

Comparison of Standards for Assessing Patients' Capacities to Make Treatment Decisions

Thomas Grisso, Ph.D., and Paul S. Appelbaum, M.D.

***Objective:** This study investigated the proportion of psychiatric and medical patients who are impaired in their decision-making abilities in relation to each of several major legal standards for determining competence to consent to treatment. **Method:** The subjects were hospitalized patients with diagnoses of schizophrenia (N=75), major depression (N=92), and ischemic heart disease (N=82) and equal numbers of community comparison subjects matched on age, race, gender, education, and occupation. Three instruments measuring abilities related to the legal standards for competence were administered to each group. Impaired functioning was defined as scores two standard deviations below the means for all subjects combined or lower. **Results:** Although similar percentages of subjects with impaired performance were found for each of the measures, different groups of patients were identified as impaired depending on the measure used. The proportion of patients identified as impaired increased when compound standards were used, i.e., when impairment was defined as poor performance on any of two or three measures. Despite previous suggestions that the legal standards might form a hierarchy of rigorosity, the data did not support this hypothesis. **Conclusions:** Choice of standards for determining competence, including compound standards, will affect the identity and proportion of patients classified as impaired. Clinicians should be aware of applicable standards in their jurisdictions. Use of compound standards requires independent evaluation of performance on abilities related to each relevant standard, because standards do not appear to be hierarchical. Policies relating to the characterization of persons as incompetent must be fashioned with caution.*

(Am J Psychiatry 1995; 152:1033-1037)

Clinicians are frequently faced with the need to assess patients' capacities to make decisions about treatment. Formal assessments may be required in general medical or psychiatric settings when questions are raised about patients' abilities to consent to or refuse treatment (1, 2). Even more common are the implicit judgments that physicians make concerning their patients' decision-making abilities when they discuss treatment options with them (3). In either case, clinicians' judgments will determine whether they accept patients' decisions or seek the assistance of alternative decision makers.

The process of assessing patients' capacities, either formally or in routine discussions with patients, has been hampered by confusion over the standards to be

applied. Legal standards for decision-making capacity (i.e., standards for assessing legal competence to consent to treatment) are often vague and inconsistent, appearing to vary across and even within jurisdictions (4). Clinicians are therefore uncertain about which standard or standards to apply and what methods to use in assessing patients' performance with respect to those standards.

Commentators have recognized in recent years that courts generally select from among a small number of possible standards for decision-making capacity, applying one or more of them in particular cases (3-5). Four standards may be found in the law, each addressing a somewhat different type of ability: 1) to express a choice; 2) to understand information relevant to the decision about treatment; 3) to appreciate the significance for one's own situation of the information disclosed about the illness and possible treatments; and 4) to manipulate the information rationally (or reason about it) in a manner that allows one to make comparisons and weigh options.

What has remained unclear to clinicians and courts alike is the impact of selecting one or another of these

Received Sept. 19, 1994; revision received Dec. 27, 1994; accepted Jan. 3, 1995. From the Law and Psychiatry Program, Department of Psychiatry, University of Massachusetts Medical Center. Address reprint requests to Dr. Grisso or Dr. Appelbaum, Department of Psychiatry, University of Massachusetts Medical Center, 55 Lake Avenue North, Worcester, MA 01655.

Supported by the Research Network on Mental Health and the Law of the John D. and Catherine T. MacArthur Foundation.

standards or of applying several simultaneously. If patients' abilities tend to be impaired uniformly across the functions addressed by each of the standards, whichever of the standards is selected will have no effect on the outcome. On the other hand, if conditions that impair mental functioning differentially affect the abilities associated with the standards, opinions about whether particular patients are capable of making their own decisions might be affected greatly by the applicable standard or standards. Moreover, if the standards are hierarchically organized, such that some are more rigorous than others, the proportion of patients allowed to make their own decisions will vary according to the standard chosen.

The potential empirical effect of the various standards has important implications for both policy and practice. Selection of a rigorous standard of decision-making capacity risks depriving a substantial number of persons of their rights to make decisions about their lives, as well as subjecting patients to potential delays in much-needed treatment while the capacities of a large number of patients are being reviewed by administrative boards or the courts. If a less rigorous standard is selected, however, the process will fail to identify some patients who should be protected from the potentially deleterious effects of their decision-making incapacities.

