

Reliability of mental capacity assessments in psychiatric in-patients[†]

RUTH CAIRNS, CLEMENTINE MADDOCK, ALEC BUCHANAN,
ANTHONY S. DAVID, PETER HAYWARD, GENEVRA RICHARDSON,
GEORGE SZMUKLER and MATTHEW HOTOPF

Background Previous work on the reliability of mental capacity assessments in patients with psychiatric illness has been limited.

Aims To describe the interrater reliability of two independent assessments of capacity to consent to treatment, as well as assessments made by a panel of clinicians based on the same interview.

Method Fifty-five patients were interviewed by two interviewers 1–7 days apart and a binary (yes/no) capacity judgement was made, guided by the MacArthur Competence Assessment Tool for Treatment (MacCAT–T). Four senior clinicians used transcripts of the interviews to judge capacity.

Results There was excellent agreement between the two interviewers for capacity judgements made at separate interviews ($\kappa=0.82$). A high level of agreement was seen between senior clinicians for capacity judgements of the same interview (mean $\kappa=0.84$).

Conclusions In combination with a clinical interview, the MacCAT–T can be used to produce highly reliable judgements of capacity.

Declaration of interest None. The study was funded by the Wellcome Trust.

Mental capacity or competence forms the cornerstone of consent to treatment. Until recently it was commonly presumed that serious mental illness, by definition, rendered a patient incapable of consenting to treatment (Grisso & Appelbaum, 1998). This has now been challenged (Appelbaum *et al*, 1995; Grisso & Appelbaum, 1995a,b; Kitamura *et al*, 1998; Wong *et al*, 2000) and lawyers and some psychiatrists have voiced concerns that the legal framework for the treatment of those with severe mental illness is outdated. In contrast to treatment for a physical disorder, where the decision of a capable adult must be respected, mental health legislation in many jurisdictions can override ‘competent’ psychiatric patients’ decisions to withhold consent for treatment of their disorders (Bellhouse *et al*, 2003). In other words, respect for patient autonomy is not absolute in the same way as in legislation for the treatment of physical illnesses.

It is against this background that attention has turned towards the assessment of mental capacity in individuals with mental disorder. The Expert Committee that advised the British Government on reform of the England and Wales Mental Health Act 1983 suggested that capacity should be a significant criterion in a new Mental Health Act (Expert Committee, 1999). This would bring mental health legislation more in line with established principles governing other healthcare decisions. In general, an individual would have to lack capacity before involuntary powers could be used and this absence of capacity would presumably have to be established on the basis of the independent judgements of two mental health clinicians applying the same test. Although the recommendation was rejected and was not included in the original or revised Draft Mental Health Bill (Department of Health, 2002, 2004), one criticism of a capacity-based Mental Health Act has been that assessments of capacity in the mental health setting are no less fraught than those

of say, risk or treatability (Fulford & Sayce, 1998).

METHOD

Aims

The study aimed to establish the interrater reliability when two research interviewers (RI 1 and RI 2) made capacity assessments, at different times, on the same patient. The interrater reliabilities between the binary capacity ratings of RI 1 and RI 2 against ratings by a panel of experts for the same interview were estimated. The panel’s ratings were based on transcripts of the MacArthur Competence Assessment Tool for Treatment (MacCAT–T; Grisso *et al*, 1997) interviews. We hypothesised that providing the panel with additional clinical information would improve the level of agreement compared with when only the transcript was available. For RI 1, only the transcript was used. For RI 2, additional clinical information was provided. We estimated the interrater reliabilities for the MacCAT–T sub-scales between research interviewers and an expert clinician, based on the same interview. We compared the assessment of capacity by the research interviewers and the clinical team. The scheme of ratings is illustrated in Fig. 1.

Participants

A consecutive sample of patients newly admitted to three acute admission wards for general adult psychiatric patients at the Maudsley Hospital, London, was approached for inclusion in this study between October 2003 and February 2004. These wards cover the catchment area of South Southwark, an inner-city deprived area, with a large population of diverse ethnic groups. Participants were seen by a research interviewer within 6 days of admission. The second interview (by the second research interviewer) was completed within 1–7 days of the first interview. The order of the interviews by the two interviewers (C.M. and R.C.) varied, as C.M. and R.C. were each responsible for recruitment from one of two wards and responsibility for recruitment from the third ward alternated on a weekly basis.

