Probability, Danger, and Coercion: A Study of Risk Perception and Decision Making in Mental Health Law

Paul Slovic

Decision Research and

Department of Psychology, University of Oregon

and

John Monahan

School of Law

University of Virginia

Running head: PROBABILITY, DANGER, AND COERCION

In press, *Law and Human Behavior* 06-24-94

Probability, Danger, and Coercion: A Study of Risk Perception and Decision Making in Mental Health Law

Paul Slovic

Decision Research and

Department of Psychology, University of Oregon

and

John Monahan

School of Law

University of Virginia

Running head: PROBABILITY, DANGER, AND COERCION

Law and Human Behavior, in press.

06-24-94

Abstract

Young adults were shown hypothetical stimulus vignettes describing mental patients and were asked to judge a) the probability that the patient would harm someone else, b) whether or not the patient should be categorized as "dangerous," and c) whether coercion should be used to insure treatment. Probability and dangerousness judgments were systematically related and were predictive of the judged necessity for coercion. However, judged probability was strongly dependent upon the form of the response scale, suggesting that probability was not represented consistently and quantitatively in our Ss'minds. Study 2 replicated these findings with forensic clinicians as Ss. These results underscore the importance of violence to others in mental health law and have important implications for the manner in which risk assessments are formulated for use by the legal system.

Probability, Danger, and Coercion: A Study of Risk Perception and Decision Making in Mental Health Law

Analyses of how people perceive risk of harm and how they make decisions based on those perceptions have proven fruitful in many topical areas such as management of technological and natural hazards (e.g., Krimsky & Golding, 1992; National Research Council, 1989; Slovic, Kunreuther, & White, 1974) and the promotion of safe sex, use of seat belts, and other personal protective behaviors (Weinstein, 1987). One area in which, to date, risk perception and decision-making methodologies have not often been applied is mental health law (see Wexler & Winick, 1991, for an exception). This is unfortunate, since some of the key issues in mental health law appear highly amenable to being enlightened by a risk perception and decision-making framework. In this paper, we explore the manner in which people perceive one kind of risk—the risk that a person with mental disorder will be violent—and the effects of those perceptions on their willingness to impose involuntary mental hospitalization upon the individual.

Perhaps the clearest example of an issue in mental health law that might be illuminated by a risk perception and decision-making framework is what has come to be known as "the dangerousness standard" for involuntary mental hospital admission. In 1969, California changed its legal test for admitting a person to a mental hospital against his or her will to be that the person must be mentally disordered and "dangerous to self or others" (see Brooks, 1978). Most states have adopted the California "dangerous" language (Brakel, Parry, & Weiner, 1985), but others refer to the "likelihood" that the individual will cause "serious harm" (Brakel et al., 1985, p. 34). The American Psychiatric Association (1983) recommends similar language—"likely to cause harm to others"—in one of its model tests for civil commitment. The National Center for State Courts' Guidelines for Involuntary Civil Commitment (1986), on the other hand, speak of "predictions of violence," and the American Bar Association (1989) recommends that the civil commitment of persons acquitted of crime by reason of insanity be limited to those who are mentally disordered and "as a result, pose a substantial risk of serious bodily harm to others" (Standard 7-7.3). Finally, one influential court decision phrased the issue in terms of a "high probability of substantial injury" (Cross v. Harris, 1969, p. 1097).

"Dangerousness," "likelihood," "prediction," "risk," and "probability," therefore, have been used interchangeably to refer to the undesirable outcomes that are anticipated to occur if some mentally disordered persons are left at liberty. However, the extensive literature in the area of risk perception and behavioral decision theory has uncovered many subtle and anomalous effects which suggest that these various terms may not be fungible. They may, in fact, have differential effects on the judgments that are rendered by clinicians and courts (e.g., Slovic, Fischhoff, & Lichtenstein, 1982). The equivalence of concepts such as probability of doing harm and dangerousness is examined in the present study.

A second pivotal issue in mental health law that follows directly from the "dangerousness standard" issue is the use of the coercive power of the state to impose mental health treatment (Bennett et al., 1993). Once a person is perceived as mentally disordered and as dangerous—the legal test for commitment in most American states—a decision must be made as to whether he or she is "sufficiently" mentally disordered and dangerous to qualify for involuntary mental hospitalization, unless the person voluntarily admits him- or herself into the hospital (Hoge et al., 1993; Rogers, 1993).

The present study represents an initial attempt to provide insight into two questions fundamental to the relationship between dangerousness and coercion in mental health law:

- How are judgments of probability of doing harm and dangerousness formed and to what extent are these two concepts equivalent?
- How much danger is enough danger to trigger coercive intervention?

Study 1

Method

Hypothetical case vignettes were constructed from the following eight dichotomous cues:

1. Gender (M/F)
2. Prior Hospitalization (1 or 3)
3. Delusions (No/Yes)
4. Prior Assaultiveness (Never/Yes)
5. Anger (Often/Rarely)
6. Impulsivity (Frequent/Seldom)
7. Psychopathy (Little Empathy or Concern/Much . . .)
8. Social Support (Many Friends & Family Members/Few . . .)

