Original Research Article

Dementia and Geriatric Cognitive Disorders

Dement Geriatr Cogn Disord 2012;33:73–83
DOI: 10.1159/000336864

Accepted: January 26, 2012
Published online: March 9, 2012

The Assessment of Changes in Cognitive Functioning in the Elderly: Age- and Education-Specific Reliable Change Indices for the SIDAM

Janine Stein[a] Melanie Luppa[a] Wolfgang Maier[b] Franziska Tebarth[b] Kathrin Heser[b] Martin Scherer[c] Thomas Zimmermann[c] Marion Eisele[c] Horst Bickel[e] Edelgard Mösch[e] Siegfried Weyerer[f] Jochen Werle[f] Michael Pentzek[g] Angela Fuchs[g] Birgitt Wiese[h] Jana Prokein[h] Hans-Helmut König[d] Hanna Leicht[d] Steffi G. Riedel-Heller[a] for the AgeCoDe Study Group

[a]Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, [b]Department of Psychiatry, University of Bonn, Bonn, [c]Institute of Primary Medical Care and [d]Department of Medical Sociology and Health Economics, University Medical Center Hamburg-Eppendorf, Hamburg, [e]Department of Psychiatry, Technical University of Munich, Munich, [f]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Mannheim, [g]Department of General Practice, University Medical Center Duesseldorf, Duesseldorf, and [h]Institute for Biometrics, Hannover Medical School, Hannover, Germany

Key Words

Reliable change indices · SIDAM · Neuropsychological assessment · Cognitive disorders · Cognitive change · Change norms · Dementia

Abstract

Background/Aims: The diagnostic criteria for dementia include reliable evidence of cognitive deterioration over time measured by cognitive tests. The Structured Interview for the Diagnosis of Dementia of the Alzheimer Type, Multi-infarct Dementia and Dementia of other Etiology according to DSM-III-R, DSM-IV and ICD-10 (SIDAM) is a neuropsychological instrument to determine cognitive status in patients with mild cognitive impairment (MCI) and dementia. Normative data for changes in cognitive functioning that normally occur in cognitively healthy individuals are required to interpret changes in SIDAM test scores. ***Methods:*** A sample of 1,090 cognitively healthy individuals participating in the German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe) aged 75 years and older was assessed four times at 1.5-year intervals over a period of 4.5 years using the SIDAM. Age- and education-specific reliable change indices (RCIs) accounting for probable measurement error and practice effects were computed for a 90% confidence interval. ***Results:*** Across different age and education subgroups, changes from at least 3–5 points indicated significant (i.e. reliable) changes in SIDAM test scores at the 90% confidence level. ***Conclusion:*** This study offers age- and ed-

Further members of the AgeCoDe Study Group: Heinz-Harald Abholz, Cadja Bachmann, Wolfgang Blank, Sandra Eifflaender-Gorfer, Annette Ernst, Frank Jessen, Hanna Kaduszkiewicz, Teresa Kaufeler, Mirjam Köhler, Carolin Lange, Tobias Luck, Manfred Mayer, Julia Olbrich, Anna Schumacher, Susanne Steinmann, Michael Wagner, Klaus Weckbecker, Dagmar Weeg and Steffen Wolfsgruber. Principal investigators: Wolfgang Maier, Martin Scherer, Hendrik van den Bussche (2002–2011).

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1420–8008/12/0333–0073$38.00/0

Accessible online at:
www.karger.com/dem

Janine Stein
Institute of Social Medicine, Occupational Health and Public Health
University of Leipzig, Philipp-Rosenthal-Strasse 55
DE–04103 Leipzig (Germany)
Tel. +49 341 972 4494, E-Mail Janine.Stein@medizin.uni-leipzig.de

ucation-specific normative data for the SIDAM based upon established RCI methods. The RCI scores provided in this study may help clinicians and researchers to interpret cognitive changes in SIDAM test scores and may contribute to the early detection and diagnosis of MCI and dementia in the elderly.

Introduction

Progressive cognitive deterioration is a major diagnostic marker of dementia and disorders leading to dementia in the elderly [1]. The evaluation of cognitive status is a key component of establishing a diagnosis and monitoring dementia patients during the course of the illness for both clinical practice and research purposes. Therefore, repeated neuropsychological assessments are often used to monitor changes in a patient's cognitive functioning over time. In the context of changing neurological and neuropsychological conditions and in order to plan sufficient treatment and long-term care options for patients, it is most important that reliable information on the rate of deterioration or improvement of cognitive functioning in the development and the course of dementia disease is available.

The critical task when performing repeated testing to evaluate cognitive functioning in geriatric patients over time is how to obtain accurate and reliable data on the rate of changes in the cognitive abilities of a person and how to differentiate clinically significant cognitive change from cognitive changes that normally occur in the process of natural aging. Moreover, varying test results with consecutive testing may also occur due to multiple other factors and statistical artefacts such as measurement error or the occurrence of practice effects and regression to the mean [2, 3]. Thus, interpreting cognitive changes measured by standardized neuropsychological instruments requires knowing how much change is expected to occur normally in cognitively healthy individuals. In order to discriminate between clinically significant changes in cognitive test performance and other factors, there is a need for information on test-retest characteristics of the neuropsychological instruments used. However, this type of normative data, especially for the older population, is available only for a small number of psychometric instruments [4].

To determine reliable and significant changes in test scores, Jacobson et al. [5, 6] were the first to propose the statistical procedure of the reliable change index (RCI) for estimating the probability that an individual's change from pre-test to post-test performance can be ascribed to clinically significant changes in underlying cognitive abilities and is not simply due to chance. In recent years, there have evolved more refined statistical methods for modelling change [7–9] than that suggested by Jacobson and colleagues. According to the RCI methods of Chelune et al. [10] and Hsu [11], a reliable change in a subject's test scores is a change that is unlikely to have occurred by measurement error or practice effects (Chelune et al. [10]) or by regression to the mean (Hsu [11]).

