Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination

A Hensel, M C Angermeyer, S G Riedel-Heller

J Neurol Neurosurg Psychiatry 2007;**78**:1298–1303. doi: 10.1136/jnnp.2006.109074

See end of article for authors' affiliations

Correspondence to:
Dr S G Riedel-Heller,
Department of Psychiatry,
University of Leipzig,
Johannisallee 20, 04317
Leipzig, Germany; Steffi.
Riedel-Heller@medizin.
uni-leipzig.de

Received 16 October 2006

Revised 7 March 2007

Accepted 12 March 2007

Published Online First

18 April 2007

Background: In clinical and research settings, the Mini-Mental State Examination (MMSE) is commonly used to measure cognitive change over time. The interpretation of changes in MMSE is often difficult. They do not necessarily result from true clinical change. Their interpretation requires comparison with normative data for change. However, MMSE change norms are lacking for long intervals.

Objective: To examine what is a reliable change in MMSE for long follow-up periods commonly used in clinic. To provide normative data for change.

Methods: A sample of 119 cognitively normal individuals, aged 75 years and over, who participated in the Leipzig Longitudinal Study of the Aged (LEILA 75+). All participants were tested six times at 1.5 year intervals with the MMSE over a mean period of 7.1 years. Reliable change indices were computed for a common confidence interval (90%).

Results: In repeated assessments with 1.5 year intervals, a change in MMSE of at least 2–4 points indicated a reliable change at the 90% confidence level.

Conclusion: Small changes in MMSE can be interpreted only with great uncertainty. They have a reasonable probability of being caused by measurement error, regression to the mean or practice.

The Mini-Mental State Examination (MMSE)¹ is one of the most widely used screening instruments for dementia. In clinical as well as in research settings, it is often used to measure cognitive change over time in older adults. The detection of decline or improvement is crucial for diagnosis and therapy. Progressive cognitive decline is a hallmark of diseases leading to dementia. Improvement is an indicator of the response to therapy.

However, it has been recognised that not all changes in the MMSE score reflect true clinical improvement or decline (eg, see Tombaugh² and Schmand and colleagues³). Changes in MMSE test score may also be the result of measurement error, regression to the mean, practice effect as well as of normal aging. Thus the interpretation of individual changes in test scores requires judging whether the change is probably a result of a true clinical change (ie, is reliable) or can be explained by other factors with reasonable probability. For this distinction, information on test–retest characteristics of the MMSE is indispensable. In particular, knowledge is required on how much change can be expected to occur normally. Individual changes in MMSE test scores “need to be compared against some type of normative standard”.²

Norms for change in neuropsychological tests are ideally derived from cognitively normal individuals⁴ investigated in population based studies. By definition, changes in normal samples are not caused by clinical change and are thus appropriate reference values. However, to the best of our knowledge, only two population based studies have published change norms for the MMSE. Change scores by the Canadian Study of Health and Aging (CSHA)² covered four visits with different interval lengths (65 days to 5 years). Change scores by the Amstel Project³ covered two visits with a 1 year interval. MMSE change norms for longer follow-up intervals are lacking. Those norms are required because change scores are probably related to the length of the follow-up period. Preferably, the interpretation of individual changes should rely on norms for change with similar follow-up lengths that were actually used. Beyond it there is a wealth of longitudinal studies using the

MMSE (for review see Park and colleagues⁵ and Tombaugh and colleagues⁶), which generally lack detailed tables with change scores.

The change in test scores can be modelled using different statistical methods (for overview see for example Collie and colleagues,⁷ Frerichs and Tuokko⁸ and Maassen⁹). One method is the so-called Reliable Change Indices (RCI), which provide estimates of the probability that an individual's change in test scores is not due to chance (ie, that it is reliable). The RCI, as suggested by Chelune *et al*, accounts for measurement error and practice effect.¹⁰ The RCI, as suggested by Hsu, accounts for regression to the mean.¹¹ According to these concepts, a reliable change is a change that is unlikely to have occurred by measurement error or by practice effect (Chelune's concept) or by regression to the mean (Hsu's concept). Because they correct for different effects, the two methods produce slightly different results. Methodological studies comparing different change score methods (eg, Frerichs and Tuokko⁸ and Heaton and colleagues¹²) found these two RCI^{10 11} to be suitable to accurately classify normal change in older adults.

