

Evaluation of a Summary Score of Cognitive Performance for Use in Trials in Perioperative and Critical Care

Christoph S. Burkhardt^a Dagmar Birkner-Binder^a Alexa Gagneux^b
Manfred Berres^d Stephan P. Strebel^a Andreas U. Monsch^b Luzius A. Steiner^c

^aDepartment of Anesthesia and Intensive Care Medicine and ^bMemory Clinic, Department of Geriatrics, University Hospital Basel, Basel, and ^cDepartment of Anesthesia, University Hospital Centre and University of Lausanne, Lausanne, Switzerland; ^dRheinAhrCampus Remagen, University of Applied Sciences Koblenz, Remagen, Germany

Key Words

Cognitive testing · Consortium to Establish a Registry for Alzheimer's Disease · Neuropsychological tests · Postoperative cognitive dysfunction · Practice effects

Abstract

Background/Aims: Cognitive dysfunction after medical treatment is increasingly being recognized. Studies on this topic require repeated cognitive testing within a short time. However, with repeated testing, practice effects must be expected. We quantified practice effects in a demographically corrected summary score of a neuropsychological test battery repeatedly administered to healthy elderly volunteers. **Methods:** The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Neuropsychological Assessment Battery (for which a demographically corrected summary score was developed), phonemic fluency tests, and trail-making tests were administered in healthy volunteers aged 65 years or older on days 0, 7, and 90. This battery allows calculation of a demographically adjusted continuous summary score. **Results:** Significant practice effects were observed in the CERAD total score and in the word list (learning and recall) subtest. Based on these volunteer data, we developed a threshold for diagnosis of postoperative cognitive dysfunction (POCD) with the CERAD total score. **Conclu-**

sion: Practice effects with repeated administration of neuropsychological tests must be accounted for in the interpretation of such tests. Ignoring practice effects may lead to an underestimation of POCD. The usefulness of the proposed demographically adjusted continuous score for cognitive function will have to be tested prospectively in patients.

Copyright © 2011 S. Karger AG, Basel

Introduction

There is growing interest in cognitive deficits occurring in the context of therapeutic interventions. In particular, elderly patients are at risk of developing such deficits, which can be found in a surprisingly large number of patients. Monk et al. [1] have documented such changes in 41% of patients aged 60 years or older after major non-cardiac surgery. After intensive care treatment, the incidence is less clear, as typically no baseline cognitive testing is available for patients admitted to the intensive care unit. However, cognitive deficits have been described in 30–80% of patients 6–12 months after intensive care

Christoph S. Burkhardt and Dagmar Birkner-Binder contributed equally to the manuscript.

unit discharge [2, 3]. The diagnosis of cognitive changes requires repeated neuropsychological assessments within a short period of time. Time intervals of 7 days and 3 months are often used for postoperative cognitive dysfunction (POCD) [1, 4]. Apart from the heterogeneity of the test batteries used to diagnose POCD, which makes comparisons between studies difficult, two important points need to be considered. First, repeated cognitive testing may lead to practice effects, at least in some of the tested domains. Ignoring practice effects may lead to an underestimation of the incidence of POCD, as potentially not only a deterioration is relevant, but also the failure to show a learning curve with repeated testing may be indicative of a cognitive deficit. Second, the diagnosis of POCD is made if a patient shows a certain predefined grade of deterioration in the cognitive tests used. So far, to the best of our knowledge, only binary criteria have been used to define POCD. However, as already suggested by Newman et al. [5], postoperative cognitive changes are not likely to be an all-or-nothing phenomenon, but rather represent a continuum of subtle changes in cognitive function. Accordingly, test batteries should report the results on a continuous scale rather than in a binary fashion. Furthermore, to avoid the need for comparisons with controls, it would be desirable to have an easy-to-use neuropsychological test battery that can adjust for demographic factors such as age, gender, and education.

We quantified practice effects with repeated administration of the extended version of the Consortium to Establish a Registry for Alzheimer's Disease – Neuropsychological Assessment Battery [6] (CERAD-NAB, validated German version) [7] for which a summary score (the CERAD total score) [8] was calculated using healthy volunteers. This score ranges from 0 to 100 and potentially allows cognitive dysfunction to be quantified as a continuous variable. Data are available that allow adjustment of test results for the demographic factors age, gender, and education. We also assessed the commonly used tests for phonemic fluency [9] and the trail-making tests (TMT) [10] for practice effects. These tests are not included in the CERAD battery. However, they are frequently used in the assessment of POCD to test attention and executive function.

Methods

Participants

This study was approved by the regional ethics committee (Ethikkommission beider Basel, registered at www.clinicaltrials.gov, NCT00512200), and all participants gave written informed

consent prior to testing. Participants were given feedback on their test scores if they requested it, and referral was made when possible problems were detected. Healthy volunteers aged 65 years or older were eligible, and recruited from the audience of a popular lecture series on various topics organized by the University of Basel, Switzerland, and offered specifically to elderly people irrespective of their educational background. Exclusion criteria were a history of cerebrovascular disease, a Mini-Mental State Examination (MMSE) Score <24, and long-term psychiatric medication use. Sixty participants were included, all of whom were native German speakers, none of whom was institutionalized, and none of whom had previously met clinical criteria for dementia or mild cognitive impairment. All participants were administered the German version of the CERAD-NAB, phonemic fluency test, and TMT on days 0, 7, and 90. Administration of tests was standardized as previously described [6]. Additionally, the participants were screened for subjective symptoms of depression using the Geriatric Depression Scale [11], and autonomy was assessed with the Instrumental Activities of Daily Living Scale [12] on days 0 and 90. Education, medical history, medication, alcohol and tobacco consumption, and blood pressure were recorded.

