Differential diagnosis in the memory clinic: Exploring the value of improved neuropsychological examination

Rienstra, A.

Citation for published version (APA):
CHAPTER 4

Symptom validity testing in memory clinics:
hippocampal-memory associations and relevance
for diagnosing MCI

Anne Rienstra, Paul F.C. Groot, Pauline E.J. Spaan, Charles B.L.M. Majoie,
Aart J. Nederveen, Gerard J.M. Walstra, Jos F.M. de Jonghe, Willem A. van Gool,
Silvia D. Olabarriaga, Vladimir V. Korkhov and Ben Schmand

ABSTRACT
Patients with mild cognitive impairment (MCI) do not always convert to dementia. In such cases, abnormal neuropsychological test results may not validly reflect cognitive symptoms due to brain disease, and the usual brain-behavior relationships may be absent. This study examined symptom validity in a memory clinic sample, and its effect on the associations between hippocampal volume and memory performance.

Eleven of 170 consecutive patients (6.5%; 13% of patients younger than 65 years) referred to memory clinics showed non-credible performance on symptom validity tests (SVTs, viz. Word Memory Test and Test of Memory Malingering). They were compared to a demographically matched group (N=57) selected from the remaining patients. Hippocampal volume, measured by an automated volumetric method (Freesurfer), was correlated with scores on six verbal memory tests. The median correlation was $r = 0.49$ in the matched group. However, the relation was absent (median $r = -0.11$) in patients who failed SVTs.

Memory clinic samples may include patients who show non-credible performance, which invalidates their MCI diagnosis. This underscores the importance of applying SVTs in evaluating patients with cognitive complaints that may signify a pre-dementia stage, especially when these patients are relatively young.
1. INTRODUCTION

The concept of Mild Cognitive Impairment (MCI) is often used to diagnose Alzheimer’s disease (AD) in its pre-dementia stages (Albert et al., 2011; Petersen et al., 1999). Patients who satisfy MCI criteria are at high risk of developing dementia. Conversion rates from MCI to AD are about 10-15% per year. Approximately 80% show progression to dementia within six years (Petersen, 2007). These impressive conversion rates, however, have been observed in selected samples of relatively uncomplicated patients. In a broader context, the association between MCI and AD is far from unidirectional. Some patients do not progress to dementia at the long term, and some patients even return to normal upon re-examination (Maioli et al., 2007; Mitchell & Shiri-Feshki, 2009; Visser, Kester, Jolles, & Verhey, 2006). Although these patients satisfied the MCI criteria at baseline, a diagnosis of MCI in the sense of a pre-dementia stage is inaccurate because their temporary cognitive decline was probably not due to a degenerative brain disease.

Until now, the limited predictive validity of the MCI diagnosis has been related mainly to differences in research methods (diagnostic criteria, type of cohort) or to several reversible causes of cognitive decline (such as psychiatric conditions, and a variety of metabolic, nutritional or sensory impairments) (Gauthier & Touchon, 2005). Little attention has been paid to another cause: invalidity of the diagnosis due to non-credible performance during neuropsychological evaluation.

Non-credible performance is a threat to the validity of neuropsychological test results (Bush et al., 2005). If patients are unable or unwilling to invest the required amount of effort while doing the tests, abnormal test results may not validly reflect cognitive impairments due to cerebral dysfunction. In other words, it has become uncertain whether the tests measure what they intend to measure, i.e. cognitive functioning or cognitive impairment. Symptom validity testing may detect this lack of validity.

Two symptom validity tests frequently used in clinical practice are the Word Memory Test (WMT; Green, 2003) and the Test of Memory Malingering (TOMM; Tombaugh, 1996). The effort subtests of the WMT have high sensitivity in discriminating between persons asked to make a good effort and those instructed to feign a believable memory impairment (Brockhaus & Merten, 2004; Iverson, Green, & Gervais, 1999; Jing, Slick, Strauss, & Hultsch, 2002). Moreover, compensation-seeking individuals were shown to suppress their performance on this task (Green, Lees-Haley, & Allen, 2002). Also, the WMT was found to be sensitive in differentiating between carefully defined malingering and non-malingering groups (Greve, Ord, Curtis, Bianchini, & Brennan, 2008). The effort subtests of the WMT are more
sensitive to non-credible performance than the TOMM (Gervais, Rohling, Green, Ford, 2004; Sollman & Berry, 2011).