This study aims to determine the amount of test score change in MMSE that is necessary to be deemed statistically reliable. In a normative sample of 119 cognitively normal individuals from the population based Leipzig Longitudinal Study of the Aged (LEILA 75+) study, norms for reliable change in MMSE were computed using procedures suggested by Chelune *et al*¹⁰ and by Hsu.¹¹

METHODS

Sample

Data were derived from a sample of all cognitively normal older persons who had been followed-up in the LEILA 75+ (n = 119).

Abbreviations: ADL, activities of daily living; CSHA, Canadian Study of Health and Aging; LEILA 75+, Leipzig Longitudinal Study of the Aged; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; RCI, Reliable Change Indices; SED, standard error of a difference; SEM, standard error of measurement; SIDAM, Structured Interview for the Diagnosis of Dementia

All subjects gave written informed consent to participate in this study. The study was approved by the local ethics committee and was therefore performed in accordance with the ethics standards laid down in the 1964 Declaration of Helsinki. LEILA 75+ is an ongoing population based study on the epidemiology of dementia, which was initiated in January 1997. To date, one baseline assessment and five follow-up assessments have been completed. The baseline cohort comprised a total of 1692 community dwelling individuals aged 75 years and over and resident in the Leipzig-South district, which were identified by systematic random sampling from an age ordered list provided by the local registry office. The study design has been described in detail elsewhere.¹³

For this study, cognitively normal subjects were identified using the following inclusion criteria: (i) SIDAM (Structured Interview for the Diagnosis of Dementia)¹⁴ score above age and education specific cut-off scores (1 SD) for mild cognitive impairment¹⁵ at each visit; (ii) no diagnosis of dementia during the course of the study according to DSM-IV diagnostic criteria; and (iii) valid test data at each of the six assessments. Thus only individuals without any clinically meaningful cognitive decline were included in the study and the norm data presented here reflect optimal aging. Individuals with severe sensory impairment leading to inability to complete neuropsychological tests were excluded. A total of 119 subjects met all of the inclusion criteria. Of the overall sample of 1692 subjects in the LEILA 75+, the main reasons for excluding subjects from this study were death during the course of the study (n = 707) and mild cognitive impairment (MCI) diagnosis or dementia diagnosis during the course of the study (n = 599). MCI was the unique reason for exclusion in only 25 persons. All other MCI cases fulfilled at least a second exclusion criterion.

Instruments

At each assessment, participants were investigated in their home environment by trained psychologists and physicians. Each participant was investigated with the neuropsychological test part of the SIDAM. The SIDAM test part consists of 55 items, including all 30 items of the MMSE. In our MMSE version, the spatial orientation question required the participants to specify the address of their home. The words “apple”, “table” and “penny” were used for the three word recall. All participants were asked to perform the serial sevens task. The same MMSE version was used at every visit. Furthermore, with

each participant, a fully structured interview was conducted and data on sociodemographic variables as well as possible risk factors for dementia were collected. The capacity to perform activities of daily living (ADL) was assessed with the SIDAM ADL-Scale¹⁴ and the ADL/IADL-Scale according to Schneekloth and Pothoff.¹⁶

Statistics

All statistical computations were performed using SPSS for windows (v 12.0.1.). If not otherwise stated, the significance level was set at 0.05 for all analyses. Differences in MMSE test performance between the six assessments were investigated using the Friedman test followed by pairwise comparisons with the Wilcoxon test. For pairwise comparisons, the α level was adjusted for multiple testing by Bonferroni–Holm. Non-parametric statistics were chosen throughout because the distributions of the MMSE scores were significantly different from normal distribution at all six time points. To investigate a possible effect of regression to the mean, we divided the sample into three groups according to MMSE percentiles (table 3) at each assessment. Differences between the MMSE groups in the mean change in MMSE were tested using non-parametric Kruskal–Wallis H testing. The test–retest reliability of the MMSE was analysed by computing non-parametric correlation coefficients (Spearman Rho). RCI were computed according to the procedure outlined by Chelune *et al*¹⁰ where

$$RCI = ((X2 - X1) - (M2 - M1)) / SED$$
with X1 = observed pre-test score, X2 = observed post-test score, M1 = group mean pre-test score and M2 = group mean post-test score. The difference (X2–X1) is the individual change in MMSE score (individual level). The difference (M2–M1) is the mean change in MMSE score within the normative sample (group level). An average improvement within the normative sample (M2–M1 >0) is interpreted as practice effect on a group level; thus the individual change is corrected for the average practice effect using the term ((X2–X1) – (M2–M1)).