These cues are believed to be, in fact, related to the likelihood of future violence. They are among the cues being researched in the MacArthur Risk Assessment Study (Steadman et al., 1994). There are 256 patterns of "cases" that can be formed from all 2^8 combinations of these cues. Thirty-two of these patterns were selected for study, based on a ⅛ fractional replication design. Across these 32 cases, all pairs of cues are uncorrelated and each one takes each of its two levels 16 times. In addition to the eight cues, each case was assigned an age (between 25 and 29 years) and an occupation. One of the 32 cases is shown at the top of Figure 1. Subjects were told that each vignette provided a brief description of a person who has been examined by a psychiatrist or psychologist and has been diagnosed as having some form of mental illness.

Subjects were asked to make three consecutive judgments about each vignette, as shown in Figure 1. First they were asked to indicate "the probability that this person will harm someone else during the three years following the examination." This judgment, designated P_h, was made on an 11-category scale ranging from 0 (no chance) to 100% (certain to harm) in increments of 10.

Second, subjects were asked to indicate whether they would describe the patient as dangerous (yes or no). Finally, they were asked whether they believed coercion should be used to insure hospitalization and treatment for the patient in the event the patient refused to enter the hospital. This, too, was a dichotomous judgment (yes or no). The judgments of dangerousness and coercion will hereafter be labeled D and C, respectively.

Insert Figure 1 about here

A second group of subjects judged the 32 cases in the same three ways except that the probability scale included five small probabilities: less than 1 chance in 1,000, 1/1000, 1%, 2%, and 5%. These changes were made to allow the respondent to differentiate at the low end of the probability scale instead of being forced to use either 0% or 10% to express a small probability. In addition, the high end of the scale was labeled "greater than 40%." The instructions for this scale are shown in Figure 2. Hereafter, we shall differentiate the two probability formats by referring to the first as Condition LP (for large probability) and the second as condition SP (for small probability).

In a medical judgment task, Mazur and Merz (in press) found that use of a small-probability scale (such as that used in condition SP) led to much lower probabilities being assigned to complications from surgery described as "rare" (in comparison with results obtained from a scale such as that in Condition LP). It is important, therefore, to examine possible probability scale effects in the present study.

Insert Figure 2 about here

Subjects. Subjects were 95 women and 96 men who answered an ad in the University of Oregon student newspaper. Most of these individuals were university students. Their mean age was 21 years (range = 18 to 49 years). They were paid for participating in the study. Ninety-five subjects judged the vignettes using the 0-100 probability scale (Condition LP) and 96 used the scale that allowed various small probabilities to be assigned (Condition SP).

Results

Probability and dangerousness judgments. Table 1 presents the relationship between judged probability of harming someone (denoted P_h) and the judgment of dangerousness in Condition LP. It is apparent that there is a close functional relationship between these two judgments. When judged probability of harm is .20 or less, the person is rarely judged dangerous. When the judged probability is .70 or greater, dangerousness is almost always ascribed to the individual in the vignette. As P_h increases from .20 to .70, the probability that the label "dangerous" will be assigned increases quite rapidly.

Insert Table 1 about here

Table 1 presents group data, from all subjects in Condition LP. What might the relationship look like for individual respondents? One possibility is that individual subjects will exhibit functions that look more or less similar to the group function. A second hypothesis is that the individual data resemble step functions, with the stimulus patient being judged not dangerous (D = 0) whenever judged probability of harm was less than some transition level (denoted P_{ht}) and dangerous whenever $P_h \geq P_{ht}$. After arranging all the P_h judgments in ascending order, a step function might look like this (hypothetical data):

Judge 1.	P_h judgment	0	10	10	20	30	30	30	40	40	50	50	50
	Dangerousness judgment	0	0	0	0	1	1	1	1	1	1	1	1
						↑ P_{ht}							

where 0 represents not dangerous and 1 represents dangerous and the transition point (P_{ht}), occurs at .30.

Another judge might have a different transition point, say P_{ht} = .50, as in the following example.

Judge 2.	P_h judgment	10	20	20	30	30	40	40	50	60	70
	Dangerousness judgment	0	0	0	0	0	0	0	1	1	1
									↑ P_{ht}		

Combining data across judges whose step functions differ only with respect to the point of transition between not dangerous and dangerous could produce group data similar to the data in Table 1.

The step function hypothesis was tested by examining respondents' dangerousness judgments to see whether they indeed made a clean transition as shown in the examples above. The 32 responses for each judge were coded as 0 (not dangerous) or 1 (dangerous) and arranged in a vector ordered in terms of the probability of harm judgments that preceded each dangerousness judgment, as in the examples just given.[1] Guttman's coefficient of reproducibility (Guttman, 1944) was computed to indicate the percentage of responses that would have to be changed in order to produce a perfect step function for each respondent. This was done for each respondent. The results provided strong support for the step function hypothesis. The median coefficient of reproducibility was .93; 64 of 86 subjects (74%) had coefficients of .88 or greater, meaning that reversal of 3 or fewer responses out of 32 would give them a perfect step function relating P_h to D.

