Population-Based Distribution and Psychometric Properties of a Short Cognitive Performance Measure in the Population-Based Heinz Nixdorf Recall Study

Natalia Wege, Martha Dlugaj, Johannes Siegrist, Nico Dragano, Raimund Erbel, Karl-Heinz Jöckel, Susanne Moebus, Christian Weimar on behalf of the Heinz Nixdorf Recall Study Investigative Group

*Department of Medical Sociology, University of Düsseldorf, Düsseldorf, and *Department of Neurology, *Institute for Medical Informatics, Biometry and Epidemiology, and *Clinic of Cardiology, West-German Heart Centre, University Hospital of Essen, University Duisburg-Essen, Essen, Germany

**Key Words**
Cognitive functioning · Ageing · Population-based study · Psychometric properties · Population-based norms · Mild cognitive impairment

**Abstract**

**Background:** Providing a valid and sensitive measure of different domains of cognitive performance in epidemiologic studies of early old-age populations presents a methodological challenge, given the broad range of variability in cognitive functioning in this age group. **Objectives:** (1) To provide data on the distribution of cognitive performance scores in a representative sample of an early old-age population, and (2) to assess psychometric properties of a short cognitive performance measure developed within the framework of a cohort study. **Design:** Population-based cohort study. **Setting and Participants:** As part of the second examination of the Heinz Nixdorf Recall Study, 4,145 participants aged 50–80 years underwent a short cognitive performance assessment composed of 5 subtests with a mean duration of 7.31 min. Additionally, a subsample of 656 participants had a detailed neuropsychological and neurological examination. **Methods:** Age- and education-specific cognitive performance scores in the total sample were calculated. Based on data from the subsample, concurrent validity was examined by comparing findings with a clinically validated neuropsychological assessment. **Results:** In the total sample, younger and more highly educated participants had higher scores of cognitive performance. In the subsample, a good accuracy [area under the curve (AUC) = 0.81 (0.74–0.87)] of the short cognitive performance assessment compared with results from a clinically established Alzheimer disease assessment scale and diagnosis of mild cognitive impairment [AUC = 0.82 (0.78–0.82)] was observed. **Conclusion:** This brief, cognitive performance measure, documenting good psychometric properties, can be useful in future epidemiological investigations exploring different domains and overall cognitive functioning in early old-age populations.

**Introduction**

With the rapid increase in ageing populations, a growing burden of mental disorders including cognitive impairment is observed. While a decline of cognitive functioning is commonly considered a normal consequence of ageing [1], a high interindividual variability in performance is evident, in particular with respect to speed of
processing and memory [2–4]. Population-based research on cognitive functioning has largely focused on old-age and very old-age groups, in response to the challenges provided by a growing number of people with severe cognitive impairment and related consequences for the provision of health care in terms of personnel and costs [5]. Yet, examining the distribution of cognitive performance in early old-age populations is an important task enabling the identification, and possibly prevention, of early signs of cognitive decline. Moreover, this approach provides new insights into the population-based values reflecting ‘normal’ cognitive ageing and the variability in cognitive performance according to socioeconomic, biomedical and genetic factors.

This task is not easy to perform, given the difficulties in assessing cognitive functioning in large epidemiologic investigations of people at early old age. The difficulties are due to lack of time needed for a comprehensive assessment by established neuropsychological tests [6, 7], low acceptance among healthy adults of undergoing cognitive testing, and high costs of recruiting and training experienced personnel. Ideally, a cognitive performance measure for epidemiologic studies should be easily applicable in a relatively short time, it should be well accepted by a healthy early old-age population and it should incorporate core dimensions of cognitive functioning. Up to now, such measures are rarely available as most respective instruments target populations at advanced old age [8–11]. For instance, in the Cardiovascular Determinants of Dementia (CASCADE) study the age range was 65–75 years [8], in the Memory and Morbidity in Augsburg Elderly (MEMO) study it was 65+ years [9], in the Leipzig Longitudinal Study of Ageing (LEILÀ75+) it was 75+ years [10], and in the Asset and Health Dynamics among the Oldest Old (AHEAD) study it was 65+ years [11].

Given the fact that only a small proportion of early old people suffer from severe cognitive impairment, it is important to focus on those with ‘normal’ or moderately impaired cognitive functioning. To assess respective variations, a sensitive – and yet easily applicable – short measure representing core cognitive domains is required. Therefore, to avoid ceiling effects, we set out to construct an appropriate multidimensional measure of cognitive performance for our relatively young study cohort (age range: 50–80 years). This paper aims at investigating the distribution of scores of a newly developed cognitive performance measure in a general unselected population, and at testing its concurrent validity with a clinically validated neuropsychological assessment.

