J Neurosurg 112:399–409, 2010

Cognitive and motor improvement after retesting in normal-pressure hydrocephalus: a real change or merely a learning effect?

Clinical article

Elisabeth Solana,[1] Maria Antonia Poca, M.D., Ph.D.,[1,2] Juan Sahuquillo, M.D., Ph.D.,[1,2] Bessy Benejam,[1] Carme Junqué, Ph.D.,[3,4] and Mithilesh Dronavalli, B.Med.Sci., M.Bios.[5]

[1]Neurosurgery and Neurotraumatology Research Unit and [2]Department of Neurosurgery, Vall d'Hebron University Hospital and Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona; [3]Department of Psychiatry and Clinical Psychophysiology, University of Barcelona; [4]Institut d'Investigations Biomèdiques August Pi-Sunyer, Barcelona, Spain; and [5]Royal Melbourne Hospital, University of Melbourne and Data Clinic, Sydney, Australia

Object. The test-retest method is commonly used in the management of patients with normal-pressure hydrocephalus (NPH). One of the most widely used techniques in the diagnosis of this condition is evaluation of the patient's response to CSF evacuation by lumbar puncture (a so-called tap test or spinal tap). However, interpretation of improved results in subsequent evaluations is controversial because higher scores could reflect a real change in specific abilities or could be simply the result of a learning effect.

Methods. To determine the effect of testing-retesting in patients with NPH, the authors analyzed changes documented on 5 neuropsychological tests (the Toulouse-Pieron, Trail Making Test A, Grooved Pegboard, Word Fluency, and Bingley Memory tests) and several motor ability scales (motor performance test, length of step, and walking speed tests) in a series of 32 patients with NPH who underwent the same battery on 4 consecutive days. The same tests were also applied in 30 healthy volunteers. In both groups, the authors used the generalized least-squares regression method with random effects to test for learning effects. To evaluate possible differences in response depending on the degree of cognitive impairment at baseline, the results were adjusted by using the Mini-Mental State Examination scores of patients and controls when these scores were significant in the model.

Results. In patients with NPH there were no statistically significant differences in any of the neuropsychological or motor tests performed over the 4 consecutive days, except in the results of the Toulouse-Pieron test, which were significantly improved on Day 3. In contrast, healthy volunteers had statistically significant improvement in the results of the Toulouse-Pieron test, Trail Making Test A, and Grooved Pegboard test but not in the remaining neuropsychological tests. Patients in the healthy volunteer group also exhibited statistically significant improvement in the motor performance test but not in step length or walking speed.

Conclusions. No learning effect was found in patients with NPH on any of the neuropsychological or motor tests. Clinical improvement after retesting in these patients reflects real changes, and this strategy can therefore be used in both the diagnosis and evaluation of surgical outcomes. *(DOI: 10.3171/2009.4.JNS081664)*

Key Words • **normal-pressure hydrocephalus** • **practice effect** • **neuropsychological assessment** • **test-retest** • **learning effect** • **diagnosis**

Neuropsychological testing is commonly used in the diagnosis and follow-up of several disorders associated with cognitive impairment and in the assessment of subsequent improvement after medical and/or surgical treatment. The use of most of these tests requires retesting when the diagnosis is unclear, when the disorder is thought to be dynamic, and when there are other reasons for tracking cognitive status.[2,5,11–13,29] Retesting is also common to evaluate changes in cognitive competence after pharmacological and/or surgical treatment. In general, multiple tests are given to the same patient at different time points after treatment. In this situation, spurious statistically significant improvement might be detected, but these changes do not always reflect a clinically significant change, and may simply reflect random variability.

When a neuropsychological test is readministered, the results can be expected to differ even when the patient's cognition shows no real change. The reason for

Abbreviations used in this paper: ICP = intracranial pressure; IQR = interquartile range; MMSE = Mini-Mental State Examination; NPH = normal-pressure hydrocephalus.

this well-known phenomenon is that there is no perfectly reliable test and people naturally fluctuate in their functioning from time to time.[11,26] Consequently, a major clinically relevant issue in retesting is to understand the psychometric properties of each test to identify whether a change in score reflects a real change in cognitive abilities or is merely the result of a learning effect and random variability. In this scenario, better scores do not necessarily indicate clinical improvement because the practice effect can partly counterbalance cognitive or motor decline.[19]

What is known as the learning effect can occur when the same item or test is presented to individuals on repeated occasions or when individuals gain experience in solving certain types of problems.[3,6,12,19] Test-retest change may depend on the individuals' and tests' characteristics: younger and more educated individuals with higher general neuropsychological competence at baseline and with a short interval between tests are more likely to exhibit greater practice effects.

Normal-pressure hydrocephalus is a treatable cause of dementia,[10,15] manifested by gait disturbance, progressive dementia, and urinary incontinence combined with ventricular enlargement,[1,10,15,20,27,28] which can be partially or completely reversed by implanting a CSF shunt. A major challenge in the management of suspected NPH is to identify which patients will benefit from shunt therapy. One of the most widely used techniques in the diagnosis of patients with NPH is evaluation of their response to CSF evacuation by lumbar puncture (the so-called tap test). Before and after the tap test, a battery of neuropsychological and motor tests are performed. Some authors perform the tap test more than once, with the consequent replication of multiple neuropsychological and motor tests. The potential effect of retesting may mask the patient's real response to the tap test and hence jeopardize the validity of this test in patients with NPH.

The aim of our study was to determine the effect of testing-retesting on cognitive and motor performance in a cohort of 32 patients with NPH who underwent the same battery of neuropsychological and motor tests on 4 consecutive days. The patients were evaluated without the application of any therapeutic maneuvers. The results were compared with those obtained in a control group of 30 healthy volunteers who underwent the same protocol. The hypothesis we aimed to verify or refute was that learning effect can exist in normal populations (controls) but is clinically irrelevant in NPH patients because of cognitive impairment. To understand whether learning effects exist in NPH patients is clinically important to establish the individual cutoff thresholds that can reflect real patient changes and not simply random variability and/or practice effects.

