

Determining when impairment constitutes incapacity for informed consent in schizophrenia research

SCOTT Y. H. KIM, PAUL S. APPELBAUM, JEFFREY SWAN, T. SCOTT STROUP, JOSEPH P. McEVOY, DONALD C. GOFF, DILIP V. JESTE, J. STEVEN LAMBERTI, ADRIAN LEIBOVICI and ERIC D. CAINE

Background Although people with schizophrenia display impaired abilities for consent, it is not known how much impairment constitutes incapacity.

Aims To assess a method for determining the categorical capacity status of potential participants in schizophrenia research.

Method Expert-judgement validation of capacity thresholds on the sub-scales of the MacArthur Competence Assessment Tool – Clinical Research (MacCAT–CR) was evaluated using receiver operating characteristic (ROC) analysis in 91 people with severe mental illness and 40 controls.

Results The ROC areas under the curve for the understanding, appreciation and reasoning sub-scales of the MacCAT–CR were 0.94 (95% CI 0.88–0.99), 0.85 (95% CI 0.76–0.94) and 0.80 (95% CI 0.70–0.90). These findings yielded negative and positive predictive values of incapacity that can guide the practice of investigators and research ethics committees.

Conclusions By performing such validation studies for a few categories of research with varying risks and benefits, it might be possible to create evidence-based capacity determination guidelines for most schizophrenia research.

Declaration of interest D.C.G. is on the advisory boards of several pharmaceutical companies. Funding detailed in Acknowledgements.

The ethics of research involving adults with impaired decision-making capacity remains a focus of policy discussions in the USA (Kim *et al*, 2004), of policy statements internationally (UNESCO, 2005), and a subject of new legislation in the UK (Adults with Incapacity (Scotland) Act 2000; Mental Capacity Act 2005) and two US states (Kim *et al*, 2004). In particular, research involving people with schizophrenia has been controversial because as a group they have greater decisional impairment than healthy controls (Carpenter *et al*, 2000; Kovnick *et al*, 2003; Palmer *et al*, 2004). However, diagnosis cannot be equated with decisional incapacity because there is too much heterogeneity in decisional abilities (Grisso & Appelbaum, 1995b; Carpenter *et al*, 2000; Palmer *et al*, 2004). Although there are now instruments for assessing decisional ability, we currently lack an evidence-based method for translating those dimensional data into categorical judgements (Kim, 2006).

In this study, we used the judgements of independent clinicians experienced in capacity assessments to address the following question: given that people with schizophrenia exhibit a range of decisional abilities, how can we use a standardised instrument to distinguish those who are capable from those who are incapable of informed consent? We asked the question in the context of a unique opportunity presented by a multisite clinical trial, funded by the National Institute of Mental Health, the Clinical Antipsychotic Trials of Intervention Effectiveness – Schizophrenia (CATIE; Stroup *et al*, 2003), which used as part of its research protocol the most widely tested measure of decisional ability, the MacArthur Competence Assessment Tool – Clinical Research (MacCAT–CR; Appelbaum & Grisso, 2001).

METHOD

Participants

In line with the aim of the project, the goal of recruitment was to ensure that a

sufficient spectrum of decision-making abilities was represented in our sample, rather than a random sample of a particular population. Participants included 91 people with severe mental illness and 40 people in the community comparison group. The group with severe mental illness consisted of two subgroups: 55 participants in the CATIE–Schizophrenia study at six different sites across the USA; and 36 people who were not part of the CATIE study but were recruited specifically for this interview study from two out-patient clinics serving people with severe and persistent mental illnesses, and from in-patient units at a state hospital in Rochester, New York, USA. Those who were not part of the CATIE study were added to ensure a sufficient spectrum (i.e. to avoid spectrum bias; Zhou *et al*, 2002) of decision-making ability; we noticed in the early part of the study that the performance of those in the CATIE study tended to cluster in the upper end – a trend that was ultimately borne out in the overall CATIE–Schizophrenia sample (Stroup *et al*, 2005). The participants in the control group were all without psychosis and were recruited in Rochester through advertisements in the community, in support staff work areas of a general hospital and at an out-patient substance misuse recovery programme.

