U.S. Department of Veterans Affairs

Public Access Author manuscript

Am J Geriatr Psychiatry. Author manuscript; available in PMC 2016 May 14.

Published in final edited form as:

Am J Geriatr Psychiatry. 2007 February ; 15(2): 168–173. doi:10.1097/JGP.0b013e31802e705c.

Agreement Between Instruments for Rating Treatment Decisional Capacity

Ronald J. Gurrera, M.D., Michele J. Karel, Ph.D., Armin R. Azar, M.A., and Jennifer Moye, Ph.D.

Mental Health Service, VA Boston Healthcare System, Boston, MA, and Harvard Medical School Department of Psychiatry, Boston, MA (RJG, MJK, JM); and the Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY (ARA)

Abstract

Objective—The objective of this study was to measure agreement between three treatment decisional capacity assessment instruments in mild to moderate dementia.

Method—Subjects (N = 79) were recruited from the community. Rating agreement was evaluated with kappa statistics.

Results—Three-way agreement was fair for overall capacity ($\kappa = 0.451$), very good for understanding (0.618), very poor for choice (0.158), and no better than chance for reasoning and appreciation. Pairwise agreement showed a similar pattern.

Conclusions—With the exception of understanding, current treatment decisional capacity assessment instruments do not consistently agree with one another in assessing treatment decision abilities.

Keywords

Treatment decisions; decisional capacity; interrater agreement

The right of individuals to make treatment choices is a cornerstone of modern medical ethics, yet conventional clinical assessments of treatment decisional capacity yield inconsistent results.¹ Factors that may contribute to these variable outcomes include misconceptions about notions of capacity and its assessment, incorrect or biased application of capacity standards, and intrusion of raters' personal values and perspectives into the process.²

Three assessment instruments that have been developed to address this widely recognized problem are the Capacity to Consent to Treatment Interview (CCTI),³ the Hopemont Capacity Assessment Interview (HCAI),⁴ and the MacCarthur Competence Assessment Tool–Treatment (MacCAT-T).⁵ However, it is not known how well they agree with one another in rating treatment decisional capacity.

METHODS

Subjects

Thirty-eight men and 41 women with mild to moderate dementia, mean $\pm$ standard deviation (SD) age 74.9 ± 6.3 years, mean $\pm$ SD educational achievement 13.9 ± 3.1 grade levels, and intact primary attentional ability were recruited from the community (see 6 for details of recruitment, screening, and group assignment procedures). All subjects were included in one or more prior reports that examined correlations among these instruments,² rates of impairment at various levels of dementia,⁷ correlations with neuropsychologic scores,⁶ and correlation of neuropsychologic scores with nine-month decline on the MacCAT-T.⁸

Informed Consent Procedure

Study information was presented in simple direct language verbally and in writing. Participants were informed they could discontinue participation at any time. Written informed consent was obtained after a complete description of the study was provided. Subjects were compensated financially for their time. The institutional human studies subcommittee approved this study.

Assignment to the Dementia Group

Initial telephone screening identified 154 potential dementia group subjects. Consensus clinical diagnoses of *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition* dementia were based on Telephone Interview for Cognitive Status scores, Dementia Diagnostic Screening Questionnaire responses, and medical records. A dementia diagnosis could not be confirmed for 34 (22.1%) callers, 28 (18.2%) were unable or unwilling to participate further, and 9 (5.8%) failed to complete the entire assessment battery for various reasons. Based on Telephone Interview for Cognitive Status cutoff scores of 30 of 31, the remaining 79 subjects had mild ($N = 45$) or moderate ($N = 34$) dementia.

Screening Procedure for Comparison Group

Potential comparison group subjects ($N = 120$) completed the Health Screening Questionnaire to identify health problems that might cause cognitive impairment. Seventeen (14.2%) individuals were excluded for health reasons and 15 (12.5%) were unable or unwilling to participate further, leaving 88 comparison subjects.

Decision Capacity Assessment Instruments

The Clinical Competency test Interview,³ HCAI,⁴ and MacCAT-T⁵ use hypothetical vignettes (CCTI and HCAI) or *actual* clinical situations (MacCAT-T) followed by probe questions to score each decisional ability (understanding, reasoning, appreciation, and expression of choice). For research purposes, a standardized hypothetical vignette was developed for the MacCAT-T.²

Instruments were administered according to instructions. Items were scored 0, 1, or 2 using detailed scoring manuals created in consultation with each instrument's developers; subscale scores were computed according to guidelines established in consultation with instrument

developers as needed. Instruments were administered in one session with order counterbalanced.

