

HEINONLINE

Citation: 56 Emory L.J. 275 2006-2007

Content downloaded/printed from
HeinOnline (<http://heinonline.org>)
Fri Jul 27 15:03:32 2012

- Your use of this HeinOnline PDF indicates your acceptance of HeinOnline's Terms and Conditions of the license agreement available at <http://heinonline.org/HOL/License>
- The search text of this PDF is generated from uncorrected OCR text.
- To obtain permission to use this article beyond the scope of your HeinOnline license, please use:

[https://www.copyright.com/ccc/basicSearch.do?
&operation=go&searchType=0
&lastSearch=simple&all=on&titleOrStdNo=0094-4076](https://www.copyright.com/ccc/basicSearch.do?&operation=go&searchType=0&lastSearch=simple&all=on&titleOrStdNo=0094-4076)



DiscoverArchive

Retrieved from DiscoverArchive,
Vanderbilt University's Institutional Repository

This work was originally published in
56 Emory L.J. 275 2006-2007

EMORY LAW JOURNAL

Volume 56

2006

Number 2

ARTICLES

DANGEROUSNESS AND EXPERTISE REDUX

Christopher Slobogin *

ABSTRACT

Civil commitment, confinement under sexual predator laws, and many capital and noncapital sentences depend upon proof of a propensity toward violence. This Article discusses the current state of prediction science, in particular the advantages and disadvantages of clinical and actuarial prediction, and then analyzes how the rules of evidence should be interpreted in deciding whether opinions about propensity should be admissible. It concludes that dangerousness predictions that are not based on empirically derived probability estimates should be excluded from the courtroom unless the defense decides otherwise. This conclusion is not bottomed on the usual concern courts and commentators raise about expert prediction testimony—that the associated false-positive rates are too high. In fact, because they produce better-than-chance results, both clinical and actuarial risk assessments are, with certain caveats, sufficiently probative to meet the expertise threshold, whether that threshold is defined by *Frye* or *Daubert*. Rather, the rationale for the conclusion reached in this Article is that prediction testimony should only be inadmissible when its prejudicial impact outweighs its (admittedly

* Stephen C. O'Connell Professor of Law, University of Florida Fredric G. Levin College of Law. The author would like to thank John Monahan, Randy Otto, Norman Poythress, and workshop participants at Hastings and Florida law schools for their comments on earlier drafts of this Article.

weak) probative value. Clinical prediction testimony should be excluded from the government's case-in-chief under this formula, because of its relative invulnerability to adversarial testing. But the government should be able to use actuarial prediction testimony, as empirical evidence suggests that this type of testimony is likely to be taken for what it is worth. Finally, because of judges and juries' demonstrably strong predilection toward findings of dangerousness in commitment and sentencing proceedings, the defense should always be able to use both clinical and actuarial testimony to rebut the state's case.

INTRODUCTION

Over two decades ago, I wrote an article entitled *Dangerousness and Expertise* that made a number of suggestions about how the rules of evidence should apply to expert testimony aimed at predicting future antisocial behavior.¹ This Article addresses the same issue and reaches the same basic conclusion: Expert prediction testimony based on empirically derived probability assessments usually should be admissible, whether proffered by the government or the defense, but clinical prediction testimony, which is still the predominant method of proving dangerousness, should be excluded unless the defense uses it in its case-in-chief. Despite the similarity in end result between the two articles, a number of intervening developments require a new analysis of the way courts make determinations about future violence.

First, as society grows increasingly frustrated with the traditional punishment system's failure to protect its citizens, dangerousness has become increasingly popular as a criterion for depriving people of liberty. The most obvious example of this phenomenon is the "sex predator law" movement. Since 1990, about one third of the states have enacted laws that permit indeterminate post-sentence commitment of sex offenders considered to be "predisposed" to violent behavior.² On authority of these statutes, well over 2,500 people have been committed as sexual predators and thousands of others

¹ Christopher Slobogin, *Dangerousness and Expertise*, 133 U. PA. L. REV. 97 (1984).

² See W. Lawrence Fitch & Debra A. Hammen, *The New Generation of Sex Offender Commitment Laws: Which States Have Them and How Do They Work?*, in PROTECTING SOCIETY FROM SEXUALLY DANGEROUS OFFENDERS: LAW, JUSTICE, AND THERAPY 27, 33 (Bruce J. Winick & John Q. LaFond eds., 2003).

have been subjected to sexual predator hearings.³ Even more expansive “dangerous offender” laws, which permit post-sentence confinement of *any* offender perceived to have violent tendencies, have also been enacted.⁴

Meanwhile, dangerousness determinations remain important in many other areas of the law. While criminal sentencing since 1970 has generally moved toward fixed terms based on backward-looking culpability assessments,⁵ recent Supreme Court decisions could push the states back toward indeterminate dispositions grounded on risk assessments,⁶ and several states continue to make the risk of future violent crime an aggravating circumstance in death penalty cases.⁷ Finally, the civil commitment laws of virtually every state permit involuntary hospitalization based on the likelihood of imminent harm to others,⁸ and similar provisions authorize commitment of those found not guilty by reason of insanity.⁹

A second reason why a fresh look at methods of proving dangerousness is needed is that prediction methodology has significantly improved. Two decades ago, actuarial models for predicting violent behavior were in their infancy and the dominant method of evaluating dangerousness was essentially seat-of-the-pants, “clinical” speculation. Today the development of several sophisticated, empirically validated risk assessment instruments has made prediction much more of a science.¹⁰

A third reason why the topic addressed in this Article is worth revisiting is that the law of evidence has changed significantly. Until 1993, most courts

³ W. Lawrence Fitch, *Sexual Offender Commitment in the United States: Legislative and Policy Concerns*, 989 ANNALS N.Y. ACAD. SCI. 489, 492 (2003) (reporting that, as of the summer of 2002, 2,478 individuals had been committed or were being held as sexually violent predators in the United States).

⁴ See, e.g., CAL. PENAL CODE § 2962 (West 2006).

⁵ 28 U.S.C. § 994(c) (2000) (stating that sentences should be based only on offense characteristics and deterrence); § 994(k) (rejecting rehabilitation as a goal of sentencing); 18 U.S.C. § 3624(a)–(b) (2000) (abolishing parole).

⁶ Susan R. Klein, *The Return of Federal Judicial Discretion in Criminal Sentencing*, 39 VAL. U. L. REV. 693, 726–27, 731–33 (2005) (describing the impact of the *Apprendi*, *Blakely*, and *Booker* cases—U.S. Supreme Court cases restructuring sentencing that were all decided since 2000—and predicting much more variation in sentencing based in part on predictive criteria).

⁷ At least six states have such provisions. Mitzi Dorland & Daniel Krauss, *The Danger of Dangerousness in Capital Sentencing: Exacerbating the Problem of Arbitrary and Capricious Decision Making*, 29 L. & PSYCHOL. REV. 63, 64 n.5 (2005).

⁸ RALPH REISNER ET AL., LAW AND THE MENTAL HEALTH SYSTEM: CIVIL AND CRIMINAL ASPECTS 670 (4th ed. 2004).

⁹ *Id.* at 830.

¹⁰ See *infra* Part II.

evaluated the admissibility of expert testimony using one of two tests: The “balancing test” (derived from the balancing required by Federal Rules of Evidence 401, 402 and 403),¹¹ which weighs the probative value of the evidence against the likelihood it will mislead or confuse the fact finder or unfairly prejudice one of the parties, and the *Frye* test, derived from *Frye v. United States*,¹² which looks at whether the basis of the testimony is “generally accepted” among professionals in the relevant field.¹³ Today, however, the federal courts and more than half the states follow the test announced in *Daubert v. Merrell Dow Pharmaceuticals*,¹⁴ which, at least on its face, is much tougher to meet. *Daubert* requires that expert testimony be validated or verified, preferably using scientific methodology that produces error rates rather than simply through assessment of whether the basis of the testimony is relevant or generally accepted.¹⁵

The reason *Daubert*, *Frye*, and other tests of expert admissibility are important in this context is that dangerousness assessments are often bottomed, if not wholly dependent, on the opinions of psychiatrists and psychologists. In some settings, particularly in connection with noncapital sentencing proceedings, dangerousness determinations may be based entirely on lay judgments, usually made by probation officers and judges. But in numerous other settings—in particular, capital sentencing, police power civil commitment proceedings, and sexual predator hearings—mental health professionals and other “experts” on dangerousness are almost always involved in assisting the decision makers who predict antisocial conduct.¹⁶

A number of commentators, writing both before and after *Daubert* raised the admissibility threshold, have asserted that this practice of relying on mental

¹¹ FED. R. EVID. 401–03 (Rule 401 defines relevance, Rule 402 states that “[a]ll relevant evidence is admissible, except as otherwise provided,” and Rule 403 states that “relevant, evidence may be excluded if its probative value is substantially outweighed” by the types of factors mentioned in the text.).

¹² 293 F. 1013 (D.C. Cir. 1923).

¹³ See Paul Giannelli, *Daubert in the States*, 34 CRIM. L. BULL. 154, 155–56 (1998), for a description of these tests and their history.

¹⁴ 509 U.S. 579 (1993). See David E. Bernstein & Jeffrey D. Jackson, *The Daubert Trilogy in the States*, 44 JURIMETRICS J. 351, 356–61 (2004), for a discussion of the number of states that have adopted *Daubert*.

¹⁵ *Daubert*, 509 U.S. at 592–94. In 1999, the Court made clear that this test governed the admissibility of “technical” and “specialized” knowledge, as well as “scientific” knowledge, preventing courts from avoiding *Daubert*’s dictates by classifying testimony as “non-scientific.” *Kumho Tire Co. Ltd. v. Carmichael*, 526 U.S. 137 (1999).

¹⁶ See GARY MELTON ET AL., *PSYCHOLOGICAL EVALUATIONS FOR THE COURTS: A HANDBOOK FOR MENTAL HEALTH PROFESSIONALS AND LAWYERS* 258–67, 323 (2d ed. 1997), for a discussion of the role of mental health professionals in commitment and sentencing proceedings.

health professionals to assist in making legally relevant predictions should stop or be severely circumscribed.¹⁷ The official organs of the mental health professions have long voiced similar views. Back in 1974, the American Psychiatric Association stated that “[p]sychiatric expertise in the prediction of ‘dangerousness’ is not established.”¹⁸ Four years later, the American Psychological Association came to much the same conclusion, asserting that “the validity of psychological predictions of violent behavior, at least in . . . sentencing and release situations . . . , is extremely poor, so poor that one could oppose their use on the strictly empirical grounds that psychologists are not professionally competent to make such judgments.”¹⁹ Today, despite improvements in prediction science, forensic professionals are still cautious about endorsing opinion testimony concerning violent propensities.²⁰

The courts, however, have paid no attention to such assertions. In 1983, despite amicus briefs incorporating statements like those quoted above, as well as John Monahan’s 1981 conclusion, based on a meta-review of the studies, that two out of three long-term predictions of violence are wrong, the Supreme Court upheld the admissibility of prediction testimony from mental health professionals in *death penalty* proceedings.²¹ That holding, in *Barefoot v. Estelle*,²² still stands. If executions may be predicated on prediction testimony, presumably sex offenders and mentally ill people subject to commitment may be confined based on such opinion evidence as well.

¹⁷ Bruce J. Ennis & Thomas R. Litwack, *Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom*, 62 CAL. L. REV. 693, 701–08 (1974) (arguing that expert prediction testimony should be banned); Stephen J. Morse, *Crazy Behavior, Morals, and Science: An Analysis of Mental Health Law*, 51 S. CAL. L. REV. 527, 621–22 (1978) (stating that only “hard” actuarial prediction should be permitted); David Faust & Jay Ziskin, *The Expert Witness in Psychology and Psychiatry*, 241 SCI. 31, 33–35 (1988) (only actuarial data should be admissible); Daniel W. Shuman & Bruce D. Sales, *The Admissibility of Expert Testimony Based upon Clinical Judgment and Scientific Research*, 4 PSYCHOL. PUB. POL’Y & L. 1226, 1250 (1998) (same); Erica Beecher-Monas & Edgar Garcia-Rill, *Danger at the Edge of Chaos: Predicting Violent Behavior in a Post-Daubert World*, 24 CARDOZO L. REV. 1845, 1848 (2003) (same, but advocating that actuarial devices be “used with caution”).

¹⁸ AM. PSYCHIATRIC ASS’N TASK FORCE, CLINICAL ASPECTS OF THE VIOLENT INDIVIDUAL 33 (1974).

¹⁹ *Report of the Task Force on the Role of Psychology in the Criminal Justice System*, 33 AM. PSYCHOLOGIST 1099, 1110 (1978).

²⁰ RICHARD ROGERS & DANIEL W. SHUMAN, FUNDAMENTALS OF FORENSIC PRACTICE: MENTAL HEALTH AND CRIMINAL LAW 298, 306 (2005) (noting “the marked limitations of risk assessment in predicting recidivism for sex offenders” and the “significant errors” produced by modern risk assessment measures).

²¹ See *Barefoot v. Estelle*, 463 U.S. 880, 899 n.7 (1983) (describing APA amicus brief and Monahan’s conclusion).

²² *Id.* at 880.

Barefoot, however, only announced the constitutional minimum; evidentiary rules can demand more of experts, as *Frye* and *Daubert* do. Using a four-part analytical framework derived from traditional evidence law,²³ I take an intermediate position, as I did in my first article on the subject, between complete exclusion and unquestioned acceptance of expert testimony on dangerousness. Traditional clinical testimony regarding a person's future behavior, while material, probative, and, on its face, helpful, is so prejudicial (and thus ultimately unhelpful) that in most settings it should be admissible only if the person first seeks to use clinical testimony to show that he or she is not dangerous. If the individual eschews such testimony, then the state should be limited to proving dangerousness using appropriately normed actuarial instruments or structured interview instruments that are tied to explicit probability estimates.

This Article develops this argument in eight parts. Part I describes the current state of prediction science. It explains the difference between the three primary prediction methodologies—clinical, actuarial, and structured professional judgment—and then evaluates their relative reliability. Part II examines the relevant evidentiary case law, which for the most part has not differentiated between prediction methodologies, and in any event has usually permitted any type of prediction testimony, regardless of its basis. Part III begins my own evidentiary analysis with an assessment of the materiality of prediction testimony, in particular the extent to which materiality is undermined by nomothetic (group-based) prediction or prediction based on characteristics over which the person has little or no control. Parts IV and V of the Article argue that prediction testimony is also usually probative and helpful, despite its relatively low accuracy rate. It is probative whenever it is derived from a methodology that produces predictions that are better than chance, and it is helpful whenever it is based on the literature about violence risk and avoids ultimate issue language.

Those conclusions establish a presumption in favor of admitting prediction testimony. However, the fourth component of traditional evidentiary analysis—which looks at potential prejudicial impact—requires significant limitations on clinical, as opposed to actuarial, testimony, at least outside the civil commitment context. Part VI asserts that when offered by the government in its case-in-chief, testimony that is not linked to specific risk

²³ For a fuller description of this framework, see Christopher Slobogin, *Psychiatric Evidence in Criminal Trials: To Junk or Not to Junk?*, 40 WM. & MARY L. REV. 1, 29–30 (1998).

probabilities is too prone to misinterpretation and misuse, and thus whatever probative value it has is substantially outweighed by its prejudicial impact. In contrast, when proffered by the subject of the prediction (the criminal offender or the respondent in a commitment proceeding), the prejudicial impact of probative prediction testimony is likely to be slight, given the natural inclination of fact finders to assume those subject to sentencing and commitment are dangerous. Thus, the subject should be permitted to introduce clinical prediction testimony whenever he or she sees fit. Only if the subject decides to do so should the government be able to use softer prediction testimony as a method of responding to the subject's claims.