This study was approved by the research ethics committees (institutional review boards) of all participating institutions, and all participants provided written informed consent after full disclosure of study elements. The CATIE participants provided separate informed consent for this ancillary study. For the group with severe mental illness, as has been done in other studies of this kind (Moser *et al*, 2002; Stroup *et al*, 2006), given the low risk of this interview study, a relatively undemanding standard for capacity to consent was used.

Measures

Participants were videotaped during their assessment with the MacCAT–CR (Appelbaum & Grisso, 2001). The MacCAT–CR has been extensively used in people with schizophrenia (Carpenter *et al*, 2000; Dunn *et al*, 2002; Moser *et al*, 2002; Stroup *et al*, 2005) and people with major depression (Appelbaum *et al*, 1999) and dementia (Kim *et al*, 2001), and is a companion instrument to the MacArthur Competence Assessment Tool for Treatment (MacCAT–T) (Cairns *et al*, 2005a,b).

The MacCAT-CR contains pertinent disclosure elements of informed consent and is designed to be adapted to specific research protocols, to reflect the task-specific nature of decisional capacity (Appelbaum & Grisso, 2001). The version used in the CATIE-Schizophrenia study was used for all participants in this study; thus, the non-CATIE and control participants were asked to imagine being invited to participate in the CATIE study as their decisional abilities were assessed. This procedure is commonly employed in capacity research (Carpenter *et al*, 2000; Moser *et al*, 2002).

The MacCAT-CR is structured according to the four-abilities model of decision-making capacity (Grisso & Appelbaum, 1998). These include 'understanding [emphasis added] of disclosed information about the nature of the research project and its procedures (13 items for a possible total score of 26 – each item in the MacCAT-CR has a score range of 0–2 with objective scoring criteria); appreciation of the effects of research participation (or failure to participate) on subjects' own situations (3 items for a possible total score of 6); reasoning about participation (4 items for a possible total score of 8); and ability to communicate a choice (one item for a possible total score of 2)' (Appelbaum *et al*, 1999). Data on the ability to communicate a choice will not be discussed here as almost everyone received a full score. The MacCAT-CR does not provide a global score because requirements for each ability related to capacity can vary by jurisdiction and according to the decisional demands of a given study (Grisso & Appelbaum, 1995a). However, it is important to note that the four-abilities model is based on an extensive review of laws, court decisions and ethics literature, such that it provides a reasonable approximation of the standards for capacity broadly laid out in statutes. Thus, researchers have been able to use the MacCAT instruments to approximate, for example, the criteria of the Mental Capacity Act 2005 (Cairns *et al*, 2005a,b).

The final MacCAT-CR sub-scale ratings for all 131 participants used for analysis were made by J.S. During the course of the project, the principal investigator (S.K.) independently scored 36 out of the 131 interviews. This was the basis for calculations of interrater reliability. For those 36 participants, discrepancies arising after independent scoring of MacCAT-CR items by the two raters were

resolved through discussion between the two raters. The intraclass correlation coefficients for total scores of MacCAT-CR subscales were 0.93 for understanding ($F=29.3$, $d.f.=35.0$, 35 , $P<0.0001$), 0.89 for appreciation ($F=16.7$, $d.f.=35.0$, 35 ; $P<0.0001$), and 0.84 for reasoning ($F=11.3$, $d.f.=35.0$, 35 , $P<0.0001$).

Psychiatric diagnoses were made using medical records and the Structured Clinical Interview for DSM-IV (SCID-IV; First *et al*, 1997). Severity of psychiatric symptoms was measured using the Positive and Negative Syndrome Scale (PANSS; Kay *et al*, 1987), which includes positive, negative and general psychopathology sub-scales. Control participants were administered the SCID only.

Expert judgements

Three psychiatrists with experience in assessing decisional capacity (two consultation psychiatrists and one board-certified geriatric psychiatrist) were recruited to serve as expert judges and a fourth judge was added as a back-up. The judges were prepared for their task by informing them of the basic outlines of the CATIE-Schizophrenia study (the rationale, the medications to be tested, the total number of participants to be enrolled, and the fact that treatment failures would lead to re-randomisation with a new study drug). They were told that their job was to render a categorical judgement based on viewing an interview of a semi-structured capacity assessment (but they were unaware of the actual MacCAT-CR scores). The ultimate goal of deriving a final judgement was explained as: 'Your task is to review the tapes carefully and make a categorical judgement (definitely capable, probably capable, probably not capable, and definitely not capable). In the real world, decisions need to be made even if things aren't clear, as we reduce complex clinical data into a yes/no judgement'. They also rated the statement: 'The videotaped interview gave me sufficient basis to make my decision in this case' on a 5-point Likert scale ranging from 'strongly agree=1' to 'strongly disagree=5.'