This article uses data from a large, multisite controlled study of decision-making capacities of patients with mental and medical illnesses to consider the effects of selecting among standards to determine decision-making capacity. The study examined patients' performance on instruments that assessed abilities related to legal standards that are commonly applied when questions regarding patients' decision-making capacities are raised.

METHOD

Three groups of patients hospitalized for acute illness participated in the study: two with mental illness (schizophrenia or schizoaffective disorder, $N=75$; major depression, $N=92$) and one with a medical illness (ischemic heart disease [angina pectoris], $N=82$). Data were collected in psychiatric and medical facilities at three sites (Worcester, Mass.; Pittsburgh, Pa.; and Kansas City, Mo.); each group of patients was recruited from two of the sites. Patients' participation was voluntary, written informed consent was obtained, and testing occurred in most cases 2–7 days after hospital admission.

Three comparison groups were composed of volunteers who were living in the communities served by the hospitals and who were not ill. Each community comparison subject was matched to a hospitalized subject on age, gender, race, and socioeconomic status (based on education and highest-rated lifetime occupation). Potential community subjects were excluded if they had a history of experiencing the illness for which they were to participate as a comparison subject or, in the case of the comparison groups for the mentally ill groups, if they met screening criteria for schizophrenia or depression (6).

Further details of the composition and recruitment of the groups have been provided elsewhere (7). Patients' clinicians asked that certain patients not be approached for participation, usually because they were too ill (1%–20% of admitted patients, depending on the site). Other sources of attrition included unavailability of potential subjects when interviewers were present. Ultimately, 43%–56% of patients who were eligible on the basis of admitting diagnosis were

approached for recruitment to the study. Among patients who were approached, 15%–20% declined to participate. Despite various sources of attrition, the demographic composition of the study groups (age, race, gender) was similar to that of the admission populations from which they were drawn.

Three instruments were developed specifically to assess the abilities associated with four legal standards for competence to consent to treatment. In contrast to other measures that have been developed for assessing competence of patients (5), these instruments are the only set of measures designed to assess abilities related to all of the legal standards. Moreover, in contrast to other instruments, the three measures were developed to maximize standardization of administration, provide highly structured procedures, and offer scoring criteria that minimized differences between scorers in judgments about the quality of patients' responses. The content, administration, and scoring reliability of the three instruments are described in an earlier publication (8) and in the instruments' manuals, which are available from the first author on request. Interrater reliability was found to be quite good for all measures.

The instrument measuring *understanding* ("Understanding Treatment Disclosure") presents subjects with a standardized set of information in five brief parts, describing 1) a disorder and its symptoms, 2) a proposed medication as treatment, 3) the potential benefits of the medication, 4) the potential risks and discomforts of the medication, and 5) an alternative treatment (hospital milieu and psychotherapy, surgery) and its potential benefits and risks/discomforts. Each hospitalized subject received a version of the instrument describing his or her own type of disorder, and each community comparison subject received the same version as was given the patient to whom he or she was matched. After each of the five parts of the disclosure, subjects were asked to paraphrase the information in their own words, and their responses were scored according to criteria provided in the instrument's manual. Scores range from 0 to 10.

The instrument measuring *reasoning* ("Thinking Rationally About Treatment") presents subjects with a hypothetical story about a person who developed an illness (the same as the patient's own or, for comparison subjects, the same as the illness of the group for which they were acting as comparisons) and then offers a description of three treatment choices, their benefits, and their risks/discomforts. Subjects are asked to "advise" the person in the story regarding a choice of treatment and to explain their reasons. The instrument provides for objective scoring of the presence or absence in subjects' explanations of several aspects of decision making (e.g., considering consequences, making comparisons between options). Other tasks in the reasoning instrument include standardized tests of subjects' grasp of probability and transitive thinking and their ability to generate consequences of the available treatment options. Scores range from 0 to 14. A measure of ability to express a choice is included in this instrument; scores range from 0 to 2.