Part VII of the Article spells out how this "subject-first" regime would work in practice. Where material, empirically derived probability estimates do not exist, the government would have to rely on lay testimony (presumably regarding past conduct), unless the subject of the prediction opens the door to use of clinical prediction testimony. That door would be opened when the offender-respondent uses clinical testimony to contradict claims that he or she is dangerous, volitionally impaired, or untreatable, but not when the defense expert focuses solely on culpability issues in a sentencing proceeding. This regime allows the government to prove dangerousness in the most accurate, least confounding manner, while permitting the offender-respondent to attack the state's attempt at preventive detention on the ground that the "numbers" do not accurately reflect his or her violence potential.

Part VIII of the Article briefly describes why the foregoing analysis may not be applicable to emergency civil commitment. Simply put, short-term clinical dangerousness predictions of the type at issue in civil commitment hearings are probably more accurate than the long-term predictions required in other settings. Even if they are not, practical reasons might dictate continued reliance on clinical prediction in this one setting.

As this summary indicates, this Article will focus, as my previous article did, on four types of proceedings: civil commitment, "criminal" commitment (commitment and release hearings of sexual predators and people acquitted by reason of insanity), noncapital sentencing hearings, and capital sentencing proceedings. Although much of what is said here could apply to other proceedings,²⁴ these four are most likely to involve expert assessments of dangerousness to others.

²⁴ "Shah has identified 11 other stages of the legal process at which dangerousness assessments are made, including pretrial release hearings, juvenile transfer decisions, and transfer of prisoners to special prisons for

Even limited to its use in these settings, “dangerousness” can take on a number of different meanings. For example, in capital sentencing, “the term usually refers solely to one’s propensity to cause serious bodily injury to another.”²⁵ In the civil or criminal commitment contexts, in contrast, it might also refer to the likelihood that a person will cause damage to property,²⁶ and can even encompass emotional harm.²⁷ Civil commitment focuses on imminent danger, while the other three contexts are more concerned with long-term dangerousness. Finally, the degree of risk that justifies intervention may well differ depending upon the type of intervention.²⁸ For instance, aggravators in death penalty cases usually must be proven beyond a reasonable doubt, whereas civil commitment is permissible if the state meets the lesser clear and convincing evidence standard.²⁹

I will emphasize these distinctions where appropriate. But for the most part I will not distinguish between them. Thus, I will usually use the term “dangerousness” without further explanation, trusting the reader to recognize that the term can vary in meaning depending on the context.

disruptive offenders.” *Id.* at 101 n.15 (citing Saleem A. Shah, *Dangerousness: A Paradigm for Exploring Some Issues in Law and Psychology*, 33 AM. PSYCHOLOGIST 224, 225 (1978)).

²⁵ *Id.* at 102; *see, e.g.*, IDAHO CODE ANN. § 19-2515(9)(i) (2006) (requiring “propensity to commit murder which will probably constitute a continuing threat to society”); TEX. CODE CRIM. PROC. ANN. art. 37.071 § 2(b)(1) (Vernon 2006) (requiring “probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society”); VA. CODE ANN. § 19.2-264.4C (2004) (same).

²⁶ *See, e.g.*, ARK. CODE ANN. § 5-2-315(a)(1) (2006) (release of insanity acquittees required if they “no longer create a substantial risk of bodily injury to another person or serious damage to the property of another”); DEL. CODE ANN. tit. 16, § 5001(6) (2003) (permitting civil commitment if individual poses a threat of substantial damage to property); HAW. REV. STAT. § 334-1 (2005) (same); KAN. STAT. ANN. § 59-2946(3)(a) (Supp. 2002) (same); N.D. CENT. CODE § 25-03.1-02(12)(b) (1987) (same). In *Jones v. United States*, 463 U.S. 354, 364 (1983), the Supreme Court stated, in the context of criminal commitment proceedings, “We do not agree with petitioner’s suggestion that the requisite dangerousness is not established by proof that a person committed a non-violent crime against property.” *Id.* at 364–65.

²⁷ *See, e.g.*, HAW. REV. STAT. § 334-1 (2005) (dangerousness to others defined as “likely to do substantial physical or emotional injury on another”); IOWA CODE ANN. § 229.1(16)(b) (2006) (dangerousness to others defined as “serious emotional injury on members of the person’s family or others who lack reasonable opportunity to avoid contact with the person”).

²⁸ Jay Katz & Alan Goldstein, *Dangerousness and Mental Illness*, 131 J. NERVOUS & MENTAL DISEASES 404, 410–11 (1960).

²⁹ *Compare* IDAHO CODE ANN. § 19-2515(9) (2006) (requiring proof beyond a reasonable doubt for all aggravating factors in capital cases), *with* *Addington v. Texas*, 441 U.S. 418 (1979) (permitting civil commitment by clear and convincing evidence).

I. THE SCIENCE OF PREDICTION

Mental health professionals use a number of different prediction techniques. Research evaluating the accuracy of these techniques suggests that clinicians are not as poor at prognostication as the critics suggest, although it also leaves no doubt that expert prediction is far from perfect. The discussion below describes the various prediction methodologies and their strengths and flaws, including their relative accuracy.

A. Prediction Methodologies

Until the late 1980s, almost all expert testimony regarding dangerousness was *clinical* in nature. That is, it relied on whatever information the individual clinician deemed pertinent. In the past twenty years, more structured approaches to prediction have been developed. An *actuarial* approach relies, like insurance actuaries do, on a finite number of pre-identified variables that statistically correlate to risk and that produce a definitive probability or probability range of risk. A third, hybrid prediction method, known as *adjusted actuarial* assessment, begins with an actuarial assessment that the professional then adjusts, based on individualized factors not considered in the actuarial formula. A final prediction methodology—called *structured professional judgment* or guided clinical assessment—also relies, like actuarial prediction, on an evaluation of a finite number of predefined factors that have been associated with risk, but neither the factors nor the final conclusions about risk are mathematically obtained.³⁰

An exemplary illustration of the clinical approach to prediction, back in the days when it was king, is provided by the work of Dr. Harry Kozol and his associates in 1972. Every prediction made by Kozol and his colleagues required “independent examinations by at least five clinicians, a battery of psychological tests, and ‘a meticulous reconstruction of the life history [of the subject] elicited from multiple sources.’”³¹ A second example of clinical prediction—but one that sits at the other end of the spectrum—comes from the case of *Estelle v. Smith*, in which James Grigson, a psychiatrist in Texas,

³⁰ For descriptions of these techniques, see Kevin S. Douglas & Jennifer L. Skeem, *Violence Risk Assessment: Getting Specific About Being Dynamic*, 11 PSYCHOL. PUB. POL’Y & L. 347, 352–53 (2005); R. Karl Hanson, *What Do We Know About Sex Offender Risk Assessment?*, 4 PSYCHOL. PUB. POL’Y & L. 50, 52–53 (1998).

³¹ Harry L. Kozol et al., *The Diagnosis and Treatment of Dangerousness*, 18 CRIME & DELINQ. 371, 383 (1972), quoted in Slobogin, *supra* note 1, at 111.

purported to be able to deliver an opinion on competency to stand trial, insanity, and dangerousness based solely on a ninety minute interview with Smith, the defendant.³²

The types of information underlying clinical predictions vary immensely, from examiner to examiner and case to case. Of particular importance to Kozol and his associates were “details in the description of the [triggering crime].”³³ In addition, they attempted to ascertain the extent to which the individual

harbors anger, hostility, and resentment; enjoys witnessing or inflicting suffering; lacks altruistic and compassionate concern for others; sees himself as a victim rather than as an aggressor; resents or rejects authority; is primarily concerned with his own satisfaction and with the relief of his own discomfort; is intolerant of frustration or delay of satisfaction; lacks control of his own impulses; has immature attitudes toward social responsibility; lacks insight into his own psychological structure; and distorts his perception of reality in accordance with his own wishes and needs.³⁴

This detailed focus on themes or commonalities in the examinee’s behavior in order to identify risk or protective factors is sometimes called anamnestic assessment.³⁵ Dr. Grigson again provides a contrast. In *Barefoot v. Estelle*, another case in which he was the prosecution’s key witness, he stated there was a “one hundred percent and absolute” chance that Barefoot would commit acts of criminal violence, despite never having interviewed him.³⁶ Rather, Grigson’s testimony rested entirely on a hypothetical question describing Barefoot’s four nonviolent offenses, his arrest on charges of statutory rape and unlawful restraint of a child, his escape from prison after that arrest, and the events surrounding the capital murder,³⁷ which admittedly were horrific.³⁸

Actuarial prediction is much more structured than clinical prediction. Probably the most prominent actuarial prediction device is the Violence Risk

³² *Estelle v. Smith*, 451 U.S. 454, 457–60 (1981).

³³ Harry Kozol, *The Diagnosis of Dangerousness*, in *VIOLENCE AND VICTIMS* 3, 8 (Stefan A. Pasternack ed., 1975).

³⁴ Kozol et al., *supra* note 31, at 379.

³⁵ MELTON ET AL., *supra* note 16, at 284–85.

³⁶ *Barefoot v. Estelle*, 463 U.S. 880, 917–19 (1983) (Blackmun, J., dissenting).

³⁷ *Id.* at 917–19.

³⁸ Barefoot shot at point-blank range a police officer who wanted to ask him questions about a fire (that Barefoot apparently had started to divert attention from a planned robbery). *Barefoot v. State*, 596 S.W.2d 875, 878–79 (Tex. Crim. App. 1980).

Appraisal Guide (VRAG).³⁹ The VRAG focuses on twelve empirically derived and relatively narrow variables—no other information is considered in making the assessment—and its goal is to produce a score that indicates a particular probability of recidivism.⁴⁰ For example, a score of 0–6 on the VRAG is associated with at least a 35% probability of “violent recidivism” within seven years, while a score of 21–27 is associated with a 76% risk of such recidivism over that period.⁴¹ A key variable on the VRAG is the individual’s score on the Psychopathy Checklist-Revised (PCL-R) (described in more detail below); all by itself, a rating of 35 or more on this measure results in 12 points on the VRAG, which indicates a recidivism risk of 44% (whereas, for example, a rating of 0–4 on the PCL-R results in -5 points on the VRAG, associated with a 17% risk).⁴² The eleven other variables in the VRAG focus on past misconduct, psychiatric diagnosis, substance abuse, age at the time of the offense, parental presence, and the gender of and harm to the victim.⁴³ Again, the points received on each variable are added and the composite score is then used to ascertain a recidivism risk, based on study of the criminal histories of others with similar scores.⁴⁴

Another actuarial approach, initially known as the multiple Iterative Classification Tree (ICT) and now called Classification of Violence Risk (COVR), requires the examiner to analyze risk using what amounts to risk flow charts.⁴⁵ For instance, under one classification tree, the evaluator first determines whether the person demonstrates low or high psychopathy. If psychopathy is low, the evaluator then determines whether the individual has been arrested only a “few” times or “many” times; if the former, the evaluator then looks at recent violence, if the latter, the examiner ascertains whether the

³⁹ The best description of the VRAG and pertinent research is found in VERNON L. QUINSEY, *VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK* (2d ed. 2006). For more information, see Grant T. Harris et al., *Violent Recidivism of Mentally Disordered Offenders: The Development of a Statistical Prediction Instrument*, 20 CRIM. JUST. & BEHAV. 315 (1993).

⁴⁰ See QUINSEY, *supra* note 39, at 160–63; Harris et al., *supra* note 39, at 378.

⁴¹ A more recent prospective study found that over a five-year period these two categories were associated with a 42% and 71% chance of reoffending, respectively. Grant T. Harris et al., *Prospective Replication of the “Violence Risk Appraisal Guide” in Predicting Violent Recidivism Among Forensic Patients*, 26 L. & HUM. BEHAV. 377, 385 (2002).

⁴² Harris et al., *supra* note 41, at 380.

⁴³ *Id.* at 378.

⁴⁴ *Id.*

⁴⁵ A good description of the conceptual work for this instrument is found in JOHN MONAHAN ET AL., *RETHINKING RISK ASSESSMENT: THE MACARTHUR STUDY OF MENTAL DISORDER AND VIOLENCE* 133–34 (2001); see also Henry J. Steadman et al., *A Classification Tree Approach to the Development of Actuarial Violence Risk Assessment Tools*, 24 L. & HUM. BEHAV. 83 (2000).

individual abuses drugs or alcohol. If the individual instead demonstrates strong psychopathic tendencies, the evaluator first determines whether the individual suffered serious abuse as a child; if so, then inquiry into substance abuse occurs.⁴⁶ At each step, a particular answer is associated with a particular recidivism probability.

Other actuarial devices focus specifically on sex offenders. For instance, the Minnesota Sex Offender Screening Tool relies on sixteen variables (for front-end commitments) or twenty-one variables (if the offender is being considered for release).⁴⁷ In addition to scoring points for convictions, arrests, and other items similar to those on the VRAG, this instrument requires the examiner to look at the age of the victim, the length of sexual offending history, and employment history.⁴⁸ Living up to its name, the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR) relies on only four variables (prior sex offenses, age at release, victim gender, and relationship to victim).⁴⁹ Each of these instruments also produces a score that is associated with a particular risk of recidivism.

In an adjusted actuarial approach, the examiner would use one of these devices as a baseline but then raise or lower the risk prediction based on other considerations, typically those known to relate to offending. For instance, an examiner administering the VRAG might lower the probability of recidivism if he or she knows the offender has done well in treatment or is about to get married. Or the risk level might be raised if the individual has made threats or has stated that, if released, he will join a gang.⁵⁰

One of the best examples of the final methodology—structured professional judgment—is the HCR-20 violence risk assessment scheme, which consists of 20 items relating to three categories of information:

⁴⁶ For a basic description, see Steadman et al., *supra* note 45, at 84.

⁴⁷ David L. Epperson et al., *Minnesota Sex Offender Screening Tool—Revised (MnSOST-R): Development, Validation, and Recommended Risk Level Cut Scores*, (Dec. 2003), <http://www.psychology.iastate.edu/faculty/epperson/TechUpdatePaper12-03.pdf>); *see also* Richard Hamill, *Recidivism of Sex Offenders: What You Need to Know*, 15 CRIM. JUST. 24, 30 (2001).

⁴⁸ Epperson et al., *supra* note 47, at 13.

⁴⁹ *See* R. KARL HANSON, PUB. WORKS & GOV'T SERVS. CAN., *THE DEVELOPMENT OF A BRIEF ACTUARIAL SCALE FOR SEXUAL OFFENSE RECIDIVISM 13–14* (1997), http://ww2.psepc-sppcc.gc.ca/publications/corrections/199704_e.pdf.