The final categorical status of each participant was determined by collapsing the 'definite' and 'probable' categories of the experts' responses to create a dichotomous variable and then using a majority (2 out of 3) or better (3 out of 3) agreement among the three expert judges to determine

the final status (Kim *et al*, 2001). Owing to unavoidable circumstances, two of the three original raters were not able to complete all of the interviews. However, we had a back-up expert judge (a psychiatrist trained in both internal medicine and psychiatry, who primarily works with people with schizophrenia) whose scores were used whenever there was a missing judgement among the first three judges. The experts rendered their categorical judgements independently of one another.

Categorical capacity judgements were rendered for 101 participants: 90 with severe mental illness and 11 controls. The videotape for one participant with mental illness was not used because the sound quality was poor. Moreover, because of lower variance with higher performance (ceiling effect) in the comparison group, we only used 11 of the 40 tapes, including the two lowest scoring control participants. Expert judge 1 reviewed all 101 interviews, judge 2 reviewed 79 interviews and judge 3 reviewed 91 interviews. There were no participants who had a missing judgement from more than one judge. The back-up expert judge rendered judgements for 72 interviews and, of those, 32 judgements in which there was a missing judgement from either judge 2 or judge 3 were used in the final determination of capacity status (the back-up judge rated more than the 32 participants with missing ratings to assess reliability among all four judges).

The rationale and methodology for the expert judgement criterion method have been described elsewhere (Kim, 2006), including its advantages over an *a priori* cut-off criterion (Wirshing *et al*, 1998; Moser *et al*, 2002) and a psychometric criterion (Marson *et al*, 1995; Grisso *et al*, 1997; Schmand *et al*, 1999; Kovnick *et al*, 2003). Given that most societies look to clinicians' judgement about such decisions, expert judgement offers an arguably more valid standard against which to measure participant performance. Methodologically, expert judgement provides an independent assessment criterion, since the experts are not affiliated with the schizophrenia research studies.

Statistical analyses

Group comparisons of demographic, symptom severity and MacCAT-CR summary data were conducted using parametric or non-parametric tests. Pairwise and group

kappa coefficients were calculated to assess categorical agreement among expert judges. Receiver operating characteristic (ROC) analysis was conducted to assess the test characteristics of each of the three sub-scales (understanding, appreciation and reasoning) of MacCAT-CR against the final categorical judgements made by the expert judges. To demonstrate how the sensitivity and specificity data generated from the ROC analysis can be applied to potential research scenarios, we calculated the positive and negative predictive values (PPVs and NPVs) for a range of hypothetical prior probabilities for three cut-off points on the understanding sub-scale.

Data were analysed using SPSS version 12.0 and Stata version 8.0 (both for Windows).

RESULTS

The group with severe mental illness and controls showed no significant differences in age, gender and race distribution, or educational level (Table 1). None of the controls had a psychotic disorder; 6 had a mood disorder, 1 a substance dependence disorder and 1 an anxiety disorder. Seventy-five of the group with severe mental illness had schizophrenia, 14 had schizoaffective disorder and 2 had affective

disorders with psychosis. Among the 55 participants from the CATIE study, 49 had schizophrenia and 6 had schizoaffective disorder.

Performance on MacCAT-CR

Those with severe mental illness performed significantly worse than the comparison group on the MacCAT-CR sub-scales (except for choice). Within this group, the 55 participants from the CATIE study performed better than the other participants on all sub-scales of the MacCAT-CR: understanding, mean score (s.d.) 21.3 (3.7) *v.* 19.1 (5.6), $t=2.1$, *d.f.*=54.7, $P=0.04$; appreciation, 4.1 (1.5) *v.* 3.4 (1.8), $t=2.0$, *d.f.*=66.0, $P=0.05$; reasoning, 5.4 (1.5) *v.* 4.8 (2.3), $t=1.4$, *d.f.*=54.2, $P=0.17$. This is consistent with our goal of avoiding spectrum bias by expanding the range of scores in the group with severe mental illness.