The Structured Interview for the Diagnosis of Dementia of the Alzheimer Type, Multi-infarct Dementia and Dementia of other Etiology according to DSM-III-R, DSM-IV and ICD-10 (SIDAM) is a mixed semi-structured and structured instrument which has been developed for diagnosing mild cognitive impairment (MCI) and dementia according to ICD-10, DSM-III-R and DSM-IV diagnostic criteria [12–14]. The SIDAM considers both the dimensional and the categorical aspects of MCI and dementia. The cognitive section of the SIDAM provides a global cognitive score named 'SISCO', consisting of 55 questions and subscores for orientation, memory, intellectual abilities and higher cortical functions. The SISCO score includes all 30 items of the Mini-Mental State Examination [15]. Change norms (i.e. RCIs) for the accurate classification of cognitive change in the elderly according to SIDAM test scores have been published [16]. However, that study was based on a relatively small sample of older adults (n = 119), and the computation of RCIs did not include correction for the potential influences of sociodemographic variables such as age, education and gender. Nevertheless, SIDAM scores are known to be significantly influenced by age and education [17]. Since the SIDAM represents a frequently used instrument for the screening and diagnosis of dementia in clinical routine and research settings, including epidemiological and cohort studies, in German-speaking areas [17–19], there is a great need to provide age- and education-specific reliable change norms for this instrument.

The objective of the present study was to investigate RCI scores for a large community-based sample of cognitively healthy individuals aged 75 years and older tested on the SIDAM at 1.5-year intervals over a time period of approximately 4.5 years. Specifically, the impact of participants' age, education and gender on cognitive test scores was examined in order to adjust RCI scores for sociodemographic factors. Following the methods suggested by Chelune et al. [10] (accounting for measurement error and practice effects) and Hsu [11] (accounting for regression to the mean), the present study aims to provide reliable rates of changes in SIDAM test scores (SISCO) for clinical and research purposes and to demonstrate how to calculate RCI scores for individual patients.

Methods

Study Design and Sample

Data were derived from a sample of cognitively healthy older individuals participating in the German Study on Ageing, Cognition and Dementia in Primary Care Patients (AgeCoDe), a prospective longitudinal study on the early detection of MCI and dementia in primary care established by the German Research Network on Dementia. The study was approved by the local ethics committee, and all subjects gave written informed consent. The study was conducted in six study centres: Bonn, Düsseldorf, Hamburg, Leipzig, Mannheim and Munich. The subjects were recruited between January 1, 2003 and November 30, 2004. Altogether, 138 general practitioners (GPs) participated in the recruitment process, 19–29 GPs in each centre. Inclusion criteria were age of 75 years or older, the absence of dementia based on diagnostics done by the GPs and at least one contact with the GP within the last 12 months. Exclusion criteria were GP consultations only by home visits, residence in a nursing home, a severe illness which the GP deemed would be fatal within 3 months, an insufficient knowledge of the German language, deafness or blindness, lack of ability to consent and status as only an occasional patient of the participating GP.

Figure 1 shows detailed information about the sample selection process. For our study sample, we selected cognitively unimpaired participants using the following criteria for inclusion: (1) SIDAM score at each visit above age- and education-specific cut-off scores (1 SD) for MCI as given by Busse et al. [17]; (2) no diagnosis of dementia during the course of the study according to DSM-IV diagnostic criteria, and (3) valid test data for the SIDAM at each of the four assessments. In all, 2,237 of the interviewed patients were excluded from the sample for the calculation of RCIs. Of these, 1,359 (60.8%) had incomplete test data or did not participate in all four assessments. Of these 1,359 subjects, 511 (37.6%) were excluded from the study sample because of death during the course of the study and 848 (62.4%) because of incomplete participation in all four assessments. Of the 848 subjects who did not participate in all four assessments, 717 (52.76%) refused to participate in at least one of the four assessments during the course of the study, 52 (3.82%) moved house or moved to an unknown address and 30 (2.21%) refused to participate because of health reasons. The remaining subjects refused to participate because of other reasons. The subjects who were excluded from the study (n = 1,359) and included in the study (n = 1,090) did not differ significantly with regard to gender [$\chi^2 = 2.735$, degrees of freedom (d.f.) = 1, p = 0.098] or level of education ($\chi^2 = 1.192$, d.f. = 1, p = 0.275). However, the comparison indicated that included subjects were significantly younger than excluded subjects (79.1 vs. 80.2 years; $\chi^2 = 27.879$, d.f. = 1, $p < 0.001$). Furthermore, 29 subjects (1.3%) were excluded from the study sample because they were younger than 75 years, 174 (7.8%) because they were demented, 494 (22.1%) because they had MCI, 10 (0.4%) who did not have German as their first language, 155 (6.9%) for a diagnosis of a severe illness with potential impact on cognitive test performance (i.e. debility, Parkinson's disease, stroke, brain tumour, organic brain syndrome, neurological failures, psychosis/depression, alcohol abuse/alcoholism) and 16 (0.7%) who had severe sensory impairment. The calculation of RCI values is based on the remaining 1,090 subjects.

Procedure and Instruments

The neuropsychological evaluation, including a comprehensive cognitive assessment to determine the cognitive status of study participants, was performed by trained physicians and psychologists administering fully structured clinical interviews to participants in their home environment. At each assessment, participants were assessed using the neuropsychological test part of the SIDAM, consisting of 55 items (SISCO score). The SIDAM includes all 30 items of the Mini-Mental State Examination. All items of the SIDAM rely on DSM-III-R, DSM-IV and ICD-10 algorithms, which allow a reliable separation of persons with dementia from those without such a disorder or MCI, respectively. The diagnoses of dementia were based on DSM-IV criteria [1] following the diagnostic algorithm of the SIDAM and the Global Deterioration Scale [20], which had to be scored with at least 4. Unclear diagnoses were discussed in a standing diagnostic clearing platform consisting of multiple experts in the field. Furthermore, the 15-item version of the Geriatric Depression Scale [21, 22] was employed to determine depressive symptoms. Patients with Geriatric Depression Scale scores of 10 or more points (cut-off for severe depressive symptoms) were excluded from our study sample because of the possible impact of depressive symptoms on cognitive test performance. Data on sociodemographic variables as well as possible risk factors for dementia were also collected. In addition, the GPs filled in a questionnaire for each patient to determine health status, clinical diagnoses and comorbidity.