The SED (standard error of a difference) was computed in two steps. Firstly, the standard error of measurement (SEM) was computed by

$$SEM = SD[(1 - r)^{1/2}]$$
where SD = pre-test SD of the MMSE and r = reliability coefficient. Secondly, the SED was computed using the formula $SED = (2 * SEM^2)^{1/2}$

A reliable change was defined by setting α at 0.10 (two-tailed) (ie, RCI values exceeding +1.645 were defined as reliable improvement, RCI values below –1.645 were defined as reliable deterioration). To find the MMSE difference (X2–X1) which is equivalent to an RCI = 1.645, we solved the above formula for (X2–X1). The results were rounded to the subsequent integer, as individuals can change only in points (integers). Data are provided to enable the reader to compute RCI, as suggested by Hsu,¹¹ with the formula

$$RCI = ((X2 - M2) - (r * (X1 - M1))) / SEP$$
where X1, X2, M1 and M2 are as previously defined. We computed SEP (standard error of prediction) = $SD * (1 - r^2)^{1/2}$ where SD and r are as previously defined.

RESULTS

Table 1 summarises the baseline characteristics of the sample. In common with the total LEILA 75+ cohort, women outnumbered men. The 119 subjects were younger than the remainder of the LEILA 75+ population which was probably because only cognitively normal subjects were included who completed all six assessments. On average, study subjects completed more years of education than the remainder of the LEILA 75+ sample.

Table 1 Baseline sample characteristics (n = 119)

Participants	
Women (n (%))	90 (76)
Age (y) (mean (SD) [range])	79 (3.6) [75–94]
Education (n (%))	
Low	22 (19)
Middle	74 (62)
High	23 (19)
Comparison with baseline (interval (y)): (mean (SD))	
Time1–Time2	1.6 (0.2)
Time1–Time3	2.9 (0.1)
Time1–Time4	4.5 (0.1)
Time1–Time5	6.1 (0.2)
Time1–Time6	7.1 (0.3)
Adjacent assessments (interval (y)): (mean (SD))	
Time2–Time3	1.4 (0.2)
Time3–Time4	1.5 (0.2)
Time4–Time5	1.6 (0.2)
Time5–Time6	0.9 (0.3)

Table 2 Mini-Mental State Examination scores in six repeated assessments (n = 119)

	Mean (SD)	Median	range
Time1	28.3 (1.3)	28	24–30
Time2	28.4 (1.3)	29	25–30
Time3	28.8 (1.4)	29	23–30
Time4	28.4 (1.4)	29	23–30
Time5	28.3 (1.7)	29	23–30
Time6	28.4 (1.6)	29	22–30

Table 2 summarises the average MMSE scores from six assessments. They were relatively high at all visits. As only cognitively normal individuals were included, the range of MMSE scores was 22–30 points instead of 0–30 at every examination and the variances in MMSE scores were low (SD 1.3–1.7). In all visits at least 45% of the sample received maximal MMSE scores of either 29 or 30. This corresponds to the well known ceiling effect of the MMSE.

We investigated whether there was an indication for the presence of a practice effect and/or for an age related normal decline on a group level. The non-parametric Friedman test showed that the MMSE scores were significantly different in the six repeated assessments (Friedman's chi-quadrat = 19.4, df = 5, p = 0.002). Pairwise Wilcoxon analyses showed that Time3 performance was, on average, significantly higher than Time1 performance. This improvement can be interpreted as a practice effect (ie, average improvement on the group level attributed to prior exposure to test content and procedures). To avoid misinterpretation it should be emphasised that the average significant improvement (Time1 to Time3) was found on a *group level* using statistical significance testing. On an *individual level*, participants differed in the actual amount of improvement (practice); some participants showed a deterioration in MMSE score (n = 23) while others (n = 24) experienced an improvement larger than +1 point within the first assessments. Individuals who experienced a gain in MMSE points outnumbered individuals who experienced a loss in MMSE points (compare Time1 with Time3). After Time3, there was no significant difference between subsequent visits on a group level (ie, we found no significant age related normal decline).