Expert judgements

Of the 101 people reviewed, 25 (including 7 of the 55 CATIE participants) were deemed probably or definitely incapable of consent. The pairwise kappa coefficients among the four judges ranged from 0.56 to 0.90; the group kappa coefficient for the four expert judges was 0.69 ($Z=14.1$,

$P<0.001$). When asked whether or not the videos provided a sufficient basis for them to make their capacity determinations, the mean rating ranged between strongly agree=1 and agree=2 for three of the experts, with mean (s.d.) ratings of 1.4 (0.9), 1.4 (0.8) and 1.9 (0.9), and between agree=2 and neutral=3 for the remaining expert judge, whose mean rating was 2.5 (1.1).

Predictive values of MacCAT-CR scores

Table 2 summarises the sensitivity and specificity using various cut-off points on the three sub-scales of the MacCAT-CR. The area under the ROC curve was higher for the understanding sub-scale at 0.94 (95% CI 0.88–0.99) than for the appreciation sub-scale (0.85, 0.76–0.94) and the reasoning sub-scale (0.80, 0.70–0.90), indicating that MacCAT-CR scores, especially for understanding, were significant predictors of categorical capacity status. However, none of the sub-scales had a single cut-off score with a very high sensitivity and specificity.

Sensitivity and specificity are features of tests, not populations, and cannot guide decisions without information about prevalence. For the purpose of determining the acceptable capacity scores that might be recommended (for instance to a research ethics committee reviewing a research protocol to be used to screen people with impaired capacity), the results of the ROC analysis were used to generate positive predictive values (PPV, the probability that a person found to perform at or below a MacCAT-CR sub-scale cut-off score will in fact be incapable) and negative predictive values (NPV, the probability that a person performing above the cut-point will in fact be capable), as shown in Table 3.

A high PPV implies a low false-positive rate (i.e. low likelihood of mistakenly excluding a capable person); a high NPV implies a low false-negative rate (i.e. low likelihood of mistakenly enrolling an incapable person). In determining what degree of decisional capacity to require of research participants, it would be undesirable to use a high cut-off score when prevalence is low (e.g. understanding score of 21 at 10% prevalence of incapacity) because 76% of persons excluded as too impaired will in fact be capable (given the PPV of 24%). Such a practice would not only be inefficient but also would unfairly exclude willing and capable persons from participating in research. It would also be

Table 1 Participants' characteristics and performance on the MacArthur Competence Assessment Tool—Clinical Research

	Participants with severe mental illness (n=91)	Controls (n=40)	Test statistic		P
			χ^2	t	
Female gender, n (%)	23 (25)	14 (35)	1.3		0.26
Black or minority, n (%)	31 (34)	8 (20)	2.6		0.11
Age, years: mean (s.d.)	42.2 (10.2)	39.9 (10.9)		1.15	0.25
Education, ¹ years: mean (s.d.)	11.8 (2.1)	12.5 (2.1)		-1.64	0.10
PANSS score ² : mean (s.d.)					
Positive	16.8 (5.9)				
Negative	18.3 (6.2)				
General	37.2 (10.6)				
MacCAT-CR score: mean (s.d.)					
Understanding	20.4 (4.6)	25.0 (2.4)		-7.4	<0.001
Appreciation	3.8 (1.6)	5.1 (1.0)		-5.3	<0.001
Reasoning	5.2 (1.9)	7.0 (1.4)		-6.2	<0.001
Choice	1.97 (0.18)	2.0 (0)		-1.8	0.08

MacCAT-CR, MacArthur Competence Assessment Tool—Clinical Research; PANSS, Positive and Negative Syndrome Scale.

1. Missing data for one participant.

2. The CATIE participants' scores were lower (indicating a lower degree of psychopathology) than other participants with severe mental illness: positive, 14.7 *v.* 19.8 ($P<0.001$); negative, 17.4 *v.* 19.6 ($P=0.07$); general, 33.2 *v.* 43.1 ($P<0.001$).