The local research ethics committee approved the study. After complete description of the study to the participants, written informed consent was obtained. There are potential problems in conducting research on patients who may lack the capacity to

[†]See pp. 379–385, this issue.

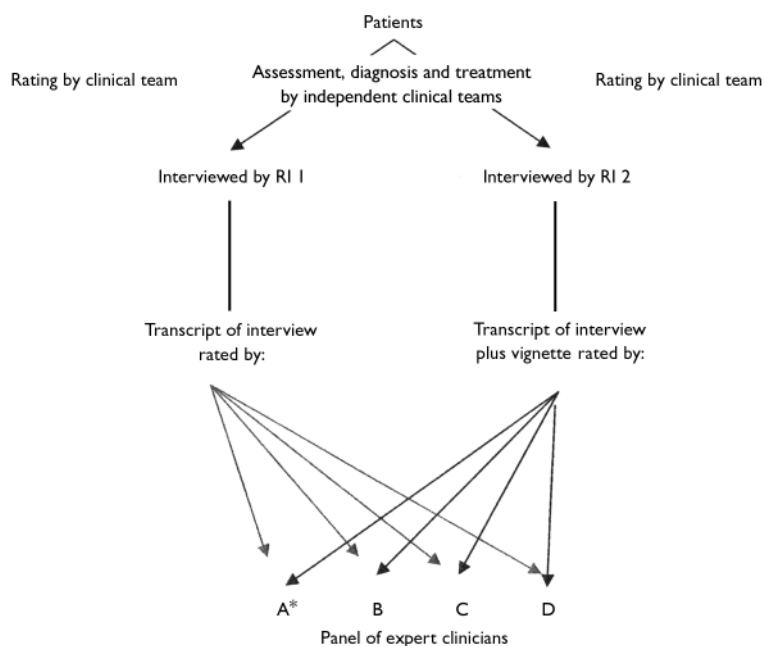


Fig. 1 Scheme of ratings. *Expert clinician A also rated MacArthur Competence Assessment Tool for Treatment (MacCAT-T) sub-scales.

consent (Doyal, 1997; Osborn, 1999; Gunn *et al*, 2000). A small number of patients were deemed too disturbed to participate by medical or senior nursing staff and it was therefore not possible to infer assent to the study. Other reasons for exclusion were being on no prescribed psychotropic medication or receiving medication for the sole purpose of a medically assisted alcohol detoxification, and speaking no English.

Measurement of capacity

The MacCAT-T was administered to the patient in both interviews. It is a semi-structured interview that provides relevant treatment information for the patient and evaluates capacity in terms of its different components. As such it can detect impairment in four areas: the patient’s *understanding* of the disorder and treatment-related information; *appreciation* of the significance of that information for the patient, in particular the benefits and risks of treatment; the *reasoning* ability of the patient to compare their prescribed treatment with an alternative treatment (and the impact of these treatments on their everyday life); and ability of the patient to *express a choice* between their recommended treatment and an alternative treatment. The interview was modified for

the purpose of our study. Instead of offering an alternative treatment, patients were given the option of ‘no treatment’ as the alternative to their prescribed or ‘recommended’ medication. This was to avoid confusion about the patient’s current treatment and also to prevent potential problems in the relationship between the participant and the treating clinician. This constituted another sub-scale ‘understanding alternative treatment option’.

Before each interview, relevant information about the patient’s diagnosis, presenting symptoms and recommended treatment was obtained from the case notes and discussion with the clinical team. Where a patient was prescribed more than one form of psychotropic medication, the interview focused on the medication that was judged to be the patient’s main treatment. This information was disclosed to the patient during the MacCAT-T interview (which took approximately 20 min to complete) together with standardised information about the features, benefits and risks of the particular recommended treatment (based on UK Psychiatric Pharmacy Group Information leaflets; <http://www.ukppg.org.uk>). The benefits and risks of no treatment were then given. All MacCAT-T interviews were audiotaped and transcribed.