We investigated for all adjacent assessments whether there was an indication for regression to the mean (table 3) on a group level. When adjacent assessments were compared, participants with low pre-test MMSE scores had higher post-test scores. Participants with high pre-test scores had lower post-test scores. Kruskal–Wallis H testing showed that participants with a different pre-test MMSE level (low vs high) had

significantly different mean change scores in MMSE which can be interpreted as regression to the mean.

Table 4 (second and third columns) presents the average raw difference scores of the six assessments. Only the comparisons of clinical interest are presented (ie, comparisons of adjacent assessments and comparisons with baseline). On average, there was only a small change in MMSE (≤ 0.5 point) in any comparison of assessments. However, that only reflected the fact that the proportion of participants who experienced a gain in MMSE roughly equalled the proportion of participants who experienced a loss in MMSE score. Changes in MMSE up to 2 points occurred commonly (the range of SD was 1.4–1.9 overall comparisons). Few participants actually experienced changes up to 7 points in MMSE.

Non-parametric test–retest reliability analysis (Spearman rho) showed that all reliability coefficients were relatively low (table 4; fourth column). This might be explained by the long time intervals between visits and to the effect of regression to the mean. Correlations between adjacent visits were higher than correlations between visits with longer intervals. We next computed RCI, as described by Chelune and colleagues¹⁰ (table 4; eighth column). For almost all comparisons, changes in MMSE of at least 3 points were needed to conclude with 90% confidence that there was a reliable change (ie, RCI ≥ 1.645). The RCI at the 90% confidence interval defines a change, which by definition occurs in only 10% of cognitively normal elderly (5% improvement and 5% decline) and is therefore considered to be reliable.

We checked how many participants actually experienced reliable changes in MMSE according to the RCI (Chelune) at the 90% level. By definition (see above), this should be 10% of the sample. Table 4 (right columns) shows the percentage of participants with reliable changes. The percentages had roughly the order of magnitude expected.

We computed RCI, as suggested by Hsu,¹¹ for a number of MMSE scores (24–30 because this was the range of MMSE values in the normative population) for the comparison of Time1 with Time2 (table 5). As the tabulation of RCI of Hsu for all comparisons and MMSE scores would be very space-consuming, we demonstrated using two examples how to compute the RCI, as suggested by Hsu¹¹ (see formula in statistics section).

(i) In case 1, an individual has a pre-test MMSE of 30, a post-test MMSE of 28 and was assessed with a time interval of 1.5 years. In this case, the RCI is $((28 - 28.4) - (0.31 * (30 - 28.3))) / 1.28 = -0.72$. This is inside the ± 1.645 (90% confidence interval). There is not enough evidence to conclude with sufficient certainty that this is a real decline.

(ii) In case 2, an individual has a pre-test MMSE of 28 and a post-test MMSE of 26 and the time interval is again 1.5 years.

Table 3 Raw difference scores according to pre-test MMSE score: regression to the mean (n = 119)

Compared assessments	Mean change in MMSE			Inference statistic		
	Low pre-test MMSE	Medium pre-test MMSE	High pre-test MMSE	Chi-quadrat	df	p Value
Time1–Time2	1 (–1 to 5)	0 (–3 to 2)	–1 (–5 to 0)	30.8	2	<0.001
Time2–Time3	1 (0 to 4)	1 (–6 to 2)	0 (–3 to 0)	29.9	2	<0.001
Time3–Time4	1 (–2 to 4)	0 (–6 to 1)	–1 (–5 to 0)	24.1	2	<0.001
Time4–Time5	2 (–4 to 5)	0 (–6 to 2)	–1 (–4 to 0)	21.6	2	<0.001
Time5–Time6	1 (–4 to 6)	0 (–3 to 2)	–1 (–2 to 0)	23.0	2	<0.001

MMSE, Mini-Mental State Examination.

Data are given as median (range).

Definition of pre-test MMSE levels was according to percentiles (see statistics): low = MMSE $\geq$ lowest value and $\leq$ 25 percentile; medium = MMSE > 25 percentile and < 75 percentile; high = MMSE $\geq$ 75 percentile.

