CHAPTER 2

Recommended measures for the assessment of cognitive and physical performance in older patients with dementia: a systematic review.¹

Willem Bossers, Lucas van der Woude, Froukje Boersma, Erik Scherder, Marieke van Heuvelen

¹See also: Dementia and Geriatric Cognitive Disorders EXTRA 2012 Jan,2(I):589-609 doi 10.1159/000345038
ABSTRACT

Aim
Recommend a set of neuropsychological and physical exercise tests for researchers to assess cognitive and motor function in clinical trials with older patients with dementia; to create consensus, decrease heterogeneity, and improve research quality.

Methods
A literature search (2005 - 2011) yielded 89 randomized controlled trials (RCTs). To provide information for test recommendations, the frequency of test use, effect size of the test outcome, study quality, and clinimetric properties of tests were analyzed.

Results
Fifty-nine neuropsychological tests (cognitive function domains: global cognition, executive functioning, memory, and attention) and ten physical exercise tests (motor function domains: endurance capacity, muscle strength, balance, and mobility) were found.

Conclusion/recommendations
To measure global cognition, the Mini Mental State Examination, Alzheimer Disease Assessment Scale-cognitive subscale, and the Severe Impairment Battery were recommended. To measure executive functioning, the Verbal Fluency Test Category/Letters, Clock Drawing Test, and Trail Making Test-B were recommended. No specific memory test could be recommended. The Digit Span Forward, Digit Span Backward, and Trail Making Test-A were recommended to measure attention. The Timed Up and Go test and Six Meter Walk test for mobility, the Six Minute Walk Distance for endurance capacity, and the Tinetti Balance Scale for balance were recommended as physical exercise tests.

INTRODUCTION

With the worldwide trend of an aging population, the number of patients with dementia will increase dramatically in the coming decades. Dementia is characterized by a loss of neurons and atrophy of brain tissue. Eventually, this leads to limitations in cognitive performance, including a decline in executive functioning, memory, and attention. The neurodegenerative processes in the brain go hand in hand with a decline in motor function, including a loss of endurance capacity, muscle strength, balance, and mobility. Eventually, decline in cognitive and motor function result in problems in activities of daily living, leading to institutionalization and a decreased quality of life. Therefore, prevention of decline, and preferably an improvement in both cognitive and motor performance, in patients with dementia is of utmost importance.

With the growing impact of dementia on today’s society, new treatments need to be developed that effectively reduce the limitations that are caused by a decline of cognitive and motor performance. Meta-analysis and systematic reviews reported that pharmacological (e.g., medication) and non-pharmacological (e.g., exercise) interventions in patients with dementia may have a positive effect on cognitive and/or motor function. However, the individual studies in these reviews showed ambiguous results and the tests that measured cognitive and motor function appeared to show large heterogeneity. Consequently, the comparability of the outcomes of clinical trials is hampered. Therefore, future intervention studies that aim to improve cognitive and/or motor function should strive to use a limited number of generally accepted, feasible, reliable, and valid tests that adequately cover the domains of cognitive and motor function in patients with dementia.

Recommendations on cognitive tests for the purpose of diagnosing dementia were recently provided. However, information regarding which cognitive tests to use in measuring treatment effects are not available yet. To the author’s knowledge, recommendations regarding the use of physical exercise tests for evaluating motor function are fully lacking.

The aim of this systematic review was to give up-to-date recommendations of neuropsychological and physical exercise tests to investigate treatment effects on cognitive and motor function in older patients with dementia. The current search was based on recent randomized controlled trials (RCTs) to give a comprehensive overview of neuropsychological and physical exercise tests. Frequently used neuropsychological and physical exercise tests were evaluated in relation with the study quality of the RCTs in which the tests were used, the nature of the interventions, the type of dementia of the included participants, and the sensitivity to change of the tests. Furthermore, the clinimetric properties (i.e., reliability and validity) of frequently used tests were reviewed.
Chapter 2

Recommended measures for the assessment of cognitive and physical performance: a systematic review