⁵⁰ *See generally* Hanson, *supra* note 30, at 65–67 (discussing the types of factors that might adjust an actuarial prediction).

historical, clinical, and risk management.⁵¹ The history scale contains ten items: previous violence, age at first violent incident, relationship instability, employment problems, substance use problems, major mental illness, psychopathy, early maladjustment, personality disorder, and prior supervision failure.⁵² The five items on the clinical scale are lack of insight, negative attitudes, active symptoms of major mental illness, impulsivity, and unresponsiveness to treatment.⁵³ Finally, the five risk management scale items are infeasibility of plans, exposure to destabilizers, lack of personal support, noncompliance with remediation attempts, and stress.⁵⁴ Examiners rate each item on a scale of 0–2, making 40 the maximum possible score. Although this scheme resembles the actuarial devices described above, there is no algorithm that statistically correlates a given score with a risk probability. Rather, the examiner arrives at risk ratings of low, moderate, or high based on his or her clinical assessment of the various items in the protocol.⁵⁵

Of course, the scores obtained using the HCR-20 can be and have been correlated with risk. For instance, a follow-up study of patients who were evaluated using this instrument found that while only 11% of those who scored in the 0–14 range committed or threatened to commit a physically violent act during a two-year period, 35% of those who scored over 26 did so, and 75% of those who received the highest scores did so.⁵⁶ When combined with such studies, the HCR-20 is not very different from an actuarial device. But a distinction still exists: Whereas actuarial devices like the VRAG were developed empirically (through studies as to which items statistically differentiated violent and nonviolent offenders), the items on the HCR-20 came from observational and theoretical research as to the likely correlates of risk.

The same can be said for the Psychopathy Checklist Revised (PCL-R), which consists of twenty factors scored on a 0–2 scale, designed to identify the

⁵¹ See Kevin S. Douglas & Christopher D. Webster, *The HCR-20 Violence Risk Assessment Scheme: Concurrent Validity in a Sample of Incarcerated Offenders*, 26 CRIM. JUST. & BEHAV. 3, 8 (1999).

⁵² *Id.*

⁵³ *Id.*

⁵⁴ *Id.*

⁵⁵ *Id.*

⁵⁶ Kevin S. Douglas et al., *Assessing Risk for Violence Among Psychiatric Patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version*, 67 J. CONSULTING & CLINICAL PSYCHOL. 917, 925 (1999) (noting that only 45% of the latter group actually committed a violent act, while 35% only threatened one).

extent to which an individual is psychopathic.⁵⁷ As with the HCR-20, the items on the PCL-R were developed through clinical observation rather than mechanistically. Further, the ultimate score obtained indicates degree of psychopathy, not risk potential. But, as noted above, additional research has found that the instrument is also a good predictor of violence, with a score of greater than 30 (on a 40-point scale) indicating a strong propensity for violence.⁵⁸

B. A Comparison of Actuarial and Clinical Prediction

Because both the validity research reported below and my proposed approach to prediction testimony make distinctions between these various methodologies, a closer look at them will be useful. Each methodology has advantages and disadvantages.

One advantage of both the actuarial and structured judgment techniques is that they combine variables more reliably (that is, in a manner that is more easily replicated) than clinical opinions do.⁵⁹ Each of the first two approaches transparently designates the specific factors the predictor must consider. Compared to actuarial prediction, the structured clinical approach gives the examiner more flexibility as to how much weight to assign to each predictor, but it still limits the examiner to the listed predictors and requires reference to a specified scale (e.g., 0–2). An unstructured clinical prediction, in contrast, “must ultimately be based upon an overall subjective impression which is based upon an understanding of the interrelatedness of many facts.”⁶⁰ Because these subjective impressions are likely to differ across clinicians and even across cases with respect to the same clinician, each clinical prediction is

⁵⁷ Robert D. Hare, *The Hare PCL-R: Some Issues Concerning Its Use and Misuse*, 3 LEGAL & CRIM. PSYCHOL. 99, 100 (1998).

⁵⁸ Steven Wong, *Recidivism and Criminal Career Profiles of Psychopaths: A Longitudinal Study*, 24 ISSUES IN CRIMINOLOGICAL & LEGAL PSYCHOL. 147 (1995); see also Marnie E. Rice et al., *An Evaluation of a Maximum Security Therapeutic Community for Psychopaths and Other Mentally Disordered Offenders*, 16 L. & HUM. BEHAV. 399, 406 (1992) (reporting that 77% of those who scored higher than 25 on the PCL-R committed a violent offense despite treatment).

⁵⁹ See Slobogin, *supra* note 1, at 122. “Unlike the actuarial predictor, [a] clinical decisionmaker is not committed in advance of decision to the factors that will be considered and the rule for combining them.” *Id.* at 122 n.94 (quoting Barbara D. Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment*, 88 YALE L.J. 1408, 1423 (1979)).

⁶⁰ SEYMOUR L. HALLECK, *PSYCHIATRY AND THE DILEMMAS OF CRIME* 314 (1967), *quoted in* Slobogin, *supra* note 1, at 122–23.

highly likely to be based on a “different constellation of factors,” some of which may be irrelevant or based on erroneous stereotypes and prejudices.⁶¹

Furthermore, as indicated above, the actuarial and structured approaches can provide more precise probability estimates than is possible through the clinical process. Because it is not standardized, a clinical evaluation of dangerousness cannot be reliably compared to other predictions. At best, the evaluator can give the court information about his or her own accuracy rate (which is rarely available) and the accuracy rate of other clinicians who make predictions (which is not particularly helpful in assessing this evaluator’s work). As a consequence of this paucity of data, honest clinical evaluators asked to state the violence proneness of a particular individual can at best rate subjects as high or low risk, and in doing so they have only their experience and clinical wisdom to back them up. In contrast, the VRAG and similar devices definitively state the probability of reoffending within a certain time period for people who have the subject’s characteristics.

As I noted in my previous article,⁶² however, an actuarial prediction does have at least two drawbacks. First, actuarial information is not presently available for some of the populations that are likely to be evaluated. For instance, minimal data exist about the reoffending rates of insanity acquittees or offenders sentenced to death. Whereas a clinical prediction does not rely on this type of information, an actuarial prediction *depends* on it, as do structured approaches to the extent their results are tied to specific recidivism probabilities.

An even more trenchant criticism of actuarial prediction, to a lesser extent also true of structured judgment approaches, is that it inevitably neglects pertinent characteristics of the individual evaluated.⁶³ An actuarial prediction may give us the most explicit information we can obtain about sex offenders who committed their first sex offense before age twenty-six against a female victim. But it tells us nothing about the extent to which other factors that were not included in the instrument and not researched by its developers⁶⁴ might

⁶¹ See MELTON ET AL., *supra* note 16, at 278–79 (discussing judgment errors and biases that can affect clinical prediction).

⁶² Slobogin, *supra* note 1, at 123.

⁶³ *Id.* at 124–25 (citing PAUL MEEHL, *What Can the Clinician Do Well?*, in PSYCHODIAGNOSIS: SELECTED PAPERS 165, 169–70 (1973) (discussing factors favoring clinical prediction)); Underwood, *supra* note 59, at 1427–28).

⁶⁴ Note that the developers of the VRAG considered fifty variables and found that only twelve were significantly related to violence. Grant T. Harris et al., *Psychopathy and Violent Recidivism*, 15 L. & HUM.

increase or decrease the potential for reoffending. As I noted in my previous article, “It is in discovering these idiosyncratic characteristics that the clinical process provides information, however imprecise, that an actuarial prediction cannot.”⁶⁵ Thus, leading researchers in the area have stated that “actuarial instruments . . . are best viewed as ‘tools’ for clinical assessment [R]eliance on clinical judgment—aided by an empirical understanding of risk factors for violence and their interactions—reflects, and in our view should reflect, the standard of care at this juncture in the field’s development.”⁶⁶

One final difference between the methodologies—but one that is not as great as some have suggested—has to do with the types of variables consulted. Actuarial instruments tend to be more objective than more clinical approaches, in the sense that they rely heavily on “hard” variables that are relatively easy to ascertain, such as prior convictions and age. But it should also be apparent from the above descriptions that actuarial devices can incorporate more subjective, “soft” variables as well. The VRAG’s reliance on the PCL-R illustrates this point. In addition to delving into specific types of antisocial conduct, the latter instrument requires assessment of the individual’s glibness, grandiosity, need for stimulation, penchant for pathological lying, manipulateness, lack of remorse, affect, callousness, and similar variables.⁶⁷ Although the scoring criteria for these variables are fairly tightly defined and can be reliably scored by trained examiners,⁶⁸ they do not eliminate the subjectivity of the prediction process.

C. *The Validity of Prediction Methodologies*

Attempts to gauge the validity of a given prediction confront a number of obstacles. Two in particular are important to note at the outset: First, those predicted to be at high risk are generally confined, or at least treated, making ambiguous whether subsequent failure to offend means the prediction—presumably focused on behavior while unconfined—was inaccurate. Second, even if release of the purportedly dangerous person occurs for some reason, follow-up to ascertain the accuracy of a prediction can be very difficult;

BEHAV. 625, 631 tbl.1 (1991). Thus, for the thirty-eight excluded variables, there is some basis for considering them irrelevant to prediction.

⁶⁵ Slobogin, *supra* note 1, at 125.

⁶⁶ MONAHAN ET AL., *supra* note 45, at 134.

⁶⁷ Hare, *supra* note 57, at 100.

⁶⁸ Grant T. Harris et al., *The Construct of Psychopathy*, 28 CRIME & JUST. 197, 217 (2001) (“The PCL-R has yielded high inter-rater reliability and test–retest reliability on prisoners and forensic psychiatric patients.”).

conviction and arrest records seriously underreport crime,⁶⁹ and self-reports about incriminating acts are obviously problematic as well. Thus, any research evaluating predictive validity is likely to underestimate it.

Even taking these caveats into account, early research on dangerousness predictions was not encouraging. The initial studies measured validity in terms of false positives (the proportion predicted to be violent who were not) and false negatives (the proportion predicted to be nonviolent who were not). While finding relatively small false-negative rates, these studies reported false-positive rates of anywhere from 40 to 90%, with the usual finding somewhere in the 60 to 70% range.⁷⁰ That led to Monahan's conclusion in 1981 that the best clinical prediction methods produce erroneous predictions of dangerousness two out of three times.⁷¹

In light of more recent studies that try to correct for the methodological difficulties mentioned above, however, the typical false-positive rate for expert predictions should probably be revised downward to around 50%. For instance, a study of the COVR's validity that relied on record checks and self-reports of those who were released from the hospital showed that 49% of those predicted to be high risk committed at least one violent act within 20 weeks of release.⁷² Other well-done studies, using both clinical and actuarial methodologies, produce similar or even somewhat more positive results.⁷³ As early as 1984, only one year after *Barefoot* and only three years after his one-out-of-three statement, Monahan conceded that even clinical predictions of violence might be correct one out of two times.⁷⁴

⁶⁹ See MELTON ET AL., *supra* note 16, at 283 (reporting a study finding that self-reports increased base rates from 3% to 40%).

⁷⁰ See Slobogin, *supra* note 1, at 110–11, 117–18 (describing eight clinical prediction studies and three actuarial prediction studies finding false-positive rates ranging from 54% to 92%).

⁷¹ JOHN MONAHAN, *THE CLINICAL PREDICTION OF VIOLENT BEHAVIOR* 47 (1981).

⁷² John Monahan et al., *An Actuarial Model of Violence Risk Assessment for Persons with Mental Disorders*, 56 *PSYCHIATRIC SERVICES* 810, 814 tbl.2 (2005).

⁷³ For studies of clinical prediction, see L. Joy Apperson et al., *Short-Term Clinical Prediction of Assaultive Behavior: Artifacts of Research Methods*, 150 *AM. J. PSYCHIATRY* 1374 (1993) (25%); Charles W. Lidz et al., *The Accuracy of Predictions of Violence to Others*, 269 *J. AM. MED. ASS'N* 1007 (1993) (47%); Diana S. Sepejak et al., *Clinical Predictions of Dangerousness: Two-Year Follow-up of 408 Pre-Trial Forensic Cases*, 11 *BULL. AM. ACAD. PSYCHIATRY & L.* 171 (1983) (44%). For studies of actuarial prediction, see Deidre Klassen & William A. O'Connor, *A Prospective Study of Predictors of Violence in Adult Male Mental Health Admissions*, 12 *L. & HUM. BEHAV.* 143 (1988) (40%), and the research reported in the following text.

⁷⁴ John Monahan, *The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy*, 141 *AM. J. PSYCHIATRY* 10, 11 (1984); see also Randy Otto, *On the Ability of Mental Health*

The false-positive rate, however, is only part of the story. The accuracy of expert predictions can be fully understood only if base rates of recidivism are taken into account. If, for instance, the base rate for violence among the population studied (say, serial sex offenders) is 50%, then a 50% false-positive rate would be no better than chance selection. Those carrying out the prediction would do just as well simply assigning, with no evaluation, every second individual to the dangerous category. On the other hand, a false-positive rate of 50% is quite impressive if only 1 out of 100 people in the relevant population commit crime during the follow-up period; under these circumstances, a correct prediction of violence 1 out of 2 times would be 50 times more accurate than random assignment.⁷⁵

In recognition of this interplay between accuracy and base rates, researchers developed a second way of measuring predictive validity. The “receiver operating characteristic” curve represents the true positive rate (sensitivity) as a function of the false-positive rate (specificity).⁷⁶ The “area under the curve” (AUC) produced by this function provides a measure of accuracy as it relates to base rates. If the AUC value is .50, the predictive power of the methodology is no better than chance, while an AUC value of 1.0 represents perfect accuracy, and a value of 0 means complete inaccuracy. An AUC value of .75 for a given prediction methodology indicates a 75% chance that a recidivist will receive a higher risk rating than a nonrecidivist.⁷⁷

Douglas Mossman reanalyzed 58 early studies (from 1972 to 1991) and found that, despite their often high false-positive rates, 47 of them demonstrated prediction accuracy “significantly better than chance,” with an average AUC value of .67 for clinical prediction and .71 for actuarial

Professionals to “Predict Dangerousness”: A Commentary on Interpretations of the “Dangerousness” Literature, 18 L. & PSYCHOL. REV. 43, 63 & n.63 (1994).

[W]hereas first generation research suggested that perhaps one out of three people predicted to engage in some kind of violent behavior will actually go on to do so, more recent studies suggest that one out of every two people predicted to be violent would go on to engage in some kind of legally relevant, violent behavior.

Id. at 63.

⁷⁵ I first pointed this out in 1984. Slobogin, *supra* note 1, at 112–14.

⁷⁶ The more true positives and the fewer false positives, the more accurate the predictive method. For general background on ROC curves in the context of predicting dangerous sexual behavior, see, e.g., R. Karl Hanson & David Thornton, *Improving Risk Assessments for Sex Offenders: A Comparison of Three Actuarial Scales*, 24 L. & HUM. BEHAV. 119, 125 (2000).

⁷⁷ Kevin Douglas & John Weir, HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography 4 (2005), <http://www.sfu.ca/psyc/faculty/hart/Guides.htm> (follow “Download PDF” hyperlink).

prediction made on a cross-validation group.⁷⁸ Representative AUC values for modern actuarial devices are generally as good or better: .76 to .80 for the VRAG⁷⁹ and .63 for the COVR.⁸⁰ Studies of the HCR-20 have found AUC values ranging from .69 to .89, depending on the population studied,⁸¹ and research on the PCL-R has obtained a value of around .72.⁸² These findings obviously do not indicate perfect accuracy, but they do demonstrate that professional predictions are much better than chance selection.

A separate validity issue is whether adjusting an actuarial prediction using additional, nonactuarialized factors improves the ultimate prediction. Some commentators have stated that “[a]ctuarial methods are too good and clinical judgment too poor to risk contaminating the former with the latter.”⁸³ But, as already noted, others hold an opposing view, with some, for example, pointing to research indicating that “considering current clinical conditions (especially regarding the presence of heightened anger or violent fantasies) can be an important contribution to assessments of dangerousness.”⁸⁴ It seems fair to say that this issue remains unresolved at the present time.