Table 4 Raw difference scores and Reliable Change Indices according to Chelune for the Mini-Mental State Examination (n = 119)

	Raw change in MMSE		Test-retest reliability	SED	SEP	Correction for practice/decline*	Reliable change in MMSE†	Per cent with reliable change	
	Mean (SD)	Range						– (5th)	+ (95th)
Comparison with baseline									
Time1–Time2	0.2 (1.6)	–5 to 5	0.31	1.58	1.28	0	–3 ≤, ≥+3	3.4	6.7
Time1–Time3	0.5 (1.5)	–5 to 6	0.44	1.42	1.21	1	–2 ≤, ≥+4	5.9	2.5
Time1–Time4	0.1 (1.6)	–4 to 4	0.31	1.58	1.28	0	–3 ≤, ≥+3	8.4	5.9
Time1–Time5	0 (1.8)	–6 to 6	0.33	1.55	1.26	0	–3 ≤, ≥+3	5.9	7.6
Time1–Time6	0.1 (1.8)	–7 to 6	0.26	1.63	1.29	0	–3 ≤, ≥+3	6.7	5.9
Adjacent assessments									
Time2–Time3	0.3 (1.4)	–6 to 4	0.42	1.40	1.18	0	–3 ≤, ≥+3	4.2	5.9
Time3–Time4	–0.4 (1.6)	–6 to 4	0.40	1.51	1.26	0	–3 ≤, ≥+3	9.2	4.2
Time4–Time5	–0.1 (1.9)	–6 to 5	0.36	1.64	1.35	0	–3 ≤, ≥+3	12.6	7.6
Time5–Time6	0.1 (1.7)	–4 to 6	0.45	1.82	1.55	0	–3 ≤, ≥+3	4.2	6.7

MMSE, Mini-Mental State Examination; SED, standard error of a difference; SEP, standard error of prediction.

*The correction constant was computed by subtracting the post-test MMSE from the pre-test MMSE mean and then rounding to the nearest integer.

†The reliable change intervals were computed using the Reliable Change Index (RCI) according to Chelune and colleagues,¹⁰ which corrects for practice effects and/or aging, but not for regression to the mean. A reliable change in MMSE was defined by setting the RCI at 1.645 (the equivalent of the 90% confidence interval).

The RCI is $((26 - 28.4) - (0.31 * (28 - 28.3))) / 1.28 = -1.80$. This is outside the ± 1.645 (90% confidence interval) and is therefore considered to be a significant change.

In both examples the raw difference score was 2. But in the first example, the pre-test MMSE was an extreme value (30) and the decline was towards the mean. By way of contrast, in the second example, the pre-test MMSE was at the mean and the decline was off from the mean. Both situations are interpreted differentially. A change towards the mean resulted in a lower RCI than a change off from the mean which resulted in a higher RCI and indicated a reliable change.

DISCUSSION

This study has investigated what difference in MMSE score in repeated assessments is needed to conclude that an individual's change in cognitive function is reliable. It was shown that a certain amount of test-retest change is normal. Changes in the range of 2–4 points in MMSE were necessary to conclude with 90% confidence that an individual has experienced a "true" or reliable change. Smaller changes can be interpreted only with great uncertainty. It is possible that they *may* reflect true change in some cases but this interpretation can be made only with insufficient confidence (<90%).

Our study is best comparable with two other population based studies: within the CSHA study,² regression based RCI

were published which were very similar to our study. A reliable change was defined by 3 MMSE points for short intervals (few months), and by 3–4 MMSE points for long intervals (up to 5 years). In common with our study, only individuals who did not develop any cognitive impairment were included. Within the Amstel Project,³ it was suggested that "only a change score of more than 5 points is suspect for disease". This range is slightly higher than in our study. The main reason for the discrepancy might be that in the Amstel Project individuals were included who developed dementia within 1 year. Generally it can be assumed that change scores are higher if subjects are included who develop dementia at follow-up.

Non-population based MMSE change scores have been published.^{17, 18} One study included individuals with progression to dementia and found "that changes of 3–5 points might be associated with change in affective state and psychological conditions at the time of testing".¹⁷ The other study published cut-offs for reliable change that varied between 1 and 4 points in MMSE dependent on age and education.¹⁸ The major shortcoming of this study is that normative subjects were tested only once (the author used previously published population normative data¹⁹) and RCI were computed using the reliability coefficients of Folstein *et al*'s original publication.¹ Consequently, the theoretical interval length was only 24 h.

