The neuropsychology of iNPH: Findings and evaluation of tests in the European multicentre study

Per Hellström a,+, 1, Petra Klinge b, 1, Jos Tans c, 1, Carsten Wikkelso a, 1

a Hydrocephalus Research Unit, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Sweden
b Department of Neurosurgery, Rhode Island Hospital, Brown Medical School, Providence, RI, USA
c Department of Neurology, Medical Centre Haaglanden, The Hague, Netherlands

**Abstract**

Objective: Neuropsychological dysfunction is common in patients with idiopathic normal pressure hydrocephalus (iNPH). Shunt treatment is beneficial, some patients reaching complete or almost complete recovery, while others show only minor improvement. We aimed to assess the efficacy of a small selection of well characterized and sensitive neuropsychological tests in the context of the European multicentre study on iNPH (Eu-INPH).

Methods: One hundred and forty-two iNPH patients included in Eu-INPH were tested with the Rey Auditory Verbal Learning Test (RAVLT), the Grooved Pegboard and the Stroop test before and after three and twelve months of treatment with a ventriculoperitoneal shunt. Their performance was compared to that of 108 healthy individuals (HI).

Results: iNPH patients performed significantly worse than HI on all of the neuropsychological measures at entry. The discriminative capacities of the eight variables were similar, with areas under the curve (AUC; ROC analysis) ranging between .86 (Delayed Recall) and .95 (Grooved Pegboard). The most usable test was RAVLT (Learning and Delayed Recall), administered to ≥90% of the patients at all occasions. However, the Grooved Pegboard and the Stroop test were more sensitive to treatment effects.

Conclusion: The three neuropsychological tests used in the Eu-INPH are expedient, highly diagnostically discriminative, and well suited to evaluate changes following shunt treatment.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Idiopathic normal pressure hydrocephalus (iNPH) is a condition of disturbed cerebrospinal fluid (CSF) dynamics, the causes of which are not known. The ventricular system is enlarged but the intracranial pressure is within normal limits. iNPH is characterized by a slowly progressive impairment of gait and balance, cognitive deterioration, and urinary incontinence.

Characterizations of the neuropsychological functions of patients with iNPH suggest that several areas are compromised (information processing speed, attention, memory and learning, visuo-spatial functions, dexterity, pragmatic aspects of language, abstract thought, working memory, executive functions) [1–7]. In a recent study we showed that iNPH patients perform worse than age matched HI on tests of memory and learning, dexterity, reaction time and executive functions, that the deficits are strongly interrelated, associated with neurological signs, and aggravated by the presence of risk factors for cerebrovascular disease [3]. While examining the changes after three months of treatment we found significant, yet incomplete recovery of neuropsychological functions. The presence of cerebrovascular risk factors influenced the performance both pre- and postoperatively, but it did not reduce the effect of treatment [8].

The recently conducted prospective European multicentre study on iNPH (the Eu-INPH study) aimed to assess the predictive value for the outcome of shunting of the CSF tap test, resistance to outflow of CSF, and compliance of the CSF space. The study required a small number of tests, covering different domains, available in different languages and possible to administrate by different professions under the supervision of a neuropsychologist. Hence, three well known, valid and reliable neuropsychological tests assessing iNPH relevant functions were selected; the Grooved Pegboard, the Rey Auditory Verbal Learning Test (RAVLT) and the Stroop test. These tests had also been found to be promising regarding efficacy in previous studies [3, 8].

In the present article we aimed to prospectively assess this small selection of neuropsychological tests in terms of their...
expediency, discriminability, and ability to depict changes following shunt treatment.

2. Patients and methods

Thirteen centres in 9 European countries (Belgium, the Czech Republic, Denmark, Germany, Hungary, Italy, the Netherlands, Spain, and Sweden) included 142 patients with the diagnosis of iNPH between September 2004 and January 2008. One hundred and forty-one patients received ventriculoperitoneal shunts (Codman adjustable) and were re-examined three months (n = 123) and one year (n = 115) after surgery.