Table 2 Sensitivity and specificity of cut-off scores on sub-scales of the MacArthur Competence Assessment Tool–Clinical Research¹

MacCAT–CR sub-scale cut-off score	Sensitivity ²	Specificity ³
Understanding (scale range 0–26)		
2	0.000	1.000
6	0.040	1.000
9	0.080	1.000
10	0.160	1.000
11	0.200	1.000
12	0.240	1.000
13	0.360	1.000
14	0.400	1.000
15	0.560	1.000
16	0.680	1.000
17	0.760	0.961
18	0.800	0.921
19	0.840	0.803
20	0.840	0.800
21	0.880	0.684
22	0.960	0.618
23	1.000	0.474
24	1.000	0.368
25	1.000	0.171
26	1.000	0.00
Appreciation (scale range 0–6)		
0	0.080	1.000
1	0.360	0.987
2	0.480	0.921
3	0.840	0.750
4	0.920	0.553
5	0.960	0.276
6	1.000	0.00
Reasoning (scale range 0–8)		
0	0.120	1.000
1	0.160	1.000
2	0.200	0.974
3	0.360	0.947
4	0.600	0.842
5	0.840	0.592
6	0.960	0.368
7	0.960	0.171
8	1.000	0.00

MacCAT–CR, MacArthur Competence Assessment Tool–Clinical Research.

1. Categorical capacity status determined by expert reviews of 101 participants (90 with severe mental illness and 11 controls).

2. Of those deemed incapable of providing consent, the proportion who scored at or below the cut-off score.

3. Of those deemed capable of providing consent, the proportion who scored above the cut-off score.

Table 3 Positive and negative predictive values for three potential cut-off scores on the MacArthur Competence Assessment Tool–Clinical Research understanding sub-scale, for a range of prevalence values

Prevalence of incapacity, %	Cut-off score 15 ¹		Cut-off score 18 ²		Cut-off score 21 ³	
	PPV, %	NPV, %	PPV, %	NPV, %	PPV, %	NPV, %
10	100	95	53	98	24	98
20	100	90	72	95	41	96
30	100	84	81	91	54	93
40	100	77	87	87	65	90
50	100	69	91	82	74	85

PPV, positive predictive value; NPV, negative predictive value.

1. Sensitivity 0.56, specificity 1.0.

2. Sensitivity 0.80, specificity 0.92.

3. Sensitivity 0.88, specificity 0.68.

ethically undesirable to use a low cut-off when the prevalence of incapacity is high (e.g. at 50% incapacity, almost a third of those who test capable will in fact be incapable given the NPV of 69%).

DISCUSSION

Despite increasing research on the decision-making abilities of people with neuropsychiatric disorders, there are few data on how to translate information about impairment into categorical determinations. In the real world, it is necessary to determine the categorical capacity status of the potential participant, i.e. whether they are capable of providing independent informed consent. This information is needed for excluding those who are incapable, for identifying those in need of surrogate decision-makers, or for identifying those who may require remedial education. Thus, an important goal of capacity research in schizophrenia is to inform policies and practices that help guide the determination of categorical capacity status of potential participants (Kim, 2006).

In the research context, informed consent disclosures are relatively consistent across participants, since the relevant information, including the risk–benefit ratio, is determined by the characteristics of a research protocol which is applicable to all potential participants. This is in contrast to the treatment context in which the procedures, risks, benefits and hence disclosures might be unique to each individual's treatment situation, and for whom the assessment of decision-making capacity requires individualised patient information (Cairns *et al*, 2005a). Further, whereas in

the treatment context the welfare of the patient is the physician's paramount concern, in the research context, the investigator's priority is the advancement of science, thus increasing the need for a more transparent and objective process for determination of capacity. Therefore, the research context provides an opportunity as well as an imperative to create a standardised capacity assessment by using an assessment instrument that can be benchmarked against ethically appropriate, methodologically rigorous independent validation provided by experienced clinicians.

Objective determination of capacity for research

Our study establishes the feasibility of an objective assessment of capacity for the research context. By validating the MacCAT–CR sub-scales against an expert judgement standard, we can go beyond mere descriptions of participants' performance on a scale. For example, given that the prior probability of incapacity among those screened for the CATIE–Schizophrenia study was probably quite low (Stroup *et al*, 2005), we can surmise that even a low cut-off score on the understanding sub-scale such as 15 (which was in fact used by the CATIE study) would rarely include people lacking capacity (e.g. at an estimate of 10% prevalence, there would only be a 5% false-negative rate), with virtually no chance of mistakenly identifying those with capacity as incapable.