METHODS

Data sources
Between August 2010 and August 2011, computer databases Pubmed, EMBASE, Biological abstracts, Web of Science (ISI), PsycINFO, CINAHL, and Cochrane Library were searched for relevant studies published between 2005 and 2011. Limits for the searches in the computer databases were set on: clinical trial, humans, and age: 65+ years. Keywords in the search included terms from Medical Subject Headings (MESH) and Embase thesaurus (EMTREE). The following terms were used in the MESH database and EMTREE thesaurus: dementia, Alzheimer Disease, Vascular Dementia, Frontotemporal Dementia, Lewy Body Disease, neuropsychological tests, and exercise tests. Keywords for dementia (Dementia OR Alzheimer Disease OR Vascular Dementia OR Frontotemporal Dementia OR Lewy Body Disease) were combined (with 'AND') with terms that expressed the use of neuropsychological or exercise tests (neuropsychological tests OR exercise tests). In addition, reference lists of reviews regarding the subject were hand searched for additional studies.

Inclusion criteria
Studies were included if they met the following criteria: RCT design, participants were diagnosed with dementia, participants were on average older than 65 years, neuropsychological tests and/or exercise tests were used to measure intervention effects, and the study was written in English, German, French, or Dutch.

Selection process
After the literature search, a first selection of studies was made on title, followed by a selection after reading the abstracts. Two reviewers (WB and MvH) independently performed both steps to identify those studies that met the inclusion criteria (agreement 94%, disagreement 6%). Disagreement was solved with full text screening. Full text analysis, to check the inclusion criteria, was performed for the studies that were identified in the preceding steps. Subsequently, reviews were hand searched for clinical trials that were not already found in the literature search. Finally, full text analysis and data extraction was performed from the selected studies.

Data extraction
From the selected RCTs the following data were extracted: neuropsychological tests, physical exercise tests, type of dementia, sample size, and data regarding the intervention description (e.g., pharmacological, exercise). For each neuropsychological or physical exercise test the overall means and standard deviations were calculated from all RCTs that used a given test. Based on the selected RCTs the overall means and standard deviations were calculated for age, baseline scores, and posttest scores.

Effect size
In order to express the sensitivity to change for each neuropsychological or physical exercise test, Cohen’s $d$ effect sizes per test were calculated over the selected RCTs. If the mean and standard deviation of pretest and posttest were presented in the RCT, the following formula was used:

$$d = \frac{(\text{post} - \text{pre})}{\sqrt{\text{S}^2_{\text{pre}} + \text{S}^2_{\text{cont}}}}$$

If the means and standard deviations were not presented in the RCT, the $F$ statistic was used with the following formula:

$$d = \sqrt{\text{S}^2_{\text{pre}} + \text{S}^2_{\text{cont}}} \cdot \frac{\text{N}_{\text{exp}}}{\text{N}_{\text{exp}} + \text{N}_{\text{cont}}}$$

The overall effect size $d$ was calculated as the mean of individual effect sizes weighted for the sample size. Cohen’s $d$ benchmarks were used to indicate small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$) effect sizes.

Study quality
Study quality of each RCT that used a given test was assessed with the Physiotherapy Evidence Database (PEDro). According to the PEDro scoring system a score of 9 - 10 was considered as excellent, a score of 6 - 8 as good, a score of 4 - 5 as moderate, and a score of 0 - 3 as poor. Further clinometric analysis of neuropsychological and physical exercise tests was done when the study quality of at least five RCTs were good or excellent.

Reliability and validity of frequently used tests
After identifying the tests that were used in five or more good or excellent quality RCTs, a second search in Pubmed was conducted in September 2011 to select the studies aimed at reporting the reliability and validity of these tests as evidenced in a population with dementia. Searches were performed by combining the terms ‘reliability’ OR ‘validity’ OR ‘reproducibility of results’ and (with AND) keywords for dementia and the selected neuropsychological and physical exercise tests. By means of references, additional reliability and validity studies were searched.
RESULTS

Study characteristics
The literature search for RCTs yielded a total of 840 publications. Eventually, 178 publications were full text screened, of which 89 were included for further analysis. A flowchart of the process is presented in Figure 2.1. The results are described in two separate sections: neuropsychological tests and exercise tests. These sections describe the test use in RCTs (no. of RCTs), test use related to intervention type, test use related to dementia type, effect sizes, and study quality of RCTs. Table 2.1 describes 59 neuropsychological tests that covered the cognitive domains global cognition, executive functioning, memory, and attention. Thereafter, Table 2.2 describes the clinimetric data of the neuropsychological tests that were most often used. Finally, ten exercise tests that covered the motor function domains endurance capacity, muscle strength, balance, and mobility are presented in Table 2.3.