Statistics

All statistical analyses were carried out using PASW Statistics 18 for Windows. If not otherwise stated, the level of α error was set to 0.05 in all computations. The impact of the sociodemographic variables age, education and gender on SIDAM test performance was examined via analyses of covariance. Differences in SIDAM test performances between the four assessments were investigated using the Friedman test (nonparametric equivalent of a repeated-measure ANOVA) followed by pair-wise comparisons using the Wilcoxon test. For pair-wise comparisons, the level of significance was adjusted for multiple testing by the Bonferroni-Holm method. Nonparametric statistical tests were used, because SIDAM test scores within the study sample were not subject to normal distribution at all time points. Effects of regression to the mean were investigated by dividing the sample into four groups according to SIDAM percentiles (25th, median, 75th), which represent above-average versus below-average levels in test performances on a group level. Regression to the mean refers to the phenomenon that a variable that is extreme on its initial measurement (e.g. high test score) will, on average, tend to move towards the mean of the population distribution when reassessed in a later measurement [23]. For all adjacent assessments [follow-up 1 (FU1)-baseline (BL), follow-up 2 (FU2)-FU1, follow-up 3 (FU3)-FU2], differences between the groups with regard to the mean

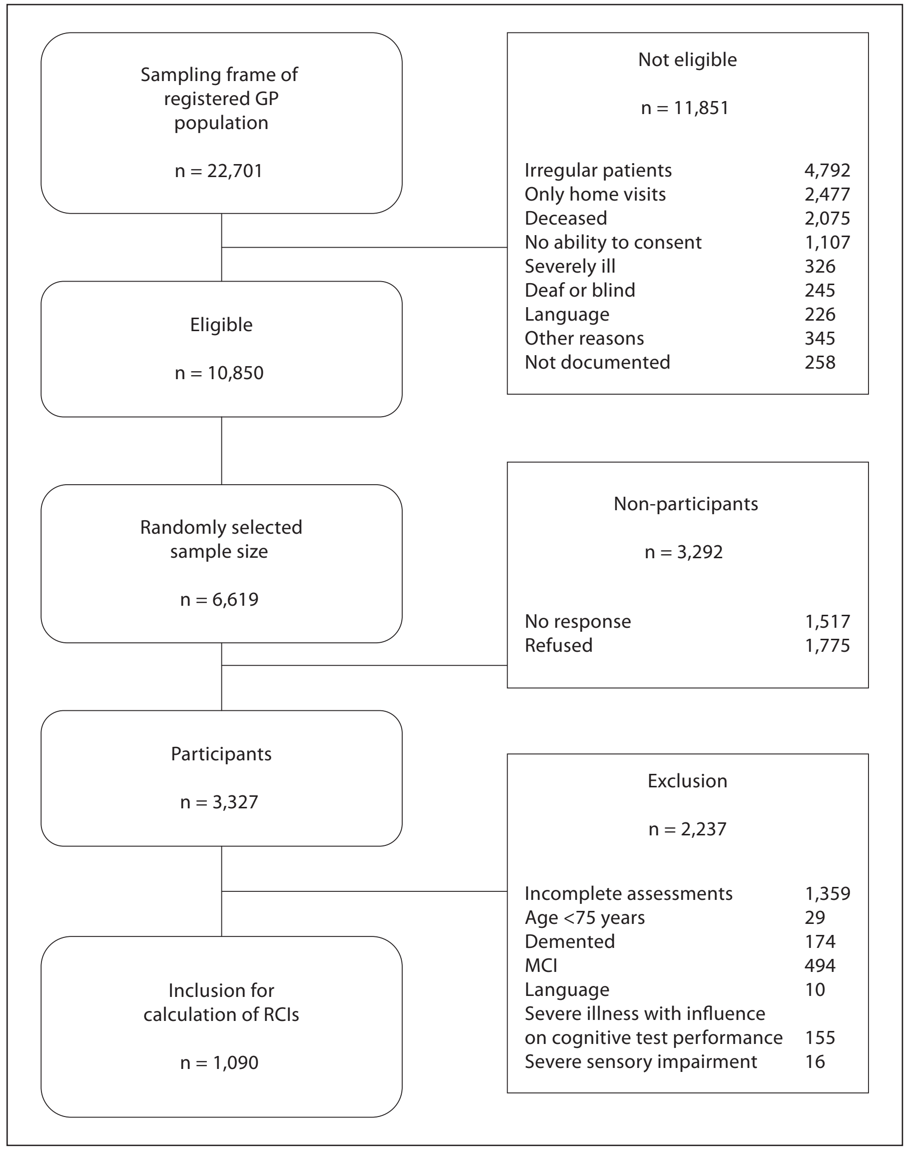

Fig. 1. Sample selection flowchart.

change in SIDAM test scores were analysed using nonparametric Kruskal-Wallis testing. The test-retest reliability of the SIDAM was determined in terms of nonparametric Spearman's r. The calculation of age- and education-specific RCIs was based on the procedures of Chelune et al. [10] and Hsu [11]; a detailed description of both methods is provided in the Appendix.

Results

The characteristics of the study sample are displayed in table 1. Educational groups were determined based on the new Comparative Analysis of Social Mobility in Industrial Nations educational classification [24]. The study sample was divided into the two educational categories 'low' and 'middle/high'; the latter group was combined into one group because of small sample sizes.