Examination of the transition values of P_h, calculated for subjects with coefficients of reproducibility above .88, indicated that the switch from not dangerous to dangerous occurred as low as $P_h = .30$ for some subjects and as high as $P_h = .90$ for others, with the median and modal value being $P_h = .60$.

Examination of the step function relationship between P_h and D for individual subjects showed that virtually all had the characteristics of the hypothetical step functions shown earlier. That is, there were multiple values of P_h assigned to patients judged not dangerous (up to the threshold possibility), and multiple values of P_h within the set of patients judged as dangerous (i.e., above the threshold). Thus, whereas a judgment of probability leads rather precisely to a judgment of dangerousness, the converse does not hold. Judgments of dangerousness do not lead to a precise judgment of probability.

Coercion judgments. The relationship between coercion judgments (C) and judged P_h is shown in Table 1. The relationship between P_h and C would be S-shaped if plotted. Some patients (15%) were judged in need of coercion even when $P_h = 0$ and some (again 15%), for whom P_h was 100, were not judged to need coercion.

An analysis was carried out to determine whether the relationship between P_h and coercion could also be modeled by a step function. Twenty-three subjects could not be modeled because they did not vary their judgment across cases. Of the remaining 73 subjects, only 58% had coefficients of reproducibility $\geq .88$ for coercion (compared to 74% for the dangerousness judgments). The median coercion threshold was .60 for these individuals, the same median as the dangerousness threshold.

We next examined the relationship between the two dichotomous judgments, dangerousness and coercion, for each of the 32 stimulus cases. The relationship is shown in Table 2. Table 2 indicates that across all subjects, the mental patient was judged dangerous in 37.0% of the cases and was judged in need of coercion in 48.6% of the cases. There was a strong relationship between dangerousness and coercion judgments. However, patients judged not dangerous were still judged in need of coercion in 29.6% of the cases and patients judged dangerous were not judged in need of coercion 19.2% of the time, indicating that dangerousness was not the sole criterion for the coercion judgment.

Insert Table 2 about here

Influence of the P_h response scale. In condition LP, P_h was judged by marking an 11-point scale going from 0 to 100 in units of 10. This scale gave the judge no opportunity to provide small values of P_h other than 0 or 10. What happened to judgments of P_h, D, and C, and the relationship between them in Condition SP, in which the scale was changed to allow distinctions to be made among small values of P_h?

Table 3 gives the distribution of P_h judgments and the proportion of cases at each P_h level that were judged dangerous and in need of coercion in Condition SP. If we compare this table with the data from condition LP in Table 1 we see an immense effect due to the change in the probability response scale. Condition SP offered six probability response categories equal to or less than 10% ($<\frac{1}{1000}$, $\frac{1}{1000}$, 1%, 2%, 5%, and 10%) and these six categories accounted for 67.7% of the total responses. This can be compared to 10.8% of cases assigned a $P_h \leq 10$ in Condition LP. Offering small-probability responses led our subjects to make many more assignments of small values of P_h. Across the 32 case vignettes, the mean probability of doing harm was .44 in Condition LP and .12 in Condition SP.[2]

Insert Table 3 about here

The relationship of dangerousness judgments and coercion judgments to P_h was also greatly altered by the change in response scale. For example, when $P_h = .10$, only 1% of the cases in Condition LP were judged dangerous and only 14% were judged in need of coercion (see Table 1) compared to 25% judged dangerous and 55% coerced in Condition SP when $P_h = .10$ (Table 3). Similar differences between Conditions LP and SP occurred when P_h was .20, .30, and .40.

Table 3 shows that the relationships between P_h and D and between P_h and C were strongly affected by the response categories for P_h. Does the overall incidence of dangerous and coercion judgments also change? The answer to this question is "yes."

There were fewer judgments of dangerousness in Condition SP (30.5% vs. 37.0%). Coercion judgments were also less frequent in Condition SP (40.4% vs. 48.6%). However, the relationship between dangerousness and coercion was quite similar within each condition. Of those judged dangerous in Condition SP, 79.4% were judged in need of coercion (compared to 80.8% in Condition LP); of those judged not dangerous in Condition SP, only 23.3% were judged in need of coercion (compared to 29.6% in Condition LP).

Would the step function relationships between P_h and D, and P_h and C hold as well for Condition SP, with its many small probabilities? The step function model fit the dangerousness judgments moderately well, as indicated by the fact that 67% of the subjects had coefficients of reproducibility equal to or greater than .88. As in condition LP, precise step functions were less frequent among the coercion judgments, with only 53% of the subjects exhibiting coefficients ≥ .88.[3]

The subjects in Condition SP whose judgments were well-fit by step functions had threshold values that were similar for both dangerousness and coercion. The switch from not dangerous to dangerous and from do not coerce to coerce occurred as low as $P_h = \frac{1}{1000}$ for some subjects and as high as $P_h > 40$ for others. The median threshold was .20 for both types of judgments. These median thresholds were considerably lower than the median threshold of .60 obtained for both dangerousness and coercion using the 0, 10, 20, . . . 100 response scale in Condition LP. In both conditions, the median threshold was about 60% up the length of the scale.

