffend.

---

<sup>22</sup> John Furlong, an attorney in private practice, has defended approximately 20 civil-commitment cases over the years, while Joan Van Pelt, an attorney first with the New Jersey Office of the Public Defender and later with the New Jersey Office of the Public Advocate, headed those Offices’ SVP Civil Commitment Units from 2001 to 2010 and oversaw hundreds of these cases.

To the extent that state experts rely on other risk factors to elevate recidivism risk estimates, the findings suggested by the review of the Appellate Court decisions noted above found support in the experiences of the first author, as well as both attorneys; apparently state experts rarely, if ever, describe the empirical evidence justifying reliance on those factors.

### **Application of Professional and Legal Standards to SORA in New Jersey**

FMHPs in New Jersey, at least as revealed in the appellate record of SVP cases, appear to conduct a form of SORA that employs a hybrid mix of unstructured clinical judgment and actuarial risk assessment, while eschewing any form of structured professional judgment. In no reported case did it appear from the record that an FMHP reported the scientific evidence in support of the supposed link between a named factor and recidivism, how the factor was defined and assessed, and the degree to which that “factor” enhanced the examinee’s risk over the base rate for the population to which the examinee belonged (Hanson, 1998; Wollert, 2004). For example, factors reported by FMHPs as increasing reoffense risk included an absence of a “fundamentally sound understanding of his deviant sexual arousal” and a “conscience impairment” (*IMO the Civil Commitment of M.L.V.*, 2006, p. 467); being “deceitful and manipulative” (*IMO the Civil Commitment of J.P.*, 2007, p. 14); having limited “insight into the role of planning and grooming” of a victim (*IMO the Civil Commitment of A.H.B.*, 2006, p. 20); and lack of “victim empathy” (*IMO the Civil Commitment of P.Z.H.*, 2005, p. 461). In each case, however, the appellate court record is devoid of any description of the empirical support for or objective measurement of these factors. It is possible that FMHPs may have proffered such scientific evidence and the trial and appellate judges merely failed to report that evidence in their opinions. However, that seems unlikely as most of the factors relied upon by the FMHPs were, as the examples above illustrate, so idiosyncratically or vaguely worded as to be unavailable for scientific scrutiny (indeed, none of the factors listed above was identified as significantly correlated with recidivism in any of the large, recently published meta-analytic studies by Hanson and his colleagues; see, e.g., Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004, 2005). Moreover, as noted above, the experience of the first author and of two attorneys with experience defending these cases, confirms that state experts rarely, if ever, provide evidence in support of the validity of risk factors they identify (J. S. Furlong, personal communication, May 23, 2011; J. Van Pelt, personal communication, June 21, 2011).

Thus, based on this preliminary data, the SORA method that FMHPs employ (or, at least, the method some attorneys and Appellate Court judges perceive that they employ) may be described. FMHPs in New Jersey appear to combine actuarial scores and unstructured clinical judgment to generate a RPER which they then report as a normative prediction of sex-offense recidivism risk. So defined, this SORA method can be evaluated under the pertinent legal and professional standards described above.

**Evaluation of New Jersey SORA under the admissibility standard.** As noted above, to be admissible, the methods employed must be relevant and reliable, as measured by their general acceptance in the relevant professional community. The

New Jersey courts have already ruled that actuarial scales as a class of SORA methods are admissible in New Jersey because they have been generally accepted by the professional community (*IMO the Civil Commitment of R.S.*, 2001). It also appears that the court has endorsed the “clinically adjusted actuarial method” described in *R.S.* That does not end the analysis, however. Whether the manner in which actuarials are currently used in New Jersey continues to satisfy the general-acceptance standard turns on two issues: (a) the general acceptance of reporting only normative descriptions of risk based on combining actuarial scores and unstructured clinical judgment; and (b) the current capability of actuarial instruments to generate risk estimates at or above the critical threshold set by the state’s likelihood standard. Whether the clinically adjusted actuarial method in which structured professional judgment is replaced by unstructured clinical judgment also enjoys general acceptance in the field depends on a review of the extensive body of research published since the court’s 2001 ruling and a determination of the reliability and validity of particular factors considered.

**Admissibility of actuarial-based assessment.** Reporting that an examinee’s raw score on an actuarial scale falls into a range designated by the scale’s developer as “high risk” would not seem to satisfy the general-acceptance test. Whether to assign the label “high risk” to a 30%, or 45%, or 70% probability of recidivism requires no application of specialized knowledge, training, or experience and arguably does not even represent knowledge beyond the ken of the average person. Certainly, nothing in the peer-reviewed literature supports this practice; indeed, the same probability of risk may be characterized as “low,” “medium,” and “high” by different evaluators (Hilton, Carter, Harris, & Byrans, 2005, unpublished manuscript, as cited in Rice & Harris, 2005; see also Campbell, 2007; Monahan, 1996). Given that different scale developers assign very different risk labels to the same probability of risk (14% on the Static-99R equating to “high” risk while MnSOST-R authors classify virtually the same percentage as “low” risk), and that the labels associated with various risk probabilities have varied over different versions of the same scale (14% risk classified as “low risk” in the original Static-99 sample [Hanson & Thornton, 1999] but “high risk” in the revised Static-99R sample [Phenix et al., 2009]), normative labels cannot be considered to possess either scientific or evidentiary reliability.

Rather, normative labels such as these incorporate norms and values reflective of the expert’s (or the author of the scale’s) personal tolerances for risk (how high is high risk) and/or personal weighing of the relative costs to society and to the individual of erroneous decisions (Schopp et al., 1999). However, FMHPs are obligated to limit their testimony to areas of professional expertise and competence (*APA Ethical Principles*, Standards 1.01, 2.01, 2.03, 2.04, 9.06; *Forensic Specialty Guidelines*, Standards II.B, III.A, III.B, III.E, IV.A.4, VII.A & VII.D [3.01, 4.05, 4.07, 4.09, 13.01]; *APA Psychiatric Ethics*, Section 2 and 4). It would appear then that, where state experts have failed to report the APERS for obtained actuarial raw scores, as may have occurred in some of the reported appellate decisions, and as noted in the experiences of the first author and an attorney experienced in SVP cases (J. S. Furlong, personal communication, May 23, 2011), such forensic practice may fall short of satisfying professional practice standards.

