Use of Serial Mini-Mental State Examinations to Diagnose and Monitor Delirium in Elderly Hospital Patients

Shaun T. O'Keeffe, MD FRCPI, Eamon C. Mulkerrin, MB FRCPI, Kayser Nayeem, MB MRCPI, Matthew Varughese, MB MRCPI, and Isweri Pillay, MB MRCPI*

OBJECTIVES: To determine the responsiveness of serial Mini-Mental State Examinations (MMSEs) for the diagnosis and monitoring of delirium in elderly hospital patients.

DESIGN: Prospective study.

SETTING: University teaching hospital.

PARTICIPANTS: One hundred sixty-five people admitted to an acute geriatric service.

MEASUREMENTS: Subjects were assessed using the MMSE and the Confusion Assessment Method on hospital Days 1 and 6. Changes in scores were compared between patients who remained free of delirium (n = 124) and those who by Day 6 had developed delirium (n = 14) or had resolution of delirium present on admission (n = 22).

RESULTS: A number of measures of responsiveness confirmed that serial MMSE scores were responsive to resolution and to development of delirium. A fall of 2 or more points on the MMSE was the best determinant for detecting development of delirium (93% sensitivity, 90% specificity, positive likelihood ratio (LR) = 8.9 (95% confidence interval (CI) = 5.2–15.1) and negative LR = 0.08 95% CI = 0.01–0.53)). A rise of 3 or more points was the best determinant for detecting resolution of delirium (77% sensitivity, 75% specificity, positive LR = 3.1 (95% CI = 2.1–4.5) and negative LR = 0.30 (95% CI = 0.14–0.66)).

CONCLUSION: The MMSE is responsive to short-term changes in cognitive function in elderly patients. Serial MMSE tests should be helpful in monitoring the development and resolution of delirium in this population. **J Am Geriatr Soc 53:867–870, 2005.**

Key words: delirium; responsiveness; Mini-Mental State Examination; older persons

The Folstein Mini-Mental State Examination (MMSE) is widely used as a brief objective measure of cognitive function in older people.[1] The MMSE is successful at discriminating between patients with and without cognitive impairment,[2] but a single MMSE score does not help in distinguishing between acute, potentially reversible cognitive decline due to delirium and chronic impairment due to dementia.[3] This is not surprising, because the MMSE chiefly examines memory and orientation, and abnormalities in these cognitive domains are common to both disorders.

Serial testing using the MMSE might be of greater value in distinguishing between delirium and dementia. Thus, an abrupt decline from a previously established baseline would suggest that delirium is present, whereas improvement in test score after treatment of acute illness could provide retrospective support for a diagnosis of delirium. However, psychometric instruments that are highly reliable in measuring between-subject differences may be poor at measuring changes within an individual over time (i.e., they may show a poor responsiveness to change).[4,5] Validity in cross-sectional studies is based on measurement of between-subject differences, which are often large in heterogenous samples, but validity in longitudinal research is based on measuring between-person differences, which may be small and hard to detect.

The aim of this study was to determine the responsiveness to acute change in cognitive function of the MMSE. The optimum determinants for change in MMSE scores for detecting development and resolution of delirium were also examined.

PATIENTS AND METHODS

Patients

Consecutive patients aged 65 and older admitted from the accident and emergency department to an acute geriatric medicine service were recruited for the study. Patients with severe aphasia or deafness, those who were unwilling to participate, and those expected to die or to be discharged before the sixth hospital day were excluded.

Assessment of Cognitive Status

Separate examiners who were blinded to each other's results administered the MMSE and determined cognitive status

From the Department of Geriatric Medicine, Galway Regional Hospitals, Galway, Ireland.
Address correspondence to Dr. S. O'Keeffe, Unit 4, Merlin Park Regional Hospital, Galway, Ireland. E-mail: s.okeeffe@whb.ie

DOI: 10.1111/j.1532-5415.2005.53266.x

0002-8614/05/$15.00

(delirium, dementia, both, or neither) on the same day. The order in which these assessments were performed was not predetermined and depended on availability of the research staff on any particular day.