Methods

General Protocol for the Study and Treatment of Patients With NPH

From May 2006 to May 2007, 32 of the 45 consecutive patients admitted to our department for evaluation of suspected NPH were included in this study. In these 32 patients NPH was diagnosed exclusively by continuous ICP monitoring. All patients had at least 2 of the following clinical symptoms, unexplained by other neurological or nonneurological conditions: gait abnormalities, sphincter dysfunction, cognitive deficits, and/or Parkinson disease refractory to treatment. In all patients, CT scanning or MR imaging showed ventricular dilation (Evans index ≥ 0.30).[8] Of the 45 patients, 3 were excluded because they were unable to perform any of the study tasks, 3 because they received a lumbar puncture as a diagnostic test, and 7 because they refused to participate in the study or because they were uncooperative due to severe dementia.

Our protocol for the study and management of patients with suspected NPH syndrome has previously been detailed.[20,21] Briefly, patients were clinically graded according to the NPH scale,[23] which registers the clinical status of patients in the 3 main symptoms. In this scale the minimum possible score is 3, indicating that the patient is bedridden or unable to walk, has no or minimal contact with the environment, and has urinary and fecal incontinence. A maximum score of 15 points indicates normal functioning in the 3 domains.

Patient evaluation also included a battery of neuropsychological tests[20,21] with which we evaluated several aspects of verbal (Rey Auditory-Verbal Learning Test) and visual (Wechsler Memory Scale III) memory, scanning and visual-motor tracking (Trail Making Test), divided attention and cognitive flexibility (Trail Making Test B), auditory attention (Digits forward), working memory (Digits backward, phonetically and category Fluency), visual-hand coordination (Purdue Pegboard Test), and global cognitive impairment (MMSE). Several scales to assess the patients' functional behavior and to evaluate changes in everyday activities were used but not included in this analysis (Table 1).

The evaluation was completed by continuous ICP monitoring, using an epidural sensor[20,21] and/or CSF dynamics studies (the Marmarou bolus test and the Katzman constant rate infusion test). According to our criteria for shunt therapy,[20,21] all patients with > 10% of B-waves in the continuous ICP recording with or without an abnormal resistance to outflow (Rout) of CSF (> 10 mm Hg/ml/min in the Katzman infusion test) underwent shunt placement. All shunt-treated patients underwent neurological, neuroimaging, and neuropsychological examinations before and 6 months after surgery. Outcome was independently assessed by a neurosurgeon and a research psychologist.

Written informed consent to participate in the study was obtained from all patients or from the next of kin of patients whose cognitive impairment precluded them from understanding or signing the written informed consent.

Psychometric and Motor Assessments to Evaluate the Learning Effect

Before shunt insertion, in addition to the aforementioned general protocol, a selected battery of 5 psychometric tests and 1 motor test was applied at the same time daily for 4 consecutive days in all patients to analyze the learning effect. During these 4 days, patients did not receive any treatment or therapeutic maneuvers.

TABLE 1: Tests used in the protocol for the clinical and neuropsychological examination of patients with suspected NPH in our center

Type of Assessment	Assessment Tool
screening test	MMSE,* Frontal Assessment Battery
functional scales	Normal Pressure Hydrocephalus Scale, Rapid Disability Rating Scale–2, Modified Stein & Langfitt Scale, Daily Life Activities Scale
motor assessment	Motor Performance Test (MPT)
neuropsychological assessment	
memory	
visual memory	Wechsler Memory Scale–R*
immediate verbal memory	Digits subtests (Wechsler Adult Intelligence Scale III)
auditive learning	Rey Auditory-Verbal Learning Test*
executive function	Trail Making Test A & B,* Verbal Fluency Test
psychomotor velocity	Purdue Pegboard Test
language	Boston Diagnostic Aphasia Examination
praxis	Boston Diagnostic Aphasia Examination
behavior	Frontal Behavioral Inventory

* Most widely used neuropsychological tests in the literature.

The neuropsychological battery evaluated attention and visual scanning (Toulouse-Pieron Test and Trail Making Test A), motor speed and manual coordination (Grooved Pegboard Test and Trail Making Test A), executive functions (Word Fluency Test) and immediate recall visual memory (Bingley Memory Test). In the Word Fluency Test, the individual was asked to say the maximum number of words beginning with a designated letter in 1 minute; a different letter—P, R, S, and T—was chosen on each of the 4 different days). In the Bingley Memory Test, the individual was asked to remember as many common pictures as possible from a chart containing 12 previously presented pictures; a different chart with different pictures was used every day.

Cognitive assessment was completed with the administration of the Motor Performance Test, which is a composite test that evaluates 6 motor tasks: 1) time to get up from a standard-height chair (in seconds); 2) getting up and down a 23-cm-high step repeatedly as quickly as possible in 10 seconds; 3) the number of faltering steps made in a 2-m tandem walk; 4) time in seconds that the patient can stand on a single leg (twice with each leg); 5) time in seconds required to walk 5 m and the mean length of steps; and 6) the number of steps required to turn 180°. Each task was scored from 2 points (capable) to 0 (incapable), except for the time to get up from a chair, which had a maximum score of 4 points. The maximum possible score in the overall Motor Performance Test is 14 points (Table 2).

Control Group

Thirty volunteers in the same age range were assessed on 4 consecutive days by a neuropsychological team and they used the same psychometric and motor battery applied in patients with NPH. The inclusion criteria for these healthy volunteers were as follows: 1) age between 65 and 90 years; 2) absence of any neurological diseases; 3) absence of any motor or joint disorders that could interfere with movement; 4) independence for Daily Life Activities; and 5) a minimum score of 24 on the MMSE.

Statistical Analysis

All descriptive statistics were analyzed using the SPSS package for Windows (version 15, SPSS Inc.). The assumption that data were normally distributed was tested using the Kolmogorov-Smirnov test. In normally distributed data, the mean ± 1 SD was used to summarize the variables. In skewed samples, the median and the IQR were used. Due to the sample size, nonparametric tests were selected to analyze the results. Data obtained on the 1st day of the study in patients with NPH and in those in the control group were compared using the Mann-Whitney U-test.