Where FMHPs do report APERs yielded by the Static-99 and MnSOST-R are these actuarials alone capable of yielding APERs at or above a likelihood standard of at least greater than 50% reoffense risk? The recidivism base rate in the development sample for the MnSOST-R (5- to 6-year recidivism rate of 35%; Epperson et al., 1999) and the original Static-99 (5-year recidivism rate of 19%; Hanson & Thornton, 1999) substantially exceed contemporary local (New Jersey) and national rates. A series of studies published in the last decade reveal a sex-offense recidivism rate in New Jersey ranging from 6%-7% for 5-6 year follow-up periods (Sager, 2000, 2001; Zgoba et al., 2008), and from 3%-12% for 10-year follow-up periods (Sager, 2002; Zgoba et al., 2003).<sup>23</sup> These rates are consistent with more recently published rates from across the country (Hanson & Bussiere, 1998 and Hanson & Morton-Bourgon, 2005, 2007, 2009 reporting 5- to 6-year recidivism rates of 11%-14%; Wollert & Waggoner, 2009, summarizing state department of corrections data,  $n = 17,697$ , revealing an average 5-year recidivism rate of 6.5%). Consequently, reliance on unadjusted risk estimates from the Static-99 and MnSOST-R experience tables may result in overestimates of risk.

The base-rate differences between these scales' development samples and the local New Jersey population were not mentioned in any of the reported decisions. FMHPs who proffer unadjusted or unqualified APERs generated by the MnSOST-R and Static-99 are likely violating numerous standards or guidelines of the *APA Ethical Principles* (Standards 2.04, 3.04, 9.01(a), 9.02, 9.06, and 9.08), *Forensic Specialty Guidelines* (Standards II.A, II.B, IV.A.4, VI.A, VII.A, and VII.D [3.01, 4.05, 4.08, 12.02]), *Testing Standards* (Standards 1.1, 1.4, 1.5, 1.19, 1.22, 2.15, 11.2, 12.6, 12.13, 12.17) and *APA Psychiatric Ethics* (Sections 2 and 5). Consequently, use of unadjusted risk estimates from these scales should not be considered generally accepted by the professional community.

On the other hand, the renorming of the Static-99 scales and the addition of two new scale revisions (the Static-99R and Static-2002R, routine samples) were derived with development samples whose base rates (6% and 5% respectively over 5 years Campbell & DeClue, 2010a) were more consistent with the local and national norms noted above. However, all three Static scales still fall short of meeting the admissibility standard as none of these instruments appears capable of yielding a risk estimate for any examinee that even approaches the New Jersey likelihood standard of "highly likely," let alone exceed the more-likely-than-not standard.

The highest risk estimates, reported as logical regression estimates, yielded by the routine samples of the Static-99R and Static-2002R scales are, for a five-year follow-up period, 29.5% (95% CI [22.8, 37.2]) and 10.5% (95% CI [5.0, 20.6]) respectively. Despite failing to approach the likelihood standard, because these sample base rates

---

<sup>23</sup> Significantly, the recidivism rate for those presumably at higher risk because they were classified by the New Jersey courts as compulsive and repetitive sex offenders in need of intensive specialized treatment and committed to the Adult Diagnostic & Treatment Center (ADTC population) in Avenel, New Jersey, obtained lower, albeit not statistically significant, recidivism rates than sex-offending inmates released from the general prison population (8.6% versus 12.7% respectively; Zgoba et al., 2003). See Zgoba et al. (2003) for a detailed description of the ADTC program and inmates.

closely approximate the New Jersey's sample base rate, these routine samples may be the most appropriate reference group for the New Jersey SVP candidates (while the nonroutine and high risk/need samples for these scales provide probability estimates in excess of the 50% [the more-likely-than-not] threshold, they are derived from samples with 10-year reoffense base rates ranging from 20% to 29%—as reported by Campbell & DeClue, 2010a—two to three times higher than reported New Jersey base rates).

Despite these shortcomings, the continued reliance upon these scales in SVPA assessments in New Jersey may reflect: (a) a lack of knowledge or understanding of the base-rate differences or its significance by lawyers, judges and experts; or (b) implicit use of a likelihood standard far below that set forth by the New Jersey Supreme Court in *IMO the Civil Commitment of W.Z.* (2002).

Also problematic is FMHPs' reliance on unadjusted APERs from *both* the MnSOST-R development sample, and the development samples of the Static group of actuarials when assigning normative labels of risk (e.g., see *IMO the Civil Commitment of R.S.*, 2001, p. 517; *IMO the Civil Commitment of W.Z.*, 2001, p. 560). For example, given 5- to 6-year base rates for the MnSOST-R and Static-99R of 35% and 7% respectively, and the fact that an examinee can't be drawn from a population that has both a 35% and 7% recidivism base rate, one of the scales will inevitably overestimate or underestimate risk. Best practice dictates selecting the single most reliable scale, given the similarities between the scale's development sample and the local population from which the examinee is drawn (Boccaccini et al., 2009; Seto, 2005).<sup>24</sup>

It would appear then that no current actuarial scale, standing alone, can yield APERs above the likelihood standard given the low sex-offense recidivism base rates in New Jersey.

***Admissibility of clinically adjusted actuarial method using clinical judgment.*** Can either unstructured or structured clinical judgment alone or in combination with actuarial assessment yield APERs in excess of the likelihood standard and thus satisfy the admissibility standard? Clearly unstructured clinical judgment alone lacks sufficient reliability and validity to be used to assess recidivism risk. Moreover, it can only yield RPERs, rendering its estimates irrelevant to the likelihood standard. Structured clinical judgment (or structured professional judgment), when limited to empirically validated risk factors, may have sufficient reliability and validity to justify their use from an ethical perspective. However, the highest correlations reported in the literature for a single factor (stranger victim,  $r = .22$ , Hanson & Bussiere, 1998) or multiple factor models (Hanson & Harris, 2000, nine-factor model,  $r = .60$ ) only elevate reoffense risk by about 11% and 30% respectively over base-rate levels (using the method described by Hanson, 1998, described above). No factor or combination of factors empirically validated as correlating with sex reoffending appears capable of elevating recidivism risk to

---

<sup>24</sup> According to Joan Van Pelt, Esq. (personal communication, June 19, 2011), for several years state experts appeared to abandon any reliance on the MnSOST-R when estimating reoffense risk, but that recently, reports of MnSOST-R-based recidivism estimates have begun to reappear in some state experts' reports.

a likelihood standard of more likely than not or better, given New Jersey's sex reoffense base rates (the highest 10-year sex-offense recidivism rate reported for New Jersey's general sex-offending population is 12.7% while the rate for the treated ADTC population [see footnote 23] is 8.6%; Zgoba et al., 2003).

Use of nonquantifiable structured clinical/professional judgment methods (such as recommended by Douglas & Skeem, 2005) with actuarial risk estimates, results in an RPER that fails to provide discriminative information for use in reclassifying SVP candidates into the likely-to-reoffend bin. On the other hand, empirically validated factors with known correlations with recidivism, when combined with actuarially derived APERS, could, conceivably, yield estimated APERS in excess of the more-likely-than-not threshold. Unfortunately, FMHPs have no validated model for combining actuarially-derived APERS with the correlation coefficients of validated factors, rendering this approach currently unavailable in forensic assessments.