II. THE COURTS’ NONCHALANCE TOWARD PREDICTION TESTIMONY

Until very recently, none of these nuances affected judicial analysis of prediction testimony. As noted in the Introduction, the Supreme Court’s decision in *Barefoot v. Estelle* refused to put constitutional limitations on such testimony, at least as long as it is presented in an adversarial proceeding.⁸⁵ Justice White’s majority opinion gave two reasons for this stance. First, Justice White characterized the suggestion that clinical prediction testimony be prohibited in death penalty proceedings as “somewhat like asking us to

⁷⁸ Douglas Mossman, *Assessing Predictions of Violence: Being Accurate About Accuracy*, 62 J. CONSULT. & CLIN. PSYCHOL. 783, 789 (1994). I focus on the medians from studies using cross-validation samples because AUC values so derived tend to reflect more accurately the real-world validity of a device than AUC values obtained on the sample used to derive the instrument.

⁷⁹ VERNON L. QUINSEY ET AL., *Violent Offenders: Appraising and Managing Risk* 148 (1st ed. 1998); Harris et al., *supra* note 41, at 386.

⁸⁰ Monahan et al., *supra* note 72, at 814.

⁸¹ Douglas & Weir, *supra* note 77, at 5–9.

⁸² M. Dolan & M. Doyle, *Violence Risk Prediction: Clinical and Actuarial Measures and the Role of the Psychopathy Checklist*, 177 BRIT. J. PSYCHIATRY 303, 305 (2000).

⁸³ QUINSEY ET AL., *supra* note 79, at 171.

⁸⁴ Thomas R. Litwack, *Actuarial Versus Clinical Assessments of Dangerousness*, 7 PSYCHOL. PUB. POL. & L. 409, 435–36 (2001).

⁸⁵ 463 U.S. 880, 896–97 (1983).

disinvent the wheel,” given the multiple contexts in which dangerousness predictions, by laypersons and experts alike, have always been permitted.⁸⁶ This unquestioning reliance on precedent—in effect saying “we will not change because we’ve always done it this way”—represents legal reasoning at its most primitive.⁸⁷ The second reason given by the Court, examined in more detail in Part VI of this Article, provides a more substantial, if still highly debatable, rationale: “We are unconvinced . . . that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness, particularly when the convicted felon has the opportunity to present his own side of the case.”⁸⁸

The Court has not revisited the *Barefoot* issue since *Daubert* was decided, so we do not know if the latter decision’s emphasis on reliability, described earlier in this Article,⁸⁹ would affect the Court’s constitutional analysis. But it is unlikely to do so. *Daubert*, *Frye*, and similar cases interpret the rules of evidence, not the Constitution. The Court is cautious about using the Constitution to mandate evidentiary practices that must be followed by every jurisdiction in the country. Rather, the Court usually sets a low constitutional floor, above which courts and legislatures are free to experiment.⁹⁰

To date, only a handful of courts and no legislatures have exercised this freedom, even when the context is the death penalty. In the 1998 decision of *Nenno v. State*, for instance, the Texas Court of Criminal Appeals concluded that any defect in clinical prediction testimony introduced in a capital proceeding “affects the weight of the evidence rather than its admissibility.”⁹¹ Four years later, the Fifth Circuit Court of Appeals found this position so clearly supported by precedent that it was willing to characterize any objection to such testimony as “frivolous” (a particularly strong statement given the fact that the testimony at issue was offered by the infamous Dr. Grigson).⁹² Testimony based on actuarial and structured professional judgments is also routinely admitted in death penalty cases.⁹³

⁸⁶ *Id.* at 896–97.

⁸⁷ See Slobogin, *supra* note 1, at 159–60.

⁸⁸ *Barefoot*, 463 U.S. at 901.

⁸⁹ See *supra* text accompanying notes 14–15.

⁹⁰ See generally CASS R. SUNSTEIN, ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT (1999).

⁹¹ 970 S.W.2d 549, 562 (Tex. Crim. App. 1998).

⁹² *Johnson v. Cockrell*, 306 F.3d 249, 254 (5th Cir. 2002).

⁹³ See, e.g., *Martinez v. Dretke*, 99 F. App’x 538, 542 (5th Cir. 2004); *United States v. Barnette*, 211 F.3d 803 (4th Cir. 2000) (holding that testimony in a capital case based on the PCL-R is admissible under *Daubert*).

If prediction testimony is admissible to support an execution, it presumably should be admissible in any case. A survey of the case law largely confirms this surmise. A review of sexual predator cases conducted in 2002 found virtually no appellate decisions upholding challenges to expert prediction testimony,⁹⁴ and decisions since then have continued that trend.⁹⁵ A 2003 survey of civil commitment cases likewise concluded that “[j]udicial opinion, split on virtually every other form of behavioral or psychic expertise, has so far unanimously accepted predictive expertise in civil commitments.”⁹⁶

There are signs of judicial discontent, but they have been scattered and un influential. In his dissent in *Barefoot*, Justice Blackmun argued that, given the inaccuracy of clinical prediction testimony and its potential for overinfluencing the jury, courts should hear only “[l]ay testimony, frankly based on statistical factors with demonstrated correlations to violent behavior.”⁹⁷ Judge Emilio Garza, countering his colleagues on the Fifth Circuit, made the same argument seventeen years later and bolstered it with the assertion that, because it is so often wrong, expert prediction testimony “fails” *Daubert*.⁹⁸ A federal district court judge echoed the latter view in 2004, albeit in dictum.⁹⁹ And in the past decade, a smattering of state courts have forthrightly concluded that some types of prediction testimony are inadmissible.¹⁰⁰

⁹⁴ Randy K. Otto & John Petrila, *Admissibility of Testimony Based on Actuarial Scales in Sex Offender Commitments: A Reply to Doren*, 3 SEX OFFENDER L. REP. 1 (2002). Trial courts in Iowa, Arizona, Florida, and Missouri have excluded predictive opinion, but appellate courts in the first two states overruled those decisions, and appellate courts in the latter two states have yet to address the issue. See *In re Detention of Holtz*, 653 N.W.2d 613, 619 (Iowa Ct. App. 2002) (“Our research has revealed no state appellate court decision which has found actuarial instruments inadmissible at [sexually violent predator] proceedings.” (quoting *In re Commitment of R. S.*, 773 A.2d 72, 96 (N.J. Super. Ct. App. Div. 2001))).

⁹⁵ See, e.g., *In re Detention of Traynoff*, 831 N.E.2d 709 (Ill. App. Ct. 2005); *In re Detention of Holtz*, 653 N.W.2d at 619 (permitting actuarial evaluation only when used “in conjunction with a full clinical evaluation”); *Commonwealth v. Bradway*, 816 N.E.2d 152, 157–58 (Mass. App. Ct. 2004).

⁹⁶ Alexander Scherr, *Daubert & Danger: The “Fit” of Expert Predictions in Civil Commitments*, 55 HASTINGS L.J. 1, 60 (2003). For a long list of cases in which prediction testimony has been held admissible, see John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 410 n.74 (2005).

⁹⁷ *Barefoot v. Estelle*, 463 U.S. 880, 938 (1983) (Blackmun, J., dissenting).

⁹⁸ *Flores v. Johnson*, 210 F.3d 456, 464–65 (5th Cir. 2000) (Garza, J., specially concurring).

⁹⁹ *United States v. Sampson*, 335 F. Supp. 2d 166 (D. Mass. 2004).

¹⁰⁰ *People v. Taylor*, 782 N.E.2d 920, 932 (Ill. App. Ct. 2002); *Collier v. State*, 857 So. 2d 943, 945–46 (Fla. Dist. Ct. App. 2003). See generally Eric S. Janus & Robert A. Prentky, *Forensic Use of Actuarial Risk Assessment with Sex Offenders: Accuracy, Admissibility and Accountability*, 40 AM. CRIM. L. REV. 1443, 1459 & n.90 (2003).

A noteworthy fact about the latter decisions is that almost all of them have taken aim at the actuarial method, while continuing to permit clinical opinion, despite its lesser accuracy. Ironically, these courts exempt traditional, more speculative expert opinion from screening rules but are willing to subject newer, more scientific testimony to *Daubert* or *Frye* analysis. To date, only one court has concluded that clinical prediction testimony is so unreliable it must be excluded; the force of this decision was undercut, however, by that court's subsequent willingness to allow clinical opinion as long as it is combined with actuarial findings.¹⁰¹

In short, the courts have paid virtually no attention to the critics' plea that expert prediction testimony be barred or severely limited. No court has pronounced a complete ban on expert prediction testimony. Those very few cases where exclusion has occurred have usually involved actuarial-based testimony, not more suspect clinical testimony. The rest of this Article critiques these positions, looking in turn at each of the four components of evidentiary analysis identified in the introduction: materiality, probative value, helpfulness, and prejudicial impact.

III. MATERIALITY

If dangerousness is at issue—as it often is at sentencing, and always is in criminal and civil commitment proceedings—the materiality, or logical relevance, of prediction testimony would seem to be a foregone conclusion. The issue is not that simple, however. Especially when, as is true for actuarial prediction, the basis for the prediction is transparent, three types of questions can be raised: First, is prediction that is based on studies of groups ever applicable to a given individual? Second, assuming so, is the particular prediction methodology used applicable to this particular individual? Third, assuming so, is the prediction nonetheless immaterial because it relies on factors that are not legally cognizable (such as race)?

¹⁰¹ Compare *In re Care and Treatment of Coffel*, 117 S.W.3d 116, 129 (Mo. Ct. App. 2003) (excluding clinical prediction testimony), with *In re Care and Treatment of Kapprelian*, 168 S.W.3d 708, 715 (Mo. Ct. App. 2005) (distinguishing *Coffel* and admitting clinical testimony because the expert's opinion was not based "solely on clinical judgment").

A. *Nomothetic Prediction Data*

Justice Coyne of the Minnesota Supreme Court succinctly raised the first issue in his dissenting opinion in *In re Linehan*: “Not only are . . . statistics concerning the violent behavior of others irrelevant, but it seems to me wrong to confine any person on the basis not of that person’s own prior conduct but on the basis of statistical evidence regarding the behavior of other people.”¹⁰² Assume that a person convicted of a sex offense—call him John—receives a score of 21 on the VRAG. The most accurate characterization of this score is that John shares a number of traits with a group of individuals, 76% of whom are known to have engaged in violent recidivism. Justice Coyne’s position is that this information is immaterial in a proceeding to determine whether John is dangerous, because it speaks of other people, not John. Mental health professionals have long expressed similar concerns, with the psychologist Gordon Allport making the following statement over sixty years ago:

A fatal nonsequitur occurs in the reasoning that if 80% of the delinquents who come from broken homes are recidivists, then this delinquent from a broken home has an 80% chance of becoming a recidivist. The truth of the matter is that this delinquent has either 100% certainty of becoming a repeater or 100% certainty of going straight.¹⁰³

Neither of these objections renders group-based predictions legally immaterial, however. First, Justice Coyne mischaracterizes the nature of prediction testimony, because a prediction based on research examining the behavior of others can still be directly linked to the individual who is the subject of the prediction. It is John’s age, prior record, marital status, and psychopathic personality traits that place him in the 76% recidivism category. Moreover, these types of characteristics are all logically related to recidivism, allowing us, as John Monahan has pointed out, to make “the inferential leap from membership in a class that has in the past been violent to the prediction that this member of the same class will in the future be violent.”¹⁰⁴ While it is true, as Allport’s comment above notes, that any given individual either will or will not offend, it is not incoherent to say that the VRAG score means there is a 76% chance that John will recidivate. This type of statement is no different in kind than an assertion that John probably committed a criminal act in the past (even though he either did or did not) or a prediction, based on a clinical

¹⁰² *In re Linehan*, 518 N.W.2d 609, 616 (Minn. 1994) (Coyne, J., dissenting).

¹⁰³ GORDON ALLPORT, *THE USE OF PERSONAL DOCUMENTS IN PSYCHOLOGICAL SCIENCE* 156 (1942).

¹⁰⁴ MONAHAN, *supra* note 71, at 66.

interview, that he is highly likely to reoffend, a type of statement that mental health professionals make all the time.¹⁰⁵

The latter comparison suggests another reason why prediction testimony cannot be considered immaterial simply because it is nomothetic in nature. Such a position would prohibit not only actuarial prediction testimony but clinical prediction testimony as well, and indeed would bar virtually any type of opinion evidence from mental health professionals and similar experts, because all of these opinions inherently are based on inferences drawn from observation of other individuals. As I stated in my previous article:

While clinicians look at individual patterns, they do not do so in a vacuum. Rather they make comparisons—sometimes implicit, sometimes explicit—between these patterns and the patterns of other individuals or groups of individuals that they know about through experience, training, or education. In this regard, most clinical predictions differ from actuarial ones only in the sense that the link between past groups and present individuals is not statistically correlated.¹⁰⁶

B. Nomothetic Prediction and Individual Cases

A willingness to permit reliance on predictions based on nomothetic information does not necessarily mean that such predictions will be material in every case. Much depends upon the extent to which the underlying data can be generalized. Sometimes the group data may not be applicable to the individual who is the subject of the prediction (the norming problem). Even if it is, it may not address the question the law wants answered (the criterion variable problem). Finally, even if it does answer the legal question, the prediction may provide misleading information about the individual in question (the lack-of-individualization problem). All of these points, which deal with what *Daubert* called the “fit” issue,¹⁰⁷ can be illustrated by the case of John, the hypothetical offender introduced above.

¹⁰⁵ Joseph T. McCann, *Standards for Expert Testimony in New York Death Penalty Cases*, 68 N.Y. St. B. J. 30, 30 (1996) (stating that in every one of the 144 capital cases in which Grigson testified up to 1992, he stated he was sure the defendant would kill again).

¹⁰⁶ Slobogin, *supra* note 1, at 126.

¹⁰⁷ *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 591 (1993) (“‘Fit’ is not always obvious, and scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes.” (quoting *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985))).

Recall that John's score on the VRAG is a 21, which supposedly indicates a 76% chance of reoffending. But note that the VRAG was normed on a group of mostly white offenders released from a maximum-security psychiatric hospital in northern Ontario, Canada, in the 1970s.¹⁰⁸ What if John is an African-American sex offender in the United States at the present time? The applicability of the VRAG data to his case is open to serious question. In fact, the VRAG's AUC value when cross-validated on sex offenders is only .60.¹⁰⁹ One commentator has suggested that this type of norming problem is "vast and potentially insurmountable" and that without finely tuned cross-validation, "absolute risk predictions based upon [actuarial instruments] are meaningless."¹¹⁰

Second, in determining who in their sample population engaged in "violent recidivism," researchers for the VRAG included not only those who committed felonies, but also those who engaged in two or more simple assaults.¹¹¹ As a result, the risk ratios reported by the developers of VRAG do not refer to the probability of serious violence. This fact substantially reduces the materiality of the VRAG score for capital sentencing and other contexts in which the legal inquiry is usually focused on precisely that issue. Consider, for instance, how a jury should interpret John's VRAG score of 21—even assuming it is appropriately normed—if the question is whether John exhibits a "propensity to commit murder that will probably constitute a continuing threat to society" or a "probability that [he] would commit criminal acts of violence that would constitute a continuing threat to society" (aggravating factors under the Idaho and Texas capital statutes, respectively).¹¹²

Finally, as noted above, actuarial devices consider only a limited number of variables. John's score on the VRAG does not take into account whether he is undergoing treatment, is about to get married, has recently lost functioning in one or more limbs, or has found religion. The proponents of the VRAG argue

¹⁰⁸ QUINSEY ET AL., *supra* note 79, at 141.