Since most research studies could probably be categorised into a handful of categories in terms of their risk–benefit ratio (Maryland Attorney General's Research Working Group, 1998; National Bioethics

Advisory Commission, 1998; New York Department of Health Advisory Work Group on Human Subject Research Involving the Protected Classes, 1999), a limited number of targeted validation studies would likely provide a sufficient evidence base for ethically appropriate yet efficient practice for a variety of research studies involving participants with impairment in decision-making. In effect, a series of tables such as Table 3 could provide a systematic and objective guide to research ethics committees and investigators.

How important is the fact that the sub-scales of the MacCAT-CR do not seem to have a single cut-off point that has both high sensitivity and specificity? First, it might simply be unrealistic to expect extreme precision and predictability from a standardised instrument when it is applied to making complex, value-laden judgements about a person's decision-making capacity. Second, this limitation might not be a problem as long as the purpose of assessing capacity is clear; for instance, one might focus more heavily on the PPV or the NPV, depending on the situation. So for an early-phase study of an invasive intervention likely to yield no benefit to the participants but which may pose some risk, the thresholds could be set high enough to eliminate any one who lacks capacity (false-negatives). Alternatively, if such a method eliminates too many potential participants as false-positives, a two-step approach could be used: a lower MacCAT-CR threshold to decrease false-positives and then individual in-depth capacity assessments to ensure that only competent persons are enrolled. As this last example illustrates, we believe that our method should be used flexibly to meet the ethical requirements of the situation, rather than rigidly adhering to a formula.

Strengths

To our knowledge, this is the most thorough validation against expert judgement of capacity thresholds on a widely used capacity assessment tool for clinical research, and the first such validation study involving people with schizophrenia. This study has several strengths. The majority of participants with severe mental illness were in an actual clinical trial. This group exhibited a wide spectrum of impairment, allowing us to conduct a meaningful ROC analysis. The expert judges achieved

relatively high levels of non-chance agreement, and they felt that in general they had sufficient information to make the capacity determinations. The expert judges based their judgements on video recordings of interviews, which provided more information than written transcripts.

Limitations

There are however some limitations to the study and caveats. First, our sample consisted of both CATIE participants and people with severe mental illness who were not in the CATIE study. The latter were included to ensure a sufficient spectrum of performance for the ROC analysis. Thus, no generalisations regarding the relative performance of the subgroups should be drawn from our data. The CATIE participants might have performed better on the MacCAT-CR because they had a less severe illness but also because, being involved in the study, the study protocol had been previously explained to them in more depth.

Second, the experts' judgements were based on their viewing the taped MacCAT-CR interviews rather than performing their own independent assessments (which was not feasible in this multisite study involving multiple expert judges). Thus, our method is susceptible to incorporation bias that can falsely increase the accuracy of the test (Zhou *et al*, 2002). However, this limitation must be weighed against the following countervailing considerations. Currently, there is a lack of standardised procedures for capacity determinations. The MacCAT-CR covers the essential elements of a capacity assessment (Appelbaum & Grisso, 2001) and its standardised nature mitigates the variability of capacity assessments. In the absence of agreed procedures for capacity assessments, a criterion standard based on various experts' evaluations (even if it were feasible in a multisite study such as this) would involve a variety of methods, creating uncertainties regarding the nature of the standards used. Thus, although we cannot rule out the possibility of incorporation bias, we believe our results represent a reasonable balance between feasibility and validity.

Third, before the results of our study are generalised to other contexts, one must take into account the potential adverse effects of focusing on 'cut-off scores' of capacity assessment instruments (Grisso & Appelbaum, 1996), especially the danger

that the cut-off scores will be seen as inherent features of the assessment instrument (i.e. anyone scoring above a certain level has adequate capacity for any decision), rather than needing context-by-context validation and context-sensitive application. To avoid such misuse, any generalisation of our validation method must take into account two points.