Although obtaining such a result for any given SVP candidate would be an infrequent occurrence, conceivably, use of the clinically adjusted actuarial model based on the use of actuarial assessment and quantifiable structured clinical judgment passes the admissibility standard in that it uses methods that are generally accepted and it is capable of yielding results relevant to making a classification under the likelihood standard. Unfortunately, despite the enthusiasm for this model reported in *IMO the Civil Commitment of R.S.* (2001), which found it admissible under the *Frye* standard, the review of published cases and the anecdotal evidence above reveals little evidence this method is used as intended. Rather, in New Jersey, on the record reviewed, experts and the courts appear to rely upon a combination of wholly unstructured clinical judgment which is then combined in a nonsystematic, unspecified manner with normative judgments of risk derived from actuarial assessment. Not only does such a method lack any support in the SORA literature, or in the field of risk assessment, FMHPs' use of such a method, absent admission as to its lack of reliability and validity as an SORA method, violates the ethical standards described above. There is no evidence this approach is generally accepted, a finding confirmed by at least two jurisdictions noted above (*State of New Hampshire v. Ploof*, 2009; *U.S. v. Shields*, 2008).

Thus, given New Jersey's low sex-offense recidivism base rate, neither clinical judgment (in any form) nor sole reliance on currently available actuarial instruments, appear capable of generating an APER in excess of the State's likelihood standard, rendering sole use of either method irrelevant and thus inadmissible for SVP determinations that require a likelihood standard of at least more likely than not or greater. Moreover, reporting normative labels while withholding APERS from fact finders when describing reoffense risk is not generally accepted in the field of SORA. Combining actuarial assessment with structured clinical/professional judgment, using empirically validated factors with demonstrable incremental validity, may satisfy the admissibility standard. However, a review of the published appellate decisions of SVP proceedings and the anecdotal experiences of two SVP attorneys produced no evidence this method was actually employed in New Jersey.

***Evaluation of New Jersey SORA under the reasonable certainty standard.***

APERs lend themselves readily to a reasonable-certainty analysis. As explained above, APERs expressed in the form of Statement B incorporate information about the degree of confidence with which a risk estimate is expressed (in the form of the confidence interval). FMHPs employing actuarial assessment instruments in New Jersey SVP commitment proceedings should be able to satisfy the reasonable certainty standard by reporting the CIs associated with the risk estimates they report.

Satisfying this standard becomes considerably more difficult with unstructured clinical judgment. Indeed, there is no known method for evaluating whether the idiosyncratic application of a set of internal heuristics derived from nonsystematic, unrepresentative recollections of prior contacts with sex-offending individuals, to a determination of a new sex-offending individual's reoffense risk, has any knowable degree of certainty. Because the rates of error associated with risk estimation based on unstructured clinical judgment are unknown and unknowable, FMHPs in New Jersey relying on this method, if asked if they are reasonably certain as to their risk estimates, have only one ethically permissible and nonperjurious answer available to them: "I don't know."

Presentation of a risk estimate grounded on the clinically adjusted actuarial method may compromise an FMHP's reasonable certainty in proffered risk estimates. For example, if FMHPs make no attempt to evaluate the degree of covariance or the extent to which addition of an empirically validated risk factor adds incremental validity to a risk estimate generated from actuarial assessment, they may have difficulty ethically asserting a reasonable certainty in a risk estimate borne of a combination of these methods. As DeClue and Campbell (2010) note, the addition of six items to the Rapid Risk Assessment for Sex Offense Recidivism (RRASOR: Hanson, 1997) to create the Static-99 actuarial scale increased predictive accuracy by only a small degree (AUC increased from .68 to .71; DeClue and Campbell, p. 332). Circumstances under which FMHPs may be able to ethically assert with reasonable certainty that enhancement of an actuarially derived risk estimate based on additional factors is justified may be exceedingly rare.

***Evaluation of New Jersey SORA under the likelihood standard.*** As explained above, the New Jersey Supreme Court adopted a likelihood standard that requires a likelihood of sexual reoffense at some unquantified degree above 51% probability. The analysis presented in this article demonstrates that only in the rarest of cases will an FMHP assessing an SVP candidate in New Jersey be able to present an expert opinion to the fact finder that is both admissible and which the FMHP is reasonably certain will satisfy the likelihood standard. Similar to Boccaccini et al.'s (2009) findings in Texas, the base rate of sex-offense recidivism in New Jersey (8-12% over 10 years) makes it very difficult to accurately identify the approximately one in 10 sex-offending individuals who will recidivate, without misclassifying large numbers of the more than nine out of 10 sex-offending individuals who will not recidivate.<sup>25</sup>

---

<sup>25</sup> While a 10-year recidivism rate fails to anticipate reoffenses occurring after 10 years, the New Jersey Courts limited risk predictions to the reasonable foreseeable future (*IMO the Civil Commitment of W.Z.*, 2002). In more traditional civil commitment proceedings, FMHPs typically consider long-term predictions

As of 2009, the Special Treatment Units in New Jersey to which individuals are committed after being found to be SVPs housed about 405 civil committees (Gallagher, 2009). Ninety percent of those who appear before a judge for a final commitment hearing are found to be SVPs (Gallagher). In light of the recidivism base rate, three possible explanations for this commitment rate can be offered: (a) the screening process that determines who will be subject to SVP commitment proceedings may be extraordinarily accurate at identifying SVPs; (b) the final commitment process is extraordinarily inaccurate at identifying non-SVPs; or (c) in actual practice, New Jersey trial judges adjudicating SVP hearings are not utilizing the likelihood standard identified by the New Jersey Supreme Court but rather are using a standard well below the more-likely-than-not standard.

To test the first possibility, we estimated the likelihood ratio (LR) associated with a screening process by which 90% of those initially screened for possible civil commitment were correctly identified as true recidivists (in this scenario, the final commitment of an individual as an SVP by a court is being treated as a true positive; in New Jersey, all those convicted of a sex offense and about to be released from prison, are screened by the Attorney General's office for possible civil commitment).