Main Statistical Analyses

The impaired range for each ability was defined by 2.5 standard deviations below the corresponding comparison group mean. We chose statistically defined thresholds for this research study because they are objective, and capacity may be distributed continuously. This contrasts with how these instruments are intended for use in clinical applications: no specific guidance is provided for mapping scores onto dichotomous ratings, and clinicians must incorporate scores into a broader assessment that considers each patient separately. The clinical approach emphasizes the uniqueness of each patient and situation and is consistent with the important ethical aim of maximizing every patient's decision-making autonomy, but it is less compatible with research methodology for which standardized methods are essential.

"Overall" capacity was scored as impaired if at least one individual ability score on that instrument was in the impaired range. Thus, each subject had 12 dichotomous ability scores (3 instruments $\times$ 4 abilities) and three overall capacity scores (one for each instrument). Abilities and overall capacity were scored as 0 = impaired or 1 = intact. Kappa coefficients were used to measure three-way agreement beyond chance between instruments; each instrument was treated as a different rater. Kappas and corresponding χ^2 significance levels for three-way comparisons were computed manually according to Bartko and Carpenter.⁹ Pairwise kappas were computed by SPSS, version 12.0. Kappa greater than 0.75 represents excellent agreement beyond chance, values between 0.40 and 0.75 indicate fair to good agreement beyond chance, and values below 0.40 represent poor agreement beyond chance (see,¹⁰ p 218). Supplemental statistics (bias index and prevalence index¹¹; indices of proportional positive and negative agreement¹²) were computed manually for the purpose of informing the main analyses.

RESULTS

Dementia group scores were virtually identical to those previously published.⁶ Across instruments, understanding scores showed the widest, and expression of choice the narrowest, range. Overall capacity was impaired least frequently when assessed by HCAI (11 [13.9%]) and most frequently when assessed by CCTI (24 [30.4%]); 13 subjects (16.5%) lacked capacity by MacCAT-T assessment.

The three-way kappa for overall capacity was .451 ($\chi^2 = 150.32$, $df = 78$, $p < 0.001$), indicating only fair agreement beyond chance. Kappa for understanding was 0.618 ($\chi^2 = 176.70$, $df = 78$, $p < 0.001$), indicating very good agreement across all instruments. Kappa for expression of choice differed significantly from zero (0.158, $\chi^2 = 104.01$, $df = 78$, $p = 0.026$), but not for appreciation (-0.039) or reasoning (0.047).

In pairwise comparisons (Figure 1), MacCAT-T and HCAI demonstrated excellent agreement for understanding ($\kappa = 0.802$) and good agreement for overall capacity (0.607), whereas CCTI showed only fair agreement for these variables (kappa range = 0.400–0.580).

MacCAT-T and HCAI showed poor agreement for reasoning, and MacCAT-T and CCTI showed poor agreement for expression of choice; all other pairwise comparisons demonstrated only chance agreement. CCTI scores fell in the impaired range more frequently than those for MacCAT-T and HCAI on understanding (13 versus 8 and 9), appreciation (4 versus 3 and 2), and choice (9 versus 3 and 6). Seven MacCAT-T reasoning scores were in the impaired range compared with two each for HCAI and CCTI.

The bias index (BI) estimates the tendency of two observers to differ in how frequently they observe the occurrence of a condition in the same sample; a nonzero BI indicates kappa may be inflated by rater bias.¹¹ The BI departed minimally from zero for CCTI evaluations of overall capacity and was close to zero for HCAI/MacCAT-T evaluations (Figure 1). The prevalence index (PI) estimates the relative likelihoods of “present” and “absent” ratings as a function of the prevalence of the condition being rated; as |PI| increases, kappa decreases.¹¹ PI deviated moderately from zero in CCTI evaluations of overall capacity and more so for the HCAI/MacCAT-T pair. The index of average proportional negative agreement (both observers agree condition is absent; “P-neg” in Figure 1) and the index of proportional positive agreement (both observers agree condition is present; “P-pos” in Figure 1) mirrored corresponding kappas for overall capacity. Taken together, these supplementary analyses indicate that the relatively poorer agreement observed with CCTI evaluations may reflect a tendency of that instrument to produce different outcomes with respect to both “present” and “absent” findings of overall impairment.

For individual abilities, BI was consistently near zero, but PI was very elevated ($|PI| = 0.722$) for all pairwise comparisons (Figure 1). P-neg was consistently very high across all pairwise evaluations of individual abilities, but P-pos was virtually identical to kappa throughout. These results suggest that the low prevalence of impaired abilities in this sample may have pervasively and uniformly reduced kappa across comparisons but that virtually all of the variance was the result of disagreement between instruments regarding whether a particular ability was impaired.