The results of the analyses of covariance are summarized in table 2 and revealed that the sociodemographic variables age and education had a significant impact on SIDAM test performance. At different assessments, younger age and higher education were significantly associated with better test performances on the SIDAM. The gender of the study participants did not impact on SIDAM test scores.

The mean raw scores on the SIDAM across different age and education subgroups and for the total sample are

Table 1. Baseline characteristics of the sample (n = 1,090)

Characteristic	
Age, years	79.1 ± 3.17
Range	75–92
Age groups	
75–79 years	641 (58.8)
≥80 years	449 (41.2)
Gender	
Female	735 (67.4)
Male	355 (32.6)
Education	
Low	743 (68.2)
Middle/high	347 (31.8)
Comparison with baseline, interval in years	
BL–FU1	1.63 ± 0.18
BL–FU2	3.07 ± 0.16
BL–FU3	4.64 ± 0.21
Adjacent assessments, interval in years	
FU1–FU2	1.44 ± 0.17
FU2–FU3	1.57 ± 0.21

Values represent means ± SD or numbers (percentage), as appropriate. Education: educational classification according to the new Comparative Analysis of Social Mobility in Industrial Nations [24], as follows: low = inadequately completed general education, general elementary education, basic vocational qualification or general elementary education and vocational qualification; middle = intermediate vocational qualification or intermediate general qualification and vocational qualification, intermediate general qualification, general maturity certificate, vocational maturity certificate/general maturity certificate and vocational qualification; high = lower tertiary education – general diplomas/diplomas with vocational emphasis, higher tertiary education – lower level/higher level.

displayed in table 3. The Friedman test revealed significant differences between SIDAM test scores in the four repeated assessments (Friedman χ^2 = 9.980, d.f. = 3, p = 0.019). Following the Friedman test, pair-wise analyses via Wilcoxon tests indicated significant differences in test performances between BL and FU1 assessments (Z = –2.306, p = 0.021) and BL and FU2 assessments (Z = –3.046, p = 0.002) as well as BL and FU3 assessments (Z = –2.848, p = 0.004). Altogether, the results indicated the presence of practice effects and/or normal age-related cognitive decline at a group level.

The investigation of possible effects of regression to the mean for adjacent assessments via Kruskal-Wallis testing showed significant differences between mean SIDAM test performance levels in adjacent assessments at a group level. Participants with different pre-test SIDAM levels (above-average versus below-average levels) yielded significantly different mean SIDAM change scores in compared assessments (FU1–BL: χ^2 = 194.781, d.f. = 3, p < 0.001; FU2–FU1: χ^2 = 168.794, d.f. = 3, p < 0.001; FU3–FU2: χ^2 = 105.057, d.f. = 2, p < 0.001), which can be considered as effects of regression to the mean. Table 4 summarizes the mean raw difference in SIDAM scores for compared assessments. Observed differences in SIDAM were rather low (≤0.37), indicating that the ratio of individuals showing a gain in SIDAM test score approximately equalled the ratio of individuals showing a loss in SIDAM test score. Reliable changes in SIDAM test scores (improvement and decline), according to the method suggested by Chelune et al. [10], are also displayed in table 4. From BL to FU3, for example, 9.1% of our cognitively healthy study participants experienced either a reliable deterioration (4.5%) or a reliable improvement (4.6%). Overall, these results are in line with the definition of the RCI suggested by Chelune et al. [10] with regard to the proportion of individuals experiencing a reliable change in SIDAM test scores (reliable deterioration or reliable improvement).

Test-retest reliability coefficients ranged from 0.47 to 0.65 and appeared to be higher for adjacent assessments (table 5). Calculated RCI scores according to the method of Chelune et al. [10] are also shown in table 5. Across different age and education subgroups, changes in SIDAM from at least 3 to up to 5 points were needed to conclude at the 90% confidence level that a reliable change in SIDAM test scores (improvement or deterioration) had occurred without being evoked by factors other than true cognitive changes, such as measurement error, practice effects or natural aging effects. On average, RCI ranges appeared to be larger for younger persons (age 75–79 years) with a lower educational level than for older persons (≥80 years) with a higher educational level.

The presented data allow the reader to calculate RCIs for a given individual according to the method of Hsu [11], which accounts for effects of regression to the mean (for the formula, see the Appendix). Two examples of how to compute the RCI are demonstrated, because tabulation of RCIs according to Hsu [11] for the SIDAM would be too space-consuming. In the first example, a person aged 77 years who completed general elementary education (8 years) has a pre-test SIDAM score (X_1) of 55, a post-test SIDAM score (X_2) of 50 and was assessed after a time interval of 1.5 years. According to table 5 (see first column, comparison BL–FU1), the standard error of prediction (SEP) is 2.39 and r_{xx} is 0.57 (test-retest reliability). According to table 3 (see first column), the pre-test SIDAM

Table 2. Analysis of covariance of the influences of age, education and gender on SIDAM test performance (n = 1,090)

Time interval	Age			Education			Gender		
	F	d.f.	p	F	d.f.	p	F	d.f.	p
FU3–BL	24.368	1	0.000**	30.877	1	0.000**	0.320	1	0.571
FU2–BL	12.264	1	0.000**	27.104	1	0.000**	0.673	1	0.412
FU1–BL	8.553	1	0.004*	14.771	1	0.000**	1.938	1	0.164
FU2–FU1	10.524	1	0.001**	40.025	1	0.000**	0.110	1	0.740
FU3–FU2	18.761	1	0.000**	31.265	1	0.000**	0.162	1	0.687

* $p \leq 0.05$; ** $p \leq 0.001$.

