How Bad Is Civil Commitment? A Study of Attitudes Toward Violence and Involuntary Hospitalization

Douglas Mossman, MD; and Kathleen J. Hart, PhD

Civil commitment statutes throughout the nation authorize involuntary hospitalization for persons who are believed dangerous to others, even though clinicians' ability to predict violence is imperfect. Decision-makers faced with ambiguous evidence about future violence must make either-or decisions about involuntary hospitalization. Such decisions can be characterized as "true positives" (hospitalization of a person who would have acted violently if released), "true negatives" (nonviolent person is not hospitalized), "false positive" (nonviolent person is hospitalized), or "false negative" (person is released and subsequently acts violently). This paper presents two pilot studies of attitudes about false negative and false positive decisions, and explains how Decision Theory can use information gleaned from such studies to establish optimal decision thresholds for initiating involuntary hospitalization. Subjects expressed a broad range of implicit tolerances for false negative and false positive predictions. Though most subjects preferred being hospitalized for three days to being the victim of a knife-wielding attacker, a substantial minority preferred being attacked to being hospitalized. The article briefly explores the practical implications of these findings, which include an implicit endorsement of stringent commitment policies that would release a large fraction of potentially violent persons.

For centuries, legal authorities have recognized that the pursuit of justice entails a balancing of errors. Sir William Blackstone's oft-quoted statement, "It is better that ten guilty persons escape than that one innocent suffer," implicitly concedes that erroneous acquittals and convictions will occur.[1] Although other historical authorities have proposed error ratios both lower (5:1)[2] and higher (20:1)[3] than Blackstone's, all seem to agree that erroneous convictions are worse than wrongful acquittals, and that legal fact-finders should err on the side of releasing those who are actually guilty.

The usual standard of proof in criminal trials is guilt "beyond a reasonable doubt," and legal literature offers many

Dr. Mossman is attending physician, Psychiatry Service, Department of Veterans Affairs Medical Center, Cincinnati, Ohio, and associate professor of clinical psychiatry, University of Cincinnati College of Medicine. Dr. Hart is associate professor of psychology, Xavier University, Cincinnati, Ohio. Address correspondence to Dr. Mossman at Psychiatry Service (116A), Department of Veterans Affairs Medical Center, 3200 Vine Street, Cincinnati, OH 45220.

examples of metaphors and paraphrases intended to clarify the meaning of this potentially ambiguous phrase.[4] Empirical studies of judges' and potential jurors' efforts to assign a level of probability to the phrase "reasonable doubt" have yielded very broad ranges of values,[5,6] a finding that reflects both imprecision of everyday English usage and variation in individuals' interpretations of commonly used words.[7] Such imprecision also characterizes words used to express the probability of risk in informed consent litigation: a recent study of 110 published opinions showed that terms such as 'common,' 'possible,' 'small,' and 'rare' "represent broad ranges of numeric probabilities."[8]

In many instances, clinicians who make decisions about possibly violent patients face substantial clinical and verbal uncertainty. One source of uncertainty recognized by most clinicians is the inaccuracy of their predictions about future behavior. A second source stems from ambiguity about the degree of risk that triggers a clinical duty, e.g., the duty to take protective action to prevent harm to third parties. The *Tarasoff* court held that the duty to protect arises when a patient poses "a serious threat of violence" to another individual.[9] Later decisions emphasize that the 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."[10]

When involuntary hospitalization is among the contemplated responses to possible future violence, Monahan[11] has recommended an approach that paraphrases Blackstone's rule about convictions and acquittals. The low base rate of violent behavior, and the relative costs of false positive predictions of violence (deprivation of a few days' liberty) and false negative predictions (physical injury or death), suggest to Monahan that it is best to overpredict violence. Commitment policy should favor false positive prediction errors because "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."[12, p. 370] Chief Justice Warren Burger used very similar reasoning to devise a standard of proof for civil commitment hearings.[13] To our knowledge, however, there have been no published efforts to assess public attitudes about the relative value of false negative and false positive predictions of violence.

This article presents portions of a larger study that applies the methods of formal Decision Theory to evaluate dangerousness-based involuntary hospitalization. We describe here our findings about persons' attitudes toward wrongful hospitalizations and wrongful releases, and discuss briefly how such findings might affect clinical and legal decisions. (For additional discussion of the mathematical features of the theory and an application to psychiatry decision-making, see reference 14.)