Interrater reliability

On completion of each interview, the research interviewer (C.M. or R.C.) made a judgement about whether the patient did or did not have capacity to make a treatment decision. We describe this as a ‘binary’ assessment of capacity, to distinguish it from performance on the various sub-scales of the MacCAT-T. This binary judgement was based on both the MacCAT-T and a clinical interview with the patient and was withheld from the other interviewer until both assessments had been made. A member of the clinical team, usually from the nursing staff, was then asked whether they judged the patient to have capacity to make a treatment decision. The interviewer also scored understanding, appreciation, reasoning and expression of choice according to MacCAT-T guidelines for each patient she had interviewed.

The anonymised, typed transcripts of all the MacCAT-T interviews conducted by RI 1 were distributed to a panel of three consultant psychiatrists (A.S.D., M.H., G.S.) and one consultant psychologist (P.H.). Each panel member independently rated whether they judged each patient to have capacity to make a decision about their own treatment. The binary rating was based on the definition of ‘inability to make decisions’ proposed in the Draft Mental Incapacity Bill (England and Wales) (Department for Constitutional Affairs, 2003) (now the Mental Capacity Act 2005). This states that persons are unable to make a decision for themselves if:

- (a) they are unable to understand the information relevant to the decision;
- (b) they are unable to retain the information relevant to the decision;
- (c) they are unable to use the information relevant to the decision as part of the process of making the decision; or
- (d) they are unable to communicate the decision.’

The panel’s training consisted of a brief discussion about using this definition to make a capacity judgement. For each case the judgement about capacity was rated as ‘very easy’, ‘moderately easy’, ‘moderately difficult’ or ‘difficult’, where 1 was ‘very easy’ and 4 was ‘difficult’. When the participant was judged to lack capacity, the panel member indicated in which area they had performed poorly (a–d). One panel member (M.H.) also rated each typed transcript according to MacCAT-T criteria.

The anonymised typed transcripts from RI 2 were distributed to panel members

once all the transcripts from RI 1 had been rated and returned. This time clinical information was provided with these transcripts in the form of brief summaries (about 200 words) that outlined the reason for admission, details of previous contact with psychiatric services and risk of harm to self or others. Finally, after all the transcripts had been rated, the sources of disagreement for cases in which opinion had been divided were explored in a discussion between the panel members, a lawyer with a special interest in mental capacity (G.R.), and the interviewers.

Statistical analyses

Data analysis was performed using the Statistical Package for the Social Sciences Version 11 (SPSS, 2001) and STATA (release 8.0; Stata Corporation, 2003). Cohen's kappa correlation coefficient and weighted kappa values (using STATA) were calculated to examine the correlations between the different assessments of capacity.

RESULTS

Participant characteristics

One hundred and twenty-seven newly admitted patients were approached and 55 (43%) of these completed both interviews. Of the remaining 72, 8 (11%) agreed to take part but did not complete the second interview, 39 (54%) refused to take part, 21 (29%) were excluded and 4 (6%) were eligible but not included either because there was judged to be a high risk of violence to the interviewer (3) or the patient had absconded from the ward (1). The valid participation rate was 54%. Of the 21 patients who were excluded, 10 were deemed too disturbed to participate by medical or senior nursing staff or were unable to assent to research, 8 were on no prescribed medication and 3 spoke no English. The main reason for not completing the second interview was being discharged (5) but 1 patient refused, 1 patient absconded without leave and was subsequently discharged and 1 patient was arrested and then discharged.

The sample comprised 38 men and 17 women with a mean age of 36 years (s.d.=12.4). Of these, 31 (56.3%) had the following psychotic illnesses (ICD-10 F20-F29; World Health Organization, 1993): schizophrenia (19), schizoaffective disorder (5) and other psychotic disorder (7). Seven patients (12.7%) had a diagnosis

of bipolar affective disorder (ICD-10 F31), 16 patients (29.1%) had a diagnosis of depression (ICD-10 F32-F33) and 1 patient (1.8%) had borderline personality disorder (ICD-10 F60.3). Nineteen patients (34.5%) had been admitted involuntarily under the Mental Health Act 1983 whereas the remaining 36 had agreed to voluntary admission. There were no significant demographic differences between the group of patients that participated and the 'non-participants' (comprising excluded and ineligible patients, those for whom only one interview was completed, and those who refused to take part) except that the latter tended to be older and there was a trend for non-participants to be female. The groups did not differ in terms of diagnosis, admission status (including type of section under the Mental Health Act 1983) or number of previous admissions. A comparison of the two groups is shown in Table 1.