Cognitive Function Testing

The German version of the CERAD-NAB is composed of 6 subtests:

Verbal Fluency

In the verbal fluency test [13], impairments in verbal production, semantic memory, and language are tested. Subjects are asked to name as many animals as possible in 1 min (maximum score = 24).

Boston Naming Test

The Boston Naming Test is a naming task where the subject is asked to name 15 objects presented as line drawings; a maximum of 10 s is allowed for each drawing (maximum score = 15) [6, 14].

Mini-Mental State Examination

The MMSE is a widely used and well-known brief cognitive battery testing immediate and delayed memory, orientation, concentration, language, and praxis [15].

Word List Test

The word list test measures the ability to remember newly learned information (word list learning) and delayed memory (word list recall and word list recognition) [16]. Ten printed words are presented at a frequency of 1 every 2 s, and the subject is asked immediately after the tenth word to recall as many of the words as possible. There are 2 subsequent trials in which the same 10 words are presented in a new random order, and after each trial the subject is again asked to recall as many words as possible (word list learning, maximum score = 30, the sum of the number of words learned over the 3 trials). After a delay of 5–10 min (time taken for the constructional praxis test), the patient is asked to recall as many of the 10 words presented in the word list memory task as possible (word list recall, maximum score = 10). In the word list recognition test, the patient is presented with a list of 20 printed words and asked after each word to say whether the word was one of the words s/he was presented with earlier in the word list learning test (maximum score = 10: number of true positives

minus number of false positives) [6]. In accordance with Ehrensperger et al. [17], we used 2 additional variables: (1) 'intrusions', the number of word responses given in the word list learning trials and the word list recall that were not in the original list, and 'savings', the proportion of correctly recalled words during the word list recall trial compared with word list learning trial 3; (2) 'discriminability', calculated by dividing the number of true positives by the number of true negatives in the word list recognition trial.

Constructional Praxis Test

In the constructional praxis [18] test, the subject is instructed to copy 4 line drawings of increasing complexity: a circle, a diamond, intersecting rectangles, and a three-dimensional cube (maximum score = 11) [6]. After a delay of 5 min (time taken for the word list recall test), the patient is asked to again recall and draw as many of the line drawings as possible (constructional praxis recall). According to Ehrensperger et al. [17], 'constructional praxis savings' was calculated as the proportion of correctly recalled figures during the constructional praxis recall trial compared to constructional praxis.

Phonemic Fluency and TMT

Phonemic fluency test ('s' words) [9] and TMT (parts A and B) [10] were added to the CERAD-NAB to include some tests focusing upon attention and executive function. In the phonemic fluency test, the patient is asked to name as many words starting with an 's' as possible in 1 min (numbers and proper names are not allowed). In the TMT part A, the subject is instructed to draw lines connecting circles numbered 1–25 in ascending order. In part B, the circles include both numbers (1–13) and letters (A–L) and the subject is instructed to connect the circles in an ascending pattern, alternating between the numbers and letters (i.e. 1-A-2-B-3-C, etc.). Time for completion of the test is recorded.

All tests were administered by two authors (D.B.-B. and C.S.B.) who were trained and supervised by a certified psychologist working at our memory clinic. The tests were administered in a standardized fashion in a quiet room. The unadjusted CERAD total score was calculated as described by Chandler et al. [8], and demographic adjustment (for age, gender, and education) was performed according to the formula developed for the German version of the CERAD by Ehrensperger et al. [17]: Chandler raw total score - (-0.391 × age + 0.886 × education + 4.447 × gender) with gender coded as male = 0 and female = 1.

Statistical Analysis

Raw scores of the CERAD-NAB subtests were adjusted for age, gender, and years of education as previously described [7]. To avoid the difficulties resulting from comparing raw test scores of different tests, z-scores (i.e. a dimensionless quantity indicating standard deviations above or below the mean) were calculated based on normative values from 1,100 elderly Swiss individuals allowing calculation of demographically adjusted (age, gender, and education) standard scores. For statistical analysis of practice effects, repeated-measures ANOVA of 16 test scores was used. To address the problem of multiple testing, p values of the within-subjects factors of the repeated-measures analysis were adjusted according to the Bonferroni procedure. The level of significance derived from this procedure was 0.003 (adjusted $\alpha = 0.05/16$). The Bonferroni-Holm procedure testing the next lower p value (0.01)

Table 1. Characteristics of participants (n = 57)

Age, years	72 (65–89)
Males	30 (53)
Education, years	14 (9–20)
Diabetes mellitus	0
Arterial hypertension	25 (44)
Cardiac disease	7 (12)
Atrial fibrillation	2 (4)
Chronic pulmonary disease	0
Peripheral vascular disease	1 (2)
Treatment with statins	7 (12)
Current smoker	6 (11)
Nicotine, pack years	4 (0–70)
Alcohol consumption, glasses/week	6 (0–28)
MMSE (baseline)	30 (26–30)
CERAD total (baseline)	91 (62–100)
CERAD total (baseline adjusted)	105 (73–115)
IADLS (baseline)	8 (8–8)
GDS (baseline)	1 (0–10)

Data presented as n (%) or medians (range).