The instrument measuring *appreciation* ("Perceptions of Disorder") has two parts. One part ("Nonacknowledgment of Disorder") describes to patients their own current diagnosis and symptoms (based on descriptions in their hospital charts), then asks them to rate the degree to which they agree that these statements accurately describe them. Low scores (lack of appreciation) are obtained on the instrument when patients' perceptions are markedly discrepant from clinical observations in their charts. The second part ("Nonacknowledgment of Treatment Potential") elicits patients' opinions about whether they believe that treatment in general, and medication specifically, might be of some benefit to them. The measure does not "penalize" subjects merely for disbelieving the value of treatment. A procedure for analyzing their explanations for their belief ensures that subjects will obtain low scores (failure to appreciate the possible benefit of treatment) only when their explanations are based on grossly distorted or delusional premises. Scores on each of the two subtests range from 0 to 6.

Research assistants administered all three measures to the hospitalized subjects. Community comparison subjects, however, received only two instruments; the appreciation measure was not appropriate for them, because its format requires that reference be made to the subject's own disorder and symptoms.

For understanding and reasoning, performance was classified as impaired if scores were two standard deviations below the mean of

TABLE 1. Subjects Whose Scores Indicated Impairment on Three Measures of Capacity to Make Treatment Decisions^a

Measure	Subjects With Scores Indicating Impairment							
	Patients With Schizophrenia (N=75) ^b		Patients With Depression (N=92)		Patients With Angina (N=82)		Comparison Subjects (N=249)	
	N	%	N	%	N	%	N	%
Understanding	21	28.0	5	5.4	6	7.3	6	2.4
Reasoning	18	24.0	7	7.6	0	0.0	5	2.0
Appreciation ^c	17	22.7	11	12.0	4	4.9		

^aCriteria for impairment were the following: for understanding, scores of 0–4; for reasoning, scores of 0–5; for appreciation, a score of 0 on nonacknowledgment of disorder and/or a score of 0–2 on nonacknowledgment of treatment potential.

^bOn all measures, the group with schizophrenia manifested a significantly greater proportion of subjects in the impaired range than did the other groups. For example, in the comparison between the patients with schizophrenia and the patients with depression, for understanding, $\chi^2=14.52$, $df=1$, $p<0.001$; for reasoning, $\chi^2=7.48$, $df=1$, $p<0.01$; for appreciation, $\chi^2=4.15$, $df=1$, $p<0.05$.

^cThe appreciation measure was not administered to the comparison subjects.

TABLE 2. Subjects With Adequate Performance on One Measure of Capacity to Make Treatment Decisions Who Showed Impairment on Other Measures

Subjects With Adequate Performance on Measure	Subjects With Impaired Performance on Measure					
	Understanding		Reasoning		Appreciation ^a	
	N	%	N	%	N	%
Understanding						
Schizophrenic patients (N=54)			8	14.8	13	24.1
Depressed patients (N=87)			6	6.9	12	13.8
Angina patients (N=76)			0	0.0	4	5.3
Comparison subjects (N=243)			4	1.6		
Reasoning						
Schizophrenic patients (N=55)	10	18.2			12	21.8
Depressed patients (N=84)	4	4.8			10	11.9
Angina patients (N=82)	6	7.3			4	4.9
Comparison subjects (N=244)	5	2.0				
Appreciation						
Schizophrenic patients (N=57)	17	29.8	14	24.6		
Depressed patients (N=80)	5	6.3	6	7.5		
Angina patients (N=79)	5	6.3	0	0.0		

^aThe appreciation measure was not administered to the comparison subjects.

scores for the total study group (i.e., for all hospitalized patients and community comparison subjects combined) or lower. With the use of this criterion, impaired understanding was defined as a score of 4 or less and impaired reasoning as a score of 5 or less. Impaired performance on appreciation was defined as having either the lowest possible score (score=0) on the subtest for nonacknowledgment of disorder or a score of 0–2 on the subtest for nonacknowledgment of treatment potential. These cutoffs provide a conservative estimate of impaired performance, a choice that is consistent with the policy that individuals should be considered capable of making decisions for themselves in the absence of substantial evidence to the contrary (9).

RESULTS

Table 1 summarizes the percentage of subjects in each hospitalized patient group and in the combined community comparison groups who scored in the impaired range on each of the instruments. Scores on ability to express a choice showed that few subjects performed poorly; these data, therefore, are not discussed further in this article.