Wollert (2007) reported the Bayesian formulas for determining the LR<sup>26</sup> associated with a particular base rate, or P(R), and the percentage of classifications that were true positives (Positive Predictive Value or PPV).

$$\frac{[PPV \times P(R)] - PPV}{[PPV \times P(R)] - P(R)} = LR$$

By inserting into the above formula the percentage of SVP candidates screened for civil commitment who are actually committed (PPV = 90%), the highest base rate of sex-offense recidivism reported for New Jersey, or P(R) = 12.7% (Zgoba et al., 2003, 10-year recidivism rate, untreated population), one can estimate the LR associated with the screening process.

$$\frac{[.90 \times .127] - .90}{[.90 \times .127] - .127} = 62$$

---

of risk as encompassing months, not years (Garb, 1998; Monahan, 1996). As recidivism rates beyond 1 to 5 years post-release drop precipitously (e.g., Hanson & Thornton, 1999; Rice & Harris, 1997; Song & Lieb, 1995), the capacity to reliably predict who will reoffend declines as well. Moreover, confining someone for a 10-year period during which he would not reoffend, because he may commit a sex offense over the next five to ten years, arguably becomes indistinguishable from preventative detention. The New Jersey Supreme Court's decision in *IMO the Civil Commitment of W.Z.* (2002, p. 132) appears consistent with this view as the Court emphasized the focus of risk assessment should be on "present" risk of reoffense. Remaining offense free for up to 10 years strongly suggests the absence of a "present" risk.

<sup>26</sup> The likelihood ratio is a factor of the ability of the screening process to correctly identify recidivists while minimizing incorrect identification of nonrecidivists: the higher the LR the more accurate the classification process (Wollert, 2007).

A LR of 62 for the SVP screening process in New Jersey is extraordinarily implausible.<sup>27</sup> As Wollert (2007) reported, the LR for actuarial assessment of sex-offense recidivism is about 3.1. Campbell and DeClue (2010a) reported the LRs for different cut scores on the Static family of actuarial scales. For the 36 LRs reported, we calculated the mean LR ( $M = 3.34$ ;  $SD = 3.20$ ). Given actuarially based LRs in this range, FMHPs would not be justified in concluding that an unresearched and undocumented screening process could generate an LR eighteen times greater than that of well-researched and demonstrably reliable actuarial measures.

To test the second possibility, we estimated the true-positive rate or PPV that one could expect for a screening process with a LR consistent with LRs typically obtained in actuarial assessment of recidivism risk, and assuming a sex-reoffense base rate consistent with local New Jersey rates.

We can estimate the PPV (i.e., the accuracy of the SVP process; Wollert & Waggoner, 2009) that can be expected given a  $P(R)$  of .127 and an LR of 3.34 (equivalent to the average of the best actuarially based assessments; Campbell & DeClue, 2010a),<sup>28</sup> using another Bayesian formula provided by Wollert and Waggoner (2009).

$$\frac{\left(\frac{P(R)}{1 - P(R)}\right) \times LR}{1 + \left[\left(\frac{P(R)}{1 - P(R)}\right) \times LR\right]} = PPV$$

Substituting the values above for the formula yields a PPV of 32.7%.

<sup>27</sup> LR is also a function of a scale's true-positive fraction or sensitivity (probability the scale will correctly classify a recidivist [T]) and its false-alarm rate or 1 - specificity (probability the scale will incorrectly classify a nonrecidivist [F]) such that  $LR = T / F$  (Wollert, 2006). As neither T nor F can exceed 1.00, knowing the LR and assuming perfect sensitivity ( $T = 1.00$ ), we can determine the highest false alarm rate possible for an LR of 62 by solving for F:  $62 = 1.00 / F$ ;  $62 * F = 1.00$ ;  $F = 1.00 / 62$ ;  $F = .016$  or 1.6%. Thus, for an assessment method to attain an LR of 62, it could misclassify no more than 1.6% of nonrecidivists as recidivists while correctly identifying 100% of recidivists. No assessment protocol could conceivably attain such high rates of accuracy.

<sup>28</sup> In actually, this average LR will tend to overclassify individuals as likely to reoffend as the average PPV associated with the 36 LRs reported by Campbell and DeClue (2010a) was only .375 or 37.5% ( $SD = .11$ ), well below a more-likely-than-not standard. Moreover, the 36 LRs and PPVs included one instance of an LR of 20 derived from a cut score yielding one true-positive and one false-positive recidivism classification. Eliminating this unstable LR from the analysis yields a mean LR of 2.87 ( $SD = 1.47$ ,  $n = 35$ ) and a mean PPV of .372 ( $SD = .11$ ,  $n = 35$ ).

$$\frac{\left(\frac{.127}{1-.127}\right) \times 3.34}{1 + \left[\left(\frac{.127}{1-.127}\right) \times 3.34\right]} = .327$$

Thus, given a presumed New Jersey sex-offense base rate of 12.7% over 10 years, and assuming a classification process at least as reliable as the best actuarial instruments available today (noting that adding clinical judgment to actuarially assessed predictions lowers overall predictive accuracy; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009), one can expect that only 33% of those predicted to be SVPs during the screening process were actually SVPs. In other words, given a 12.7% sex-offense recidivism base rate and a LR of 3.34, one would expect that only 33% of the individuals identified as putative SVPs (true positives) by the initial screening process, would actually turn out to be recidivists. Comparing the 33% rate to the 90% actual commitment rate at final SVP hearings indicates that there is a very high probability that three times as many non-SVPs are being civilly committed in New Jersey as true SVPs (so far we've only considered the likelihood to reoffend prong of the SVP statutes; problems with the reliability of determinations of the mental abnormality prong may also be contributing to this commitment rate). Thus, it seems highly probable that the process in New Jersey misclassifies as SVPs a significant proportion of nonrecidivists.

### Limitations and Alternative Considerations

Clearly, relying upon appellate court decisions and anecdotal evidence significantly limits the inferences we can draw about the type of expert testimony proffered at SVP hearings. Appellate court decisions condense and summarize an evidentiary record that often runs to several hundred pages of testimony. Consequently, it is impossible to know whether failure of an appellate court to reference a risk estimate expressed as a percentage, or to relate the scientific evidence in support of a risk factor identified by an expert, is due to the absence of that information in the expert's reports and testimony, or evidence of what the Appellate Court, in its judgment, considered the most salient data. In addition, recollections of attorneys defending these cases may not always comport to actual evidence or testimony proffered due to erroneous recollections.

Moreover, it may be that judges struggle with some of the probability data presented in SVP hearings and thus latch on to the easier-to-understand normative labels such as "high risk." Or the New Jersey Courts may have a more liberal approach to the admission of clinically grounded testimony under the *Frye* standard than some of the jurisdictions which have rejected such testimony under the *Daubert* standard (such as *State of New Hampshire v. Ploof*, 2009; *U.S. v. Shields*, 2008). Or perhaps courts at the appellate level refrain from summarizing scientific evidence presented at the trial level.

Additionally, given the small number of reported appellate decisions compared to the number of civil-commitment hearings actually conducted,<sup>29</sup> drawing inferences about the quality and content of expert testimony in the larger sample of unpublished commitment orders and appeals, may be difficult to justify.