IADL = Instrumental Activities of Daily Living Scale; GDS = Geriatric Depression Scale.

delivered a level of significance of $0.05/(16 - 4) = 0.05/12 = 0.004$. As no p value was between the Bonferroni limit and the Bonferroni-Holm limit for the next level of significance, only the Bonferroni correction was used. For any subtest that was significant in the Bonferroni procedure, pairwise comparisons between baseline, day 7, and 3 months could be done without further adjustment by a closed multiple-testing procedure. Regression analysis was used to calculate threshold values for POCD at follow-up depending on baseline result. Statistical calculations were performed with SPSS 17.0 for Windows (SPSS, Chicago, Ill., USA).

Results

Sixty subjects were included and tested. Three were excluded from analysis, 2 because they did not complete all the follow-up tests and 1 because of sedatives taken prior to the second test. Characteristics of the remaining 57 participants are shown in table 1.

Tests were performed at baseline and were repeated after 7 (range 6–7) and 91 (range 81–123) days (medians).

Results of the unadjusted and demographically adjusted CERAD total score and the subtests are shown in table 2.

The p values for differences in the CERAD total score, CERAD subtests, phonemic fluency, and the TMTs are shown in table 3. Significant practice effects were observed when results of the unadjusted and the adjusted

Table 2. Test results of participants (n = 57)

	Baseline score	Score on day 7	Score at 3 months
CERAD total score (raw)	91 (62–100)	96 (76–100)	98 (67–100)
CERAD total score (adjusted)	105 (73–115)	109 (91–118)	109 (82–121)
Semantic fluency	24 (13–32)	25 (13–40)	25 (13–39)
Boston naming test	15 (11–15)	15 (13–15)	15 (14–15)
MMSE	30 (26–30)	30 (28–30)	30 (27–30)
Word list (learning) ^a	25 (12–30)	29 (16–30)	29 (16–30)
Word list (recall)	9 (2–10)	10 (4–10)	10 (2–10)
Word list (intrusions) ^b	0 (0–4)	0 (0–4)	0 (0–2)
Word list (savings) ^c	100 (40–125)	100 (67–114)	100 (40–125)
Word list (discriminability) ^d	100 (85–100)	100 (90–100)	100 (90–100)
Constructional praxis (copy)	11 (8–11)	11 (8–11)	11 (9–11)
Constructional praxis (recall)	10 (6–11)	11 (6–11)	11 (1–11)
Constructional praxis (savings) ^e	100 (55–125)	100 (60–125)	100 (10–111)
Phonemic fluency	14 (5–33)	17 (6–30)	17 (6–31)
TMT (part A)	42 (21–84)	40 (20–96)	40 (21–87)
TMT (part B)	88 (39–203)	86 (33–162)	81 (42–207)
TMT (B/A)	1.96 (1–4)	1.93 (1–4)	2.00 (1–4)

Data presented as medians (ranges).

^a Sum of words learned after the 3 word list learning trials.

^b Word responses given in the word list learning trials and word list recall that were not in the original list.

^c Proportion of correctly recalled words during the word list recall trial compared with word list learning trial 3.

^d Number of true positives divided by the number of true negatives in the word list recognition trial.

^e Proportion of correctly recalled figures during the constructional praxis recall trial compared with constructional praxis.

Table 3. Differences in adjusted CERAD total score and subtests expressed as p values (n = 57, $\alpha = 0.003125$)

	Repeated-measures ANOVA	Baseline to day 7	Day 7 to 3 months	Baseline to 3 months
CERAD total score	<0.001	<0.001	0.109	<0.001
Semantic fluency	0.024	0.011	0.379	0.070
Boston naming test	0.244	0.182	0.871	0.230
MMSE	0.547	0.457	0.210	0.728
Word list (learning) ^a	<0.001	<0.001	0.041	<0.001
Word list (recall)	<0.001	0.012	0.022	<0.001
Word list (intrusions) ^b	0.270	0.643	0.230	0.096
Word list (savings) ^c	0.126	0.516	0.147	0.043
Word list (discriminability) ^d	0.350	0.903	0.151	0.161
Constructional praxis (copy)	0.070	0.523	0.103	0.037
Constructional praxis (recall)	0.011	0.097	0.280	0.006
Constructional praxis (savings) ^e	0.120	0.157	0.938	0.101
Phonemic fluency	0.002	<0.001	0.338	0.018
TMT (part A)	0.402	0.574	0.510	0.214
TMT (part B)	0.257	0.107	0.484	0.369
TMT (B/A)	0.533	0.365	0.289	0.764

^a Sum of words learned after the 3 word list learning trials.

^b Word responses given in the word list learning trials and the word list recall that were not in the original list.

^c Proportion of correctly recalled words during the word list recall trial compared with word list learning trial 3.

^d Number of true positives divided by the number of true negatives in the word list recognition trial.

^e Proportion of correctly recalled figures during the constructional praxis recall trial compared with constructional praxis.