Number of publications from literature search:

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>407</td>
</tr>
<tr>
<td>EMBASE</td>
<td>301</td>
</tr>
<tr>
<td>Biological abstracts</td>
<td>23</td>
</tr>
<tr>
<td>Web of Science</td>
<td>28</td>
</tr>
<tr>
<td>PsycINFO</td>
<td>20</td>
</tr>
<tr>
<td>CINAHL</td>
<td>101</td>
</tr>
<tr>
<td>Cochrane</td>
<td>570</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1350</strong></td>
</tr>
</tbody>
</table>

Removing doubles (-510)
Excluded after screening title and abstract (-663)
Hand search of 4 reviews (+5)

178 full-text publications screened

Excluded after full-text screening (-89)
No intervention study (30), no tests used (27), no randomized controlled trial (19), no validated dementia diagnosis (7), mean age c60 (5), commentary (1)

89 studies included after full-text screening

59 neuropsychological tests and 10 motor function tests

45 clinimetric studies

Figure 2.1. Flowchart of literature search and study selection.

Neuropsychological tests

Frequency of test use
As shown in table 2.1, global cognitive functioning was measured most often with the Mini Mental State Examination (MMSE) (54 times), Alzheimer’s Disease Assessment Scale-cognitive subscale (ADAS-cog) (43 times), and the Severe Impairment Battery (SIB) (8 times). Tests for global cognition were used more than neuropsychological tests that covered a specific cognitive area.

Thirty-two domain specific neuropsychological tests were used in 63 RCTs, of which 7 tests were used in 5 or more RCTs. Executive functioning was measured with the Verbal Fluency Test Category (12 times), Clock Drawing Test (6 times), Verbal Fluency Test Letters (6 times), and the Trail Making Test-B (5 times). Attention was measured with the Digit Span Forward (8 times), Digit Span Backward (7 times), and Trail Making Test-A (6 times).

In sum, global cognitive tests were used more often than neuropsychological tests that covered a specific cognitive area. Frequently used neuropsychological tests that were used in five or more RCTs covered the cognitive domains executive functioning and attention. Memory tests that were used in more than five RCTs were not found.
Table 2.1: Descriptive data of 63 randomized clinical trials that used 59 different neuropsychological tests.

<table>
<thead>
<tr>
<th>Neuropsychological test</th>
<th>No. of studies</th>
<th>N</th>
<th>Intervention type (No. of RCTs)</th>
<th>Age Mean (SD)</th>
<th>Gender (%)</th>
<th>Dementia type Mean (SD)</th>
<th>Baseline Mean (SD)</th>
<th>ES</th>
<th>PDBro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini Mental State Examination</td>
<td>32</td>
<td>1290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer’s Disease Assessment Scale –</td>
<td>22</td>
<td>666</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cognitive subscale(s)^1^, 2^, 3^, 4^, 5^, 6^, 7^, 8^, 9^, 10^</td>
<td>43</td>
<td>10133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Fluency Test Category^1^, 2^, 3^, 4^, 5^, 6^, 7^, 8^, 9^, 10^</td>
<td>12</td>
<td>726</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span Forward^3^, 4^, 5^, 6^, 7^, 8^</td>
<td>8</td>
<td>342</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span Backward^3^, 4^, 5^, 6^, 7^, 8^</td>
<td>7</td>
<td>280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gock Drawing Test^1^, 2^, 3^, 4^, 5^, 6^, 7^</td>
<td>6</td>
<td>1674</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making Test^1^, 2^, 3^, 4^, 5^, 6^, 7^</td>
<td>6</td>
<td>1502</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Fluency Test Letter^1^, 2^, 3^, 4^, 5^, 6^, 7^</td>
<td>5</td>
<td>319</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logical Memory Test – immediate recall^6^, 7^, 8^, 9^, 10^</td>
<td>4</td>
<td>277</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logical Memory Test – delayed recall^6^, 7^, 8^, 9^, 10^</td>
<td>4</td>
<td>277</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eight Word Test – immediate recall^10^, 3^</td>
<td>4</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eight Word Test – delayed recall^10^, 3^</td>
<td>4</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBMT – face recognition^10^, 3^, 4^, 5^, 6^</td>
<td>3</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBMT – picture recognition^10^, 3^, 4^, 5^, 6^</td>
<td>3</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mattis Dementia Rating Scale^1^, 2^, 3^, 4^</td>
<td>3</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Boston Naming Test^1^, 2^, 3^</td>
<td>3</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syndrome Kurtz Test^1^, 2^, 3^, 4^, 5^, 6^</td>
<td>2</td>
<td>290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Symbol Test^1^, 2^, 3^, 4^, 5^, 6^, 7^</td>
<td>2</td>
<td>279</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop Interference Test^3^, 4^, 5^</td>
<td>2</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambridge Neuropsychological Test Battery^2^, 3^</td>
<td>2</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Memory Span^1^, 2^, 3^, 4^, 5^, 6^</td>
<td>2</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective Reminding Test^1^, 2^, 3^, 4^</td>
<td>2</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dementia type
Table 2.1 shows that a majority of the participants were diagnosed with Alzheimer’s disease (AD) (84%) or vascular dementia (VaD) (7%). Neuropsychological tests that were used only in RCTs with AD patients were the SIB (global cognitive functioning), Verbal Fluency Test Letters (executive functioning), Trail Making Test-A, Digit Span Forward, and Digit Span Backward (attention). Tests used in RCTs with AD or VaD patients were the MMSE and ADAS-cog (global cognitive functioning), Verbal Fluency Test Category, Clock Drawing Test, and Trail Making Test-B (executive functioning). Tests that were used in RCTs with only VaD, Lewy Body Disease, Pick’s Disease, and frontotemporal dementia patients were not found.