¹⁰⁹ Mamie E. Rice & Grant T. Harris, *Cross-Validation and Extension of the Violence Risk Appraisal Guide for Child Molesters and Rapists*, 21 L. & HUM. BEHAV. 231, 235 tbl.2 (1997).

¹¹⁰ Donna Cropp Bechman, *Sex Offender Civil Commitments: Scientists or Psychics?*, 16 CRIM. JUST. 24, 29 (2001).

¹¹¹ Litwack, *supra* note 84, at 428.

¹¹² IDAHO CODE ANN. § 19-2515(9)(i) (2006); TEX. CODE CRIM. PROC. ANN. art. 37.0711 § 3(b)(2) (Vernon 2006).

that, based on their data, these factors are irrelevant.¹¹³ But their samples were certainly not large enough to encompass statistically significant numbers of every type of treatment or religious conversion experience or every type of individual's reaction to these sorts of events.¹¹⁴ Unless data exist demonstrating that these factors do not lower the score of people like John, reliance *solely* on the actuarial information might be inappropriate because alone that information may be considered immaterial to John's propensity toward violence.

These are all worthy concerns, but none should lead to wholesale exclusion of prediction testimony. Difficulties concerning norming, criterion variables, and lack of individualization can all be brought to the attention of the fact finder. The subject of the prediction can also be given wide latitude to question the generalizability of the research underlying the actuarial prediction and suggest why that prediction may be off base in his or her case. With these precautions, any weakness in the materiality of prediction testimony usually can be exposed.

Sometimes, however, these difficulties—in combination or alone—will be so significant that they should go to admissibility and not weight. Further discussion of when that might occur is best left to analysis of the prejudice factor in Part VI.

C. *Illegitimate Bases for Prediction*

The materiality of a prediction may be threatened not only by the factors it does not consider, but also by the variables it does. At one time, racial traits were considered a good predictor of behavior.¹¹⁵ Currently, the gender and age of the perpetrator play crucial roles in many prediction methodologies—explicitly when the prediction is actuarial and implicitly when it is clinical. Yet in many areas of the law, differentiating between individuals on the basis of race, gender, and age often runs afoul of antidiscrimination principles, in

¹¹³ QUINSEY ET AL., *supra* note 79, at 177–78 (“[T]here is good evidence that clinicians’ appraisals of patients’ current clinical conditions are unrelated to recidivism [I]t is now clear that the ‘getting to know’ individuals that occurs in typical interviews does not improve the prediction of behavior in any domain.”).

¹¹⁴ Note, for instance, that none of the variables excluded by the VRAG researchers involved the factors mentioned in the text or any of the other factors that were earlier hypothesized in discussing adjusted actuarial assessments (e.g., treatment, gang membership, specific threats) or clinical variables (e.g., anger). Harris et al., *supra* note 64, at 631 tbl.1.

¹¹⁵ See MONAHAN, *supra* note 71, at 74–75.

large part because these are immutable characteristics that usually have very little to do with any legitimate purpose.¹¹⁶

Some commentators have gone further, suggesting that *any* characteristic of an individual over which he or she has little or no control—diagnosis, personality traits, abuse as a child—should be anathema as a basis for a prediction.¹¹⁷ In arguing that actuarial prediction should be barred from the sentencing setting, Daniel Goodman put this point as follows:

It is a fundamental orthodoxy of our criminal justice system that the punishment should fit the crime and the individual, not the statistical history of the class of persons to which the defendant belongs. To allow a criminal defendant's sentence to be determined to any degree by his unchosen membership in a given [group] denies the very premise of self-determination upon which our criminal justice system is built. It raises the threat that defendants will be sentenced not only on the basis of their personal merit or conduct, but on the basis of their "status."¹¹⁸

Others have echoed this view, with John Monahan recently arguing that “[p]ast criminal behavior is the only scientifically valid risk factor for violence that unambiguously implicates blameworthiness, and therefore the only one that should enter the jurisprudential calculus in criminal sentencing.”¹¹⁹ If these prescriptions are followed, then both actuarial and clinical prediction testimony would be immaterial at capital and noncapital sentencing proceedings to the extent that their predicate ventures beyond criminal history.

A first response to this argument is that even suspect classifications such as race and gender are constitutionally permissible when they significantly further a compelling government interest.¹²⁰ The identification of dangerous individuals is a very important government interest, and gender and age, at least, are extremely useful in making that determination. Race, on the other

¹¹⁶ See generally *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 360 (1978) (stating “race, like gender . . . is an immutable characteristic which its possessors are powerless to escape or set aside” and thus subject to constitutional protection under the Equal Protection Clause). One is also powerless to escape one’s age, although the Supreme Court has held that discrimination on the basis of relative youth—the type of discrimination most relevant to violence prediction—is not cognizable under the Age Discrimination Employment Act. *Gen. Dynamics Land Sys., Inc. v. Cline*, 540 U.S. 581, 600 (2004).

¹¹⁷ See Daniel S. Goodman, Note, *Demographic Evidence in Capital Sentencing*, 39 STAN. L. REV. 499, 521 (1987).

¹¹⁸ *Id.*

¹¹⁹ Monahan, *supra* note 96, at 428; see also Underwood, *supra* note 59, at 1416–17.

¹²⁰ See JOHN E. NOWAK & RONALD D. ROTUNDA, *PRINCIPLES OF CONSTITUTIONAL LAW* 378–79 (2d ed. 2005).

hand, is not a particularly good predictor,¹²¹ and reliance on it for prediction purposes should probably be barred in any event because of the societal and symbolic repercussions such reliance would occasion.¹²²

A second and more important response is that when the government relies on a dangerousness prediction to enhance a sentence, it is not pursuing punishment based on blameworthiness and retribution, but rather is seeking prevention based on assessment of risk, a completely different enterprise. When the government is engaged solely in assessing blameworthiness for past conduct, as is the case at trial, then Goodman and Monahan are correct in their assertion that immutable factors should play no role. But at sentencing proceedings in those states that view the goal of punishment as a mix of retributive and other objectives, blameworthiness is not the only issue. For instance, retributive considerations might be considered relevant only in setting the outer limit of the sentence, with its precise length in a given case dependent upon an evaluation of dangerousness and rehabilitative potential.¹²³ In such instances, limiting the basis of any prediction to factors that indicate blameworthiness—such as prior crimes—is *inconsistent* with the purpose of punishment.

What Goodman and Monahan are really saying is that dangerousness is not a legitimate basis for a sentence. That is a defensible position.¹²⁴ But if one accepts the substantive law as a given—as this Article does—and if the relevant law permits sentences to be based on dangerousness—as is the case in many states¹²⁵—it does not undermine the criminal justice system’s “premise

¹²¹ MONAHAN ET AL., *supra* note 45, at 163 (showing a correlation between race and violence of .12).

¹²² As I argued in a different context:

[A government] action which depends upon factors such as race denigrates the *state's* interest in maintaining a democratic society and the allegiance of the populace Some citizens might see the state's behavior as a justification for using race as a surrogate in their own decision making. Other, more sensitive, citizens who experience or hear about such [actions] will question the legitimacy not only of the [actions] themselves, but of the government that would permit them. In either case, the democratic state's interests are severely damaged.

Christopher Slobogin, *The World Without a Fourth Amendment*, 39 UCLA L. REV. 2, 85–86 (1991).

¹²³ Norval Morris, *Predicting Violence with Statistics*, 34 STAN. L. REV. 249, 253 (1981).

¹²⁴ See, e.g., Paul Robinson, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARV. L. REV. 1429, 1438 (2001) (arguing that desert and incapacitative principles are inherently incompatible and that the latter principle should play no role in the criminal justice system, which should be based entirely on desert).

¹²⁵ Even many of the states with sentencing guidelines, which are normally an attribute of a desert-based system, permit dangerousness assessments to influence the length of sentence, usually through a parole board

of self-determination” (a premise that has already been honored at trial) to permit predictions based on immutable or quasi-immutable factors. Furthermore, as a practical matter, it is hardly protective of the individual’s interests to make prediction a sentencing issue and then deny the fact finder the best means of making the prediction.

Monahan himself acknowledges the force of these points in concluding that, outside the criminal setting, any prediction factor (other than race) is material.¹²⁶ Because blameworthiness is “irrelevant to imposing civil hospitalization,” for instance, Monahan would allow all risk factors to be considered in that arena, which for him includes civil commitment and, to the extent it is not punishment in disguise, sexual predator commitment.¹²⁷ Given governing law, this conclusion is clearly correct. In both commitment settings, blameworthiness is clearly not at issue as far as the Supreme Court is concerned.¹²⁸ Thus, the principle of self-determination that the law seeks to implement when blameworthiness is germane is not threatened by reliance on unchangeable or difficult-to-change risk factors in commitment proceedings.

IV. THE PROBATIVE VALUE OF PREDICTION TESTIMONY

Probative value, as used here, is a measure of accuracy and is generally considered a second aspect of relevancy (the first being materiality).¹²⁹ Prediction testimony may not be very accurate in an absolute sense, but it is sufficiently accurate to meet the law’s test of probativeness, whether that term is defined in the traditional manner or in a *Daubert* sense. A separate, but closely related, issue is whether prediction testimony, even if probative, can ever be enough to satisfy the relevant standard of proof. Although not technically an admissibility issue, a negative answer to this question could result in exclusion as well, so it is considered here.

determination. Richard S. Frase, *State Sentencing Guidelines: Diversity, Consensus, and Unresolved Policy Issues*, 105 COLUM. L. REV. 1190, 1200 (2005).

¹²⁶ Monahan, *supra* note 96, at 429.

¹²⁷ *Id.*

¹²⁸ See *Kansas v. Hendricks*, 521 U.S. 346, 361–62 (1997) (holding that sexual predator laws “[do] not implicate either of the two primary objectives of criminal punishment: retribution or deterrence”).

¹²⁹ MCCORMICK ON EVIDENCE § 185, at 276–77 (John W. Strong ed., 5th ed. 1999) (“There are two components to relevant evidence: materiality and probative value [Materiality] looks to the relation between the propositions for which the evidence is offered and the issues in the case [P]robative value [is] the tendency of evidence to establish the proposition that it is offered to prove.”).

A. *The Evidentiary Threshold*

Under the Federal Rules of Evidence, evidence is probative if it has “any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.”¹³⁰ Expert prediction testimony, even if clinical in nature, virtually always satisfies this test when dangerousness is the “fact” to be determined. As explained in Part II, most expert prediction methodologies produce predictions that exceed chance or random selection. Thus, assuming no objection on materiality grounds, a prediction using these methodologies makes a correct resolution of the dangerousness issue more probable than if the prediction were made without reference to any evidence.

Critics of prediction testimony have been fond of comparing it unfavorably to coin flipping, apparently on the assumption that the best possible false-positive rate for such testimony is 50%.¹³¹ Even accepting that assumption, the discussion in Part II should make clear why the coin flipping analogy does not work. Experts who are wrong more than one out of two times are only outdone by a coin toss when the base rate for harmful conduct is 50% or more. Yet the base rate in the groups subject to prediction is seldom that high.¹³² As the AUC values reported in Part II suggest, today’s experts can correctly identify those who will be violent at an accuracy rate—ranging from 45 to 75%—that is considerably higher than the base rate for violence within the prediction group—usually in the 15 to 35% range. Thus, the coin-flipping analogy is specious and misleading.

Nor is expert prediction testimony so unreliable that exclusion is required under *Daubert*, despite that decision’s focus on “error rates.”¹³³ The main thrust of *Daubert* is not that expert testimony should be admitted only when it is provably accurate (a rule that would require exclusion of virtually all expert testimony), but rather that its basis should be subjected to some sort of verification process—preferably scientific in nature—so that it reflects “the knowledge and experience of [the relevant] discipline.”¹³⁴ Thus, for instance, Michael Saks, long an advocate for a rigorous interpretation of *Daubert*, has nonetheless counseled against excluding expert evidence simply because “the

¹³⁰ FED. R. EVID. 401.

¹³¹ See, e.g., *Barefoot v. Estelle*, 463 U.S. 880, 931 (1983) (Blackmun, J., dissenting); Ennis & Litwack, *supra* note 17, at 737.

¹³² Slobogin, *supra* note 1, at 113 n.61.

¹³³ *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 594 (1993).

¹³⁴ *Id.* at 592.

witnesses practicing in that field assert erroneous conclusions with some regularity.”¹³⁵ As long as the fact finder is “informed about the likelihood of error in the opinions, and the court [is] satisfied that the fact-finder is capable of properly adjusting the weight to be given to the evidence,” Saks would permit such testimony.¹³⁶ On this reasoning, even if research evidences false-positive and false-negative rates well above 50%, material prediction testimony derived from a methodology that does better than chance should be admitted, as long as the relevant error rates are also provided to the fact finder.

Daubert additionally suggests that courts look at whether the methodology employed by the expert is generally accepted among those in the relevant field. This is, of course, also the inquiry required by the *Frye* test applicable in most of the jurisdictions that have not adopted *Daubert*. This definition of probative value may, at first glance, appear to pose more of a problem for prediction testimony. As noted earlier, even though clinical prediction testimony has been the mainstay of sentencing and commitment hearings for some time, the relevant official organizations have expressed considerable concern about its dependability, as have many adherents of the actuarial approach to prediction.¹³⁷ Nor has the latter methodology always escaped criticism under *Frye*: At least one court has excluded actuarial prediction testimony because only a small proportion of the relevant profession is familiar with it.¹³⁸

The fact remains that both methods of prediction are routinely used by large segments of the mental health profession, in practice and in court. The general acceptance notion cannot meaningfully be employed as a means of determining whether prediction evidence, writ large, should be banned. Rather, it is best used as a tool for ferreting out good and bad *methods* of obtaining such evidence, consistent with Rule 703’s requirement that the facts or data forming the basis for expert testimony be “of a type reasonably relied upon by experts in the particular field.”¹³⁹

In this regard, recall the contrast between the clinical prediction process employed by Dr. Kozol and his associates and that of Dr. Grigson. When, as

¹³⁵ Michael J. Saks, *The Legal and Scientific Evaluation of Forensic Science (Especially Fingerprint Expert Testimony)*, 33 SETON HALL L. REV. 1167, 1168 (2003).

¹³⁶ *Id.*; see also Richard D. Friedman, *Squeezing Daubert Out of the Picture*, 33 SETON HALL L. REV. 1047, 1048 (2003).

¹³⁷ See *supra* text accompanying notes 17–18.

¹³⁸ *Collier v. State*, 857 So. 2d 943, 945–46 (Fla. Dist. Ct. App. 2003); see also *In re Valdez*, No. 99-000045CI (Fla. 2000) (reported in Otto & Petrla, *supra* note 94, at 5).

¹³⁹ FED. R. EVID. 703.

was the case with Grigson's testimony in *Smith*, the facts come from a short interview with no attempt to obtain corroborative information, or, as was the case with his testimony in *Barefoot*, the facts are taken entirely from a hypothetical question, they should not be deemed "of a type reasonably relied upon by experts in the field." As I argued previously, "such procedures do not, on their face and according to most mental health professionals, appear to afford adequate protection against inaccuracy."¹⁴⁰ Similarly, there are accepted and unacceptable ways of using actuarial instruments and structured professional judgment.

An analysis of prediction testimony's probative value, then, reduces to two simple prescriptions: First, a court should permit a mental health professional to testify on the dangerousness issue only if it is established that he or she has followed generally accepted assessment procedures that attempt to insure a high degree of reliability, in both the social science and legal sense of the word.¹⁴¹ Second, if the testimony is admitted, error rate information must be provided to the fact finder, either through the expert or through some other means. If these conditions are met, the testimony should be considered sufficiently probative.