Interrater agreements

Interrater reliability between two interviewers making separate capacity assessments, at different times, on the same patient

There was near-perfect agreement (Landis & Koch, 1977) between the two interviewers' binary judgements of mental capacity using two separate interviews, each based on both MacCAT-T and a clinical interview, with a kappa value of 0.82. The two interviewers agreed on binary capacity judgements in 91.0% of cases and rated 43.6% (24) and 45.5% (25) of patients as lacking capacity, respectively.

The weighted kappa values for the MacCAT-T sub-scale scores from two separate interviews were as follows: understanding, 0.65; understanding alternative treatment option, 0.56; reasoning, 0.54; appreciation, 0.71; expressing a choice, 0.33. According to Landis & Koch's (1977) interpretation of kappa, this translates to a substantial level of agreement for understanding and appreciation, a moderate level for understanding the alternative treatment and reasoning, and a fair level of agreement for expressing a choice.

Interrater reliabilities between interviewers against expert clinicians, for the same interview

There was a moderate level of agreement (Landis & Koch, 1977) for binary capacity judgements between a panel of experts and

RI 1 using typed transcripts from the same MacCAT-T interviews, with a mean kappa value of 0.60 (Table 2). However, in line with our hypothesis, there was near-perfect agreement (Landis & Koch, 1977) for binary capacity judgements when brief summaries (outlining the reason for admission, past psychiatric history and risk issues) were supplied in addition to the typed MacCAT-T transcripts, with a mean kappa value of 0.84 (Table 3).

Interrater reliabilities for the MacCAT-T sub-scales between the interviewers and an expert clinician, based on the same interview

The level of agreement (weighted kappa values) for the individual MacCAT-T sub-scale scores from the same interview scored by the interviewer and a senior clinician are shown in Table 4 for RI 1 and RI 2. For RI 2 additional clinical information was provided. Under these conditions all kappas were above 0.8.

Interrater agreement of capacity judgements about a patient between the interviewers and the clinical team

There was a moderate level of agreement (Landis & Koch, 1977) for binary capacity judgements between the interviewers and members of the clinical teams responsible for the patients' care (mean kappa=0.51).

Sources of disagreement between judgements

As hypothesised, the disagreement about capacity judgements was less when the panel members were provided with additional clinical information. This is reflected in the mean kappa values for binary capacity judgements (0.82 compared with 0.60). For the capacity ratings based on MacCAT-T transcripts and clinical vignettes, there was a clear consensus (at least four of the five raters agreed with each other) in 53 out of 55 cases. The panel members' mean difficulty rating was 2.65 (s.d.=0.21) for cases where the judgement was split compared with 1.92 (s.d.=0.74) when the consensus was clear. This difference was not statistically significant ($t=1.39$, d.f.=53, $P=0.17$). For ratings based solely on MacCAT-T transcripts there was a clear consensus in 48 cases.

When ratings had been completed, all raters met to discuss cases where there had been disagreement. We identified variations in the panel members'