Condition SP, with its numerous small-probability response categories, caused many more judges to rate all 32 cases as not dangerous and all cases as not in need of coercion. Specifically, the percentage of subjects judging all 32 cases as "not dangerous" was 17.7% in Condition SP and 3.2% in Condition LP. The percentage of subjects wanting none of the 32 cases coerced was 19.8% in Condition SP and 9.5% in Condition LP.

Case-by-case mean values of P_h, D, and C were computed by averaging across all judges in each condition (LP and SP). The mean P_h value was far higher in Condition LP for every case. However, the percentage of judges who rated a case as dangerous or in need of coercion was not as dramatically different in the two conditions, despite the widely discrepant judgments of P_h. Take Case 4, for example: the mean P_h was .51 in Condition LP and only .13 in Condition SP yet the proportion of subjects who judged this patient as dangerous was .44 vs. .32 in the two conditions and the coercion percentages were .48 and .39, respectively.

Although the mean values of P_h varied greatly between conditions LP and SP,their relative magnitudes across cases were almost identical. The correlation between P_h in Condition LP and in Condition SP was .95. The correlation between the dangerousness percentages in the two conditions was also .95; for the two sets of coercion percentages it was .89. In sum, our subjects' probability responses were much higher in Condition LP but in a relative sense (relative to P_h in Condition SP and to the dangerousness and coercion judgments) interrelationships across cases were little changed.

Modeling judgments of P_h, D, and C. We developed an equation to predict the mean judgments of P_h, D, and C for each case, based upon a weighted additive combination of the eight attributes in the vignette (see, e.g., Hoffman, Slovic, & Rorer, 1968). Because the 32 cases were constructed by means of a functional replication design, pairs of attributes are uncorrelated and standardized regression weights for each of the eight attributes are equivalent to the correlations between the attributes and the judgments across the 32 vignettes. These weights are shown in Table 4, along with the R^2 value indicating the overall predictability of the mean judgments from the model. The results indicate that the 32 vignette means were highly predictable from these simple weighted additive models, with R^2 values ranging between .85 and .96. Prior assaultiveness was the most heavily weighted attribute in every model, with anger being next most important. Gender and support from friends and family were given relatively small weights. The models for P_h and D were quite similar. The weight given to delusions was substantially higher in the coercion model for Condition LP, and slightly higher for the coercion model for Condition SP. The weight given to prior hospital admissions was substantially higher in the coercion model for condition SP.

Insert Table 4 about here

Discussion of Study 1

Study 1 produced a number of substantive findings relevant to the central issues of violence and coercion in mental health law.

First, the judged probability that a patient will harm someone was strongly related to the judgment that the patient is "dangerous."

Second, whereas group data on the relationship between judged probability of harm and dangerousness could be represented nicely by an S-shaped curve, the relationship appeared to be a step-function for many of our individual subjects: below a given probability value, the patient was rarely designated dangerous, and above that value, the patient was almost always designated dangerous. That "transition value" varied greatly across subjects, from as low as a .30 probability of harm to as high as a .90 probability of harm when the response options included larger probabilities and from a probability of 1/1000 to .40 when the response options included smaller probabilities.

Third, the judged probability that a patient will harm another, and the ascription of dangerousness, were both strongly related to the judgment that the patient should, if necessary, be coerced into mental hospital admission. In the condition employing large probabilities, for example, subjects would coerce into treatment 29.6% of the patients seen as not dangerous, and 80.8% of the patients seen as dangerous.

Fourth, it was possible using attributes of the case vignettes to create highly accurate models to predict judgments of probability of harm, dangerousness, and the need to coerce. Prior assaultiveness was the most heavily weighted attribute, with anger being the next most important, in models predicting all three judgments. The three models were similar except that prior hospital admissions and the presence of delusions weighed more heavily in judgments that coercive treatment was necessary than in judgments of probability of harm or dangerousness.

Fifth, the response scale used to structure judgments of the probability of harm had an enormous effect on the judgments that are rendered. When two response categories equal to or less than 10% (i.e., 0% and 10%) were provided, 10.8% of the cases were assigned to one of these categories. When six response options equal to or less than a 10% probability of harm were provided, those response categories together accounted for 67.7% of all responses. Rating probability of harm on a scale with small-probability options reduced the tendency to subsequently judge the patient as being dangerous and reduced the tendency to want coercion used. Scales with small probabilities seemed to induce more people never to label someone as dangerous and never to desire any patient to be coerced. Although the numerical probability judgment was greatly affected by the response scale, mean judgments of probability for each case correlated quite highly (.95) across the 32 cases despite being elicited with two different response scales.

Study 2

Subjects in Study 1 had no professional expertise in judging the dangerousness of mental patients. It is perhaps not surprising that they exhibited little consistency in assigning probabilities to cases when the response scale was altered. Would mental health professionals behave similarly? We addressed this question in Study 2, which was a partial replication of Study 1.

Subjects

The subjects in Study 2 were 137 forensic clinicians (55% men; 45% women) attending the semi-annual Forensic Symposium of the Institute of Law, Psychiatry, and Public Policy at the University of Virginia. The majority of these clinicians (68%) were psychologists; 13% were psychiatrists, and 11% were social workers. The number of years these people had been in practice ranged from 1 to 45 with a mean and median of 12 years.