First, the prevalence of incapacity in schizophrenia studies other than the CATIE study might be different for a variety of reasons. For instance, in studies that target people with refractory illness or those who are long-term in-patients (Kovnick *et al*, 2003), the prevalence rates will be higher and the estimation of PPV and NPV will need to take that into account. Second, any attempt to generalise our validation method to other schizophrenia studies must take into account the risk-sensitive nature of capacity thresholds (Brock, 1991; Grisso & Appelbaum, 1998; National Bioethics Advisory Commission, 1998). In our study, the expert raters made judgements regarding the level of capacity that was adequate for a relatively low-risk clinical trial. However, the risk-benefit ratio might be different for studies involving placebos, symptom provocation, or phase I tests of invasive interventions. The ROC curves for such studies might look quite different from those in this study.

Finally, the fact that 7 of 55 CATIE study participants were deemed to lack capacity by our experts needs to be interpreted with caution. Our CATIE sample was not intended to be representative of the overall CATIE study. The ratings of the expert judges were not available to the CATIE investigators at the time that they made their judgements regarding admission to the CATIE study. Further, a number of unique safeguards (Stroup *et al*, 2005), including independent participant advocates (Stroup & Appelbaum, 2003), were built into the CATIE project. Finally, in the absence of a true 'gold standard' for determining categorical status, we are proposing the expert judgement-based method as a provisional criterion standard that needs to be further studied and improved (Kim, 2006).

Future directions

The results of our study provide an evidence-based decision framework for how to use instruments for measuring

decisional abilities to guide valid categorical judgements about a potential participant's capacity to give informed consent. We believe that as long as its limitations and caveats are kept in mind, future research employing the framework provided in our study could have important practical implications. By performing validation studies for a few categories of risk-benefit situations, it might even be possible to interpolate reasonable guidelines for most schizophrenia research studies. Such an approach would make the crucial task of determining a potential participant's capacity status much more transparent, objective and evidence-based than it is today.

ACKNOWLEDGEMENTS

We thank Linda Ryan, MD, Lior Givon, MD and Telva Olivares, MD for their expert ratings, Sonia Davis, PhD for assistance with database management and Jayendra Patel, MD for assistance with recruitment. The study was supported by the National Institute of Mental Health USA (grants K23 MH64172 and NOI MH90001).