Effect size
Pooled effect sizes ranged from small ($d = -0.16$) to large ($d = 1.58$). The global cognitive test Rapid Evaluation of Cognitive Functioning measured a large effect size ($d = 1.12$). The global cognitive tests SIB ($d = 0.34$), ADAS-cog ($d = 0.19$) and MMSE ($d = 0.09$) showed small effect sizes. Overall effect sizes were small for both pharmacological and non-pharmacological RCTs. Furthermore, two neuropsychological tests that measured memory revealed large pooled effect sizes with the Visual Reproduction test ($d = 1.58$) and the Syndrome Kurtz Test ($d = 1.12$). The global cognitive test Rapid Evaluation of Cognitive Functioning measured a large effect size ($d = 1.58$).

Study quality
According to the PEDro scale, the study quality of RCTs that used neuropsychological tests ranged from 2 (poor) to 10 (excellent). Three RCTs with poor study quality used the Verbal Fluency Test Category, Digit Span Forward, Digit Span Backward, Clock Drawing Test, Trail Making Test-A, and Verbal Fluency Test Letters. Because these tests were also found in RCTs with excellent and good study quality, this had no effect on the selection process of these neuropsychological tests.

Reliability and validity
Table 2.2 presents clinimetric data (i.e., reliability and validity) of ten neuropsychological tests that were used in five or more high-quality randomized controlled trials.

<table>
<thead>
<tr>
<th>Neuropsychological Test</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Fluency Test Category</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>Clock Drawing Test</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>Trail Making Test-B</td>
<td>0.4 – 0.5</td>
<td>0.3 – 0.4</td>
</tr>
<tr>
<td>Digit Span Forward</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>Digit Span Backward</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>MMSE</td>
<td>0.7 – 0.8</td>
<td>0.6 – 0.7</td>
</tr>
<tr>
<td>ADAS-cog</td>
<td>0.6 – 0.7</td>
<td>0.5 – 0.6</td>
</tr>
<tr>
<td>Syndrome Kurtz Test</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>Visual Reproduction Test</td>
<td>0.6 – 0.7</td>
<td>0.5 – 0.6</td>
</tr>
<tr>
<td>National Hospital Examination</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
<tr>
<td>Verbal Fluency Test Letters</td>
<td>0.5 – 0.6</td>
<td>0.4 – 0.5</td>
</tr>
</tbody>
</table>

Note: DC, Clinimetric Correlation; MMSE, Mini-Mental State Examination; ADAS-cog, Alzheimer’s Disease Assessment Scale—Cognitive.
Exercise tests

Frequency of test use
Ten different exercise tests were used in 13 RCTs (Table 2.3). These tests measured the motor function domains endurance capacity with the Six Minute Walk Distance and Two Minute Step Test; muscle strength with the Five Times Sit To Stand and 30 Second Chair Stand; balance with the Tinetti Balance Scale, Abnormal One-Leg Balance, and Berg Balance Scale; mobility with the Timed Up and Go and Six Meter Walk; and flexibility with the Functional Reach Test. All physical exercise tests were used in non-pharmacological RCTs, except for the Tinetti Balance Scale, which was also used in one pharmacological RCT.