Table 3. SIDAM mean raw scores in four assessments (n = 1,090)

	Age 75–79 years		Age ≥80 years	
	low education (n = 437)	middle/high education (n = 306)	low education (n = 204)	middle/high education (n = 143)
Mean raw score				
BL	49.70 (2.91)	52.12 (2.09)	48.65 (3.36)	51.22 (2.58)
FU1	49.97 (2.87)	52.01 (2.12)	48.92 (3.11)	51.07 (2.83)
FU2	50.06 (2.78)	52.16 (1.94)	48.89 (3.31)	51.31 (2.57)
FU3	50.07 (2.91)	52.41 (2.01)	48.68 (3.69)	51.14 (2.74)

	Total sample				
	mean	range	percentiles		
			25th	median	75th
BL	50.06 (3.14)	33–55	48	51	52
FU1	50.20 (3.02)	36–55	48	51	52
FU2	50.29 (3.02)	34–55	49	51	52
FU3	50.26 (3.26)	33–55	49	51	53

Values in parentheses represent SDs.

mean is 49.70 [BL, group mean pre-test score (M_1)] and the post-test SIDAM mean is 49.97 [FU1, group mean post-test score (M_2)]. In this case, the RCI can be calculated as follows:

$$RCI = \frac{\left(\left(X_2 - M_2\right) - r_{xx}\left(X_1 - M_1\right)\right)}{SEP}$$

$$RCI = \frac{\left(\left(50 - 49.97\right) - 0.57\left(55 - 49.70\right)\right)}{2.39} = -1.251$$

This RCI value is inside the 90% confidence interval of ±1.645 and therefore expected to occur without a real change in post-test scores. In other words, there is not enough evidence to conclude with sufficient certainty that this reflects a reliable decline in post-test scores. In the second example, a person aged 77 years who completed general elementary education (8 years) has a pre-test SIDAM score of 50, a post-test SIDAM score of 45 and the time interval is again 1.5 years. Analogously, the RCI can be calculated as follows:

$$RCI = \frac{\left(\left(45 - 49.97\right) - 0.57\left(50 - 49.70\right)\right)}{2.39} = 2.151$$

Table 4. SIDAM mean raw difference scores for compared assessments and reliable change in SIDAM test scores according to the method suggested by Chelune et al. [10]

	Age 75–79 years		Age ≥80 years		Total sample (n = 1,090)	Reliable change in total sample (n = 1,090)		
	low education (n = 437)	middle/high education (n = 306)	low education (n = 204)	middle/high education (n = 143)		RD	normal	RI
Mean raw difference[1]								
BL–FU1	0.27 ± 2.52	–0.11 ± 2.08	0.27 ± 2.74	–0.15 ± 2.69	0.14 ± 2.53	50 (4.6)	997 (91.5)	43 (3.9)
BL–FU2	0.36 ± 2.81	0.04 ± 2.06	0.24 ± 2.70	0.08 ± 2.33	0.23 ± 2.59	39 (3.6)	1,005 (92.2)	46 (4.2)
BL–FU3	0.37 ± 2.76	0.30 ± 2.18	0.04 ± 3.31	–0.08 ± 2.52	0.20 ± 2.80	49 (4.5)	991 (90.9)	50 (4.6)
FU1–FU2	0.09 ± 2.54	0.15 ± 2.00	–0.03 ± 2.69	0.24 ± 2.22	0.09 ± 2.45	44 (4.0)	997 (91.5)	49 (4.5)
FU2–FU3	0.01 ± 2.50	0.25 ± 1.83	–0.21 ± 3.02	–0.17 ± 2.51	–0.03 ± 2.56	67 (6.1)	987 (90.6)	36 (3.3)

Values represent means ± SD or numbers (percentage), as appropriate. RD = Reliable deterioration; RI = reliable improvement.
[1] Positive values reflect improvements and negative values reflect decline in cognitive test scores.

This RCI value exceeds the 90% confidence interval of ±1.645 and indicates by definition a significant (i.e. reliable) change in post-test scores. In both examples, the raw difference in test scores was 5. Note that in the first example, the pre-test score was an extreme value (55) and the decline was towards the mean. In contrast, in the second example, the pre-test score was near the mean and the post-test score declined away from the mean. Thus, both cases reflect different situations and are interpreted differentially. A change in test scores towards the mean resulted in a lower RCI, and a change of test scores away from the mean resulted in a higher RCI, which indicated a significant (i.e. reliable) change.

Discussion

The purpose of this study was to evaluate RCI scores for the SIDAM in consideration of sociodemographic factors and their impact on SIDAM test scores. Clinically meaningful (i.e. reliable) changes in SIDAM test scores were evaluated by computing age- and education-specific RCI scores for a large sample of cognitively healthy persons in older age groups (mean age 79.1 years) who were tested at 1.5-year intervals over a period of approximately 4.5 years. The educational level of each participant was classified on the basis of the new Comparative Analysis of Social Mobility in Industrial Nations educational classification, enabling the comparison of different European educational systems [24].

In accordance with other findings, the sociodemographic factors age and education had a significant impact on SIDAM test performance [17]. Generally, test scores of comparable neuropsychological instruments, for example the Alzheimer's Disease Assessment Scale cognitive subtest [25] and its revised version, the Vascular Dementia Assessment Scale cognitive subscale [26], are known to be influenced by age and education [27, 28]. Consequently, the impact of sociodemographic variables on test scores should be considered when calculating and interpreting RCIs.

Consistent with the findings of previous research [16], group level analyses via Friedman and Wilcoxon tests indicated small but significant evidence for practice effects. The small magnitude of practice effects (<1 point) could be explained by the fact that older adults' abilities to benefit from learning were found to have declined in comparison to middle-aged adults' abilities to benefit from previous exposure to test materials due to age-related changes in fluid intelligence, perceptual speed and attention/concentration capacity [29–31]. Nevertheless, practice effects are likely to occur in repeated neuropsychological assessments and should be considered when calculating RCI scores.

