Reliability of a Standardized Mini-Mental State Examination Compared With the Traditional Mini-Mental State Examination


Objective: The objective of this study was to compare the reliability of the Mini-Mental State Examination with that of a new Standardized Mini-Mental State Examination, which has expanded guidelines for administration and scoring. Method: The subjects were 32 stable elderly residents of a nursing home and 16 elderly residents of a chronic care hospital unit. Six raters administered the Folstein Mini-Mental State to 22 of these stable elderly subjects, and five raters administered the standardized version to 26 of these subjects. Each subject was tested on three different occasions 1 week apart. Each rater tested 4–6 subjects at the first and third weeks and 4–6 different subjects at the second week. The analytic technique used was one-way analysis of variance to estimate the interrater variance and the intrarater variance. Results: The intrarater variance on all occasions was reduced by 86% and the interrater variance was reduced by 76% when the Standardized Mini-Mental State was used; the reductions in variance were significant (p<0.003). The intraclass correlation for the Mini-Mental State was 0.69; for the standardized version it was 0.90. It took less time to administer the Standardized Mini-Mental State than the Mini-Mental State. Conclusions: The Standardized Mini-Mental State had better reliability than the Mini-Mental State in this study group. Although the improved reliability of the Standardized Mini-Mental State was achieved by reducing measurement noise, this advantage would likely occur in a broad spectrum of patients.


Screening instruments for cognitive impairment are frequently used in the elderly (1–3). The Folstein Mini-Mental State is the most widely used instrument to measure cognitive impairment in the elderly (1). This instrument has gained wide acceptance and has been used to assess cognitive function in clinical and epidemiologic settings (1, 4–8).

The Mini-Mental State measures orientation to time and place, immediate recall, short-term memory, calculation, language, and constructive ability. The maximum Mini-Mental State score is 30. The guidelines for its administration are short, and the scores for each item are recorded on the test form itself.

The Mini-Mental State is widely used because of its ease of administration and scoring. A substantial volume of literature has now accumulated about its validity (1, 4, 5). However, because the guidelines for its use are brief, the interpretation and scoring of answers are broad and subjective and may vary from one individual to the next. Differences between individuals scoring the test may affect the reliability of its scores.

Some of the Mini-Mental State items must be changed to accommodate different settings such as clinics, hospitals, and the subjects’ homes. When it is administered at home, the word “hospital” has to be changed to “place” and the term “floor” may be irrelevant. Scoring of the praxis item (interlocking 5-sided figures) and calculation is subjective, and an individual’s performance is affected by education (5). Because there is no time limit suggested for any reply, one could wait seconds or minutes for an answer. It is also not clear how one should score near misses, e.g., when a subject gives the last day of the month when it is the first day of the next month or gives a date of the month that is wrong by 1 day.

Received June 20, 1989; revision received May 30, 1990; accepted July 5, 1990. From the Department of Medicine, McMaster University and Hamilton Civic Hospitals Research Center. Address reprint requests to Dr. Molloy, Hamilton Civic Hospitals, Henderson Division, 711 Concession St., Hamilton, Ont., Canada L8V 1C3.

The authors thank Dr. M. Folstein, Dr. D. Streiner, and A. Mochide for their assistance in the development of the Standardized Mini-Mental State Examination.

The Mini-Mental State may be improved by developing more precise guidelines for its execution. We have developed a standardized version of this test (the Standardized Mini-Mental State) with explicit detailed guidelines for its administration and scoring (available on request from the first author). These stricter instructions were developed to improve the reliability of the instrument.

The goal of this study was to examine the effects of imposing stricter guidelines for administration and scoring on the reliability of the Mini-Mental State. The interrater reliability and the intrarater reliability of the traditional Mini-Mental State were compared with those of the standardized version.

The hypothesis tested was that the stricter directions in the Standardized Mini-Mental State would improve the reliability of the instrument.

METHOD

The study group comprised 48 elderly people, 32 of whom were residents of a nursing home and 16 of whom were residents of a chronic care hospital unit. Fourteen of the subjects were men and 34 were women. Their age range was 60–95 years. They suffered from a variety of conditions, including Alzheimer’s disease, stroke, arthritis, and osteoporosis. Twenty-two were given the Mini-Mental State and 26 were given the Standardized Mini-Mental State. Eight of the subjects given the Mini-Mental State and six of the subjects given the Standardized Mini-Mental State were men.

