1-1-1995

# Dangerous Decisions: An Essay on the Mathematics of Clinical Violence Prediction and Involuntary Hospitalization

Douglas Mossman MD
*University of Cincinnati College of Law*, douglas.mossman@uc.edu

# Dangerousness Decisions: An Essay on the Mathematics of Clinical Violence Prediction and Involuntary Hospitalization

DOUGLAS MOSSMAN

That this Article on the mathematics of violence predictions and involuntary hospitalization decisions is being prepared for a symposium on "Domestic Violence, Child Abuse, and the Law" requires that I begin by providing readers with an explanation of the connection between what may initially appear to be quite different subjects.

Scientific studies and evaluations of mental health professionals' predictions of violence typically have focused on the incidence of violence,[1] the characteristics of perpetrators,[2] and the ability of clinicians to distinguish those persons

1. See, for example, Thomas J. Craig, *An Epidemiologic Study of Problems Associated with Violence among Psychiatric Inpatients*, 139 Am J Psychiatry 1262 (1982); Bruce G. Link, Howard Andrews, and Francis T. Cullen, *The Violent and Illegal Behavior of Mental Patients Reconsidered*, 57 Am Soc Rev 275 (1992); Jeffrey W. Swanson, et al, *Violence and Psychiatric Disorder in the Community: Evidence from the Epidemiologic Catchment Area Surveys*, 41 Hosp & Community Psychiatry 761 (1990). The last two studies both conclude that persons with mental disorders have higher rates of violent and illegal behavior than persons without such disorders.

2. See, for example, Kenneth Tardiff and Attia Sweillam, *Assault, Suicide, and Mental Illness*, 37 Arch Gen Psychiatry 164 (1980); William B. Lawson, Jerome A. Yesavage, and Paul D. Werner, *Race, Violence, and Psychopathology*, 45 J Clin Psychiatry 294 (1984);

95

who will be violent from those who will not.[3] In the vast majority of these reports, the identity of the *victims* of violence is not mentioned. But the relatively few studies that focus on the targets of psychiatric patients' assaults suggest that family members are the most at risk.[4] One study of three hundred patients admitted to a short-term psychiatric ward found that forty-six (fifteen percent) patients assaulted another person in the two weeks before admission, that over half of these forty-six patients had assaulted family members, and "that even after assaulting a family member, the majority of patients planned to return to live with their families after discharge."[5] A recent review of studies about clinical violence prediction found similar rates of recent violence among newly admitted psychiatric patients.[6] Thus, decisions about admissions to psychiatric hospitals frequently have important and immediate implications for the bodily safety (as well as the feelings) of the family members with whom patients live.[7]

Especially in publicly-funded psychiatric facilities, a large fraction of patients are admitted involuntarily in accordance with state civil commitment statutes,[8] which anticipate that mental health professionals will make decisions

---

Kenneth Tardiff, *Characteristics of Assaultive Patients in Private Hospitals*, 141 Am J Psychiatry 1232 (1984).

3. The mental health and criminological literature contains *scores* of publications about this subject. I summarize the results of forty-four studies in Douglas Mossman, *Assessing Predictions of Violence: Being Accurate about Accuracy*, 62 J Consulting & Clin Psych 783 (1994).

4. See, for example, Tardiff, 141 Am J Psychiatry at 1232 (cited in note 2); Andrew E. Skodol and Toksoz B. Karasu, *Emergency Psychiatry and the Assaultive Patient*, 135 Am J Psychiatry 202, 203 (1978); Renee L. Binder and Dale E. McNiel, *Victims and Families of Violent Psychiatric Patients*, 14 Bull Am Acad Psychiatry & L 131 (1986).

5. Binder and McNiel, 14 Bull Am Acad Psychiatry & L at 137 (cited in note 4). Twenty-five patients had committed twenty-nine assaults toward family members: ten assaults were toward a parent, seven toward a spouse, six toward a sibling, three toward a child, and three toward a nephew or niece. Id at 133.

6. Randy K. Otto, *Prediction of Dangerous Behavior: A Review and Analysis of "Second-Generation" Research*, 5 Forensic Rptr 103, 111 (1992).

7. A recent study suggests that family members may account for a large fraction of the victims killed by seriously mentally ill persons. Family members were the victims of fifty-four percent of those individuals found not guilty by reason of insanity in Connecticut over the years 1985-92. By way of comparison, fourteen percent of all persons murdered in the U.S. in 1990 were killed by family members. Carl Sherman, *Half of "'Insane'" Murderers Kill within the Family*, 21 Clin Psychiatry News 2 (Dec 1993) (describing an October 1993 presentation at the American Academy of Psychiatry and the Law's Annual Meeting by Deborah Scott and colleagues).

8. For example, in a study of 666 patients who presented themselves for care at a large, urban psychiatric emergency service, forty-nine percent were hospitalized and sixty-three percent of these admissions were involuntary. Sarah Rosenfield, *Race Differences in Involuntary Hospitalization: Psychiatric vs. Labeling Perspectives*, 25 J Health & Soc Beh 14, 15-16 (1984). Although police involvement is a major factor in predicting whether a petition for commitment will result in involuntary hospitalization, approximately seventy percent of the persons who initiate commitments are family members. Ralph Slovenko, *Criminal Justice Procedures in Civil Commitment*, 28 Hosp & Community Psychiatry 317,

about initiating involuntary psychiatric hospitalization based on judgments about the imminent risk that their patients pose to others.[9] The inadequacy of clinical predictions of violence has been the subject of extensive literature[10] that consistently criticizes psychiatrists' ability to gauge individuals' long-term future dangerousness.[11] Evaluations of this criticism have noted that emergency commitment proceedings are not concerned with *long*-term dangerousness, but with behavior that will occur over a relatively short time following the prediction.[12] As part of the evolution of "a second generation of thought on violence prediction,"[13] recent studies report that involuntarily hospitalized patients have a high degree of violence immediately before and after hospitalization[14] and that professional judgments about these patients' dangerousness

---

819 (1977).

9. State laws generally allow for the hospitalization of mentally ill persons under court order provided that such persons meet the statutory definition of mental illness and, because of that illness, represent a danger to themselves or others. Beginning with *Lessard v Schmidt*, 413 F Supp 1318 (E D Wis 1976), courts across the United States struck down then-existing state commitment statutes as being too broad or too vague. In response, all states eventually passed commitment statutes that made dangerousness to self or others (as opposed to simple need for treatment) a necessary condition for involuntary psychiatric hospitalization. See generally Michael L. Perlin, *Mental Disability Law: Civil and Criminal* §§ 2.06, 2.17-2.19 (Michie 1989 & Supp 1993); Edward Beis, *State Involuntary Commitment Statutes*, 7 Ment Disab L Rptr 358 (1983); Richard Lonsdorf, *The Involuntary Commitment of Adults: An Examination of Recent Legal Trends*, 6 Psychiatric Clin N Am 651 (1983); John Monahan and Saleem A. Shah, *Dangerousness and Commitment of the Mentally Disordered in the United States*, 15 Schizophrenia Bull 541 (1989). These more restrictive statutes have been criticized (chiefly by clinicians and family members) for not making possible the commitment of persons who desperately need treatment but who are not imminently dangerous to themselves or others. See, for example, Committee on Government Policy, Group for the Advancement of Psychiatry, *Forced Into Treatment: The Role of Coercion in Clinical Practice* 31-43, 59-62 (1994). For a discussion and summary of recent statutory changes, see Robert D. Miller, *Need-for-Treatment Criteria for Involuntary Civil Commitment: Impact in Practice*, 149 Am J Psychiatry 1380 (1992). For a scathing criticism of professionals' typical approaches to making commitment decisions, see Michael L. Perlin, *Pretexts and Mental Disability Law: The Case of Competency*, 47 U Miami L Rev 625, 644-52 (1993).

10. For a review, see Perlin, *Mental Disability Law* §§ 2.14–2.15 (cited in note 9).

11. Bruce J. Ennis and Thomas R. Litwack, *Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom*, 62 Cal L Rev 693 (1974); David Faust and Jay Ziskin, *The Expert Witness in Psychology and Psychiatry*, 241 Sci 31 (1988).

12. John Monahan, *The Clinical Prediction of Violent Behavior* 56 (US Dept of Health & Human Serv, 1981); Otto, 5 Forensic Rptr at 106 (cited in note 6).

13. The "second generation" of violence prediction studies derive their designation from Monahan's call for and description of research that would overcome shortcomings discussed in Monahan, *Clinical Prediction* (cited in note 12). See John Monahan, *The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy*, 141 Am J Psychiatry 10, 11 (1984). For a recent summary of "second generation" findings, see Otto, 5 Forensic Rptr at 103 (cited in note 6).

14. Otto, 5 Forensic Rptr at 111 (cited in note 6); Dale E. McNiel and Renee L. Binder, *Violence, Civil Commitment, and Hospitalization*, 174 J Nerv & Ment Disease 107 (1986); A. Michael Rossi, et al, *Characteristics of Psychiatric Patients Who Engage in*

have "a relatively high degree of short-term predictive validity."[15]

The *accuracy* of violence prediction may be less important to practicing clinicians than a closely related but distinct question: when *should* mental health professionals predict violence? Two decades after the initial *Tarasoff* decision,[16] psychotherapists in most jurisdictions[17] can anticipate being held accountable for not having warned victims about, or taken measures to prevent, the behavior of patients whose violence should have been anticipated.[18] Even where their liability has been limited through legislative initiatives that define adequate responses to potentially dangerous patients,[19] clinicians still are expected to take action[20] when the potential for violence, as indicated by some combination of their patients' acts and thoughts, reaches a threshold of "foreseeability."[21] As Dr. Paul Appelbaum has noted, decisions following

---

*Assaultive or Other Fear-Inducing Behaviors*, 174 J Nerv & Ment Disease 154, 157-59 (1986).

15. Dale E. McNiel and Renee L. Binder, *Predictive Validity of Judgments of Dangerousness in Emergency Civil Commitment*, 144 Am J Psychiatry 197, 197 (1987).

16. *Tarasoff v Regents of the University of California*, 13 Cal3d 177, 529 P2d 553 (1974) (*"Tarasoff I"*), reargued 17 Cal3d 425, 551 P2d 334 (1976) (*"Tarasoff II"*). *Tarasoff I* found that psychotherapists have a duty to *warn* potential victims of their patients. *Tarasoff I*, 529 P2d at 553. The California Supreme Court subsequently took the unusual step of rehearing the case and, in *Tarasoff II*, held that psychotherapists have a duty to *protect* potential victims. *Tarasoff II*, 551 P2d at 334.

17. In most states where cases concerning therapists' duty to warn or protect have been heard, courts have found that therapists do have a duty to intervene in some fashion prior to violent action by their patients. See Mark J. Mills, *The So-Called Duty to Warn: The Psychotherapeutic Duty to Protect Third Parties from Patients' Violent Acts*, 2 Beh Sci & L 237, 242-43 (1984). But in some states, courts have considered the issue and have ruled that therapists do not have a *Tarasoff*-like duty. See, for example, *Hopewell v Adibempe*, No. GD73-28756 ([Pa] CP, Allegheny County, 1981); *Shaw v Glickman*, 45 Md App 718, 415 A2d 625 (1980).

18. "After a decade of litigation, the public policy, as it stands to date, dictates that the psychotherapist is required to use reasonable care to protect a third party from a potentially dangerous patient. The psychotherapist should use reasonable care in assessing the patient's potential for violence, identifying and notifying the possible victim or victims, and informing a law enforcement agency, sometimes even when no specific victim can be identified." Mark J. Mills, Greer Sullivan, and Spencer Eth, *Protecting Third Parties: A Decade After* Tarasoff, 144 Am J Psychiatry 68, 71 (1987).

19. By 1989, twelve states had legislative definitions of the duty (with Ohio law precluding liability). Paul S. Appelbaum, et al, *Statutory Approaches to Limiting Psychiatrists' Liability for Their Patients' Violent Acts*, 146 Am J Psychiatry 821, 823 (1989).

20. Such actions may take various forms, including warning the potential victim(s), informing police, attempting to commit the patient, voluntarily hospitalizing the patient, or taking other reasonable measures (e.g., providing medication). In those states where liability has been statutorily defined, laws often specify what interventions will discharge the duty. See Appelbaum, et al, 146 Am J Psychiatry at 824-28 (cited in note 19).

21. For example, in *Tarasoff II*, 551 P2d at 345, the threshold for taking action to protect a third party who is a foreseeable victim of the dangerous patient is reached "once a therapist does in fact determine, or under applicable professional standards reasonably should have determined, that a patient poses a serious danger of

*Tarasoff* have emphasized that a duty to protect arises "only when a threshold of probability is crossed, . . . [but] the terms used to define that threshold have varied, and never has it been specified with any precision."[22]

Although professional standards, and not prediction accuracy, provide the criteria by which *Tarasoff* liability is often judged,[23] these standards ideally should reflect actual features of professionals' predictions. If clinicians are to be faulted for the violent acts of their patients, then evaluations of their prediction-based decisions should reflect our knowledge about the intrinsic characteristics and accuracy of those predictions coupled with our judgments about the moral and societal interests affected by those decisions.

This Article has two major purposes. First, it provides a mathematical description of an ideal procedure for making clinical decisions about patients' future violence, a description that provides a context for evaluating clinicians' "dangerousness decisions." For purposes of illustration, the Article uses a specific clinical situation—deciding whether to hospitaize involuntarily a patient based on his risk of harming another. The Article argues that the decision involves balancing potential risks to third parties (often the patient's family members[24]) with the "massive deprivation of liberty"[25] and other potential harms[26] to the patient that could result from confinement. The mathematical description of the decision procedure consists of a comprehensive method for describing the accuracy of predictions or prediction instruments,[27] a method for assigning values to correct and incorrect predictions,[28] a method for adjusting predictions based on those values,[29] and—most importantly—an ex-

---

violence to others." I know of no psychiatric professional standard that tells clinicians how great a danger must be in order for it to be deemed "serious." For a discussion of the difficulty in specifying a useful *statutory* definition of foreseeability, see Appelbaum, et al, 146 Am J Psychiatry at 822-24 (cited in note 19).

22. Paul S. Appelbaum, *Ask the Experts*, 17 Am Acad Psychiatry L Newsl 19 (1992).

23. In a recent article informing clinicians about violence risk assessment, Dr. Tardiff notes,

Whether psychiatrists can predict violence by patients is a controversial topic. . . . [I]t has been my experience and that of my colleagues who serve as expert witnesses in lawsuits that evidence in a patient's record reflecting consideration of violence is more important than the accuracy of the prediction. That the psychiatrist gathered relevant data, made a decision about the potential for violence and recorded that decision assures the law that there was serious consideration of violence. The law does not expect psychiatrists to be accurate all the time, but it does expect us to try.

Kenneth Tardiff, *How to Recognize a Potentially Violent Patient*, Psychiatric Times 13 (Mar 1993).

24. See notes 1-7 and accompanying text.

25. *Addington v Texas*, 441 US 418, 425 (1979) (citations omitted) (holding that "clear and convincing" evidence of mental illness is needed to meet federal constitutional due process guarantees in involuntary commitment proceedings).

26. See notes 57-59 and accompanying text.

27. See notes 64-66 and accompanying text.

28. See notes 48-63 and accompanying text.

29. See notes 67-70 and accompanying text.

plicit means for expressing uncertainty in those values.[30]

Second, the Article evaluates the actual impact of uncertainty on an ideal decision procedure. When we combine our uncertainty about moral valuations of right and wrong decisions, our uncertainty about base rates, and our uncertainty about the relevant time periods over which predictions should apply, what results is an uncertainty about the correctness of prediction-based decisions that makes most criticism of those decisions untenable.[31] This Article shows that our uncertainty about the factors intrinsic to a hypothetical, best-case prediction procedure usually would preclude valid *post hoc* criticism of wrong decisions about dangerousness; *a fortiori*, most real-life prediction errors also should be beyond criticism.