Method

Each subject was given the same task used in Study 1 except that only the first eight of the 32 vignettes were included in the questionnaire. Seventy persons were assigned to Condition LP; 67 to Condition SP.

Results

Influence of the P_h response scale. Tables 5 and 6 present the distribution of responses in Conditions LP and HP respectively. As in Study 1, the scale format had a strong influence on the judged values of P_h. In Condition SP, the six response categories equal to or less than 10% accounted for 49.3% of the total responses. This contrasts sharply with the 20.0% of the cases assigned to $P_h \geq 10$ in Condition LP. Offering small-probability responses led even these professionals to make many more assignments of small values of P_h.

Insert Tables 5 and 6 about here

The relationship between dangerousness judgments and coercion judgments to P_h was also greatly altered by the change in response scale. In Condition HP, few cases were judged dangerous or in need of coercion until P_h reached 50. In Condition LP, the rapid increase in dangerousness and coercion judgments began when P_h reached 15. Additional perspective on the response-scale effect is gained by examining specific levels of P_h. For example, when $P_h = .20$, only 2% of the cases in Condition LP were judged dangerous and only 6% were judged in need of coercion, compared to 53% judged dangerous and 55% in need of coercion at $P_h = .20$ in Condition SP. Providing small-probability response options greatly increased judgments of dangerousness and coercion at specific values of P_h.

As in Study 1, there was a strong relationship between the two dichotomous judgments, dangerousness and coercion, and this relationship was almost identical in the two probability conditions. Moreover, despite the different values of P_h assigned in the two conditions, the overall percentage of cases judged dangerous and the percentage judged in need of coercion were almost identical in the two conditions (about 33% dangerous; 35% in need of coercion).

The case-by-case assignments of P_h, D, and C across all judges in each condition are shown in Table 7. The mean value of P_h is considerably higher in Condition LP for each case.[4] However, as in Study 1, the percentage of clinicians who judged a case to be dangerous or in need of coercion was not much different in the two conditions, despite the large mean differences in P_h. Across the 8 cases, the mean values of P_h in the two conditions correlated .99, showing that the ordering was preserved although the mean values differed.

Insert Table 7 about here

In Study 1, Condition SP appeared to cause more judges to rate none of the 32 cases as dangerous and none of the cases as needing coercion. This finding did not replicate with the experienced clinician subjects in Study 2. The proportion of subjects who judged none of the eight cases as dangerous or none in need of coercion was almost identical in the two response-scale conditions.

General Discussion

The results of Study 2, with experienced clinicians as subjects, were remarkably similar to the results obtained in Study 1 with naive subjects. Together, these studies underscore the central importance of violence to others in current mental health law (Appelbaum, 1988). Whether expressed as "probability of harm" or "dangerousness," the higher a person's perceived risk of violence, the more likely both the public and the clinicians will want the person treated in a mental hospital, coercively so if voluntary admission is refused.

At the same time, it should be recognized that violence, while a dominating concern, is not the only issue in involuntary hospitalization. In Study 1, between one-quarter and one-third of the cases that were not judged as dangerous to others were nonetheless recommended as candidates for involuntary hospital admission (the comparable percentage in Study 2 was about 15%). Among the attributes of these "not dangerous to others, but coercible" cases in Study 1 were the presence of delusional beliefs and multiple hospital admissions. Subjects responding to such cases may have been concerned more with "danger to self." Such a "parens patriae" concern with danger to self (or the passive form of danger to self called "grave disability") has always been part of mental health law (Monahan & Shah, 1989).

Perhaps the most striking result in the two studies was the effect of the response scale on judged probability of harm. A scale that finely differentiated among small probabilities led to far lower probabilities being assigned to the cases, much as had been found earlier in a medical judgment context (Mazur & Merz, in press). The response-scale effect was almost as large among experienced clinicians in Study 2 as with naive subjects in Study 1. Naive subjects, who were induced by the response scale to assign lower P_h values to a case, were somewhat less likely to label the patient as dangerous or in need of coercion. However, the mental health professionals appeared immune to such influence. The likelihood that a clinician would judge a patient to be dangerous or in need of coercion was not systematically affected by the response scale and its effect on P_h.

Our subjects appear to have used the probability response scales as if they were rating scales, numbered from 1 to 11 or 1 to 13, with no meaning to the numbers other than their rank. This behavior is more than just a methodological artifact. It strongly implies that the concept "probability of harm" was not represented in our subjects' minds in a consistent, quantitative way. The numbers circled on the probability scale appear to have been meaningless in an absolute sense, though they were consistent and meaningful in a relative sense. In other words, if a subject circled 10% as the probability for Case 1 and 20% for Case 2, we can be confident that he or she perceived Case 2 as having a greater probability of harming someone. We cannot have confidence in the values 10% and 20%. Small changes in the response scale can cause these two cases to be judged as having very different values of P_h (though the order will likely be preserved).