Table 5 Reliable Change Indices according to Hsu which account for regression to the mean: comparison of Time1 to Time2 (n = 119)

		Post-test MMSE score					
		30	29	28* normative mean	27	26	25
Pre-test MMSE score	30		0.06‡	–0.72‡	–1.51‡	–2.29‡	–3.07‡
	29	+1.08		–0.48‡	–1.26‡	–2.04‡	–2.83‡
	28†	+1.32	+0.54		–1.02‡	–1.80‡	–2.58‡
	normative mean						
	27	+1.56	+0.78	+0.00		–1.56‡	–2.34‡
	26	+1.81	+1.03	+0.24	–0.54		–2.10‡
	25	+2.05	+1.27	+0.49	–0.29	–1.08	
	24	+2.29	+1.51	+0.73	–0.05	–0.83	–1.61

*Mean post-test MMSE score of the normative population (exact group mean was 28.4).

†Mean pre-test MMSE score of the normative population (exact group mean was 28.3).

‡Bold type = reliable change; ‡values indicate decline in MMSE score; all other values indicate improvement in MMSE score.

The interpretation of changes in MMSE score needs answers to three important questions: might there be (i) an effect of regression to the mean; (ii) a practice effect; or (iii) normal age related decline which should be taken into account for interpretation? Following the logic of the RCI, these effects were first analysed on a group level using statistical hypothesis testing. Afterwards, conclusions were drawn for the individual level and the RCI were computed.

On a group level, we found a significant effect of regression to the mean. On average, participants with extreme pre-test MMSE scores had less extreme post-test MMSE scores. On an individual level, changes off from the mean have a higher probability of being a result of a “real” change than changes towards the mean. Earlier studies also indicate an effect of regression to the mean for repeated assessments of the MMSE.^{2 20}

On a group level, we found on average a significant but small (on average +0.5 point) practice effect within the first three assessments. The RCI interval for the comparison between visits 1 and 3 was centred at +1 (≤ -2 to $\geq +4$). On an individual level, a gain in MMSE score (up to 4 points) within the first assessments has a reasonable probability of being due to practice and/or measurement error. This does not exclude the possibility that such an improvement might be because of a “real” change but this interpretation is uncertain. Significant but small practice effects (<1 point) in the MMSE have also been found in earlier studies excluding impaired individuals.^{2 21}

On a group level, we found only a small non-significant decline in MMSE after assessment 3 (-0.0 to -0.4 points). The RCI intervals were centred at 0 for all comparisons after assessment 3 and were always -3 to $+3$. On an individual level, a loss in MMSE greater than 3 points is unlikely a result of normal decline or measurement error. This again does not exclude the possibility that a smaller decline (<3 points) is due to a real change. Our results are in line with previous studies. A systematic review of studies in the general elderly population⁹ found a mean annual decline in MMSE of -0.16 to -0.56 points per year, if studies excluded patients with dementia (in common with ours).

The estimates of reliability of the MMSE in this study were relatively low. In an earlier review,⁶ test–retest correlations “generally fell between 0.80 and 0.95” for test–retest intervals of 2 months or less. However, reliability coefficients probably depend on the design of the study: lower coefficients have been found for longer intervals^{3 17 20} and in samples excluding impaired individuals.² Test–retest reliability varied between 0.48 and 0.65 in the CSHA study² and was 0.55 in the Amstel study.³ For a 2 year interval, the reliability coefficient was 0.38.¹⁷ Because of the relatively low reliability coefficients that occurred for normal individuals, it has been suggested “that small changes in MMSE scores should be interpreted with caution”.⁶ It can be concluded that the MMSE is not very sensitive in detecting cognitive change (ie, the presence of a reliable change may be indicative of decline while the absence may not exclude cognitive decline). The low test–retest reliability stresses the importance of repeated testing.

In this study, we used two different RCI methods to examine changes in MMSE score: the RCI, as suggested by Chelune and colleagues¹⁰ and the RCI, as suggested by Hsu.¹¹ The choice of the appropriate change score method is non-trivial. There are different approaches for measuring cognitive change (for overview see for example Collie and colleagues,⁷ Frerichs and Tuokko⁸ and Maassen⁹) including standardised regression based methods,²² the standard deviation method and RCI. There is no agreement on the “right” way to measure test–retest score changes (not least because comparative studies are lacking about the effectiveness of different change scores for

specific clinical questions, see end of discussion). The decision to use the RCI method in this study was motivated by findings that RCI, as suggested by Chelune and colleagues¹⁰ and Hsu,¹¹ perform at least as well as the standardised regression based methods in classifying normal change.^{8 12 23} We used those two RCI, which account for the effects which we actually found (practice and regression to the mean). We presented two different RCI methods (instead of only one) because it seemed premature to abandon one RCI method as long as their use for specific clinical questions needs to be investigated.