Table 1 Comparison of participants and non-participants

Variable	Participants	Others (excluded, ineligible, one interview, refused)	χ^2	d.f.	P
Total group, n	55	72			
Male gender, n (%)	38 (69.1)	38 (52.8)	3.45	1	0.06
Age, years: mean (s.d.)	36.2 (12.4)	40.6 (12.0)	2.07 ¹	124	0.04
Ethnicity, n (%)					
White European	30 (54.5)	36 (50.0)	4.14	4	0.39
Black British	6 (10.9)	3 (4.2)			
Black African	9 (16.4)	20 (27.8)			
African–Caribbean	4 (7.3)	4 (5.6)			
Other	6 (10.9)	7 (9.7)			
Unknown		2 (2.8)			
Education, n (%)					
No qualifications	25 (45.5)	14 (19.4)	1.06	2	0.59
GCSEs or equivalent ²	13 (23.6)	7 (9.7)			
A levels or higher ³	12 (21.8)	11 (15.3)			
Unknown	5 (9.1)	40 (55.6)			
Marital status, n (%)					
Single	45 (81.8)	50 (69.4)	0.91	1	0.34
Married/cohabiting	10 (18.2)	17 (23.6)			
Unknown		5 (6.9)			
Employment, n (%)					
Employed	13 (23.6)	15 (20.8)	2.03	2	0.36
Unemployed	37 (67.3)	49 (68.1)			
Student	5 (9.1)	2 (2.8)			
Unknown		6 (8.3)			
Diagnosis, n (%)					
Depression	16 (29.1)	12 (16.7)	9.38	8	0.31
Schizophrenia	19 (34.5)	26 (36.1)			
BPAD	7 (12.7)	6 (8.3)			
Schizoaffective disorder	5 (9.1)	7 (9.7)			
Psychotic disorder	7 (12.7)	15 (20.8)			
Other	1 (1.8)	6 (8.3)			
Previous admissions, n (%)					
0	15 (27.3)	13 (18.1)	0.72	3	0.87
1–2	12 (21.8)	15 (20.8)			
3–5	9 (16.4)	11 (15.3)			
> 5	19 (34.5)	24 (33.36)			
Unknown		9 (12.5)			
Detained under Mental Health Act 1983, n (%)	19 (34.5)	34 (47.2)	2.06	1	0.15

BPAD, bipolar affective disorder.

1. t value (independent samples t -test).

2. Examinations taken in UK at age 16 (earliest school-leaving age).

3. Examinations taken in UK at 18 for those who choose to stay in education.

interpretations of the participants’ reasoning and appreciation abilities to be the main source of disagreement in reaching binary capacity judgements. The less stringent view was that evidence of good reasoning at some point in the interview, with some sensible answers and some consistency with

the end decision, was sufficient evidence of preserved reasoning ability. The alternative view was that anything more than trivial internal inconsistencies in the patients’ arguments was evidence of poor reasoning and sufficient to deem the patient incompetent. Similarly, the more lenient

Table 2 Agreement between individual expert panel members (A–D) and research interviewer 1 (RI 1) for the same interview (transcripts only)

Interviewer/ panel member	Kappa values			
	Rater A	Rater B	Rater C	Rater D
RI 1	0.63	0.74	0.62	0.46
Rater A		0.74	0.70	0.53
Rater B			0.58	0.71
Rater C				0.32

Table 3 Agreement between individual expert panel members (A–D) and research interviewer 2 (RI 2) for the same interview (transcripts and clinical vignettes)

Interviewer/ panel member	Kappa values			
	Rater A	Rater B	Rater C	Rater D
RI 2	0.78	0.93	0.85	1.0
Rater A		0.71	0.78	0.78
Rater B			0.78	0.93
Rater C				0.85

interpretation of patients’ fluctuations in the appreciation of their disorder and need for treatment was that even temporary glimpses of insight suggested they were at some level able to appreciate the relevance of this information for themselves. The more stringent view was that any significant fluctuations meant that a patient’s capacity was impaired. Underlying these different views was an uncertainty as to the precise degree of inconsistency in reasoning and appreciation required to establish incapacity.

Other issues were also identified. First, there was probably a bias towards judging a patient as having capacity if they made the apparently ‘correct’ decision, agreeing to treatment. Second, panel members felt that for more difficult capacity judgements it would have been important to ask the patient additional questions outside the constraints of the MacCAT–T interview, and also to reassess the patient at another time. Finally, a difficulty arose in one case from uncertainty about whether odd use of language was attributable to the patient speaking English as a second language or to the patient’s psychopathology.

Table 4 Agreement for the MacCAT-T sub-scale scores from the same interview scored by the interviewer and a senior clinician

MacCAT-T sub-scale	Research interviewer 1	Research interviewer 2 (additional clinical information given to senior clinician)
Understanding	0.87	0.95
Understanding alternative treatment option	0.83	0.88
Reasoning	0.59	0.81
Appreciation	0.86	0.86
Expressing a choice	0.46	0.82

MacCAT-T, MacArthur Competence Assessment Tool for Treatment.

DISCUSSION

Our study aimed to measure the level of agreement between raters, under a number of circumstances, assessing a patient's capacity to make a treatment decision. We found the agreement to be high, especially when the MacCAT-T was used in association with additional clinical information. The MacCAT-T is probably the most widely used of the clinical and research tools that help inform the clinical judgement of capacity.