Research on the specificity of the WMT reported no differences in mean SVT scores between neurological patients with normal versus impaired memory (Green, Allen, & Astner, 1996). Even amnesic subjects with bilateral hippocampal damage scored above the WMT cut-offs (Goodrich-Hunsaker & Hopkins, 2009). However, some studies demonstrated lower specificity in cognitively impaired patients (Gorissen, Sanz, & Schmand, 2005; Greve et al., 2008; Merten, Bossink, & Schmand, 2007). Furthermore, concern over the specificity of the WMT has also been raised in a recent meta-analytic review (Sollman & Berry, 2011). Yet, if the so-called “dementia-profile” is taken into account, the specificity of the WMT is high even in groups with severe memory impairment, including dementia (Green, Montijo, & Brockhaus, 2011). Patients who show a dementia profile perform below the cut-off on the primary SVT variables, but their extremely low scores on the conventional memory subtests and the pattern of these scores suggest that this is due to genuine amnesia. Therefore, concerns over the WMT specificity appear to be resolved when the effort and ability subtest results are considered together. However, the effect of using the effort and the ability decision rules on the sensitivity of the WMT has not been studied yet.

The specificity of the TOMM in samples of older patients varies across studies. In a study of Teichner & Wagner (2004), all elderly normals (100%) and the majority (91.7%) of non-demented cognitively impaired patients (including those with MCI) made fewer errors than the suggested cut-off for non-credible performance. Previous studies for demented groups ranged from a high of 82% (Greve et al., 2006) to relatively low specificity rates of 45% (Dean, Victor, Boone, Philpot, & Hess, 2009) and 24% (Teichner & Wagner, 2004). Specificity of the TOMM is high in cognitively impaired patients who are able to obtain a normal MMSE score (Merten et al., 2007).

If abnormal test results of patients who fail SVTs are indeed not reflecting cognitive impairments due to brain disease, then it follows that brain – behavior relationships are weakened, or perhaps even completely absent, in these subjects. In patients with memory complaints this may affect the usual association between memory performance and medial temporal lobe structures, the hippocampus in particular (Dickerson & Eichenbaum, 2010).

The present study examined the associations between hippocampal volume and memory performance in elderly patients as a function of SVT results. We hypothesized that these associations would be quite strong in patients with credible SVT performance, whereas they would be absent in patients with non-credible SVT scores.
2. METHODS
2.1. PARTICIPANTS
This study was part of the longitudinal project 'Improving the early Diagnosis of Alzheimer’s Disease and Other dementias' (IDADO). Between February 2007 and October 2009 170 patients were recruited from the neurological and geriatric outpatient clinics, day clinics and memory clinics of six general and psychiatric hospitals in the Netherlands (Academic Medical Center, Medical Center Alkmaar, Slotervaart Hospital, GGZ-Noord-Holland-Noord, and Geriant Noord-Kennemerland). Patients between 50 and 85 years of age were included if they presented a) complaints of decline in cognitive or behavioral functioning (as expressed by the patient or a close relative) not normal for age, and b) essentially intact instrumental activities of daily living. Patients were referred to the study by dementia specialists (neurologists, geriatricians, psychiatrists) after an initial consultation consisting of patient history, collateral information from a relative, a clinical examination including a dementia screening instrument (MMSE), laboratory analyses, MRI or CT scan if deemed necessary, and application of the clinical dementia criteria of the Diagnostic and Statistical Manual of Mental Disorders–IV criteria (DSM-IV; American Psychiatric Association, 1994). For all participants the differential diagnosis included a possible early stage of dementia at the time of referral. Exclusion criteria were dementia as established during initial consultation by the dementia specialist according to the DSM-IV criteria, other brain disease or systemic disease sufficient to cause the mental complaints, current substance abuse or addiction, a medical condition or handicap that prevented neuropsychological evaluation, pre-existent mental retardation, contra-indications for MRI scanning, and insufficient command of the Dutch language. Psychiatric (co)-morbidty was not an exclusion criterion.

Eleven participants with non-credible scores on SVTs were identified and included in the present study. However, for three of these patients MRI data were not available due to claustrophobia. Therefore, all analyses have been performed with the total group of 11 participants, except those involving MRI data. A credible control group (N=60), comparable to the SVT non-credible group with respect to age, gender and education, was selected from the patients who had normal SVT results.

2.2. PROCEDURE
Participants underwent a comprehensive neuropsychological test battery to evaluate several cognitive domains. Tests were administered in a fixed order. In intervals of verbal memory tests, other non-verbal neuropsychological tests were administered (and
vice versa) in order to avoid interference. In this study we only report results on memory tests and symptom validity measures. Furthermore, a structured psychiatric assessment was administered and a high-resolution 3-Tesla structural brain MRI scan was made. Patients are being followed-up biannually until a clinical diagnosis is established.