Demonstrating these points will require me to make explicit and (I hope) non-controversial assumptions about a clinical decision process that is often governed by implicit assumptions or unconscious heuristics.[32] This Article's as-

---

30. See notes 71-86 and accompanying text.

31. See notes 97-103 and accompanying text.

32. "Research has been done on issues such as interclinician agreement, but little on which criteria they actually use in forming their opinions." Vernon L. Quinsey and Anne Maguire, *Maximum Security Psychiatric Patients: Actuarial and Clinical Prediction of Dangerousness*, 1 J Interpersonal Violence 143, 148 (1986). No systematic studies of how clinicians make violence predictions have been published. Thomas R. Litwack, Stuart M. Kirschner, and Renate C. Wack, *The Assessment of Dangerousness and Predictions of Violence: Recent Research and Future Prospects*, 64 Psychiatric Q 245, 266 (1993). However, several authors have reported their observations concerning clinicians' use of faulty decision-making procedures and stereotypes. See, for example, Virginia A. Hiday and Lynn N. Smith, *Effects of the Dangerousness Standard in Civil Commitment*, 15 J Psychiatry & L 433, 449 (1987) (finding that aberrant behavior by a few patients distorted perceptions of outcome, that mental health professionals overstate percentage of commitments that begin as police referrals or that jeopardize staff safety, and that the "drama of a few cases caused their retelling while the mundane cases faded from memory"); see also George E. Dix, *Determining the Continued Dangerousness of Psychologically Abnormal Sex Offenders*, 3 J Psychiatry & L 327 (1975); Stephen J. Pfohl, *Predicting Dangerousness: The Social Construction of Psychiatric Reality* (Lexington, 1978); Christopher D. Webster and Robert J. Menzies, *The Clinical Prediction of Dangerousness*, in David N. Weisstub, 3 *Law and Mental Health: International Perspectives* (Pergamon, 1987).

"'Heuristics' is a cognitive psychology construct that refers to implicit thinking devices that individuals use to oversimplify complex, information-processing tasks. The use of these heuristic devices often leads to distorted and systematically erroneous decisions, and causes decision-makers to 'ignore or misuse items of rationally useful information.'" Perlin, 47 U Miami L Rev at 660 (cited in note 9) (quoting Michael L. Perlin, *Psychodynamics and the Insanity Defense: "Ordinary Common Sense" and Heuristic Reasoning*, 69 Neb L Rev 3, 14-15 (1990)). Examples of such cognitive devices are "vividness" (i.e., memory of a single, highly memorable event outweighs voluminous but bland data), "availability" (i.e., probability of an event is judged based on ease with which we can recall a similar event), and "hindsight bias" (i.e., exaggerating in retrospect the ease with which an event could or should have been predicted). Perlin, 69 Neb L Rev at 15-17 (cited in this note); David B. Wexler and Robert F. Schopp, *How and When to Correct for Juror Hindsight Bias*, 7 Beh Sci & L 485, 487-89 (1989). The major scholarly work on the subject is Daniel Kahneman, et al, eds, *Judgment under Uncertainty: Heuristics and*

sumptions include an open recognition that prediction mistakes are inevitable.[33] In describing a mechanism for equitably balancing the negative consequences of prediction errors, this Article takes the viewpoint that such balancing should reflect a decision-making strategy governed by public attitudes about suffering violence and suffering involuntary hospitalization.[34] Because they are explicit, this Article's assumptions can be evaluated for their normative value or their agreement with empirical findings.

## I. Dangerousness Decisions: A Mathematical Description

### A. DEFINITIONS

This Section describes the implementation and the mathematical properties of prediction methods that might be used to make decisions about future dangerousness.[35] In the following discussion, the term *dangerousness decision* refers to a decision to initiate a certain clinical action—for example, the involuntary hospitalization of a patient—based on an estimate of the likelihood of violence during some future time period. *Dangerousness* refers to the likelihood or probability that someone will act violently. Thus someone can be described as having a "low level" or a "high level" of dangerousness, implying, respectively, a low or a high chance of acting violently. *Violence detection instrument* refers to what might be regarded as a "diagnostic test" of future violence; that is, a method or technique used to sort those persons who will commit acts of violence toward others over some future time period from those who will not. A *prediction of violence* is a judgment that someone's dangerousness is high enough that he should be treated as though he will act violently. A *dangerous* person is one whose dangerousness is high enough to warrant a prediction of violence.

---

*Biases* (Cambridge, 1982).

33. In clinical practice, one generally does not receive unbiased feedback about one's judgments and this may explain why clinicians consistently overestimate the accuracy of their diagnoses and predictions and systematically ignore information that would improve accuracy. David Faust, *Data Integration in Legal Evaluations: Can Clinicians Deliver on Their Premises?*, 7 Beh Sci & L 469, 480 (1989).

34. This decision-making strategy and its empirical investigation are discussed in detail in Douglas Mossman and Kathleen J. Hart, *How Bad Is Civil Commitment? A Study of Attitudes toward Violence and Involuntary Hospitalization*, 21 Bull Am Acad Psychiatry & L 181, 182 (1993) and explained briefly in notes 60-63 and accompanying text.

35. Several writers have lamented the ambiguity and multiple meanings of the term "dangerousness" and have offered clarifying statements similar to those in this paragraph. For an oft-cited example, see Saleem A. Shah, *Dangerousness: A Paradigm for Exploring Some Issues in Law and Psychology*, 33 Am Psych 224, 224-25 (1978).

## B. DEVELOPING A VIOLENCE DETECTION INSTRUMENT[36]

We can understand the theory behind mathematical descriptions of violence detection methods by considering the process of calibrating an airport metal detector. The purpose of the detector is to render a "diagnosis" about passengers who walk through; the condition to be "diagnosed" is whether a passenger is carrying a weapon. Suppose that the sensitivity of the detector is not fixed, that its alarm is controlled by a dial marked from 0 to 100, and that the lower the dial setting, the smaller the amount of metal that will trigger the alarm. If the dial is set very low, the detector will be so sensitive that dental fillings will trigger the alarm; if the dial is set too high, annoying false alarms (for pocket change and belt buckles) will be uncommon, but some small guns and knives might evade detection. Given that false alarms are a small annoyance compared to the problems caused by letting an armed passenger board, airport security personnel will probably set the dial so that false alarms are at the minimum level that still detects just about all weapons.

Mental health professionals do not have "violence detectors" through which their patients walk, but their task in evaluating patients for involuntary hospitalization based on possible future violence poses the same problems faced by airport personnel. We can think of clinicians as having their patients "pass through" an evaluation process that (among other things) attempts to sort those who will be violent from those who will not, and we can imagine that clinicians try to adjust their threshold for admission to effect the best balance between unnecessary hospitalizations (i.e., hospitalizations of persons who would not be violent if released) and letting violent persons go. An adjustable decision threshold is a *general* feature of violence detection methods, and this means that we can describe many different kinds of violence prediction instruments—whether they be rating scales,[37] "actuarial" assessments,[38] blood tests,[39] or even situa-

---

36. This Section summarizes several more extensive expositions, which for readers' sakes I shall not cite repeatedly. For more detailed discussions and explanations, see Eugene Somoza, Louis Soutullo-Esperon, and Douglas Mossman, *Evaluation and Optimization of Diagnostic Tests Using Receiver Operating Characteristic Analysis and Information Theory*, 24 Intl J Biomed Computing 153 (1989); Mossman, 62 J Consulting & Clin Psych at 783 (cited in note 3); Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 183-85 (cited in note 34).

37. A rating scale might rank subjects according to their likelihood of future violence. For an example, see Monahan, *Clinical Prediction* at 68-71 (cited in note 12) (discussing the Michigan Department of Corrections' "Assaultive Risk Screening Sheet").

38. By "actuarial" assessments, I mean those methods that utilize combinations of explicit facts about an individual to estimate his risk of acting violently. A typical example would involve entering demographic facts (i.e., age, sex, or socio-economic status), personal history (i.e., previous arrests or age of first contact with criminal justice system), and psychological factors (i.e., drug abuse, psychosis, or character pathology) in an equation that models probability of acting violently during some future period of time. See, for example, Deidre Klassen and William A. O'Connor, *A Prospective Study of Predictors of Violence in Adult Male Mental Healths Admissions*, 12 L & Human Beh 143 (1988); Deidre

tions where the "instrument" of prediction is a mental health professional who uses "clinical judgment" (an "educated guess") about a patient's future behavior[40]—using a single mathematical framework that captures this adjustability. The framework I shall describe is suitable for use with any definition of violent act, in any population, so long as methodological and statistical rigor are maintained.[41]

For clarity of presentation, I shall refer to a hypothetical future violence test (FVT), which ranks patients' short-term dangerousness from 0 (lowest) to 100 (highest). To simplify our discussion, let us assume that the methodological problems that usually plague investigators[42] can be solved. We assume, for

---

Klassen and William A. O'Connor, *Assessing the Risk of Violence in Released Adult Mental Patients: A Cross-Validation Study*, 1 J Consulting & Clinical Psych 75 (1989). For additional discussion, see Robyn M. Dawes, et al, *Clinical Versus Actuarial Judgment*, 243 Sci 1668 (1989). I review the accuracy of several such assessment methods (and challenge the perceived view that they are more accurate than clinical judgments) in Mossman, 62 J Consulting & Clinical Psych at 783 (cited in note 3).

39. See, for example, Matti Virkkunen, et al, *Relationship of Psychobiological Variables to Recidivism in Violent Offenders and Impulsive Fire Setters*, 46 Arch Gen Psychiatry 600, 602 (1989) (employing discriminant function to predict dangerousness that incorporates measurements of blood glucose and cerebrospinal fluid levels of serotonin metabolites). The relationship between serotonin and impulsive or violent behavior is reviewed in Gerald L. Brown and Markku I. Linnoila, *CSF Serotonin Metabolite (5-HIAA) Studies in Depression, Impulsivity, and Violence*, 51 J Clin Psychiatry 31, 31-41 (Apr 1990 Supp). For another example of using physical factors to predict violence, see E. Kandel, et al, *Minor Physical Anomalies and Recidivistic Adult Violent Criminal Behavior*, 79 Acta Psychiatrica Scandinavica 103 (1989) (finding that external signs of defects in fetal central nervous system development are correlated with violent offenses).

40. Because the dangerousness assessments made by most clinicians do not use any formal tools for measuring likelihood of violence, "clinical judgment" is the most commonly used method for gauging dangerousness in everyday practice. For criticisms of this approach, see Dawes, et al, 243 Sci at 1668 (cited in note 38) (suggesting that actuarial judgments are more accurate than clinical judgments); Thomas Grisso and Paul S. Appelbaum, *Is it Unethical to Offer Predictions of Future Violence?*, 16 L & Human Beh 621, 623-28 (1992) (suggesting that clinicians should avoid yes-no dangerousness predictions and should attempt to make actuarially-based risk assessments). But see Litwack, Kirschner, and Wack, 64 Psychiatric Q at 267-68 (cited in note 32) (qualifying position of Grisso and Appelbaum); Mossman, 62 J Consulting & Clin Psych at 783 (cited in note 3) (showing that properly-evaluated actuarial methods are no more accurate than clinical judgment).

41. I describe a methodologically-perfect imaginary "study" of the accuracy of violence predictions in Mossman, 62 J Consulting & Clinical Psych at 783 (cited in note 3). Among the study's investigators are omniscient observers who are able to determine with certainty which patients act violently after a clinical assessment. Real-life studies are plagued by a host of problems, such as defining what "counts" as violence (verbal threats? drunk driving? hostile gestures?) and determining who actually acted violently, as well as ethical and legal restrictions that make ideal studies of violence prediction impossible. For additional discussion of these methodological issues, see Monahan, *Clinical Prediction* at 50-56 (cited in note 12); Otto, 5 Forensic Rptr at 107-08 (cited in note 6); Litwack, Kirschner, and Wack, 64 Psychiatric Q at 262-66 (cited in note 32).

42. That is, those problems discussed in note 41.

example, that what constitutes a "dangerous act" is crisply defined and that individuals' behavior can be ascertained reliably, so that our knowledge of the "truth" about subjects—i.e., whether they acted violently during the time period—is beyond reasonable doubt.

The first step in evaluating the FVT would involve applying the test to an appropriate group of individuals, who would be monitored over a period of time thereafter. An "appropriate group" would be one comparable to those persons to whom the test ultimately would be applied. For example, if clinicians planned to use the FVT to make hospitalization decisions, they would evaluate its performance on individuals subject to involuntary hospitalization, who could be observed reliably for a period of several days.[43]

Second, the results of the test would be evaluated. The scores of the non-violent subjects would be compared with those who acted violently. Assuming that the FVT had some predictive value, we would expect the non-violent subjects' scores to be lower than those of the violent subjects'. We would also expect that the FVT would not be a perfect predictor of future dangerousness, so that the groups' scores would form overlapping distributions.

For the sake of exposition, let us imagine that the FVT was applied to 2,200 persons of whom 200 acted violently during their observation periods. The distributions of their FVT scores are shown in Figure 1. In looking at Figure 1, one sees that the violent patients, whose distribution peaks at 60, generally scored higher than the non-violent patients, whose distribution peaks at 40. These results suggest that the FVT provides information that is helpful in distinguishing violent from non-violent patients. However, the scores of the violent and non-violent patients overlap, implying that the FVT does not sort the two groups perfectly. The likelihood that a patient will act violently given a certain FVT score is represented by the dashed line in Figure 1. Notice that between 40 and 65—the portion of the scale where many of the violent patients' scores lie—clinicians using the FVT would obtain estimates of dangerousness that were far from certainty.

This imperfection leads us to the *third* part of the FVT's development, which requires choosing some value along the scale on which to base decisions about its future use, some threshold beyond which persons with higher scores might be committed involuntarily. I shall refer to this third step—choosing a decision threshold or operating point—as *operationalizing* the test.

Consider three possible cut-offs—40, 50, and 60—for the rating scale. In each case, persons with scores falling *above* the threshold are test-positive, and those falling *below*, test-negative. As one moves the threshold higher (from 40 to 50 to 60) the kinds of correct identifications change: the fraction of actually violent persons correctly identified by the scale (the scale's *sensitivity* or *true positive rate* (TPR)) decreases, but the probability of correctly identifying a non-violent person (the scale's *specificity* or *true negative rate* (TNR)) increases.

---

43. For an example of one such study, see McNiel and Binder, 144 Am J Psychiatry at 197 (cited in note 15).
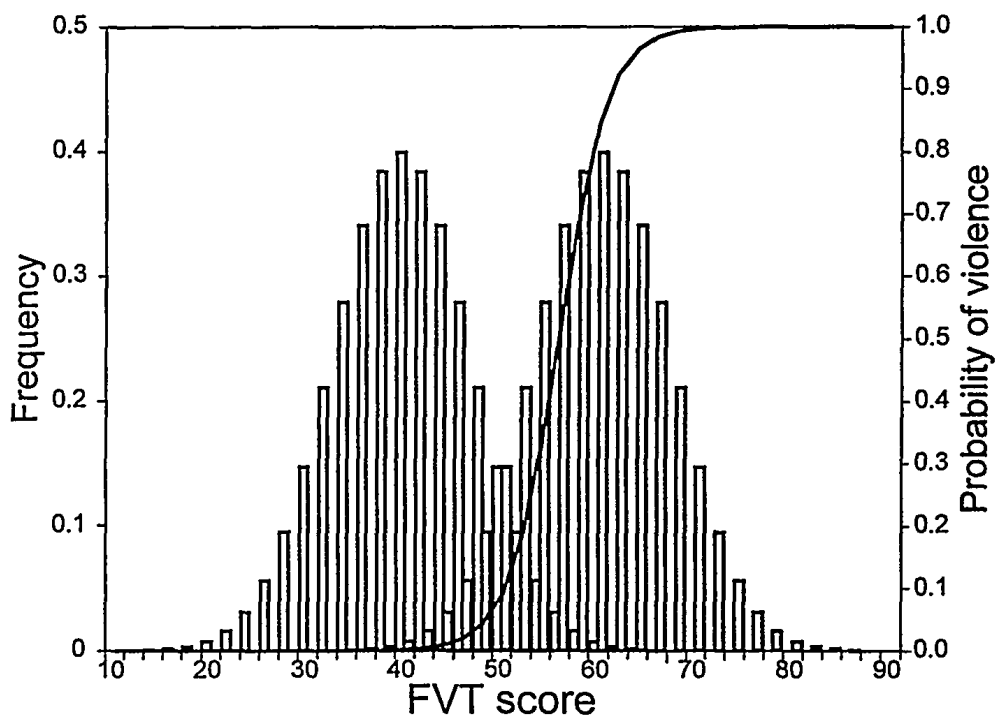
Fig. 1. A hypothetical "Future Violence Test" for predicting violence. The left-most distribution (mean ± s.d. = 40 ± 10) represents test results of individuals who actually are non-violent; the right-most distribution (mean = 60, s.d. = 10), actually violent persons. The S-shaped curve shows the likelihood of violence associated with each FVT score, assuming that the base rate is 0.10.