However, for a number of reasons, we believe this limited review of expert testimony as revealed in appellate court decisions and the anecdotal data reported above, should raise red flags for both FMHPs and attorneys and judges. First, though the question of whether or not state experts report risk estimates expressed as percentages in their reports and/or testimony remains unclear (as the competing experiences of the first author and two attorneys specializing in SVP cases as described above, demonstrated), given that the risk estimates the scales can reliably generate are relatively low (and usually well under 50%), their use by the courts in supporting a finding that the likelihood standard has been met, suggests that, at the least, experts for both the state and defense are not adequately explaining how these data should be interpreted.

Second, even a cursory review of the risk factors cited by appellate courts supports the conclusion that the named factors are not of the type the field has operationally defined and subjected to empirical validation (see examples provided above). The experiences of the first author and of the two attorneys noted above support the inference drawn from the appellate record that state experts don't usually cite scientific research to support the risk factors they employ.

Third, while some of the probability evidence may not be intuitive, the New Jersey Courts have demonstrated a capacity for understanding this type of testimony in the past. In *State of New Jersey v. Spann* (1993), the Court demonstrated a highly nuanced understanding of probabilistic reasoning, Bayes Theorem, prior probabilities, likelihood ratios, and positive predictive values in the context of paternity testing. Nor have the state courts been overly lenient in admitting clinically based testimony that lacks a research basis, or reticent to summarize the relevant literature. For example, in *State of New Jersey v. Cavallo* (1982), *State of New Jersey v. Fortin* (2000), and *State of New Jersey v. Raso* (1999), the New Jersey appellate courts extensively discussed and weighed the relevant published research and demonstrated a willingness to reject clinically based testimony that was not grounded in scientific evidence.

Fourth, even if FMHPs are providing the appropriate information to the courts as to risk percentages and empirically validated factors, the absence of any reference to this information in most appellate decisions suggests that their importance may not have been stressed to a sufficient degree to gain the court's attention, or the courts have failed to recognize the significance of these data.

---

<sup>29</sup> Not all of the 600 petitions for orders of temporary civil commitment filed as of February 2011 necessarily led to a full hearing. Some individuals are released prior to a hearing when the state determines the temporary order was sought in error, some individuals die, and others might be rearrested on another charge and be returned to state prison.

Whatever the limitations of the data set, we do not believe these limitations blunt the force or importance of the recommendations offered below for two reasons.

First, if, in fact, FMHPs are providing scientifically validated evidence in support of their SORAs, it may be the courts are misinterpreting or misunderstanding this data and need to be better educated as to its use by FMHPs. In this case, attorneys and judges need to incorporate the recommendations below when considering expert testimony and when questioning the expert about his or her findings.

Second, if FMHPs are, in fact, reporting risk estimates solely as normative labels or percentile rankings and/or relying upon risk factors that have not been operationally defined and empirically validated as correlated with recidivism risk, then the recommendations provide a roadmap for ethical practice.

### **Recommendations**

This review and analysis suggests that the SVP screening and process in New Jersey may be misclassifying individuals as SVPs at a significant rate (up to 67% of individuals misidentified as SVPs). Based on the data of the appellate cases reviewed and the experiences of a nonrandom sample of two attorneys involved in the majority of SVP cases over the last decade, we tentatively offer several explanations. First, judges and attorneys may lack the technical knowledge to recognize and process the most salient aspects of FMHPs' reports and testimony, relying on normative labels such as "high" risk as a shorthand method of processing the probability data presented.

Second, FMHPs may be systematically and consistently rendering expert testimony that: a) fails to express risk estimates in terms of percentages or fails to place that information in the proper context if it is reported; b) relies on putative risk ". . . factors based only on [the] plausibility" of the factor as a predictor of risk (Hanson, 1998, p. 61) rather than factors empirically linked to recidivism risk; c) fails to consider whether risk factors identified add any incremental validity, and, if so, how much to the actuarially assessed risk estimate; and/or d) fails to report or consider sex-reoffense base-rate data for New Jersey releasees and the impact of those base rates on predictive validity.

Third, judicial fact-finders may be implicitly or explicitly using a likelihood standard well below the more-likely-than-not standard required under New Jersey law.

The third alternative is beyond the scope of this paper and data set. And while deficiencies in the manner in which the mental abnormality standard is interpreted and applied may also contribute to inaccurate SVP classifications, focusing reforms on the first and second alternatives above, may significantly enhance the valid application of the likelihood standard.

Whether classification errors in SVP proceedings originate with legal professionals (judges and attorneys) or FMHPs, the recommendations that follow should inform the practices and decisions of both groups of professionals. Attorneys and judges can

demand testimony from FMHPs that conform to these recommendations and experts can insist on communicating assessment results and behavioral predictions in accord with these recommendations.

First, FMHPs should refrain from proffering normative descriptions of risk and courts should refuse to accept them. Whether a particular probability of risk is considered low, moderate, or high depends on no professional expertise, is not beyond the ken of the average layperson, and may tend to carry unjustified weight when propounded by an expert.

Moreover, fact finders have no means to ascertain whether two different FMHPs' pronouncements that an examinee is high risk are equivalent (recalling that the Static-99R authors consider 14% risk to be high while the same probability was deemed low risk by the MnSOST-R authors). Were FMHPs to rely solely on reporting the APER associated with their actuarially reported results, fact finders may find it substantially more difficult to classify examinees as "highly likely to reoffend," as most examinees would be expected to obtain APERs well below the likelihood standard.

Moreover, FMHPs cannot ethically defend continued use of the MnSOST-R in New Jersey. The MnSOST-R's experience table is based on a very small ( $n = 256$ ), nonrandom sample that is nearly two decades old. The MnSOST-R's five- to six-year recidivism rate is nearly three times higher than the highest 10-year rate reported for any New Jersey sample (12.7%; Zgoba et al., 2003). Several recent studies have also cast doubt on the validity of the MnSOST-R for estimating recidivism in U.S. samples (Bartosh, Garby, Lewis, & Gray, 2003; Boccaccini et al., 2009). Moreover, the discriminative properties of the MnSOST-R may not be stable across populations (Vrieze & Grove, 2008). With much larger, more contemporary data sets, the Static-99R and Static-2002R are likely to produce more reliable and valid APERs than the MnSOST-R. Finally, the MnSOST-R does not appear to add any incremental validity to risk estimates over that afforded by the Static-99R (Boccaccini et al., 2009; Seto, 2005).

With regard to the clinically adjusted actuarial method, what steps might FMHPs take to increase the validity of adjusting actuarially based APERs (while also overcoming the objections of Campbell & DeClue, 2010b, regarding the validity of such adjustments)? First, in keeping with *APA Ethical Principles*, Standard 9.06 exhorting FMHPs to ". . . indicate any significant limitations of their interpretations," FMHPs need to communicate to fact finders the reality that, except in rare circumstances (e.g., an examinee expressing an inability to control his sexually deviant urges and behaviors), most of the time the APERs and CIs yielded by either the Static-99R or the Static-2002R will fail to exceed a more-likely-than-not threshold.