Assessments were carried out in the hospital or at the participant’s home, in two sessions of 2–3 hours with suitable rest periods. MRI scans were made within a month of the neuropsychological examination. The local ethics committees of participating hospitals approved the study. Written informed consent was obtained from all patients after the nature of the study was fully explained.

2.2.1 Memory assessment
Verbal episodic memory was evaluated using the Rey Auditory Verbal Learning Test (RAVLT; memorizing a series of 15 unrelated concrete nouns in 5 learning trials, 20-minute delayed free recall, and recognition of the 15 items between 15 distracters; Rey, 1964); the Rivermead Behavioral Memory Test Logical Memory (RBMT LM) subtest (immediate and 20-minute delayed recall of two news stories; Wilson, Cockburn, & Baddely, 1985); the Visual Association Test (VAT; participants are shown pictures of two common objects, representing an unusual combination, e.g. an ape carrying an umbrella; recall is tested without a delay by showing one object and asking what other object is missing; Lindeboom & Schmand, 2003); the enhanced cued recall test (ECR; immediate recall of 16 items with help of category cues; after a short distraction task subjects are asked to freely recall as many of the pictures as possible; category cues are provided for the remaining items; Grober, Buschke, Crystal, Bang, & Dresner, 1988); the Paired-Associate Learning Test (PALT; cued recall of seven semantically related and seven semantically unrelated word pairs; Spaan, Raaijmakers, & Jonker, 2005); and the Paired-Associate Recognition Test (PART; explicit recognition of the word pairs from the Paired-Associate Learning Test; each first word is presented along with the target word, two semantically related distracters and an unrelated word; Spaan & Schmand, 2010).

2.2.2 Assessment of symptom validity
The Word Memory Test (WMT; Green, 2003) requires immediate (IR) and 30 minute-delayed recognition (DR) of 20 semantically linked word pairs (e.g., dog-cat), in which participants are shown new word pairs and are asked to select the words they have seen in the original list (e.g., dog from the pair dog-rabbit). IR, DR and the consistency (CNS) between IR and DR are the SVT variables of the test. The SVT
measures are followed by a series of memory tests of gradually increasing difficulty, sensitive to genuine verbal memory impairment.

The Test of Memory Malingering (TOMM; Tombaugh, 1996) is a 50-item visual recognition test consisting of two learning trials, each followed by a recognition trial in which participants make two-alternative forced choice decisions to identify the objects they have seen previously. After 15 minutes a delayed recognition trial (Retention Trial) is administered. The last two trials are used as SVT measures.

2.2.3. Group allocation: credible versus non-credible
Non-credible SVT performance was defined using the norm-based cut-off scores published in the TOMM and WMT manuals (Green, 2003; Tombaugh, 1996). Patients were considered non-credible performers if they produced non-credible scores on one or more of the WMT symptom validity subtests, unless they showed the dementia profile. Patients who showed the WMT dementia profile were reallocated to the non-credible group if they failed the TOMM and scored more than 23 points on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975). This latter criterion was applied because bona fide neurological patients with normal MMSE scores are able to pass the TOMM (Merten et al., 2007).

2.2.4. Psychiatric assessment
Current and lifetime Axis I diagnoses according to DSM-IV (American Psychiatric Association, 1994) and International Classification of Diseases-10 (World Health Organization, 1992) were assessed with the MINI International Neuropsychiatric Interview (M.I.N.I. Dutch version 5.0.0; Sheehan et al., 1998). The Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) was used to determine the presence and severity of affective disturbances.

2.2.5. MRI acquisition
Imaging was performed on a 3.0 Tesla MRI system (Philips Intera, Best, The Netherlands) with a 6-channel SENSE head coil. A gradient echo 3D FFE, T1-weighted, sagittal sequence was used with the following pulse sequence parameters: echo time [TE] = 3.5 ms, repetition time [TR] = 9 ms, field of view [FOV] = 256, 232, 170 mm (FH,AP,RL), scanning matrix = 256x231, flip angle = 8°, 170 slices, voxel size = 1x1x1 mm, slice thickness = 1 mm (no gap).
2.2.6. Image processing
Volume measurements of the hippocampus were calculated automatically using the Freesurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/) version 4.5.0, installed on local and remote computers (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999; Olabarriaga, Glatard, & de Boer, 2010). MR images were processed using the default analysis settings (https://surfer.nmr.mgh.harvard.edu/fswiki/FreesurferMethodsCitation), with the exception that surface statistics were not taken into account. Details of the procedures for subcortical segmentation are described elsewhere (Fischl et al., 2002). Validation studies showed reasonable overlap in absolute hippocampal volumes between Freesurfer and manual volumetry (Lehmann et al., 2010; Tae, Kim, Lee, Nam, & Kim, 2008). All images were visually inspected for gross structural abnormalities, accuracy of registration, and presence of artifacts. The hippocampal volumes in each individual were normalized using the total intracranial volume to control for variation in head size.