Similarly, the kinds of misidentifications change: as the threshold is moved higher, fewer patients are wrongly deemed violent (false positives), but more actually violent patients are deemed non-violent (false negatives).

The types of errors and the error rates at three different thresholds are shown in Table 1. Table 1 provides a good summary of diagnostic performance at these thresholds but does not tell us which, if any, of these thresholds is the best one to choose.

### C. OPERATIONALIZING A PREDICTION METHOD[44]

The task of operationalizing a test requires that one effect a balance between sensitivity and specificity by choosing a threshold that reflects the risks and benefits of test outcomes. To put this another way: given that diagnostic

---

44. This section is adapted from a discussion in Eugene Somoza and Douglas Mossman, *"Biological Markers" and Psychiatric Diagnosis: Risk-Benefit Balancing Using ROC Analysis,* 29 Biol Psychiatry 811, 812-18 (1991).

Table 1

Performance of the "Future Violent Test" at Three Decision Thresholds
(Assuming Base Rate of Violence = Pr = 0.10)

| Decision Threshold | 40 | 50 | 60 |
|---|---|---|---|
| Number of True Positives[1] | 195 | 168 | 100 |
| Number of True Negatives[2] | 1,000 | 1,683 | 1,954 |
| Number of False Positives[3] | 1,000 | 317 | 46 |
| Number of False Negatives[4] | 5 | 32 | 100 |
| False Positive Rate (FPR)[5] | 0.500 | 0.159 | 0.023 |
| True Positive Rate (TPR)[6] | 0.975 | 0.841 | 0.500 |
| FP:TP Ratio[7] | 5.1 | 1.9 | 0.46 |
| Marginal Tradeoff[8] | 74 | 10 | 0.74 |

1. "True Positives": violent patients who are correctly identified.
2. "True Negatives": non-violent patients who are correctly identified.
3. "False Positives": non-violent patients who are misidentified as violent.
4. "False Negatives": violent patients who are misidentified as non-violent.
5. FPR = fraction of non-violent patients wrongly predicted to be violent.
6. TPR = fraction of violent patients correctly predicted to be violent.
7. FP:TP Ratio = number of false positives divided by number of true positives.
8. Marginal Tradeoff = number (at this threshold) of addtional false positives needed to correctly identify one additional violent patient. See note 55.

tests—even very accurate ones—are imperfect, diagnostic errors are inevitable. Actual use of tests therefore requires the adoption of a strategy for balancing the consequences of erroneous judgments and the benefits of correct decisions. In the case at hand, this involves balancing the costs of a false-positive prediction (committing a non-violent person) and the costs of a false-negative prediction (not committing a violent person) with the benefits yielded by correct (true-positive and true-negative) predictions.

There is, in theory, a rational way of finding the optimal operating point (OOP), the point on the decision scale that optimally balances the likelihoods and the values of test outcomes.[45] One would like to find the point along the

---

45. Actually, there are *two* rational approaches to finding the OOP. If one prefers all outcomes equally, or if one has no way to assign values and costs to the various outcomes, one might choose a decision threshold that *maximizes information* obtained from the detection instrument. See Douglas Mossman and Eugene Somoza, *Maximizing Diagnostic Information from the Dexamethasone Suppression Test: An Approach to Criterion Selection Using Receiver Operating Characteristic Analysis*, 46 Arch Gen Psychiatry

scale where the expected net gain from the test is a maximum, which obviously would be the OOP for the scale. Once this point is known, a "correct" dangerousness decision would be one that maximized the FVT's net gain. This would be accomplished if persons with scores above the OOP were deemed dangerous (and subject to involuntary hospitalization), and those with scores below, not dangerous. Not all of these judgments would be correct, but they would represent the best balance of errors and correct judgments.

To suggest that a decision should be made so as to maximize net expected gain is to recommend, in a formal way, the use of decision procedure in which all practicing physicians informally engage.[46] In deciding on treatments for almost any ailment, physicians frequently weigh the risks and benefits that will accrue to their patients, taking into consideration the possibility that their diagnoses or decisions may be wrong. Physicians often make an explicit or implicit judgment about the likelihood and seriousness of toxic side effects, the potential costs of leaving a disorder untreated, and the consequences of mistreating due to diagnostic error.[47]

To approach the problem of operationalization more formally, we can examine the following equation:

---

653 (1989); Eugene Somoza and Douglas Mossman, *Comparing and Optimizing Diagnostic Tests: An Information-Theoretical Approach*, 12 Med Dec Making 179 (1992). As I explain in notes 53-59 and the accompanying text, the outcomes of an involuntary hospitalization decision are probably not equal in desirability: most of us would regard the failure to detect a violent person as a worse error than deeming a non-violent person violent. For this reason, cost-benefit balancing is clearly the best conceptual approach to optimizing diagnostic performance.

46. This is not to suggest that physicians always act so as to maximize net expected gain. In some situations, for example, they seem to prefer employing a strategy that minimizes the possible losses or costs. For example, physicians commonly treat viral pharyngitis (i.e., a sore throat) with antibiotics despite diagnostic information that it is not due to a streptococcal infection (i.e., is not "strep throat" and that no benefit will accrue from antibiotic therapy); doctors seem to prefer to adopt a treatment policy that always minimizes the chance that a patient will suffer rheumatic fever, an uncommon sequela of untreated streptococcal pharyngitis. See Scott D. Holmberg and Gerald A. Faich, *Streptococcal Pharyngitis and Acute Rheumatic Fever in Rhode Island*, 250 JAMA 2307 (1983); Jerris R. Hedges, Benita M. Singal, and Jennifer L. Estep, *The Impact of a Rapid Screen for Streptococcal Pharyngitis on Clinical Decision Making in the Emergency Department*, 11 Med Dec Making 119 (1991). This general strategy is termed the "minimax regret," because the decision-maker chooses the option that minimizes the maximum regret associated with all the outcomes. For a discussion of this and other decision strategies, see Simon French, *Decision Theory: An Introduction to the Mathematics of Rationality* 16-17, 36-39 (Ellis Horwood Ltd, 1988); R. Duncan Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* 278-86 (John Wiley & Sons, 1967). The minimax regret strategy is not a viable option for making involuntary admission decisions, since it would dictate admitting *everyone* to avoid the risk of a released person acting violently.

47. This is paradigmatically the case in the treatment of cancer, where both the disease and its therapy can be fatal. For an example and discussion, see J. Robert Andrews, *Benefit, Risk, and Optimization by ROC Analysis in Cancer Radiotherapy*, 11 Intl J Radiation Oncology Biol Physics 1557 (1985).

[1]  *Net Expected Gain* = $(Pr)(\text{TPR})(G_{TP}) + (Pr)(1-\text{TPR})(C_{FN})$
$\quad + (1-Pr)(\text{FPR})(C_{FP}) + (1-Pr)(1-\text{FPR})(G_{TN})$.

The net expected gain from a decision is the sum of four terms, which are themselves the products of the benefits or costs associated with test outcomes ($G_{TP}$ = gain from a true positive, $C_{FN}$ = cost of a false negative, etc.) and the probability of those outcomes. The probability of a true positive test (in this case, a correct prediction of violence) equals the base rate, or prior probability ($Pr$) of violence in the population, multiplied by the TPR, the sensitivity of the test; probabilities for other test outcomes can be calculated similarly. If we can quantify gains and costs, specify the rate of violence, and determine the true and false positive rates associated with each value of the FVT, we can find the threshold that maximizes net expected gain.

## D. ASSIGNING VALUES TO OUTCOMES

The values associated with decision outcomes might be estimated in several ways. In theory, one might calculate the average dollar costs and benefits associated with the outcomes, and use these values in the equation. For example, the cost of a false negative (incorrectly predicting non-violence) would include estimates of medical costs associated with the death or injury of victims, costs to society (e.g., court costs and the cost of incarcerating the perpetrator), the cost of pain and suffering, losses from missing work, malpractice liability, and so on. The problem with this approach should already be obvious: even if we could enumerate all the outcomes associated with a false positive and could determine how likely each one of these outcomes would be given a false positive decision, and even if it were appropriate to rate these outcomes using monetary values, it still would be extremely difficult to agree on what monetary values should be assigned.[48] The practical and conceptual difficulties with this approach[49] have led decision theorists to advocate use of a different quantitative method that rates outcomes according to their relative *utility*.[50] Accordingly, we can rewrite Equation 1 as follows:

[1′]    *Expected Utility* = EU = $(Pr)(\text{TPR})(U_{TP}) + (Pr)(1-\text{TPR})(U_{FN})$
$\quad + (1-Pr)(\text{FPR})(U_{FP}) + (1-Pr)(1-\text{FPR})(U_{TN})$,

---

48. For these reasons, most medical decision theorists find it more meaningful or sensible to use utility theory—discussed further in notes 50-61 and the accompanying text—to rank the relative value of outcomes. Dennis D. Patton and James M. Woolfenden, *A Utility-Based Model for Comparing the Cost-Effectiveness of Diagnostic Studies*, 24 Investigative Radiol 263, 265 (1989).

49. Indeed, for most clinical decisions, systematic consideration of all the separate possible outcomes seems unfathomable. Somoza and Mossman, 12 Med Dec Making at 185 (cited in note 45).

50. At this point in the exposition, I rely on readers' intuitive understanding of the term "utility." In decision analysis, the term has a formal definition, which I explain in note 60 and the accompanying text.

where $U_{TP}$ is the utility associated with a true positive, $U_{FN}$ is the utility associated with a false negative, etc. Customarily, the outcome-associated utilities are rated on a scale, which is set so that 0 represents the worst possible outcome and 1 represents the best possible outcome.[51] Subjective though this method may be, it forces us to assign relative values to the available options and to rank those options in accordance with our fears, hopes, and moral preferences. Pegging the best and worst outcomes at 1 and 0 normalizes the scale and permits relative valuations of intermediate outcomes. Assigning values to outcomes also permits comparison of two or more prediction methods.[52]

If we assume that not detecting an individual who goes on to commit an act of violence (a false negative judgment) is the worst outcome, then $U_{FN}$ is assigned a utility value of 0. Correctly detecting a person who would otherwise commit an act of serious violence (a "true positive") and correctly detecting a non-violent person (a "true negative") both allow for the preservation of public safety without needless infringements on individual liberty. I shall assume in this discussion that both correct outcomes can be assigned a utility value of 1, implying that they are equally desirable. A "false positive" judgment that involuntarily hospitalizes a person who would not have been violent needlessly deprives an individual of his liberty. I shall assume that this outcome is neither as desirable as a correct judgment nor as undesirable as a false negative; therefore, $U_{FP}$ would be assigned a utility value intermediate between 0 and 1.[53] The uncertainties (inter- or intra-personal) in utility estimates can be expressed explicitly by assigning utilities a range, i.e., a value ± error. This allows us to give a mathematical characterization—albeit a rough one—to the impact of uncertainty on the operationalizing of a violence prediction tool.[54]

---

51. Methods for assigning utilities along such a scale have been discussed extensively in the decision analysis literature. See, for example, Luce and Raiffa, *Games and Decisions* at 12-38 (cited in note 46); French, *Decision Theory* at 16-17, 36-39 (cited in note 46); Robert D. Behn and James W. Vaupel, *Quick Analysis for Busy Decision Makers* (Basic, 1982); Robin M. Hogarth, *Judgment and Choice: The Psychology of Decision* (John Wiley & Sons, 1980); Barbara J. McNeil and S. J. Adelstein, *Determining the Value of Diagnostic and Screening Tests*, 17 J Nucl Med 439 (1976).

52. Patton and Woolfenden, 24 Investigative Radiol at 265 (cited in note 48).

53. Readers may differ on this point. One alternative procedure for assigning utilities might look strictly at the outcomes, which are release of a non-violent person (TN), release of a violent person (FN), and involuntary hospitalization (TP, FP). Because they have been hospitalized, the false positives (who would not have been violent had they been released) are indistinguishable from the true positives (those who would have been violent); both groups suffer confinement against their will, which is not as desirable as releasing someone who is non-violent ($U_{TN}$ = 1) or as undesirable as releasing a violent person ($U_{FN}$ = 0). Another alternative procedure might assign the *highest* utility to TP judgments because they both protect the public and provide hospitalization for a person who needs it but will not voluntarily accept it. Fortunately, the main point of this paper is not affected by these different approaches so long as $U_{TP}$, $U_{FP}$, and $U_{TN}$ all have values close to 1, that is, so long as false negatives are seen as much less desirable than the other outcomes. Interested readers may rework my calculations to demonstrate this to themselves.

54. As we shall see at notes 85-86 and the accompanying text, even a rough char-

Monahan has suggested that "it may be possible ethically to justify short-term commitment even if the predictions of imminent violence on which it is based are *less* accurate than the long-term research indicates. Paraphrasing Blackstone, it may be better that ten 'false positives' suffer commitment *for three days* than that one 'false negative' go free to kill someone during that period."[55] I suspect most readers would believe it reasonable to commit more than ten non-violent patients to prevent one patient from murdering someone. Not all acts of violence, however, are murders.[56] Some assaults frighten victims but cause little

---

acterization of this uncertainty is sufficient for describing *crucial* difficulties in actually putting a prediction tool into use.

55. John Monahan, *Strategies for an Empirical Analysis of the Prediction of Violence in Emergency Civil Commitment*, 1 L & Human Beh 363, 370 (1977) (emphasis in original). Other historical authorities have suggested other ratios for wrongful acquittals and convictions. See, for example, Matthew Hale, *Pleas of the Crown: A Methodical Summary* (reprint of 1678 ed, Professional Books, 1972) (suggesting 5:1 ratio); John Fortescue, *A Learned Commendation of the Laws of England* (reprint of 1567 ed, W. J. Johnson, 1969) (suggesting 20:1 ratio). In establishing the appropriate burden of proof in civil commitment hearings, Chief Justice Warren Burger used reasoning very similar to Monahan: "One who is suffering from a debilitating mental illness and in need of treatment is neither wholly at liberty nor free of stigma. It cannot be said, therefore, that it is much better for a mentally ill person to 'go free' than for a mentally normal person to be committed." *Addington v Texas*, 441 US 418, 429 (1979) (citations omitted). For a discussion of the legal literature's treatment of the "balancing" of wrongful commitments and wrongful releases of mentally ill persons, see Perlin, *Mental Disability Law* at § 3.38 (cited in note 9).

Monahan's analytical approach has been applied to other kinds of decisions and is termed the "preferred marginal tradeoff." In the medical context, "[t]his metric is *the number of treatment errors* [i.e., treatments of persons without disease] *that are acceptable in order to treat correctly one additional person with the disease*. In the framework of utility theory, the preferred marginal tradeoff is equivalent to the ratio of the net benefit of treating a diseased person to the net harm of treating a well person, so it is independent of disease prevalence." Peter DeNeef and Daniel L. Kent, *Using Treatment-Tradeoff Preferences to Select Diagnostic Strategies: Linking the ROC Curve to Threshold Analysis*, 13 Med Dec Making 126, 126 (1993) (emphasis in original).

56. A moment's thought reveals that the outcome of a violent act is necessarily ambiguous, and this ambiguity is one of "the conditions under which clinicians must make dangerousness decisions. When evaluating a potentially dangerous patient, clinicians typically are not sure whether, how, or when the patient might engage in violence. And even if a clinician did know, for example, that a patient would use a knife in an effort to harm someone, the clinician could only guess whether the victim would escape harm, receive minor wounds, be severely injured, or be killed." Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 191-92 (cited in note 34).

In their nineteen-month follow-up of 1,938 former psychiatric patients, Cocozza and colleagues found that 183 patients were arrested a total of 230 times. Joseph J. Cocozza, Mary Evans Melick, and Henry J. Steadman, *Trends in Violent Crime among Ex-mental Patients*, 16 Criminol 317, 322 (1978). Thirty-three patients were arrested for violent crimes (i.e., murder, manslaughter, or assault) and seven were arrested for sex crimes. Id. The remaining arrests were for robbery, burglary, property crimes, drug crimes, and "minor crimes." Id. The felony arrest rate per thousand patients for this cohort was 98.50; the arrest rates for violent crimes, potentially violent crimes, and sex crimes were 12.03, 6.18, and 2.60, respectively. Id at 323.

physical harm, and their evil approximates the evil of needless hospitalizations more closely than does an act of murder. Moreover, involuntary hospitalization of persons who, if left alone, would have been harmless is more than a mere annoyance.[57] It can be stigmatizing[58] and damaging to an individual's career, marriage, and nervous system.[59] These considerations suggest that assigning a numerical value to $U_{FP}$ is an inherently imprecise task.

Let us now give this imprecision a mathematical characterization. To do this, we must first define utility more formally.[60] We can state that an individual associates the utility $U_M$ with a state of affairs M when we know that he is indifferent between two alternatives: (1) having that state of affairs M occur or (2) engaging in a lottery in which state of affairs L with known utility $U_L$ has chance $\gamma$ of occurring, and state of affairs N with known utility $U_N$ has chance $(1-\gamma)$ of occurring. In the mathematical terms of formal decision theory, $U_M = \gamma U_L + (1-\gamma)U_N$.

---

A study of 867 patients admitted to Bellevue Hospital from July 1969 through June 1971 found that 202 were arrested during the two years before and two years after admission: thirty were arrested for violent crimes and fifty-five were arrested for potentially violent crimes. Arthur Zitrin, et al, *Crime and Violence among Mental Patients*, 133 Am J Psychiatry 142, 144 (1976). Arrest rates in this sample were slightly higher than those for the general population in the Bellevue catchment area. Id at 147 tbl 6.

57. One way to quantify antipathy toward civil commitment would be to ask persons whether they would prefer to be involuntarily hospitalized or subjected to an act of violence. Mossman and Hart did just this and found (to their surprise) that twenty-eight percent of a group of college undergraduates preferred being attacked by a man wielding a knife to being hospitalized involuntarily for three days. Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 188 (cited in note 34).

58. The Supreme Court notes that civil commitment "can engender adverse social consequences to the individual. Whether we label this phenomenon 'stigma' or choose to call it something else is less important than that we recognize that it can occur and that it can have a very significant impact on the individual." *Addington*, 441 US at 426. In requiring a series of procedural safeguards in the transfer of convicted felons to mental hospitals, the Court emphasized that "the stigmatizing consequences of a transfer to a mental hospital for involuntary psychiatric treatment, coupled with the subjection of the prisoner to mandatory behavior modification as a treatment for mental illness, constitute the kind of deprivations of liberty that requires procedural protection." *Vitek v Jones*, 445 US 480, 494 (1980).

59. Central nervous system damage may result from administration of antipsychotic medications or "neuroleptics." Neuroleptic malignant syndrome is an extremely serious idiosyncratic reaction to these medications that can be associated with seizures, coma, and death. Tardive dyskinesia is a usually-mild disorder that frequently results from lengthy exposure to neuroleptic medications; it is characterized by permanent or long-standing abnormal involuntary movements. George W. Arana and Steven E. Hyman, *Handbook of Psychiatric Drug Therapy* 29 (Little, Brown, 2d ed 1991). The chief use of neuroleptics is the treatment of psychoses, id at 13-19, but they are also employed in a variety of emergencies including those where patients are violent or assaultive. Harold I. Kaplan and Benjamin J. Sadock, *Synopsis of Psychiatry* 560 (Williams & Wilkins, 6th ed 1991).

60. This paragraph presents an abbreviated discussion of formal utility theory. For more extensive introductions, see Luce and Raiffa, *Games and Decisions* at 12-38 (cited in note 46); French, *Decision Theory* at 149-209 (cited in note 46).

From the standpoint of the public at large, a clinician's predictions about violence can result in either: (1) state of affairs L—no one is harmed, which occurs when clinicians make correct positive or correct negative predictions of violence or (2) state of affairs N—a person is harmed by a violent attack following what turned out to be a false negative prediction of violence. But considerations of equity require us not to ignore state of affairs M, the harm done a non-violent person who undergoes needless hospitalization as a result of a false positive prediction of violence. Considerations of equity and fairness require that, in gauging the utility that a person ascribes to a false positive judgment about violence, we ask that person to "universalize" his judgment. We ask him, in other words, to regard any possible outcome affecting *any* individual in society as though it happened to him. An individual's evaluation of the utility of involuntary hospitalization should therefore incorporate the notion that *he* is the one undergoing hospitalization.[61]

Actual durations of pre-hearing involuntary psychiatric hospitalizations vary between jurisdictions and depend on a variety of hard-to-predict factors. Let us assume, however, that the period of involuntary hospitalization under consideration is three to seven days.[62] To find an individual's $U_{FP}$, we would want

---

61. Although the requirements just outlined seem obvious to me, I feel obliged to offer a brief justification. (A complete justification would take me far beyond the intended scope of this Article.) I am at heart a deontological ethicist, and I view these requirements as being mandated by the categorical imperative: "*Act only on that maxim through which you can at the same time will that it should become a universal law.*" Immanuel Kant, *Groundwork of the Metaphysic of Morals* 88 (2d ed 1785, H. J. Paton trans, Harper & Row, 1964) (emphasis in original). Assigning utilities in the manner described—that is, requiring all of us to view ourselves as potentially deprived of liberty—also seems consonant with the spirit of modern contractarian political theory. See, for example, John Rawls, *A Theory of Justice* 11-22 (Harvard, 1971) (explaining that the notion of justice as fairness dictates that we imagine ourselves to be parties in an "original position," in which persons choose principles of self-governance from behind a "veil of ignorance").

This is not to say that Rawls would support the scheme outlined in this Article. He endorses restricting the liberty of mentally ill persons in ways "designed to undo the unfortunate consequences of the imprudent behavior," so long as "with the development or the recovery of his rational powers the individual in question will accept our decision on his behalf and agree with us that we did the best thing for him." Id at 249. He believes, however, that parties to the original agreement, having "no basis for determining the probable nature of their society, or their place in it" would "have strong reasons for being wary of probability calculations if any other course is open to them." Id at 155. One indeed can argue that the "balancing" scheme being discussed is totally at odds with my stated preference for a deontological approach to ethical decision-making. However, given (1) the nature of current commitment laws (which require a determination of a level of dangerousness and which do not permit involuntarily hospitalizations whose sole basis is paternalistic (see note 9), (2) this Article's formulation of the decision problem, and (3) the inevitability of false positive and false negative mistakes, I see no course but to adopt "probability calculations" as part of a decision-making approach that attempts to balance optimally the costs of mistakes and the benefits of correct decisions.

For a description of a method for determining how persons would assign utilities from the standpoint described in the text, see Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 185 (cited in note 34).

62. For example, the Ohio statute requires that a hearing be conducted "within five

to know for what value of $\gamma$ he would be indifferent between (1) being committed for three to seven days (state of affairs M) and (2) engaging in a lottery in which he had chance $\gamma$ of not being attacked (state of affairs L, the result of a true negative decision) and chance $(1-\gamma)$ of experiencing a violent attack (state of affairs N). L and N have utilities $U_L = U_{TN} = 1$ and $U_N = U_{FN} = 0$, and state M is expected to have a utility intermediate between 1 and 0. By definition, $U_M = \gamma U_L + (1-\gamma)U_N$, so $U_M = U_{FP} = \gamma$.

Obviously, finding $\gamma$ for an individual or a population is a matter for empirical determination. Here I shall assume that most persons would give $\gamma$ a value close to 1 (because they would prefer hospitalization unless the risk of attack were small), but they would have trouble giving $\gamma$ a precise value. Moreover, individuals no doubt differ markedly in their feelings about undergoing involuntary hospitalization and being the victim of a violent attack. There is no absolute or "right" value for $\gamma$, since $\gamma$ (and for that matter, all utilities) merely reflect persons' preferences. Therefore, the value of $\gamma$, whether for an individual or for a group of persons, should be described as extending over a *range* of values.

Let us suppose that we can be ninety-nine percent certain that $0.9 \leq \gamma \leq 0.999$. That is, we are virtually sure that persons would prefer being hospitalized involuntarily to taking a ten percent chance that they would be the victim of a serious violent attack, but they would prefer not to be hospitalized if the alternative were taking a 0.1% risk of being attacked. Let $U_x = 1-\gamma = 1-U_{FP}$. $U_x$ then will be 1/10 to 1/1000 of $U_{TP} = U_{TN} = 1$; that is,

[2]   $10^{-3} \leq U_x \leq 10^{-1}$,

with ninety-nine percent probability. Assuming that a population's probability distribution of $U_x$ is log-normal[63] with mean $= 10^{-2}$, we can rewrite the 99% confidence limits as

[3]   $U_x = 10^{-2\pm1}$.

---

court days from the day on which the respondent is detained [involuntarily hospitalized] or an affidavit is filed, whichever occurs first. . . . [F]or good cause shown, the court may order a continuance . . . for no more than ten days" from the date of admission or filing. Ohio Rev Code Ann § 5122.141(B) (Baldwin 1989). In my experience, hearings are usually held between three and seven days after a patient has been involuntarily hospitalized.

63. This distributional assumption is convenient and intuitively sensible when a variable is constrained to a finite interval. This assumption also is consistent with empirical findings measuring persons' values for $U_x$. See Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 187, 190 (cited in note 34).

## E. ROC ANALYSIS: A METHOD FOR DESCRIBING PREDICTION ACCURACY[64]

In addition to the utilities just discussed, Equation 1 refers to false and true positive rates, which are characteristics of the test used to discriminate violent from non-violent patients. Examination of Figure 1 shows that we could choose a threshold for the FVT that would allow FPR or TPR to take on any value between 0 (threshold very high) and 1 (threshold very low). It is therefore valuable to have an explicit means of relating FPR and TPR to each other and to the total expected utility.

The interrelationships of these quantities can be given a precise description through the use of receiver operating characteristic (ROC) analysis. Figure 2 contains a ROC graph representing the FVT (the solid curve labeled 2). In a ROC graph, a test's TPR (sensitivity) is plotted as a function of its FPR (1–specificity). A full ROC curve provides a pictorial description of diagnostic performance across a test's entire range of possible thresholds. In Figure 2, relatively strict hospitalization thresholds (corresponding to high FVT scores) are represented by portions of the curve lying in the lower left corner of ROC space, and relatively lenient hospitalization thresholds (low FVT scores) are represented by the upper right portions of the curve. The ROC curve thus displays a common feature of diagnostic tests: as TPR increases, so does FPR. In other words, *increased* sensitivity is purchased at the cost of *decreased* specificity. Several

---

64. Receiver operating characteristic (ROC) analysis has attained "a central or unifying position in the process of assessing and using diagnostic tools" in clinical medicine. Mark H. Zweig and Gregory Campbell, *Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine*, 39 Clin Chemistry 561 (1993). My Medline database search covering the period 1989 to mid-1993 revealed over two hundred articles about or incorporating ROC methods. The broad range of subjects where ROC analysis has been used to select the optimal test operating point is described in John A. Swets, *The Science of Choosing the Right Decision Threshold in High-Stakes Diagnostics*, 47 Am Psych 522 (1992).

The concepts discussed in this and the following sub-section are developed in more detail in the following series of articles: Eugene Somoza and Douglas Mossman, *Introduction to Neuropsychiatric Decision Making: Binary Diagnostic Tests*, 2 J Neuropsychiatry & Clin Neurosci 297 (1990); Douglas Mossman and Eugene Somoza, *Neuropsychiatric Decision Making: The Role of Disorder Prevalence in Diagnostic Testing*, 3 J Neuropsychiatry & Clin Neurosci 84 (1991); Eugene Somoza and Douglas Mossman, *Neuropsychiatric Decision Making: Designing Nonbinary Diagnostic Tests*, 3 J Neuropsychiatry & Clin Neurosci 197 (1991); Douglas Mossman and Eugene Somoza, *ROC Curves, Test Accuracy, and the Description of Diagnostic Tests*, 3 J Neuropsychiatry & Clin Neurosci 330 (1991); Eugene Somoza and Douglas Mossman, *ROC Curves and the Binormal Assumption*, 3 J Neuropsychiatry & Clin Neurosci 436 (1991); Douglas Mossman and Eugene Somoza, *Diagnostic Tests and Information Theory*, 4 J Neuropsychiatry & Clin Neurosci 95 (1992); Eugene Somoza and Douglas Mossman, *Comparing Diagnostic Tests Using Information Theory: The INFO-ROC Technique*, 4 J Neuropsychiatry & Clin Neurosci 214 (1992); Douglas Mossman and Eugene Somoza, *Balancing Risks and Benefits: Another Approach to Optimizing Diagnostic Tests*, 4 J Neuropsychiatry & Clin Neurosci 331 (1992).

possible thresholds for the FVT depicted in Figure 1 are marked along the FVT's ROC curve in Figure 2.

Figure 2 also contains two other ROC curves (dashed lines) that represent the discrimination capacity of two other hypothetical tests for future violence whose performances can also be compared in Figure 3. By examining the curves in Figure 2, we can learn how the relative performance of diagnostic tests can be evaluated using ROC graphs. One can easily see that, at each FPR, Curve 3 has a higher TPR than Curves 1 or 2. This is equivalent to saying that, at each level of specificity, Test 3 is a more sensitive detector of violence than the other two tests. The test represented by Curve 3 typifies an excellent diagnostic test, in that its ROC curve rises sharply from the lower left corner (FPR=0,TPR=0) and bends to the right very near the upper left corner (FPR=0,TPR=1) of the unit square. A poorly-performing diagnostic test would be represented by a ROC curve that lay close to a diagonal line running from (FPR=0,TPR=0) to (FPR=1,TPR=1) across the unit square. The FVT's ROC curve represents a fairly good test; the test represented by Curve 1 is modestly accurate. Readers may find it interesting to compare the accuracy of the FVT to some radiologic tests: the performance of a conventional chest x-ray would be represented by a ROC curve lying midway between Curves 1 and 2;[65] a computerized tomographic head scan's accuracy in detecting disease[66] would be represented by a ROC curve lying slightly above Curve 2.

### F. FINDING THE OPTIMAL THRESHOLD

We now return to the problem of operationalizing the hypothetical future violence test in Figure 1. I noted above that the ROC curves in Figure 2 depict the performance of the three tests whose performance is also shown in Figure 3. Figure 3 indicates that a test's "quality," i.e., its ability to sort violent from non-violent patients, is a function of its ability to separate the distributions of the two groups' test scores. The best test is the one that most separates these two distributions. It turns out that ROC curves such as those shown in Figure 2 can be summarized by mathematical indices that describe the shape of test results' distributions and the separation between the means (the "peaks") of the distributions.[67] Using these mathematical indices, one can differentiate Equation 1' with

---

65. Jung-Tsuoe Ho and Robert A. Kruger, *Comparison of Dual-Energy and Conventional Chest Radiography for Nodule Detection*, 24 Investigative Radiol 861 (1989).

66. Barbara J. McNeil, et al, *Paired Receiver Operating Characteristic Curves and the Effect of History on Radiographic Interpretation*, 149 Radiol 75 (1983).

67. The indices referred to are derived from the "binormal assumption" of ROC curve fitting. For a broad variety of empirical diagnostic data, when (FPR,TPR) pairs associated with various thresholds are transformed to normal deviates or "z-scores" and are plotted, the thresholds tend to fall along a straight line of the form

[67-1]    $Z_{TPR} = BZ_{FPR} + A$,

where $Z_{TPR}$ and $Z_{FPR}$ are the normal deviates of TPR and FPR, $B$ is the slope of the line,
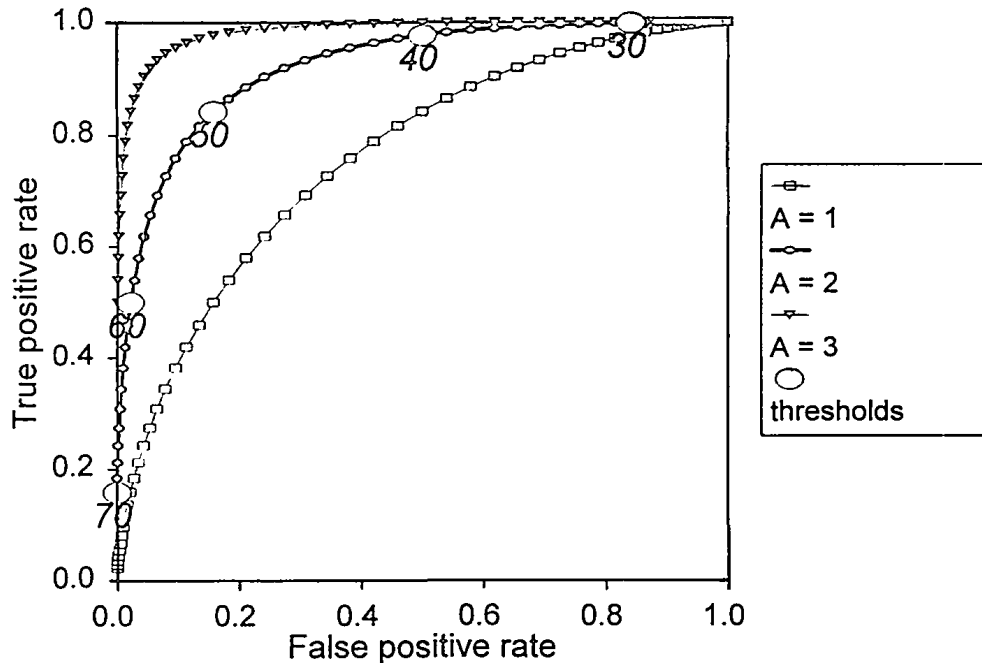
Fig. 2. ROC curves for three tests of differing discriminating ability. The curves represent tests in which the distribution of the violent population is offset by 1, 2, or 3 standard deviations from the distribution of the non-violent population, as shown in fig. 3. The FVT shown in fig. 1 is represented by the middle curve. Several FVT thresholds are marked along the middle curve.

respect to the threshold, set this derivative equal to 0, and derive an analytic solution for the utility-maximizing threshold. For the purposes of this Article, let us assume that, for a sufficiently large population, the FVT results of violent and non-violent patients do conform to two normal (bell-shaped or Gaussian) distributions with different means but *equal* variances.[68] This means that the ac-

---

and $A$ is the intercept. See John A. Swets, *Forms of Empirical ROCs in Discrimination and Diagnostic Tasks: Implications for Theory and Measurement of Performance*, 99 Psych Bull 181 (1986). This finding leads to the assumption that, on some monotonic transformation of the original decision axis, test results conform to two normal distributions whose standard deviations have the ratio $B$ and whose means are separated by $A$ standard deviations. The indices $A$ and $B$ thus summarize the performance of a diagnostic test throughout its entire range of thresholds. For the FVT shown in Figure 1, $A = 2$ and $B = 1$. For additional discussion of the justification and hypothesis testing for the binormal assumption, see Charles E. Metz, *Some Practical Issues of Experimental Design and Data Analysis in Radiological ROC Studies*, 24 Investigative Radiol 234 (1989); James A. Hanley, *The Robustness of the "Binormal" Assumptions Used in Fitting ROC Curves*, 8 Med Dec Making 197 (1988).

68. The use of equal variances greatly simplifies the mathematical discussion that follows, without compromising any of the fundamental conclusions. The more general and
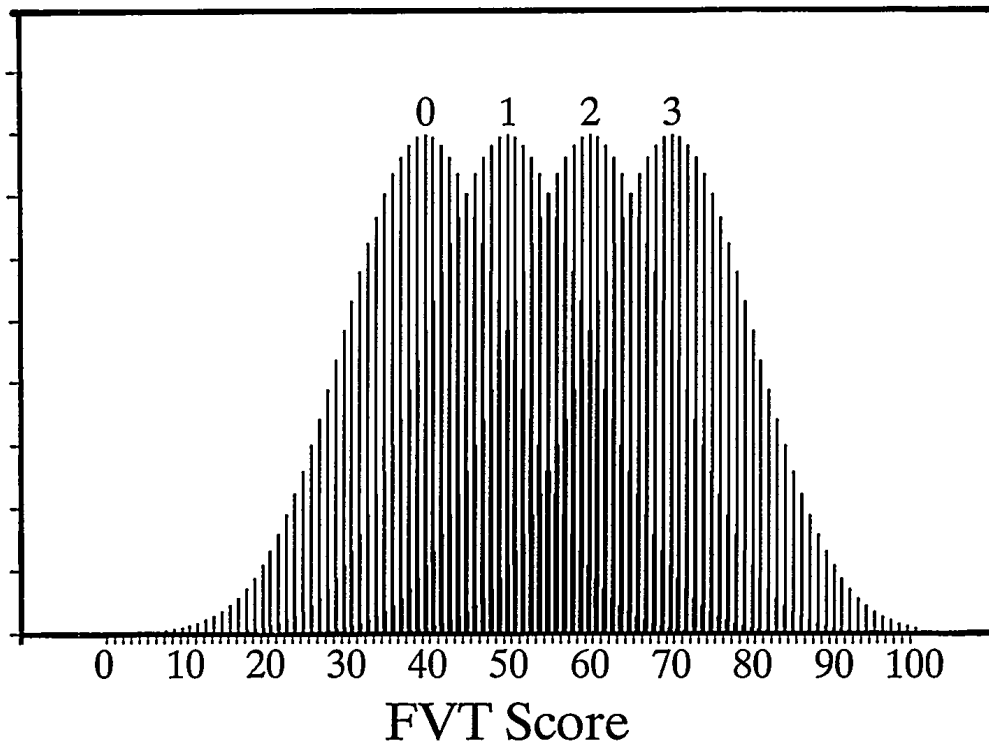
Fig. 3. Separation of violent and non-violent populations by three different Future Violence Tests, whose ROC curves are shown in fig. 2. The non-violent population (labelled 0) is centered at 40; increasingly accurate FVTs separate the violent population by 1, 2, or 3 standard deviations.

curacy of the FVT can be summarized by a single index $A$, the distance between the means of the two groups expressed in units of standard deviation. Let $Z_{FPR}$ be the normal deviate of the false positive rate of the FVT, and recall that $U_{TP}=U_{TN}=1$, $U_{FN}=0$, and $1-U_{FP} = U_x$. Under these conditions, expected utility $EU$ in Equation 1' will be maximized when $\partial EU/\partial Z_{FPR} = 0$. One can show that when this is the case:

$$[4] \qquad Z_{FPR} = -\frac{1}{A}\ln\left(\frac{[1-Pr]}{Pr}U_x\right) - \frac{A}{2}.$$

When the threshold is chosen so that $Z_{FPR}$ satisfies Equation 4, the future violence test will be operationalized to allow the ideal balance of correct and incorrect predictions. To strike this balance, one need only know the base rate

---

somewhat more complicated case—different means and *un*equal standard deviations—is discussed in Somoza and Mossman, 29 Biol Psychiatry at 815-18, 823-24 (cited in note 44).

(*Pr*), the discriminating capacity of the test (represented by *A*), and the relative utility of false positive predictions.[69]

Because the distance in standard deviations between the violent and non-violent distributions is *A*, the true and false positive rates for a diagnostic test will have the relationship $Z_{TPR} = Z_{FPR} + A$.[70] In both Figure 1 and Figure 3, the population distributions have standard deviations equal to 10 FVT points. Thresholds in both figures are easily converted to values of $Z_{FPR}$ using the formula $Z_{FPR} = (40-FVT) \div 10$. The non-violent distribution is centered at FVT = 40, and on the normal deviate scale, the mean of the non-violent population is (by definition) $Z_{FPR} = 0$. In Figure 1, the distributions are separated by two standard deviations, or 20 FVT points, so at FVT = 60, the normal deviate for the mean of the violent population occurs where $Z_{FPR} = [40-60] \div 10 = -2$. Note finally that the difference between mean FVT scores for the violent and non-violent populations is 20, and that $A = [60-40] \div 10 = 2$. This allows us to rewrite Equation 4 for FVT*, the FVT score in Figure 1 that yields maximum utility, as follows:

$$[4^*] \qquad FVT^* = 50 + 5 \ln\left(\frac{[1-Pr]}{Pr} U_x\right).$$

## II. The Impact of Uncertainty and Imprecision

### A. UNCERTAINTY IN UTILITY

The uncertainty in our judgments about utilities of test outcomes has an important effect on how certain we can be that we have chosen the best operating point for a test. A concrete example will illustrate this point. Suppose that having recognized that we have assigned $U_x$ a range of values, we wish to know in what range of values lies the utility-maximizing threshold for the FVT used to make decisions covering a specified time period. Assume that we shall be evaluating emergency room patients for whom the base rate of serious violence during the time period is 0.4% (i.e., four of one thousand patients act violently),[71] and the FVT is quite accurate (*A* = 2) such as the FVT in Figure 1. Using Equations 4 and 4*, and letting $U_x = 0.01$, we find that the violence prediction test is optimized when $Z_{FPR} = -1.456$, $Z_{TPR} = +0.544$, and FVT = 54.6. From a table of the standard normal distribution, we find that, at this cut-off, FPR = (1-specificity) = 0.0727, and TPR = sensitivity = 0.707. In other words, about seventy-one percent of the actually violent individuals would be correctly classified, and about ninety-three percent of the non-violent ones would be correctly identified.

---

69. The full derivation is given in the Appendix.

70. Notice that this is the same as Equation 67-1 in note 67, with *B* = 1.

71. My reason for choosing this base rate will become clear shortly. See text accompanying note 84.

Equations 2 and 3 express the distributional characteristics and the 99% confidence limits of $U_x$, then the 95% confidence limits are

[5] $10^{-2.76} \leq U_x \leq 10^{-1.24}$, or $0.00174 \leq U_x \leq 0.0577$.[72]

The 95% confidence interval for the OOP can be determined by substituting these limits into Equations 4 and 4*, and calculating $Z_{FPR}$, $Z_{TPR}$,[73] and the FVT score. Thus, given our above assumptions, we could be ninety-five percent sure that

[6] $-2.332 \leq Z_{FPR} \leq -0.579$, $-0.332 \leq Z_{TPR} \leq +1.421$, and $45.8 \leq FVT \leq 63.3$,

and that

[7] $0.00985 \leq FPR \leq 0.281$ and $0.370 \leq TPR \leq 0.921$.

One can show that when $Pr = 0.004$, $FPR = 0.0727$, and $TPR = 0.707$, a FVT will classify about twenty-six non-violent persons incorrectly for every correctly-classified violent person. This does *not* imply that the test is inaccurate—indeed, our FVT is, by hypothesis, a good one.[74] The FP:TP ratio is the result of the test's imperfection coupled with the assumed low base rate of violence and a preference for involuntary hospitalization over releasing violent individuals. Notice that our uncertainty about the relative values of outcomes requires us to tolerate a large variation in acceptable test operating points and decision outcomes. A clinician who hospitalizes seventy-five non-violent persons for every violent one may not be making poorer assessments than one whose ratio is 7:1; they may just be using different operating points within the ranges described in Equations 6 and 7.[75]

---

72. In a normal distribution, ninety-nine percent of the population is contained within 2.576 standard deviations of the mean, and ninety-five percent of the population lies within 1.96 standard deviations. If $U_x$ is distributed log-normally and we are 99% sure that $10^{-3} \leq U_x \leq 10^{-1}$, the standard deviation of $\log_{10}U_x$ is $1 \div 2.576 = 0.388$.

We could calculate other confidence intervals similarly. A 95% confidence level is the conventional (though entirely arbitrary) point used in statistical inference to accept or reject hypotheses. See, for example, Wayne W. Daniel, *Biostatistics: A Foundation for Analysis in the Health Sciences* 165-200 (John Wiley & Sons, 1983).

73. The relationship between $Z_{TPR}$ and $Z_{FPR}$ is discussed in note 67.

74. See note 76 and accompanying text (discussing typical accuracy of short-term predictions of violence).

75. An even more striking comparison involves differences in a clinician's preferred marginal tradeoff at the mid-point and extremes of the 95% confidence limits. Recall that this tradeoff represents the number of non-violent patients a clinician would hospitalize to avoid releasing one actually violent patient (see note 55). When $U_x = 10^{-1.24}$, the tradeoff is 17.3; when $U_x = 10^{-2}$, the tradeoff is 100; and when $U_x = 10^{-2.76}$, the tradeoff is 808!

## B. UNCERTAINTY IN BASE RATE

Thus far we have assumed that uncertainty arises only from ambiguity in assigning values to test outcomes. In fact, error in optimizing the FVT also arises from uncertainties in estimates of the base rate ($Pr$) and the accuracy index $A$. In my review of several studies of violence prediction, I found that the reported accuracy for short-term predictions (those covering periods of one to seven days) exhibited accuracies such that $0.34 \leq A \leq 1.78$.[76] Interstudy variation in the inferred value of $A$ comes, in part, from different study methods.[77] For a given violence detection instrument, however, one can show that variance in the accuracy index $A$ can be expected to account for only a small portion of the variance in $Z_{FPR}$.[78] For simplicity's sake, I shall ignore variance in accuracy indices in the following discussion. We shall consider variation in $Pr$ as the sole additional source of potential error in estimating the FVT's OOP.

To assign a value to $Pr$, we should recall that most studies report higher levels of violence during periods shortly after evaluation than during later periods.[79] If a certain fraction of persons committing their first act of violence

---

76. See Mossman, 62 J Consulting & Clin Psych at 789 tbl 4 (cited in note 3). In that study, prediction accuracy is described in terms of the *area under the ROC curve* (AUC), which is the most widely used ROC index of test accuracy. See, for example, John A. Swets and Ronald M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* 31-33 (Academic, 1982). Use of the binormal indices, see note 67 for definition, allows for the development of expressions for specifying specific thresholds such as Equation 4 in the text. The relationship between AUC and the binormal ROC indices is given by

$$\text{AUC} = \Phi\left(\frac{A}{\sqrt{1+B^2}}\right),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. In both this paper and Mossman, 62 J Consulting & Clin Psych at 789 (cited in note 3), I use the assumption that $B = 1$, so that $A$ may be obtained from AUC using the relationship $A = [\Phi^{-1}(\text{AUC})] \times \sqrt{2}$.

77. In general, predictions that were derived from past behavior alone (i.e., was the patient violent just before admission) were more accurate than clinicians' judgments or properly-evaluated discriminant functions. See Mossman, 62 J Consulting & Clin Psych at 789 tbl 4 (cited in note 3).

78. This is because the error in $A$ for an individual test is generally on the order of 25-50% of its value, whereas the errors in $U_x$ and $Pr$ are represented by *orders of magnitude*. Curious readers can rework Equation 4 and demonstrate this to themselves. However, the value of $A$ has a major effect on the impact of errors in $U_x$ and $Pr$ on the OOP. The lower $A$ is, the more uncertainty is produced by a given error in $U_x$ and Pr. See text accompanying note 86, Equations 10-13, and Table 3.

79. See, for example, McNiel and Binder, 144 Am J Psychiatry at 198 (cited in note 15) (finding that 14/101 patients became physically assaultive during first twenty-four hours of hospitalization and that an additional 3/101 became assaultive on days two or three); Robert Tillman, *The Size of the "Criminal Population": The Prevalence and*

is fixed, then the absolute number of persons committing their first violent act will decline with time, and the fraction of patients who become violent should therefore be represented as a curvilinear function of time.[80]

The daily rates of violence per person shown as powers of 10 in Table 2 were calculated from published data by assuming that a fixed fraction of patients who had not previously been violent became violent each day. The rates are expressed as exponentials to emphasize the order-of-magnitude inter-study variation and for calculational purposes discussed below. The enormous variation in implicit or reported rates of violence among various psychiatric and non-psychiatric populations implies that a clinician who wished to estimate the base rate of violence for an emergency room population would have trouble achieving better than order-of-magnitude precision.[81]

---

*Incidence of Adult Arrest*, 25 Criminol 561, 568-69, 571-72 (1987) (graphing the fraction of persons arrested as a function of age); Stuart B. Silver, Marcia I. Cohen, and Michael K. Spodak, *Follow-up after Release of Insanity Acquittees, Mentally Disordered Offenders, and Convicted Felons*, 17 Bull Am Acad Psychiatry & L 387, 397 (1989) (graphing the cumulative percentage of rearrests of released insanity acquittees and mentally disordered offenders).

80. A curvilinear function avoids the impossibility of having more than 100% of a population act violently, which is what would be implied if we assumed that a fixed fraction (say $1/m$) of patients acted violently per time period, and the patients were followed for more than $m$ periods. The general model I use in this Article assumes that $Pr$ is a function of the rate of violence, $r$, over a short time period, and of the number of periods of observation $m$:

[80-1]     $Pr = F(r,m) = 1-(1-r)^m$.

However, if $r \ll 1$, then expanding the polynomial on the right side of Equation 80-1 gives us

[80-2]     $Pr \approx mr$,

which is the assumption used to produce Equation 8 in the text.

One could conceive of other reasonable curvilinear functions to describe $Pr$. For example, if one thought that, even after an extended length of time, only a certain fraction $Pr_0$ of patients would act violently, one might utilize the relationship

[80-3]     $Pr = Pr_0 m/(m + h)$,

where $Pr_0$ and $h$ would be empirically-determined constants. The appropriateness of any particular curvilinear function should be the subject of empirical investigation, but a full treatment of this topic is not possible here.

81. In an effort to explain the apparent discrepancies in Table 2, one might note that the studies with the lowest daily rates of violence look at patients in the community (often following hospitalization) and often rely on arrest data that probably underrepresents actual rates of violent acts. The studies with highest rates, by contrast, principally examine behavior of patients who are hospitalized and who therefore may be more disturbed, more violent, and better observed than if they were living in the community.

To my mind, however, this does little to reduce error in estimating the "true" rate of violence, because it underscores the ambiguity in establishing the truth about a patient's

To simplify the following discussion, I shall assume that the fraction of patients who become violent during a short time period—the "base rate" or prevalence (*Pr*) for the time period — is a linear multiple of the daily rate of violence ($10^D$) and the number of days (*N*) over which the patients are observed.[82] Thus

[8]  $Pr = N \cdot 10^D$ .

If we assume that the 99% confidence interval for the daily rate of psychiatric patients' violent acts was $10^{-3\pm1}$ per patient per day, then $D \pm$ s.d.[83] = -3.0 ± 0.39. (Note that this range estimate is conservative.) Suppose that we can narrow the substantial societal disagreement about the number of days over which professionals' predictions should apply to a 99% confidence interval of one to seven days.[84] Then $N \pm$ s.d. = 4.0 ± 1.16 days, and our central estimate of the

---

behavior. Years of work in public and private hospitals has taught me that what gets interpreted as "violence," and the likelihood of aggressive outbursts by any particular hospitalized patient, clearly are functions of the ward "atmosphere" or "*milieu*," that is, the availability, maturity, and equanimity of the nursing staff, the relationships among staff members of various disciplines, the quality of physician leadership, the pleasantness of the surroundings, medication usage, and the number of highly disturbed patients on the ward. For more systematic empirical confirmation, see Miriam Sheridan, et al, *Precipitants of Violence in a Psychiatric Inpatient Setting*, 41 Hosp & Community Psychiatry 776, 776-77, 779 (1990) (finding that behavior leading to restraint was "more likely to relate to external situations than to the patient's internal psychiatric symptoms," that "[t]he most frequent external event that precipitated restraint was patient-staff conflict," and that "in only two [of seventy-three restrained] patients were [] hallucinations directly related to the events leading to restraint").

All of this is to suggest that not all the events (even if they are termed "assaults") recorded as acts of violence by hospitalized patients are serious (or serious enough to justify taking away someone's liberty), and some may be brought about by events associated with, or patients' responses to, being hospitalized and being around other disturbed patients. Moreover, the high rate of violence in the two weeks before hospitalization reported by Binder and McNiel, 14 Bull Am Acad Psychiatry & L 131 (cited in note 4), likely reflects *causes* of or *desires* for hospitalization, and may not be representative of what patients would do had they been left at home. On the other hand, it seems unreasonable to require that only behavior severe enough to generate an arrest be counted as violent—especially given the biases associated with arrest data. See, for example, Bureau of Justice Statistics, *Reporting Crimes to the Police* 1 (US Dept of Justice, 1985) (finding that crimes against younger persons and offenses that do no involve injury are relatively underreported).

The estimates used in the text are the best I can produce after much contemplation of both the ambiguity of published data and the ambiguity of the moral issues involved. Perhaps my discussion will lead other investigators to efforts to reduce this particular source of uncertainty further. For additional discussion of this issue in the context of assessing predictions of violence, see Monahan, *Clinical Prediction* at 52-56 (cited in note 12).

82. For justification, see note 80.

83. "s.d." = standard deviation.

84. This may reasonably represent the consensus of psychiatrists. See, for example,

Table 2
Calculated Rates of Violence (per Person per Day) in
Psychiatric Populations

| Study | Type/Criterion of Violence | Rate |
|---|---|---|
| Cocozza and Steadman[1] (Baxtrom) | assaults | $10^{-3.81}$ |
| Thornberry and Jacoby[2] (Dixon) | assaults | $10^{-3.99}$ |
| Klassen and O'Connor[3] | arrests | $10^{-3.34}$ |
| Klassen and O'Connor[3] | readmitted or arrested | $10^{-2.83}$ |
| Zitrin, et al[4] | arrests for index offenses | $10^{-3.78}$ |
| Zitrin, et al[4] | arrests, violent offenses | $10^{-4.15}$ |
| Rofman, et al[5] | battery | $10^{-2.37}$ |
| Kozol, Boucher, and Garofab[6] | serious assault | $10^{-4.18}$ |
| Steadman[7] (Patuxent) | violent crime | $10^{-3.42}$ |
| McNeil and Binder[8] | physical assault | $10^{-1.22}$ |
| Cocozza, Melich, and Steadman[8] | violent crime | $10^{-4.48}$ |
| Hiday[10] | arrests for any offense | $10^{-3.13}$ |
| Lidz, Mulvey, and Gardner[11] | hit, strike, or serious violence | $10^{-2.61}$ |
| Link, Andrews, and Cullen[12] | weapons use | $10^{-4.25}$ |
| Link, Andrews, and Cullen[12] | fighting | $10^{-3.79}$ |

1.  J. J. Cocozza and Henry J. Steadman, *Some Refinements in the Measurement and Predictions of Dangerous Behavior*, 131 Am J Psychiatry 1012 (1974).

2.  T. P. Thornberry and Joseph E. Jacoby, *The Uses of Discretion in a Maxium Security Hospital: The Dixon Case*, Presentation to the Annual Meeting of the American Society of Criminology (1978).

3.  Deidre Klassen and William A. O'Connor, *A Prospective Study of the Predictors of Violence in Adult Male Mental Health Admissions*, 12 L & Human Beh 143 (1988).

4.  Arthur Zitrin, et al, *Crime and Violence among Mental Patients*, 133 Am J Psychiatry 142 (1976).

5.  Ethan S. Rofman, et al, *The Predictions of Dangerous Behavior in Emergency Civil Commitment*, 137 Am J Psychiatry 1061 (1980).

6.  H. L. Kozol, R. J. Boucher, and R. F. Garofalo, *The Diagnosis and Treatment of Dangerousness*, 18 Crime & Delinquency 371 (1972).

7.  Henry J. Steadman, *A New Look at Recidivism among Patuxent Inmates*, 5 Bull Am Acad Psychiatry L 200 (1977).

8.  Dale E. McNeil and Renee L. Binder, *Predictive Validity of Judgments of Dangerousness in Emergency Civil Commitment*, 144 Am J Psychiatry 197 (1987).

9.  Joseph J. Cocozza, Mary Evans Melick, and Henry J. Steadman, *Trends in Violent Crime among Ex-Mental Patients*, 16 Criminology 317 (1978).

10.  Virginia A. Hiday, *Arrest and Incarceration of Civil Commitment Candidates*, 42 Hosp Comm Psychiatry 729 (1991).

11.  Charles W. Lidz, Edward P. Mulvey, and William Gardner, *The Accuracy of Predictions of Violence to Others*, 269 JAMA 1007 (1993).

12.  Bruce G. Link, Howard Andrews, and Francis T. Cullen, *The Violent and Illegal Behavior of Mental Patients Reconsidered*, 57 Am Soc Rev 275 (1992).

Tardiff, Psychiatric Times at 13 (cited in note 23) ("I believe that we can predict the short-term potential for violence using a model analogous to that used for predicting suicide potential. 'Short-term' refers to no more than a few days to a week . . . "). However, there is no reason to think that the general public agrees. See, for example, Mossman and Hart, 21 Bull Am Acad Psychiat & L at 187 (cited in note 34) (finding that 46/217 undergraduates thought psychiatrists should have no liability for released patients' violence, but 29/217 thought they should be liable for 180 days or more).

the base rate is $Pr = N \cdot 10^D = 0.004$.

To calculate the variance in the base rate estimate, var(*Pr*), we can use the method of differentials:[85]

$$[9] \qquad \mathrm{var}(Pr) \approx (dPr)^2 \approx \left(\frac{\partial Pr}{\partial D}\right)^2 \mathrm{var}(D) + \left(\frac{\partial Pr}{\partial N}\right)^2 \mathrm{var}(N)$$

$$= \left(N \cdot 10^D \cdot (\ln 10)\right)^2 \mathrm{var}(D) + (10^D)^2 \mathrm{var}(N) \ .$$

If $[\mathrm{var}(D)]^{\frac{1}{2}} = \mathrm{s.d.}(D) = 0.39$ and $[\mathrm{var}(N)]^{\frac{1}{2}} = \mathrm{s.d.}(N) = 1.16$, then var(*Pr*) = 0.0000142.

To estimate the variance in $Z_{FPR}$, we again use the method of differentials:

$$[10] \qquad \mathrm{var}(Z_{FPR}) \approx (dZ_{FPR})^2 \approx \left(\frac{\partial Z_{FPR}}{\partial U_x}\right)^2 \mathrm{var}(U_x) + \left(\frac{\partial Z_{FPR}}{\partial Pr}\right)^2 \mathrm{var}(Pr)$$

$$= \left(\frac{-1}{A \cdot U_x}\right)^2 \mathrm{var}(U_x) + \left(\frac{1}{A \cdot (1-Pr) \cdot Pr}\right)^2 \mathrm{var}(Pr) \ .$$

When $A = 2$, $U_x = 0.01$, $Pr = 0.004$, and $U_x$, $D$, and $N$ have the 99% confidence intervals given above, one can show that var($Z_{FPR}$) = 0.4234.

### C.  CONFIDENCE INTERVALS FOR THE DECISION THRESHOLD

One can use Equations 4 and 4* and the finding that var($Z_{FPR}$) = 0.4234 to show that the 95% confidence interval for the FVT's OOP would be

$$[11] \qquad Z_{FPR} \pm 1.96 \times [\mathrm{var}(Z_{FPR})]^{\frac{1}{2}} = Z_{FPR} \pm 1.96 \times [0.4234]^{\frac{1}{2}} = -1.456 \pm 1.275.$$

Therefore,

$$[12] \qquad -2.73 \le Z_{FPR} \le -0.181, \ -0.731 \le Z_{TPR} \le +1.82,$$
$$\text{and } 41.3 \le FVT \le 67.3,$$

and

$$[13] \qquad 0.0032 \le FPR \le 0.428, \text{ and } 0.232 \le TPR \le 0.966 \ .$$

---

85. Regina C. Elandt-Johnson and Norman L. Johnson, *Survival Models and Data Analysis* (John Wiley & Sons, 1980). Equations 9 and 10 are derived from squaring the first terms in the Taylor series expansion. See Angus E. Taylor and W. Robert Mann, *Advanced Calculus* 222-24 (John Wiley & Sons, 2d ed 1972). Equations 9 and 10 do not include "cross terms" because the covariances of $D$ and $N$ and $U_x$ and $Pr$ can be expected to be zero.

Thus far we have examined the effect of error on operationalizing a fairly accurate test, where $A = 2$. As was noted earlier,[86] however, studies of violence prediction suggest that accuracy usually falls well below this level. If $A = 1$, and $Pr$, $U_x$, $D$, and $N$ are unchanged, the 95% confidence interval for the OOP is expanded, and Equations 11-13 become:

[11a]    $Z_{FPR} \pm 1.96 \times [\text{var}(Z_{FPR})]^{\frac{1}{2}} = -1.412 \pm 2.551;$

[12a]    $-3.96 \leq Z_{FPR} \leq +1.139, -2.96 \leq Z_{TPR} \leq +2.239,$
            and $28.6 \leq \text{FVT} \leq 79.6;$

[13a]    $0.000038 \leq \text{FPR} \leq 0.872,$ and $0.0015 \leq \text{TPR} \leq 0.987.$

Table 3 shows how dangerousness decisions for a population of 100,000 would be affected by moving the hospitalization threshold across the 95% confidence ranges when $A = 1$ or 2, assuming 400 (0.4%) of patients actually were violent. For the less accurate test, virtually any cut-off falls within the range of acceptable thresholds, including one that detects only 0.15% of the violent persons and another that deems eighty-seven percent of the non-violent persons violent. The 95% confidence range for the more accurate test (where $A = 2$) requires identification of at least twenty-three percent of the violent individuals, that is, the twenty-three percent having the highest scores on the FVT. In other words, very accurate predictors who miss seventy-seven percent of violent individuals would be performing within the 95% confidence interval; they would also be *mis*identifying 3.4 non-violent individuals for each correct prediction of violence. Equally accurate but more "cautious" predictors would misidentify 110 non-violent individuals for each correctly identified violent individual, and yet they also would be operating within the confidence limits. Notice that even at this high FP:TP ratio, these cautious predictors would still miss 3.5% of the violent individuals.

## III. Discussion

The following sections suggest several points about the description, evaluation, accuracy, and uses of violence predictions.

### A. DESCRIBING PREDICTION ACCURACY

#### 1. Use of ROC methods.

In future studies, the accuracy of violence predictions should be described using ROC methods.[87] This Article demonstrates two important reasons for

---

86. See note 76 and accompanying text.
87. For a full treatment of this methodological issue, see Mossman, 62 J Consulting & Clin Psych at 783 (cited in note 3).

Table 3

Acceptable Decision Thresholds, FVT Scores, and Numbers of Correct and Incorrect Predictions of Violence When $U_x \pm$ s.d. $= 10^{-2} = 0.39$, $Pr = N \cdot 10^D$, $N \pm$ s.d. $= 4 \pm 1.16$, and $D \pm$ s.d. $= -3 \pm 0.39$; Upper Limits, Means, and Lower Limits of the 95% Confidence Intervals When $A = 1$ and $A = 2$.

| | | | 100,000 Evaluated Patients | | | | |
|---|---|---|---|---|---|---|---|
| | | | 99,600 Non-Violent | | 400 Violent | | |
| A | $Z_{FPR}$ | FVT Score | True Negatives | False Positives | True Positives | False Negatives | FP:TP Ratio |
| 2 | -2.73 | 67.3 | 99,281 | 319 | 93 | 307 | 3.4 |
| | -1.456 | 54.6 | 92,363 | 7,237 | 283 | 117 | 25.6 |
| | -0.181 | 41.8 | 56,971 | 42,629 | 386 | 14 | 110 |
| 1 | -3.962 | 79.6 | 99,596 | 4 | 0.6 | 399.4 | 6.7 |
| | -1.412 | 54.1 | 91,734 | 7,866 | 136 | 264 | 57.8 |
| | +1.138 | 28.6 | 12,705 | 86,895 | 395 | 5 | 220 |

this. First, ROC analysis captures one of the essential features of violence predictions—the trade-offs between sensitivity and specificity inherent in those predictions. It thereby provides an assessment of intrinsic accuracy that is independent of biases favoring sensitivity (identification of violent individuals) or specificity (identification of non-violent individuals). Second, ROC methods provide a means for analyzing the relationship between sensitivity and specificity distinct from preferences for certain outcomes and from the prevalence of violence.[88]

## 2. Misinterpretation of prediction research.

The oft-repeated observation that clinicians' predictions of violence are wrong two-thirds to ninety-five percent of the time[89] is misleading, because it ignores the distinction between intrinsic accuracy and a preference for certain types of results—usually a bias in favor of detecting or preventing violence.[90]

---

88. Readers interested in a more detailed explanation of these general issues should consult my articles cited in note 64 and Swets and Pickett, *Evaluation of Diagnostic Systems* at 15-45 (cited in note 76).

89. See, for example, Monahan, *Clinical Prediction* at 48 tbl 3 (cited in note 12) (finding that fifty-nine percent to eighty-six percent of predictions that patients would be violent were wrong); Monahan, 141 Am J Psychiatry at 11 (cited in note 13) (characterizing such statements as "ACLU-type evaluations of the field" of violence prediction); Bruce Ennis and Richard Emery, *The Rights of Mental Patients* 20 (Avon, 1978) (finding that "predictions of dangerous behavior are wrong about 95% of the time"); Amicus brief of the American Psychiatric Association, *Barefoot v Estelle*, 463 US 880 (1983); *Barefoot*, 463 US at 916 (Blackmun dissenting).

90. As Monahan, *Clinical Prediction* at 48 tbl 3 (cited in note 12), shows, clinicians' predictions of *non*-violence are correct most of the time. For an excellent discussion of this

The impórtance of this distinction was demonstrated in the previous section, where we found that, if experiencing needless commitment is deemed a much lesser harm than experiencing violence, there *ought* to be a high ratio of false to true positive violence predictions.[91] When a low base rate of violence is coupled with a preference for protecting the public, then even the use of a very accurate FVT should yield many incorrect predictions of violence for each correct one.

### 3. Predictions are modestly accurate.

Re-examining earlier and "second generation"[92] studies of violence prediction suggests that clinicians are able to distinguish violent from non-violent individuals with a modest degree of accuracy.[93] Recent studies have shown a higher fraction of correct predictions than earlier reports,[94] but only because the base rate of violence was much higher.[95] Ofttimes, statistical methods have appeared to be more accurate than clinical methods, but only because these "predictions" were based on retrospective analyses of data, using "prediction"

---

and other ways that prediction research has been misinterpreted, see Albert W. Alschuler, *Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process*, 85 Mich L Rev 510, 539-46 (1986). In addition to the problems discussed by Professor Alschuler, a good deal of confusion has resulted from authors' ambiguous or inconsistent use of epidemiological terminology. For example, in Monahan, *Clinical Prediction* at 48 (cited in note 12), the term "percent false positive" (%FP) refers to the likelihood that a prediction of violence will be incorrect (%FP = 100 × {FP + [FP + TP]}); in Otto, 5 Forensic Rptr at 103 (cited in note 6), the term "false positive rate" (FPR) refers to the number of false positive predictions divided by the sum of the false positive and true *negative* predictions (FPR = FP + [FP + TN]). In this Article, I use the same definition of FPR as Otto, and I use the FP:TP ratio to refer to the notion that Monahan was addressing. For further discussion, see Stephen D. Hart, Christopher D. Webster, and Robert J. Menzies, *A Note on Portraying the Accuracy of Violence Predictions*, 17 L & Human Beh 695, 697 (1993).

91. The conclusion is premised on the notion that it is legitimate to attempt to balance deprivations of liberty with the harm they prevent. See note 61 and accompanying text. For alternative views, see Henry J. Steadman, *The Right Not to Be a False Positive: Problems in the Application of the Dangerousness Standard*, 52 Psychiatric Q 81, 84-86, 95-98 (1980) (arguing that false positive rates exceed any accepted criminal evidentiary standards and must be the subject of additional policy analysis); Leslie T. Wilkins, *Current Aspects of Penology: Directions for Corrections*, 118 Proc Am Phil Soc 235 (1974) (arguing that for persons without a history of violence, no false positives should be tolerated).

92. For reviews and discussion, see generally Monahan, 141 Am J Psychiatry at 10 (cited in note 13); Otto, 5 Forensic Rptr at 103 (cited in note 6); Mossman, 62 J Consulting & Clin Psychology at 783 (cited in note 3) (discussing distinction between first- and second-generation studies).

93. See Mossman, 62 J Consulting & Clin Psych at 787, 789 tbls 3-4 (cited in note 3).

94. See, for example, McNiel and Binder, 144 Am J Psychiatry at 198 (cited in note 15) (finding that sixty-one percent to seventy-one percent of predictions were correct); Klassen and O'Connor, 12 L & Human Beh at 153 (cited in note 38) (finding that eighty-five percent of classifications were correct).

95. Mossman, 62 J Consulting & Clin Psych at 787 tbl 3 (cited in note 3).

equations tailored to that data. Where cross-validation has been used to assess statistical techniques, their performance appears comparable to clinical judgments.[96]

### 4. Criticizing clinicians' judgments.

These observations should not allow us to lose sight of an important implication of the decision-making model described above: even if clinicians could agree on the best method for assessing violence, almost all decisions (for example, to hospitalize involuntarily, to warn, or to release from custody) based on judgments about the likelihood of future violence should be deemed acceptable. This is the case not because such decisions are likely to be wrong, nor because clinicians utterly lack ability to assess dangerousness. Rather, the task of judging whether a decision was made incorrectly is clouded by our collective ambiguity and disagreement about the value of right and wrong decisions and the time periods over which those decisions should apply.[97]

### B. IMPLICATIONS FOR CLINICIAN LIABILITY

Let me clarify this last point by asking the reader to imagine a lawsuit brought against a psychiatrist by someone who was injured by a patient seven days after the patient had been seen for an emergency room evaluation. The plaintiff claims the patient, who had psychotic symptoms and seemed hostile, should have been deemed dangerous and involuntarily hospitalized. Let us imagine testimony by two experts who are honest and objective, who agree with the defendant on how violence should be predicted (all three use a FVT of typical accuracy [$A = 1$]), but who exemplify the moral uncertainty described in previous sections. Let us also suppose that the jury is comprised of honest, objective citizens unswayed by hindsight bias,[98] irrational fears of the mentally ill,[99] or desire to compensate victims.

---

96. Id at 789 tbl 4.

97. Thus, this Article has provided a mathematical justification for Tardiff's belief that courts expect psychiatrists to collect data and give serious consideration to violence potential, but do not expect dangerousness decisions to be right all the time. Tardiff, Psychiatric Times at 13 (cited in note 23). But this Article calls into question the reasonableness of this expectation: if the use of almost any decision threshold would be acceptable, why bother gathering data or letting dangerousness play a role in clinical decision-making?

98. See note 32 and the sources cited therein.

99. "Mental patients are feared, in part because they are thought to be violent." Charles W. Lidz, Edward P. Mulvey, and William Gardner, *The Accuracy of Predictions of Violence to Others*, 269 JAMA 1007, 1007 (1993) (citation omitted). See also Henry J. Steadman and Joseph J. Cocozza, *Selective Reporting and the Public's Misconceptions of the Criminally Insane*, 41 Pub Op Q 523 (1978). Several years ago, however, Henry Steadman pointed out that research data supported public fears, and that mental health professionals should not attempt "to assuage prospective neighbors of a hostel [for psychiatric patients] by assuring them that mental patients are less dangerous statistically than their present neighbors." Henry J. Steadman, *Critically Reassessing the Accuracy of Public Perceptions of the Dangerousness of the Mentally Ill*, 22 J Health & Soc Beh 310,

The expert retained by the plaintiff argues that the patient displayed several features indicative of future violence, that the patient scored 55 on the FVT, and that hospitalization for potential violence should occur when a patient's score exceeds 50. This witness acknowledges that this cut-off yields a large number of needless hospitalizations, but argues that the rate of violence among psychiatric patients is higher than that of the general population, and that psychiatrists have an obligation to use their legal authority in this risk-averse manner.

A defense expert agrees that the patient scored 55 but argues that this implies the patient should *not* have been hospitalized because the appropriate FVT cut-off is 60. The defense expert believes that dangerousness assessments are valid for only a short time horizon (i.e., one to two days), that the base rate of violence is lower than that stated by the plaintiff's expert, and that clinicians should attend to the deprivation of liberty attendant to involuntary hospitalization. He also points out that reasonable people disagree about these issues and that by not hospitalizing the patient, the defendant used a decision strategy that lay well within the limits of what he could expect his profession and society rationally to endorse.

Knowing that the 95% confidence limits for the FVT's proper cut-off are described by Equation 4 and Table 3, sensible jurors should conclude that most decisions about future dangerousness ought to be acceptable. Because a broad range of decision thresholds is reasonable, jurors might feel they should adopt a position concerning professional judgment analogous to the Supreme Court's stance in *Youngberg v Romeo*[100] concerning the constitutionality of such judgments: so long as they are making a professionally-based[101] judgment about

---

314 (1981).

As was noted in note 1, recent reports find that *ceteris paribus*, persons with psychotic symptoms are more violent than other persons. However, the contribution of such symptoms to the likelihood of acting violently is quite small compared to socio-demographic factors, such as age, income, race, sex, and education, and the contribution of substance abuse. See Link, Andrews, and Cullen, 57 Am Soc Rev at 290 (cited in note 1); John Monahan, *Mental Disorder and Violent Behavior: Perceptions and Evidence*, 47 Am Psych 511, 519 (1992).

100. 457 US 307, 323 (1982) (holding that in institutional settings, decisions by professionals violate constitutional rights only if "the decision by the professional is such a substantial departure from accepted professional judgment, practice, or standards as to demonstrate that the person responsible actually did not base the decision on such a judgment").

101. That is, a judgment based on "legitimate," relevant criteria. Jurors who accepted this principle might find that a doctor who was bribed to make a decision not to hospitalize had not exercised "professional" judgment and therefore might be liable in tort for damages resulting from the decision. For a recent discussion of courts' efforts to define departures from professional judgment under the *Youngberg* standard, see Susan Stefan, *What Constitutes Departure from Professional Judgment?*, 17 Ment & Phys Disab L Rptr 207 (1993). Professor Stefan examines the impact (and, in her view, misapplication) of the professional judgment standard in Susan Stefan, *Leaving Civil Rights to the "Experts": From Deference to Abdication under the Professional Judgment Standard*, 102 Yale L J 639 (1992).

violence, psychiatrists should not be liable for their prediction errors (save, perhaps, in the grossest circumstances, for example, failing to deem dangerous a blatantly psychotic person who enters the emergency room firing a gun).[102] Of course, after learning that the uncertainties in an ideal prediction process still allow for a huge range of variability in decision-making, jurors might wonder why the law permits mental health professionals to make *any* dangerousness-based commitment decisions. The jurors might realize, however, that given the nature of current commitment laws, psychiatrists working in emergency rooms (and other settings) *cannot avoid* making decisions based on their beliefs about likelihood of future violence, no matter how questionable those decisions may be.[103]

---

102. As Professor Perlin pointed out in a personal communication, "the ambulance attendant, the gift shop host(ess), and Alex and Julie [Professor Perlin's pre-teen children] could figure this out too. . . ." Letter from Michael Perlin, Professor, New York Law School, to Douglas Mossman, Professor, Wright State University (Jan 10, 1994) (on file with author). Although one might argue that psychiatrists ought to have an advantage in predicting violence because they have expertise in detecting mental symptoms, there is no evidence to support this. Mossman, 62 J Consulting & Clin Psych at 783 (cited in note 3). This Article, it should be noted, makes an argument different from the "best judgment rule" which has served in some jurisdictions to shield a psychiatrist from malpractice liability "for a mere error of judgment, provided [the psychiatrist] does what he thinks is best after careful examination." *Littleton v Good Samaritan Hospital*, 39 Ohio St 3d 86, 529 NE2d 449, 457 (Ohio 1988).

> Under such a 'psychotherapist judgment rule,' the court would not allow liability to be imposed on therapists for simple errors in judgment. Instead, the court would examine the 'good faith, independence and thoroughness' of a psychotherapist's decision not to commit a patient. . . . Factors in reviewing such good faith include the competence and training of the reviewing psychotherapists, whether the relevant documents and evidence were adequately, promptly and independently reviewed, whether the advice or opinion of another therapist was obtained, whether the evaluation was made in light of the proper legal standards for commitment, and whether other evidence of good faith exists.

Id at 458 (citing *Currie v United States*, 644 F Supp 1074, 1083 (MD NC 1986)). See also *Schrempf v State*, 66 NY2d 289, 487 NE2d 883 (NY 1985) (holding that a state is not liable for the results of psychiatrist's reasonable but erroneous decision not to commit patient); *Soutear v United States*, 646 F Supp 524, 536 (ED Mich 1986) (*"a psychiatrist will not be held liable for his patient's violent behavior simply because he failed to predict it accurately"*) (emphasis in original), quoting *Davis v Lhim*, 124 Mich App 291, 301, 335 NW2d 481 (1983).

The language in *Littleton* assumes that even though clinicians thoroughly and conscientiously review available data, they will make errors in commitment decisions (i.e., release violent patients) because they are imperfect predictors despite their expertise. *Littleton*, 529 NE2d at 457. This Article suggests that professional expertise should have little impact on the acceptability of dangerousness decisions, because the ambiguity in the moral judgments we attach to the consequences of those decisions overwhelms the at-best-trivial contribution of expertise to prediction accuracy.

103. I am not here referring to psychiatrists' need or desire to avoid *Tarasoff* liability, discussed in notes 16-20 and the accompanying text. I simply mean that following any emergency room encounter, a psychiatrist can either admit a patient involuntarily or not. (If the psychiatrist does not admit the patient involuntarily, the patient may be admitted voluntarily or released, but the patient then is *not* an involuntarily-hospitalized patient.)

## C. ASSUMPTIONS AND QUALIFICATIONS

A major issue—one left largely unaddressed in my discussion—involves the need to use actual numerical values to obtain "real" results about decision thresholds, error rates, and variances. As noted above, persons' preferences and utilities and the relevant time period over which predictions should apply are, for the most part, empirical matters and must reflect a social consensus about the role of mental health professionals in ensuring public well-being and safeguarding patients' civil liberties. This Article suggests an important role for empirical studies that would help gauge the actual distribution of outcome utilities among various populations,[104] as well as specific professional or legal guidance as to the relevant time period for which clinicians should be deemed responsible for their decisions about future dangerousness. Setting precise guidelines for the time period over which predictions should apply would decrease clinicians' uncertainty about their responsibility; indeed, this Article allows one to make an estimate of how great this decrease would be.[105]

In the absence of such guidance, this discussion has utilized what I believe are very conservative estimates of the uncertainties associated with the variables that make up Equation 5.[106] A large disparity in outcome valuations should come as no surprise, for it reflects the disparity in valuations already expressed in published commentary on civil commitment. This commentary includes the views of those mental health professionals who see treatment of mental disorders as an inherent good, who know that involuntary hospitalization is often the only means of providing treatment to seriously disabled persons,[107] and who realize that they are the ones held liable for their patients' violent acts. It also includes constitutional scholars, civil libertarians, and mental health advocates who view involuntary hospitalization as, first and foremost, a revocation of liberty that is justifiable only as an expression of limited state police power,[108] that is likely

---

Even if a clinician never thinks about one of these mutually-exclusive options during a clinical encounter (for example, when a clinician briefly sees an obviously-stable patient presenting himself to an emergency room for a medication refill because the mental health center is closed), one of these two options *must* be exercised.

104. See Mossman and Hart, 21 Bull Am Acad Psychiatry & L 181 (cited in note 34).

105. A precise time period would eliminate the contribution to uncertainty from the variable $N$ in Equations 8-10. Of course, this would still leave considerable uncertainty due to the broad range of possible values for $U_x$ and $D$.

106. Mossman and Hart, 21 Bull Am Acad Psychiatry & L at 187 (cited in note 34), showed that even in very homogeneous populations, perceived ranges for duration of liability and for the relative utilities of false positive and false negative decisions were *orders of magnitude* larger than the ranges used in deriving Equations 11-13 and 11a-13a. One must conclude that people differ greatly about the magnitude of mental health professionals' responsibility for patients' acts and about the desirability of involuntary hospitalization as a means for protecting the public.

107. See, for example, Alan Stone, *Broadening the Statutory Criteria for Civil Commitment: A Reply to Durham and LaFond*, 5 Yale L & Policy Rev 412 (1987).

108. See, for example, Mary L. Durham and John Q. LaFond, *The Empirical Conse-*

to be over-used absent strict safeguards,[109] and that should not necessarily imply a committed person's treatment with pharmacotherapy.[110]

The present attempt to provide a mathematical framework for involuntary hospitalization views such hospitalization as the outgrowth of a policy that treats patients as a population about whom decisions are made, and thus differs from an approach to hospitalization decisions that assumes that each patient can be evaluated under unique circumstances. My treatment of involuntary hospitalization has required, in addition to assumptions about empirical matters, assumptions about what the nature of the decision process should be. I have assumed that for a patient who poses no threat to himself, the justification for involuntary hospitalization derives solely from the desire to prevent harm to others,[111] and I have ignored entirely the closely-linked issues of mentally ill persons' amenability and ability to recognize their need for treatment.[112] I have assumed that making a dangerousness decision involves making a determination that a patient's characteristics exceed a certain threshold;[113] the threshold is derived

---

*quences and Policy Implications of Broadening the Statutory Criteria for Civil Commitment*, 3 Yale L & Policy Rev 395, 431 (1985).

109. Id. See also Stephen J. Morse, *A Preference for Liberty: The Case against Involuntary Commitment of the Mentally Disordered*, 70 Cal L Rev 54, 67-79 (1982).

110. For a summary of the rationale behind "right-to-refuse-treatment" litigation, see Perlin, *Mental Disability Law* §§ 5.01-5.69 (cited in note 9). See also *Rogers v Commissioner, Dept of Mental Health*, 390 Mass 489, 458 NE2d 308 (1983) (holding that the involuntary commitment of a mental patient is not a determination that he is incompetent to make treatment decisions).

111. In practice, clinicians would base admission decisions on their assessment of the potential for harm to both self and others. Treating just the subject of violence simplifies discussion. But as I point out in notes 122-30 and the accompanying text, I suspect that many of the same problems with establishing a threshold of "dangerousness to others" would be found in a similar effort to define a threshold of "dangerousness to self" for purposes of involuntary hospitalization.

112. Several proposed models for commitment law reform have suggested that statutes should incorporate incompetence and/or availability of effective treatment. See, for example, Alan A. Stone, *Mental Health and Law: A System in Transition* 69, 83-89 (Natl Inst of Mental Health, 1975); Clifford D. Stromberg and Alan A. Stone, *A Model State Law on Civil Commitment of the Mentally Ill*, 20 Harv J Leg 275 (1983). Although commentators have been concerned that adoption of such statutes might lead to increased use and abuse of commitment, see Durham and LaFond, 3 Yale L & Policy Rev at 395 (cited in note 108), empirical studies suggest this might not be the case. See Steven K. Hoge, Paul S. Appelbaum, and Alexander Greer, *An Empirical Comparison of the Stone and Dangerousness Criteria for Civil Commitment*, 146 Am J Psychiatry 170, 173 (1989) (finding that Stone's criteria prove to be more restrictive); and Miller, 149 Am J Psychiatry at 1380, 1383 (cited in note 9) (finding that changes in commitment statutes have had minimal practical impact on rates of commitment).

113. See notes 19-20. In an important criticism of this assumption, Dr. Marshall Ginsburg has suggested that my "threshold"-based analysis is entirely irrelevant to the *type* of hospitalization decisions that clinicians operating under the laws of many states are supposed to make. Professor Marshall Ginsburg, conversation with author, 30 Oct 1991. He cites Ohio Rev Code § 5122.01(B)(2) as permitting clinicians to hospitalize mentally ill persons who pose "a substantial risk of physical harm to others as manifested by

from a mathematical computation that incorporates base rates, test characteristics, and the utility associated with correct and incorrect decisions.[114] I have assumed that any prediction instrument will be imperfect in that on some monotonic transformation of the instrument's decision axis non-violent and violent populations will form overlapping, readily-characterizable distributions.[115] I have assumed that rates of violence can be characterized as well.[116] I have assumed that under the best circumstances, prediction mistakes are inevitable, and that the task of the decision-maker is to balance mistakes in a fair and equitable manner. I have assumed that this balancing ideally is achieved by a policy or strategy that maximizes expected utility, where utility is defined so as to take into account our diversity of views about the value of individual liberty and the need to preserve public safety.[117]

All these assumptions are intended to have some normative power. The hypothetical decision process I have described seems preferable to ones that fail

---

*evidence* of recent homicidal or other violent behavior, *evidence* of recent threats that place another in reasonable fear of violent behavior and serious physical harm, or other *evidence* of present dangerousness. . . ." Id (emphasis added). Thus, Dr. Ginsburg argues, Ohio law expects clinicians simply to "obtain evidence" concerning a patient's recent actions, rather than decide (as this Article assumes) whether any particular decision threshold was exceeded. Id.

I offer three counterarguments. First, commitment laws in several states are much less specific and seem to *require* threshold-based thinking. See, for example, 405 Ill Rev Comp Stat 5/1-119(1) (West 1993) (authorizing involuntary hospitalization when a patient is "mentally ill and . . . because of his illness is *reasonably expected* to inflict serious physical harm upon himself or another in the near future" [emphasis added]). Second, the Ohio statute does not limit what might constitute "other evidence of present dangerousness." Ohio Rev Code Ann § 5122.01 (Baldwin 1989). Presumably, a clinician could adduce as evidence a host of demographic factors—age, sex, socioeconomic status, and substance abuse—that are clearly associated with violence, see Swanson, et al, 41 Hosp & Community Psychiatry at 769 tbl 7 (cited in note 1), as well as clinical findings that have been associated with violence. For examples and discussions of such findings, see Stephen H. Dinwiddie and Sean Yutzy, *Dangerous Delusions? Misidentification Syndromes and Professional Negligence*, 21 Bull Am Acad Psychiatry & L 513 (1993) (suggesting that professional literature has linked certain types of delusional syndromes with violence, but that this link is a statistical artifact). Third, even in states, such as Pennsylvania and Wisconsin, that make recent "overt acts" necessary conditions for involuntary hospitalization, there is still room for interpretation and clinical discretion—i.e., threshold analyses—concerning the significance of such acts for future conduct. See 50 Pa Stat § 4405(a)(2) (Purdon 1993) ("The acts or threats which give cause to believe the person to be mentally disabled and in need of immediate care are overt, demonstrate a clear and present danger to self or others. . . ."); Wis Stat § 51.20(1)(a)2b (1993) (" . . . by evidence that others are placed in reasonable fear of violent behavior and serious physical harm to them, as evidenced by a recent overt act, attempt or threat to do serious physical harm. . . .").

114. See Equation 4 and text accompanying note 69.

115. This is the standard "binormal" assumption of ROC analysis. See notes 67-68, the accompanying text, and Figures 1 and 3.

116. See notes 79-84, accompanying text, and Table 2.

117. See notes 48-62 and accompanying text.

to acknowledge the inevitability of mistakes, or that rely on personal moral views, clinical lore, availability heuristics, unconscious race and class biases,[118] correlations between violence and pathology whose predictive power is unquantified,[119] or fears about how a mistake would look in retrospect.[120]

To properly defend all these assumptions would require discussion far beyond the scope of this Article. Mentioning them here only helps emphasize that in judging dangerousness decisions, assumptions about values—either mine or better ones—are necessary. Mathematical approaches to decision problems provide a vehicle for helping us make such assumptions known to *ourselves* as well as others.

## D. ADDITIONAL COMMENTS

I have confined this discussion to decisions made by mental health professionals about violence that their patients might do to third parties. One can readily appreciate that similar imprecisions and problems would frustrate efforts to evaluate clinical decisions involving dangerousness to others in other contexts,[121] or to evaluate decisions about dangerousness to oneself, where similar moral uncertainties about the values of right and wrong judgments, base rates, and time frames no doubt exist.[122]

---

118. For a summary of these phenomena as they affect decision-making about involuntary hospitalization and civil commitment, see Perlin, 47 U Miami L Rev at 640-89 (cited in note 9).

119. For example, Tardiff, Psychiatric Times at 13 (cited in note 23), provides one of many currently-available, excellent summaries of the dozens of factors that scientific studies have shown to increase the risk of violence. The problem for clinicians, however, is that they almost always have no way of knowing how much each of these factors increase risk of violence, and they have no simple way to mathematically combine all these varying factors into anything beyond a vague, global judgment of risk. And even if clinicians could integrate all of this information and calculate risks precisely, the present Article has shown that there is no agreement—and there may never be any agreement—about what level of risk should trigger a particular decision. Well-intended efforts such
as Tardiff's may actually perpetuate the notion that psychiatrists can offer courts and society something more useful than what any intelligent member of the public might contribute.

120. See, for example, Michael L. Perlin, *Morality and Pretextuality, Psychiatry and Law: Of "Ordinary Common Sense," Heuristic Reasoning, and Cognitive Dissonance*, 19 Bull Am Acad Psychiatry & L 131, 137 (1991); *Barefoot*, 463 US at 922 n 4 (discussing the fear of under-predicting violence as determinant of psychiatrists' behavior); *Francois v Henderson*, 850 F2d 231, 234 (5th Cir 1988) (testifying doctor conceded that he "hedged" testimony "because he did not want to be criticized should [the patient] be released and then commit a criminal act").

121. Such contexts include decisions about bail, waiver to adult court of accused juveniles, parole decisions, decisions to release "sexual psychopaths" and persons found not guilty by reason of insanity, and decisions about imposing capital punishment. See Shah, 33 Am Psych at 225 (cited in note 35), and Norval Morris and Marc Miller, *Predictions of Dangerousness*, 6 Crime & Justice 1, 4 nn 2, 7-10 (1985).

122. Moreover, even using *retrospective* statistical models in a population at far-above-average risk for suicide, it was not possible to select out a group for which suicide

I should also note that such uncertainties probably affect a host of legal issues that do not involve mental health professionals or psychiatric treatment. In a series of cases relying on Justice Harlan's concurrence to *In re Winship*,[123] the Supreme Court has justified its view concerning procedural requirements or allocation of burden of proof in various types of litigation using a utilitarian, interest-balancing approach that reflects its views of the possibility of erroneous outcomes and its "assessment of the comparative social disutility of each" outcome.[124] The Court has held that in cases where Bill of Rights provisions are not applicable, procedural "due process generally requires consideration of three distinct factors": private interests, risk of error, and governmental interests.[125] The Court has applied this test in a variety of decisions involving the right to counsel in parental status termination proceedings,[126] the burden of proof in findings of permanent parental neglect,[127] the constitutionality of pre-trial detention of juveniles,[128] and the constitutionality of the Federal Bail Reform Act.[129]

Once again, it would take me far beyond this Article's scope to discuss even minimally the merits of this approach to legal decision-making. I can only point out here that the difficulties we found in defining a threshold for a relatively clear-cut clinical decision with only a limited number of outcomes would be greatly magnified in the less well-defined and much more complex circumstances dealt with in the above-mentioned Supreme Court cases.[130]

---

reached even a fifty percent likelihood. Rise L. Goldstein, et al, *The Prediction of Suicide: Sensitivity, Specificity, and Predictive Value of a Multivariate Model Applied to Suicide among 1,906 Patients with Affective Disorders*, 48 Arch Gen Psychiatry 418, 420 (1991).

123. 397 US 358, 368 (1970) (Harlan concurring).

124. Id at 371.

125. *Mathews v Eldridge*, 424 US 319, 335 (1976) (determining proper procedures for establishing initial and continued entitlement to disability benefits under the Social Security Act).

126. *Lassiter v Dept of Social Services*, 452 US 18, 26-27 (1981) (holding that the Fourteenth Amendment's due process clause does not require appointment of counsel in every proceeding).

127. *Santosky v Kramer*, 455 US 745, 747-48 (1982) (holding that due process requires a state to support its allegations by at least clear and convincing evidence).

128. *Schall v Martin*, 467 US 253, 256-57 (1984) (holding that a New York statute permitting pre-trial detention of juveniles does not violate Due Process Clause).

129. *United States v Salerno*, 481 US 739, 741 (1987) (upholding constitutionality of the Bail Reform Act, 18 USC §§ 3141-3156 (1982 & 1985 Supp II), which allows pre-trial detention without bond for persons accused of certain crimes). The Bail Reform Act itself asks judges to determine whether a defendant's likelihood of endangering the community (especially prospective witnesses and jurors) reaches a threshold of "serious risk." 18 USC § 3142(f) (1982 & 1985 Supp II). For a criticism of the interest-balancing approach generally and the Bail Reform Act in particular, see Alschuler, 85 Mich L Rev at 512-20 (cited in note 90).

130. In his dissent in *Santosky*, Justice Rehnquist makes a similar point:

New York's adoption of the preponderance-of-the-evidence standard reflects its conclusion that the undesirable consequence of an erroneous finding of parental unfitness . . . is roughly equal to the undesirable consequence of an erroneous finding of parental fitness. . . . Such a conclusion is well within the province of

## IV. Conclusion

Psychiatric predictions of violence are enshrined in public policies that reflect our collective reluctance to incarcerate seriously disordered persons when they need medical care[131] and our belief that psychological factors ought to play a role in determining culpability.[132] This Article argues that clinical decisions about dangerousness are fraught with ambiguity that can be given a mathematical characterization. This characterization strongly suggests clinicians should not be held accountable for the consequences of most incorrect violence predictions.

Given the nature of current commitment laws and clinicians' reasonable expectations about civil liability,[133] clinicians will continue to make prediction-based decisions. This Article does not imply that they can or should cease doing so. Clinicians should bear in mind that whether they utilize their customary "intuition" or formalized prediction instruments, predictions about future violence currently have only modest intrinsic accuracy,[134] and decisions based on those predictions are subject to a host of moral ambiguities that are not easily reducible. Extremely accurate prediction tools would mitigate the impact of these ambiguities,[135] but the inherent nature of violent behavior, with its complex interplay of individual and environmental variables, makes it unlikely that very accurate predictions will soon be possible. Research and commentary on violence prediction should bear in mind that prediction accuracy and the blameworthiness of decisions are separate issues. Even if significant improvements in prediction accuracy are achieved, such improvements should not, by themselves, provide reasons to hold clinicians responsible for dangerousness decisions that go wrong.

---

state legislatures. It cannot be said that the New York procedures are unconstitutional simply because a majority of the Members of this Court disagree with the New York Legislature's weighing of the interests of the parents and the child in an error-free factfinding hearing.

*Santosky*, 455 US at 788 n 13 (Rehnquist dissenting).

131. Otto, 5 Forensic Rptr at 106 (cited in note 6); Arthur R. Matthews, *Observations on Police Policy and Procedures for Emergency Detention of the Mentally Ill*, 61 J Crim L Criminol & Police Sci 283, 288-90 (1970).

132. Twentieth Century Fund, *Fair and Certain Punishment* (Priority, 1976).

133. See notes 17-18 and accompanying text.

134. See Mossman, 62 J Consulting & Clin Psych 783, 787-89 tbls 3-4 (cited in note 3).

135. One commercially available version of the enzyme-linked immuno-assay for detecting HIV, the "AIDS virus," has an AUC > 0.999, see Somoza, et al, 24 Intl J Biomed Computing at 164-66 (cited in note 36), and Mossman and Somoza, 3 J Neuro-psychiatry & Clin Neurosci at 332 (cited in note 64), which is equivalent to a diagnostic test in which $A \approx 5$, see note 76. The use of such a prediction tool would greatly reduce var($Z_{FPR}$) calculated in Equation 10, and would change the 95% confidence interval in Equation 11a to $Z_{FPR} = -2.682 \pm 0.510$. Clinicians could then be expected to correctly identify 96.5% to 99.8% of all violent patients and 98.5% to 99.9% of all non-violent patients, even with no change in the errors in $U_x$ and $Pr$. There is *no* reason, however, to believe that clinicians will *ever* be able to predict future violence with this kind of accuracy. FVTs where $A = 2$ seem to be the upper limit of what we can soon hope for.

## V. Appendix—Derivation of Equation 4

We rewrite Equation 1′:

[A1]    $EU = (Pr)(\text{TPR})(U_{TP}) + (Pr)(1\text{-TPR})(U_{FN}) + (1\text{-}Pr)(\text{FPR})(U_{FP})$
        $+ (1\text{-}Pr)(1\text{-FPR})(U_{TN})$

from Equation 67-1, with $B = 1$,

[A2]    $Z_{TPR} = Z_{FPR} + A$

because $Z_{TPR}$ and $Z_{FPR}$ are the normal deviates of TPR and FPR,

[A3]    $\text{TPR} = \Phi(Z_{TPR}) = \Phi(Z_{FPR} + A)$ and $\text{FPR} = \Phi(Z_{FPR})$,

where $\Phi(\cdot)$ is the unit normal cumulative distribution function:

[A4]    $\Phi(y) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt$ .

Setting $U_{TP} = U_{TN} = 1$ and $U_{FN} = 0$, Equation A1 becomes

[A5]    $EU = (Pr)(\text{TPR}) + (1\text{-}Pr)(\text{FPR})(U_{FP}) + (1\text{-}Pr)(1\text{-FPR})$.

We now differentiate Equation A5 with respect to $Z_{FPR}$:

[A6]    $\dfrac{\partial EU}{\partial Z_{FPR}} = \dfrac{Pr}{\sqrt{2\pi}} e^{-\frac{(Z_{FPR}+A)^2}{2}} + \dfrac{(U_{FP})(1\text{-}Pr)}{\sqrt{2\pi}} e^{-\frac{Z_{FPR}^2}{2}} - \dfrac{(1\text{-}Pr)}{\sqrt{2\pi}} e^{-\frac{Z_{FPR}^2}{2}}$ .

Setting this derivative equal to 0 and rearranging terms:

[A7]    $(Pr)e^{-\frac{(Z_{FPR}+A)^2}{2}} = (1\text{-}Pr)(1\text{-}U_{FP})e^{-\frac{Z_{FPR}^2}{2}}$ .

We now rearrange further, and substitute $U_x$ for $(1\text{-}U_{FP})$:

[A8] $$\frac{e^{-\frac{(Z_{int}+A)^2}{2}}}{e^{-\frac{Z_{int}^2}{2}}} = \frac{1-Pr}{Pr} U_x \; .$$

We expand the left side of Equation A7 and take the natural logarithm of both sides:

[A9] $$- AZ_{FPR} - \frac{A^2}{2} = \ln\left(\frac{[1-Pr]}{Pr} U_x\right) \; .$$

When Equation A9 is solved for $Z_{FPR}$, we obtain Equation 4 in the text.