The following diagnostic criteria for iNPH were used:

Clinical findings: The presence of a gradually developed gait disturbance was mandatory. Symptoms of mental deterioration or incontinence were not required.

Radiological findings: A symmetrical communicating quadriventricular enlargement with an Evan’s index ≥ 30 was mandatory. Single cortical infarcts, moderate cortical atrophy and moderate to severe leukoaraiosis were accepted.

Exclusion criteria: Patients with obstructive, secondary and/or high pressure (>18 mm Hg) hydrocephalus were excluded. Further, patients who otherwise fulfilled the inclusion criteria, but who were unwilling to participate, had a restricted life-expectancy, had contra-indications for surgery, or were unable to participate fully in the tests or investigations were also excluded.

Apart from the observation of signs necessary for the diagnosis of INPH and the administration of the MMSE, the neurological examination at study entry also included grading of gait, balance and bladder function. Further, risk factors for and indicators of cerebrovascular disease (hypertension, smoking, diabetes, cardiac disease, previous strokes, and peripheral vascular disease) were registered as present or absent.

A control group of 108 Swedish HI (aged 50–87), recruited from senior associations and church communities, who volunteered to undergo neuropsychological testing during the spring of 2006, was used for comparison. Only individuals without neurological or severe psychiatric disorders were included in the control group, but stable medical conditions (e.g., successfully treated diabetes, hypertension, or hypothyreosis), and/or minor psychiatric disorders (i.e., mild symptoms of anxiety or depression and/or treatment with SSRIs or minor tranquilizers reported during interview) were allowed. Demographic descriptions of the patients and the HI are presented in Table 1.

2.1. Neuropsychological methods

The following tests were administered:

Grooved Pegboard (Lafayette Instrument Co.), measuring manual dexterity as time to fit 25 pegs into holes with randomly positioned slots measured as time in seconds for the Dominant and Non-dominant hand and the sum of both;

The Rey Auditory Verbal Learning Test (RAVLT; [9]) measuring verbal learning and memory measured as the total number of recalled words over five learning trials (Learning) and the number of recalled words after 30 min of distraction (Delayed Recall);

The Stroop test [10] measuring Color Naming speed and Response Selection which requires the patient to name the colors (blue, red, green or yellow) of 100 rectangles as fast as possible, and then to name the printed color of 100 incongruent color words (e.g., the word blue printed in red). The scores are the number of seconds for each trial (Color Naming and Response Selection, respectively) and the increase in time from the first to the second trial (Increment).

2.2. Statistical analysis

Nonparametric statistical methods were used for all analyses to adhere to the level of measurement and/or distributional properties. Thus, the Spearman rank correlation coefficient (ρ) was used to estimate associations between variables, whereas the Mann–Whitney U-test and Fishers exact test (or the χ² test) were used for comparisons between groups regarding scores and proportions, respectively.

Receiver-operating characteristic (ROC) curves were used to evaluate the discriminative capacities (INPH versus HI) of the neuropsychological tests. The area under each ROC curve (the AUC value) reflects the probability that a randomly selected INPH patient will perform at a lower level than a randomly selected HI on the specific test [11,12]. Statistical analyses were performed with SPSS 17.0 for Windows (SPSS, Inc., Chicago, IL).

3. Results

3.1. Expediency

The most usable test was RAVLT (both Learning and Delayed Recall) which could be administered to 94, 90 and 92% of the patients at the different occasions. The least usable test was the Stroop test (Response Selection) which was administered to 68, 76 and 70% of the patients at the different examinations. The time needed to administer the three tests is approximately 20 min including the time needed for the Delayed Recall trial after 30 min of distracting activity (e.g., neurological examination).

The numbers of patients who received scores on all of the three tests (but not necessarily on all of the variables) was 108 (out of 142 at entry), 101 (out of 123 at 3 months), and 94 (out of the 115 returning after one year). The corresponding numbers of patients who did not receive scores on any of the tests were 3, 6 and 5.