The same examiner administered the MMSE on the first full day of hospitalization (Day 1) and on hospital Day 6. This examiner was one of two (KN and MV) registrars (fellows) in geriatric and general internal medicine. A training session was held before the study to standardize administration and scoring of the MMSE. Interclass correlation coefficient for assessment of 20 patients by both examiners (interrater reliability) was 0.98 (range 0.95–0.99). The version of the MMSE used in this study was adapted and validated for use in an Irish population.[6] To minimize any practice effect, different sets of words were used for testing registration and recall in the two administrations of the MMSE analyzed in this study: "ball, tree, and flag" and "apple, penny, and table." On the same hospital days, an experienced consultant geriatrician (SOK or EM) interviewed and examined all patients and determined the presence or absence of delirium, dementia, or both. Delirium was diagnosed using the Confusion Assessment Method diagnostic algorithm[7] and required the presence of acute onset and fluctuating course, inattention, and disorganized thinking or altered level of consciousness. This clinician did not have access to the MMSE results and was asked not to perform a formal MMSE. Instead, he was encouraged to base his assessment of cognitive function primarily on an interview with the patient focussing on the presenting illness and medical history, supplemented if necessary by simple cognitive tests. Information about previous cognitive status and the pattern of onset of cognitive impairment was sought from carers and the general practitioner and by inspection of old medical and nursing notes. A diagnosis of dementia was made if there was evidence of cognitive impairment sufficient to interfere with social functioning of at least 6 months' duration.

Statistical Analysis

Changes in MMSE score between Day 1 and Day 6 of hospitalization were examined. Patients without delirium on admission who were delirious on Day 6 were considered to have worsened, and patients who were delirious on admission but were not delirious on Day 6 were improved. Patients were stable if they were free of delirium between Day 1 and Day 6 or had delirium on both days.

There is no consensus as to how best to measure responsiveness. Four different but related approaches were used: mean observed change, effect size statistic, responsiveness coefficient, and area under receiver operating characteristic (ROC) curves.[8] Separate responsiveness statistics were calculated in patients who deteriorated and in patients who improved, because scores in the two directions are not necessarily the same.[9] Patients with delirium on Day 1 were excluded when calculating responsiveness to development of delirium, whereas those who first developed delirium on Day 6 were excluded when calculating responsiveness to improvement in delirium.

MEAN OBSERVED CHANGE

Responsiveness was assessed by checking whether change scores and paired-sample *t* tests showed a gradient across those who worsened, were unchanged, and improved.

EFFECT SIZE

The effect size relates the mean change in score (Day 1 to Day 6) in unstable subjects to the standard deviation of the measure at Day 1 in the same subjects.[10] An effect size of 0.2 has been defined as small, one of 0.5 as moderate, and one of 0.8 or greater as large.[11]

RESPONSIVENESS COEFFICIENT

The responsiveness coefficient proposed in an earlier study relates the mean change in score in unstable subjects to the standard deviation of the change in score in stable subjects.[4] Like the effect size, this coefficient represents a standardized measure of change, and scores of 0.2, 0.5, and 0.8 can be interpreted as indicating small, moderate, and large responsiveness, respectively.[12]

When assessing the effect size and the responsiveness coefficient, the numerator was modified, as suggested previously, by subtracting the mean change in stable patients from the mean change in unstable patients.[13] This would reduce the influence on results of a predicted practice effect from performing serial MMSE tests at a relatively short interval.

AREA UNDER ROC CURVES

These areas were calculated using the nonparametric approach.[14] A value of 1 for area corresponds to perfect prediction, whereas a value of 0.5 is equivalent to that expected by chance. ROC curves were also used to plot the sensitivity and specificity of change on the MMSE for detecting development or resolution of delirium.