We used repeated measures for each individual on the same dependent variable, and therefore, differences between and within individuals occurred. Consequently, the observations of the dependent variable were no longer independent, and statistically we needed to take this nonindependence into account. Therefore, a generalized least-squares regression method with random effects was used to test for learning effects. In this model each individual's test result is 1 observation. The observations were grouped by subject. In the results, the statistic Sigma within-value represents the variation of test results within a subject and the Sigma between-value represents the variation of test results between subjects. In this method, Rho is defined as the percentage of total variation that is due to intrasubject variation. This analysis is optimal when the total variation is mainly due to intrasubject variation and not to intersubject variation, and therefore, within-patient changes in test results can be better assessed in repeated measures.

In this analysis, Tests 1–3 were the scores obtained in repeated tests compared with the scores obtained by patients or controls in the baseline test. Therefore, Test 1 was calculated as the difference between baseline and the score obtained in the same test on Day 1. The coefficient of Tests 1–3 was the learning effect compared with baseline of each respective test. The results were adjusted by the MMSE when MMSE was statistically significant as a predictor. Models were built separately for patients and controls. Combining the data and comparing the learning effect between patients and controls was not performed because this was not the aim of our study; hence, the requirement for a power-consuming interaction term was avoided. The main assumption in this model was that within-patient/subject variation would be similar for all patients/subjects. All generalized least-squares regression analyses were done using Stata version 9.0 (Stata Corp.) and repeated-measures analysis was done with random

TABLE 2: Motor performance in patients with NPH syndrome

Motor Tasks	Description	Score*
time to get up from a standard-height chair	patient is asked to get up from a chair (an armless, stiff- & straight-backed, standard-height chair), possibly w/o using hands (hands can be used if is necessary)	4: < 2 secs (w/ hands) 3: ≥ 2 secs (w/ hands) 2: < 2 secs (w/o hands) 1: ≥ 2 secs (w/o hands) 0: not feasible
getting up & down a step	patient is asked to go up & down a 23-cm-high step repeatedly as quickly as possible	2: ≥ 3 steps 1: < 3 steps 0: not feasible
tandem walk	patient is instructed to walk touching the toes of 1 foot w/ the heel of the other, following a 2-m-long, 5-cm-wide line	2: < 8 errors 1: ≥ 8 errors 0: not feasible
1-leg stand time	patient is instructed to stand, alternately, on the rt & lt leg for as long as possible, for total of 4 attempts	2: ≥ 2 secs 1: < 2 secs 0: not feasible
time to walk 5 m	patient is asked to walk 5 m at a normal pace; 2 attempts are made; mean speed (m/sec) & mean length of step (cm) are also calculated	2: ≥ 0.6 m/sec 1: < 0.6 m/sec 0: not feasible
no. of steps to turn around (180°)	patient stands w/ feet together & is asked to turn around	2: < 5 steps 1: ≥ 5 steps 0: not feasible

* The score ranges from a minimum of 0 (most severe motor impairment) to a maximum of 14 (normal motor performance).

effects using the XTREG time-series regression command specifying the patient identity as the subject identity and specifying the random effects model. Statistical significance was considered at p ≤ 0.05.

Results

Patients With NPH

The patient group included 18 women and 14 men, age 74.6 ± 6.3 years (range 61–85 years). Fourteen patients had not finished basic education (up to the age of 14 years), 14 had finished basic education but did not attend secondary school, 3 had completed secondary school, and 1 had university education. The mean score of patients on the MMSE was 20.8 ± 6.6 (range 6–30). The diagnosis of NPH was confirmed in all patients (mean percentage of B-waves 41.6 ± 20%, range 10–81%). A differential low-pressure valve system was implanted in 29 patients. Surgical treatment was not performed in 3 patients because of the family's refusal. A programmable Hakim Medos valve (Medos S.A.), with a closing pressure range selected at between 30 and 70 mm H_2O, was implanted in 22 patients. In all 22 patients, this valve was combined with an infraclavicular low-pressure gravity-compensating accessory (NMT Neurosciences Implants S.A.). A gravitational Miethke Dual-Switch valve with an opening pressure of 5/40 cm H_2O (distributed by Aesculap AG & Co., KG) was used in the remaining 7 patients. Of the 29 treated patients, 26 (90%) improved after shunt placement, showing an increase of ≥ 1 point on the NPH scale.

Control Group

Cognitive skills were assessed in all 30 volunteers, age 71.3 ± 5.4 years (range 65–90 years) (17 men and 13 women). Of these, 16 had not finished basic education, 10 had finished basic education but did not attend secondary school, and 4 had completed secondary school. All volunteers had an MMSE score ≥ 24 points (mean 27.6 ± 1.8, range 25–30). The Motor Performance Test was performed in 28 volunteers, age 70.3 ± 3.9 years (range 65–79 years) (16 men and 12 women).

Cognitive Results in Patients and Controls

Complete neuropsychological evaluation could not be performed in all patients with NPH due to their clinical status. Table 3 summarizes the results of the neuropsychological tests performed in patients and in healthy volunteers on the 1st day of the study. In all, tests scores were lower in patients with NPH than in healthy volunteers. When data obtained from neuropsychological tests performed on 4 consecutive days in patients with NPH were compared, no statistically significant differences were found for any of the tests, except in the Toulouse-Pieron test when the results from Day 3 were compared with those from Day 1 (Table 4).

When we compared data obtained from neuropsychological tests performed on 4 consecutive days in healthy volunteers, statistically significant differences were found in the Toulouse-Pieron test, the Trail Making Test A, and the Grooved Pegboard test for the dominant and nondominant hands. No statistically significant differences were found in the Word Fluency test (Table 5). For the Bingley Memory test, the results were significantly worse on Day 2 than Day 1 (p = 0.006). However, no statistically significant differences were found when Days 3 and 4 results were compared with those obtained on Day 1 (p = 0.125

TABLE 3: Cognitive and motor results of healthy volunteers and NPH patients on Day 1 of the study*

	Patients w/ NPH			Healthy Volunteers		
Test	No. of Cases	Mean ± SD	Range	No. of Cases	Mean ± SD	Range
Toulouse-Pieron	30	4.1 ± 3.5	0–14	30	15.8 ± 6.2	5–27
TMT A (secs)	17	198.7 ± 179.4	77–840	30	81.6 ± 33.5	37–175
Grooved Pegboard						
dominant hand	29	8.7 ± 6.6	0–25	30	21.3 ± 7.8	6–36
nondominant hand	27	8.2 ± 5.6	0–22	30	19.6 ± 6.7	5–33
Bingley Memory	32	3.7 ± 2.3	0–9	30	7.6 ± 1.9	4–11
Word Fluency	28	3.5 ± 3.4	0–13	30	8.8 ± 4	2–18
MPT	20	7 ± 3.6	1–13	28	14† (IQR 13–14)	11–14
length of steps (cm)	20	30.9 ± 11	13.7–55.6	28	51.9 ± 8.3	38.5–71.4
speed velocity (m/sec)	20	0.45 ± 0.19	0.14–1	28	0.89 ± 0.3	0.48–1.67

* MPT = Motor Performance Test; TMT = Trail Making Test.
† Median.

and p = 0.733, respectively). Figures 1 and 2 summarize the results obtained in patients with NPH and healthy volunteers for all tests performed on all 4 days of the study.