Second, where FMHPs adjust their actuarially derived risk estimates based on other factors, both FMHPs and legal professionals should ensure that any such factors have been empirically validated as correlating with sex-offense recidivism. Legal professionals should inquire about, and FMHPs should discuss the incremental validity of any such factors.

Fact finders may choose to disregard validity concerns and perform their own idiosyncratic analysis to satisfy themselves that the likelihood standard has been met, but they should not be able to ground that determination on inaccurate estimates and interpretations provided by the FMHP. Most importantly, FMHPs should not encourage such ad hoc analysis by themselves combing through criminal histories, treatment records, or a clinical interview in an effort to identify recidivism-related factors that have neither been operationally defined nor empirically linked with recidivism.

As noted by Faust (2005) when discussing FMHPs' use of clinically derived factors to predict behavior, "mistakenly including a weak or invalid variable in the prediction mix usually does considerably more harm than mistakenly overlooking or disregarding a good predictor" (p. 26). Best practice demands a conservative approach where FMHPs rely on a few, well-validated, nonredundant factors (Faust).

Finally, FMHPs must attend to local base-rate data, when available, and explain their significance to fact finders. As Boccaccini et al. (2009) showed, FMHPs cannot assume that published experience tables of actuarial scales will generalize to local populations (see also, Mossman, 2006). Indeed, in the Texas sample that Boccaccini examined, the rank order of recidivism risk did not follow the progression of raw scores (1 to 6+) in the original Static-99 table (rank order raw scores of the Texas sample, from lowest to highest risk, reported as 2, 4, 3, 1, 5 and 6+; thus, those with the lowest Static-99 score [raw score = 1] were the third most likely to reoffend). Local recidivism base rates matter (Boccaccini et al., 2009; Helmus et al., 2009; Mossman, 2006; Wollert, 2006) and failure to take them into account violates ethical standards of practice (*APA Ethical Principles*, Standard 9.02; *Forensic Specialty Guidelines*, Standards IV.A.4, VI.A [4.05, 12.02]; *Testing Standards*, Standards 1.1, 1.4, 1.5, 2.1, 2.2, 11.2, 12.6, 12.13; *APA Psychiatric Ethics*, Section 5).

Thus, testifying experts need to be explicit as to the local base rate of sex reoffense recidivism they presumed when conducting their risk analysis. If experts are not forthcoming about base-rate assumptions, counsel and/or the court should demand that they specify the base-rate presumption they employed and the basis for its selection. If the base rates that opposing experts relied on differ significantly, the court should make a factual finding as to the proper base rate based on the best available evidence. If necessary, both experts should then adjust their risk estimates based on this adjudicated base rate, adding any caveats as to validity of the new estimates as they deem appropriate.

If the reported appellate court decisions accurately reflect how SORA is conducted and presented to fact finders in New Jersey, the above analysis suggests strongly that SORAs previously conducted in New Jersey SVP proceedings using the MnSOST-R and the original Static-99 samples, likely yielded significant overestimates of reoffense risk for substantial numbers of SVP candidates over the past decade.<sup>30,31</sup> With the addi-

---

<sup>30</sup> A score of 6+, which the Static group authors established as the lower threshold for the label "high" risk, yields a 10-year recidivism risk of 16.7% on the revised, routine sample of the Static-99 (Helmus et al.,

tional threat posed by overreliance on unvalidated clinical factors, there is a substantial risk that large numbers of individuals who would not have reoffended if released have been erroneously civilly committed as SVPs (even if all these individuals satisfied the mental abnormality standard, the SVP prongs are conjunctive; failure to satisfy the likelihood standard precludes classifying an individual as an SVP regardless of the mental-abnormality prong). This possibility should not be ignored and new SORAs should be conducted based on the Static-99R and Static-2002R as may be ethically adjusted by a predetermined, limited list of factors, if and when research demonstrates such factors add incremental validity to risk estimation.

Some will complain that this analysis has left FMHPs with virtually no tools with which to assist the courts to make SVP determinations. To argue that FMHPs should render forensic opinions, even in the absence of supporting empirical evidence, because doing so advances some desired societal policy, represents a form of the “utility” argument espoused by pragmatic psychology (Fishman, 2003).

However, utility stripped of empirical verity can quickly degrade into tyranny; and just because someone asks FMHPs for an opinion doesn’t mean they have to offer one, especially if they cannot ethically and empirically defend its assertions and conclusions. As Vrieze and Grove (2008, p. 275) point out, taking SVP “cases is hardly compulsory.”

Substantial political and social forces emphasize and prioritize reduction of risk of false negatives over reduction of risk of false positives (Janus & Nudell, 1999; Prentky et al., 2006). A commitment to empirically validated methods may inoculate forensic psychologists and psychiatrists against such pressures, thus preserving the integrity of the assessment process and professional practice. Conversely, failure to resist these forces threatens to undermine the principles of both the law and science (Prentky et al., 2006).

By admitting the limits of their knowledge, forensic experts—as scientists first—acknowledge that an empirically grounded extension of their reach must precede the exercise of their grasp. Without such humble clarity, science, the law, and ultimately society, will suffer.

---

2009), compared to the 45% risk estimate reported for a 6+ on the original sample (Hanson & Thornton, 1999).

<sup>31</sup> As of 2006, New Jersey's population of civilly committed SVPs totalled 350. Nationally, 2,627 individuals were civilly committed as SVPs that year. Thus, approximately 12% of all SVPs in the twenty jurisdictions with SVP laws, resided in New Jersey. By 2006, Pennsylvania, by comparison, had only civilly committed 8 residents as SVPs (Deming, 2008).

## References

*Addington v. Texas*, 441 U.S. 327 (1979).

American Academy of Psychiatry and the Law. (2005). *American Academy of Psychiatry and the Law ethics guidelines for the practice of forensic psychiatry*. Bloomfield, CT: Author.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

American Psychiatric Association. (2006). *The principles of medical ethics with annotations especially applicable to psychiatry - 2006 edition*. Arlington, VA: Author.

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

Aron, A., & Aron, E. N. (1999). *Statistics for psychology* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, 47, 422-438.

Berlin, F. S., Galbreath, N. W., Geary, B., & McGlone, G. (2003). The use of actuarials at civil commitment hearings to predict the likelihood of future sexual violence. *Sexual Abuse: A Journal of Research and Treatment*, 15, 377-382.

Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy, and Law*, 15, 278-314.

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1998). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. Lutz, FL: Psychological Assessment Resources.

*Bourjaity v. United States*, 483 U.S. 171 (1987).