Test-retest reliability was assessed in stable patients using the intraclass correlation coefficient calculated from a one-way analysis of variance.

RESULTS

Of 289 patients admitted during the study period, 61 were excluded because they were expected to die or to be discharged before the sixth hospital day, 20 because of communication problems, and eight because they did not wish to participate. Of the 200 patients initially included in the study, 35 (18%) did not have a second assessment on Day 6 because of death (n = 6), discharge (n = 14), or error (n = 15). The mean age ± standard deviation of the remaining 165 patients was 79 ± 8; 159 (96%) were Caucasian, 110 (81%) had been admitted from the community, 76 (56%) were women, and 36 (27%) had dementia. The most common admitting diagnoses were respiratory (n = 68 patients), cardiovascular, (n = 35), gastrointestinal (n = 20), and cerebrovascular (n = 19) disease. Patients had an average of 4.1 ± 1.9 comorbid conditions.

By Day 6, 22 of the 27 patients who had been delirious on admission no longer satisfied criteria for delirium (improved patients), and 14 of the 138 patients who had not been delirious on admission were delirious (worsened patients). Of the remaining 129 stable patients, five were

Table 1. Mini-Mental State Examination (MMSE) Scores and Responsiveness and Reproducibility Statistics

	Change in Cognition between Day 1 and Day 6		
MMSE Score	Worsened (n = 14)	Stable (n = 129)	Improved (n = 22)
Day 1, mean ± SD	21.3 ± 5.3	22.2 ± 7.4	17.8 ± 7.0
Day 6, mean ± SD	16.9 ± 4.9	23.1 ± 7.1	22.6 ± 6.8
Change, mean ± SD	−4.4 ± 2.6	0.9 ± 2.1	4.8 ± 2.9
t score	6.3	5.1	7.8
P-value	<.001	<.001	<.001

SD = standard deviation.

delirious on Days 1 and 6, and 124 were not delirious on either day.

MMSE scores on Days 1 and 6 are shown in Table 1. The differences between MMSE scores on Days 1 and 6 are significant for all patient groups. In particular, there was a significant increase in MMSE scores in those considered stable. Nevertheless, the intraclass correlation coefficient for stable patients was 0.95 (95% confidence interval (CI) = 0.93–0.97), indicating high reproducibility.

For measuring development of delirium, the responsiveness coefficient was 2.65 (2.2 without Deyo correction of numerator), effect size was 1.0, and the area under the ROC curve was 0.97 (95% CI = 0.95–1.0) (Figure 1). For measuring resolution of delirium, the responsiveness coefficient was 1.95 (2.4 without Deyo correction), effect size was 0.56, and the area under the ROC curve was 0.87 (95% CI = 0.79–0.95) (Figure 2).

A fall of 2 or more points on the MMSE was the best determinant for detecting development of delirium (93% sensitivity, 90% specificity, positive likelihood ratio (LR) = 8.9 (95% CI = 5.2–15.1), and negative LR = 0.08 (95% CI = 0.01–0.53)). A rise of 3 or more points was the best determinant for detecting resolution of delirium (77% sensitivity, 75% specificity, positive LR = 3.1 (95% CI = 2.1–4.5) and negative LR = 0.30 (95% CI = 0.14–0.66)).