Subjects who were medically stable, without aphasia, and without hearing impairment were included in the study. Subjects with acute illnesses or those taking drugs affecting cognitive function were excluded. Subjects who could not understand English, those with behavior problems, and those who had previously received a Mini-Mental State examination were also excluded.

Eleven volunteer gerontology students administered the instruments. These raters were recruited from advertisements posted at a university. Six of the 11 raters were randomly assigned to administer the Mini-Mental State and five were randomly assigned to administer the Standardized Mini-Mental State. None of the raters had been exposed to the Mini-Mental State previously. All were given a 15-minute talk on the Mini-Mental State before they were randomly assigned to administer the Mini-Mental State or the Standardized Mini-Mental State.

Each subject was tested at weeks 1, 2, and 3. Each rater tested 4–6 subjects on two occasions and 4–6 different subjects once. All subjects were tested in the mornings (between 9:00 a.m. and 12:00 noon). If a subject developed any problems that may have affected his or her cognitive function, the assessment was postponed until the next visit (a week later). Raters recorded the time for each assessment and response of the subject to each question. The raters were asked not to discuss the assessments with their colleagues until the study was completed.

Each subject tested with the Mini-Mental State was asked to do serial 7s or, alternatively, to spell “world” backwards as stated on the test itself. Those subjects who were tested with the Standardized Mini-Mental State were asked to perform serial 7s during the test and to spell “world” backwards at the end. Total scores for the Standardized Mini-Mental State reported here do not include the score from the spelling task.

The assessment schedule was arranged so that the same rater assessed a patient at weeks 1 and 3 and a different rater assessed the patient at week 2. The variability between scores for any particular subject from week 1 to week 2 or from week 2 to week 3 represents the between-rater (interrater) variation, and the variability from week 1 to week 3 represents the within-rater (intrarater) variation. The analytic technique used was one-way analysis of variance (ANOVA) (9, pp. 167–170) to estimate the intrarater variance and the interrater variance. Direct comparisons between the variance components (both within and between observers) resulting from the Mini-Mental State and the Standardized Mini-Mental State were achieved by using the variance ratio (F) test (9, pp. 37–41). The intraclass correlation (9, pp. 184–185) was computed as a measure of overall reliability.

RESULTS

The mean±SD scores at weeks 1, 2, and 3 for subjects given the Mini-Mental State were 17.71±8.49, 19.8±8.09, and 20.11±7.25. The scores for those given the Standardized Mini-Mental State were 25±4.54, 23.50±6.67, and 24.50±6.20. The mean differences over time were small, and formal statistical comparison with two-way ANOVA indicated nonsignificant Time differences for both the Mini-Mental State (F=1.66, df=2, 32, p=0.21) and the Standardized Mini-Mental State (F=1.61, df=2, 50, p=0.21).

Since there were no consistent changes over time, the variability of scores among the three measures for any subject represents the inherent test-retest reproducibility of the instrument when administered by different raters (week 1 versus week 2 and week 2 versus week 3) or by the same rater on two different occasions (week 1 versus week 3).

The simplest measure of the level of reproducibility is the within-subject standard deviation—the variation in a series of scores from the same subject. If these scores were all derived from the same rater we would expect the variation (standard deviation) to be smaller than if a different rater was used to test the subject on each occasion. Similarly, if one form of the test, say the Standardized Mini-Mental State, was more reproducible than the other, it would have lower within-subject standard deviations, both in the situation where the
same rater was used and when different raters were used.

The pooled within-subject standard deviation was estimated by using one-way ANOVA separately for the situation when the same rater or different raters were involved and separately for the Mini-Mental State and the Standardized Mini-Mental State. For example, the Mini-Mental State was administered to 18 subjects at week 1 and week 3 by the same raters. One-way ANOVA with Subjects as the grouping factor yields a pooled within-subject standard deviation of 4.8 (see table 1). The equivalent data for the Standardized Mini-Mental State yield a within-subject standard deviation of only 1.8, indicating that the Standardized Mini-Mental State had better reproducibility. A comparison of these two standard deviations (i.e., the corresponding variances) by an F test indicates a highly significant difference in reproducibility between the two tests (see table 1).

The Standardized Mini-Mental State Examination had significantly lower within-subject standard deviation for all comparisons in table 1, both when the same rater was involved and when different raters assessed the subjects (see table 1).