As part of this study, possible effects of regression to the mean on SIDAM test scores were examined. Results of Kruskal-Wallis testing may reflect that extreme pre-test SIDAM test scores may yield to less extreme post-test SIDAM test scores closer to the SIDAM group mean score, a statistical phenomenon that was also observed

Table 5. Age- and education-specific test-retest reliability and RCIs for the SIDAM according to the method of Chelune et al. [10]

	Age 75–79 years		Age ≥80 years	
	low education (n = 473)	middle/high education (n = 204)	low education (n = 306)	middle/high education (n = 143)
Test-retest reliability				
BL–FU1	0.57	0.53	0.62	0.58
BL–FU2	0.47	0.48	0.65	0.56
BL–FU3	0.53	0.45	0.54	0.47
FU1–FU2	0.56	0.52	0.63	0.62
FU2–FU3	0.60	0.57	0.61	0.48
SED				
BL–FU1	2.70	2.04	2.93	2.36
BL–FU2	2.99	2.12	2.79	2.42
BL–FU3	2.83	2.18	3.21	2.65
FU1–FU2	2.69	2.09	2.68	2.47
FU2–FU3	2.50	1.80	2.91	2.62
SEP				
BL–FU1	2.39	1.78	2.64	2.10
BL–FU2	2.57	1.83	2.54	2.14
BL–FU3	2.47	1.86	2.82	2.27
FU1–FU2	2.38	1.82	2.42	2.23
FU2–FU3	2.23	1.59	2.62	2.25
RCIs				
BL–FU1 (1.63)[1]	–4≤, ≥+5	–3≤, ≥+3	–5≤, ≥+5	–4≤, ≥+4
BL–FU2 (3.07)	–5≤, ≥+5	–3≤, ≥+4	–4≤, ≥+5	–4≤, ≥+4
BL–FU3 (4.64)	–4≤, ≥+5	–3≤, ≥+4	–5≤, ≥+5	–4≤, ≥+4
FU1–FU2 (1.44)	–4≤, ≥+5	–3≤, ≥+4	–4≤, ≥+4	–4≤, ≥+4
FU2–FU3 (1.57)	–4≤, ≥+4	–3≤, ≥+3	–5≤, ≥+5	–4≤, ≥+4

RCIs were calculated according to the method of Chelune et al. [10], which accounts for measurement error and practice effects. Test-retest reliability was calculated in terms of Spearman's r. Correction for practice/decline was determined by subtracting the post-test SIDAM mean score from the pre-test SIDAM mean score and then rounding to the nearest integer.

[1] Mean time intervals in years are given in parentheses.

and reported in other studies [16, 32]. In this study, data are provided to calculate individual RCI scores according to the method of Hsu [11], which takes into consideration possible effects of regression to the mean and may substantially contribute to the precise measurement of changes in test scores.

The test-retest reliability coefficients observed in this study were acceptable and similar to the test-retest reliabilities observed in a previous study [16]. In the next step, RCIs were derived from the psychometric qualities of the SIDAM to estimate changes. In this study, changes in SIDAM from at least 3 to up to 5 points across different age and education subgroups were needed to conclude at the 90% confidence level that clinically meaningful and reliable changes in SIDAM test scores had occurred. Smaller changes in SIDAM test scores are expected to occur without 'real' (i.e. reliable) changes in cognitive test performance and underlying cognitive abilities and most likely reflect measurement error, practice effects and/or normal aging effects. In comparison, Hensel et al. [16] reported RCI scores ranging from at least 4 to up to 7 points on the SIDAM, indicating reliable change. Presumably, the larger sample size in this study allowed more precise estimations, leading to smaller ranges of RCI scores. Furthermore, the magnitude of RCI ranges in this study seems to vary across different age and education subgroups, with older and more highly educated individuals tending to yield smaller RCI ranges (≤1 point). Research suggests that different cognitive abilities, including fluid and crystallised abilities, show different trajectories of change across the life span. Several moderating lifestyle factors, including educational, occupational and mental activity, have been identified, which may have a protective impact on the development of normal age-associated or pathological cognitive decline [29, 33, 34]. Accordingly, the high performance of older individuals in this study may be mediated by their high educational level. The results of this study may support the concept of 'cognitive reserve', suggesting that higher educational attainment may lead to a higher cognitive reserve capacity, which may lower the risk for cognitive decline, pathological brain changes and incident dementia [35].

In recent studies, the SIDAM was found to significantly discriminate between subjects with and without dementia and to be useful for the detection of pre-clinical stages of dementia such as MCI [13, 19, 36, 37]. There is evidence from many studies that especially persons in older age groups suffering from MCI show an increased risk of developing dementia in the future [38–40]. In clinical practice, the use of neuropsychological tests, for example the SIDAM, is essential for the diagnosis of MCI and dementia [41]. Insufficient psychometric quality characteristics of the instrument and a lack of normative data can diminish diagnostic accuracy and lead to misclassification. Against this background, change norms for the SIDAM to diagnose and monitor cognitive changes over longer time periods are indispensable, and RCI scores may enhance the reliability of MCI and dementia diagnoses. Hence, this study extends the results of previous investigations by providing age- and education-specific RCI reference values accounting for factors such as measurement error, practice effects (method of Chelune et al. [10]) and regression to the mean (method of Hsu et al. [11]) for cognitively healthy individuals aged 75 years and older. In a comparison of different methods for measuring cognitive change in older adults, both RCI methods were found to be diagnostically useful in the elderly population [8]. Moreover, RCI methods were found to contribute to the prediction and early diagnosis of dementia [18].