### Methods

#### Study Sample

Data were collected during the second examination of the Heinz Nixdorf Recall (HNR) Study, an ongoing prospective population-based cohort study of a middle to early old-age urban population in Germany. The main aim of this study was to evaluate the predictive value of coronary artery calcification, using electron beam computed tomography, for myocardial infarction and cardiac death, in combination with new and established cardiovascular risk factors [12–15]. While incident coronary heart disease was the original study outcome, additional health measures were included, such as cognitive functioning.

The initial sample of the HNR study consisted of 4,814 men and women aged 45–74 years (response rate: 55.8%) recruited from statutory registers of 3 cities in Western Germany. The baseline examination was conducted between December 2000 and July 2003 [12], and participants were followed over a 5-year period. A detailed analysis of the non-responders has previously been published [14]. Briefly, there was no age difference between the participants and the non-participants, but elderly women were less likely to participate. Furthermore, a school degree at university entrance qualification level was more often reported among participants than among non-participants. The follow-up examination (n = 4,359; response rate: 90.1%) included a brief assessment of cognitive functioning, which was accomplished in 4,145 study participants (95.1%). In addition, a randomised subsample of the participants with low cognitive performance (n = 701) was recruited to undergo a neuropsychological and neurological examination (participation rate: 62.6%; n = 445). Data from this subsample were compared with data from a further subsample of participants with age-appropriate cognitive performance (participation rate: 66.6%; n = 211) [16]. Forty-one participants with dementia, severe depression [Alzheimer’s Disease Assessment Scale (ADAS) depression subscale score >4], Parkinson disease, mental retardation, severe alcohol consumption (for women: >20 g/day; for men: >40 g/day), known brain cancer, severe problems with the German language (foreign persons) and severe sensory impairment leading to invalid cognitive testing were excluded from the further analyses. Compared with the full sample of participants, the subsample (n = 615) was significantly older (full sample: 64.4 ± 7.6 years; subsample: 68.6 ± 6.9 years; p < 0.001) and had a lower education level (Pearson’s $\chi^2 = 192.478$; p < 0.001). The HNR Study followed established guidelines of good epidemiological practice and was repeatedly certified by an external board.

#### Measurements

#### Cognitive Functioning

As a part of the second examination, all participants underwent an abbreviated cognitive performance assessment, conceptualised as a multidimensional test, using established measures of immediate and delayed verbal memory, verbal fluency, speed of processing and visuospatial ability [17–20]. The choice of subtests to be included was based on the following criteria. Measures should (1) be sensitive to normal ageing, (2) be less vulnerable to the bias of educational attainment, (3) be characterised by sound psychometrical properties, and (4) be easily administered to a large study sample within a short period of time.
Composition of the Short Cognitive Performance Measure

Immediate and Delayed Verbal Memory. An 8-word list from the Nuremberg Geriatric Inventory (Nürnberg Alters-Inventar, NAI) [17] was applied. A first trial of the list was conducted to assess immediate recall, and a second trial was carried out at the end of the interview. Results were evaluated on a score ranging from 0 to 8, indicating the number of words correctly recalled [17].

Visuospatial Ability. The clock drawing test (CDT) [18] was administered, in which participants were asked to fix the time '10 min past 11' as precisely as possible on a predrawn circle representing a clock. Results were evaluated on a 5-point scale according to Shulman [18], with a range from 0 (perfect performance) to 5 (poor performance).

Speed of Processing/Executive Functioning. The labyrinth test from the NAI was applied to this aim, in which time spent was measured in seconds (range: 12–180 s) [17].

Verbal Fluency. The animal naming task – measuring verbal production and semantic memory – was applied, scoring the number of animal names mentioned within 1 min (range: 1–53) [20].

Low Cognitive Functioning. A summary index of cognitive functioning was defined, based on age-specific norms of cognitive subtests for the German general population. Subtest results were rated as 'impaired' if the performance was 1 SD below the age-adjusted mean [17]. Results of the abbreviated cognitive performance assessment were used to stratify participants into a group with impaired cognitive performance and a group with age-appropriate results. Impaired cognitive functioning was stated if participants scored low in at least 2 subtests (n = 1,125), whereas age-appropriate functioning required normal values in at least 4 of the 5 subtests administered (n = 3,020). Other studies have shown that combining subtests for several cognitive domains, as in our cognitive performance assessment, improved the predictive validity of this assessment [21].