Background

We use the term "dangerousness decision" to refer to a clinical action—e.g., the involuntary hospitalization of a pa-

tient—based on a prediction about a person who, over some future period of time, has a certain likelihood of committing an act of violence toward another. Swets[15] points out that decisions of this type are common to many fields, including weather prediction, detection of aircraft flaws, industrial quality control, tax auditing, oil exploration, and drug testing. Such decisions can be characterized as "true positives" (hospitalization of a person who would have acted violently if released), "true negatives" (nonviolent person is not hospitalized), "false positive" (nonviolent person is hospitalized), or "false negative" (person is released and subsequently acts violently). Making a decision in this situation is, essentially, a diagnostic problem common to many fields, where "making a positive or negative decision in a systematic way requires selecting a threshold along the scale of evidence, such that values above the threshold uniformly lead to a positive decision and values below it lead to a negative decision."[15]

The language in *Tarasoff* suggests that, under certain circumstances, patients who attain a threshold value of "serious" potential for future violence create a clinical duty to make a dangerousness decision about their need for hospitalization. Though clinicians usually don't make conscious, mathematical calculations to arrive at decisions, we think it is illuminating to think of this threshold as a mathematical quantity, i.e., as a specific point along a decision scale representing a specific probability of future violence.

To aid readers in thinking mathematically about decisions to hospitalize involuntarily, we have prepared Figure 1, which portrays the performance of the "Future Violence Test" (FVT). The FVT ranks from 0 (lowest) to 100 (highest) the likelihood of future violence in patients presenting to a psychiatric emergency room. Nonviolent individuals tend to score lower on the FVT than do violent individuals. The scores are distributed as shown in Figure 1, with the nonviolent individuals' scores comprising the left-most distribution, and the violent individuals' scores forming the right-most distribution. We "designed" this FVT so that it separates violent and nonviolent persons by one standard deviation. This level of accuracy is very near the average accuracy noted in published reports about violence predictions,[16] an average that includes short- and long-term predictions using a variety of techniques such as clinical intuition, discriminant functions, "actuarial" assessments,[11] and blood tests.

To use the FVT, one simply needs to choose some value or cut-off score at which a person might be hospitalized involuntarily. (We present here a simple discussion of the process of operationalizing a test. Readers will find more extensive discussions in references 14, 17, and 18.) While viewing Figure 1, the reader can imagine three possible cut-offs (40, 50, and 60) for the FVT. In each case, persons with scores falling *above* the cut-off would be "test positive" (that is, would be predicted to be violent and therefore subject to involuntary hospitalization), and those falling

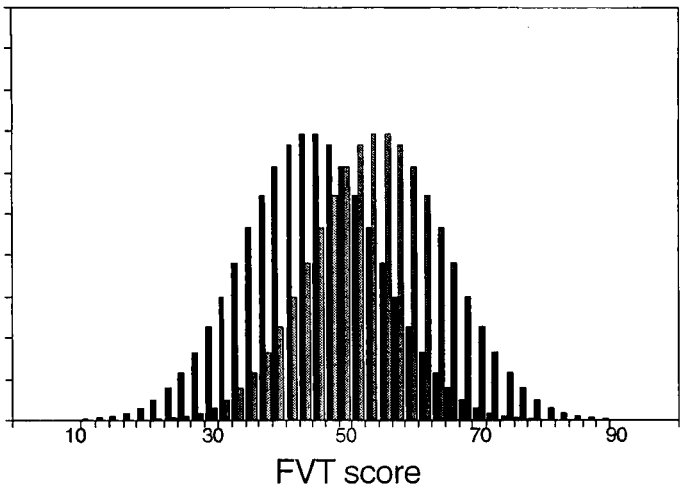

Figure 1. Results of using a hypothetical "Future Violence Test" to rank violent and nonviolent individuals. The violent persons (right-most distribution) tend to have higher scores than the nonviolent persons (left-most distribution). The sensitivity and specificity of the test depend on the cut-off chosen. The distributions have a good deal of overlap, implying that a substantial portion of decisions will be erroneous.

below, "test negative" (that is, predicted to be nonviolent). As one moves the cut-off higher (40 to 60) the performance of the test changes: the fraction of actually violent persons correctly identified by the test (the test's sensitivity) decreases, but the probability of correctly identifying a nonviolent person (the test's specificity) increases.

The task of operationalizing a test requires that one effect a *balance* between sensitivity and specificity by choosing a cut-off that reflects the risks and benefits of test outcomes.[14,15,19] Because even very accurate diagnostic tests 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.