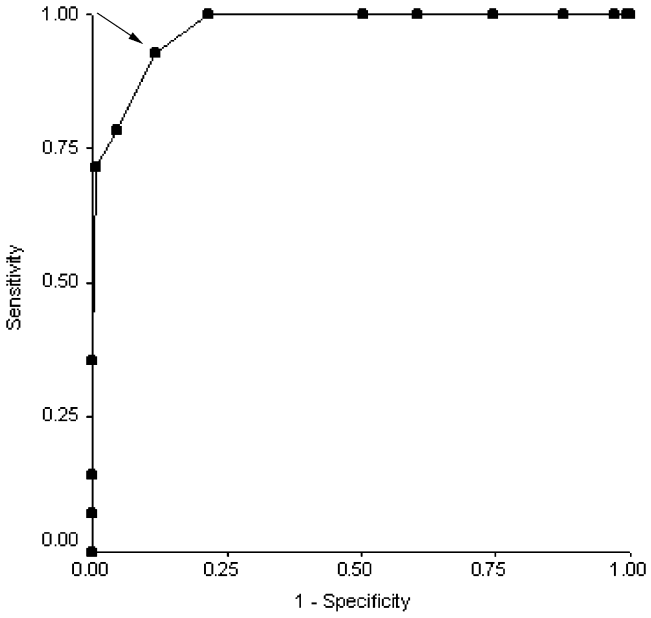

Figure 1. Receiver operating characteristic curve for use of serial Mini-Mental State Examination (MMSE) scores to monitor development of delirium. The arrow indicates the best determinant—a fall of 2 or more points.

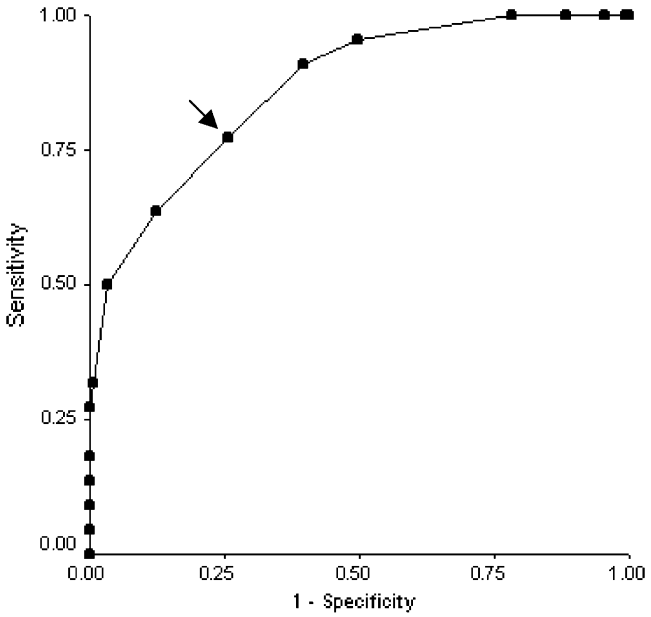

Figure 2. Receiver operating characteristic curve for use of serial Mini-Mental State Examination (MMSE) scores to monitor resolution of delirium. The arrow indicates the best determinant—a rise of 3 or more points.

DISCUSSION

This article shows that the MMSE is responsive to the short-term changes in cognitive function characteristic of delirium. This confirms the suggestion in the original description of the MMSE that it could be used for "serially documenting cognitive change."[1] The current study is consistent with work showing that serial MMSE scores are useful in monitoring long-term decline in cognitive function in patients with mild or moderate dementia.[2]

Delirium is a serious condition in elderly people, and early diagnosis and investigation are vital, yet recognition of delirium by medical and nursing staff on general wards is notoriously poor.[15] High sensitivity is the most important requirement of any diagnostic test in this population. The results of this study suggest that a decline in MMSE score of 2 or more points is a sensitive and specific indicator of delirium. Routine serial MMSE tests in vulnerable patients during the first few days of hospitalization may prove valuable in these settings.

These results also suggest that serial MMSE tests will be useful for monitoring the progress of delirious patients in clinical practice and, as an adjunct to specific delirium assessment instruments, in research studies. Although the MMSE was more responsive to deterioration than to improvement in cognition in this study, measures of responsiveness to improvement were satisfactory. The systematic increase in MMSE score on retesting in patients defined as

stable may partly explain the difference in responsiveness to deterioration and to improvement. There are a number of possible reasons for this increase. A practice effect is common on retesting, even after long intervals.[2] Improvement in general health and alertness or in mild cognitive symptoms during the first few days of hospitalization may also be important, but the increase in MMSE score in this study was modest, and test-retest stability was high, supporting the value of the MMSE for serial testing.