If this interpretation of the response-scale effect is correct, it has profound implications for the manner in which risk assignments are formulated by psychiatrists and psychologists for use by the legal system. Judged probabilities that an individual will be violent cannot be trusted, except in an ordinal sense. It has long been advocated (e.g., Monahan & Wexler, 1978) that mental health professionals offer courts their probability estimates that violent behavior will occur, and not impose summary labels of "dangerous" or "not dangerous" on these estimates. From our data, however, ascriptions of "dangerousness" actually appear more stable than numerical judgments of probability! We do not, however, recommend a return to the era when clinicians offered conclusory statements about "dangerousness" to courts. Rational decision making must ultimately depend upon an appreciation of the probability that the patient (or prospective patient) will become violent. To use an example from Study 2—with experienced clinician subjects—it should make a difference in decision making, at least in the marginal case, whether the probability of violence is .27 or .52 (which were the obtained mean probabilities in conditions SP and LP, respectively, for vignette 4; see Table 7). Ordinal judgements alone (e.g., "Of the 8 people I have recently evaluated, I would rank him as the third or fourth most likely to be violent," which was also true for vignette 4, in both conditions SP and LP) will not suffice.

Rather than resurrect the respectability of summary labels, the conclusion that numerical probability judgments by clinicians are unreliable points to the need for probability assessments guided by actuarial studies (e.g., Klassen & O'Connor, 1988; Lidz, Mulvey, & Gardner, 1993; Steadman et al., 1994). Indeed, if numerical probability judgments are as unstable as our studies have shown them to be, one might question the ethicality of making such judgments in a clinical setting without relying on actuarial guidelines, despite the acknowledged difficulties of the actuarial approach (Grisso & Appelbaum, 1992, 1993; Litwack, 1993).

Such a strong conclusion may, however, be premature. One can argue that Study 2 placed clinicians at a severe disadvantage by asking them to rate hypothetical vignettes in an unnatural context. Clinicians, in their actual practice, judging real patients, might be better able to apply their knowledge and experience to the task of assessing probabilities. We suspect, however, that our well-structured hypothetical cases are actually easier to judge than are real patients, whose history and attributes are not so clearly specifiable. Nevertheless, it is important that additional studies be conducted to determine the impact of scale effects in the clinician's natural assessment setting.

Additional studies should also employ procedures for structuring the assessment task that have been found to facilitate probabilistic thinking. Such procedures range from instruction about the meaning of probability and practice in making probability judgments (e.g., Fong, Krantz, & Nisbett, 1986; Lichtenstein & Fischhoff, 1980; Morgan & Henrion, 1990) to asking about relative frequencies ("of 100 patients with this set of characteristics or symptoms, how many would be violent?") instead of asking for the "probability that this particular person will be violent" (e.g., Cosmides & Tooby, in press; Gigerenzer, in press)

There may be other structural changes in the assessment task that would infuse quantitative meaning into clinicians' probability assessments. For example, assuming that the clinician could judge probability reliably at the ordinal level, he or she could be asked to insert the stimulus case into a probability scale whose numerical values were linked with (or calibrated against) other kinds of events whose probabilities or relative frequencies were well known and agreed upon. Alternatively, the scale might be calibrated with verbal expressions of probability (Hamm, 1991) or descriptions of other "marker" patients located on the scale by expert consensus.

In sum, the strong effect of scale format has important implications for research and practice in mental health law. It remains to be seen whether the continued use of quantitative clinical judgments of the probability of violent behavior can be justified.

References

American Bar Association. (1989). ABA criminal justice mental health standards. Chicago: American Bar Association.

American Psychiatric Association (1983). Guidelines for legislation on the psychiatric hospitalization of adults. American Journal of Psychiatry, 140, 672-679.

Appelbaum, P. (1988). The new preventive detention: Psychiatry's problematic responsibility for the control of violence. American Journal of Psychiatry, 145, 728-732.

Bennett, N., Lidz, C., Monahan, J., Mulvey, E., Hoge, S., Roth, L., & Gardner, W. (1993). Inclusion, motivation, and good faith: The morality of coercion in mental hospital admission. Behavioral Sciences and the Law, 11(3), 295-306.

Brakel, S., Parry, J., & Weiner, B. (1985). The mentally disabled and the law (3rd ed.). Chicago: American Bar Foundation.

Brooks, A. (1978). Notes on defining the dangerousness of the mentally ill. In C. Frederick (Ed.), Dangerous behavior: A problem in law and mental health (pp. 87-60). Washington, DC: Government Printing Office.

Cosmides, L., & Tooby, J. (In press). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty. Cognition.

Cross v. Harris, 418 F.2d 1095 (1969).

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.

Gigerenzer, G. (In press). Why the distinction between single-event probabilities and frequencies is important for psychology and vice versa. In G. Wright & P. Ayton (Eds.), Subjective probability. New York: Wiley.

Grisso, T., & Appelbaum, P. S. (1992). Is it unethical to offer predictions of future violence? Law and Human Behavior, 16, 621-633.

Grisso, T., & Appelbaum, P.S. (1993). Structuring the debate about ethical predictions of future violence. Law and Human Behavior, 17, 482-485.

Guttman, L. (1944). A basis for scaling qualitative data. American Sociological Review, 9, 139-150.

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. Organizational Behavior and Human Decision Processes, 48, 193-223.