2.3. STATISTICAL ANALYSES
Differences between the groups were analyzed with independent t-tests, Mann-Whitney tests or chi-square test. Pearson’s r correlations were calculated to document associations between memory tests and hippocampal volume (Spearman’s rho rank-ordered for non-normally distributed variables). Partial correlations were used to explore the relationships between the memory tests and the hippocampal volume of each cerebral hemisphere, partialling out the effect of the other side. Fisher’s exact test and phi coefficient were used to evaluate relationships between groups and psychiatric diagnoses assessed with the M.I.N.I. Effect sizes were expressed as Cohen’s d (Cohen, 1988).

Analyses were performed using SPSS for Windows (version 18.0). One-tailed p values ≤0.05 were considered significant. Missing VAT data (N = 2) were imputed with the SPSS Missing Value Analysis EM method.

3. RESULTS
3.1. DESCRIPTION OF STUDY SAMPLE
The cohort consisted of 170 patients who satisfied older or more recent MCI criteria (Albert et al. 2011; Petersen, 2004), as well as patients with uncertain diagnosis, patients with a deferred diagnosis at the time of referral, and patients who turned out to be worried well. Seven percent of the patients (11 out of 170; 95% confidence interval: 4 - 11) scored positively on SVTs. Ten of these patients were younger than 65 years.
Thus, of the 76 patients in our sample who were younger than 65 years of age, 13% showed non-credible performance (95% confidence interval: 6 – 23).

Distribution plots of the hippocampal volumes revealed a definite outlier in the SVT non-credible group. This patient was excluded from further analysis. In the credible control group, 18 patients scored below the cut-off score on at least one of the WMT SVT measures. All these patients showed the dementia-profile. Three of these patients also failed the TOMM. Although their MMSE scores were less than 24 points, we decided to remove these patients from further analysis since we could not rule out non-credible performance in these individuals.

Table 1 shows the demographic and clinical characteristics. No significant differences for any of these variables were found between the SVT non-credible group and the credible control group.

### TABLE 1
Demographic and Clinical Characteristics of the Samples

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Non-credible group N = 11</th>
<th>Credible control group N = 57</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>57.55 (7.44)</td>
<td>61.12 (9.75)</td>
<td>.25</td>
</tr>
<tr>
<td>Male, %</td>
<td>45.50</td>
<td>31.60</td>
<td>.37</td>
</tr>
<tr>
<td>Education, y</td>
<td>11.09 (2.17)</td>
<td>12.33 (2.48)</td>
<td>.13</td>
</tr>
<tr>
<td>MMSE</td>
<td>26.18 (3.76)</td>
<td>27.33 (2.05)</td>
<td>.15</td>
</tr>
</tbody>
</table>

Values are expressed as mean (SD) unless otherwise indicated.

MMSE = Mini Mental State Examination.

### 3.2. RELATIONS BETWEEN HIPPOCAMPAL VOLUME AND MEMORY

There were no significant differences between the SVT non-credible group and the credible control group in hippocampal volumes and memory performance, except for the PART, where SVT non-credible subjects scored less (see Table 2). Table 3 shows the correlations between memory tests and hippocampal volume. Significant, positive correlations were found for all memory measures in the credible control group. These correlations were mostly negative in the SVT non-credible group, and they were not significant. Figure 1 shows the correlation between performance on RAVLT trial 1-5 and left hippocampal volume in the SVT non-credible and credible control groups. This illustrative example is representative for the correlations between hippocampal volume and the other memory tests.
Partial correlations between the verbal memory tasks and the left hippocampal volume in the credible control group, controlling for the volume of the right hippocampus, showed a median correlation coefficient of $r = 0.24$ (range $0.06 - 0.38$). Associations for the right hippocampus, controlling for the volume of the left side, revealed a median correlation coefficient of $r = -0.02$ (range $-0.13 - 0.03$).