To determine whether agreement was sensitive to the thresholds used to define impairment, supplementary analyses were conducted using more lenient (2 SD below the comparison group mean) and more stringent (3 SD below the comparison group mean) thresholds for scoring an ability as impaired. With the more lenient definition, three-way capacity agreement slipped from fair to poor ($\kappa = 0.350$, $\chi^2 = 134.34$, $df = 78$, $p < 0.001$) and understanding agreement fell from good to fair (0.514 , $\chi^2 = 160.28$, $df = 78$, $p < 0.001$); kappa for choice was essentially unchanged but reached statistical significance for reasoning (0.165 , $\chi^2 = 105.15$, $df = 78$, $p < 0.025$). For the more stringent definition, capacity agreement improved slightly ($\kappa = 0.489$, $\chi^2 = 156.22$, $df = 78$, $p < 0.001$) as did agreement for understanding (0.687 , $\chi^2 = 187.47$, $df = 78$, $p < 0.001$) and choice (0.196 , $\chi^2 = 109.92$, $df = 78$, $p = 0.010$). Kappas for appreciation and reasoning were unchanged.

CONCLUSIONS

The major findings of this study are that these instruments demonstrate only fair to good agreement beyond chance in evaluating overall treatment decisional capacity and, except for understanding, poor to chance agreement for individual decisional abilities.

In pairwise evaluation of overall capacity, HCAI and MacCAT-T agreed better with each other than with CCTI. Kappas fell within the same range reported¹ for pairwise comparisons among five experienced physicians who had received a standardized orientation to legal competency standards (kappa range = 0.32–0.63). However, clinicians who make these determinations in routine clinical practice are not always well-informed and experienced.

Kappa is generally an excellent measure of chance-corrected agreement for categorical variables, but prevalence effects can produce lower estimates of agreement when the underlying variable is continuous.¹³ If decisional capacity is continuously distributed, like it may be, the low prevalence of incapacity in this sample may have produced smaller coefficients of agreement than would be observed in other samples with higher prevalence of incapacity. However, our purpose was to examine how these instruments perform in a community-dwelling sample of individuals, because this is a population actually encountered in clinical practice. Thus, the measures of agreement reported here are valid when generalized to this population.¹⁴

All instruments rated understanding more consistently than other abilities. This finding mirrors results for physician raters viewing CCTI interviews¹ and for two research assistants and a capacity expert using the MacCAT-T.¹⁵ Understanding is more closely related to neuropsychologic test performance than other abilities⁶ so may be more readily operationalized and less sensitive to measurement approaches, personal biases, and cultural nuances. Our previous investigations in these subjects have demonstrated that understanding exhibits better convergent validity,² more frequent initial impairment,⁸ and more consistent impairment across instruments.⁷ In addition, neuropsychologic performance variables most closely associated with understanding and reasoning⁶ are also the best predictors of impaired capacity initially and at 9-month follow up.⁸ Notably, agreement for reasoning, which is most strongly related to neuropsychologic performance after understanding,⁶ was poor.

Poor agreement between instruments on reasoning, appreciation, and choice may have several possible explanations. First, the instruments differ in how decisional abilities are defined and measured.² For example, MacCAT-T evaluates reasoning based on an ability to compare options and anticipate consequences in a logically consistent manner, whereas CCTI bases scores on the number and accuracy of reasons stated and HCAI scores the ability to articulate how the individual balanced risks and benefits. To evaluate appreciation, MacCAT-T asks about “any reason to doubt” the information provided and the possibility of benefit, whereas CCTI focuses on preparing for the chosen treatment and anticipating its consequences 1 year later and HCAI spotlights the physician’s rationale for recommending treatment.

These instruments also emphasize somewhat different cognitive processes. The conjectural scenarios used by CCTI and HCAI may bias scores in the direction of greater impairment

because executive functions such as loss of task and loss of detachment may emerge relatively early in dementia, and these deficits can impede the evaluation of hypothetical situations.¹⁶ However, the hypothetical vignette used for the MacCAT-T in this study should have minimized this difference. Another consideration is that the psychometric properties of these instruments are incompletely known,² and psychometric limitations could reduce interrater agreement.

A more complete comprehension of the relationship between neuropsychologic function and decisional capacity may help to refine structured assessment approaches in ways that enable more consistent outcomes. The development of more reliable and objective assessment methods is an important goal of capacity research because it reflects our cultural commitment to the ethical principal that individuals should have the opportunity to participate in treatment decision-making to the extent allowed by their abilities.

Acknowledgments

This study was supported by NIMH R29 MH57104 (Jennifer Moye, PI).