There is, in theory, a rational way of finding the optimal operating point (OOP) that balances the likelihood and the values of test outcomes. One would find the point along the scale where the overall *expected utility* from the test is a maximum; that point, by definition, would be the OOP for the scale. A proper dangerousness decision would be one made such that the FVT's utility was maximized; 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 decisions would reflect correct predictions about violence, but they would represent the best balance of erroneous and correct predictions.

Formal methods for quantifying utilities have been discussed extensively in the decision analysis literature.[20,21] These techniques allow one to make mathematical calculations that yield an OOP for a test.[14,18] Our focus in this study is a simple comparison of the two

possible errors associated with involuntary hospitalization decisions. Monahan suggests that decisions be made bearing in mind that false negative mistakes (releasing dangerous persons) are much worse than false positive mistakes (hospitalizing nondangerous persons). But we wondered: Do others share this opinion? Is there a consensus on how bad violence is compared with involuntary hospitalization? How might we get persons to compare experiencing violence with experiencing involuntary hospitalization? What guidance might public opinion give mental health professionals as to when a risk of violence is "serious" enough to justify involuntary hospitalization?

From the standpoint of the public at large, a clinician's predictions about violence can result in either (1) no one's being harmed, which occurs when clinicians make correct positive or correct negative predictions of violence, or (2) 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 the harm done to a nonviolent person who undergoes a needless involuntary hospitalization as a result of a false positive prediction of violence.

Central to this study is the notion that, in gauging the relative merits of negative and positive judgments about violence, equity and fairness require individuals to "universalize" their judgment. That is, in thinking about releasing or involuntarily hospitalizing someone, persons should regard any possible outcome affecting *any* individual in society as though it happened to *them*. Our evaluations of false negative and false positive judgments about violence or involuntary hospitalization should therefore incorporate the notion that *we* are the ones experiencing violence or undergoing hospitalization. (A full discussion of this assumption would take us far beyond the scope of this article. The intent here is to create a frame of reference for considering civil commitment policies analogous to John Rawls's "initial position." For additional discussion, see reference 22, especially pp. 11–22.)

Methods

We asked undergraduate students in introductory psychology classes and medical students participating in their third-year psychiatry rotations a series of questions, and told them to imagine that their responses would be used to design and implement a "Future Violence Test." The test would be used at a busy urban psychiatric emergency service to help determine whether adults should be hospitalized. We explained that persons with a mental disorder who are believed likely to harm others in the near future can be hospitalized involuntary for a few days for evaluation, and that a substantial fraction of the patients who come to the emergency service behave irrationally, hallucinate, and/or are delusional. We told the students that the FVT is imperfect and that there is no way of knowing in advance who will be misidentified by the test. We also described the consequences of wrong predictions (i.e., a false negative leads to release of a person who subsequently

commits violence, a false positive, to involuntary hospitalization of a nonviolent person). The students were to assume that their answers would be used by the FVT's designers to balance the test's mistakes, *i.e.*, to balance the violations of liberty caused by unnecessary hospitalizations with the violations of public welfare wrought by those mentally ill individuals who act violently.

Task A The students were told that in many jurisdictions, psychiatrists can be held legally responsible for a patient's violent acts, and can be required by courts to compensate victims of violence and their families, if those acts occur within a *limited* time after the psychiatrist evaluated or treated the patient. We asked the students to use a time line to indicate their opinions as to the time period following evaluation that psychiatrists should be liable if patients they wrongfully release later commit acts of violence. The students were given an anchored time line extending from 0 to 365 days; they were asked to briefly explain their answers.

Task B The students were asked a series of questions concerning whether they would prefer being attacked by a man armed with a knife, or spending a certain time period as a patient in a state psychiatric hospital. The students could state they preferred being attacked, hospitalization, or that they felt "about the same" regarding the choices. The time periods offered were 3, 6, 12, and 48 hours; 3, 5, 7, and 10 days; 2, 3, and 6 weeks; 3 and 6 months; and 1, 2, 5, and 10 years. To minimize "anchoring" toward either end of the spectrum,[23] the time periods were presented in a fixed order, starting with the extreme periods and working toward the intermediate ones (i.e., 3 hours, 10 years, 6 hours, 5 years, and so on).

We used the following rules to interpret and code some of the students' answers:

1. In Task B, many students reported more than one time period of indifference, indicating, for example, that they felt being attacked was "about the same" as spending 10, 14, or 21 days hospitalized. When students indicated more than one such indifference point, we coded the geometric mean answer as their response.

2. Many students indicated no point of indifference on Task B. In such instances, we coded the geometric mean time as the student's point of indifference. For example, a student who preferred a six-week hospitalization over being attacked and who preferred being attacked to a three-month hospitalization had 62 days coded as his point of indifference.