3.2. Sensitivity and specificity

The patients with iNPH performed significantly worse than HI on all of the neuropsychological measures at entry (Table 2, columns 2 and 3).

Prior to surgery, the iNPH patients used twice the time needed for HI in both of the Stroop’s tasks and the Grooved Pegboard and scored half of what HI did in memory tests. The difference between iNPH patients and HI was largest in Increment on the Stroop test and Delayed Recall where they only scored 25% of what the HI did.

The discriminative capacities of the eight variables were similar, the AUC values ranging between the lowest .86 (Delayed Recall) and the highest .95 (Grooved Pegboard, Sum) as illustrated by the ROC curves in Fig. 1.

3.3. Ability to detect changes

The iNPH patients were significantly improved in all 8 variables 3 and 12 months after surgery but still performed below HI. No significant changes occurred between the two follow up examinations (Table 2, columns 3–5). The proportions of patients who improved more than 1 sd of the distribution of the HI ranged between 62%

Table 1

Demographic characteristics of iNPH patients and healthy individuals (HI).

<table>
<thead>
<tr>
<th>Demographic variables</th>
<th>iNPH (n = 142)</th>
<th>HI (n = 108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range)</td>
<td>72.5 (30–87)</td>
<td>72 (50–87)</td>
</tr>
<tr>
<td>Sex, male/female (%)</td>
<td>73/69 (51/49)</td>
<td>52/56 (48/52)</td>
</tr>
<tr>
<td>Education, median (range)</td>
<td>10 (3–20)</td>
<td>12.5 (5–22)</td>
</tr>
<tr>
<td>MMSE, median (range)</td>
<td>25 (9–30)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2
Neuropsychological performances of healthy individuals and iNPH patients before and after shunt surgery.

<table>
<thead>
<tr>
<th>Neuropsychological test results Md (range)</th>
<th>Healthy individuals</th>
<th>iNPH patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entry</td>
<td>3 months</td>
</tr>
<tr>
<td>Grooved Pegboard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant hand</td>
<td>80 (50–130)</td>
<td>166 (59–710)*</td>
</tr>
<tr>
<td>Non-dominant hand</td>
<td>90 (58–187)</td>
<td>184 (64–875)*</td>
</tr>
<tr>
<td>Sum of both hands</td>
<td>171 (110–258)</td>
<td>333 (129–1408)*</td>
</tr>
<tr>
<td>Rey AVLT Learning</td>
<td>43 (21–65)</td>
<td>22 (4–54)*</td>
</tr>
<tr>
<td>Delayed Recall</td>
<td>8 (0–15)</td>
<td>2 (0–13)*</td>
</tr>
<tr>
<td>Stroop task, s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color Naming</td>
<td>67 (43–133)</td>
<td>119 (56–535)*</td>
</tr>
<tr>
<td>Response selection</td>
<td>127 (77–252)</td>
<td>260 (80–840)*</td>
</tr>
<tr>
<td>Increment</td>
<td>56 (25–152)</td>
<td>142 (64–681)*</td>
</tr>
</tbody>
</table>

* Significantly impaired in comparison to HI (p < .0001).

b Significantly improved in comparison to entry (p < .0001).

c Significantly improved in comparison to entry (p < .01).

3.4. Associations between changes in different tests

There were weak but significant correlations between changes between entry and 3 months examination in Grooved Pegboard sum and Grooved Pegboard (ρ = .25, p = .03), Rey AVLT learning and Delayed Recall (ρ = .30, p = .002) and between Grooved Pegboard and Rey Delayed Recall (ρ = .42, p = .0003) and between Rey Delayed Recall (ρ = .48, p = .0001) otherwise no significant correlations were found.