For each of the three measures, the proportion of subjects with schizophrenia who scored in the im-

paired range—about one-quarter—was significantly higher than the proportion in the other two hospitalized groups. Nonetheless, within any given diagnostic group, the proportion of subjects who scored in the impaired range was not significantly different across instruments. To explore whether each instrument identified the same subset of subjects, we examined subjects' patterns of impairment scores across all three measures.

As shown in table 2, the results did not support the hypothesis. A substantial number of participants who performed adequately on one measure manifested impaired performance on another. For example, of the 72.0% of subjects with schizophrenia who performed adequately on the understanding measure, 24.1% had impaired performance on appreciation and 14.8% on reasoning. Given the lower overall proportions of subjects in the impaired range in the depression, angina, and comparison groups, the figures for these groups were proportionately smaller.

It is apparent, therefore, that the use of compound standards (for example, a court's use of more than one

TABLE 3. Subjects Whose Scores Indicated Impairment on Compound Measures of Capacity to Make Treatment Decisions

Measures	Subjects With Scores Indicating Impairment							
	Patients With Schizophrenia (N=75)		Patients With Depression (N=92)		Patients With Angina (N=82)		Comparison Subjects (N=249)	
	N	%	N	%	N	%	N	%
Understanding and/or appreciation	34	45.3	17	18.5	10	12.2	6	2.4 ^a
Understanding and/or reasoning	29	38.7	11	12.0	6	7.3	10	4.0
Appreciation and/or reasoning	30	40.0	17	18.5	4	4.9	5	2.0 ^a
Understanding, appreciation, and/or reasoning	39	52.0	22	23.9	10	12.2	10	4.0 ^a

^aPercentage does not reflect performance on the appreciation measure, which was not administered to the comparison subjects.

standard in any given case) would produce a greater proportion of persons identified as impaired in decision-making abilities than would the use of any single standard. Table 3 demonstrates this effect with various compound standards. For example, while only about one-quarter of the subjects with schizophrenia showed impaired performance on any one measure, 52.0% of them would be considered impaired if a compound standard that defined nonimpairment as adequate performance on all three measures were used. Comparable figures for the other groups were 23.9% for the group with depression, 12.2% for the group with angina, and 4.0% for the three community comparison groups combined.

The possible hierarchical relationships of the standards are seen in tables 1 and 3. If the standards were hierarchical, such that they could be ranked in order of increasing stringency, then different proportions of subjects would score in the impaired range on different standards. As can be seen in table 1, however, the application of single standards failed to result in the appearance of a hierarchy; the percentage of subjects in the impaired range was not markedly different across standards. Moreover, as shown in table 3, no clear hierarchical effect arose when combinations of two or more standards were used.

DISCUSSION

This study suggests that the three major standards for evaluating decision-making capacities related to treatment do not differ in the proportion of patients they identify as performing relatively poorly on treatment decision tasks. Nonetheless, if a single standard is to be used, the choice of standards is important in determining which particular patients may be classified as relatively impaired in their decision-making ability, because a substantial proportion of patients who perform adequately according to any one standard may do poorly when judged according to another standard. Additional data are needed to allow identification of patients' characteristics that are associated with poor performance on each standard.

Often, however, clinicians or courts use a combination of standards to assess decision-making capacity.

Our data speak to both the wisdom and the risks of that approach. Clearly, no single standard is sufficiently broad to identify all persons whose decision making may be substantially impaired. Given that the cognitive functions associated with each standard appear to be affected differently by psychiatric and medical illnesses, a compound standard is useful in preventing many decision makers with substantial cognitive deficiencies from "slipping through the net." On the other hand, as standards are added to one another, an ever-increasing proportion of persons may be excluded from making their own decisions about treatment. Substantially raising the threshold of demands for decision-making abilities could compromise the principle—fundamental to the view of individuals' rights that is dominant in our society—that almost all persons should be permitted to make decisions affecting their well-being.

There have been suggestions in the literature on decision-making capacities and legal competence that the single standards commonly used may form a hierarchy (10, 11). If this is true, it would allow us to talk about more and less stringent standards for competence and to fashion our policies accordingly. The data described in this article, however, generally do not support this assertion. Although the standard of evidencing a choice (which we have not examined in detail here) is clearly less demanding than the three primary standards, there is no clear hierarchy of rigor among the remaining three. Each of them appears to tap functionally different areas, although data we have presented elsewhere (7) indicate a somewhat greater overlap between understanding and reasoning than between either of these and appreciation.