The limitations of this study are related to the applicability of the calculated RCIs. Firstly, the presented RCI scores may only be appropriate for persons aged 75 years and older. However, research indicates a higher prevalence of dementia diseases with increasing age [42–44]. Therefore, RCI scores for older age groups may be particularly useful. Secondly, test-retest intervals substantially longer or shorter than 1.5 years may not be interpretable. Additional research examining the utility of the RCI methods with different test-retest intervals, for example 1 or 2 years, is needed. On the other hand, other investigations suggest that RCI methods should be used after a sufficient amount of time between assessments to account for normal variability in natural aging and to avoid additional practice effects, which were found to be most likely to occur with shorter test-retest intervals [7, 8, 45]. Furthermore, the study did not systematically investigate the potential impact of other variables on test scores, for example motivation, intellectual capacity, medication or comorbidity. Thus, cognitively healthy participants could have been excluded from the study sample due to low test scores from factors other than cognitive impairment. However, the comprehensive cognitive assessment applied in this study widely ensured that only participants with no MCI or dementia were included in the study sample.

In conclusion, nonpathological cognitive decline usually occurs as part of normal aging processes. The assessment of reliable changes in cognitive abilities in older adults is crucially important in order to differentiate these normal aging effects from pathological cognitive decline probably leading to dementia. Cognitive screening and neuropsychological testing supported by RCI scores may help clinicians to determine and interpret cognitive changes in the elderly population over time and enhance early detection and reliable diagnosis of MCI and dementia. Future research should address the publication of change norms for further established neuropsychological instruments. Future research should also include the examination and validation of change score methods, particularly the RCI, in other clinical samples of older adults and for study designs covering different test-retest intervals and overall time periods. In addition, further analyses of the sensitivity and prognostic value of RCIs would be desirable.

Appendix

Age- and education-specific RCIs were calculated following the procedure of Chelune et al. [10]:

$$RCI = \frac{\left(\left(X_2 - X_1\right) - \left(M_2 - M_1\right)\right)}{SED}$$

where X_1 is the observed pre-test score, X_2 is the observed post-test score, M_1 is the group mean pre-test score, M_2 is the group mean post-test score and SED is the standard error of a difference. The difference $(X_2 - X_1)$ represents the individual change in SIDAM test score (individual level), and the difference $(M_2 - M_1)$ reflects the mean change in SIDAM test score within the normative sample (group level). An average improvement in test performance within the normative sample $(M_2 - M_1 > 0)$ is interpreted as a practice effect at a group level; thus, the individual change is corrected for the average practice effect using the term $((X_2 - X_1) - (M_2 - M_1))$. The SED is taken as the standard deviation of the mean observed difference score using the formula $SED = (2 \times SEM^2)^{1/2}$, where $SEM = SD[(1 - r)^{1/2}]$, SD represents the pre-test SD of the SIDAM and r is the reliability coefficient.

RCI values larger than ± 1.645 ($p < 0.10$, two-tailed prediction) were defined as representing a reliable change (i.e. reliable improvement, reliable deterioration) in SIDAM test scores and would be expected to occur without real change only 10% of the time by chance (two-tailed prediction). In order to determine the SIDAM difference $(X_2 - X_1)$ which is equivalent to an RCI of 1.645, the formula above was solved for $(X_2 - X_1)$. The results were rounded to the subsequent integer, as an individual can change only in points (integers). Furthermore, data are provided to enable the reader to calculate age-, education- and gender-specific RCIs according to the method suggested by Hsu [11] with the following formula:

$$RCI = \frac{\left(\left(X_2 - M_2\right) - r_{xx}\left(X_1 - M_1\right)\right)}{SEP}$$

where X_1, X_2, M_1 and M_2 are as previously defined and $SEP = SD\,(1 - r^2_{xx})^{1/2}$. SD is the SD of the pre-test score and r_{xx} is the reliability coefficient [11].

Acknowledgments

This study/publication is part of the German Research Network on Dementia (KND) and the German Research Network on Degenerative Dementia (KNDD) and was funded by the German Federal Ministry of Education and Research (KND grants

01GI0102, 01GI0420, 01GI0422, 01GI0423, 01GI0429, 01GI0431, 01GI0433, 01GI0434; KNDD grants 01GI0710, 01GI0711, 01GI0712, 01GI0713, 01GI0714, 01GI0715, 01GI0716, 01ET1006B).

We wish to thank all participating patients and their GPs for their good collaboration.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

1 American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, ed 4 (Text Revision). Washington, American Psychiatric Association, 2000.

2 Temkin NR, Heaton RK, Grant I, Dikmen SS: Detecting significant change in neuropsychological test performance: a comparison of four models. J Int Neuropsychol Soc 1999;5:357–369.

3 Tombaugh TN: Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. Arch Clin Neuropsychol 2005;20:485–503.

4 Stein J, Luppa M, Brähler E, König HH, Riedel-Heller SG: The assessment of changes in cognitive functioning: reliable change indices for neuropsychological instruments in the elderly – a systematic review. Dement Geriatr Cogn Disord 2010;29:275–286.

5 Jacobson NS, Follette WC, Revenstorf D: Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. Behav Ther 1984;15:336–352.

6 Jacobson NS, Truax P: Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59:12–19.

7 Collie A, Maruff P, Makdissi M, McStephen M, Darby DG, McCrory P: Statistical procedures for determining the extent of cognitive change following concussion. Br J Sports Med 2004;38:273–278.

8 Frerichs RJ, Tuokko HA: A comparison of methods for measuring cognitive change in older adults. Arch Clin Neuropsychol 2005;20:321–333.

9 Maassen GH: Principles of defining reliable change indices. J Clin Exp Neuropsychol 2000;22:622–632.

10 Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA: 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;67:594–598.

12 Zaudig M, Mittelhammer J, Hiller W, Pauls A, Thora C, Morinigo A, Mombour W: 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;21:225–236.

13 Zaudig M: A new systematic method of measurement and diagnosis of 'mild cognitive impairment' and dementia according to ICD-10 and DSM-III-R criteria. Int Psychogeriatr 1992;4(suppl 2):203–219.