Campbell, T. W. (2000). Sexual predator evaluations and phrenology: Considering issues of evidentiary reliability. *Behavioral Sciences and the Law*, 18, 111-130.

Campbell, T. (2007). When prophecy fails: Retreating from prediction. *Journal of Sexual Offender Civil Commitment: Science and the Law*, 2, 1-11.

- Campbell, T. W., & DeClue, G. (2010a). Maximizing predictive accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2*, 148-232.
- Campbell, T. W., & DeClue, G. (2010b). Flying blind with naked factors: Problems and pitfalls in adjusted actuarial sex offender risk assessment. *Open Access Journal of Forensic Psychology, 2*, 75-101.
- Committee on Ethical Guidelines for Forensic Psychologists. (1991). Forensic specialty guidelines for forensic psychologists. *Law and Human Behavior, 15*, 655-665.
- Committee on the Revision of the Specialty Guidelines for Forensic Psychology. (2009). *Specialty guidelines for forensic psychology* (4th draft). Washington, DC: Author. Retrieved from <http://www.ap-ls.org/links/professionalsgfp.html>
- Dallas v. Burlington N., Inc.*, 212 Mont. 514, 689 P.2d 273 (1984).
- Daubert v. Merrell Dow Pharmaceutical, Inc.*, 509 U.S. 579 (1993).
- DeClue, G., & Campbell, T. W. (2010). Still maximizing accuracy in sexually violent predator evaluations. *Open Access Journal of Forensic Psychology, 2*, 322-336.
- Deming, A. (2008). Sex offender civil commitment programs, current practices, characteristics and resident demographics. *Journal of Psychiatry & Law, 36*, 438-450.
- Donaldson, T., & Wollert, R. (2008). A mathematical proof and example that Bayes's Theorem is fundamental to actuarial estimates of sexual recidivism risk. *Sexual Abuse: A Journal of Research and Treatment, 20*, 206-217.
- Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitment and beyond*. Thousand Oaks, CA: Sage.
- Doren, D. M. (2006). Inaccurate arguments in sex offender civil commitment proceedings. In A. Schlank (Ed.). *The sexual predators' law and public policy—clinical treatment* (Vol. III). Kingston, NJ: Civic Research Institute.
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383.
- Epperson, D., Kaul, J., & Hesselton, D. (1999). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R): Development, performance, and recommended risk level cut scores*. Unpublished Manuscript.

- Fabian, J. M. (2005). The risky business of conducting risk assessments for those already civilly committed as sexually violent predators. *William Mitchell Law Review*, 32, 81-159.
- Faigman, D. L. (1999). *Legal alchemy: The use and misuse of science in the law*. New York, NY: W. H. Freeman & Co.
- Faust, D. (2005). Increasing the accuracy of clinical judgment (and thereby treatment effectiveness). In G. P. Koocher, J. C. Norcross, & S. S. Hill III (Eds.). *Psychologists' desk reference* (2nd ed., pp. 23-27). New York, NY: Oxford University Press.
- Fishman, D. B. (2003). Background on the "Psycholegal Lexis Proposal": Exploring the potential of a systematic case study database in forensic psychology. *Psychology, Public Policy, and Law*, 9, 267-274.
- Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).
- Gallagher, M. P. (2009, November 2). Sex offenders sue over state refusal to allow set-up of legal-help group. *New Jersey Law Journal*, 198, 468.
- Garb, H. N. (1998). *Study the clinician: Judgment, research and psychological assessment*. Washington, DC: American Psychological Association.
- Garner, B. A. (2004). *Black's law dictionary* (8th ed.). St. Paul, MN: Thomson, West.
- Grove, W. M., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19-30.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sex offense recidivism* (User Report 1997-04). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.
- Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law*, 4, 50-72.
- Hanson, R. K. (2005). *The validity of the Static-99 with older sexual offenders*. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada. Retrieved October 1, 2005 from <http://www.sppc-psepc.gc.ca/publications/corrections/20050630>

- Hanson, R. K., & Bussiere, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior, 27*, 6-35.
- Hanson, R. K., & Morton-Bourgon, K. E. (2004). *Predictions of sexual recidivism: An updated meta-analysis* (User Report 2004-02). Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sex offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-63.
- Hanson, R. K., & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis* (Cat. No. PS4-36/2007E). Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 1999-02). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.
- Harcourt, B. E. (2004). *Against prediction: Sentencing, policing, and punishing in an actuarial age*. Unpublished manuscript.
- Heilbrun, K., DeMatteo, D., & Marczyk, G. (2004). Pragmatic psychology, forensic mental health assessment, and the case of Thomas Johnson. *Psychology, Public Policy, and Law, 10*, 31-70.
- Helmus, L., Hanson, R. K., & Thornton, D. (2009). Reporting Static-99 in light of new research on recidivism norms. *The Forum, 21(1)*, 38-45.
- IMO the Civil Commitment of A.E.F.*, 377 N.J. Super. 473, 873 A.2d 604 (App. Div. 2005).
- IMO the Civil Commitment of A.H.B.*, 386 N.J. Super. 16, 898 A.2d 1027 (App. Div. 2006).
- IMO the Civil Commitment of E.S.T.*, 371 N.J. Super. 562, 854 A.2d 936 (App. Div. 2004).

*IMO the Civil Commitment of G.G.N.*, 372 N.J. Super. 42, 855 A.2d 569 (App. Div. 2004).

*IMO the Civil Commitment of J.M.B.*, 197 N.J. 563, 964 A.2d 752 (2009).

*IMO the Civil Commitment of J.P.*, 339 N.J. Super. 443, 772 A.2d 54 (App. Div. 2001).

*IMO the Civil Commitment of J.P.*, 393 N.J. Super. 7, 922 A.2d 754 (App. Div. 2007).

*IMO the Civil Commitment of M.L.V.*, 388 N.J. Super. 454, 909 A.2d 286 (App. Div. 2006).

*IMO the Civil Commitment of P.Z.H.*, 377 N.J. Super. 458, 873 A.2d 595 (App. Div. 2005).

*IMO the Civil Commitment of R.S.*, 339 N.J. Super. 507, 773 A.2d 72 (App. Div. 2001).

*IMO the Civil Commitment of T.J.N.*, 390 N.J. Super. 218, 915 A.2d 53 (App. Div. 2007).

*IMO the Civil Commitment of W.Z.*, 339 N.J. Super. 549, 773 A.2d 97 (App. Div. 2001).

*IMO the Civil Commitment of W.Z.*, 173 N.J. 109, 801 A.2d 205 (2002).

*In re Paoli R.R. Yard PCB Litigation*, 35 F.3d 717 (3d Cir. 1994).

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse*, 19, 425-448.

Janus, E. S., & Meehl, P. E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy, and Law*, 3, 33-64.