3.5. Associations between neuropsychological performance and other variables

Age was weakly associated with pegboard at entry (ρ = -.26, p < .05), with the Stroop Color Naming at the one year examination (ρ = -.20, p < .05) and with the Stroop Response Selection at both postoperative examinations (ρ = -.27, p < .05 after three months, and ρ = -.37, p < .01 after one year). Age was weakly negatively associated with the change from entry to one year on RAVLT Learning (Spearman r = -.304, p = .002) but not with the changes of the other variables.

Education was weakly associated with performance on the Pegboard sum and the Stroop Color Naming at all occasions, and with the Stroop Response Selection at entry and after three months (r ranging from .21 p < .05 to .36 p < .01). Education was not correlated with the magnitude of change following treatment on any of the neuropsychological variables.

Female patients performed better than male patients on the RAVLT Delayed Recall at the two follow up examinations (Mann–Whitney, Z = −2.210, p = .027 after three months and Z = −1.961, p = .050 after one year). Correspondingly, women improved more than men on this single variable (Mann–Whitney, Z = −2.032, p = .042 from entry to three months and Z = −2.283 p = .022 from entry to one year). Apart from these findings there were no associations between sex and neuropsychological performance.

The presence of risk factors for cerebrovascular disease or previous cerebrovascular incidents (smoking, hypertension, diabetes, cardiac disease, peripheral vascular disease and previous stroke, singular or in combinations) had a negative impact on the performance on the three Grooved Pegboard variables (e.g.,...
Sum of both hands; Mann–Whitney, $Z = -2.623$, $p = .009$ and the Stroop Response Selection and Increment ($Z = -2.960$, $p = .003$ and $Z = -2.780$, $p = .005$) at entry, but no other effects on neuropsychological performance at any of the occasions. The presence of these factors (again, alone or in combinations) had no statistically significant impact on the magnitude of change in neuropsychological performance following surgery. This means that the changes due to treatment presented in Table 2 still remain statistically significant when patients without vascular risk factors are excluded from the analyses.

4. Discussion

The aim of this study was to evaluate the efficacy of the 3 neuropsychological tests used in the Eu-iNPH study in terms of expediency (referring to the appropriateness of a test for use within the specific clinical population and the time for administration), sensitivity, specificity, and ability to detect changes following treatment.

The 3 neuropsychological tests also constitute the cognitive part of the new iNPH scale consisting of mostly continuous variables, also measuring gait, balance and bladder functions. This scale is intended to provide a more fine graded assessment of features of iNPH than previously used ordinal/categorical scales in this field. A manuscript presenting the scale in detail is underway (manuscript submitted to Acta Neurologica Scandinavica).

The tests varied with regard to usability, RAVLT being the most frequently administered test and the Stroop Response selection being the least often administered subtask. This part of the Stroop test is the most cognitively demanding of the tasks and the low administration rate reflects that many iNPH patients lack the executive skill needed to achieve a valid score. Albeit qualitatively revealing (none of HI failed in this manner), it renders the test less expedient, since it can be expected to provide less data for quantification of severity and change.

All 3 selected tests proved to be sensitive and specific. They all showed significant differences between HI and iNPH patients, not only at baseline but also after three and twelve months of treatment. Admittedly, the Grooved Pegboard test was the most discriminative test at baseline with an AUC value of .95, but all tests had values exceeding .85 (values $>.80$–.90 are considered to be good, whereas values $>.90$ are considered excellent). These results are in close agreement with the results of previous AUC calculations on data from a different iNPH group in comparison to the same HI [3].

Interestingly, there were large differences between the tests regarding the proportion of patients who were normal before and after surgery (with a limit for normality set to within 1.5 sd from the mean for an appropriate normative sample). These differences partly depend on the distributional properties of the specific tests, but the main interpretation is that the motor impairment and the mental speed/executive deficits due to iNPH are more severe and salient than the memory disturbances. Hence the distributions of iNPH patients and those of HI for the former two are quite distinct, whereas there is a considerable overlap at the distributions of scores on Delayed Recall.