Dementia type
A majority of the participants was diagnosed with AD (84%) or VaD (6%). Six exercise tests were used only in AD patients, and covered the motor function domains endurance capacity (Two Minute Step Test), muscle strength (30 Second Chair Stand), balance (Tinetti Balance Scale, Abnormal One-Leg Balance), and mobility (Timed Up and Go, Six Meter Walk). In RCTs that included both AD and VaD, three physical exercise tests measured the motor function domains endurance capacity (Six Minute Walk Distance), flexibility (Functional Reach Test), and balance (Berg Balance Scale).

Study quality
The study quality of RCTs ranged from 5 (moderate) to 9 (excellent). Only the Tinetti Balance Scale was used in a RCT with excellent study quality (PEDro 9).

Effect size
Pooled effect sizes of RCTs ranged from small ($d = 0.02$) to large ($d = 0.87$). A large effect was found with the Tinetti Balance Scale ($d = 0.87$). Medium effect sizes were found with the Six Meter Walk test ($d = 0.58$) and the Six Minute Walk Distance test ($d = 0.51$).

Reliability and validity
The Timed Up and Go test (Intra Class Correlation (ICC) = .985 - .988), Six Minute Walk Distance (ICC = .982 - .987), and Six Meter Walk test (ICC = .973 - .977) showed excellent test retest values in patients with dementia.\textsuperscript{53} No further clinimetric data were available in patients with dementia for the remaining physical exercise tests that were presented in Table 2.3.

### Table 2.3. Descriptive data of 13 randomized clinical trials that used 10 different motor function tests.

<table>
<thead>
<tr>
<th>Physical exercise test</th>
<th>No. of studies</th>
<th>N</th>
<th>Intervention type (No. of RCTs)</th>
<th>Age $\pm$ (SD)</th>
<th>Gender (%)</th>
<th>Dementia type (%)</th>
<th>Baseline $\pm$ (SD)</th>
<th>ES $\pm$ PEDro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timed Up &amp; Go\textsuperscript{53, 90, 103}</td>
<td>3</td>
<td>179</td>
<td>Exers(5)</td>
<td>81.9 (7.3)</td>
<td>71%</td>
<td>AD(100)</td>
<td>17.1 (7.5)</td>
<td>small</td>
</tr>
<tr>
<td>Six Minute Walk Test\textsuperscript{37}</td>
<td>2</td>
<td>105</td>
<td>Exers(2)</td>
<td>77.6 (6.6)</td>
<td>65%</td>
<td>VaD(164) LB (16)</td>
<td>221.0 (82.8)</td>
<td>medium</td>
</tr>
<tr>
<td>Functional Reach Test\textsuperscript{34, 35}</td>
<td>2</td>
<td>94</td>
<td>Exers(2)</td>
<td>76.6 (6.6)</td>
<td>52%</td>
<td>VaD(188)</td>
<td>20.4 (8.1)</td>
<td>small</td>
</tr>
<tr>
<td>Six Meter Walk Test\textsuperscript{31}</td>
<td>1</td>
<td>134</td>
<td>Exers(1)</td>
<td>83.0 (7.4)</td>
<td>75%</td>
<td>AD(100)</td>
<td>0.4 (0.2)</td>
<td>medium</td>
</tr>
<tr>
<td>One-leg Balance Test\textsuperscript{31}</td>
<td>1</td>
<td>134</td>
<td>Exers(1)</td>
<td>83.0 (7.4)</td>
<td>75%</td>
<td>AD(100)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tinetti Balance Scale\textsuperscript{72}</td>
<td>1</td>
<td>116</td>
<td>Pharm(1)</td>
<td>75.4 (2.5)</td>
<td>62%</td>
<td>AD(100)</td>
<td>8.5 (2.5)</td>
<td>large</td>
</tr>
<tr>
<td>Five Times Sit to Stand\textsuperscript{30}</td>
<td>1</td>
<td>29</td>
<td>Exers(1)</td>
<td>76.9 (6.7)</td>
<td>51%</td>
<td>AD(10) LB(28)</td>
<td>18.9 (7.2)</td>
<td>small</td>
</tr>
<tr>
<td>Berg Balance Scale\textsuperscript{77}</td>
<td>1</td>
<td>85</td>
<td>Exers(1)</td>
<td>76.6 (6.5)</td>
<td>52%</td>
<td>VaD(220) LB (20)</td>
<td>47.5 (16.9)</td>
<td>small</td>
</tr>
<tr>
<td>30 Second Chair Stand\textsuperscript{52}</td>
<td>1</td>
<td>16</td>
<td>Exers(1)</td>
<td>74.5 (-)</td>
<td>37%</td>
<td>AD(100)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two Minute Step Test\textsuperscript{52}</td>
<td>1</td>
<td>16</td>
<td>Exers(1)</td>
<td>74.5 (-)</td>
<td>37%</td>
<td>AD(100)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. $f$, Weighted data as a function of the number of participants; ES, effect size; PEDro, Physiotherapy Evidence Database; E, excellent (9-10); G, Good (6-8); M, Moderate (4-5); P, Poor (0-3); Pharm., Pharmacological; Exers., Exercise; AD, Alzheimer’s Disease; VaD, Vascular Dementia; LB, Lewy Body Disease; n.r., not reported.
DISCUSSION