Neuropsychological and Neurological Assessment

In the subsample of 656 participants, 445 with low and 211 with age-appropriate cognitive performance, a standardised neuropsychological assessment was conducted by a psychologist with expertise in dementia research. The assessment included the ADAS [19], the number connection test from the NAI [17] and the verbal fluency test [20] (subtests with a formal lexical category and 2 subtests with a semantic category) as well as instrumental activities of daily living (instrumental activities of daily living score to 1 SD above the age-adjusted mean for normal individuals) [27–29]. A cut-off of 1 SD was chosen because it was found to be associated with a higher relative prognostic power in predicting the development of dementia compared with a cut-off of 1.5 SD [30]. Borderline memory was defined as memory performance between 1 and 0.7 SD below the age-adjusted norm [31].

Education

Education was classified according to the International Standard Classification of Education [32] as total years of formal education, with the highest category of ≥18 years of education (equivalent to a university degree) and the lowest category of ≤10 years (equivalent to a basic school degree and no vocational training).

Statistical Analysis

Descriptive statistics were produced for basic study variables. Means and SD of test scores by age and education groups were calculated. As no considerable differences in cognitive tests by gender were observed, results are presented for the whole sample. Comparison of the variables for descriptive purposes was done using the t-test, and mean differences between extreme age and education groups were calculated.

To describe the underlying factor structure of the short cognitive performance measure, a theory-based approach [33] was taken, with an expected 3-factorial structure: (1) memory, (2) language and (3) visuconstructive abilities. Since all factors were expected to be correlated (as indicators of general cognitive ability), a non-orthogonal rotation was chosen [34].

The χ² test (ANOVA) was used for comparing cognitive test scores between the following 3 groups: (1) participants with MCI, (2) participants with borderline memory impairment, and (3) participants without any cognitive impairment (control group). Results are presented in online suppl. tables A1, A2, fig. A1, www.karger.com/doi/10.1159/000328262). To estimate the diagnostic accuracy of our brief cognitive assessment, receiver operating characteristic (ROC) curves were obtained by plotting sensitivity on the y-axis against (1 – specificity) on the x-axis. The area under the ROC curve (AUC) and associated 95% confidence intervals (CI) were calculated, with an AUC of 1.0 indicating a perfect test, and an AUC of 0.5 indicating unsatisfactory (chance) performance. In addition to the ROC curve analysis, 4 indicators were calculated to identify best cut-off scores for predicting MCI and borderline memory impairment: sensitivity, specificity, Youden index [35] and the point on the ROC curve closest to 0.1 [36] (online suppl. tables A1, A2; online suppl. fig. A1). The Youden index was calculated as: (sensitivity + specificity) – 1, with values close to 1 indicating that the test’s effectiveness was relatively high [35]. A point on the ROC curve closest to 0.1 indicates the minimum value of the square root of {[(1 – sensitivity)² + (1 – specificity)²]} [36]. Statistical analyses were provided using the statistical software PASW Statistic 17.0 and STATA version 10.0 (STATA Corp., College Station, Tex., USA).

Short Cognitive Performance Measure in HNR Study

Neuroepidemiology 2011;37:13–20
Results

Of the 4,359 individuals that participated in the follow-up examination, 4,145 (95.6%) had agreed to complete the short cognitive performance measurement, indicating the high acceptance of this instrument among participants. The mean duration of the measurement was 7.31 min. Of the 4,145 participants who underwent the short cognitive performance measurement, 524 men (25.6%) and 601 women (28.6%) showed a low performance in at least 2 subtests. 157 individuals of the subsample (23.9%) met the MCI modified criteria, and 162 (26.3%) met the criteria for borderline memory performance, with a relatively higher prevalence of MCI in older participants (p < 0.001) and among individuals with low education (p < 0.001). Based on this MCI prevalence in a subsample with a high proportion of subjects with poor cognitive performance (defined by the short cognitive performance measure), an overall MCI prevalence of 12.1% was estimated for the total sample (n = 4,145) [16].

| Table 1. Distributional characteristics of test scores by age and education groups |