### 3.3. RELATIONSHIPS BETWEEN HIPPOCAMPAL VOLUME AND SVT PERFORMANCE

Correlation coefficients between the volume of the left hippocampus and SVT scores are presented in Table 4. There were no significant correlations in the non-credible SVT group, and again, the correlations were negative. In the credible control group, no significant correlations were found for the TOMM. However, all WMT scores were strongly related to the volume of the left hippocampus.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Non-credible group</th>
<th>Credible control group</th>
<th>$p$ Value</th>
<th>Cohen's $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mm$^3$ Hippocampus left/TIV*100</td>
<td>26.62 (2.74)</td>
<td>23.26 (5.21)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>mm$^3$ hippocampus right/TIV*100</td>
<td>26.40 (4.11)</td>
<td>23.61 (4.98)</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>RAVLT trial 1-5$^a$</td>
<td>33.43 (6.92)</td>
<td>39.25 (10.36)</td>
<td>09</td>
<td></td>
</tr>
<tr>
<td>RAVLT delayed recogniti$^a$</td>
<td>5.71 (2.87)</td>
<td>7.39 (3.99)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>RBMT LM immediate recall$^b$</td>
<td>17.78 (6.51)</td>
<td>15.77 (6.62)</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>RBMT LM delayed recall$^b$</td>
<td>12.00 (7.56)</td>
<td>10.62 (6.68)</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Visual Association Test$^b$</td>
<td>19.14 (6.74)</td>
<td>20.36 (4.79)</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Enhanced Cued Recall$^b$</td>
<td>13.71 (2.49)</td>
<td>14.09 (2.84)</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>PALT$^a$</td>
<td>14.00 (7.44)</td>
<td>20.42 (10.07)</td>
<td>06</td>
<td></td>
</tr>
<tr>
<td>PART$^a$</td>
<td>8.14 (3.72)</td>
<td>11.51 (3.78)</td>
<td>02</td>
<td></td>
</tr>
<tr>
<td>HADS total$^b$</td>
<td>22.91 (11.85)</td>
<td>14.13 (7.83)</td>
<td>003</td>
<td>1.03</td>
</tr>
<tr>
<td>HADS-A$^b$</td>
<td>11.82 (5.84)</td>
<td>7.66 (4.27)</td>
<td>004</td>
<td>0.73</td>
</tr>
<tr>
<td>HADS-D$^b$</td>
<td>11.09 (6.50)</td>
<td>6.46 (4.27)</td>
<td>007</td>
<td>1.18</td>
</tr>
</tbody>
</table>

TIV = Total Intracranial Volume; RAVLT = Rey Auditory Verbal Learning test; RBMT LM = Rivermead Behavioral Memory Test Logical Memory; PALT = Paired-Associate Learning Test; PART = Paired-Associate Recognition Test
HADS total = Hospital Anxiety and Depression Scale Total score; HADS-A = Hospital Anxiety and Depression Scale Anxiety subscale; HADS-D = Hospital Anxiety and Depression Scale Depression subscale; Values are expressed as mean (SD)

$^a$Independent t-tests; $^b$Mann-Whitney tests
### TABLE 3
**Correlations between Memory Tests (raw scores) and Hippocampal Volume**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Credible control group</th>
<th>Non-credible group</th>
<th>Credible control group</th>
<th>Non-credible group</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAVLT trial 1-5</td>
<td>0.51</td>
<td>-0.04</td>
<td>0.48</td>
<td>0.06</td>
</tr>
<tr>
<td>RAVLT delayed recognition</td>
<td>0.59</td>
<td>0.23</td>
<td>0.54</td>
<td>0.28</td>
</tr>
<tr>
<td>RBMT LM immediate recall</td>
<td>0.27</td>
<td>-0.14</td>
<td>0.29</td>
<td>-0.25</td>
</tr>
<tr>
<td>RBMT LM delayed recall</td>
<td>0.38</td>
<td>-0.22</td>
<td>0.37</td>
<td>-0.29</td>
</tr>
<tr>
<td>Visual Association Test</td>
<td>0.53</td>
<td>-0.33</td>
<td>0.51</td>
<td>-0.33</td>
</tr>
<tr>
<td>Enhanced Cued Recall</td>
<td>0.58</td>
<td>-0.04</td>
<td>0.61</td>
<td>-0.04</td>
</tr>
<tr>
<td>PALT</td>
<td>0.44</td>
<td>-0.17</td>
<td>0.36</td>
<td>-0.14</td>
</tr>
<tr>
<td>PART</td>
<td>0.46</td>
<td>-0.02</td>
<td>0.44</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

*Pearson r; *Spearman rho; *p < 0.05 (one-tailed), *# p < 0.01 (one-tailed).