A major strength of this study is that evidence was provided that participants were truly cognitively normal. All participants in this study remained free of symptoms for 7 years, which indicates that they were truly cognitively normal. Moreover, norms were provided for six repeated assessments in intervals commonly used in clinical practice (1.5 years). Most comparable studies carried out investigations on fewer occasions or at shorter intervals. However, change scores are not constant for one instrument and probably depend on the interval length and number of visits. Cognitive changes in normal older adults seem to evolve very slowly, thus studies over longer intervals are needed.

The limits of the study include the length of the test–retest interval which may affect retest performance in cognitive tests.² Thus presented norms may not be appropriate for the interpretation of test–retest intervals substantially shorter than 1.5 years. The presented norms for change are a reference only for individuals aged 75 years and over, as this was the age range of our normative sample. Moreover, MMSE scores are known to be influenced by age and education. We did not compute age and education specific change norms because the sample of $n = 119$ was too small for a reasonable computation of those norms. We checked for all comparisons whether age, gender or education was associated with the RCI (Chelune and Hsu) using non-parametric correlation analysis (Kendall–Tau b) and U test. In almost all comparisons, neither RCI was significantly associated with demographic variables. A few correlations with age and education were significant, but these were low. Similarly, earlier studies found that most of the variance in change in MMSE was due to pre-test MMSE scores and not age or education.² Finally, the selection of normal controls could be a matter of debate. Participants with MCI were excluded; this could be criticised. The diagnostic label MCI is still debated. Persons with MCI are at increased risk of developing dementia but may also experience a stable course or improve.²⁴ Hence possibly some truly “normal” people in the LEILA 75+ study were not included. However, only 25 subjects were excluded with MCI as the unique reason for exclusion. In order to check the robustness of our results, we computed the RCI (Chelune and Hsu) for the enlarged sample with 144 subjects (119 original study sample with the additional 25 persons with MCI) for comparison of Time1 with Time2. The RCI (Chelune) remained the same (-3 to $+3$). The RCI (Hsu) changed slightly for some combinations of MMSE scores (by 1 MMSE point) because the “normative means” (to which individuals might potentially “regress” to) for Time1 and Time2 were lower if subjects with MCI were included in the normative sample.

Further research is needed to determine whether change norms in MMSE derived from cognitively intact individuals are suitable to predict the manifestations of a dementia syndrome. For other instruments, there are discrepant findings as to whether RCI methods contribute to the prediction of a diagnosis of dementia in older adults.^{8 25} Moreover, the publication of change norms for further neuropsychological instruments would be desirable, as it is impossible to adequately interpret changes in test scores without information on normal change.

ACKNOWLEDGEMENTS

This work was supported by the Interdisziplinäres Zentrum für Klinische Forschung (IZKF) Leipzig (Interdisciplinary Centre for Clinical Research Leipzig) at the Faculty of Medicine, University of Leipzig (project C07).

Authors' affiliations

A Hensel, M C Angermeyer, S G Riedel-Heller, Department of Psychiatry, Research Department of Evaluation and Epidemiology, University of Leipzig, Germany

Competing interests: None.