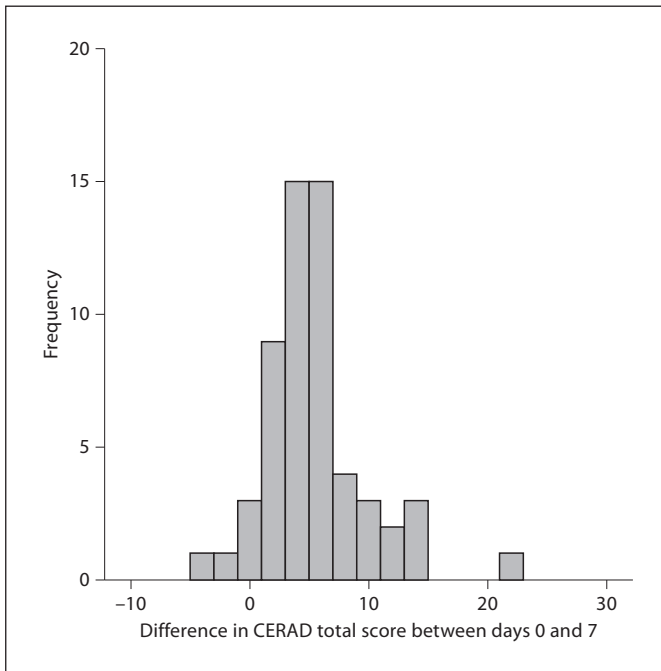


Fig. 1. Histogram for the differences in test performance between day 0 (baseline) and day 7 using the unadjusted CERAD total score (mean = 5.0, SD = 4.3, n = 57).

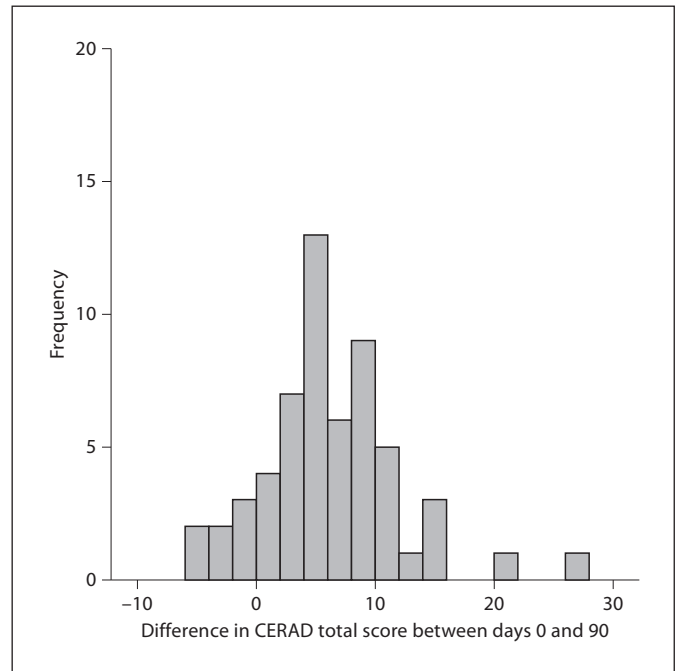


Fig. 2. Histogram for the differences in test performance between day 0 (baseline) and day 90 using the unadjusted CERAD total score (mean = 5.8, SD = 5.8, n = 57).

CERAD total score and word list (learning) tested on day 7 and after 3 months were compared to baseline test results. In the subtest word list (recall), a significant practice effect was observed comparing results at baseline and after 3 months. With phonemic fluency, a significant practice effect was observed comparing results at baseline and at day 7. In the remaining subtests, no significant practice effects were found when considering 0.003 as the level of significance. Figures 1–3 show the histograms for the differences in test performance between baseline, day 7, and 3 months (i.e. day 90) using the adjusted CERAD total score.

Based on these results from healthy volunteers, we developed a diagnostic cutoff for the adjusted CERAD total score after 7 days and after 3 months that can be used to define POCD. This cutoff is dependent on changes from baseline and, thus, takes practice effects into account. As the adjusted CERAD total score is already demographically adjusted, raw data values can be used and calculation of z-scores is not necessary. Regression analysis for test results after 7 days and after 3 months compared to baseline scores resulted in the lower 10% tolerance lines shown in figures 4 and 5, and POCD can be said to be present if test results of an individual lie below this line.

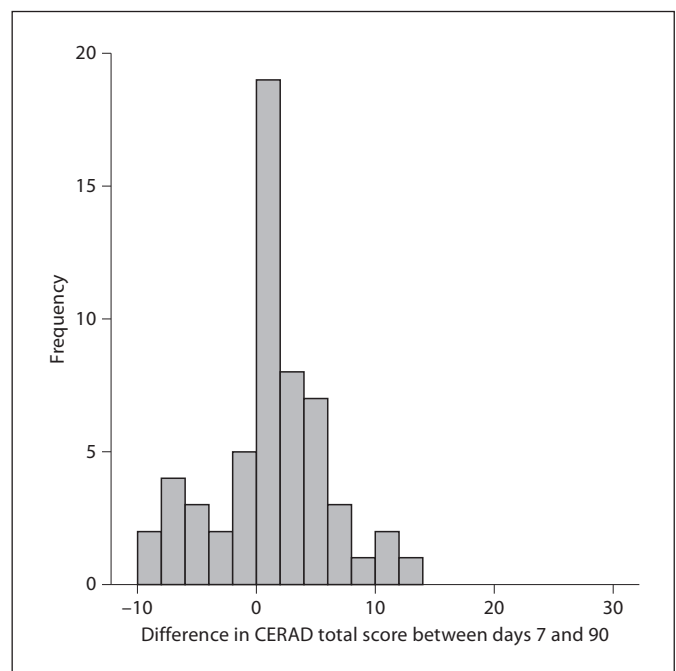


Fig. 3. Histogram for the differences in test performance between day 7 and day 90 using the unadjusted CERAD total score (mean = 0.8, SD = 4.6, n = 57).