To improve the study quality and increase comparability of clinical trials and observational studies, researchers should strive to use a limited number of generally accepted, feasible, reliable, and valid tests that cover the domains of cognitive and motor function. Following previous studies that recommended neuropsychological tests for the diagnoses of dementia\(^{20, 21}\) and studies that stated the importance of physical exercise to attenuate cognitive impairment in older patients with dementia\(^{146}\), the aim of the current review was to give up-to-date recommendations of both neuropsychological and physical exercise tests for high quality experimental research in older patients with dementia.

Neuropsychological tests

This study revealed 59 different neuropsychological tests that were used in 63 RCTs. This confirms the assumption that there is a large heterogeneity in neuropsychological test use in RCTs with older patients with dementia. The results showed that global cognitive tests were used more often in comparison with neuropsychological tests that measured one specific cognitive domain.

In particular, the global cognitive tests MMSE, ADAS-cog and the SIB were standing out from other global cognitive tests because of their excellent clinimetric properties (Table 2.2). Furthermore, these tests were frequently used in high quality RCTs (i.e., respectively 54, 43, and 8 times), which suggest that they are feasible to use in patients with dementia. However, for all three tests the sensitivity to change was low. The low sensitivity to change of the MMSE and ADAS-cog was previously challenged in other studies because changes in test scores may be caused by small measurement errors.\(^{125, 140}\) For the SIB however, previous research suggest that this test may be sensitive to change in patients with severe dementia.\(^{110}\) Based on the above, we recommend to use the SIB in RCTs to measure global cognitive treatment effects. However, when comparing effects with previous studies in patients with dementia, the MMSE and ADAS-cog are suggested.

Memory tests could not be selected in this review due to the large heterogeneity in memory test use. Earlier work on memory tests for diagnosing dementia showed that verbal memory and visual memory can be assessed with several neuropsychological tests.\(^{20}\) Most of these proposed neuropsychological tests were only used once in RCTs between 2005 and 2011. Moreover, the proposed ‘Word List’ of the Consortium to Establish a Registry for Alzheimer’s disease was not used at all in RCTs over that period. Additionally, because these tests were specifically recommended for diagnosis of dementia, we suggest that they are not suitable to measure intervention effects over time. Furthermore, studies that investigated the clinimetric properties of these memory tests in patients with dementia are lacking.

To measure executive function, we recommend using the Verbal Fluency Test Category, Clock Drawing Test, Verbal Fluency Test Letters, and the Trail Making Test-B because these tests were frequently used in high quality RCTs. However, we only found that the Verbal Fluency Test Category was able to detect change and only the Clock Drawing Test was found reliable in older dementia patients. Since information and clinimetric quality is in many cases still insufficient, the recommended selection should be used with care and further evaluation of these tests is needed.