There are a number of limitations to this study. The choice of hospital Days 1 and 6 for calculation of responsiveness in this study was based on the fact that most episodes of delirium last less than a week and most patients who are going to develop delirium do so within the first few days of admission to hospital,[16] although it is not uncommon for delirium in elderly patients to last for several weeks, and individual symptoms of delirium may persist for a long time after resolution of the full-blown syndrome.[17,18] Groups defined according to the presence or absence of a change in delirium status were used to derive responsiveness statistics. Although such a change is clinically significant, the corresponding MMSE score changes are likely to exceed the minimal clinically important difference in scores used as the numerator in the original formulation of Guyatt's responsiveness coefficient. Relatively few of these patients had severe dementia or perfect cognitive function. Floor and ceiling effects could lower the responsiveness of both indices at the extremes of cognitive function.

The syndrome of delirium comprises more than cognitive impairment. Indeed, the characteristic distractibility of the delirious patient often allows the experienced examiner to make a diagnosis from the end of the bed without formal cognitive testing. Instruments that specifically address the range of symptoms that occur in delirium may prove even more valuable for monitoring change in the delirious patient.[19] Also, the optimal changes in MMSE scores for detecting development or resolution of delirium identified in this study require confirmation in an independent patient population. Nevertheless, the MMSE is already in widespread use in healthcare settings, and the current study suggests that serial testing can is useful in the diagnosis and monitoring of delirium in older patients.

ACKNOWLEDGMENTS

We are grateful for the assistance of Dr. Laobhaise Monaghan and Dr. Michael Geoghegan in conducting this study.

REFERENCES

1. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. Apractical method for grading the cognitive state of patients for the physician. J Psychiatr Res 1975;12:189–198.
2. Tombaugh TN, McIntyre NJ. The Mini-Mental State Examination: A comprehensive review. J Am Geriatr Soc 1992;40:922–935.
3. Anthony JC, LeResche L, Niaz U et al. Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. Psychol Med 1982;12:397–408.
4. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171–178.
5. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: What are the necessary measurement properties? J Clin Epidemiol 1992;45:1341–1345.
6. O'Neill D, Condren L, O'Kelly F et al. Cognitive impairment in the elderly. Ir Med J 1988;81:11–13.
7. Inouye SK, van Dyck CH, Aless CA et al. Clarifying confusion: The confusion assessment method. A new method for detection of delirium. Ann Intern Med 1990;113:941–948.
8. Blanchard C, Feeny D, Mahon JL et al. Is the health utilities index responsive in total hip arthroplasty patients? J Clin Epidemiol 2003;56:1046–1054.
9. Siu AL, Ouslander JG, Osterweil D et al. Change in self-reported functioning in older persons entering a residential care facility. J Clin Epidemiol 1993; 46:1093–1101.
10. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(Suppl):S178–S189.
11. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.
12. Crosby RD, Kolotkin RL, Williams GR. Defining meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395–407.
13. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Contr Clin Trial 1991;12:142S–158S.
14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.
15. Inouye SK. The dilemma of delirium: Clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. Am J Med 1994;97:278–288.
16. Rockwood K. The occurrence and duration of symptoms in elderly patients with delirium. J Gerontol 1993;48:M162–M166.
17. O'Keeffe S, Lavan J. The prognostic significance of delirium in older hospital patients. J Am Geriatr Soc 1997;45:174–178.
18. Levkoff SE, Evans DA, Liptzin B et al. Delirium. The occurrence and persistence of symptoms among elderly hospitalized patients. Arch Intern Med 1992;152:334–340.
19. Trzepacz PT. A review of delirium assessment instruments. Gen Hosp Psychiatry 1994;16:397–404.