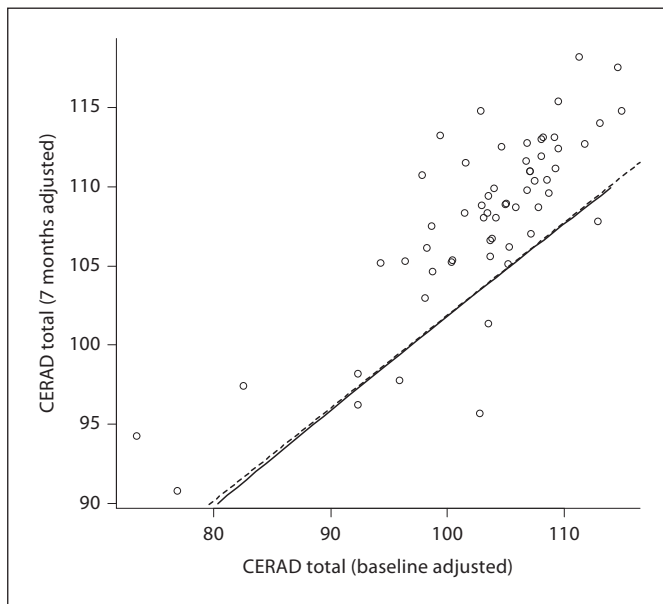


Fig. 4. Lower 10% tolerance line for adjusted CERAD total score: day 7 = $47.72 + 0.5852 \times \text{baseline} - 1.28 \times 3.39 = 43.38 + 0.5852 \times \text{BL}$. Solid line = Estimated hyperbola; dashed line = approximated straight line.

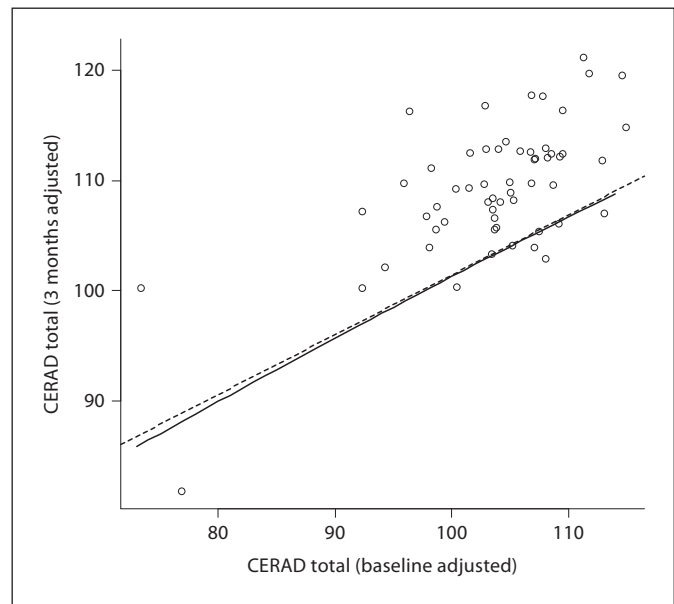


Fig. 5. Lower 10% tolerance line for adjusted CERAD total score: 3 months = $62.2 + 0.4568 \times \text{baseline} - 1.28 \times 4.27 = 56.73 + 0.4568 \times \text{baseline}$. Solid line = Estimated hyperbola; dashed line = approximated straight line.

Excluding 2 influential observations on the left margin yields a less steep line and 4 instead of 5 observations below it. We decided to keep these observations in the analysis. To give an example how to use the equation of the 10% tolerance lines: first, we assume that a patient scored 113 points in the adjusted CERAD total score at baseline (BL), 108 points on day 7, and 112 points at 3 months. Filling in the values for $\text{BL} = 113$, the equation day 7 = $47.72 + 0.5852 \times \text{BL} - 1.28 \times 3.39 = 43.38 + 0.5852 \times \text{BL}$ yields the result: day 7 = 110. Accordingly, the subject is expected to score at least 110 points on day 7 but only scored 108, and hence we assume that this patient has a cognitive deficit. Otherwise stated, inserting the BL score for the CERAD total score into the equation for day 7 or for 3 months will return the minimal score a patient has to achieve on day 7 or at 3 months in order to be considered not to have developed a cognitive deficit.