REFERENCES

- Appelbaum, P. & Grisso, T. (2001)** *MacCAT-CR: MacArthur Competence Assessment Tool for Clinical Research*. Professional Resource Press.
- Appelbaum, P. S., Grisso, T., Frank, E., et al (1999)** Competence of depressed patients for consent to research. *American Journal of Psychiatry*, **156**, 1380–1384.
- Brock, D. W. (1991)** Decisionmaking competence and risk. *Bioethics*, **5**, 105–112.
- Cairns, R., Maddock, C., Buchanan, A., et al (2005a)** Reliability of mental capacity assessments in psychiatric in-patients. *British Journal of Psychiatry*, **187**, 372–378.
- Cairns, R., Maddock, C., Buchanan, A., et al (2005b)** Prevalence and predictors of mental incapacity in psychiatric in-patients. *British Journal of Psychiatry*, **187**, 379–385.
- Carpenter, W. T., Jr, Gold, J., Lahti, A., et al (2000)** Decisional capacity for informed consent in schizophrenia research. *Archives of General Psychiatry*, **57**, 533–538.
- Dunn, L., Lindamer, L., Palmer, B. W., et al (2002)** Improving understanding of research consent in middle-aged and elderly patients with psychotic disorders. *American Journal of Geriatric Psychiatry*, **10**, 142–150.
- First, M., Spitzer, R., Gibbon, M., et al (1997)** *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-I)*. American Psychiatric Publishing.
- Grisso, T. & Appelbaum, P. S. (1995a)** Comparison of standards for assessing patients' capacities to make treatment decisions. *American Journal of Psychiatry*, **152**, 1033–1037.
- Grisso, T. & Appelbaum, P. S. (1995b)** The MacArthur Treatment Competence Study. III. Abilities of patients to consent to psychiatric and medical treatments. *Law and Human Behavior*, **19**, 149–174.
- Grisso, T. & Appelbaum, P. S. (1996)** Values and limits of the MacArthur Treatment Competence Study. *Psychology, Public Policy, and Law*, **2**, 167–181.
- Grisso, T. & Appelbaum, P. S. (1998)** *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. Oxford University Press.
- Grisso, T., Appelbaum, P. S. & Hill-Fotouhi, C. (1997)** The MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. *Psychiatric Services*, **48**, 1415–1419.
- Kay, S., Fiszbein, A. & Opler, L. (1987)** The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, **13**, 261–276.
- Kim, S. Y. H. (2006)** When does decisional impairment become decisional incompetence? Ethical and methodological issues in capacity research in schizophrenia. *Schizophrenia Bulletin*, **32**, 92–97.
- Kim, S. Y. H., Caine, E. D., Currier, G. W., et al (2001)** Assessing the competence of persons with Alzheimer's disease in providing informed consent for participation in research. *American Journal of Psychiatry*, **158**, 712–717.
- Kim, S. Y. H., Appelbaum, P. S., Jeste, D. V., et al (2004)** Proxy and surrogate consent in geriatric neuropsychiatric research: update and recommendations. *American Journal of Psychiatry*, **161**, 797–806.
- Kovnick, J. A., Appelbaum, P. S., Hoge, S. K., et al (2003)** Competence to consent to research among long-stay inpatients with chronic schizophrenia. *Psychiatric Services*, **54**, 1247–1252.
- Marson, D. C., Ingram, K. K., Cody, H. A., et al (1995)** Assessing the competency of patients with Alzheimer's disease under different legal standards. A prototype instrument. *Archives of Neurology*, **52**, 949–954.
- Maryland Attorney General's Research Working Group (1998)** *Final Report of the Attorney General's Research Working Group*. Maryland Attorney General's Office.
- Moser, D. J., Schultz, S. K., Arndt, S., et al (2002)** Capacity to provide informed consent for participation in schizophrenia and HIV research. *American Journal of Psychiatry*, **159**, 1201–1207.
- National Bioethics Advisory Commission (1998)** *Research Involving Persons with Mental Disorders That May Affect Decisionmaking Capacity*, Vol. I. National Bioethics Advisory Commission.
- New York Department of Health Advisory Work Group on Human Subject Research Involving the Protected Classes (1999)** *Recommendations on the Oversight of Human Subject Research Involving the Protected Classes* (<http://www.nysl.nysed.gov/scandoclinks/jocm49377072.htm>). New York (State) Department of Health.
- Palmer, B. W., Dunn, L. B., Appelbaum, P. S., et al (2004)** Correlates of treatment-related decision-making capacity among middle-aged and older patients with schizophrenia. *Archives of General Psychiatry*, **61**, 230–236.
- Schmand, B., Gouwenberg, B., Smit, J. H., et al (1999)** Assessment of mental competency in community-dwelling elderly. *Alzheimer Disease and Associated Disorders*, **13**, 80–87.
- Stroup, S. & Appelbaum, P. S. (2003)** The subject advocate: protecting the interests of participants with fluctuating decisionmaking capacity. *IRB Ethics and Human Research*, **25**, 9–11.
- Stroup, T., McEvoy, J., Swartz, M., et al (2003)** The NIMH–CATIE Project: schizophrenia trial design and protocol development. *Schizophrenia Bulletin*, **29**, 15–31.
- Stroup, S., Appelbaum, P., Swartz, M., et al (2005)** Decision-making capacity for research participation among individuals in the CATIE schizophrenia trial. *Schizophrenia Research*, **80**, 1–8.
- Stroup, T. S., Alves, W. M., Hamer, R. M., et al (2006)** Clinical trials for antipsychotic drugs: design conventions, dilemmas and innovations. *Nature Reviews Drug Discovery*, **5**, 133–146.
- UNESCO (2005)** *Universal Declaration on Bioethics and Human Rights*. UNESCO.
- Wirshing, D. A., Wirshing, W. C., Marder, S. R., et al (1998)** Informed consent: assessment of comprehension. *American Journal of Psychiatry*, **155**, 1508–1511.
- Zhou, X.-H., Obuchowski, N. & McClish, D. (2002)** *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons.

Correspondence: Dr Scott Y. H. Kim, 300 North Ingalls, Ann Arbor, MI 48109-0429, USA. Email: scottkim@umich.edu

(First received 13 November 2006, final revision 13 March 2007, accepted 23 March 2007)