RAVLT = Rey Auditory Verbal Learning Test; RBMT LM = Rivermead Behavioral Memory Test Logical Memory; PALT = Paired-Associate Learning Test; PART = Paired-Associate Recognition Test.

### TABLE 4
**Correlation of Left Hippocampal Volume with Symptom Validity Tests (SVTs)**

<table>
<thead>
<tr>
<th>Test</th>
<th>Non-credible group</th>
<th>Credible control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOMM Trial 1</td>
<td>-0.31</td>
<td>0.20</td>
</tr>
<tr>
<td>TOMM Trial 2</td>
<td>-0.38</td>
<td>0.11</td>
</tr>
<tr>
<td>WMT IR</td>
<td>-0.22</td>
<td>0.46</td>
</tr>
<tr>
<td>WMT DR</td>
<td>-0.02</td>
<td>0.34</td>
</tr>
<tr>
<td>WMT CNS</td>
<td>-0.36</td>
<td>0.41</td>
</tr>
<tr>
<td>WMT MC</td>
<td>-0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>WMT PA</td>
<td>-0.30</td>
<td>0.52</td>
</tr>
<tr>
<td>WMT FR</td>
<td>-0.10</td>
<td>0.56</td>
</tr>
</tbody>
</table>

*Spearman rho; *Pearson r;

#p < 0.01

TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recognition; DR = Delayed Recognition; CNS = Consistency; MC = Multiple Choice; PA = Paired Associate; FR = Free Recall.
FIGURE 1
Scatterplot Rey Auditory Verbal Learning Test Total Score on Trial 1-5 and Volume of the Left Hippocampus Corrected for Total Intracranial Volume (ICV).
3.4. EMOTIONAL-BEHAVIORAL OR PSYCHIATRIC SYMPTOMS

In the credible control group 17 patients met criteria for at least one current Axis I diagnosis versus six patients in the non-credible SVT group. Patients in the non-credible SVT group were significantly more often diagnosed with current major depressive disorder (Fisher exact test; \( p = 0.01 \)). The strength of the association was moderate (\( \phi = 0.34 \)). There were no group differences for other diagnoses assessed.

HADS scores revealed that patients in the SVT non-credible group reported significantly more symptoms of depression and anxiety than patients in the credible control group. The effect sizes for these measures were large.

Furthermore, based on the originally proposed cut-off values, the proportion of patients in the non-credible SVT group reporting increased levels of anxiety (HADS anxiety score ≥ 8 points) was significantly higher than in the credible control group (9 out of 11 patients in the SVT non-credible group versus 28 out of 60 patients in the credible control group; \( p = 0.04 \)). Patients in the SVT non-credible group tended to be more often classified as depressed than patients in the credible control group (\( p = 0.07 \)).

4. DISCUSSION

This study is part of a longitudinal project that aims to improve the prediction of conversion to dementia in memory clinic patients by application of SVTs. Whether these tests are able to do so, will be determined by longitudinal results, which are currently being collected. However, if abnormal neuropsychological test results of some patients do not reflect cognitive impairments due to brain disease, we might already find a moderating effect of SVT performance on brain-behavior relationships at baseline. The main finding of our study is indeed that the correlation between hippocampal volumes and memory performance in memory clinic patients is rather strong, whereas this relation is virtually absent when these patients exert non-credible performance.

Our findings are relevant for the discussion on the validity of the MCI diagnosis. As stated in the Introduction, the limited predictive validity of MCI has been attributed to various causes. Our results suggest another possible cause: non-credible cognitive test performance in a subgroup of patients, which, if undetected, will result in invalid MCI diagnoses.

From a researcher’s point of view, these patients cause noise in the data. The possibility that the scientific database may be contaminated by negative response bias in an unknown number of study participants has been discussed with some vigour.
Differential diagnosis in the memory clinic: before, in particular in the context of PTSD (e.g., Rosen, 2004; Rubenzer, 2009). The conclusion that arises from this discussion is, that in some scientific contexts possible response bias should be checked for, and samples should be cleaned from invalid responders. Otherwise, research data may not only be contaminated, but yield wrong conclusions, which do not correspond to the real facts. In this study the percentage of patients who failed SVTs was not excessively high. Undetected noise would not have affected our data very much. In settings where this percentage is higher (see discussion below), it could obscure brain-behavior relations. For example, if 25% of the patients had shown non-credible performance, then brain-behavior correlations in the order of 0.6 would have dropped to about 0.4 (details available on request).