Janus, E. S., & Nudell, L. J. (1999). Defending sex offender commitment cases. In A. Schlink & F. Cohen (Eds.). *The sexual predator: Law, policy, evaluation and treatment* (pp. 3.2 - 3.29). Kingston, NJ: Civic Research Institute.

Johnston, R. (2003). Reducing analytic error: Integrating methodologies into teams of substantive experts. *Studies in Intelligence*, 47, 1.

*Kansas v. Crane*, 534 U.S. 407 (2002).

*Kansas v. Hendricks*, 521 U.S. 346 (1997).

LaFond, J. Q. (1999). Can therapeutic jurisprudence be normatively neutral? Sexual predator laws: Their impact on participants and policy. *Arizona Law Review*, 41, 375-415.

- Langan, P. A., Schmitt, E. L., & Durose, M. R. (2003). *Recidivism of sex offenders released from prison in 1994* (NCJ 198281). Washington, DC: United States Department of Justice, Bureau of Justice Statistics.
- Levenson, J. S. (2004). Reliability of sexually violent predator civil commitment criteria. *Law and Human Behavior, 28*, 357-369.
- Mauro v. Raymark Industries*, 116 N.J. 126, 561 A.2d 257 (1989).
- Miller, H., Amenta, A., & Conroy, M. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior, 29*, 29-54.
- Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior, 23*(1), 107-120.
- Monahan, J. (2006). A jurisprudence of risk assessment forecasting harm among prisoners, predators, and patients. *Virginia Law Review, 92*, 391-435.
- Moore v. Ashland Chemical, Inc.*, 151 F.3d 269 (6th Cir. 1998).
- Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse, 18*, 41-63.
- Mossman, D. (2008). Analyzing the performance of risk assessment instruments: A response to Vrieze & Grove (2007). *Law and Human Behavior, 32*, 279-291.
- Phenix, A., Helmus, L., & Hanson, R. K. (2009, September 28). *Static-99R & Static-2002R: Evaluator's workbook*. Retrieved October 18, 2009 from <http://www.static99.org>
- Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B. K., & Kafka, M. P. (2006). Sexually violent predators in the courtroom: Science on trial. *Psychology, Public Policy, and Law, 12*, 357-393.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent Offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior, 21*, 231-241.
- 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.

*Rubanick v. Witco Chemical Corp.*, 242 N.J. Super. 36, 576 A.2d 4 (App. Div. 1990).

*Rutigliano v. Valley*, 929 F. Supp. 779 (D.N.H. 1996).

Sager, W. (2000, November 10). *Report - 1994 release recidivism study*. Avenel, NJ: Adult Diagnostic and Treatment Center.

Sager, W. (2001, October 10). *Report - 1995 release recidivism study*. Avenel, NJ: Adult Diagnostic and Treatment Center.

Sager, W. (2002, January 21). *Report - 1990 release recidivism study*. Avenel, NJ: Adult Diagnostic and Treatment Center.

Schopp, R. F., Scalora, M. J., & Pearce, M. (1999). Expert testimony and professional judgment: Psychological expertise and commitment as a sexual predator after Hendricks. *Psychology, Public Policy, and Law*, 5, 120-174.

Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, 17, 156-167.

Sexually Violent Predators Act, *N.J. Stat. Ann.* 30:4-27.24 *et seq.* (1998).

Song, L., & Lieb, R. (1994). *Recidivism data concerning Washington State sex offenders*. Unpublished raw data. Olympia, WA: Washington Institute for Public Policy.

Sreenivasan, S., Weinberger, L. E., & Garrick, T. (2003). Expert testimony in sexually violent predator commitments: Conceptualizing legal standards of "mental disorder" and "likely to reoffend." *Journal of the American Academy of Psychiatry and the Law*, 31, 471-85.

*State of New Hampshire v. Ploof*, No. 07-E-0238 (N.D. N.H. Super., Apr. 28, 2009). Retrieved October 18, 2009 from <http://www.static99.org/norms>.

*State of New Jersey v. Cavallo*, 88, N. J. 508, 443, A.2d 1020 (1982),

*State of New Jersey v. Fortin*, 178 N.J. 540, 843 A.2d 974 (2004).

*State of New Jersey v. Harvey*, 151 N.J. 117, 699 A.2d 596 (1997).

*State of New Jersey v. Raso*, 321 N.J. Super 5, 728 A.2d 231 (App. Div. 1999).

*State of New Jersey v. Spann*, 130 N.J. 484, 617 A.2d 247 (1993).

*U.S. v. Shields*, No. 07-12056-PBS, 2008 WL 544940, at \*1 (D. Mass Feb. 26, 2008).

- Vrieze, S. I., & Grove, W. M. (2008). Predicting sex offender recidivism. I. Correcting for item overselection and accuracy overestimation in scale development. II. Sampling error-induced attenuation of predictive validity over base rate information. *Law and Human Behavior, 32*, 266-278.
- Wecht, C. H. (Ed.). (2008). *Forensic sciences*. Newark, NJ: Matthew Bender Co., Inc.
- Wollert, R. (2004, January 16). *An analysis of actuarial risk assessment*. Paper presented at the SVP Training Program, Florida Public Defenders Association, Orlando, FL.
- Wollert, R. (2006). Low base rates limit expert certainty when current actuarials are used to identify sexually violent predators: An application of Bayes's Theorem. *Psychology, Public Policy, and Law, 12*, 56-85.
- 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.
- Wollert, R., & Waggoner, J. (2009). Bayesian computations protect sexually violent predator evaluations from the degrading effects of confirmatory bias and illusions of certainty: A reply to Doren and Levenson (2009). *Sexual Offender Treatment, 4*, 1-27.
- Woodworth, G. G., & Kadane, J. (2004). Expert testimony supporting post-sentence civil incarceration of violent sexual offenders. *Law, Probability, and Risk, 3*, 221-241.
- Zgoba, K. M., & Simon, L. N. J. (2005). Recidivism rates of sexual offenders up to 7 years later. Does treatment matter? *Criminal Justice Review, 30*, 155-173.
- Zgoba, K. M., Sager, W. R., & Witt, P. H. (2003). Evaluation of New Jersey's sex offender treatment program at the Adult Diagnostic and Treatment Center: Preliminary results. *Journal of Psychiatry, and Law, 31*, 133-164.
- Zgoba, K., Witt, P., Dalessandro, M., & Veysey, B. (2008). *Megan's Law: Assessing the practical and monetary efficacy* (NCJ No. 225370). Washington, DC: National Institute of Justice/NCJRS.
- Zonanna, H. (2000). Sex offender testimony: Junk science or unethical testimony? *The Journal of the American Academy of Psychiatry and the Law, 28*, 386-88.
- Submitted 2/18/11, revision received 8/20/11, accepted 10/4/11