The authors thank Thomas Grisso, Ph.D., for his valuable assistance in the early stages of this project.

References

1. Marson DC, Earnst KS, Jamil F, et al. Consistency of physicians' legal standard and personal judgments of competency in patients with Alzheimer's disease. *J Am Geriatr Soc.* 2000; 48:911–918. [PubMed: 10968294]
2. Moye J, Karel M, Azar A, et al. Hopes and cautions for instrument-based evaluation of consent capacity: results of a construct validity study of three instruments. *Ethics Law Aging Rev.* 2004; 10:39–61.
3. Marson DC, Ingram KK, Cody HA, et al. Assessing the competency of patients with Alzheimer's disease under different legal standards: a prototype instrument. *Arch Neurol.* 1995; 52:949–954. [PubMed: 7575221]
4. Edelstein, B. Hopemont Capacity Assessment Interview Manual and Scoring Guide. Morgantown, WV: West Virginia University; 1999.
5. Grisso, T.; Appelbaum, PS. MacArthur Competence Assessment Tool–Treatment (MacCAT-T) Manual. Worcester, MA: University of Massachusetts; 1996.
6. Gurrera RJ, Moye J, Karel MJ, et al. Cognitive performance predicts treatment decisional abilities in mild to moderate dementia. *Neurology.* 2006; 66:1367–1372. [PubMed: 16682669]
7. Moye J, Karel MJ, Azar AR, et al. Capacity to consent to treatment: empirical comparison of three instruments in older adults with and without dementia. *Gerontologist.* 2004; 44:166–175. [PubMed: 15075413]
8. Moye J, Karel MJ, Gurrera RJ, et al. Neuropsychological predictors of decision-making capacity over 9 months in mild-to-moderate dementia. *J Gen Intern Med.* 2006; 21:78–83. [PubMed: 16423129]
9. Bartko JJ, Carpenter WT. On the methods and theory of reliability. *J Nerv Ment Dis.* 1976; 163:307–317. [PubMed: 978187]
10. Fleiss, JL. Statistical Methods for Rates and Proportions. 2nd. New York: John Wiley & Sons, Inc; 1981.
11. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993; 46:423–429. [PubMed: 8501467]
12. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol.* 1990; 43:551–558. [PubMed: 2189948]

13. Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol.* 2000; 53:499–503. [PubMed: 10812322]
14. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol.* 1988; 41:949–958. [PubMed: 3057117]
15. Grisso T, Appelbaum P, Hill-Fatouhi C. The MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. *Psychiatr Serv.* 1997; 48:1415–1419. [PubMed: 9355168]
16. Marson DC, Annis SM, McInturff B, et al. Error behaviors associated with loss of competency in Alzheimer's disease. *Neurology.* 1999; 53:1983–1992. [PubMed: 10599769]

FIGURE 1. Rating Agreement for Instrument Pairs by Individual Ability and Overall Capacity

(Lower panel) The frequencies in each cell of a standard 2×2 contingency table corresponding to each pair of instruments (C-H = CCTI/HCAI; M-C = MacCAT-T/CCTI; M-H = MacCAT-T/HCAI), for each ability (U: understanding; A: appreciation; R: reasoning; C: expression of choice). For example, the bar labeled “C-H U” corresponds to a 2×2 table comparing dichotomous CCTI (columns) and HCAI (rows) ratings for understanding; “a” is the number of cases in which both instruments agreed the individual was impaired, “b” the number of cases impaired by CCTI but not by HCAI, “c” the number of cases impaired by HCAI but not by CCTI, and “d” the number of cases unimpaired by both instruments.

The lower right panel provides the same information for overall capacity (O: overall capacity). Overall capacity was rated as impaired when one or more individual abilities were impaired.

(Upper panels) The upper left panel graphically represents values of kappa; bias index (BI) = $(b-c)/N$; prevalence index (PI) = $(a-d)/N^{11}$; proportion of positive agreement (P-pos) = $2a/[N+(a-d)]$; and proportion of negative agreement (P-neg) = $2d/[N-(a-d)]^{12}$ for corresponding instrument pairs immediately below.

The upper right panel presents the same information for corresponding pairwise evaluations of overall capacity immediately below.

A single asterisk (*) indicates kappa was significantly different from zero at the 0.05 level; a double asterisk (**) denotes a kappa that differed significantly from zero at the 0.001 level. p values were based on z tests. CCTI: Capacity to Consent to Treatment Interview; HCAI: Hopemont Capacity Assessment Interview; MacCat-T: MacCarthur Competence Assessment Tool–Treatment.