REFERENCES

- 1 **Folstein MF**, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;**12**:189–98.
- 2 **Tombaugh TN**. Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. *Arch Clin Neuropsychol* 2005;**20**:485–503.
- 3 **Schmand B**, Lindeboom J, Launer L, et al. What is a significant score change on the Mini-Mental State examination? *Int J Geriatr Psychiatry* 1995;**10**:411–14.
- 4 **Ivnik RJ**, Smith GE, Lucas JA, et al. Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance. *Neuropsychology* 1999;**13**:121–7.
- 5 **Park HL**, O'Connell JE, Thomson RG. A systematic review of cognitive decline in the general elderly population. *Int J Geriatr Psychiatry* 2003;**18**:1121–34.
- 6 **Tombaugh TN**, McIntyre NJ. The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc* 1992;**40**:922–35.
- 7 **Collie A**, Maruff P, Makdissi M, et al. Statistical procedures for determining the extent of cognitive change following concussion. *Br J Sports Med* 2004;**38**:273–8.
- 8 **Frerichs RJ**, Tuokko HA. A comparison of methods for measuring cognitive change in older adults. *Arch Clin Neuropsychol* 2005;**20**:321–33.
- 9 **Maassen GH**. Reliable change assessment in sport concussion research: a comment on the proposal and reviews of Collie et al. *Br J Sports Med* 2005;**39**:483–7.
- 10 **Chelune GJ**, Naugle RI, Lüders H, et al. Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology* 1993;**7**:41–52.
- 11 **Hsu LM**. Caveats concerning comparisons of change rates obtained with five methods of identifying significant client changes: comment on Speer and Greenbaum (1995). *J Consult Clin Psychol* 1999;**67**:594–8.
- 12 **Heaton RK**, Temkin N, Dikmen S, et al. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol* 2001;**16**:75–91.
- 13 **Riedel-Heller SG**, Busse A, Aurich C, et al. Prevalence of dementia according to DSM-III-R and ICD-10. *Br J Psychiatry* 2001;**179**:250–4.
- 14 **Zaudig M**, Mittelhammer J, Hiller W, et al. SIDAM—A structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias of other aetiology according to ICD-10 and DSM-III-R. *Psychol Med* 1991;**21**:225–36.
- 15 **Busse A**, Aurich C, Zaudig M, et al. Age- and education-specific reference values for the cognitive test of the SIDAM (Structured interview for the diagnosis of dementia of the Alzheimer type, multi-infarct dementia and dementias of other etiology according to ICD-10 and DSM-IV). *Z Gerontol Geriatr* 2002;**35**:565–74.
- 16 **Schneekloth U**, Potthoff P. *Hilfe- und Pflegebedürftige in privaten Haushalten. Bericht zur Repräsentativerhebung im Forschungsprojekt "Möglichkeiten und Grenzen selbständiger Lebensführung", im Auftrag des Bundesministeriums für Familie und Senioren (Bd. 20.2 – Schriftenreihe des BMFSF)*. Stuttgart: Kohlhammer, 1993.
- 17 **Mitrushina M**, Satz P. Reliability and validity of the Mini-Mental State Exam in neurologically intact elderly. *J Clin Psychol* 1991;**47**:537–43.
- 18 **Iverson GL**. Interpretation of Mini-Mental State Examination scores in community-dwelling elderly and geriatric neuropsychiatry patients. *Int J Geriatr Psychiatry* 1998;**13**:661–6.
- 19 **Crum RM**, Anthony JC, Bassett SS, et al. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 1993;**269**:2386–91.
- 20 **Olin JT**, Zellinski EM. The 12-month reliability of the Mini-Mental-State Examination. *Psychol Assess* 1991;**3**:427–32.
- 21 **Jacqmin-Gadda H**, Fabrigoule C, Commenges D, et al. A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *Am J Epidemiol* 1997;**145**:498–506.
- 22 **Sawrie SM**, Marson DC, Boothe AL, et al. A method for assessing clinically relevant individual cognitive change in older adult populations. *J Gerontol B Psychol Sci Soc Sci* 1999;**54**:116–24.
- 23 **Temkin NR**, Heaton RK, Grant I, et al. Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc* 1999;**5**:357–69.
- 24 **Palmer K**, Fratiglioni L, Winblad B. What is mild cognitive impairment? Variations in definitions and evolution of nondemented persons with cognitive impairment. *Acta Neurol Scand Suppl* 2003;**179**:14–20.
- 25 **Ivnik RJ**, Smith GE, Petersen RC, et al. Diagnostic accuracy of four approaches to interpreting neuropsychological test data. *Neuropsychology* 2000;**14**:163–77.

Access a vast information database with Toll-Free linking

"Toll-free" linking gives you immediate access to the full text of many of the cited articles in a paper's reference list—FOR FREE. With the support of HighWire's vast journal catalogue, a huge reference library is now open to you. If HighWire hosts the journal, you can view the full text of the referenced article, completely free of charge by following the Free Full Text links in the references.