The distribution of scores did not affect the reliability of the Standardized Mini-Mental State. The main reason the Standardized Mini-Mental State was more reliable was because it reduced the intrarater variance. We stratified subjects into those with low (21 or less) and high (more than 21) scores to examine the effect of the value of the scores on reliability. The within-subject standard deviations of the higher-scoring group were 3.0 for the Mini-Mental State and 1.7 for the Standardized Mini-Mental State. The within-subject standard deviations for the lower-scoring group were 4.4 for the Mini-Mental State and 2.0 for the Standardized Mini-Mental State. Within each group the within-subject deviation was almost halved for the standardized version of the test.

Within-subject standard deviations represent the inherent "noise" associated with the measurement of cognitive function. Whether the noise is acceptable in terms of an instrument’s performance depends on the strength of the "signal" we wish to detect with the test. The term "reliability" is used to describe an instrument’s ability to distinguish differences between subjects and is analogous to a signal-to-noise ratio. The most usually quoted measure of reliability is the intraclass correlation, which is the proportion of the total variance that comprises the between-subject differences as opposed to measurement error.

The intraclass correlation for the same rater assessing a subject increased from 0.69 for the Mini-Mental State to 0.92 with the Standardized Mini-Mental State. Similar improvements in intraclass correlation were seen with the Standardized Mini-Mental State even when different raters were involved (see table 1).

The mean time required for each interview with the Mini-Mental State was 13.39 minutes, which was significantly greater than that required for the Standardized Mini-Mental State (10.47 minutes) (t = 2.92, df = 124, p = 0.004).

The mean scores for spelling "world" backwards and serial 7s for 25 of the subjects who were given the Standardized Mini-Mental State were 3.88 ± 1.43 and 2.36 ± 1.61, respectively. This difference was statistically significant (t = 3.820, df = 24, p < 0.001).

DISCUSSION

These findings show that the Standardized Mini-Mental State is more reliable than the Mini-Mental State. Different raters were used for both tests, and the assessments were done independently. This study design was used to determine interrater reliability and to avoid any contaminating effects of nonverbal communication between raters during simultaneous administrations.

There was no improvement in the scores obtained at weeks 1, 2, and 3 for both the Mini-Mental State and the Standardized Mini-Mental State, indicating no learning effect in the administration of the test three times in 3 weeks.

We compared the two tasks of spelling "world" backwards and calculating serial 7s in 25 of the subjects who were given the Standardized Mini-Mental State. The guidelines for administration of the Mini-Mental State suggest that these two tasks can be used alternatively to test attention and calculation. How to score the spelling of "world" backwards is not stated clearly in the Mini-Mental State. Since we could not obtain scores for both tasks in the Mini-Mental State, in this study we compared the two tasks only in the subjects tested with the Standardized Mini-Mental State.

All but two subjects scored less on serial 7s than on the spelling task. The difference in mean scores be-
tween these two items was statistically significant. For elderly patients who cannot spell “world” backwards, the alternative use of serial 7s may not be applicable because the task was more difficult and the elderly scored lower on this task. These two tests are not comparable. Although this finding is not generalizable to all elderly populations, further study needs to be done to resolve this issue.

It took less time to administer the Standardized Mini-Mental State than the Mini-Mental State. This may be due to the clearer guidelines for the administration, scoring, and timing of the responses in the Standardized Mini-Mental State. In the Mini-Mental State there are no clear guidelines or time limits, and the rater could wait a long time for an answer. In the Standardized Mini-Mental State we give reasonable time limits for answers, commensurate with the tasks, based on our experience with this instrument in clinical trials and in the day-to-day assessment of elderly patients with cognitive impairment (7, 8).

Providing a finite time helps the rater to limit the time taken to administer the instrument. Without these guidelines raters may have difficulty knowing how long they should wait for a response before they proceed to the next task. This may explain why the raters took longer to administer the Mini-Mental State.

We believe that the expanded guidelines in the Standardized Mini-Mental State make the Mini-Mental State easier to administer and increase the reliability of the instrument. However, some care should be exercised in interpreting our intraclass correlations. Although one might expect the measurement noise associated with cognitive function to be reasonably stable with different subgroups of geriatric subjects, the heterogeneity of these subgroups of subjects may vary markedly. This would affect the intraclass correlation in different subgroups.

This does not mean that the intraclass correlation is inappropriate as a measure of reliability, but it does suggest that the reliability of an instrument varies depending on the patient group in which it is being used. The intraclass correlations reported in this paper are typical for a chronic care-nursing home group. Although the improved reliability of the Standardized Mini-Mental State in this study was achieved by reducing measurement noise, we believe that this advantage would likely occur in a broad spectrum of patient groups.

REFERENCES