Binary capacity judgements

As far as we are aware, this is the only study of the reliability of binary capacity judgements, guided by the MacCAT-T and clinical judgement, from two separate interviews of the same patient. Previous work on the reliability of capacity assessments in mentally ill people has consisted of different individuals rating transcripts or videos of the same interview. However, in clinical practice we would expect much of the variation that occurs between raters to derive from the way in which the interview itself is conducted. There has also been more attention paid to the reliability of rating different components of capacity (sub-scale ratings) than to the overall binary (yes/no) judgement (Roth *et al*, 1977; Janofsky *et al*, 1992; Bean *et al*, 1994; Grisso *et al*, 1997). We would argue that the latter is more important clinically.

Our results suggest that, in combination with a clinical interview, the MacCAT-T can be used to produce extremely reliable binary judgements of capacity, as currently defined, under these circumstances. The weighted kappa values for the sub-scale scores also show that the MacCAT-T can be used reliably by two interviewers. The greater strength of agreement seen for binary capacity judgements

compared with sub-scale scores alone is understandable: the additional clinical interview used for overall capacity judgements allowed important clinical and contextual factors about the patient to be taken into account.

We also investigated the level of agreement for binary judgements of capacity using the same interview and found that a panel of senior clinicians was able to agree on this even after minimal training on the method of assessment. This is important for future research as it indicates that capacity can be reliably assessed on the basis of transcribed interviews. The level of agreement substantially improved when the panel members were provided with clinical information to aid the judgement. This is of course the context in which clinical assessments are made, and the authors of the MacCAT-T have not suggested that it should be used in isolation (Grisso *et al*, 1997). It seems most likely that the improved kappa values were a function of the increased information available to the panel but it is also possible that the experience gained from rating the first set of MacCAT-T transcripts may have contributed. Care was taken to prevent discussion about individuals' techniques until ratings of both sets of transcripts were completed. The weighted kappa values for the sub-scale scores rated by the interviewer and a senior clinician also suggest that the MacCAT-T can be used reliably.

Strengths of the study

The consecutive sample design included patients with a range of psychiatric diagnoses admitted both voluntarily and involuntarily and seen at an early stage in their admission. It was therefore reasonably representative of the heterogeneous mix of

patients seen in clinical practice, ill enough to warrant hospitalisation. In addition, the number of patients recruited and seen for two interviews was larger than in previous studies, conferring additional statistical power to our findings (Roth *et al*, 1977; Janofsky *et al*, 1992; Bean *et al*, 1994; Grisso *et al*, 1997; Wong *et al*, 2000; Bellhouse *et al*, 2003). By using Cohen's kappa coefficient, which takes account of chance agreements, we also employed a more rigorous measure of reliability than that used in the original study of Grisso *et al* (1997) describing the interrater reliability of the MacCAT-T for the same interview of psychiatric patients. In assessing agreement between two interviewers performing separate interviews we have attempted to reflect the likely reality of clinical practice. Our measure of agreement is effectively a hybrid of interrater and test-retest reliability, and as such we would suspect it to yield lower kappa values than more usual judgements of interrater agreement where the same interview is assessed.

Limitations of the study

Fifty-seven per cent of the admitted patients were not included in the study. However, this is unlikely to limit the validity of the results unless a significant proportion of those patients would have presented special difficulties in the assessment of their capacity. We cannot be sure about this, but it is unlikely to be the case since the clinical backgrounds of these patients did not differ significantly from those of patients who did participate.

In addition, we noticed that patients had difficulty understanding the risks and benefits of no treatment, which we used as the alternative treatment option. Similar problems have been noted in previous studies and Wong *et al* (2000) suggest it may be inadequate to rely on capacity assessments that involve more abstract and complex elements that are cognitively demanding and depend on sophisticated verbal expressive skills. For example, people in general find it more difficult to reason on the basis of lack of harm (or benefit) rather than positive benefits (or harm), even though they may be functionally equivalent (Kahneman & Tversky, 1984). In spite of this, a moderate level of agreement was seen between the two interviewers in this study for the understanding alternative treatment element of capacity.