14 Zaudig M, Hiller W: SIDAM – Strukturiertes Interview für die Diagnose einer Demenz vom Alzheimer Typ, der Multi-Infarkt-(oder vaskulären) Demenz und Demenzen anderer Ätiologie nach DSM-III-R, DSM-IV und ICD-10 (SIDAM-Handbuch). Huber, Bern, 1996.

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 Hensel A, Angermeyer MC, Zaudig M, Riedel-Heller SG: Measuring cognitive change in older adults: reliable change indices for the SIDAM. J Neurol 2007;254:91–98.

17 Busse A, Aurich C, Zaudig M, Riedel-Heller S, Matschinger H, Angermeyer MC: 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) (in German). Z Gerontol Geriatr 2002;35:565–574.

18 Hensel A, Angermeyer MC, Riedel-Heller SG: Measuring cognitive change in older adults. Do reliable change indices of the SIDAM predict dementia? J Neurol 2007;254:1359–1365.

19 Bickel H, Mösch E, Förstl H: Screening of cognitive functions and the prediction of incident dementia by means of the SIDAM. Psychiatr Prax 2007;34:139–144.

20 Reisberg B, Ferris SH, de Leon MJ, Crook T: The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982;139:1136–1139.

21 Gauggel S, Birkner B: Validity and reliability of a German version of the Geriatric Depression Scale (GDS). Z Klin Psychol Forsch Prax 1999;28:18–27.

22 Sheik JI, Yesavage JA: Geriatric Depression Scale (GDS). Recent evidence and development of a shorter version; in Brink TL (ed): Clinical Gerontology: A Guide to Assessment and Intervention. New York, The Haworth Press, 1986, pp 165-173.

23 Davis CE: The effect of regression to the mean in epidemiologic and clinical studies. Am J Epidemiol 1976;104:493–498.

24 Brauns H, Steinmann S: Educational reform in France, West-Germany and the United Kingdom: updating the CASMIN educational classification. ZUMA-Nachrichten 1999;44:7–44.

25 Rosen WG, Mohs RC, Davis KL: A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;1356–1364.

26 Ferris SH: General measures of cognition. Int Psychogeriatr 2003;15(suppl 1):215–217.

27 Ylikoski R, Jokinen H, Andersen P, Salonen O, Madureira S, Ferro J, Barkhof F, van der FW, Schmidt R, Fazekas F, Scheltens P, Waldemar G, Salvadori E, Pantoni L, Inzitari D, Erkinjuntti T: Comparison of the Alzheimer's Disease Assessment Scale Cognitive Subscale and the Vascular Dementia Assessment Scale in differentiating elderly individuals with different degrees of white matter changes. The LADIS Study. Dement Geriatr Cogn Disord 2007;24:73–81.

28 Madureira S, Verdelho A, Moleiro C, Ferro JM, Erkinjuntti T, Jokinen H, Pantoni L, Fazekas F, van der Flier W, Visser M, Waldemar G, Wallin A, Hennerici M, Inzitari D: Neuropsychological predictors of dementia in a three-year follow-up period: data from the LADIS study. Dement Geriatr Cogn Disord 2010;29:325–334.

29 Anstey KJ, Low LF: Normal cognitive changes in aging. Aust Fam Physician 2004;33:783–787.

30 Finkel D, Reynolds CA, McArdle JJ, Pedersen NL: Age changes in processing speed as a leading indicator of cognitive aging. Psychol Aging 2007;22:558–568.

31 McCaffrey RJ, Westervelt HJ: Issues associated with repeated neuropsychological assessments. Neuropsychol Rev 1995;5:203–221.

32 Hensel A, Angermeyer MC, Riedel-Heller SG: Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination. J Neurol Neurosurg Psychiatry 2007;78:1298–1303.

33 Craik FI, Bialystok E: Cognition through the lifespan: mechanisms of change. Trends Cogn Sci 2006;10:131–138.

34 Hamid TA, Krishnaswamy S, Abdullah SS, Momtaz YA: Sociodemographic risk factors and correlates of dementia in older Malaysians. Dement Geriatr Cogn Disord 2010;30:533–539.

35 Stern Y: Cognitive reserve. Neuropsychologia 2009;47:2015–2028.
36 Burkart M, Heun R, Maier W, Benkert O: Dementia screening in routine clinical practice. A comparative analysis of MMSE, SIDAM and ADAS (in German). Nervenarzt 1998;69:983–990.
37 Heun R, Papassotiropoulos A, Jennssen F: The validity of psychometric instruments for detection of dementia in the elderly general population. Int J Geriatr Psychiatry 1998;13: 368–380.
38 Amieva H, Le Goff M, Millet X, Orgogozo JM, Peres K, Barberger-Gateau P, Jacqmin-Gadda H, Dartigues JF: Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol 2008;64: 492–498.
39 Defranceso M, Schocke M, Messner HJ, Deisenhammer EA, Hinterhuber H, Marksteiner J, Weiss EM: Conversion from MCI (Mild Cognitive Impairment) to Alzheimer's disease: diagnostic options and predictors (in German). Neuropsychiatry 2010;24:88–98.
40 Mariani E, Monastero R, Mecocci P: Mild cognitive impairment: a systematic review. J Alzheimers Dis 2007;12:23–35.
41 Pentzek M, Fuchs A, Wiese B, Abholz HH: Which information do GPs use to rate the cognitive status of elderly non-demented patients? (in German). Psychiatr Prax 2010;37: 377–383.
42 Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH: Dementia incidence continues to increase with age in the oldest old: the 90+ study. Ann Neurol 2010;67:114–121.
43 Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E, Scazufca M: Global prevalence of dementia: a Delphi consensus study. Lancet 2005;366:2112–2117.
44 Jellinger KA, Attems J: Prevalence and pathology of dementia with Lewy bodies in the oldest old: a comparison with other dementing disorders. Dement Geriatr Cogn Disord 2011;31:309–316.
45 Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO: Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. J Appl Psychol 2007;92:373–385.