All of the tests were able to detect changes at group level following three and twelve months of treatment, as evidenced by statistically significant changes in comparison to baseline. No significant differences were found between the two follow up examinations. Findings in a recent study suggest that practice effects in iNPH patients are negligible [13]. If practice effects are at hand they generally influence the results on several successive test occasions (see e.g., [14]) which would have made it reasonable to expect significant changes also from three to twelve months.

The Grooved Pegboard identified the largest proportion of patients as responders or significantly improved after surgery ($\geq 1$ sd), followed by the Stroop (see Table 3), indicating that these tests are suitable for the tracking of individual changes after treatment. The RAVLT variables were less sensitive in this regard. This may express differential effects of treatment such that motor and executive functions improve more than memory and learning. However, the ability of a test to detect changes is, in part, determined by the distributional properties of the test in the population with which comparisons are made. For instance, an improvement of 1 sd on the Grooved Pegboard requires a change equal to 20% of normal (mean of HI) performance; while the corresponding figure for Rey Delayed Recall is a change of 45% of normal performance.

We choose two frequently used limits for impairment and improvement (1.5 and 1 sd respectively). Without access to data from follow up examinations among HI with the same intervals as those used for the examinations of patients the use of these or similarly arbitrary conventions are necessary.

The results from the European multicentre study corroborate and expand previously reported findings regarding the neuropsychological changes of patients with iNPH. Firstly, iNPH patients rarely perform within normal limits on neuropsychological tests at baseline, even when the administered battery is limited to a minute number of tasks. Excluding missing values, only 2% (3 out of 142) performed normally on all of the available variables and the majority (75% or 106) failed on 4 variables or more. If missing values are considered performance failures (which is the most reasonable interpretation) the corresponding numbers are 0 patients with normal performance on all variables and 89% (127 patients) performing below normal levels on 4 or more of them.
Secondly, age and education was weakly associated with neuropsychological performance (even when age adjusted values are used) as was the presence of vascular risk factors. The latter, however, less so in this study than previously described.

Thirdly, shunt treatment was again shown to be neuropsychologically beneficial, and significantly affected all of the tasks included in this study. The magnitude of improvement was not reduced by vascular comorbidity.

Finally, just like in previous studies, iNPH patients were still outperformed by HI after shunt treatment and the proportions of patients who reached normal performance levels on the specific tasks were modest.

This study has one important limitation. Our evaluation of change or improvement is based solely on the distribution of scores of healthy individuals tested on one occasion. Reliable change indices (RCIs) based on repeated measurements of a control group (healthy individuals or untreated patients) offer a more dependable approach to the assessment of change (for reviews see [15,16]) and have been used in an increasing number of studies in related areas, e.g., evaluation of neuropsychological changes following sports concussion, coronary artery bypass grafting [17], epilepsy surgery [18], and deep brain stimulation in Parkinson’s disease [19]. We believe that the introduction of RCIs would be an important step forward also in the field of iNPH outcome research, and aim to conduct studies in order to make this possible.

We conclude that three neuropsychological tests used in the European multicentre study on iNPH the Grooved Pegboard, the Rey Auditory Verbal Learning Test and the Stroop test are expedient, highly diagnostically discriminative, and well suited to evaluate changes following shunt treatment. Preferably, all of the tests and subtasks should be administered since they capture different aspects of the neuropsychological impairment of iNPH, but, evidently, at times less will have to do. Analyses of the results on this relatively large sample of patients corroborate previous findings; the neuropsychological deficits of iNPH affect several domains, are aggravated by cerebrovascular comorbidity, and are significantly reduced, yet not abolished, by shunt treatment.

Acknowledgments

We wish to express our gratitude towards Drs. R Carrillo, K Cesarini, D d’Avella, A Gabarrós, M Juhler, M Kiefer, M Kordás, U Meier, M Sames, and M Triffaux and their respective centres for contributing to this study with their knowledge, time, efforts and data. This study was sponsored by Codman. The authors declare that they have no conflict of interest.

References