3. Several students' answers for Task B implied they were strongly averse to hospitalization: they preferred being attacked to spending even three *hours* in the hospital. For these students, we coded *one* hour as the point of indifference.

4. We deemed invalid (and eliminated from our analysis) answers that were inconsistent (e.g., stating a preference for attack over three weeks in-hospital while stating a preference for six weeks in hospital over an attack) or implausible (e.g., stating indifference be-

tween attack and hospitalizations ranging from three hours to six months).

We analyzed the responses from the undergraduates and the medical students separately so that we could make inter-group comparisons and gauge the generalizability of our results.

Results

Undergraduate Students From 305 undergraduates questioned, we obtained valid responses from 217 students (125 men, 92 women; mean age ± SD 20.3 ± 2.2 years, range 17–44 years, median age = 19 years).

Task A Students' responses concerning duration of post-evaluation liability ran the gamut of alternatives offered. The 46 students who marked 0 on the time line typically commented that psychiatrists "were only human" and couldn't predict the future, or that no one should responsible for the actions of another. The comments of the 29 students who marked 180 to 365 days as the liability time often included statements to the effect that psychiatrists were paid well to make predictions about violence or were "becoming lax in their responsibilities." The distribution of responses is depicted along the horizontal axis in Figure 2. The mean ± SD liability time was 60.2 ± 94.3 days; median time was 30 days. Men's liability times did not differ from women's.

Task B The amount of hospital time that students equated with being attacked by a man wielding a knife ranged *from one hour to ten years.* The responses were distributed log-normally (χ^2 = 5.17, *df* = 4, $p \approx 0.27$). The

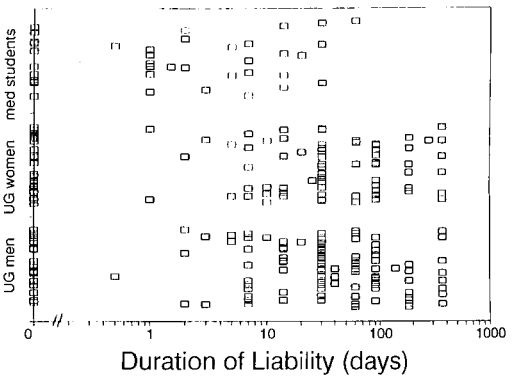

Figure 2. Amount of time for which medical students, undergraduate women, and undergraduate men felt that psychiatrists should be liable for harm done by patients to third parties.

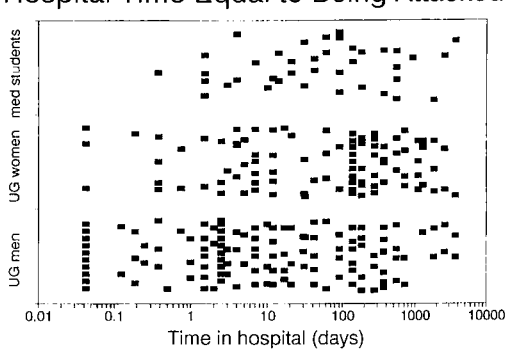

Figure 3. Amount of time in-hospital rated by medical students, undergraduate women, and undergraduate men as equivalent to being attacked by a man armed with a knife.

geometric mean ± SD hospital time equivalent to being attacked was $10^{1.309\pm 1.294}$ days (geometric mean = 20.4 days, median = 21 days, calculated 95% confidence interval = 1.4 hours to 19.2 years). The range of students' answers is depicted in Figure 3.

The women's geometric mean ± SD attack-equivalence time was $10^{1.652\pm1.211}$ days (geometric mean = 44.9 days, median = 91 days); for the men, the time was $10^{1.057\pm1.295}$ days (geometric mean = 11.4 days, median = 12 days). The

women thus equated longer periods of time with being attacked than did the men ($t = 3.437$, $df = 215$, $p < 0.001$ [two-tailed]).

Someone who is highly concerned about the potential violence wrought by psychiatric patients might want psychiatrists to assume a large amount of responsibility for preventing violence, and also might be willing to exchange time in-hospital to avoid violence. We therefore thought that there might be some relationship between the time for which a student thought psychiatrists should be liable for harm to third parties (Task A) and the time the student equated with being attacked (Task B). There was none, however ($R^2 = 0.00761$, $t = 1.284$, $df = 215$, $p \approx 0.20$ [two-tailed]), implying that the student's views about psychiatrists' responsibilities did not translate into corresponding feelings about undergoing an attack or an involuntary hospitalization. There also was no relationship between students' age and their attack-equivalent time ($R^2 = 0.0052$, $t = 1.061$, $df = 215$, $p \approx 0.29$ [two-tailed]).

A large minority of the undergraduates ($60/215 = 27.6\%$) indicated attack-equivalent hospital times of *less* than three days, implying that they thought even a short involuntary hospitalization is worse than being attacked. We did not anticipate that so many subjects would feel this way, and did not inquire further about their views. We therefore do not know whether this subgroup felt that the state's police power does not justify involuntary hospitalization, or that involuntary hospitalizations based on violence predictions are never desirable, although such opinions would be consistent with their aversion to hospitalization.

What might the students' answers tell us about balancing false negative and false positive decision errors? Using a mathematical procedure described in detail elsewhere,[18] we used the amounts of hospital time that students equated with being attacked to assign numerical values, or "utilities," to the potential outcomes of dangerousness decisions.[20, 21] Once one knows these utilities and the prevalence of violent patients in the population being evaluated, one can find the OOP for the FVT, that is, the cut-off point that maximizes the average expected utility when the scale is used to make decisions about involuntary hospitalization.

Table 1 shows the OOPs for a broad variety of views about what time in-hospital is equivalent to being attacked, assuming that the base rate of violence is one percent or ten percent. The range of attack-equivalent times shown (one-half day to three years in-hospital) is roughly equal to the 90 percent confidence limits of the undergraduates' actual range of answers. For each cut-off, the table also shows what fraction of the actually violent patients would be released (the false negative rate), and what fraction of the actually nonviolent patients would be involuntarily hospitalized (the false positive rate).

Table 1 does not include the extreme ten percent of the range of undergraduates' views about hospitalization, but still depicts an enormous range of false negative and false positive rates. The

Table 1
Fractions of Violent Patients Released (False Negative Decision Rates) and Nonviolent Patients Hospitalized (False Positive Decision Rates) Implied by Various Hospitalization-Attack Equivalence Times, Assuming Base Rates of Violence of 0.01 (1%) and 0.10 (10%)

Time in hospital equivalent to being attacked	Base rate = 0.01			Base rate = 0.10		
		Fraction of			Fraction of	
	FVT scale	Violent patients released	Nonviolent patients hospitalized	FVT scale	Violent patients released	Nonviolent patients hospitalized
½ day	≈100	1.0	0	90	0.9998	0.000005
1 day	≈100	1.0	0	83	0.9975	0.000072
3 days	96	0.99998	0	72	0.955	0.0035
1 week	87	0.99942	0.000011	63	0.80	0.032
2 weeks	81	0.9946	0.00019	56	0.56	0.12
1 month	73	0.963	0.0026	49	0.27	0.35
3 months	62	0.75	0.046	38	0.043	0.76
6 months	55	0.50	0.16	31	0.008	0.92
1 year	48	0.24	0.38	24	0.0009	0.98
2 years	41	0.081	0.66	17	0.00003	0.997
3 years	37	0.036	0.79	13	0.00001	0.9993

large variation in rates is, of course, a direct function of the large variation in attitudes about the relative desirableness of being attacked and being hospitalized. One important feature of the table is that, when the base rate of violence is 0.01 (one percent), the roughly three-fourths of the undergraduates who reported attack-equivalent times ≤ six months implicit endorse an involuntary admission policy in which *half of the violent persons are not hospitalized*; for the fourth of the undergraduates who endorsed attack equivalent times of less than three days, the best "dangerousness decision" is to hospitalize *no one* involuntarily. This finding is all the more notable given that even 0.01 may be a high three-day base rate for *serious* violence of the type described in the protocol. (Our calculations using Tillman's data[24] indicate that the three-day rate of arrests for *index* crimes in a young adult population was about 0.00016. The data of Swanson and colleagues[25] allowed us to calculate a three-day rate of weapon use of 0.000089 and a three-day rate of violence in general of 0.00030. In the latter study, however, persons with *any* psychiatric disorder had a three-day rate of violence in general of 0.00078, and persons with a substance abuse disorder had a three-day rate of 0.00205. The calculations used in determining these rates may be obtained from the authors.)

The table also allows us to make some calculations about the fraction of hospitalization and release decisions that are correct at the various cut-offs listed. For example, when the base rate is 0.01, a cut-off of 55 will assure hospitalization of about half of the actually violent patients. However, these hospitalized violent patients will amount to only three percent of all patients hospitalized; 32 nonviolent persons will be hospitalized

for each actually violent person. When the base rate is 0.10, a cut-off of 49 will lead to hospitalization of about 73 percent of the actually violent individuals, with a nonviolent-to-violent ratio of about 4:1. The observations provide pointed demonstrations of the effects of low base rates and modest prediction accuracy on error rates.