Discussion

Repeated administration of the CERAD-NAB, phonemic fluency, and TMT in healthy volunteers showed significant practice effects in the CERAD total score and in the word list subtests (learning and recall), but not in the

other subtests. With phonemic fluency ('s' words), practice effects were also found, but only on day 7 and not at 3 months. No improvement was possible in some subtests due to a ceiling effect – MMSE, Boston Naming Test, and figures (copy) – because at baseline 97% of participants reached MMSE scores ≥ 29 of a maximum of 30 points and ≥ 14 of a maximum of 15 points on the Boston Naming Test, and 89% ≥ 10 of a maximum of 11 points on constructional praxis (figure copying). We did not perform a sample size calculation or define a primary outcome parameter, as the aim of the study was exploratory; we wanted to quantify learning effects.

Internationally, one of the most widely used batteries of neuropsychological tests is the one by the CERAD. It has established itself as the standard in the evaluation of dementia syndromes in memory clinics, at least throughout German-speaking Europe. The CERAD-NAB assesses a wide variety of cognitive functions, is easy to administer after appropriate training (also by non-psychologists), is standardized, and there are normative values for normal controls from the CERAD registry [6, 8, 19] and from several observational and longitudinal studies [20–22], and from 1,100 elderly Swiss individuals allowing calculation of demographically adjusted (age, gender, and education) standard scores [17, 23]. Apart from En-

English and German, the CERAD-NAB has been translated into several languages and validated in different cultures [24–32]. This potentially allows defining POCD as a change in cognition based on demographically adjusted data. The validated German version of the CERAD-NAB is an exact translation of the English original and is usually completed in 30–45 min. The tests selected for the CERAD battery not only measure the primary cognitive manifestations of Alzheimer's disease across a range of severity of the disorder, but also discriminate well between normal subjects and those with mild or moderate dementia [6]. This battery also has a good specificity to detect minor deficits such as mild cognitive impairment in memory and practice in healthy subjects [33] and is, thus, potentially suitable for the diagnosis of POCD. The English version has been shown to have good interrater reliability (correlation coefficients from 0.92 to 1.0), longitudinal validity, and test-retest reliability, although normal controls tend to score at or near the maximum ('ceiling effect') in some subtests [6, 16, 19, 34]. A summary score for the CERAD battery ranging from 0 to 100 has been developed [8], and this continuous 'CERAD total score' has been shown to be highly accurate in differentiating independent samples of normal controls, patients with minimal cognitive impairment, and Alzheimer's disease [35], again suggesting that this battery may be suitable to diagnose POCD and other cognitive deficits following medical treatment or interventions. The cutoff differentiating normal controls from mild cognitive impairment was calculated to be 85 points in the adjusted CERAD total score [8]. The rationale for adding phonemic fluency and TMTs to the CERAD-NAB is the fact that tests for attention and executive function are not well represented in the CERAD-NAB.

Practice effects with repeated administration of neuropsychological tests have been reported for a variety of tests and after various time intervals in adults and children, as well as in brain-injured patients [23, 36, 37]. Practice effects may be due to increased familiarity with test-taking procedures, test-taking strategies, fewer feelings of fear or nervousness, a reduced sense of novelty at the time of retest, so-called test sophistication [38], and remembering specific test items. While the issue of remembering test items is likely to be more pronounced if the test is repeated after a short interval [39], test sophistication may be responsible for practice effects occurring with an interval between successive tests as long as 2–3 years.

With the German version of the CERAD-NAB, healthy volunteers showed statistically significant practice effects in verbal learning (word list) and visual mem-

ory – constructional praxis (savings) – after 2.4 ± 0.28 years, whereas there were no practice effects found in semantic fluency, Boston Naming Test, and constructional praxis (delayed recall of figures) [23]. As it is unlikely that specific test items were memorized for more than 2 years, the authors suppose that the observed improvement was due to test sophistication [38].

There are some limitations to our approach. The interpretation of our results may be limited by a recruitment bias. Healthy and motivated individuals are more likely to volunteer for a research project. Our study population may not necessarily be different from a group of patients as far as age and education are concerned, but concerning baseline cognitive function, mobility, independence, and possibly also scores on the geriatric depression scale. Furthermore, the good performance of the participants made it impossible to detect practice effects in subtests with so-called ceiling effects, because near maximum scores left no room for improvement in the repetition of these subtests. However, the CERAD total scores of some volunteers were quite low, suggesting that the investigated volunteers might represent the elderly population at large. Using the cutoff score of 85 of the adjusted CERAD total score [8], 3 of our volunteers could be diagnosed with a cognitive impairment. The CERAD-NAB and, thus, the CERAD total score do not include adequate tests for assessing attention span and executive function, which may also be important issues in POCD. Therefore, we added phonemic fluency and TMTs to the CERAD-NAB. However, the results of both phonemic fluency and TMTs are missing from the CERAD total score, and it would be desirable to have them included. On the other hand, administration of the CERAD-NAB, phonemic fluency test, and TMT takes 40–55 min. If time is critical, the CERAD-NAB could possibly be shortened to include word list learning and word list recall separated by an intervening task. Using a Bonferroni correction with $\alpha = 0.003$ may be a conservative approach, but we think it is appropriate. If the Bonferroni correction were ignored (i.e. using an $\alpha = 0.05$), additional practice effects in the subtests word list (learning) (comparing scores on day 7 and at 3 months), word list (recall) (comparing scores at baseline and on day 7 and scores on day 7 and after 3 months), and phonemic fluency (comparing scores at baseline and after 3 months) would be statistically significant. The 2 additional subtests, in which statistically significant practice effects would be observed ignoring the Bonferroni correction, are semantic fluency (comparing scores at baseline and day 7) and figure (recall) (comparing scores at baseline and after 3 months).