Results of Motor Skills Tests in Patients and Controls

Motor performance could only be evaluated in 20 of the 32 patients with NPH (the remaining 12 patients needed help to walk or to maintain stability, and in these patients the Motor Performance Test score was registered as 0). On the 1st day of the study, the mean Motor Performance Test score, in the 20 patients who were able to perform it, was 7 ± 3.6 (range 1–13), the mean step length was 30.9 ± 11 cm (range 13.7–55.6 cm), and the mean walking speed was 0.45 ± 0.19 m/second (range 0.14–1 m/second). In patients with NPH, repeated-measures analysis revealed no statistically significances in step length or walking speed on the 4 consecutive days (Table 4, Fig. 2). An increase in step length of ≥ 5 cm was observed on Day 4 in only 2 patients, whereas a decrease on Day 1 was found in 11 patients.

On Day 1 of the study, the median Motor Performance Test score in healthy volunteers was 14 (IQR 13–14, minimum 11, maximum 14), the mean step length was 51.9 ± 8.3 cm (range 38.5–71.4 cm), and the mean walking speed was 0.89 ± 0.3 m/second (range 0.48–1.67 m/second). These values were significantly higher than those documented in patients with NPH (Table 3). Repeated-measures analysis in volunteers revealed statistically significant differences between the completed Motor Performance Test but not between step length or between walking speed on the 4 consecutive days (Table 5, Fig. 2).

Discussion

The results of the present study demonstrate that the learning effect was absent in patients with NPH in all but 1 of the neuropsychological and motor tests selected for their evaluation, and this was independent of their MMSE score. The exception was the Toulouse-Pieron test, which showed improvement on Day 3 but not on Day 4. In contrast, the learning effect was present in healthy volunteers on most of the neuropsychological tests but not on tests of motor performance such as walking speed or step length. Among the healthy volunteers, the learning effect tended to increase over the 4 days of the study period (Fig. 1 and 2). The tests used in this study were selected because they are among the most widely used in the literature to evaluate outcome after CSF withdrawal in the diagnosis of patients with suspected NPH. Although the MMSE is not a suitable screening test in these patients, in the present study it was used to evaluate and adjust for possible differences in response depending on the degree of cognitive impairment at baseline.

Factors Influencing the Learning Effect

When retesting patients, the learning effect can be influenced by the patient's cognitive skills, the test's characteristics, the impact of cognitive status at earlier testing (initial intelligence or memory competence),[11,22,26] the retest interval,[2,6,9,11,13,24,26] age,[2,6,9,11,12,19,26] educational attainment,[6,12,24] and treatment effects.[13] However, the factors considered the best determinants of the learning effect are age[19,26] and performance in the first evaluation.[11,22,24,26] A study performed in healthy elderly people found that tests of psychomotor and cognitive skills, such as the Wechsler Adult Intelligence Scale performance tests and Trail Making Test A and B, showed a relationship between learning and age.[19] Nevertheless, other studies[6,22,26] reported that only the initial score had a differential practice effect for all the measures evaluated. In the present study, although age and years of education were similar between patients with NPH and controls, no learning effect was found in the patients.

The learning effect also depends of the cognitive domain being tested. Distinct cognitive abilities exhibit different degrees of temporal stability. In general, verbal

Table 4: Results of neuropsychological and motor performance tests in patients with NPH on 4 consecutive days*

Test	DS†	Coef	SE	Z Value	p Value	95% CI	Variation W/in-Patient	Variation Btwn-Patient	Rho	No. of Cases	No. of Obs
Toulouse-Pieron	2–1	0.5	0.6	0.87	0.385	−0.7 to 1.7					
	3–1	1.7	0.6	2.72	0.007‡	0.5–2.9	0.79	2.38	0.36	30	120
	4–1	1.2	0.6	1.90	0.057	0–2.4					
Grooved Pegboard§											
dominant hand	2–1	1.0	0.7	1.42	0.156	−3.4 to 2.5					
	3–1	1.3	0.7	1.83	0.067	−0.9 to 2.8	0.04	0.59	0.67	30	120
	4–1	0.4	0.7	0.60	0.552	−1.0 to 1.9					
nondominant hand	2–1	0.5	0.7	0.72	0.473	−0.9 to 1.9					
	3–1	0.3	0.7	0.48	0.633	−1.0 to 1.7	0.09	0.06	0.80	30	120
	4–1	−1.3	0.7	−1.86	0.062	−2.7 to 0.1					
Word Fluency§	2–1	0.6	0.4	1.22	0.223	−0.4 to 1.6					
	3–1	0.4	0.4	0.75	0.456	−0.6 to 1.3	0.03	0.62	0.54	30	120
	4–1	0.8	0.4	1.56	0.119	−0.2 to 1.7					
Bingley Memory§	2–1	0	0.3	0.09	0.927	−0.6 to 0.7					
	3–1	0	0.3	0.09	0.927	−0.6 to 0.7	0.04	0.71	0.52	32	128
	4–1	−0.5	0.3	−1.55	0.120	−1.2 to 0.1					
Trail Making Test A§	2–1	−13.5	21.2	−0.64	0.524	−55.1 to 28.0					
	3–1	−7.2	21.5	−0.33	0.739	−49.4 to 35.0	0.02	0.41	0.73	21	79
	4–1	9.0	21.7	0.42	0.677	−33.5 to 51.6					
Motor Performance Test	2–1	−0.2	0.4	−0.39	0.699	−1.0 to 0.7					
	3–1	−0.7	0.4	−0.15	0.877	−0.9 to 0.8	0.03	0.18	0.86	31	123
	4–1	−0.7	0.4	−1.56	0.119	−1.5 to 0.2					
length of steps (cm)	2–1	−1.51	1.4	−1.04	0.297	−4.3 to 1.3					
	3–1	−1.50	1.4	−1.04	0.299	−4.3 to 1.3	0.03	0.20	0.88	31	123
	4–1	−2.57	1.5	−1.76	0.079	−5.4 to 0.3					
walking speed (m/sec)	2–1	0	0	−1.32	0.188	−0.1 to 0					
	3–1	0	0	−1.35	0.176	−0.1 to 0	0.03	0.19	0.86	31	123
	4–1	0	0	−1.64	0.100	0–0					