About seven percent of the patients scored positively on SVTs, which corresponds to the rates that have been previously reported in medical or psychiatric cases not involving litigation or compensation (Mittenberg, Patton, Canyock, & Condit, 2002). However, this percentage was almost doubled in patients younger than 65 years of age. In fact, all patients who failed SVTs except one were under 65 years. In the Netherlands, for most people this is the retirement age. It is therefore tempting to hypothesize that SVT failure of these patients might be associated with psychosocial factors such as job dissatisfaction or wish for early retirement.

The dementia profile of the WMT seems to be a robust indicator, with a low chance of false-positives. When the dementia profile is taken into account, specificity is high in patients with severe memory impairment and probable dementia (Green et al., 2011). Furthermore, in our sample only the WMT scores were sufficient to classify patients into the non-credible and credible groups. Therefore, we used the TOMM to check if the patients with a dementia profile on the WMT were indeed credible, as confirmed by their credible TOMM score. This turned out to be the case for 15 of the 18 patients in the credible group who showed the dementia profile. Three patients in this group failed the TOMM; however, their MMSE scores were less than 24 points, and although it is likely that they were unable to pass the TOMM, we removed them from the analyses. By using two SVTs (viz. the TOMM and the WMT) we trust that sensitivity was high, and that the false positive rate was maximally reduced (due to taking the dementia profile into account, backed-up by the TOMM).

A possible explanation of the non-credible performance is that it results from a conscious attempt to suppress performance motivated by external incentives (i.e. malingering). Two patients in our non-credible SVT group told us that they were applying for compensation for their reported cognitive impairments. Another potential explanation is the "hidden agendas", which psychiatric patients may have, and that
presumably are present in some memory clinic patients (Dandachi-Fitzgerald, Ponds, Peters, & Merckelbach, 2011; Van Egmond & Kummeling, 2002).

Furthermore, it might be possible that emotional-behavioral problems reduce performance on SVTs because patients are unable to invest the required effort while doing the tests. Many studies have shown, however, that emotional problems like depression or anxiety disorders by themselves are not sufficient to cause SVT failure (Ashendorf, Constantinou, & McCaffrey, 2004; Rees, Tombaugh, & Boulay, 2001; Yanez, Fremouw, Tennant, Strunk, & Coker, 2006). Even one of the most severe psychiatric illnesses, schizophrenia, does not necessarily imply incapacity to pass SVTs (Gorissen et al., 2005). Nevertheless, it is conceivable that, compared to the general population, the risk of SVT failure is increased in individuals who present with emotional-behavioral problems or psychiatric disorders. In our patients with non-credible SVT scores, we found evidence supporting this assumption. In these patients associations between hippocampal volume and memory performance were negligible or absent, but they reported significantly more often symptoms of depression and anxiety. It is also noticeable that for one third of these patients MRI data were not available due to claustrophobia. However, these findings could also be explained by over-reporting of emotional symptoms in the SVT non-credible group. Non-credible cognitive performance and over-reporting of emotional symptoms are correlated (Dandachi-Fitzgerald et al., 2011; Haggerty, Frazier, Busch, & Naugle, 2007; Whiteside, Dunbar-Mayer, & Waters, 2009).

We did not systematically explore reasons for SVT failure. This limits any attempt to explain the causes for SVT failure. Most likely there was a mixture of reasons for failure on SVTs in the non-credible group. Yet, although the exact mechanism is important by itself, it is irrelevant to our purpose. The most important issue in this study is that the result of the neuropsychological evaluation is invalid in case of SVT failure. If we apply this logic to MCI, we may assume that MCI patients with genuine cognitive impairments as confirmed by credible performance on SVTs are likely to convert to dementia, irrespective of psychiatric co-morbidity. On the other hand, in MCI patients with non-credible SVT performance, abnormal cognitive test scores cannot validly be interpreted as signs of genuine impairments, and a degenerative brain disease is much less likely, although not ruled out.

Another result of our study is the relationship between hippocampal volume and SVT performance. We found significant correlations for the effort measures of the WMT but not for the TOMM, suggesting that the WMT is more sensitive to neurological impairment. This confirms studies reporting that the TOMM is less sensitive to
severe cognitive impairment than the WMT (Merten et al., 2007; Greve et al., 2008). In the study by Merten et al. both the WMT and the TOMM correlated with tests of declarative memory, but correlations for the WMT were higher. The difference with the present results is probably due to a lower severity of cognitive impairment in our study, which caused a ceiling effect on the TOMM.