Hoffman, P., Slovic, P., & Rorer, L. (1968). An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 69, 338-349.

Hoge, S., Lidz, C., Mulvey, E., Roth, L., Bennett, N., Siminoff, L., Arnold, R., & Monahan, J. (1993). Patient, family, and staff perceptions of coercion in mental hospital admission: An exploratory study. Behavioral Sciences and the Law, 11(3), 281-294.

Klassen, D., & O'Connor, W. (1988). A prospective study of predictors of violence in adult male mental health admissions. Law and Human Behavior, 12, 143-158.

Krimsky, S., & Golding, D. (1992). Social theories of risk. New York: Praeger.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. Organizational Behavior and Human Performance, 26, 149-171.

Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. Journal of the American Medical Associaiton, 269, 1007-1011.

Litwack, T.R. (1993). On the ethics of dangerousness assessments. Law and Human Behavior, 17, 479-482.

Mazur, C. J., & Merz, J. F. (In press). How age, outcome severity, and scale influence general medicine clinic patients' interpretations of verbal probability terms. Journal of Internal Medicine.

Monahan, J., & Shah, S. (1989). Dangerousness and commitment of the mentally disordered in the United States. Schizophrenia Bulletin, 15, 541-553.

Monahan, J., & Wexler, D. (1978). A definite maybe: Proof and probability in civil commitment. Law and Human Behavior, 2, 37-42.

Morgan, M. G., & Henrion, M. (1990). Uncertainty. Cambridge: Cambridge University.

National Center for State Courts (1986). Guidelines for involuntary civil commitment. Mental and Physical Disability Law Reporter, 10, 409-514.

National Research Council (1989). Improving risk communication. Washington, DC: National Academy Press.

Rogers, A. (1993). Coercion and "voluntary" admission: An examination of psychiatric patient views. Behavioral Sciences and the Law, 11(3), 259-268.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 463-489). New York: Cambridge University Press.

Slovic, P., Kunreuther, H., & White, G.F. (1974). Decision processes, rationality and adjustment to natural hazards. In G.F. White (Ed.), Natural hazards, local, national and global (pp. 187-205). New York: Oxford University Press.

Steadman, H., Monahan, J., Appelbaum, P., Grisso, T., Mulvey, E., Roth, L., Robbins, P., & Klassen, D. (1994). Designing a new generation of risk assessment research. In J. Monahan & H. Steadman (Eds.), Violence and mental disorder: Developments in risk assessment (pp. 297-318). Chicago: University of Chicago Press.

Weinstein, N.D. (Ed.) (1987). Taking care: Understanding and encouraging self-protective behavior. New York: Cambridge University Press.

Wexler, D., & Winick, B. (1991). Essays in therapeutic jurisprudence. Durham, NC: Carolina Academic Press.

Authors' Note

This research was supported by the Research Network on Mental Health and the Law of the John D. and Catherine T. MacArthur Foundation. We are grateful to Paul Appelbaum, Thomas Grisso, Steven K. Hoge, Sarah Lichtenstein, Allan Murphy, and Henry Steadman for their comments on the manuscript and to C. K. Mertz, Leisha Mullican, and Kari Nelson for their assistance in data analysis and preparation of the manuscript. Correspondence concerning this article should be addressed to Paul Slovic, Decision Research, 1201 Oak Street, Eugene, OR 97401-3575.

Footnotes

1. This analysis included only subjects who gave both kinds of judgments ($D = 0$ and $D = 1$). Nine persons who judged every one of the 32 cases identically were excluded because no threshold could be calculated for them.

2. When computing means in Condition SP, responses $< \frac{1}{1000}$ and $\frac{1}{1000}$ were assigned the value 0, and the response > 40 was assigned the value 70, which is above the median of the categories greater than 40 in Condition LP (see Table 1).

3. As before, these analyses included only subjects who gave both kinds of judgments (D and $\overline{D}$ and C and $\overline{C}$). Twenty-six persons who judged all 32 cases identically with respect to dangerousness and 41 persons who judged all cases identically with respect to coercion were excluded because no threshold could be calculated for them.

4. Means for Condition SP were computed as in Study 1; see footnote 2.

Table 1.

Judgments of Probability, Dangerousness, and Coercion in Condition LP

Judged Probability of Doing Harm	N	% of Total Judgments	% Judged Dangerous	% Judged Coerce
0	47	1.6	00	15
10	277	9.2	01	14
20	421	13.9	04	19
30	438	14.5	13	36
40	382	12.6	24	44
50	462	15.3	32	59
60	383	12.7	69	63
70	314	10.4	85	80
80	184	6.1	90	90
90	93	3.1	97	72
100	20	0.7	90	85
	3021			

Table 2.

Dangerousness and Coercion in Condition LP

		COERCE		
		No	Yes	
DANGEROUS	No	70.4	29.6	63.0
	Yes	19.2	80.8	37.0
		51.4	48.6	N = 3027

Note. Cell entries are row percentages. For example, of those who judged a patient not dangerous, 70.4% did not believe the patient should be coerced.

Table 3.