* Coef = coefficient; DS =days of study; Obs = total number of observations.
† Days of study: Day 2 vs Day 1, Day 3 vs Day 1, and Day 4 vs Day 1, considering Day 1 as the baseline.
‡ $p < 0.05$.
§ Results adjusted by the MMSE scores. For detailed information about the statistical methodology, refer to the *Statistical Analysis* section.

knowledge and verbal reasoning are the most stable over time.[13] Repeated neuropsychological studies in the elderly showed that nonverbal reasoning, attention, and concentration were more dynamic over time than learning and remembering new data, whereas acquiring new information was the most stable ability.[13] More pronounced learning effects have been observed in measures with problem-solving or novelty components.[6,11] Indeed, the practice effect, although small, can be found in purely motor measures or measures with alternative forms.[6]

How can the Learning Effect be Reduced?

Although the learning effect cannot be completely eliminated, it can be reduced by the use of parallel forms of the tests.[2,3,6,12,24] However, these forms may be less effective in attenuating the practice effect if they contain novel concepts, visual-spatial learning, or graphomotor reasoning,[2] and may not be equivalent to the original test in difficulty, sensitivity, and validity.[6] In our healthy volunteers, the only tests that did not show significantly better scores after retesting were the Word Fluency and the Memory Objects (Bingley Memory) tests, which involve frontal execution and immediate visual memory. These results were probably found because we used different letters and pictures on each of the 4 consecutive days, confirming the effectiveness of using alternative forms to reduce the learning effect. In contrast, patients with NPH showed no significant improvements in any of the tests analyzed. Purely motor skills did not change significantly on the 4 consecutive days of the study in either patients with NPH or control individuals.

Utility of Retesting in Patients With NPH

In patients in whom NPH is suspected, one of the

Table 5: Results of neuropsychological and motor performance tests in healthy volunteers on 4 consecutive days

Test	DS	Coef	SE	Z Value	p Value	95% CI	Variation W/in-Individual	Variation Btwn-Individual	Rho	No. of Cases	No. of Obs
Toulouse-Pieron	2–1	3.1	1.0	2.98	0.003*	1.1–5.1					
	3–1	4.2	1.0	4.12	0.000*	2.2–6.2	0.23	0.24	0.65	30	120
	4–1	4.9	1.0	4.74	0.000*	2.9–6.9					
Grooved Pegboard†											
dominant hand	2–1	3.1	0.7	4.73	0.000*	1.8–4.4					
	3–1	3.1	0.7	4.73	0.000*	1.8–4.4	0.38	0.30	0.87	30	120
	4–1	4.7	0.7	7.11	0.000*	3.4–6.0					
nondominant hand	2–1	1.2	0.6	1.95	0.052	0–2.5					
	3–1	1.4	0.6	2.21	0.027*	0.2–2.6	0.16	0.15	0.87	30	120
	4–1	2.53	0.6	4.00	0.000*	1.3–3.8					
Word Fluency	2–1	0.3	0.7	0.46	0.644	–1.1 to 1.7					
	3–1	0.4	0.7	0.60	0.549	–1.0 to 1.8	0.01	0.13	0.60	30	120
	4–1	0	0.7	0.05	0.963	–1.4 to 1.4					
Bingley Memory†	2–1	–1.1	0.4	–2.73	0.006*	–1.8 to –0.3					
	3–1	–0.6	0.4	–1.53	0.125	–1.4 to 1.2	0.10	0.58	0.08	30	120
	4–1	–0.1	0.4	–0.34	0.733	–0.9 to 0.6					
Trail Making Test A	2–1	–11.1	4.2	–2.63	0.008*	–19.3 to –2.8					
	3–1	–10.6	4.2	–2.52	0.012*	–18.8 to –2.4	0.18	0.24	0.73	30	120
	4–1	–18.0	4.2	–4.29	0.000*	–26.3 to –9.8					
Motor Performance Test	2–1	0.2	0.1	2.55	0.011*	0–0.4					
	3–1	0.2	0.1	2.55	0.011*	0–0.4	0.11	0.28	0.69	28	112
	4–1	0.2	0.1	2.55	0.011*	0–0.4					
length of steps (cm)	2–1	1.1	0.9	1.14	0.255	–0.8 to 2.9					
	3–1	0.2	0.9	0.22	0.830	–1.7 to 2.1	0.02	0.33	0.77	28	112
	4–1	0.8	0.9	0.87	0.383	–1 to 2.7					
walking speed (m/sec)	2–1	0	0	–0.08	0.934	–0.1 to 0.1					
	3–1	0	0	0.38	0.706	0–0.1	0	0.22	0.72	28	112
	4–1	0	0	0.44	0.663	0–0.1					

* $p < 0.05$.
† Results were adjusted by using the MMSE scores. For detailed information about the statistical methodology, refer to the *Statistical Analysis* section.

most commonly used diagnostic techniques is to quantify clinical changes after CSF removal by lumbar puncture (tap test) or transitory lumbar drainage.[18] This approach implies testing and retesting patients in a short period. When the tap test is performed, some authors retest patients on 3 consecutive days. In 1982, Wikkelsö et al.[27,28] proposed a diagnostic protocol for patients with suspected NPH that consists of evacuating 40–50 ml of CSF by lumbar puncture on 2 consecutive days and in quantifying clinical changes before and after each procedure. Although the percentage of false-negative results in this test is high, it is still among the most frequently used tests in a large number of neurology and neurosurgery departments.