Medical Students We obtained valid responses from 39 of the 40 medical students (29 men, 10 women; mean age ± SD 26.9 ± 3.8 years, range 23–38 years, median age = 25 years) whom we questioned. They evinced a relatively conservative view about the period of time following an evaluation during which a psychiatrist should be held liable for a patient's violence. Their median liability time was 1.5 days, and their mean ± SD liability time was 7.2 ± 12.1 days. The medical students endorsed significantly shorter liability times than did the undergraduates ($t = 3.49$, $df = 254$, $p = 0.0004$ [two-tailed]).

The medical students equated being attacked by a man wielding a knife with hospitalizations ranging from 9 hours to 10 years. The responses were distributed log-normally ($\chi^2 = 0.0769$, $df = 3$, $p \approx 0.99$). The geometric mean ± SD hospital time equivalent to being attacked was $10^{1.680\pm0.985}$ days (geometric mean = 47.9 days, median = 61 days, calculated 95% confidence interval = 13.5 hours to 11.2 years), a range that was somewhat narrower than the undergraduates' but that still spanned nearly four orders of magnitude.

The mean and women medical students had similar views about the amount of time in hospital that they felt was equivalent to being attacked (men: $10^{1.69\pm1.012}$ days, geometric mean = 49.1 days; women: $10^{1.648\pm0.901}$ days, geometric mean = 44.4 days). The difference between the undergraduates' and medical students' attack-equivalent times was not significant ($t = 1.70$, $df = 254$, $p \approx 0.09$ [two-tailed]). As with the undergraduates, there was no relationship between the time for which medical students thought psychiatrists should be liable to harmed third parties and their attack-equivalent times ($R^2 = 0.0251$, $t = 0.976$, $df = 37$, $p > 0.2$ [two-tailed]), and no relationship between medical students' age and their attack-equivalent times ($R^2 = 0.00301$, $t = 0.330$, $df = 36$, $p > 0.2$ [two-tailed]).

Discussion

Decisions to release or to hospitalize potentially violent patients entail an implicit recognition of the possibility of error and a weighting of the relative values of correct or incorrect judgments. Typical decision-makers have no way to compare directly the feelings of those who will be affected most directly by their erroneous judgments, i.e., the patients who are mistakenly deemed dangerous and needlessly hospitalized against their will, and the victims of patients who are mistakenly deemed nonviolent and released. Instead, hospitalization decisions typically are made using informal, preconscious, and often error-prone[26] heuristics and weightings of outcomes that reflect the opinions, concerns, and fears of courts or medical personnel.

This study assumed that "dangerousness decisions" might be reached using a formal, explicit decision rule that assigns values to outcomes based on a sampling of views from two nonjudicial, nonpsychiatrist populations. The study asked persons to consider the impact of decision errors from the standpoint of someone who would have to endure the consequences of errors. The views generated from this procedure provide a sampling of opinion about how hospitalization and release decisions should be made under a particular moral point of view, one that requires that judgments with ethical content be "universalizable."

A number of our findings deserve additional investigation. Student in both groups equated very diverse lengths of involuntary hospitalization time equivalent with undergoing an attack at knifepoint. How such relative preferences are translated into opinions about the proper balance of false negative and false positive decision errors depends on the ability to discriminate between violent and nonviolent patients and on the base rate of violence in the population being evaluated. Table 1 allows us to make some rate-specific calculations about the ranges of false-positive-to-false-negative (FP:FN) ratios implicitly endorsed by students, assuming that violence predictions are made with a typical level of accuracy. For example, when the base rate is 0.01, someone who equates being attacked with a week's hospitalization implicitly endorses a FP:FN ratio of 1:918, and someone who equates being attacked with a year's hospitalization implicitly endorses a FP:FN ratio of 157:1; if the base rate is 0.10, a one-week attack-equivalent time translates into a FP:FN ratio of 1:2.78, and a one-year attack-equivalent time yields a FP:FN ratio of about 9800:1. Monahan's suggestion places him roughly in the middle of our subjects: when the base rate is 0.01, an attack-equivalent time of about four months implies a 10:1 FP:FN ratio; when the base rate is 0.10, the ratio is implied by an attack-equivalent time 28 days. Many students, however, implicitly endorsed FP:FN ratios that fall orders of magnitude above or below Monahan's.

We suspect that subjects' beliefs (or fantasies) about what would happen if they were assaulted at knifepoint were a major source of variation in their relative perceptions of hospitalization and being attacked. For example, one of the undergraduates wrote on his answer sheet that he was a "black belt" in karate, presumably in explanation for his willingness to spend only 24 hours in-hospital rather than be attacked. Women equated being attacked with much longer periods of time than did men, perhaps because they thought they would be susceptible to injury by a male attacker (who is statistically likely to be larger and stronger).

Although our protocol left the outcome of the attack to each subject's imagination, we think this is a virtue rather than a fault of our study, for this ambiguity mirrors 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.

We should point out, however, that ambiguity about the attack's outcome is not the only source of variation in relative preferences. The response of two individuals to whom we presented preliminary results of this study are illustrative. One person, a law professor very familiar with psychiatric hospitalization, told us that he "would rather be chased down the street by a band of Hell's Angels than spend a night in" a local public psychiatric hospital. (He was willing to spend up to 12 hours there, however.) Another person, a psychiatry professor very knowledgeable in issues of violence prediction, clearly was astonished that persons would prefer being attacked to spending periods of time in-hospital that were well above average for our subjects. Discussions with our professional colleagues convince us that persons who are knowledgeable about psychiatric inpatient treatment have big disagreements about how willing they would be to enter a state hospital as an involuntary patient. We are confident, therefore, that a substantial portion of variation in preferences stems from genuine variation in our subjects' feelings about hospitalization at a public psychiatric facility.

There was no relationship between the time that subjects felt psychiatrists should be liable for their false negative predictions and their relative aversion to being attacked or being hospitalized. The students who felt that psychiatrists should be held liable for the violent acts of a patient evaluated three, six, or 12 months earlier were just as likely to equate a violent attack with a very short hospitalization as were students who thought psychiatrists should experience no liability at all.

One interpretation of this finding is that many of the students simply were inconsistent in their ranking of outcomes (just as were subjects in studies evaluating standard gambles and attitudes toward health outcomes[27, 28]), and this might lead us to dismiss our findings as unreliable or meaningless. Other interpretations are possible, however. Students may not have recognized that an aversion to being involuntarily hospitalized themselves morally commits them to tolerating release of others, even if those others are potentially violent persons with mental problems. Task A assessed students' attitudes about third-party liability for negligent release of *other* persons, while Task B asked the students to think about undergoing attacks or hospitalization *themselves*; the difference in who would experience an adverse consequence of a decision error may have led students to think about the problem from a different perspective. It is also possible that students who would impose a long-enduring liability harbored implicit beliefs about psychiatrists' prediction accuracy that grossly exaggerate clinicians' actual abilities.

One can show, for example, that with a base rate of 0.01, a prediction instrument accurate enough to separate violent from nonviolent patients by five standard deviations (rather than the single standard deviation separation yielded by our average-accuracy FVT) would allow for a FP:FN ratio of 5.6:1 while releasing only 2.3% of the actually violent individuals. (The enzyme-linked assay for the AIDS virus is roughly this accurate,[17] but most medical tests have accuracies equivalent to a test where the decision scale separates disordered and nondisordered populations from 1 to 3.3 standard deviations.) Were clinicians as accurate as those given to "hindsight bias"[29-31] sometimes think clinicians should be, a low false positive rate might be consistent with assigning psychiatrists a high level of liability for false negative prediction errors.

This study evaluated the responses of young, relatively homogeneous populations who may be unrepresentative of the broader public's attitudes toward and knowledge about public mental policy. All of these findings therefore should be viewed as merely provisional. Given the populations' homogeneity, it is striking to find such broad variations in feelings about experiencing violence and undergoing involuntary hospitalization. A broad range of implied balances of false negative and false positive prediction errors is the direct consequence of this variation. Particularly noteworthy was the finding that over a fourth of the undergraduates expressed an implicit preference for being attacked over undergoing a three-day hospitalization in a public psychiatric facility. One might think that medical students would feel relatively comfortable with being in a hospital, but their aversion to involuntary hospitalization was nearly as great as the undergraduates'. It would be very interesting to learn whether similar preferences would be expressed by more representative samplings of the general population, and to compare their views with those of "interested parties" in the commitment process: lawyers and judges, law enforcement personnel, mental health professionals, and psychiatric patients. Such knowledge should have an important impact on commitment decisions and the nature of the commitment process.

References

1. Connolly T: Decision theory, reasonable doubt, and the utility of erroneous acquittals Law Hum Behav 11:101–12, 1987
2. Hale M: Pleas of the Crown (reprint of 1678 ed.). Oxford, Professional Books, 1972
3. Fortescue J: A Learned Communication on the Laws of England (reprint of 1567 ed.). New York, W. J. Johnson, 1969
4. Nagel S, Lamm D, Neef M: Decision theory and juror decision-making, in The Trial Process. Edited by Sales BD. New York, Plenum, 1981
5. Simon RJ, Mahan L: Quantifying burdens of proof: a view from the bench, the jury, and the classroom. Law Society Rev 5:319–30, 1971
6. Dane FC. In search of reasonable doubt: a systematic examination of selected quantification approaches. Law Hum Behav 9:141–58, 1985
7. Behn RD, Vaupel JW: Quick Analysis for Busy Decision Makers. New York, Basic Books, 1982
8. Merz JF, Druzdzel MJ, Mazur DJ: Verbal expressions of probability in informed consent litigation. Med Decis Making 11:273–81, 1991
9. Tarasoff v. Regents of the University of California, 551 P.2d 334 (Cal. 1976)

10. Appelbaum PS: Ask the experts. Am Acad Psychiatry Law Newsletter 17(1):19, 1992
11. Monahan J: The Clinical Prediction of Violent Behavior. Washington, DC, Government Printing Office, 1981
12. Monahan J: Strategies for an empirical analysis of the prediction of violence in emergency civil commitment. Law Hum Behav 1:363–71, 1977
13. Addington v. Texas, 441 U.S. 418, 429 (1979)
14. Somoza E, Mossman D: "Biological markers" and psychiatric diagnosis: risk-benefit balancing using ROC analysis. Biol Psychiatry 29:811–26, 1991
15. Swets JA: The science of choosing the right decision threshold in high-stakes diagnostics. Am Psychol 105:522–32, 1992
16. Mossman D, Schaffner R: Violence prediction: being accurate about accuracy. Presented at the Annual Meeting of the American Academy of Psychiatry and the Law, Orlando, Florida, October, 1991; Mossman D, Schaffner R: ROC analysis & the accuracy of violence prediction. New Research poster presentation at the Annual Meeting of the American Psychiatric Association, Washington, DC, May, 1992. Data are available on request to the authors.
17. Somoza E, Soutoullo-Esperon L, Mossman D: Evaluation and optimization of diagnostic tests using receiver operating characteristic analysis and information theory. Int J Biomed Comput 24:153–89, 1989
18. Mossman D, Somoza E: Balancing risks and benefits: another approach to optimizing diagnostic tests. J Neuropsychiatry Clin Neurosci 4:331–5, 1992
19. Mossman D, Somoza E: Assessing improvements in the dexamethasone suppression test using receiver operating characteristic analysis. Biol Psychiatry 25:159–73, 1989
20. French S: Decision Theory: An Introduction to the Mathematics of Rationality. Chichester, Ellis Horwood, Ltd., 1988
21. Luce RD, Raiffa H: Games and Decisions: Introduction and Critical Survey. New York, Dover, 1957
22. Rawls J: A Theory of Justice. Cambridge, MA, Harvard University Press, 1971
23. Tversky A: Assessing uncertainty. J Roy Statist Soc Ser B 36:148–59, 1974
24. Tillman R: The size of the "criminal population:" the prevalence and incidence of adult arrest. Criminology 25:561–79, 1987
25. Swanson JW, Holzer CE III, Ganju VK, Jono RT: Violence and psychiatric disorder in the community: evidence from the epidemiologic catchment area surveys. Hosp Community Psychiatry 41:761–80, 1990
26. Tversky A, Kahnemann D: Judgment under uncertainty: heuristics and biases. Science 185:1124–31, 1974
27. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, *et al*: "The measurement of patients' values in medicine. Med Decis Making 2:449–62, 1982
28. Thorton JG, Lilford RJ: Prenatal diagnosis of Down's syndrome: a method for measuring the consistency of women's decisions. Med Decis Making 10:288–93, 1990
29. Fischhoff B: Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. J Exp Psychol Hum Percept Perform 1:288–97, 1975
30. Bursztajn H, Gutheil TG, Brodsky A, *et al*: "Magical thinking," suicide, and malpractice litigation. Bull Am Acad Psychiatry Law 16:369–77, 1988
31. Wexler DB, Schopp RF: How and when to correct for juror hindsight bias in mental health malpractice litigation: some preliminary observations. Behav Sci Law 7:485–504, 1989