Judgments of Probability, Dangerousness, and Coercion in Condition SP

Judged Probability of Doing Harm	N	% of Total Judgments	% Judged Dangerous	% Judged Coerce
$< \frac{1}{1000}$	148	4.8	00	18
$\frac{1}{1000}$	244	8.0	04	19
1%	406	16.0	09	17
2%	407	13.3	12	20
5%	457	15.0	18	35
10%	323	10.6	25	55
15%	250	8.2	31	46
20%	186	6.1	42	56
25%	151	4.9	64	64
30%	168	5.5	77	74
35%	102	3.3	85	77
40%	104	3.4	97	61
> 40%	105	2.6	95	81
	3051			

Table 4.

Standardized Weights in Linear Models to Predict Mean Judgments in Conditions LP and SP

	Condition LP			Condition SP		
Judgment	P_h	D	C	P_h	D	C
R^2	.96	.92	.90	.90	.92	.85
Gender	-.11	-.14	-.08	-.04	-.09	-.09
Prior Hospitalization	.14	.06	.13	.23	.16	.35
Delusions	.19	.20	.37	.24	.28	.29
Assaultiveness	.77	.78	.72	.69	.68	.54
Anger	-.40	-.37	-.37	-.32	-.43	-.45
Impulsivity	-.21	-.20	-.14	-.23	-.25	-.25
Psychopathy	-.28	-.21	-.13	-.35	-.30	-.26
Social Support	.16	.14	.21	.19	.09	.12

Table 5.

Judgments of Probability, Dangerousness, and Coercion in Condition LP, Study 2

Judged Probability of Doing Harm	N	% of Total Judgments	% Judged Dangerous	% Judged Coerce
0	17	3.0	00	12
10	95	17.0	00	11
20	87	15.5	02	06
30	69	12.3	10	19
40	39	7.0	13	15
50	66	11.8	41	44
60	53	9.5	74	55
70	65	11.6	74	75
80	36	6.4	83	75
90	30	5.4	83	80
100	3	0.5	100	100
	560			

Table 6.

Judgments of Probability, Dangerousness, and Coercion in Condition SP, Study 2

Judged Probability of Doing Harm	N	% of Total Judgments	% Judged Dangerous	% Judged Coerce
$< \frac{1}{1000}$	8	1.5	00	12
$\frac{1}{1000}$	37	6.9	03	11
1%	52	9.7	00	06
2%	36	6.7	03	14
5%	76	14.1	05	14
10%	56	10.4	09	20
15%	36	6.7	25	28
20%	38	7.1	53	55
25%	36	6.7	61	55
30%	40	7.5	80	70
35%	22	4.1	55	64
40%	47	8.8	64	49
> 40%	47	8.8	83	77
	531			

Table 7.

Mean Probability of Harm and Proportion of Dangerousness and Coercion Judgments in Study 2

	Condition LP			Condition SP		
Case #	P_h	Danger	Coercion	P_h	Danger	Coercion
1	.15	.00	.03	.03	.01	.06
2	.62	.64	.69	.35	.57	.73
3	.52	.54	.55	.27	.45	.60
4	.52	.44	.34	.27	.55	.39
5	.63	.69	.62	.38	.74	.55
6	.22	.06	.11	.04	.03	.03
7	.26	.09	.31	.05	.03	.25
8	.39	.23	.17	.15	.29	.21
Mean	.41	.33	.35	.19	.33	.35

Figure Captions

<u>Figure 1</u>. Stimulus case and response formats.

<u>Figure 2</u>. Instructions for the probability scale used in Condition SP.

Case 1 - 6

T.D., a 25 year old male, is employed as a bus driver. Records reveal that this is his first admission to a mental hospital. He is not experiencing any delusional beliefs. Relatives report that he has never been assaultive in the past.

Among the characteristics of his personality noted in the mental health examination are that he rarely becomes very angry, that he seldom acts impulsively, and that he has much empathy or concern for others.

T.D. has many friends and family members to help him with his problems.

a. Please indicate your judgment of the probability that this person will harm someone during the 3 years following the examination.
(Circle the appropriate probability.)

0	10	20	30	40	50	60	70	80	90	100%
no chance										certain to harm

b. Would you describe this individual as <u>dangerous</u>? (Check one.)

No ________

Yes ________

c. Sometimes it is necessary to coerce a mentally disturbed person into entering a hospital and receiving treatment. If this person were to refuse hospitalization and treatment, do you believe that coercion should be used to insure that proper treatment is received? (Check one.)

No. Do not coerce. ______

Yes. Coercion should be used. ______

Figure 1. Stimulus case and response formats.

One of the questions asks you to judge the probability that the person described would harm someone during the three years following the examination. The response scale begins with very low probabilities.

less than 1 chance in 1000
1/1000 (meaning 1 chance in 1000)
1% (meaning 1 chance in 100)
2% (meaning 2 chances in 100 or 1 chance in 50)
5% (meaning 5 chances in 100 or 1 chance in 20)

and progresses to higher probabilities, as shown below:

less than 1 chance in 1000	1/1000	1%	2%	5%	10%	15%	20%	25%	30%	35%	40%	greater than 40%

Circle the appropriate probability for each case.

Figure 2. Instructions for the probability scale used in Condition SP.