The lower 10% tolerance line for the adjusted CERAD total score (which we used as the cutoff between normal cognitive function and cognitive impairment) is arbitrarily set. Using this cutoff, 9% of our volunteers had a cognitive deficit at 7 days and 7% at 3 months. This is close to our specification of the lower limit, allowing for 10% false positives. This is somewhat higher than in previous publications. Similarly, in the study by Moller et al. [4], median age in controls was 67 years, but they reported cognitive deficits in 3%, both on day 7 and after 3 months. Two important limitations when comparing studies on POCD are the variability of the tests that are used to diagnose it and the criteria used to define it. So far, a binary definition of POCD has been used, and to the best of our knowledge, there are no studies using a continuous measurement of cognitive function to investigate the pathophysiology of POCD. This is problematic not only when the statistical power of studies is considered. It has also been suggested that it would be more appropriate to introduce continuous measures of cognition [5]. Whether or not this is an advantage will need to be demonstrated by applying this test to a group of patients. Moreover, as practice effects were demonstrated in some but not all domains, our results must not be extrapolated to other test batteries.

In conclusion, significant practice effects in the CERAD total score and in the CERAD-NAB subtest word list (learning and recall) as well as in phonemic fluency were observed with repeated administration of the CERAD-NAB, phonemic fluency, and TMT in healthy volunteers. This has to be considered if results of repeated CERAD-NAB subtests and repeated CERAD total scores are compared. The usefulness, sensitivity, and specificity of the CERAD-NAB, its summary score, and the suggested definition of POCD will need to be established in a group of patients undergoing major surgery.

Acknowledgements

A. Dwileski's assistance in preparation of the manuscript is gratefully acknowledged. This study was funded by the Swiss National Science Foundation (SNF Grant 32003B-121956).

Disclosure Statement

The authors have no conflicts of interest.

References

- 1 Monk TG, Weldon BC, Garvan CW, Dede DE, van der Aa MT, Heilman KM, Gravenstein JS: Predictors of cognitive dysfunction after major noncardiac surgery. *Anesthesiology* 2008;108:18–30.
- 2 Jackson JC, Gordon SM, Ely EW, Burger C, Hopkins RO: Research issues in the evaluation of cognitive impairment in intensive care unit survivors. *Intensive Care Med* 2004;30:2009–2016.
- 3 Jackson JC, Hart RP, Gordon SM, Shintani A, Truman B, May L, Ely EW: Six-month neuropsychological outcome of medical intensive care unit patients. *Crit Care Med* 2003;31:1226–1234.
- 4 Moller JT, Cluitmans P, Rasmussen LS, Houx P, Rasmussen H, Canet J, Rabbitt P, Jolles J, Larsen K, Hanning CD, Langeron O, Johnson T, Lauven PM, Kristensen PA, Biedler A, van Beem H, Fradakis O, Silverstein JH, Beneken JE, Gravenstein JS: Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. *International Study of Post-Operative Cognitive Dysfunction*. *Lancet* 1998;351:857–861.
- 5 Newman S, Stygall J, Hirani S, Shaefi S, Maze M: Postoperative cognitive dysfunction after noncardiac surgery: a systematic review. *Anesthesiology* 2007;106:572–590.
- 6 Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C: The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–1165.
- 7 Berres M, Monsch AU, Bernasconi F, Thalman B, Stahelin HB: Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. *Stud Health Technol Inform* 2000;77:195–199.
- 8 Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Weiner MF, Cullum CM: A total score for the CERAD neuropsychological battery. *Neurology* 2005;65:102–106.
- 9 Thurstone LL: *Primary Mental Abilities*. Chicago, University of Chicago Press, 1938.
- 10 Reitan RM: Validity of the trail making test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–276.
- 11 Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO: Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1982;17:37–49.
- 12 Lawton MP, Brody EM: Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179–186.
- 13 Isaacs B, Kennie AT: The Set test as an aid to the detection of dementia in old people. *Br J Psychiatry* 1973;123:467–470.
- 14 Howieson DB, Dame A, Camicioli R, Sexton G, Payami H, Kaye JA: Cognitive markers preceding Alzheimer's dementia in the healthy oldest old. *J Am Geriatr Soc* 1997;45:584–589.
- 15 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;12:189–198.
- 16 Welsh KA, Butters N, Hughes JP, Mohs RC, Heyman A: Detection and staging of dementia in Alzheimer's disease: use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Arch Neurol* 1992;49:448–452.