This study has several other limitations. The first concerns the small size of the non-credible SVT group. This does not imply, however, that the influence of non-credible performance on neuropsychological test results should be underestimated. It is quite possible that the low incidence of non-credible performance in our study might be caused by a selection bias. Exaggeration of symptoms or impairments is probably a common cause of non-credible performance in clinical practice when patients feel the need to get recognition for their complaints (Miller, 2001). Our patients were asked to take part in a research project, which inherently implies recognition for their complaints. This may have reduced any tendency to exert non-credible performance. Furthermore, we recruited the major part of our sample in neurological memory clinics. Although depression is a common diagnosis in this type of clinic (Hejl, Hogh, & Waldemar, 2002), one might expect a higher proportion of SVT failure in patients from psychiatric memory clinics, especially below 65 years of age.

A second limitation is the fact that our longitudinal data are not yet available. Patients with subjective memory complaints at baseline may show change in hippocampal volume at follow-up (Stewart et al., 2011). Strictly speaking, for our patients with non-credible SVTs and abnormal cognitive tests, it is uncertain whether or not they are in a pre-dementia stage. Cognitive impairment has not been validly established in these patients, but it has not been disproved either. Longitudinal results have to reveal what their final clinical outcome will be and whether they will progress to dementia.

A third possible concern is the automatic hippocampal volume measurement by FreeSurfer software. Mean hippocampal volumes measured using FreeSurfer are significantly larger than the volumes determined with manual volumetric methods (Tae et al., 2008). We do not know whether this also occurred in our study, but if so, the resulting volumetric overestimation was a systematic one that probably did not invalidate our findings. We found positive correlations between hippocampal volumes and memory, consistent with other studies using manual tracing (Petersen et al., 2000). Our findings are also consistent with reports showing that the relationship between hippocampal volume and memory is material-specific: verbal memory tests were found to correlate with the left hippocampal volume, whereas partial correlations with the right hippocampus were not significant (de Toledo-Morrell et al., 2000). Moreover,
we did not take into account hippocampal subfields and other medial temporal lobe regions that contribute to memory. Most studies find severe atrophy of the hippocampus in AD patients (Zakzanis, Graham, & Campbell, 2011), but some suggest that parahippocampal volume is even a better biomarker (Echávarri et al., 2011).

Finally, one outlier in the group of patients with non-credible SVT performance was excluded. She was a 65-year-old woman with disproportionally small hippocampi. This may be a sign of a degenerative brain disease, but for several reasons this is unlikely and therefore this patient has been removed. First, this patient had suffered since 10 years from recurring episodes of treatment resistant depression. Second, follow-up examination again revealed a pattern of SVT scores indicating non-credible performance. Finally, there were no clear-cut signs of cognitive decline over the years. Conversely, discrepancies were found between test scores on re-examination. For example, the MMSE score was 21/30 in 2005, 26/30 in 2008 and 27/30 in 2009. Furthermore, in 2005 administration of the Wisconsin Card Sorting test (Heaton, Chelune, Talley, Kay, & Curtiss, 1993), was aborted. She was barely able to complete one category and made many perseverative errors. In 2008, she completed 6 categories and made a normal number of perseverative mistakes. Furthermore, her test scores were inconsistent with her history. Results on memory tests, for example, were extremely low (and sometimes on chance level), while she was still baby-sitting her 3 grand children on regular basis.

In conclusion, the data presented here suggest that brain-behavior relations may be obscured and the diagnosis of a preclinical dementia stage may be invalid in patients who do not perform to the best of their ability during cognitive evaluation. This underscores the importance of administering formal tests of symptom validity in the examination of patients who present with cognitive complaints that may signify an early stage of dementia. If researchers of the early stages of dementia apply SVTs, they may avoid this type of ‘pollution’ of their samples, thus increasing the statistical power of their studies. If clinicians administer SVTs in their diagnostic routine, it will help to improve differential diagnosis by averting false-positive diagnoses of (early) dementia (Heilbrunner, Sweet, Morgan, Larrabee, & Millis, 2009).
Acknowledgments

We performed computations using the e-bioinfra platform at the AMC, using resources of the Dutch e-Science Grid, BiGGrid project, which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO). There are no conflicts of interest.

References


Exploring the value of improved neuropsychological examination


Exploring the value of improved neuropsychological examination


Differential diagnosis in the memory clinic:


