Using the Rey Auditory Verbal Learning Test to Compare Memory Impairment in Idiopathic Normal Pressure Hydrocephalus and Alzheimer's Disease

Josefine Mellqvist¹ and Fredrik Juhlin²
¹Geriatric, Neurology and Rehabilitation, Kungälv Hospital, Kungälv, Sweden
²Neuro and Rehabilitation Clinic, Södra Älvsborg Hospital, Borås, Sweden

Advisor: Per Hellström, Institute of Neuroscience and Physiology, Göteborg University, Göteborg, Sweden
Abstract

Idiopathic Normal Pressure Hydrocephalus (iNPH) is a treatable dementing disorder, which is under-diagnosed and often mistaken for Alzheimer’s disease (AD) or Vascular dementia (VAD). This study aims to investigate verbal memory impairment, as measured by the Rey Auditory Verbal Learning Test (RAVLT), in the early stages of dementia caused by iNPH or AD. The matched samples comprised 30 healthy individuals (HI), 84 participants with AD and 84 with iNPH. The clinical samples were divided into two subgroups based on scores on the MMSE: High performers (27 - 30 points, n = 30 respectively) and Medium performers (18 - 26 points, n = 54 respectively). Results show impaired memory performances in both clinical samples in comparison to HI, despite high MMSE scores. Despite similar results on measures capturing learning, the iNPH patients outperformed AD patients on measures of recall and recognition, the latter being on par with the results of HI.
Introduction

Dementia is a decline in cognition and functioning (McKhann et al., 2011). Estimates suggest that nearly 44 million people suffered from a dementing disorder in 2016; due to higher average age and a larger population the number that has more than doubled since 1990 (estimated to 20.2 million). Worldwide, dementia is the fifth leading cause of death (Nichols et al., 2019).

The most common of the dementing disorders is Alzheimer’s Disease (AD), alone causing between 60 and 70% of the diagnosed cases (World Health Organization, 2017). Alois Alzheimer first described the disease in the early 20th century (Alzheimer, 1907; in English translation by Stelzmann et al., 1995). In the diagnostic process of AD, at least two of the domains of memory, reasoning, visuospatial abilities, language or personality, should be altered to an extent that interferes with everyday life and constitutes a loss of function in comparison to previous levels. Neuroradiology and analyses of cerebrospinal fluid biomarkers are commonly used methods in the diagnostic procedure (McKhann et al., 2011). For the vast majority, memory is the first domain to be affected (Bastin & Salmon, 2014). The disease starts in the medial temporal lobe and early on affects the hippocampus and its connections to other parts of the brain, thereby destroying structures and circuits of importance to memory functioning (Hyman, et al., 1984). In the cerebrum, the disease mainly affects the temporal and parietal lobes, but also involves the frontal lobes, while leaving the sensory parts of the parietal and occipital lobes relatively intact. At the later stages of the disease almost all of the cerebrum is affected (Braak & Braak, 1995). The two greatest risk factors for AD are high age and female gender. The latter is associated with
the longer life expectancy in women, but also with genetic differences (Hsu et al., 2019). AD is almost twice as common in women as in men (with a ratio of 1.9 according to Cao et al., 2020; 1.6 according to Nichols et al., 2019). During the last decade, theories of AD have focused on the presence of neuropathological hallmarks of amyloid plaques and neurofibrillary tangles, while earlier research identified genetics that increases the risk for developing AD (Bondi et al., 2017). Although AD has been known for more than a century, the pathological processes are not yet fully understood, and despite extensive research, there is still no cure. Between 2002 and 2012, trials of disease-modifying treatments in AD had a failure rate of 99.6 %. No new drugs treating AD have been approved in the U.S since 2003 (Cummings et al., 2014).

Idiopathic Normal Pressure Hydrocephalus (iNPH) is another, less known and less common, illness that can cause dementia. The condition was first described in 1965 by Hakim and Adams. It is believed to be caused by decreased absorption of cerebrospinal fluid (CSF), leading to a characteristic enlargement of the cerebral ventricles. The diagnostic procedure includes neuroradiology and investigations of symptoms of gait abnormality, balance disturbance, urine incontinence and cognitive impairment (executive functions, attention, speed and memory) (Hellström et al., 2012). The disease usually has an insidious onset and a gradual progression (Hakim & Adams, 1965). A large study (N = 1,524) from 2017 by Kuriyama and colleagues, noticed a difference in initial symptomatology, where gait disturbance was the most common initial symptom for men, and cognitive decline for women. INPH is believed to be an under-diagnosed condition (Nassar & Lippa, 2016; Martín-Laez et al., 2016; Jaraj et al., 2014). In a Swedish epidemiological study by Jaraj et al. (2014), the prevalence of
possible iNPH was estimated to 0.2 % among 70-79-year-olds and 5.9 % in people over the age of 80, equally affecting men and women. Another Swedish epidemiological study (Andersson et al., 2019) showed a difference in prevalence between men (4.6 %) and women (2.9 %) (due to the size of the sample, the difference was not statistically significant). In clinical settings, an overrepresentation of men has been observed with a ratio of 1.5:1 (Martin-Laez et al., 2016; Kuriyama et al., 2017). Unlike most other dementing disorders, iNPH is treatable. By inserting a shunt, excessive CSF can be removed, which causes a reduction of the symptoms. With surgery, life expectancy increases with an average of 2.2 years (Tullberg et al., 2018), and up to 80 % of the patients showed signs of improvement (Klinge et al., 2012). The degree of recovery spans from minor improvements to total recovery (Hellström et al., 2012). The magnitude of remaining symptomatology following surgery depends on the preoperative severity, making early diagnosis essential for recovery (Andrén et al., 2014).

**Cognition in Normal Aging, AD and iNPH**

A decline in most cognitive functions characterizes aging, notable exceptions being verbal comprehension (Baxendale, 2011; Salthouse, 2019) and semantic knowledge (Kennedy et al., 2015). Memory, and especially episodic memory, is considered sensitive to aging. For most people, the decline in memory functioning starts in their sixties (Nyberg et al., 2012) but for some as early as in their thirties (van der Elst et al., 2005). A decline in memory function can occur for numerous reasons, and a well-functioning memory is dependent on a network of abilities, engaging temporal,
frontal and parietal areas. Encoding also requires attention and executive skills (Putcha et al., 2019). Among the elderly, encoding can be less robust, making free recall harder without affecting recognition (Tromp et al., 2015).

Cognitive impairment in AD has been well examined (Salmon & Bondi, 2009). Memory is the cognitive domain that is most affected by AD, often impaired as much as ten years before diagnosis (Bastin & Salmon, 2014). In a recent Swedish study (Josefsson et al., 2019), impaired memory was observed 10-15 years before clinical diagnosis. As the disease spreads from the medial temporal lobes, the symptomatology evolves. Other functions, like abstract reasoning and attention (working memory), are also affected years before diagnosis (Bastin & Salmon, 2014). Other common cognitive impairments include visuospatial functions, executive functions, language and semantic knowledge (Salmon & Bondi, 2009).

The cognitive impairments in iNPH have not been as thoroughly studied as those in AD. In iNPH, impaired psychomotor speed, executive functions and attention are the cognitive functions considered to be most affected (Hellström et al., 2012; Tarnaris et al., 2011; Ogino et al., 2006). In a review article by Picascia et al. (2015), memory impairment in iNPH seems milder than in AD, and when observed, interpreted to be a result of a poor frontal function. In the initial stages of the disease, memory functions have been believed to be fairly intact. A study comprising 64 participants, where the sample was divided into four subgroups; Little or no cognitive impairment, Mildly cognitively impaired, Frontosubcortical dysfunction, and Globally impaired (including memory deficits), respectively, showed that the longer the subjects had been suffering from symptoms of gait abnormality and loss of bladder control, the more severe were
the cognitive decline. In general, verbal memory was within the normal range, reduced only in the globally impaired group (Picascia et al., 2016). In advanced iNPH, Mathew and colleagues, (2018) singled out memory as the most impaired aspect of cognition (n = 41), at a level similar to patients with advanced AD. The Mini-Mental State Examination (MMSE) was used to assess general cognition (M = 18.4, n = 14).

However, other researchers have found memory to be equally impaired in iNPH as in AD, even in the earlier stages of disease progression. In a study of general cognition, performance on the Rey Auditory Verbal Learning Test (RAVLT) did not differ between a group with iNPH and a group with AD (matched on age, sex and education, and with similar scores on MMSE with means of 21.6 and 21.3, respectively) (Saito et al., 2011).

Using Neuropsychological Assessment in Diagnostic Processes

Neuropsychological assessment is an essential part of diagnostics and in differentiating types of dementing disorders, particularly in the early stages of cognitive decline. A recent study by Thomas et al. (2020) emphasized the importance of neuropsychological assessment in prodromal AD, even more than the detection of biomarkers (e.g. beta-amyloid). Diagnostic accuracy improves by combining neuropsychological assessment and neuroimaging (Bastin & Salmon, 2014). As cognitive impairment progresses, the usefulness of neuropsychological assessment as a diagnostic aid decreases, since it becomes increasingly challenging to identify diagnosis-specific characteristics (Bondi et al., 2017).
**The RAVLT**

A frequently used test in neuropsychological assessment is the RAVLT. The test was introduced by the Swiss psychologist André Rey in 1958 (Rey, 1958). Edith Meyer Taylor translated the test into English in 1959, but she also added a distraction list and two free recall trials, thereby giving the test the format that is generally used ever since (Taylor, 1959). The test consists of five free-recall trials of a word list consisting of 15 concrete nouns, followed by one trial with a distraction list (trial six), immediately preceding a recall trial of the original list (trial seven). It is later followed by a delayed recall trial of the original list after 20-30 minutes (trial 8). The test ends with a recognition part. The RAVLT yields several measures of memory function (Vakil & Blachstein, 1993), including declarative memory, working memory and recognition (van der Elst et al., 2005).

The test is often used in diagnosing AD and is considered useful when trying to identify future AD patients among those suffering from mild cognitive impairment (MCI) (Russo et al., 2017; Eckerström et al., 2013), differentiating patients with Lewy body dementia from patients with AD (Bussè et al., 2017) and AD from Frontal lobe dementia (Ricci et al., 2012). RAVLT is also frequently used when assessing differences in memory ability between iNPH patients and healthy individuals (Hellström et al., 2012). When using the RAVLT in a healthy population, young adults outperform older, and women outperform men. Further, more highly educated tend to do better than less educated (Sundermann et al., 2017; van der Elst et al., 2005).

**Learning:** The first part of the RAVLT provides several measures of learning, the most common being the sum of the five learning trials (Schmidt, 1996). In order to
better understand impaired memory performance, Thomas et al. (2018) stress the importance of additional analysis to provide further information, like e.g., scores for serial position effects and scores for the learning curve like learning delta (i.e., the difference between the number of words recalled on the first and the fifth trial). The sum of the five trials has been considered useful in distinguishing AD from healthy individuals, both concerning encoding efficiency and serial position effects. In a study by Mistridis and colleagues (2015), word list encoding was impaired more than five years prior to AD diagnosis. For iNPH, in comparison to healthy individuals, performance on the RAVLT learning was significantly worse (Hellström et al., 2012).

When, in a test like the RAVLT, the number of words is more than what is possible for most people to remember immediately, one is more likely to remember the initial and the ending words, this is referred to as the primacy and the recency effect respectively. The overall phenomenon of the positioning of words for the later recall is referred to as the serial position effect (Murdock, 1962). In AD, a pattern of preserved recency effect but a lack of primacy effect is a common feature (Martín et al., 2013, Bayley et al., 2000). According to Egli et al. (2014), to discriminate non-converters from AD converters in an MCI population, the most useful measure in word-list-tests was poor scores of primacy where the converters showed the typical AD pattern. A study by Sindorio et al. (2017) comparing iNPH and AD patients on the RAVLT, revealed an apparent recency effect but a lack of a primacy effect in the AD group, whereas the iNPH group displayed both primacy and recency effects.

**Recall/Recognition:** The RAVLT consists of two recall trials, immediate (i.e., trial 7) and delayed (trial 8), and one recognition trial. Estimates of forgetting can also
be calculated, e.g., by subtracting the number of words recalled in the delayed recall trial (trial 8) from those remembered at trial 5 (the ending trial of the learning part).

Measures of recall and recognition have typically been found to be sensitive to the memory deficits seen in AD. In a meta-analysis by Salmon and Bondi (2009), the delayed recall was most affected in the earliest stages of AD. According to Broadhouse et al. (2019) delayed recall is the most sensitive part to identify MCI. In a study by Estévez-González et al., (2003), both delayed recall and percentage of forgetting (75% or more) were the most useful parts of the RAVLT for predicting future AD in a group with subjective memory complaints. In another study (Welsh et al., 1992, delayed recall was considered the most effective way to differentiate healthy aging from early stages of AD, far superior to both immediate recall and recognition (as well as intrusion errors). Nevertheless, a recent meta-analysis by Weissberger et al. (2017) showed little or no difference in performance between immediate and delayed recall in an AD sample. The authors concluded, however, that neuropsychological assessment by word-list tests (like the RAVLT) has good to excellent sensitivity and specificity when using delayed (and immediate) recall, when it comes to distinguishing healthy controls from AD. De Simone et al., (2019), conducted a longitudinal study investigating memory performance among 80 participants with amnestic MCI. During three years of follow-ups, half of the participants converted to AD, while the other half remained unchanged. Both groups performed poorly overall compared to healthy controls. However, the stable MCI participants performed significantly better on recognition trials than those who converted, making recognition a better predictor than free recall for AD in an MCI population. In Hellström and colleagues’ study from 2012, there was a significant
difference in performance between healthy individuals and participants with iNPH on delayed recall both before and after shunt surgery.

**Diagnostic Challenges**

The diagnostic procedures of dementia offer challenges in interpreting symptoms; other signs of aging may complicate the evaluation. For instance, the cognitive decline in normal aging often affects the same domains as AD, complicating diagnostic decisions in an elderly population. However, in AD the loss of cognition and functioning soon becomes more extensive. Similarly, the core features of iNPH are frequently encountered in a healthy, elderly population. In a large Swedish study (N = 17,612) by Rausch et al. (2019), the prevalence of age-dependent symptoms was investigated in a sample of 65 to 84-year-olds. Nearly one-third of the females reported urine incontinence, whereas the corresponding number for men was one in five. Balance disturbance (i.e., falls) was reported in 9% of the men and 12% of the women. Since, per definition, the prevalence of such symptoms increases markedly with age, so does the magnitude of diagnostic uncertainty.

Neuroradiology constitutes an important part of the diagnostic procedure in both AD and iNPH, preferably by means of magnetic resonance imaging (MRI). Enlarged ventricles are seen in both iNPH and AD. However, the ventriculomegaly in iNPH is often more general and pronounced (Daouk et al., 2014), whereas the ventricular enlargement in AD is due to atrophy of surrounding tissue. In the diagnostic procedure of AD, biomarkers can comprise part of the investigation (McKhann et al., 2011), but findings considered to indicate AD pathology can also be found in the cerebrospinal
fluid of patients with iNPH. A theory presented by Silverberg et al. (2003) suggested that the same mechanisms that are involved in the pathology of AD could also be responsible for the cognitive decline in iNPH. Less effective clearance of harmful molecules (e.g. beta-amyloid and metabolic debris) leads to brain damage. Lower than expected levels of beta-amyloid in iNPH have also been observed in a study by Tarnaris et al. (2011), however, in that particular study, tau levels were also lower than normal (opposite to AD where tau is generally elevated). More recently, studies by Jeppsson et al. (2016 and 2019) have shown that iNPH is characterized by a pattern comprising low levels of soluble amyloid precursor protein (APP) α and β, amyloid β (isoforms 38, 40, 42), total tau (T-tau) and phosphorylated tau.

Aim of the Study

To summarize parts of what has been discussed hitherto, memory impairment in AD has been extensively studied and the decline in memory functioning is associated with the diseases’ impact on the hippocampus (Hyman et al, 1984) and other memory associated structures and circuits. The RAVLT is a commonly used test in the diagnostic procedure of AD and a valuable part of a differential diagnostic assessment (Salmon & Bondi, 2009; Bussé et al., 2017; Russo et al., 2017). Less is known about the impact of iNPH on cognition and memory. Most iNPH studies have focused on frontal lobe dysfunction, sometimes suggesting that it may be the cause of memory impairment (Picascia et al., 2015). When memory deficits do occur, they have been considered to be less severe than in AD (Picascia et al., 2016). However, the studies of iNPH have often been limited in size and few have studied learning, recall and recognition in the earliest
stages of the disease. Since it is a treatable disorder, and the outcome of treatment is dependent on early detection, diagnostic accuracy is of great importance.

This study aims to investigate verbal memory impairment, as measured by the RAVLT, in the earliest stages of dementia, caused by iNPH or AD. First, we sought to investigate performance between the clinical samples and healthy individuals. To make the comparisons relevant, only participants scoring well on the most frequently used measure of global cognition, the MMSE (scoring 27-30 points), were included. We then sought to investigate if these findings could be replicated when the global cognition dropped, i.e., among patients with MMSE scores ranging from 18 to 26 points. The main purpose was to investigate whether performance on the RAVLT can be used to distinguish AD from iNPH in mild and moderate dementia. Furthermore, if so, to identify specific parts or measures of the RAVLT to which clinical psychologists working with diagnostic assessments should pay the closest attention.

**Methods**

**Participants**

Data were collected from two ongoing studies at Sweden’s largest hospital, Sahlgrenska University Hospital (SU), in Gothenburg, Sweden. The hospital serves as the region’s specialized medical center. Participants with iNPH and HI were examined at the Neurological Clinic and AD patients at the Memory Clinic. The patients were assessed by a trained neuropsychologist, whereas psychology students, supervised by a neuropsychologist, assessed the majority of the healthy individuals. Specialized doctors
conducted the medical examinations (i.e., neurologists and psychiatrists). The iNPH patients also underwent physiotherapeutic examinations of gait and balance.

**AD**

The Gothenburg Mild Cognitive Impairment study (G-MCI) started at the Memory clinic in 1999. It aims to investigate early AD and vascular dementia (VAD) (Wallin et al., 2016). Patients included in the study are between the ages of 50 and 79 years (at inclusion) and have experienced cognitive impairment for at least six months. At baseline, 274 of the study's 664 participants were diagnosed with MCI and 195 with Subjective Cognitive Impairment (SCI). Dementia was diagnosed in 195 participants, among whom 81 were diagnosed with AD. The participants were examined once every two years, including a neuropsychological assessment. At the follow-ups, some of the participants' symptoms had progressed into dementia. The results on the RAVLT and MMSE included in this study were those acquired at the assessment generating an AD diagnosis. To avoid circularity, the RAVLT was not a part of the diagnostic procedure; instead, the diagnosis was based on MMSE, Clinical Dementia Rating (CDR), Stepwise Comparative Status Analysis (STEP) and Investigation of Flexibility (iFLEX, a short form of the executive interview EXIT) (Wallin et al., 2016). Patients with prior head trauma, stroke, brain tumor or other neurological illness, as well as patients with substance abuse, current psychiatric disorders, or somatic disorders that are known to cause cognitive impairment, were excluded. Out of the 151 participants that so far have been diagnosed with AD, 35 were excluded due to missing or insufficient data on the RAVLT protocols, reducing the number of participants to 116. In the matching procedure (age,
sex, years of education and scores on the MMSE), further reducing the number of participants to 84.

**iNPH**

Since 1978, 738 patients with iNPH have been investigated at the Neurological Clinic of the Sahlgrenska University Hospital and consecutively included in the iNPH Research Database. The patients were diagnosed with possible or probable iNPH according to international guidelines published in 2005 (Relkin et al., 2005). (The diagnostic criteria used before the guideline publication by Relkin et al. were virtually the same, i.e., based on the same clinical and imaging findings). A team including neurologists, neurosurgeons, a neuropsychologist and a physiotherapist diagnosed the patients based on MRI images and clinical signs of gait impairment (including reduced speed and stride length), urinary incontinence, balance disturbance and cognitive problems. From January 1999 to December 2018, 464 participants have undergone at least two neuropsychological assessments, one pre and one post-surgery (including the Grooved Pegboard test, the Stroop test and the RAVLT). The results of the neuropsychological assessment have not been part of the diagnostic algorithm but served descriptive purposes only. In the case of multiple assessments before surgery, the pre-assessment closest to surgery was used in this study.

The matching between the initial iNPH sample, comprising 186 patients with pre- and postoperative assessments, conducted from January 2014 to December 2018, with the smaller AD sample, was unsatisfactory. The iNPH participants were significantly older than the AD patients and the proportion of men was significantly larger. In order to compensate for this, a wider time frame, from 1999, was applied,
where women and participants younger than 70 years old were actively sought for and included, increasing the number of eligible iNPH participants to 218. Nine were excluded due to MMSE scores below 18 points. In order to match the AD sample, the number was further reduced to 84 participants.

**HI**

The Neurological Clinic recruited a group of 201 HI from senior associations and church organizations. The HI voluntarily underwent a neuropsychological assessment, including a full administration of the RAVLT. Participants with severe psychiatric disorders and neurological disorders were excluded. However, subjects with mild depression or anxiety (untreated or treated with SSRI or minor tranquilizers) and well-treated stable medical conditions (diabetes, hypertension and hypothyreosis) were included (Hellström et al., 2007). For this study, 30 participants were matched with regard to age, sex and years of education to the clinical sample of High-MMSE performers.

**Ethics**

The regional ethics committee initially approved the G-MCI study in 1999 (diary number: L091-99, March 15, 1999). After changes, a new application was approved in 2011 (diary number: T479-11, June 8, 2011). The iNPH study was approved in 2014 (diary number: 328-14, May 14, 2014). All participants gave informed consent. The healthy individuals were recruited via senior associations and churches. They all actively volunteered to be included after receiving information in the form of public lectures concerning the study.
**Exclusion and Matching**

The MMSE was used to assess global cognition and the degree of dementia. We used the spans suggested by Marioni et al. (2011), excluding participants scoring 17 points or less. The span between 18 and 22 points (Moderate impairment according to Marioni et al.) and scores from 23 to 26 points (Slight impairment), made out the subgroup we refer to as Medium-MMSE performers. Patients scoring between 27 and 30 points (by Marioni et al. called No impairment) will be referred to as High-MMSE performers.

Initial analysis revealed that the three groups (HI, AD and iNPH) differed substantially on demographic variables known to influence results on the RAVLT scores (i.e., sex, age and years of education) (Sundermann et al., 2017; van der Elst et al., 2004). Further analysis showed that correcting for these differences using analysis of covariance (ANCOVA) was not feasible because the data did not fit the assumptions of ANCOVA. To make the samples comparable on demographic characteristics, we opted to match the participants on multiple variables with propensity score matching (Stuart, 2010). High-MMSE performers and HI were matched on age, sex and education length; medium-MMSE performers were additionally matched on MMSE results.

The flow chart of Figure 1 illustrates the process of inclusions, exclusions and matching procedures. The demographic characteristics of the resultant study groups, after propensity score matching, are presented in Table 1. For a more in-depth description of the matching process, see the section on “Statistical analysis” below.
Figure 1

Participant Inclusion Flowchart
Table 1

Demographic Variables of Participants in Comparisons Between AD and iNPH.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>HI (n=30)</th>
<th>iNPH (n=30)</th>
<th>AD (n=30)</th>
<th>iNPH (n=54)</th>
<th>AD (n=54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>69 (51-83)</td>
<td>71 (53-82)</td>
<td>68.5 (57-77)</td>
<td>70 (56-85)</td>
<td>69 (52-80)</td>
</tr>
<tr>
<td>Education</td>
<td>10 (7-22)</td>
<td>9.75 (7-18)</td>
<td>10 (4-17)</td>
<td>10 (6-22)</td>
<td>10 (6-21)</td>
</tr>
<tr>
<td>MMSE</td>
<td>-</td>
<td>28 (27-30)</td>
<td>27 (27-30)</td>
<td>24 (20-26)</td>
<td>24 (20-26)</td>
</tr>
<tr>
<td>Sex, % (m/f)</td>
<td>63.3/36.7</td>
<td>63.3/36.7</td>
<td>56.7/43.3</td>
<td>59.3/40.7</td>
<td>61.1/38.9</td>
</tr>
</tbody>
</table>

Note. Variables presented as median values and interquartile ranges unless otherwise noted.

RAVLT variables

The analyses were divided into two main categories, Learning and Recall/Recognition. To assess learning, we used Learning total (the sum of trial 1 - 5), Learning delta max (the difference between the best trial and trial 1) and Serial position effects. Learning delta might seem very similar to measure as Learning total. However, according to Vakil and Blachstein (1993), Learning total represents the process of acquisition and Learning delta the process of retention (which in turn can be divided into storage and retrieval). The Serial position effects were analyzed in measures of primacy and recency effects. The 15-word list was divided into a primacy, a middle and a recency region. In previous studies, different divisions of the RAVLT have been suggested; an equally sized group of words (5-5-5; Ricci et al., 2012), smaller primacy and recency regions and a larger middle region (4-7-4; Bussè et al., 2017, Hermann et al., 1996) or smaller primacy and middle regions and a larger recency region (4-4-7; Murdock, 1962). For this study, we chose the division of 4-7-4.
To assess recall we used the performance on Delayed recall and percentage of Forgetting (the difference between trial 5 and trial 8). Note that immediate recall was not used, due to earlier studies’ conclusions of the superiority of delayed recall over immediate recall (Salmon & Bastin, 2009; Estévez-González et al., 2003; Welsh et al., 1992. The Recognition part of the test was used as a measure of recognition. The recognition task consisted of fifteen forced choices between three alternatives; the target word and two distractor words begin with the same letter and have the same number of syllables as the target word.

Statistical Analysis

As the HI, iNPH and AD groups differed substantially on demographic covariates known to affect RAVLT performance (Sundermann et al., 2017; van der Elst et al., 2005), we initially attempted to adjust for these differences using ANCOVA. However, using the ANCOVA was not possible since our data did not fit all acquired assumptions; more specifically, the regression between MMSE scores and age differed between groups (thus violating the assumption of homogeneity of regression). We ultimately decided to create matched groups from the original samples, sacrificing sample size to gain demographic balance between groups. Propensity-score matching was used to create these groups.

Briefly, a participant’s propensity score indicates the probability that it belongs to one of the two matched groups and is calculated from the observed covariates. By selecting pairs of participants with similar propensity scores from each group, the groups are balanced with respect to the covariates (Stuart, 2010). There are several different ways to select the pairwise matches; in order to maximize the resulting group
sizes while minimizing group differences on covariates, we used the nearest matching with a caliper size of 0.2 as recommended by Austin (2011). In order to match HI with high-MMSE performers, we first matched iNPH and AD high-MMSE performers and then matched the resulting iNPH group with HI.

In order to evaluate the equivalence of covariates between groups, testing for non-significance on t-tests or the like, while common, is not recommended (Kover & Atwood, 2013). To see if two matched groups are balanced on covariates, calculating the standardized difference of means of the groups for the propensity score and each covariate is preferred. The absolute value of this measure should not exceed 0.25 (Stuart, 2010). In our matched groups, the standardized difference of means ranged between 0 and 0.22, see Table 2.

### Table 2

*Standardized Difference of Means for Covariates in Matched Groups*

<table>
<thead>
<tr>
<th>Covariate</th>
<th>High-MMSE performers</th>
<th>High-MMSE iNPH and HI</th>
<th>Medium-MMSE performers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propensity score</td>
<td>0.07</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td>Age</td>
<td>-0.22</td>
<td>-0.12</td>
<td>-0.13</td>
</tr>
<tr>
<td>Education</td>
<td>0.05</td>
<td>-0.05</td>
<td>-0.04</td>
</tr>
<tr>
<td>Sex (male)</td>
<td>0.03</td>
<td>-0.15</td>
<td>-0.04</td>
</tr>
<tr>
<td>MMSE score</td>
<td>-</td>
<td>-</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

*Note.* MMSE scores were not used for matching high-MMSE performers, why the standardized difference of means was not calculated for those groups.
Since independent variables had non-symmetrical distributions, non-parametric methods were used for all analyses. Kruskal-Wallis one-way analysis of variance ($H$) was used when comparing three groups (i.e. iNPH and AD high-MMSE performers vs. HI) together with Dunn’s test of multiple comparisons using rank sums for pairwise post hoc comparisons. When comparing two independent groups (i.e. iNPH and AD medium-MMSE performers), the Wilcoxon-Mann-Whitney test ($U$) was used. The Spearman rank correlation coefficient ($\rho$) was used to estimate the relation between variables. Holm correction was used to adjust for multiple comparisons.

Effect sizes were investigated using Cliff’s delta ($d$) and Eta-squared ($\eta^2$) for two-way and three-way comparisons, respectively. The $d$ statistic was originally proposed as a way to test differences on ordinal variables between groups and shows the amount of overlap between groups. It is calculated as the probability that the scores in the first group are higher than the scores in the second group, subtracted by the reverse probability (Cliff, 1993). However, $d$ can also be used as a representation of effect size (Romano et al., 2006). The value of $d$ spans between -1 and 1; a value of -1 indicates that the scores in the second group are all larger than the scores in the first group, while a value of 1 means that the scores in the first group are all larger than the scores in the second group. A value of 0 means that the groups overlap completely (Cliff, 1993). Comparing Cliff’s delta to Cohen’s $d$, Romano and colleagues (2006) have shown that Cliff’s delta can be interpreted as an effect size, where $|d|<0.147$ should be considered "negligible", $|d|<0.33$ as "small", $|d|<0.474$ as "medium", and values beyond 0.474 as "large".
All statistical analyses were done using the R software version 3.6.2 (R Core Team, 2019). Propensity score matching was done using the MatchIt package (Ho et al., 2011).

**Results**

Kruskal-Wallis comparison between HI and High-MMSE performing AD and iNPH patients showed significant differences in all RAVLT variables with very small to small effect sizes; see Table 3. Dunn post hoc comparisons showed that HI performed significantly better on all variables compared to AD participants. INPH participants’ scores typically lay somewhere between HI and AD scores, with only some variables showing significant differences either way (as can also be seen in Figure 2). Comparisons of iNPH and AD participants show a clear pattern where scores in the Recall/recognition category are all significantly different, while differences on all but one score in the Learning category fail to reach significance.
Table 3

*Kruskal-Wallis ANOVA and Dunn Post Hoc Comparisons of Test Results Between HI and High-MMSE AD and iNPH Patients*

<table>
<thead>
<tr>
<th>RAVLT var</th>
<th>HI</th>
<th>iNPH</th>
<th>AD</th>
<th>p</th>
<th>HI vs iNPH</th>
<th>HI vs AD</th>
<th>iNPH vs AD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT</td>
<td>45 (38-51)</td>
<td>33 (23-38)</td>
<td>25.5 (21-29)</td>
<td>&lt;.001</td>
<td>0.38</td>
<td>-0.59 ***</td>
<td>-0.87 ***</td>
</tr>
<tr>
<td>LD</td>
<td>6 (5-8)</td>
<td>5 (3-7)</td>
<td>4 (2-5)</td>
<td>.001</td>
<td>0.15</td>
<td>-0.23 ns</td>
<td>-0.58 ***</td>
</tr>
<tr>
<td>PI</td>
<td>70 (60-84)</td>
<td>55 (36-65)</td>
<td>38 (26-50)</td>
<td>&lt;.001</td>
<td>0.27</td>
<td>-0.44 *</td>
<td>-0.72 ***</td>
</tr>
<tr>
<td>RI</td>
<td>65 (56-75)</td>
<td>60 (46-69)</td>
<td>55 (35-65)</td>
<td>.025</td>
<td>0.06</td>
<td>-0.22 ns</td>
<td>-0.39 *</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Recall/recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>8 (6-11)</td>
<td>5 (2-7)</td>
<td>2 (0-3)</td>
<td>&lt;.001</td>
<td>0.42</td>
<td>-0.49 **</td>
<td>-0.89 ***</td>
</tr>
<tr>
<td>FI</td>
<td>19 (8-36)</td>
<td>37 (22-74)</td>
<td>68 (60-97)</td>
<td>&lt;.001</td>
<td>0.34</td>
<td>0.34 *</td>
<td>0.83 ***</td>
</tr>
<tr>
<td>Rec</td>
<td>15 (15-15)</td>
<td>15 (14-15)</td>
<td>11 (9-14)</td>
<td>&lt;.001</td>
<td>0.30</td>
<td>-0.23 ns</td>
<td>-0.79 ***</td>
</tr>
</tbody>
</table>

*Note. LT = Learning total, LD = Learning delta, PI = Primacy index, RI = Recency index, DR = Delayed recall, FI = Forgetting index, Rec = Recognition. Variable scores are presented as median values and interquartile ranges. Post hoc comparisons are presented as effect size (Cliff’s delta) and statistical significance.

*=p<.05, **=p<.01, ***p<.001

Wilcoxon-Mann-Whitney comparisons of medium-MMSE performers’ scores reinforce the pattern seen with high-MMSE performers, as no scores in the Learning category reach significance while all scores in the Recall/Recognition category remain statistically significant, see Table 4. Comparing the scores also shows that the differences are smaller for the medium-MMSE performers compared to high-MMSE performers, especially in the Learning category (see also Figure 3 for a visual representation).
USING THE RAVLT TO COMPARE MEMORY IN INPH AND AD

Table 4

Wilcoxon-Mann-Whitney Comparison of Test Results Between Medium-MMSE Performers

<table>
<thead>
<tr>
<th>RAVLT variable</th>
<th>iNPH</th>
<th>AD</th>
<th>p</th>
<th>Cliff’s delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning total</td>
<td>25 (20-29)</td>
<td>26 (21-29)</td>
<td>.59</td>
<td>0.06</td>
</tr>
<tr>
<td>Learning delta</td>
<td>3 (2-4)</td>
<td>4 (2-5)</td>
<td>.46</td>
<td>0.17</td>
</tr>
<tr>
<td>Primacy index</td>
<td>40 (30-55)</td>
<td>30 (20-50)</td>
<td>.46</td>
<td>-0.15</td>
</tr>
<tr>
<td>Recency index</td>
<td>50 (35-65)</td>
<td>55 (40-75)</td>
<td>.46</td>
<td>0.18</td>
</tr>
<tr>
<td>Recall/Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed recall</td>
<td>3 (0-5)</td>
<td>1 (0-2)</td>
<td>.02</td>
<td>-0.31</td>
</tr>
<tr>
<td>Forgetting index</td>
<td>59 (33-100)</td>
<td>88 (62-100)</td>
<td>.02</td>
<td>0.32</td>
</tr>
<tr>
<td>Recognition</td>
<td>14 (12-15)</td>
<td>11 (9-14)</td>
<td>.001</td>
<td>-0.43</td>
</tr>
</tbody>
</table>

Note. Variable scores are presented as median values and interquartile ranges.

*=p<.05, **=p<.01

The further comparison suggests that this closing of the gap is mainly due to medium-MMSE iNPH participants performing worse than their high-MMSE counterparts, while the AD participants in the two groups perform similarly. Table 5 shows the correlation between MMSE scores and RAVLT scores within the clinical groups and confirms that there is a significant relation between MMSE and RAVLT scores within the iNPH group. In contrast, for the AD group, only the Recall category scores show a significant correlation with MMSE scores. So, for the Learning category and Recognition scores, there is no clear deterioration for the AD participants as their MMSE performance worsens.
Table 5

*Spearman Rank Correlation Coefficients Between MMSE Scores and RAVLT Variables*

<table>
<thead>
<tr>
<th>Variable</th>
<th>iNPH (n=84)</th>
<th>AD (n=84)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learning</td>
<td>Recall/Recognition</td>
</tr>
<tr>
<td>Learning score</td>
<td>0.34 (**)</td>
<td>0.11 (ns)</td>
</tr>
<tr>
<td>Learning delta</td>
<td>0.31 (**)</td>
<td>-0.04 (ns)</td>
</tr>
<tr>
<td>Primacy index</td>
<td>0.25 (*)</td>
<td>0.17 (ns)</td>
</tr>
<tr>
<td>Recency index</td>
<td>0.25 (*)</td>
<td>-0.09 (ns)</td>
</tr>
<tr>
<td>Delayed recall score</td>
<td>0.40 (***)</td>
<td>0.32 (**)</td>
</tr>
<tr>
<td>Forgetting index</td>
<td>-0.32 (**)</td>
<td>-0.31 (**)</td>
</tr>
<tr>
<td>Recognition score</td>
<td>0.34 (**)</td>
<td>0.14 (ns)</td>
</tr>
</tbody>
</table>

*Note.* *=p<.05, **=p<.01, ***=p<.001

Concerning serial position effects, the common pattern of a clear recency effect but lack of primacy effect in AD patients can also be seen in our sample, regardless of MMSE performance. However, while the iNPH participants show a clear primacy and recency effect in the high-MMSE group, both those effects diminish for medium-MMSE performers, to the point where iNPH participants have a smaller recency effect than AD participants. See also Figures 4 and 5.
Figure 2

Performance on RAVLT Trials by HI and High-MMSE Performers
Figure 3

Performance on RAVLT Trials by Medium-MMSE Performers
Figure 4

*Serial Position Graph for HI and High-MMSE Performers*
Discussion

The main findings of this study include impaired memory performance in participants with iNPH and AD, even in the earliest stages of dementia and before the MMSE indicates a decline in global cognition. There were some differences between the clinical samples; the iNPH participants, despite similar results as the AD participants on measures of Learning, scored significantly better on Recall category variables and had normal performances on Recognition. The typical AD pattern in serial position effects, clear recency but a lack of primacy effect, was also observed among the iNPH participants with Medium-MMSE scores.
An assumption was made that the High-MMSE performers would outperform the Medium-MMSE performers on the RAVLT. While this was true for the iNPH participants, the AD high-MMSE performers scored as poorly as the medium-MMSE performers on Learning and Recognition variables, achieving better results only on Recall scores. Despite having an MMSE performance that most would regard as “No impairment”, the average result on Delayed Recall among High-MMSE performers with AD was two out of 15 words; for Medium-MMSE performers, the result dropped to one out of 15 words. A recent study by Josefsson and colleagues (2019) showed a decline in memory performance 10 - 15 years before AD diagnosis. Performance on the MMSE, however, did not differ to the same extent during this period in time. Josefsson and co-authors suggest that the MMSE lacks the sensitivity to detect cognitive decline in preclinical dementia. Our findings support this notion by showing that the MMSE fails to identify a substantial decline in memory performance in AD. This finding is of importance since, in a clinical setting, MMSE is standard when screening for dementia.

**Learning**

Compared to the HI, scores on Learning total and Learning delta were significantly lower in both clinical samples of the High-MMSE performers. However, the performances did not differ between the AD and the iNPH participants, a result replicated in the group of Medium-MMSE performance. Regarding learning in the early stage of iNPH, our findings agree with other studies which have also found learning to be impaired in iNPH (Saito et al., 2011) and differ from others, which suggest that learning deficits appear only in later stages (Picascia et al., 2016).
In the analyses of the serial position effects, there was a difference among the High-MMSE performers where the AD participants lacked a primacy effect, which is consistent with most literature on AD, whereas the HI and participants with iNPH showed both primacy and recency effects. In the Medium-MMSE performers, however, there was no difference between the clinical samples in serial position effects. The typical AD pattern of a smaller primacy effect and a clearer recency effect also occurred in the iNPH sample. This pattern is not consistent with the findings of Sindorio et al. (2017), where iNPH participants (n = 12, MMSE mean 18.8), showed both primacy and recency effects. The clinical relevance of our finding is a suggestion not to exclude the possibility of iNPH when a typical AD pattern is observed.

Recall/Recognition

On the measures of Recall (i.e. Delayed recall and Forgetting index), there were significant differences between the HI and both groups of High-MMSE performers; the HI outperformed both clinical groups while the iNPH participants performed better than the AD participants. These findings were replicated among the Medium-MMSE participants. Previous studies have failed to identify memory impairment in the earlier stages of dementia caused by iNPH (Ogino et al., 2006; Picascia et al., 2016). Even if the memory impairment is not as pronounced as in AD, the difference compared to HI is significant.

On Recognition, the AD patients were outperformed by both HI and iNPH patients in the High-MMSE comparison, and subsequently by the iNPH patients, in the analyses of Medium-MMSE performers. The differences were statistically significant and the effect sizes were large for both types of comparisons. The preserved ability for
recognition in iNPH is consistent with earlier studies and constitutes an essential difference between iNPH and AD (Picascia et al., 2015).

**Limitations**

Our initial plan was to use two clinical samples regarding memory performance. To make the samples comparable on demographic variables known to affect RAVLT performance, we had to adapt the iNPH sample to match the AD participants. This adaptation makes our iNPH sample less representative of the naturally occurring clinical population of iNPH patients.

**Conclusion**

In our sample, memory impairment occurs in the earliest stages of dementia, whether caused by iNPH or AD. The memory impairment is greater among AD patients; however, the impairment compared to HI is substantial also among participants with iNPH. Despite MMSE scores in the normal range of 27-30 points, both clinical samples struggled with all parts of the RAVLT, except Recognition for the iNPH participants. The interpretation of this finding is that memory in AD is affected to a great extent, before global cognition is affected, and that the MMSE lacks the sensitivity to detect memory decline in iNPH, and even more so in AD. The RAVLT is a far more sensitive test to detect dementia, where Recognition seems to be the most useful part of the test in discriminating between AD and iNPH.
References


https://doi.org/10.1016/j.neuroimage.2014.09.056