B. Evidentiary Sufficiency

Conceptually separate from, but practically related to, the probative value issue is the claim that prediction experts do not possess the ability to answer the prediction questions the law asks. In capital sentencing cases, dangerousness must be proven beyond a reasonable doubt, a standard of proof that is also required by many sexual predator statutes.¹⁴² In all other commitment contexts, the Supreme Court has required that dangerousness be proven by at least clear and convincing evidence.¹⁴³ If one adopts the standard quantifications of "proof beyond a reasonable doubt" as a 90 to 95% degree of certainty and of "clear and convincing proof" as a 75% degree of confidence (distinguishing both from the lower "preponderance of the evidence" standard used in civil cases, which is traditionally equated with a 51% level of

¹⁴⁰ Slobogin, *supra* note 1, at 134.

¹⁴¹ While the Supreme Court tends to use the word to mean accuracy—see *Daubert v. Merrell Dow Pharm.*, 509 U.S. 579, 589–90 (1993)—social scientists use it to mean inter-rater agreement or consistency. See David H. Kaye & David A. Freeman, *Reference Guide on Statistics*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 331, 341 (1994).

¹⁴² See statutes cited *supra* note 26.

¹⁴³ *Addington v. Texas*, 441 U.S. 418, 433 (1979).

certainty), prediction experts cannot, so the argument goes, satisfy the relevant standard of proof, given their 50%-plus false-positive rate. If that is true, the argument continues, there is no point in having them testify in the first place.¹⁴⁴

A first response to this argument is that it confuses the admissibility issue with the sufficiency issue. As a leading evidence treatise states:

Whether the entire body of one party's evidence is sufficient to go to the jury is one question. Whether a particular item of evidence is relevant to the case is quite another Thus, the common objection that the inference for which the fact is offered "does not necessarily follow" is untenable. It poses a standard of conclusiveness that very few single items of circumstantial evidence ever could meet. A brick is not a wall.¹⁴⁵

One might note, however, that in a substantial majority of sentencing and commitment cases, the expert prediction testimony is the entire body of evidence. In such cases, the testimony is both the brick and the wall. Then shouldn't the sufficiency and admissibility inquiries be merged?

If so, two other responses, one empirical and one legal, are possible. As an empirical matter, expert prediction testimony could be said to meet the sufficiency threshold (at least whenever the standard of proof is below the reasonable doubt level) if, rather than looking at false-positive and false-negative rates alone, we focus on their relationship to base rates. Specifically, expert prediction testimony might be considered clear and convincing proof any time the AUC value for the methodology in question is .75 or higher. Recall that such a value means that, given two randomly selected individuals, one drawn from the population of people who reoffend and the other drawn from the population of people who do not, there is a 75% chance that the methodology at issue will designate the violent individual a higher risk than the nonviolent individual.

The legal response to the sufficiency problem is to define "dangerousness" in such a way that error rates are minimized. For instance, many state commitment laws provide that a person is "dangerous" when he or she is

¹⁴⁴ Cf. Stephen J. Morse, *A Preference for Liberty: The Case Against Involuntary Commitment of the Mentally Disordered*, 70 CAL. L. REV. 54, 74-76 (1982) (arguing that a high number of unjust incarcerations probably result from the uncertainty of prediction).

¹⁴⁵ MCCORMICK ON EVIDENCE, *supra* note 129, at 278.

“likely” or “substantially likely” to harm another.¹⁴⁶ If the word “likely” is equated with a 51% probability, then proving beyond a reasonable doubt that a person is dangerous under this definition would only require a 46% likelihood (.90 x .51) that the person will harm another. Even clinical prediction testimony can produce such proof of a “definite maybe,” as Monahan and Wexler aptly describe this way of combining the standard of proof with the legal standard of dangerousness.¹⁴⁷ Other legal responses to the sufficiency problem include requiring a heightened standard of proof only for the relevant risk factors (e.g., two prior crimes, abuse as a child)¹⁴⁸ or for the person’s particular risk level (whether it be high, moderate, or low).¹⁴⁹

All of these ploys may strike some readers as sleights of hand. The tension between society’s urge to confine those who will harm others and its inability to identify precisely who these people are is not easily resolved. At bottom, the lack-of-sufficiency argument is an attack on dangerousness as a legal criterion, not an argument about admissibility. I have previously contended that, in many settings, lowering the standard of proof when dangerousness is the issue is normatively justifiable and consistent with our current criminal justice jurisprudence.¹⁵⁰ But one can also sensibly conclude that, given the myriad proof problems, dangerousness should be eliminated as a ground for liberty deprivation. For purposes of this Article, the substantive point will simply be recognized and left at large.

¹⁴⁶ See, e.g., *People v. Superior Court (Ghilotti)*, 44 P.3d 949, 954 (Cal. 2002) (holding that commitment under California’s sexual predator law does not require proof of a “better than even” chance of violence but only proof of “substantial danger—that is, a *serious and well-founded risk*—of criminal sexual violence”).

¹⁴⁷ John Monahan & David B. Wexler, *A Definite Maybe: Proof and Probability in Civil Commitment*, 2 L. & HUM. BEHAV. 37, 38 (1978). A normative argument for this approach might be based on the idea that, in some contexts, such as sexual predator cases, we will put up with many more false positives than false negatives, given the relative costs of the first—erroneous confinement until the next hearing—and the second—one or more rapes or child molestations. See generally Richard Berk et al., *Developing a Practical Forecasting Screener for Domestic Violence Incidents*, 29 EVAL. REV. 358, 368–70 (2005).

¹⁴⁸ Cf. MONT. CODE ANN. § 53-21-126(2) (2005) (requiring “proof beyond a reasonable doubt with respect to any physical facts or evidence,” while requiring only “clear and convincing evidence as to all other matters”).

¹⁴⁹ Cf. Michael Louis Corrado, *Punishment and the Wild Beast of Prey: The Problem of Preventive Detention*, 86 J. CRIM. L. & CRIMINOLOGY 778, 792 (1996) (arguing that even when risk is low, if the detained person belongs to the designated risk group, “there is a one hundred percent chance that person presents a risk of harm”).

¹⁵⁰ Christopher Slobogin, *A Jurisprudence of Dangerousness*, 98 NW. U. L. REV. 1, 6–9, 53–58 (2003).

V. THE HELPFULNESS OF PREDICTION TESTIMONY

Under Federal Rule of Evidence 702 and the evidence rules of virtually every state, a witness may not testify as an “expert” unless his or her testimony “assists” or is “helpful” to the fact finder.¹⁵¹ The courts have paid very little attention to this requirement in evaluating the admissibility of expert testimony on dangerousness. In theory, however, helpfulness could be evaluated directly and empirically, assuming the ability to measure the accuracy of a prediction. A test of incremental validity in the prediction context would simply compare the accuracy of predictions made by experts with the accuracy of predictions made by judges and juries unassisted by experts.

To date, no such direct comparison has been made. Several commentators have speculated that, given the inaccuracy of prediction experts, lay people could do just as well as the professionals, at least when compared to decisions using the clinical methodology for prediction.¹⁵² More concretely, an analysis of the earliest studies on prediction efficacy concluded that “a nonclinician furnished with knowledge of past behavior may outperform a mental health professional relying solely on information garnered from a clinical interview.”¹⁵³

If these observations are borne out, clinical prediction testimony might not be sufficiently helpful to be admissible. In the meantime, however, common sense suggests that such testimony can, under certain conditions, be very useful to laypeople inexperienced with prediction. In particular, clinical prediction testimony can apprise them of potential risk factors they would not otherwise contemplate. It is unlikely, for instance, that jurors will be able to discern how a particular person reacts to stress and what types of stress are most likely to trigger violence in that person without some explanation by a qualified clinician.¹⁵⁴

¹⁵¹ FED. R. EVID. 702.

¹⁵² See ALAN A. STONE, MENTAL HEALTH AND THE LAW: A SYSTEM IN TRANSITION 33 (1975) (“[In] many . . . situations a lay person can predict dangerousness at least as well as a professional.” For example, “drug addicts who regularly support their habit by mugging.”); Morse, *supra* note 17, at 620 (reasoning that “[w]ithout hard, methodologically sound quantitative data, the guess of an expert is unlikely to be better than the guess of laypersons”); see also Daniel W. Shuman & Bruce D. Sales, *The Admissibility of Expert Testimony Based upon Clinical Judgment and Scientific Research*, 4 PSYCHOL. PUB. POL’Y & L. 1226, 1228 (1998) (“expert judgments that are clinically derived, as opposed to actuarially derived, are as susceptible to error as lay judgments”).

¹⁵³ Mossman, *supra* note 78, at 790.

¹⁵⁴ See Slobogin, *supra* note 1, at 136 (citing Morse, *supra* note 17, at 600–26).

As this last caveat suggests, however, only clinical prediction testimony that is based on factors derived from the risk literature should be considered helpful. Most mental health professionals are trained to diagnose and treat mental disorder, not predict harmful behavior.¹⁵⁵ As George Dix has argued, a clinician who does not have a good grasp of the research literature on risk assessment should not be permitted to offer a clinical prediction of dangerousness, even if other standard educational or experiential requirements are met.¹⁵⁶ Clinicians like Dr. Kozol are much more likely to meet these threshold requirements than someone like Dr. Grigson, whose testimony is not only deficient for the procedural reasons suggested above, but also because it often consisted of little more than a bow to the type of past behavior that laypeople can evaluate for themselves. Grigson himself appeared to agree with this negative assessment of his “expertise.” As he told a journalist, “I think you could do away with the psychiatrist in these cases. Just take any man off the street, show him what the guy’s done, and most of these things are so clear-cut he would say the same things I do.”¹⁵⁷

Compared to clinical prediction, actuarial and structured professional judgment assessments are more clearly useful to lay fact finders. Both are explicitly based on the risk literature. The specific probability estimates they provide are also beyond the ken of laypeople. Assuming such assessments are material and probative, as those terms were defined in earlier discussion, they are undoubtedly helpful as well.

Whether based on clinical or actuarial information, however, prediction testimony that goes beyond identifying risk factors and probability estimates to assert that a person is “dangerous,” a “continuing threat to society,” or “committable”—in other words, prediction testimony that addresses the ultimate legal issue—is not helpful.¹⁵⁸ The determination as to whether a person presents a risk to society sufficient to justify indeterminate commitment, a lengthier sentence, or the death penalty is solely legal in nature,

¹⁵⁵ See *id.* at 129.

¹⁵⁶ George E. Dix, *The Death Penalty, “Dangerousness,” Psychiatric Testimony, and Professional Ethics*, 5 AM. J. CRIM. L. 151, 175–77 (1977).

¹⁵⁷ John Bloom, *Killers and Shrinks*, TEX. MONTHLY, July 1978, at 68. For the reasons suggested in the text, Dr. Grigson’s career as a prosecution witness was sharply curtailed in 1995, when he was expelled from the American Psychiatric Association. Hugh Aynesworth, *Texas “Dr. Death” Retires After 167 Capital Case Trials*, WASH. TIMES, Dec. 21, 2003, at A2.

¹⁵⁸ See FED. R. EVID. 702, 704 (Rule 704 permits expert testimony on the “ultimate issue,” but Rule 702 requires that such testimony be based on “specialized knowledge” that will “assist the trier of fact to understand the evidence.”).

dependent upon many factors other than those having to do with risk. For instance, a 20% probability that a person will harm others might justify short-term civil commitment but not the death penalty. Other potential legal variables include the magnitude of the harm predicted (rape vs. simple assault), its frequency, and its imminence.¹⁵⁹ Mental health professionals might be able to provide helpful information on each of these points, but they have no specialized insight into whether that information authorizes deprivation of life or liberty; thus they cannot assist the fact finder on that ultimate issue.

Mental health professionals can sometimes helpfully address the *penultimate* issue—that is, whether an individual is “likely” or “substantially likely” to harm another—because that question is an empirical one. However, it is incumbent on the mental health professional to define what he or she means when using these types of terms. Furthermore, if such testimony is not associated with a specific probability estimate based on actuarial data or structured professional judgment, then it should usually be barred as unduly prejudicial for reasons developed in the next Part.

VI. THE PREJUDICIAL NATURE OF CLINICAL PREDICTION TESTIMONY

In his dissent in *Barefoot*, Justice Blackmun explained why he was unwilling to permit capital sentencing juries to hear expert prediction testimony (as opposed to lay testimony about past conduct): “[T]he specious testimony of a psychiatrist, colored in the eyes of an impressionable jury by the inevitable untouchability of a medical specialist’s words, equates with death itself.”¹⁶⁰ Judge Garza, the lone Fifth Circuit judge willing to endorse Justice Blackmun’s prohibition of prediction testimony, expressed similar concerns about its prejudicial impact: “[W]hen a medical doctor testifies that ‘future dangerousness’ is a scientific inquiry on which [he has] particular expertise, and testifies that a particular defendant would be a ‘continuing threat to society,’ juries are almost always persuaded.”¹⁶¹

Of course, the potent influence that Justice Blackmun and Judge Garza attribute to prediction experts would not be a significant concern if the experts were correct most or all of the time. But the research literature makes clear

¹⁵⁹ See ALEXANDER D. BROOKS, LAW, PSYCHIATRY AND THE MENTAL HEALTH SYSTEM 680 (1974) (listing magnitude, probability, frequency, and imminence as four interrelated factors that should be considered in determining dangerousness).

¹⁶⁰ *Barefoot v. Estelle*, 463 U.S. 880, 916 (1983) (Blackmun, J., dissenting).

¹⁶¹ *Flores v. Johnson*, 210 F.3d 456, 466 (5th Cir. 2000) (Garza, J., concurring).

that they are not.¹⁶² If Judge Garza's assertion that lay decision makers are "almost always persuaded" by expert prediction testimony is true, solid ground for exclusion would exist.

The *Barefoot* majority dismissed this concern, confident that the adversary system is "competent to uncover, recognize, and take due account of [prediction testimony's] shortcomings."¹⁶³ Aided by this process, Justice White contended on behalf of the majority, the fact finder will be able to "separate the wheat from the chaff."¹⁶⁴ There is good reason to believe, however, that Justice Blackmun and Judge Garza are correct and Justice White is wrong about the effect of prediction testimony in an adversarial proceeding, at least when it is clinical in nature.

As the *Barefoot* majority noted, cross-examination and rebuttal experts are usually viewed as adequate means of exposing the shortcomings of an expert witness.¹⁶⁵ For instance, the weaknesses of psychiatric testimony supporting an insanity defense can be exposed through vigorous questioning and countering expert opinions. But culpability testimony and prediction testimony differ from one another in a crucial way—one that significantly undermines the efficacy of adversarial techniques in the latter situation. Culpability testimony is proffered by a criminal defendant who has admitted, explicitly or implicitly, the commission of a crime and who must fight *de jure* or *de facto* presumptions of sanity and intentionality.¹⁶⁶ The state's burden on rebuttal is substantially eased by the natural skepticism that such a setting elicits, and thus its use of adversarial techniques is likely to have some success. Prediction testimony, in contrast, is introduced by the government, either at sentencing, which directly follows conviction for a criminal offense, or at a commitment hearing, which is also preceded by a violent act of some sort. In this type of situation, the defense's case in rebuttal is a much tougher sell. Advocates against the state's position must convince the fact finder that the individual will not do again what he or she has just done.

Of course, sentencing and commitment are not the only proceedings in which a litigant must overcome a mindset favoring the government; a criminal trial, initiated by a formal charge against the defendant, is another obvious

¹⁶² See *supra* text accompanying notes 61–63.