<table>
<thead>
<tr>
<th></th>
<th>Age and education-specific distribution, years of education</th>
<th>p1</th>
<th>Difference between means2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤10</td>
<td>11–13</td>
<td>14–17</td>
</tr>
<tr>
<td>8-word list – immediate recall score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59 years</td>
<td>4.8</td>
<td>5.4</td>
<td>5.7</td>
</tr>
<tr>
<td>60–69 years</td>
<td>4.5</td>
<td>5.0</td>
<td>5.4</td>
</tr>
<tr>
<td>70–80 years</td>
<td>4.4</td>
<td>4.5</td>
<td>4.8</td>
</tr>
<tr>
<td>p1</td>
<td>0.113</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Difference between means3</td>
<td>0.40</td>
<td>0.92</td>
<td>0.94</td>
</tr>
<tr>
<td>8-word list – delayed recall score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59 years</td>
<td>3.0</td>
<td>3.7</td>
<td>4.0</td>
</tr>
<tr>
<td>60–69 years</td>
<td>2.3</td>
<td>3.0</td>
<td>3.5</td>
</tr>
<tr>
<td>70–80 years</td>
<td>2.4</td>
<td>2.5</td>
<td>2.8</td>
</tr>
<tr>
<td>p1</td>
<td>0.018</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Difference between means3</td>
<td>0.64</td>
<td>1.22</td>
<td>1.25</td>
</tr>
<tr>
<td>Labyrinth test score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59 years</td>
<td>57.7</td>
<td>47.6</td>
<td>39.4</td>
</tr>
<tr>
<td>60–69 years</td>
<td>72.2</td>
<td>60.9</td>
<td>48.4</td>
</tr>
<tr>
<td>70–80 years</td>
<td>83.5</td>
<td>75.5</td>
<td>64.3</td>
</tr>
<tr>
<td>p1</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Animal naming task score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59 years</td>
<td>20.9</td>
<td>23.4</td>
<td>24.9</td>
</tr>
<tr>
<td>60–69 years</td>
<td>18.9</td>
<td>21.8</td>
<td>23.6</td>
</tr>
<tr>
<td>70–80 years</td>
<td>19.3</td>
<td>19.9</td>
<td>20.9</td>
</tr>
<tr>
<td>p1</td>
<td>0.052</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Difference between means3</td>
<td>1.65</td>
<td>3.37</td>
<td>3.91</td>
</tr>
<tr>
<td>CDT score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–59 years</td>
<td>1.9</td>
<td>1.48</td>
<td>1.3</td>
</tr>
<tr>
<td>60–69 years</td>
<td>2.0</td>
<td>1.7</td>
<td>1.5</td>
</tr>
<tr>
<td>70–80 years</td>
<td>2.0</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>p1</td>
<td>0.489</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Difference between means3</td>
<td>0.15</td>
<td>0.40</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Values denote means ± SD unless specified otherwise. Figures in parentheses are 95% CI.
1 Significance of the trend.
2 Difference in means between contrast education groups (≤10 vs. ≥18 years of education).
3 Difference in means between contrast age groups (50–59 vs. 70–80 years).
The means and SD of each subtest are presented by age and education for the total sample (table 1). Younger age and higher education are associated with better cognitive performance. Comparing subjects with low education with those with high education, the age effect on cognitive performance was found to be statistically significant in all educational groups, with the exception of immediate recall and the CDT in the low-education group (≤10 years of education). The differences in cognitive performance between the extreme age groups (50–59 vs. 70–80 years) were less pronounced among participants with a low educational level. Education appeared to have a more significant impact on cognitive performance among younger participants. In the oldest age group (70–80 years), the differences in education were less considerable when comparing the extreme education groups (≤10 vs. ≥18 years of education).

The factor analysis revealed 2 underlying factors of cognitive performance: (1) a ‘memory and language’ factor, and (2) a ‘visuconstruction’ factor. Factor 1 explains 71.85%, and factor 2 explains 28.15% of the total variance observed. A supposed third factor, ‘language’, as measured by the ‘animal naming task’, did not emerge separately from the first factor. The 2 factors are moderately correlated (r = 0.41).

The ROC analysis, provided for the subsample with comprehensive neuropsychological examination, showed a good accuracy of the short cognitive performance measure in identifying participants with poor cognitive performance according to the more comprehensive ADAS-cog battery as well as individuals with MCI (fig. 1).

Discussion

This paper presents age- and education-specific values for different cognitive domains in unselected early old-age men and women in Germany. Moreover, the concurrent validity of the short cognitive performance instrument was tested with regard to the more comprehensive cognitive assessment. Values were calculated for the whole study sample, excluding the few subjects with diagnosed dementia. These cognitive performance norms provide important new knowledge as they are based on a large, unselected representative population of early old men and women in Germany [12]. We therefore refer to...

Fig. 1. ROC curves. Accuracy of cognitive performance scores based on 5 and 3 subtests (after exclusion of labyrinth test and CDT) in detecting poor cognitive performance based on the definition of MCI (a) and based on the ADAS-cog battery (b). Values in parentheses denote 95% CI.
these values as a ‘norm’ in the line of Feinstein’s definition of ‘range of normal’ [37]. The strong relationships of cognitive functioning and cognitive impairment with age and educational attainment are consistent with those found in previous studies [38, 39].