We encountered some difficulties when using the MacCAT-T. First, as suggested elsewhere, it may be appropriate to use a 'staged approach', asking first for a spontaneous account of the patient's existing understanding of their condition and treatment before embarking on the MacCAT-T interview (Wong *et al*, 2000). This would identify patients with a good pre-existing understanding of their condition and treatment for whom much of the disclosure part of the interview could be shortened or omitted. Some patients who clearly had capacity found the interview somewhat demeaning as they were asked, for example, to recall information when it was already clear that they could do so without difficulty.

Other patients found that an overwhelming amount of concentration was required during the disclosure of information used to test understanding in the MacCAT-T, to the extent that it may have constituted a memory test for some rather than assessing understanding *per se*. Previous studies have shown that by reducing memory load with an information sheet, in addition to a verbal disclosure, capacity can be significantly improved in some individuals (Wong *et al*, 2000; Bellhouse *et al*, 2003). This would be another possible way of tailoring the MacCAT-T to individual needs.

Clinical judgement of capacity

Although clinical judgements of capacity are dichotomous, we think it is useful to view the underlying processes as a spectrum. In exploring the differences of opinion between capacity judgements in this study, we found the sliding scale approach, encompassing the idea of proportionality, to provide a sensible and useful rationale for tackling this problem. This approach takes the severity of the consequences of the task-specific decision (in this case refusing treatment) into account and makes a judgement of incapacity more likely as the seriousness of potential risks for the patient increases (Gunn *et al*, 1999; Wong *et al*, 1999; *Ms B v. An NHS Hospital Trust*, 2002; Buchanan, 2004). Even with this approach, for the two cases in our sample where opinion was divided about the patients' capacity we remained unable to reach unanimous decisions.

This study has shown that two clinicians can reliably agree about capacity to decide about treatment in the early stages of admission to a psychiatric hospital, using

CLINICAL IMPLICATIONS

- Two clinicians can reliably agree about decisional capacity for treatment in the early stages of psychiatric admissions using the MacArthur Competence Assessment Tool for Treatment (MacCAT-T) in conjunction with a clinical interview.
- The weighted kappa values from the sub-scale scores show that the MacCAT-T can be used reliably by two interviewers.
- The finding that a panel of senior clinicians was able to agree on binary capacity judgements for the same interview is important for future research.

LIMITATIONS

- Fifty-seven per cent of the admitted patients were not included in the study.
- We have been unable to comment on the validity of our capacity assessments because there is no gold standard for the assessment of mental capacity.
- The measure reported for assessing agreement between two interviewers is effectively a hybrid of interrater and test-retest reliability and may yield lower kappa values than would be expected from a measure of pure interrater reliability.

RUTH CAIRNS, MRCPsych, CLEMENTINE MADDOCK, MRCPsych, Institute of Psychiatry, London; ALEC BUCHANAN, PhD, MD, MRCPsych, Department of Psychiatry, Yale University, New Haven, Connecticut, USA; ANTHONY S. DAVID, FRCPsych, FRCP, MD, Institute of Psychiatry, London; PETER HAYWARD, PhD, Institute of Psychiatry/Maudsley Hospital, London; GENEVA RICHARDSON, FRCPsych, Queen Mary College, University of London; GEORGE SZMUKLER, FRCPsych, MATTHEW HOTOPF, PhD, MRCPsych, Institute of Psychiatry, London, UK

Correspondence: Professor Matthew Hotopf, Academic Department of Psychological Medicine, Institute of Psychiatry, Weston Education Centre, 10 Cutcombe Road, London SE5 9RJ, UK. Tel: +44 (0)2078480778; fax: +44(0)2078485408; e-mail: m.hotopf@iop.kcl.ac.uk

(First received 24 January 2005, final revision 11 April 2005, accepted 3 May 2005)

a combination of the MacCAT-T and a clinical interview. It has also shown that for research purposes a panel of senior clinicians can reliably assess capacity using transcribed interviews. Semi-structured interviews are intended to improve the reliability of capacity assessments and our results suggest that this is the case with the MacCAT-T interview. This reliability study has not allowed us to comment on the validity of our assessments of mental capacity. Mental capacity is a complex construct that requires consideration and assessment of a number of social and other contextual factors on an individual basis for each patient. This makes it impossible to test criterion validity of capacity assessments as there is no gold standard. The main use of the MacCAT-T might be to ensure that the full range of necessary abilities

is considered when making a capacity judgement. We now know that in combination with a clinical interview this allows a rigorous and reliable assessment of mental capacity.