¹⁶³ 463 U.S. at 899.

¹⁶⁴ *Id.* at 901 n.7.

¹⁶⁵ *Id.* at 898.

¹⁶⁶ See Christopher Slobogin, *The Structure of Expertise in Criminal Cases*, 34 SETON HALL L. REV. 105, 109–10 (2003).

example. But the potential pro-state bias at criminal trials is counteracted by the presumption of innocence and the high standard of proof and can be further combated with concrete evidence that the defendant did not commit the claimed act or that someone else did. Prediction settings are vastly different: No presumption of safeness exists, either as a legal matter or in the minds of the decision makers; the quantum of proof demanded of the state in sentencing and commitment hearings, in practice, seldom approaches that required at trial, for reasons developed above; and demonstrating that a person will not act a certain way in the future is a much more problematic enterprise than proving he or she did not act a certain way in the past.

To these general points of concern can be added disquieting conjectures about the likely efficacy of specific rebuttal techniques. One might expect that providing the fact finder with information about false-positive rates (which I argued above should always occur) would diminish any tendency to hold the state expert's opinion in undue regard. That practice, while important, fails to counteract the well-known "representativeness heuristic," which leads people to give significantly more weight to case-specific information than to generalized statistics.¹⁶⁷ Fact finders can tell themselves, in many cases quite plausibly, that despite the overall high false-positive rates associated with dangerousness predictions, the state's experts are right this time. Introduction of case-specific information through a rebuttal expert could, in theory, redress that problem. But if, as is often the case when offenders or respondents are indigent, there is no opposing expert,¹⁶⁸ or the expert can, in candor, do little more than identify a few protective factors meant to rebut the state's risk factors, then this stratagem will be unavailing as well. The one study directly on point found that even strong cross-examination and an opposing expert do

¹⁶⁷ Saks and Kidd note that the "consistent overprediction of dangerousness is in part due to experts' insensitivity to the low frequency of such behavior and reliance on the representativeness heuristic wherein the person threatened with commitment is compared to the stereotype of a dangerous person." Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 *LAW & SOC'Y REV.* 123, 133 (1980) (citing Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 *COGNITIVE PSYCHOL.* 430 (1972)). If experts do not respond to generalized data, it is unlikely that laypersons will.

¹⁶⁸ At the time of Barefoot's trial, for instance, indigent defendants in death penalty cases in Texas were entitled to \$500 for "investigation and experts." 463 U.S. at 899 n.5. The Supreme Court has held that indigent defendants are entitled to state-paid assistance on the dangerousness issue, but only to one such expert, who can be a state employee and who does not have to agree with the defense position. *Ake v. Oklahoma*, 470 U.S. 68, 83-84 (1985).

not shake the influence of a state expert willing to pronounce the defendant dangerous.¹⁶⁹

Also suggestive of clinical prediction's power are data on the outcome of proceedings in which it is offered. A survey of hearings under a Maryland criminal commitment program found that, despite the fact that the staff's prediction of dangerousness was virtually always "contested" (albeit not always by opposing experts), judges agreed with its recommendations in 88% of the cases.¹⁷⁰ Similarly, only 31 (or 12%) of the 257 patients that Kozol and his staff originally diagnosed as dangerous were released by the courts,¹⁷¹ many for reasons having nothing to do with perceived risk.¹⁷² And Texas capital sentencing juries virtually never disagreed with Dr. Grigson's clinical predictions, despite cross-examination and rebuttal experts.¹⁷³ At least in these cases, Judge Garza appears to be right that lay fact finders *are* "almost always persuaded" by clinical prediction testimony proffered by the state:¹⁷⁴ a conclusion that is particularly troubling given the very low likelihood that the judges and juries involved in them were correct in concluding that more than 85% of the individuals they committed or sentenced to death would have reoffended if released.¹⁷⁵

One might be tempted to explain these results simply as a demonstration that, when given the opportunity, laypeople will almost always conclude that offenders and individuals subject to commitment are dangerous, regardless of whether experts confirm that view. In contrast to the above data, however, research focusing on the impact of actuarial prediction suggests that the latter type of testimony is much less likely to sway judges and juries. For instance, concordance between the government's expert witnesses and the ultimate

¹⁶⁹ Shari Seidman Diamond et al., *Juror Reactions to Attorneys at Trial*, 87 J. CRIM. L. & CRIMINOLOGY 17, 38–41 & tbl.4 (1996). Diamond and colleagues found that only when the government expert admitted that he might be wrong two out of three times did study subjects tend to change their minds. *Id.*

¹⁷⁰ See Bert H. Hoff, *Patuxent and Discretion in the Criminal Justice System*, 5 BULL. AM. ACAD. PSYCHIATRY & L. 144, 154 (1977).

¹⁷¹ See Slobogin, *supra* note 1, at 147 n.182 (citing Kozol et al., *supra* note 31, at 378).

¹⁷² George E. Dix, *Clinical Evaluation of the "Dangerousness" of "Normal" Criminal Defendants*, 66 VA. L. REV. 523, 540–41 (1980).

¹⁷³ According to one report, Grigson "has testified for the prosecution in at least 140 Texas capital trials; jurors imposed death sentences in more than 98 percent of these cases." AMNESTY INT'L, THE DEATH PENALTY IN TEXAS: LETHAL INJUSTICE 18 (1998), [http://web.amnesty.org/library/pdf/AMR510101998ENGLISH/\\$File/AMR5101098.pdf](http://web.amnesty.org/library/pdf/AMR510101998ENGLISH/$File/AMR5101098.pdf).

¹⁷⁴ *Flores v. Johnson*, 210 F.3d 456, 466 (5th Cir. 2000).

¹⁷⁵ One study found that only 21% of those sentenced to death as dangerous by Texas jurors committed any type of violent act while in prison. James W. Marquart et al., *Gazing into the Crystal Ball: Can Jurors Accurately Predict Dangerousness in Capital Cases?*, 23 L. & SOC'Y REV. 449, 463 (1989).

decision in sexual predator proceedings, where actuarial testimony predominates, is much lower than 85%.¹⁷⁶ Numerous studies have confirmed that clinical prediction testimony is more persuasive to lay decision makers than actuarial testimony,¹⁷⁷ despite the latter's superior accuracy.¹⁷⁸ Of particular import, given the debate in *Barefoot* about the usefulness of the adversary process, is a study finding not only that "jurors weigh clinical opinion expert testimony more heavily than actuarial expert testimony," but that "adversarial procedures may be insufficient to remove this bias."¹⁷⁹ A second study confirmed that, while cross-examination of clinical testimony has little impact on dangerousness ratings, cross-examination of actuarial testimony does reduce those ratings.¹⁸⁰

These findings make sense. Cross-examination and rebuttal experts who attack the risk factors underlying clinical prediction testimony will, at best, be able to suggest in some vague way that the subject is a lower risk than the state's expert believes. In contrast, when risk factors are associated with a precise probability of recidivism, as occurs under the actuarial approach, cross-examination and rebuttal can suggest in more concrete terms how that probability will be lowered if particular factors are not present or particular protective factors are present. Laypeople may also be less likely to give in to their pre-existing inclination to find offenders and commitment respondents dangerous when the expert says, "this person belongs to a group that re-offends at a [particular] rate," rather than simply pronounces that the person is "a high risk" or is "likely to re-offend."

The implications of these observations and research findings are twofold: First, the government should not be permitted to introduce clinical prediction testimony in its case-in-chief. Contrary to the assertion of the *Barefoot* majority, the adversarial process cannot expose effectively the shortcomings of

¹⁷⁶ See Fitch & Hammen, *supra* note 2, at 32 (indicating that, between 1999 and 2001, only 473 individuals had been committed under sexual predator laws, despite several thousand petitions for such commitments).

¹⁷⁷ Daniel A. Krauss & Dae H. Lee, *Deliberating on Dangerousness and Death: Jurors' Ability to Differentiate Between Expert Actuarial and Clinical Predictions of Dangerousness*, 26 INT'L J.L. & PSYCHIATRY 113, 115-16 (2003).

¹⁷⁸ Randy Borum, *Improving the Clinical Practice of Violence Risk Assessment*, 51 AM. PSYCHOLOGIST 945, 946 (1996); William Gardner et al., *Clinical Versus Actuarial Predictions of Violence in Patients with Mental Illness*, 64 J. CONSULTING & CLINICAL PSYCHOL. 602 (1996).

¹⁷⁹ Daniel A. Krauss & Bruce D. Sales, *The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Cases*, 7 PSYCHOL. PUB. POL'Y & L. 267, 305 (2001).

¹⁸⁰ Daniel A. Krauss et al., *The Effects of Rational and Experiential Information Processing of Expert Testimony in Death Penalty Cases*, 22 BEHAV. SCI. & L. 801, 814 (2004).

this type of opinion evidence, with the result that lay decision makers give it too much weight. In short, the prejudicial impact of clinical expert testimony proffered by the state is greater than its relatively low probative value, and thus should lead to its exclusion.

The same balancing analysis suggests the opposite result, however, when the government seeks to introduce prediction testimony tied to empirically based probability estimates associated with actuarial assessment and some structured professional judgment instruments. As earlier parts of this Article demonstrated, the latter type of testimony is both more probative and more helpful than clinical prediction testimony.¹⁸¹ And, as just discussed, testimony based on empirically derived risk estimates is not as likely to be misused by the fact finder. Perhaps, given its quantified nature, laypeople are better able to judge the true import of such testimony, or perhaps they are simply more distrustful of information in the form of data than information that tells an individualized story about the subject of the prediction.¹⁸² Whatever its cause, this differential impact on the fact finder leads to the conclusion that the government should be able to use such testimony in its case-in-chief.

VII. THE SUBJECT-FIRST REGIME

The foregoing analysis does not require a complete ban on clinical prediction testimony. The offender-respondent (hereafter the subject) should still be able to use clinical prediction to undermine the state's claim that he or she is a menace to society. Although clinical prediction testimony is less probative and less helpful than testimony that reports empirically based risk estimates, it is also less likely to have a prejudicial impact when it is proffered by the subject as opposed to the state. That is because, when proffered by the subject, prediction testimony is aimed at dispelling preconceptions, not feeding them. Under such circumstances, it is much less likely to overinfluence the fact finder.

A second reason for this "subject-first" position is that the subjects of prediction hearings should not be prevented from telling the best story they can about their future. Indeed, their interest in advancing their unique case may be entitled to constitutional status. In *Rock v. Arkansas*,¹⁸³ the Supreme Court

¹⁸¹ See *supra* text accompanying notes 78–81.

¹⁸² See Krauss & Sales, *supra* note 179, at 301.

¹⁸³ 483 U.S. 44 (1987).

held that both the Sixth Amendment right to present evidence and the Due Process Clause dictate that testimony proffered by criminal defendants be admitted unless it “is always so untrustworthy and so immune to the traditional means of evaluating credibility that it should disable a defendant from presenting her version of the events for which she is on trial.”¹⁸⁴ Although *Rock* focused on the right of the criminal defendant to testify, it has come to stand for the broader proposition that the Constitution guarantees the accused the right, in *Rock*’s words, “to present his own version of events.”¹⁸⁵ Because *Rock* relied on the Due Process Clause as well as the Sixth Amendment, its reasoning should apply to civil proceedings as well as the “criminal prosecutions” referred to in the latter amendment. Thus, unless demonstrably unreliable or “immune” to adversarial testing, both clinical and actuarial prediction testimony should be admissible when presented by the defense in any type of proceeding.

Under a subject-first regime, that option would be available, although choosing it would open the door to rebuttal using the same kind of evidence. When the state’s actuarial evidence is weak, subjects are unlikely to opt for this strategy. When it is strong, they are more likely to do so, arguing, in effect, that the statistical estimates should be adjusted downward in light of personal characteristics that the actuarial analysis did not take into account.

As I noted in my previous article,¹⁸⁶ the subject-first approach has a time-honored analogue in the character evidence rule.¹⁸⁷ Recognized in some form in every state and the federal courts, this rule prohibits the prosecution from introducing evidence concerning the defendant’s character in a criminal trial unless the defendant does it first. The rule’s prohibition is motivated by the same concerns that were just raised in connection with clinical prediction

¹⁸⁴ *Id.* at 61.

¹⁸⁵ *Id.* at 52. For elaboration of this point, see Katherine Goldwasser, *Vindicating the Right to Trial by Jury and the Requirement of Proof Beyond a Reasonable Doubt: A Critique of the Conventional Wisdom About Excluding Defense Evidence*, 86 GEO. L.J. 621, 636–37 (1998); Janet C. Hoefel, *The Sixth Amendment’s Lost Clause: Unearthing Compulsory Process*, 2002 WIS. L. REV. 1275, 1316–50; and Christopher Slobogin, *The Admissibility of Behavioral Science Information in Criminal Trials: From Primitivism to Daubert to Voice*, 5 PSYCHOL. PUB. POL’Y & L. 100, 114–17 (1999).

¹⁸⁶ Slobogin, *supra* note 1, at 149–50.

¹⁸⁷ FED. R. EVID. 404(a)(1) (“Evidence of a person’s character or a trait of character is not admissible for the purpose of proving action in conformity therewith on a particular occasion, except: (1) Character of accused—Evidence of a pertinent trait of character offered by an accused, or by the prosecution to rebut the same . . .”). Indeed, the advisory committee note to the federal character evidence rule states that the rule “is so deeply imbedded in our jurisprudence as to assume almost constitutional proportions . . .” FED. R. EVID. 404(a)(1) advisory committee’s note.

testimony. Although evidence of bad character shows that the defendant has a tendency to commit bad acts and thus might have committed the current crime, its probative value is thought to be outweighed by the possibility that the fact finder will convict the defendant merely because he or she has been a bad person, rather than because the state has proved its case.¹⁸⁸ Thus, negative character evidence is barred unless the defendant decides that there is enough evidence of good character to counteract it.¹⁸⁹ Analogously, because the risk is great that clinical prediction testimony, although probative of dangerousness, will prompt an erroneous prediction, its introduction should be prohibited unless the subject opens the door to its use.

Accordingly, as I stated in my previous article, “the character evidence rule serves as a well-established precedent for the type of rule proposed here.”¹⁹⁰ There are, however, some unique conceptual and practical issues that arise in applying the analogous subject-first rule to clinical dangerousness testimony. The following three sections explore the most important of these issues.

A. *The Type of Evidence the State May Use to Prove Dangerousness*

Without clinical predictions, the state will be deprived of a primary means of proving dangerousness. But, for the reasons outlined above, it may still rely on empirically based risk estimates that are material to the case. As indicated in Parts II and III, for some populations and some types of harmful behavior, material statistical information is scarce. Thus, courts sometimes will have to make difficult decisions as to whether a risk estimate derived from an actuarial instrument normed on a dissimilar population or using a definition of harm that varies from the applicable legal standard can reasonably be extrapolated to cover the case at hand.

If not, the state should still be able to introduce lay evidence describing prior antisocial acts. The strong consensus of the risk literature is that the number and type of prior violent acts committed by an individual are the factors most germane to a prediction of future behavior.¹⁹¹ The type of

¹⁸⁸ See Benjamin B. Sendor, *The Relevance of Conduct and Character to Guilt and Punishment*, 10 NOTRE DAME J.L. ETHICS & PUB. POL'Y 99, 104–05 (1996).

¹⁸⁹ *Michelson v. United States*, 335 U.S. 469, 476 (1948) (“The overriding policy of excluding such evidence, despite its admitted probative value, is the practical experience that its disallowance tends to prevent confusion of issues, unfair surprise and undue prejudice.”).

¹⁹⁰ Slobogin, *supra* note 1, at 149.

¹⁹¹ *Id.* at 151–52; see also MONAHAN, *supra* note 71, at 71 (“If there is one finding that overshadows all others in the area of prediction, it is that the probability of future crime increases with each prior criminal

proceeding involved will dictate the type of prior acts that might be considered. Thus, if, as the Supreme Court suggested in *Jones v. United States*, larceny is dangerous behavior in the context of insanity acquittee commitment,¹⁹² then evidence of past thefts would be admissible in such proceedings.¹⁹³ On the other hand, such evidence would presumably not be admissible in a capital sentencing proceeding where the focus is the defendant's potential for committing serious bodily harm in the future.¹⁹⁴ Courts will also have to decide, as they do now in sentencing and other contexts, whether evidence short of conviction for an offense is sufficient proof of a particular prior bad act.¹⁹⁵

Finally, of course, if the subject opts to use clinical prediction testimony, the state may do so as well. This follows from the analogy to the character evidence rule, as well as general fairness principles.

B. *When the Subject Opens the Door*

Determining when the subject should be said to have "opened the door" to state use of clinical prediction will not always be easy. Certainly, if the defense puts a clinician such as Dr. Kozol on the stand to discuss risk factors, the state may respond in kind. At the same time, the defense should be able to respond in kind to both actuarial-type prediction testimony and lay testimony about past acts *without* fear of triggering state use of clinical prediction evidence. But other scenarios present harder questions.

act."); Julian V. Roberts, *The Role of Criminal Record in the Sentencing Process*, 22 CRIME & JUST. 303, 316–17 (1997) ("Research on the prediction of criminal behavior has repeatedly demonstrated criminal record to be the single best predictor of future offending.").

¹⁹² 463 U.S. 354, 364–65 (1983).

¹⁹³ *Id.* at 365 n.14 ("[T]o describe the theft of watches and jewelry as 'non-dangerous' is to confuse danger with violence. Larceny is usually less violent than murder or assault, but in terms of public policy the purpose of the statute is the same as to both." (quoting *Overholser v. O'Beirne*, 302 F.2d 852, 861 (D.C. Cir. 1961))).

¹⁹⁴ *Id.* at 152.

¹⁹⁵ After the Supreme Court's decision in *Apprendi v. New Jersey*, 530 U.S. 466, 490 (2000), sentencing courts are no longer permitted to rely on evidence of prior bad acts to enhance a sentence beyond the statutory or guidelines maximum unless they resulted in conviction, are stipulated to by the defense, or were found by a jury beyond a reasonable doubt. It remains to be seen whether this rule applies in commitment hearings, where the Sixth Amendment right to jury trial in criminal cases—the basis for the *Apprendi* ruling—probably does not apply.

In criminal commitment proceedings, a commonly contested threshold issue is whether the subject is mentally disordered.¹⁹⁶ Because this issue focuses on present mental state, defense use of a mental health professional to address it should not authorize state use of clinical prediction testimony. In sexual predator commitment proceedings, however, the issue of whether the subject has a “mental abnormality” is often explicitly defined in terms of whether the condition “predisposes” the individual to commit further violent acts.¹⁹⁷ If the defense decides to use a clinician to address this subject, it is in effect introducing clinical prediction testimony. In such situations, the state should be entitled to respond in similar fashion.

A second, closely related issue that arises in sexual predator proceedings (and perhaps in other commitment proceedings as well) is whether the subject is volitionally impaired. This inquiry appears to be mandated by the Supreme Court’s decision in *Kansas v. Hendricks*,¹⁹⁸ which upheld sexual predator commitment on condition that the state show that the individual is dangerous “beyond [his or her] control.”¹⁹⁹ This language implies that the state must produce some evidence of impulsivity or undeterrability in order to commit under these laws.²⁰⁰ Although one’s propensity to act impulsively might be distinguishable, in a technical sense, from one’s dangerousness, the two concepts are so closely related that, once again, a defense decision to use a clinician on the former issue should permit the state to use a clinician to address either or both.

A third potential door-opening scenario involves sentencing. In sentencing proceedings, particularly capital sentencing proceedings, offenders frequently present clinical evidence about culpability in an attempt to mitigate the disposition. Thus, a defense witness might testify that the subject was suffering extreme mental or emotional distress at the time of the offense or was

¹⁹⁶ See, e.g., FLA. STAT. § 916.15(1) (West 2006) (requiring a showing of “mental illness” as a predicate for commitment of insanity acquittees).

¹⁹⁷ See, e.g., KAN. PROB. CODE ANN. § 59-29a02(b) (West 2006) (defining “mental abnormality” as a “congenital or acquired condition affecting the emotional or volitional capacity which predisposes the person to commit sexually violent offenses in a degree constituting such person a menace to the health and safety of others”).

¹⁹⁸ 521 U.S. 346 (1997).

¹⁹⁹ *Id.* at 358.

²⁰⁰ I have argued that *Hendricks* should be construed to require true “undeterrability,” i.e., a showing that the individual will commit violent acts even when the likelihood of apprehension and significant punishment is very high. Slobogin, *supra* note 150, at 40–48.

unable to appreciate the wrongfulness of his or her criminal acts.²⁰¹ This testimony often will be based on precisely the same type of information an anamnestic approach to clinical prediction might collect. Here, however, the defense's expert is clearly focused on past mental state and culpability issues, not future acts and dangerousness. Such testimony should not authorize state use of clinical prediction testimony.

Finally, both criminal commitment and sentencing proceedings frequently focus on the treatability of the subject.²⁰² This scenario is the most difficult to resolve in a subject-first regime. As I noted in my previous article:

Both dangerousness and treatability assessments involve predictions; when the defense introduces clinical evidence based on the latter type of assessment, it is often suggesting that the individual either is not dangerous or will not be so for long, given proper treatment. Frequently, however, such testimony may be directed purely at the efficacy of specific therapeutic modalities in treating problems that are not related to past behavior (such as depression or vocational deficiencies). [Accordingly, when clinical testimony about treatability is proffered by the subject] the court would need to make a sensitive appraisal of the testimony's scope in order to determine whether the state should be able to use clinical prediction evidence.²⁰³

As an efficiency mechanism, the defense could be required, analogous to common practice with respect to testimony on insanity,²⁰⁴ to give notice to the state and the court whenever it is contemplating using a clinician at criminal commitment or sentencing. In that way, door-opening issues can be sorted out prior to adjudication, and the state can prepare accordingly. Even with this adjustment, a subject-first regime should not involve a major shake-up of current defense practice. Defense use of clinicians on mental abnormality and volitional impairment issues in commitment hearings is and probably will

²⁰¹ For instance, roughly two thirds of state capital sentencing statutes recognize these two conditions as mitigating factors. Ellen Fels Berkman, Note, *Mental Illness as an Aggravating Circumstance in Capital Sentencing*, 89 COLUM. L. REV. 291, 296-98 (1989).

²⁰² See, e.g., ALBERT J. DATZ, A.B.A. CRIMINAL JUSTICE STANDARDS COMM., A.B.A. CRIMINAL JUSTICE MENTAL HEALTH STANDARDS 7-7.4(d), at 419 (1989) ("If the court is persuaded that acquittee will continue to receive the needed treatment or habilitation, it may order . . . that acquittee be released . . ."); MODEL PENAL CODE § 7.01(2)(j) (mitigation possible if "the defendant is particularly likely to respond affirmatively to probationary treatment").

²⁰³ Slobogin, *supra* note 1, at 155-56.

²⁰⁴ Every jurisdiction requires the defense to provide notice of intent to raise an insanity defense. WAYNE R. LAFAYE ET AL., CRIMINAL PROCEDURE 919 (2d ed. 1999).

remain rare, and when the defense does decide to have a clinician testify on those issues, it will probably also want the prediction issue addressed; so, the door will be opened in any event. A similar coincidence of aims will usually be present when treatability is at issue. If not, the defense may often be able to structure the clinician's testimony about treatability so that it does not address dangerousness.

C. When the State May Use a Clinician on Other Issues

The state's prerogatives with respect to experts should mirror the subject's even on issues other than dangerousness. If the defense uses an expert to address dangerousness issues in disguise (predisposing mental abnormality, volitional impairment, or treatability), the state should be able to do so as well. However, if the defense chooses not to use an expert on dangerousness or these related subjects, the state should not be able to do so either. Instead, the state would have to rely on empirically derived probability estimates or lay testimony to address these issues.

The one exception to this rule, consistent with what was said above in connection with the defense's prerogatives, is that the government should be able to present clinical testimony about treatability when it does not directly implicate dangerousness. For instance, in an insanity acquittee commitment hearing, the state should be able to introduce clinical evidence showing that, if the subject is found to be dangerous, treatment would be most efficacious in an inpatient unit rather than on an outpatient basis.²⁰⁵ It also bears iteration that, when considering the treatability issue, the state would always be permitted to present both appropriately normed hard actuarial data on the prognosis for similarly situated patients and information about the particular person's own prior treatment successes or failures, analogous to the actuarial and lay proof of risk that the state is always permitted to adduce.

VIII. CIVIL COMMITMENT

The prohibition on state use of clinical prediction testimony in its case-in-chief and the associated subject-first rule, which I have argued should govern

²⁰⁵ Cf. DATZ, *supra* note 202, § 7-7.4(d) (providing that if the court finds that the only reason an insanity acquittee does not meet the commitment criteria is because of the effect of treatment or habilitation, "the acquittee may be committed unless the court is persuaded by a preponderance of the evidence that the acquittee will continue to receive such treatment or habilitation following release for as long as the treatment or habilitation is required").

sentencing and criminal commitment, probably should not apply to civil commitment, at least at the front end. My previous article identified two reasons for this position: one practical and one evidentiary.²⁰⁶

The practical reason is simply that the proposed regime could probably not be implemented in the emergency commitment setting. A person with mental illness who is evaluated moments after threatening to kill his spouse should not be allowed to go free while the information necessary to make an actuarial or structured professional judgment is collected. In this situation, mental health professionals will have to use their best clinical judgment in deciding immediately whether hospitalization is warranted.²⁰⁷ Given these exigencies, state use of clinical predictions of dangerousness and treatability may be unavoidable. Nor would there be time for the defense to assess its options and give pretrial notice of the type proposed above.

The second reason evidentiary restrictions on prediction testimony might be relaxed in the emergency commitment setting is that the prediction called for in such situations is likely to be more accurate than the long-term predictions that have been the focus of discussion to this point. John Monahan has summarized the theoretical grounds for this assertion as follows:

In emergency commitment . . . there is a small situational and temporal “gap” between the behavior used as a predictor and the outcome that is being predicted. One is directly sampling actions, e.g., threatening words and gestures, that are “as similar as possible to the behavior used on criterion measures,” e.g., fulfilled threats. In violence as in other areas, it is potentially true that “predictions about individual behavior can be generated accurately from knowledge of the environments in which the behavior occurs.”²⁰⁸

If in fact, as research to date suggests,²⁰⁹ clinical predictions based on recent behavior in the community can provide “clear and convincing proof” that the

²⁰⁶ Slobogin, *supra* note 1, at 172–74.

²⁰⁷ Scherr, *supra* note 96, at 15–16.

²⁰⁸ See Slobogin, *supra* note 1, at 174 (citing MONAHAN, *supra* note 71, at 59 (quoting WALTER MISCHEL, PERSONALITY AND ASSESSMENT 164 (1968))).

²⁰⁹ L. Joy Apperson et al., *supra* note 73, at 1378–79 (finding a 25% false-positive rate for short-term predictions in emergency commitment context); Renée L. Binder, *Are the Mentally Ill Dangerous?*, 27 J. AM. ACAD. PSYCHIATRY & L. 189, 197 (1999) (summarizing research indicating that short-term predictions are better than long-term predictions); Ethan S. Rofman et al., *The Prediction of Dangerous Behavior in Emergency Civil Commitment*, 137 AM. J. PSYCHIATRY 1061, 1063 (1980) (producing a 59% false-positive rate but noting that those predicted to be dangerous were immediately medicated and thus that “the probability of the patients in [the] experimental group (who would be unmedicated outside the hospital) committing assaults

individual will behave violently in the near future, the rationale for the subject-first rule is significantly weakened in connection with front-end commitments.

As I noted in my previous article, however, “both the practical and evidentiary arguments against the [subject]-first approach begin to lose their persuasiveness . . . as one moves out of the emergency detention context.”²¹⁰ In many jurisdictions, a commitment hearing designed to check the emergency detention takes place two or more days after the detention.²¹¹ In such cases, there will be ample time for the state to gather the information necessary for an actuarial assessment and for the defense to make a decision as to whether clinical prediction testimony will be proffered. Furthermore, as the “situational and temporal gap” between the precipitating behavior and the predicted outcome (that is, violence in the community) increases, the accuracy of the prediction will decrease and concerns about prejudice will mount. An additional concern is that adversarial protections are notoriously lacking at civil commitment proceedings;²¹² although previous discussion suggested cross-examination and rebuttal witnesses are seldom successful at exposing flaws in clinical prediction testimony, their absence here nonetheless enhances the need for a subject-first rule outside the emergency setting.

CONCLUSION

Expert prediction testimony—whether based on actuarial, structured professional judgment, or clinical assessment—will usually be material, probative and helpful, if certain tenets are followed. When prediction testimony is based on group data, the materiality factor requires courts to pay close attention to norming and criterion variable issues. To ensure that prediction testimony is both adequately probative and optimally helpful, courts should only qualify as experts those mental health professionals who are familiar with the risk literature and modern risk methodologies and who utilize

in the community would have far exceeded 41% without emergency commitment”). See generally Thomas R. Litwack & Louis B. Schlesinger, *Assessing and Predicting Violence: Research, Law and Applications*, in HANDBOOK OF FORENSIC PSYCHOLOGY 205, 224 (Irving B. Weiner & Allen K. Hess eds., 1987) (specifying conditions in which clinical predictions of violence can provide clear and convincing evidence of dangerousness in the commitment context).

²¹⁰ Slobogin, *supra* note 1, at 172–73.

²¹¹ See, e.g., VA. CODE § 37.1-67.3 (2006) (requiring a judicial hearing within 48 hours of admission); CAL. FAM. CODE §§ 6250–57 (Deering 2006); CAL. CT. RULES § 1498 (West 2006) (providing for physician “certification” within 72 hours and a full judicial hearing after 14 days of detention).

²¹² See REISNER ET AL., LAW AND THE MENTAL HEALTH SYSTEM: CIVIL AND CRIMINAL ASPECTS 764–66 (2d ed. 1990).

appropriate assessment procedures; courts should also ensure that those experts they permit to testify describe, or at least respond to, the pertinent error rates and avoid the ultimate legal issue of whether a subject's risk factors require legal action. Finally, even if all of these requirements are met, in sentencing and criminal commitment proceedings courts should exclude testimony that is clinical in nature (rather than grounded in empirically derived probability estimates), unless it is presented by the defense to support a claim of nondangerousness or by the state to rebut the same.

Although obviously focused on testimony about dangerousness, the foregoing analysis also demonstrates a number of more general points about the use of expertise by the courts. The admissibility of expert testimony is heavily influenced by context. Put in evidentiary terms, admissibility should not depend solely on validity estimates, but rather on a nuanced balancing of materiality, probative value, helpfulness, and prejudicial impact. Thus, admissibility of particular expert evidence will vary depending on the party presenting it, the way it is framed, and the proceeding in which it is introduced, as well as on more conspicuous factors such as the reliability of a witness's methodology and the breadth of his or her qualifications.