It is of interest to note that in the preselected subsample, the education-based cut-off points for cognitive performance did not improve age-based concurrent validity. This observation may explain the surprisingly high prevalence of MCI diagnoses among people with low education in our representative study population. It may also reflect the fact that previously published norms for MCI diagnosis [17–20] were probably based on samples where poorly educated participants were not adequately represented, or where age-specific norms were not additionally stratified according to education.

The factor analysis demonstrated that our cognitive assessment instrument represents two distinct dimensions: (1) memory and language, and (2) visuospatial abilities. Notably, the first factor with an eigenvalue of >1 was much stronger than the second factor. In view of the moderate correlation, results point to a general factor underlying cognitive functioning.

By combining data from our short cognitive performance assessment with in-depth data from a comprehensive neuropsychological examination, we were able to assess the concurrent validity of different subtests of cognitive performance. Good accuracy of the short cognitive performance measure was observed by separating subjects with poor cognitive performance according to the ADAS-cog or MCI criteria from subjects with age-appropriate functioning. When comparing the psychometric properties of the different cognitive domains, the assessed memory tests and verbal fluency task showed better diagnostic accuracy than the labyrinth test and the CDT.

Our findings are in line with results from clinical studies using neuropsychological tests for MCI and dementia. Verbal fluency deficits are a common finding in patients with Alzheimer’s disease [40]. Moreover, verbal fluency was found to be significantly reduced in amnestic MCI individuals relative to cognitively intact older adults [41, 42]. Word list recall has been shown to be sensitive to MCI and to early onset of cognitive impairment in older adults [43, 44]. The CDT and labyrinth test were found to have a poor accuracy regarding the identification of MCI and borderline memory impairment cases in our study sample. Several studies have evaluated the role of the CDT as a diagnostic tool for clinically relevant cognitive impairment [45–47]. In a review article, Shulman [18] reported that most studies achieved sensitivity and specificity values around 85%, suggesting that the CDT could be useful for early detection of dementia. In contrast, Nishiwaki et al. [47], studying the validity of the CDT in a large general population sample of persons aged 75+ years, reported a relatively poor diagnostic accuracy of the CDT in detecting milder cognitive impairment, suggesting that the CDT was better in detecting moderate/severe cognitive impairment. In a similar vein, Powishta et al. [45] concluded that the CDT did not appear to be a useful screening tool for detecting very mild dementia.

This study has several limitations. First, given the age limit of our sample, we are not able to extend the results to participants younger than 50 or older than 80 years. In view of documented age-related differences in cognitive decline before the age of 50 years [48], this limitation seems critical. Yet, as the main focus of our study was to investigate the variability in different domains of cognitive performance, this age range was appropriate to achieve our research aim. A further limitation concerns the fact that data collection on cognitive performance was restricted to one point of measurement. Thus, we have no information about change in cognitive performance over time. Moreover, previously reported age-specific ADAS-cog norms for memory performance have been used for large age groups and they are not fully appropriate to our relatively young study sample as they exclude groups younger than 55 years. Finally, it should be mentioned that in our study sample, the effect of age on cognitive performance might have been overestimated due to a higher proportion of poorly educated participants in the older age group.

These limitations are balanced by several strengths. First, and importantly, our findings are based on an unselected, large population of men and women at early old age participating in a longitudinal epidemiological investigation. This fact has enabled us to describe distributional characteristics of scores reflecting different dimensions of cognitive performance in this age group. Second, we confirmed the important role of educational degree in determining the level of cognitive performance in early old age across all age subgroups, and similarly among men and women. Third, by combining multidimensional-scale cognitive data with in-depth neuropsychological and neurological examination, the concurrent validity of the short cognitive assessment was investigated. In conclusion, this brief assessment of cognitive performance in early old age proves to be an economic, well accepted and valid tool for assessing relevant cognitive domains in large epidemiological studies.
Acknowledgement

We thank the Heinz Nixdorf Foundation for their generous support of the HNR Study, which was additionally supported by the German Federal Ministry of Education and Research (BMBF). The specific part of cognitive testing was funded by the German Research Council (DFG SI236/10-1). For the MCI substudy, we received a research grant from the Dr. Werner Jackstädt-Stiftung. We would like to give special thanks to Pablo Emilio Verde from the Coordination Centre for Clinical Trials, University of Düsseldorf, for his statistical advice. We are indebted to the investigative group, the study personnel and all study participants.

References


31 Storandt M, Grant EA, Miller JP, Morris JC: Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. Neurology 2006;67:467–473.