For the cognitive domain attention, we recommend the Digit Span Forward, Digit Span Backward, and Trail making Test-A because of their frequent use in high quality studies. However, no studies were found that investigated the clinimetric properties of these tests in patients with dementia. Furthermore, the results showed that the sensitivity to change was small.

| Table 2.4. Recommended neuropsychological tests ordered on the basis of frequency of test use, overall effect size, study quality, reliability, and validity for the cognitive domains global cognitive functioning, executive functioning, memory, and attention. |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| **Global cognitive functioning** | **Executive functioning** | **Memory** | **Attention** |
| Severe Impairment Battery\(^*\)** | Verbal Fluency Test Category\(^**\)** | Visual Reproduction Test\(^**\)** | Digit Span Forward\(^**\)** |
| Mini Mental State Examination\(^*\)** | Clock Drawing Test\(^*\)** | Eight Word Test\(^*\)** | Digit Span Backward\(^**\)** |
| Alzheimer’s Disease Assessment Scale – cognitive subscale\(^*\)** | Verbal Fluency Test Letters\(^*\)** | Logical Memory Test\(^*\)** | Trail Making Test-A\(^**\)** |
| Rapid Evaluation of Cognitive Functioning\(^*\)** | Trail Making Test-B\(^**\)** |

*Note: \(^*\) Frequently used in RCTs (feasibility); \(^\dagger\) Able to measure an effect (sensitivity to change); \(^\ddagger\) Test was used in excellent / good quality RCTs (PEDro); \(^\ddagger\) Reliable / valid in dementia patients; \(^\dagger\dagger\) more research is needed.

Based on the current results, Table 2.4 sums up the currently best available tests used in international intervention studies with older patients with dementia. Although widely applied, it was shown that the recommended neuropsychological tests lack clinimetric data. Therefore, future research into the clinimetric quality for these tests is essential. The recommended selection of currently optimal cognitive tests should be used with care. Researchers are advised to select those recommended tests that most closely fit into their study objectives.
Physical exercise tests
The current review presented ten different exercise tests, which covered the motor function domains endurance capacity, muscle strength, balance, and mobility. However, a large heterogeneity in tests used was found, and none of the tests were used frequently enough in RCTs to recommend them. Preliminary recommendations based on the results of this review may be a first step for the selection of exercise tests in future studies.

The results showed that the Six Minute Walk Distance test, a test for endurance capacity, is reliable and sensitive to change. Muscle strength was measured with the Five Times Sit to Stand test and the 30 Second Chair Stand test. However, no data was available on the feasibility and clinimetric properties of these tests. For the domain balance, results showed that the Tinetti Balance Scale was sensitive to measure change but again, no data was available on feasibility and clinimetric properties in patients with dementia. Mobility was measured with the Six Meter Walk test and the Timed Up and Go test, of which both were reliable in patients with dementia. According to the present data, only the Six Meter Walk test was sensitive to measure change. Based on the limited information at hand, the best exercise tests yet available are summed up in Table 2.5.

<table>
<thead>
<tr>
<th>Endurance capacity</th>
<th>Muscle strength</th>
<th>Balance</th>
<th>Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six Minute Walk Distance*</td>
<td>Five Times Sit to Stand*</td>
<td>Tinetti Balance Scale*</td>
<td>Timed Up and Go*</td>
</tr>
<tr>
<td>30 Second Chair Stand*</td>
<td>Six Meter Walk*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1, Able to measure an effect (sensitivity to change); 2, Test was used in excellent / good quality RCTs (PEDro); 3, Reliable / valid in dementia patients; 4, more research is needed.

Because of the importance of motor function in the disease process of dementia, it is essential that future research is done towards the feasibility, sensitivity to change, reliability, and validity of physical exercise tests that were found in this review. Since this information is in many cases still insufficient, the recommendation of optimal physical exercise tests should be used with care.

CONCLUSION
This review mapped the large heterogeneity in cognitive and physical function tests, which were used in international intervention studies with older patients with dementia. The provided neuropsychological (Table 2.4) and exercise (Table 2.5) test recommendations may lead to a more evidence-based choice of tests that better fit the research question of future studies. Because information on clinimetric quality is in many cases still insufficient, the recommended selection of currently optimal cognitive and physical exercise tests should be used with care. Researchers are advised to select those recommended tests that most closely fit into their study objectives.
REFERENCES


Chapter 2