The weight given to our conclusions should reflect the methodological strengths and weaknesses of the study. We have demonstrated elsewhere that our instruments have high levels of interrater reliability and concurrent validity (7, 8), but our conclusions depend on the accuracy with which the instruments assess abilities related to legal standards for competence. Moreover, the characterization of particular scores as indicating impaired performance does not necessarily mean that the courts would designate persons scoring in that range as legally incompetent. Finally, the actual proportion of patients in hospital populations who manifest impaired decision-making abilities may be somewhat

greater than our data would suggest, because some patients at our study sites were too acutely ill to participate as research subjects and presumably would have manifested substantial impairment of cognitive capacities. Nevertheless, it is not likely that our conclusions about the effects of applying the various individual or compound standards would be altered by moderate adjustments in where the line is drawn to designate impairment, or by the addition of a few patients to the impaired categories within the various standards.

What are the implications of these findings for clinicians? The relatively distinct populations identified by each standard suggest that the standard(s) to be applied in judgments about patients' decision-making capacity must be selected with care. Clinicians should make themselves aware of their jurisdictions' applicable standards—which will usually consist of one or more of the standards reviewed in this article—and should structure their examinations accordingly. When compound standards are used, the absence of a clear hierarchy means that each standard will have to be assessed independently; success or failure on one standard does not portend a similar performance on others.

Where to fix the threshold of competence or incompetence on each of these standards, however, is a question that is not answered by the data in this study. As Drane (10) pointed out, there may be no single answer to the question of the degree of capacity that should be required for legal competence. The correct threshold may vary from case to case, depending on the clinical consequences of acceptance or refusal of treatment, as well as the risks and benefits of treatment alternatives, in individual cases.

For policy makers, the data illustrate the importance of using caution in fashioning standards for competence. Which standards are designated by the courts or legislatures, assuming that they are applied diligently, may make a large difference in how many patients and

which patients are denied the right to make their own treatment decisions. Adoption of uniform standards across jurisdictions should be encouraged in order to allow clinicians and researchers to generalize their knowledge and refine techniques for standardized assessment. This will promote accuracy in assessment and reduce the harms associated with underidentification and overidentification of persons as lacking sufficient decision-making capacity.

REFERENCES

1. Strain JJ, Taintor Z, Gise LH, Spikes J: Informed consent—mandating the consultation. *Gen Hosp Psychiatry* 1985; 7:228–233
2. Appelbaum PS: The right to refuse treatment with antipsychotic medications: retrospect and prospect. *Am J Psychiatry* 1988; 145: 413–419
3. Appelbaum PS, Grisso T: Assessing patients' capacities to consent to treatment. *N Engl J Med* 1988; 319:1635–1638
4. Roth LH, Meisel A, Lidz CW: Tests of competency to consent to treatment. *Am J Psychiatry* 1977; 134:279–284
5. Appelbaum PS, Grisso T: The MacArthur Treatment Competence Study, I: mental illness and competence to consent to treatment. *Law and Human Behavior* 1995; 19:105–126
6. Marcus SC, Robins LN, Buchholz KK, Przybeck TR: Diagnostic Interview Schedule Screening Instrument. St Louis, Washington University School of Medicine, Department of Psychiatry, 1989
7. Grisso T, Appelbaum PS: The MacArthur Treatment Competence Study, III: abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior* 1995; 19:149–174
8. Grisso T, Appelbaum PS, Mulvey E, Fletcher K: The MacArthur Treatment Competence Study, II: measures of abilities related to competence to consent to treatment. *Law and Human Behavior* 1995; 19:127–148
9. Morreim EH: Competence: at the intersection of law, medicine and philosophy, in *Competency: A Study of Informal Competency Determinations in Primary Care*. Edited by Cutter MAG, Slep EE. Dordrecht, The Netherlands, Kluwer Academic, 1991
10. Drane J: Competency to give an informed consent: a model for making clinical assessments. *JAMA* 1984; 252:925–927
11. Tancredi L: Competency for informed consent: conceptual limits of empirical data. *Int J Law Psychiatry* 1982; 5:51–63