The rationale for the tap test is that removing CSF from the subarachnoid space produces a partial normalization of CSF dynamics. After the tap test, patients with NPH may improve, whereas those with cerebral atrophy will not.[25,28] Additionally, patients who respond to CSF evacuation after the tap test may also respond positively after shunt implantation.[10,25] However, when using the tap test, the potential effect of retesting and learning contamination bias may mask the patient's real response after CSF removal and, consequently, jeopardize the validity of this diagnostic maneuver.

Difficulty of Detecting the Learning Effect

The most commonly used technique in the literature for considering improvement after retesting patients is the standard deviation method, which describes significant change as an increase from preoperative test score by at least 1 SD of the mean baseline score of the population sample. Another widely used technique defines significant score change as an increased of at least 20% from baseline. While these methods may be simple to implement, they do not adequately evaluate the psychometric and

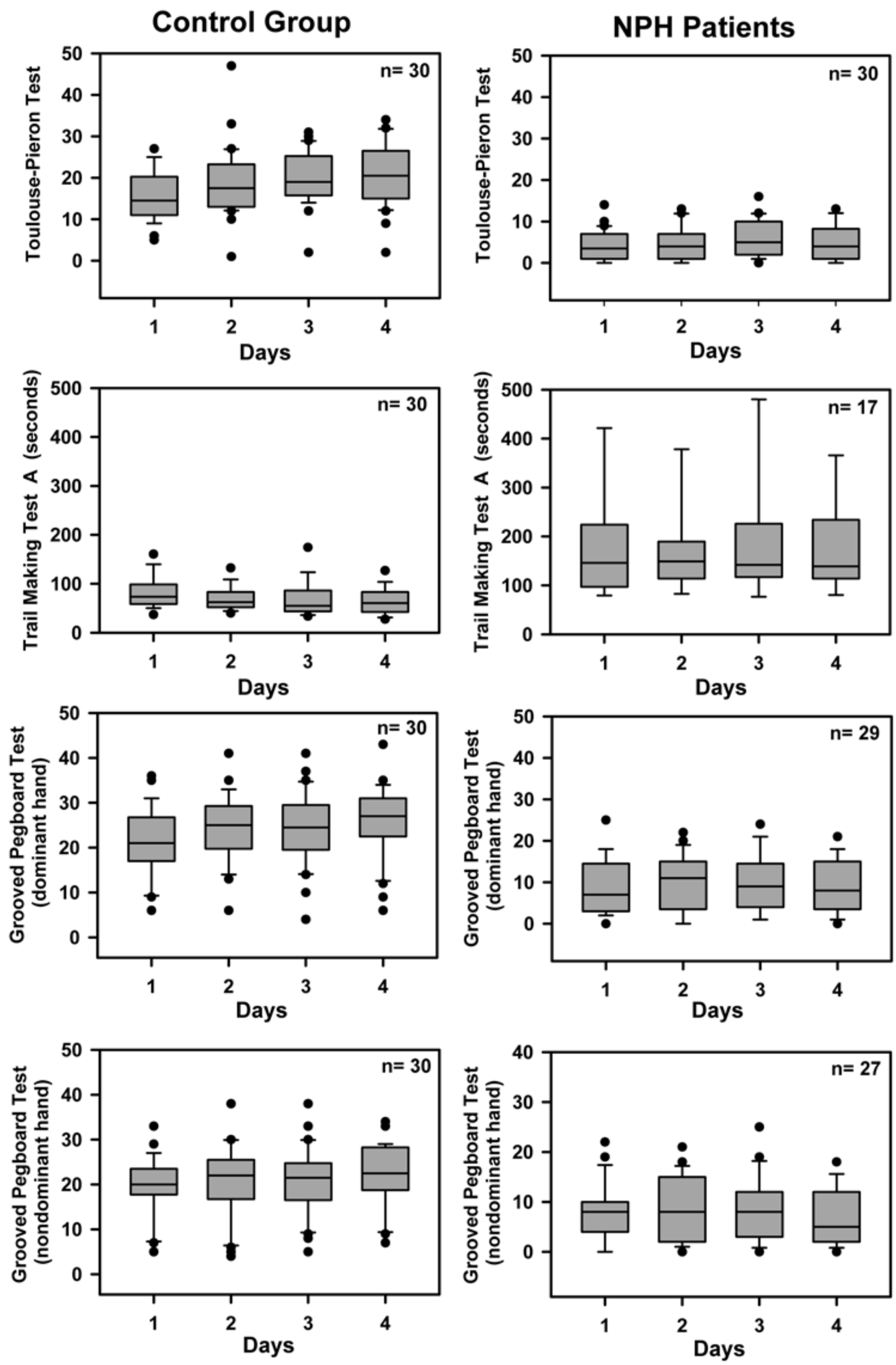

Fig. 1. Box-and-whisker plots showing the results of the Toulouse-Pieron test, the Trail Making Test A, and the Grooved Pegboard test (dominant and nondominant hands) obtained in healthy control individuals *(left column)* and in patients with NPH *(right column).* The plots show only data acquired in patients able to perform all the tests on each of the 4 consecutive days of the study. Plots show complete data ranges, except for the Trail Making Test A, in which data are limited to the 5th–95th percentiles, due to the large difference observed between the groups. In all tests, better performance was observed in healthy volunteers than in patients with NPH. In healthy volunteers, scores progressively increased over the 4 days of the study, whereas the time required to perform the Trail Making Test A was progressively reduced due to the learning effect. In patients with NPH, no statistically significant differences were observed in any of the tests, except for the Toulouse-Pieron test, which showed improvement on Day 3 but not Day 4 (see Table 4).