REFERENCES

- Appelbaum, P. S., Grisso, T., Mulvey, E., *et al* (1995) The MacArthur Treatment Competence Study: I Mental illness and competence to consent to treatment. *Law and Human Behaviour*, **19**, 105–126.
- Bean, G., Nishisato, S., Rector, N., *et al* (1994) The psychometric properties of the Competency Interview Schedule. *Canadian Journal of Psychiatry*, **39**, 368–376.
- Bellhouse, J., Holland, A. J., Clare, I. C. H., *et al* (2003) Capacity-based mental health legislation and its impact on clinical practice: 2) treatment in hospital. *Journal of Mental Health Law*, July, 24–28.
- Buchanan, A. (2004) Mental capacity, legal competence and consent to treatment. *Journal of the Royal Society of Medicine*, **97**, 415–420.

- Department for Constitutional Affairs (2003)** *Draft Mental Incapacity Bill*. Cmnd. 5859-I. London: Stationery Office.
- Department of Health (2002)** *Draft Mental Health Bill*. Cmnd 5538-I. London: Stationery Office.
- Department of Health (2004)** *Draft Mental Health Bill*. Cmnd. 6305-I. London: Stationery Office.
- Doyal, L. (1997)** Informed consent in medical research: Journals should not publish research to which patients have not given fully informed consent – with three exceptions. *BMJ*, **314**, 1107–1111.
- Expert Committee (1999)** *Review of the Mental Health Act*. London: Stationery Office.
- Fulford, K. W. M. & Sayce, L. (1998)** Invited commentary on: Mental health legislation is now a harmful anachronism. *Psychiatric Bulletin*, **22**, 666–670.
- Grisso, T. & Appelbaum, P. S. (1995a)** The MacArthur Treatment Competence Study: II Measure of abilities related to competence to consent to treatment. *Law and Human Behaviour*, **19**, 127–148.
- 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 Behaviour*, **19**, 149–173.
- Grisso, T. & Appelbaum, P. S. (1998)** *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. New York: 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.
- Gunn, M. J., Wong, J. G., Clare, I. C. H., et al (1999)** Decision-making capacity. *Medical Law Review*, **7**, 269–306.
- Gunn, M. J., Wong, J. G., Clare, I. C. H., et al (2000)** Medical research and incompetent adults. *Journal of Mental Health Law*, February, 60–71.
- Janofsky, J. S., McCarthy, R. J., Folstein, M. F. (1992)** The Hopkins Competency Assessment Test: a brief method for evaluating patients' capacity to give informed consent. *Hospital and Community Psychiatry*, **43**, 132–136.
- Kahneman, D. & Tversky, A. (1984)** Choices, values and frames. *American Psychologist*, **39**, 341–350.
- Kitamura, F., Tomoda, A., Tsukada, K., et al (1998)** Method for assessment of competency to consent in the mentally ill. Rationale, development, and comparison with the medically ill. *International Journal of Law and Psychiatry*, **21**, 223–244.
- Landis, J. R. & Koch, G. G. (1977)** The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- Ms B v. An NHS Hospital Trust [2002]** EWHC 429 (Fam).
- Osborn, D. P. J. (1999)** Research and ethics: leaving exclusion behind. *Current Opinion in Psychiatry*, **12**, 601–604.
- Roth, L. H., Meisel, A. & Lidz, C. W. (1977)** Tests of competency to consent to treatment. *American Journal of Psychiatry*, **134**, 279–284.
- SPSS (2001)** *SPSS Base 11.0.1 User's Guide*. Chicago, IL: SPSS.
- Stata Corporation (2003)** *Stata Release 8.0*. College Station, TX: Stata.
- Wong, J. G., Clare, I. C. H., Gunn, M. J., et al (1999)** Capacity to make health care decisions: its importance in clinical practice. *Psychological Medicine*, **29**, 437–446.
- Wong, J. G., Clare, I. C. H., Holland, A. J., et al (2000)** The capacity of people with a "mental disability" to make a health care decision. *Psychological Medicine*, **30**, 295–306.
- World Health Organization (1993)** *The ICD–10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva: WHO.