- 17 Ehrensperger MM, Berres M, Taylor KI, Monsch AU: Early detection of Alzheimer's disease with a total score of the German CERAD. *J Int Neuropsychol Soc* 2010;16: 910–920.
- 18 Rosen WG, Mohs RC, Davis KL: A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984;141:1356–1364.
- 19 Welsh KA, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, Heyman A: The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). V. A normative study of the neuropsychological battery. *Neurology* 1994;44:609–614.
- 20 Potter GG, Plassman BL, Burke JR, Kaboto MU, Langa KM, Llewellyn DJ, Rogers MA, Steffens DC: Cognitive performance and informant reports in the diagnosis of cognitive impairment and dementia in African Americans and whites. *Alzheimers Dement* 2009; 5:445–453.
- 21 Luck T, Riedel-Heller SG, Wiese B, Stein J, Weyerer S, Werle J, Kaduszkiewicz H, Wagner M, Mosch E, Zimmermann T, Maier W, Bickel H, van den Bussche H, Jessen F, Fuchs A, Pentzek M: CERAD-NP battery: age-, gender- and education-specific reference values for selected subtests: results of the German Study on Ageing, Cognition and Dementia in Primary Care Patients (Age-CoDe) (in German). *Z Gerontol Geriatr* 2009;42:372–384.
- 22 Lee SB, Kim KW, Youn JC, Park JH, Lee JJ, Kim MH, Choi EA, Jhoo JH, Choo IH, Lee DY, Woo JI: Prevalence of mild cognitive impairment and its subtypes are influenced by the application of diagnostic criteria: results from the Korean Longitudinal Study on Health and Aging (KLoSHA). *Dement Geriatr Cogn Disord* 2009;28:23–29.
- 23 Zehnder AE, Bläsi S, Berres M, Spiegel R, Monsch AU: Lack of practice effects on neuropsychological tests as early cognitive markers of Alzheimer disease? *Am J Alzheimers Dis Other Dement* 2007;22:416–426.
- 24 Lee JH, Lee KU, Lee DY, Kim KW, Jhoo JH, Kim JH, Lee KH, Kim SY, Han SH, Woo JI: Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. *J Gerontol B Psychol Sci Soc Sci* 2002;57:P47–P53.
- 25 Demers P, Robillard A, Lafleche G, Nash F, Heyman A, Fillenbaum G: Translation of clinical and neuropsychological instruments into French: the CERAD experience. *Age Ageing* 1994;23:449–451.
- 26 Aguirre-Acevedo DC, Gomez RD, Moreno S, Henao-Arboleda E, Motta M, Munoz C, Arana A, Pineda DA, Lopera F: Validity and reliability of the CERAD-Col neuropsychological battery (in Spanish). *Rev Neurol* 2007;45:655–660.
- 27 Bertolucci PH, Okamoto IH, Brucki SM, Siviero MO, Toniolo Neto J, Ramos LR: Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arq Neuropsiquiatr* 2001;59:532–536.
- 28 Ganguli M, Chandra V, Gilby JE, Ratcliff G, Sharma SD, Pandav R, Seaberg EC, Belle S: Cognitive test performance in a community-based nondemented elderly sample in rural India: the Indo-U.S. Cross-National Dementia Epidemiology Study. *Int Psychogeriatr* 1996;8:507–524.
- 29 Gurujie O, Unverzagt FW, Osuntokun BO, Hendrie HC, Baiyewu O, Ogunniyi A, Hali KS: The CERAD Neuropsychological Test Battery: norms from a Yoruba-speaking Nigerian sample. *West Afr J Med* 1995;14:29–33.
- 30 Unverzagt FW, Morgan OS, Thesiger CH, Eledemire DA, Luseko J, Pokuri S, Hui SL, Hall KS, Hendrie HC: Clinical utility of CERAD neuropsychological battery in elderly Jamaicans. *J Int Neuropsychol Soc* 1999;5:255–259.
- 31 Sosa AL, Albanese E, Prince M, Acosta D, Ferri CP, Guerra M, Huang Y, Jacob KS, de Rodriguez JL, Salas A, Yang F, Gaona C, Joteeshwaran A, Rodriguez G, de la Torre GR, Williams JD, Stewart R: Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: a cross-sectional survey. *BMC Neurol* 2009;9:48.
- 32 Lucca U, Tettamanti M, Quadri P: The Italian version of Consortium to Establish a Registry of Alzheimer's Disease (CERAD). *Alzheimers Dement* 2008;4:310.
- 33 Karrasch M, Sinerva E, Gronholm P, Rinne J, Laine M: CERAD test performances in amnesic mild cognitive impairment and Alzheimer's disease. *Acta Neurol Scand* 2005;111:172–179.
- 34 Welsh-Bohmer KA, Mohs RC: Neuropsychological assessment of Alzheimer's disease. *Neurology* 1997;49:S11–S13.
- 35 Seo EH, Lee DY, Lee JH, Choo IH, Kim JW, Kim SG, Park SY, Shin JH, Do YJ, Yoon JC, Jhoo JH, Kim KW, Woo JI: Total scores of the CERAD neuropsychological assessment battery: validation for mild cognitive impairment and dementia patients with diverse etiologies. *Am J Geriatr Psychiatry* 2010;18: 801–809.
- 36 Rabbitt P, Diggle P, Smith D, Holland F, McInnes L: Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. *Neuropsychologia* 2001; 39:532–543.
- 37 Lezak MD, Howieson DB, Loring DW, Hannay HJ, Fischer JS: *Neuropsychological Assessment*, ed 4. New York, Oxford University Press, 2004.
- 38 Anastasi A: Coaching, test sophistication, and developed abilities. *Am Psychol* 1981;36: 1086–1093.
- 39 Collie A, Maruff P, Darby DG, McStephen M: The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc* 2003;9:419–428.