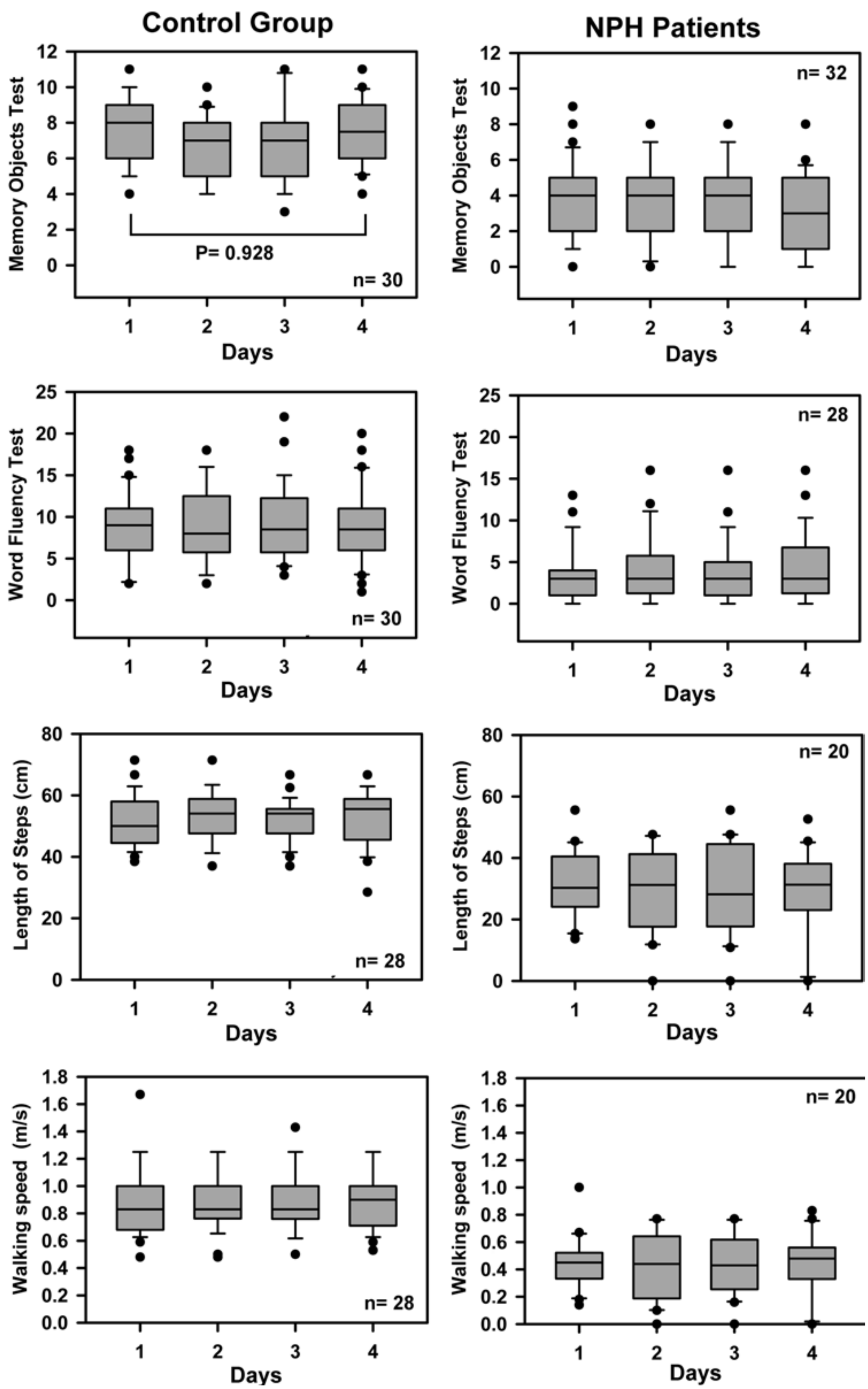

FIG. 2. Box-and-whisker plots showing the results of the Memory Objects test, Word Fluency test, step length, and walking speed obtained in healthy control individuals *(left column)* and in patients with NPH *(right column).* The plots show only data obtained in patients able to perform all the tests on each of the 4 consecutive days of the study. Plots show complete data ranges. In all tests, better performance was observed in healthy volunteers than in patients with NPH. Performance did not significantly improve over the 4 days of the study in any of the tests, in either the volunteer or patient group.

statistical issues surrounding change scores.[22] To evaluate "true" change, a method should determine whether the observed change exceeds that expected from measurement error and improvement over time due to practice or regression to the mean.[4] Consequently, many authors tried to identify real changes by using more complex and reliable predictors of follow-up scores in the retest, such as Reliable Change Index adjusted for practice or different types of regression models.[5,7,11,14,16,17,22,26,29]

In the present study, we used a generalized least squares regression method with random effects to test for learning effects. Repeated-measures analysis with random effects was the preferred analysis to distinguish within-patient variation from between-patient variation. The ANOVA and other analyses that use the entire data fail to account for repeated measures and their association with between- and within-subject patient variation and assume that all test observations are independent and that there is no added correlation within a single patient's observations. This assumption causes the standard errors to shrink and to give falsely low p values. A regression is not optimal as only pre- and postobservations can be used. There is a loss of power when only 2 of the possible 4 data points are used in the analysis, which inflates the p values. In a paired t-test, the difference between baseline and Test 1 or another test result is regressed against predictors, which is not always appropriate and only works if the baseline test result is not correlated to Test 1 or whichever posttest is used for comparison. The method used in the present study is superior to using ranks because nonparametric analyses consume more power and cannot quantify the learning effect as a continuous measure or score.

Conclusions

The results of the present study confirm that the learning effect is absent in patients with NPH and, consequently, that improvements after CSF removal should be considered to be real rather than the presence of artifacts secondary to retesting patients on several consecutive days. Given the lack of learning effect when patients were studied on 4 consecutive days, this absence is all the more likely to occur in clinical practice when the interval between tests is much greater (usually between 6 and 12 months after shunt placement). The tests selected in this study can therefore be used both in diagnosis and in the evaluation of surgical outcomes.

Disclosure

This study was supported in part by the Institut de Recerca Vall d'Hebron, Universitat Autònoma de Barcelona (E.S.), and by grant no. 07/0681 (M.A.P.) from the Fondo de Investigación Sanitaria.

Acknowledgments

The authors gratefully acknowledge Gail Craigie for editorial assistance and the collaboration of the neurosurgical nurses in the study of these patients, especially Maria Angeles Barba, R.N., and Mercedes Batlle, R.N.

References

1. Adams RD, Fisher CM, Hakim S, Ojemann RG, Sweet WH: Symptomatic occult hydrocephalus with "normal" cerebrospinal-fluid pressure. A treatable syndrome. **N Engl J Med 273:** 117–126, 1965
2. Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, et al: Practice effects and the use of alternate forms in serial neuropsychological testing. **Arch Clin Neuropsychol 20:**517–529, 2005
3. Benedict RH, Zgaljardic DJ: Practice effects during repeated administrations of memory tests with and without alternate forms. **J Clin Exp Neuropsychol 20:**339–352, 1998
4. Chelune GJ, Naugle RI, Luders H, Awad IA: Prediction of cognitive change as a function of preoperative ability status among temporal lobectomy patients seen at 6-month follow-up. **Neurology 41:**399–404, 1991
5. Collie A, Maruff P, Darby D, McStephen M: The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. **J Int Neuropsychol Soc 9:**419–428, 2003
6. Dikmen SS, Heaton RK, Grant I, Temkin NR: Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. **J Int Neuropsychol Soc 5:** 346–356, 1999
7. Erlanger D, Feldman D, Kutner K, Kaushik T, Kroger H, Festa J, et al: Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. **Arch Clin Neuropsychol 18:**293–316, 2003
8. Evans WA Jr: An encephalographic ratio for estimating ventricular enlargement and cerebral atrophy. **Arch Neurol Psychiatry 42:**931–937, 1942
9. Feinstein A, Brown R, Ron M: Effects of practice of serial tests of attention in healthy subjects. **J Clin Exp Neuropsychol 16:**436–447, 1994
10. Gallia GL, Rigamonti D, Williams MA: The diagnosis and treatment of idiopathic normal pressure hydrocephalus. **Nat Clin Pract Neurol 2:**375–381, 2006
11. Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, et al: Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. **Arch Clin Neuropsychol 16:**75–91, 2001
12. Horton AM Jr: Neuropsychological practice effects x age: a brief note. **Percept Mot Skills 75:**257–258, 1992
13. Ivnik RJ, Smith GE, Lucas JA, Petersen RC, Boeve BF, Kokmen E, et al: Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance. **Neuropsychology 13:**121–127, 1999
14. Jacobson NS, Truax P: Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. **J Consult Clin Psychol 59:**12–19, 1991
15. Kahlon B, Sundbarg G, Rehncrona S: Comparison between the lumbar infusion and CSF tap tests to predict outcome after shunt surgery in suspected normal pressure hydrocephalus. **J Neurol Neurosurg Psychiatry 73:**721–726, 2002
16. Knight RG, McMahon J, Skeaff CM, Green TJ: Reliable Change Index scores for persons over the age of 65 tested on alternate forms of the Rey AVLT. **Arch Clin Neuropsychol 22:**513–518, 2007
17. Lovell MR, Iverson GL, Collins MW, Podell K, Johnston KM, Pardini D, et al: Measurement of symptoms following sports-related concussion: reliability and normative data for the post-concussion scale. **Appl Neuropsychol 13:**166–174, 2006
18. Marmarou A, Bergsneider M, Klinge P, Relkin N, Black PM: The value of supplemental prognostic tests for the preoperative assessment of idiopathic normal-pressure hydrocephalus. **Neurosurgery 57 (3 Suppl):**S17–S28, ii–v, 2005
19. Mitrushina M, Satz P: Effect of repeated administration of a neuropsychological battery in the elderly. **J Clin Psychol 47:** 790–801, 1991

20. Poca MA, Mataréó M, Del Mar Matarín M, Arikan F, Junqué C, Sahuquillo J: Is the placement of shunts in patients with idiopathic normal-pressure hydrocephalus worth the risk? Results of a study based on continuous monitoring of intracranial pressure. **J Neurosurg 100:**855–866, 2004
21. Poca MA, Mataró M, Matarín M, Arikan F, Junqué C, Sahuquillo J: Good outcome in patients with normal-pressure hydrocephalus and factors indicating poor prognosis. **J Neurosurg 103:**455–463, 2005
22. Raymond PD, Hinton-Bayre AD, Radel M, Ray MJ, Marsh NA: Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery. **Eur J Cardiothorac Surg 29:**82–88, 2006
23. Sahuquillo J, Rubio E, Codina A, Molins A, Guitart JM, Poca MA, et al: Reappraisal of the intracranial pressure and cerebrospinal fluid dynamics in patients with the so-called "normal pressure hydrocephalus" syndrome. **Acta Neurochir (Wien) 112:**50–61, 1991
24. Salinsky MC, Storzbach D, Dodrill CB, Binder LM: Test-retest bias, reliability, and regression equations for neuropsychological measures repeated over a 12-16-week period. **J Int Neuropsychol Soc 7:**597–605, 2001
25. Sand T, Bovim G, Grimse R, Myhr G, Helde G, Cappelen J: Idiopathic normal pressure hydrocephalus: the CSF tap-test may predict the clinical response to shunting. **Acta Neurol Scand 89:**311–316, 1994
26. Temkin NR, Heaton RK, Grant I, Dikmen SS: Detecting significant change in neuropsychological test performance: a comparison of four models. **J Int Neuropsychol Soc 5:**357–369, 1999
27. Wikkelso C, Andersson H, Blomstrand C, Lindqvist G: The clinical effect of lumbar puncture in normal pressure hydrocephalus. **J Neurol Neurosurg Psychiatry 45:**64–69, 1982
28. Wikkelso C, Andersson H, Blomstrand C, Lindqvist G, Svendsen P: Normal pressure hydrocephalus. Predictive value of the cerebrospinal fluid tap-test. **Acta Neurol Scand 73:**566–573, 1986
29. Woods SP, Childers M, Ellis RJ, Guaman S, Grant I, Heaton RK, et al: A battery approach for measuring neuropsychological change. **Arch Clin Neuropsychol 21:**83–89, 2005

Manuscript submitted December 12, 2008.

Accepted April 27, 2009.

Please include this information when citing this paper: published online May 29, 2009; DOI: 10.3171/2009.4.JNS081664.

Part of this study was presented at the Hydrocephalus 2008 Congress (oral presentation), Hannover, Germany, September 17–20, 2008.

Address correspondence to: Maria Antonia Poca, M.D., Ph.D., Department of Neurosurgery, Vall d'Hebron University Hospital, Universitat Autònoma de Barcelona, Passeig Vall d'Hebron, 119-129 08035 Barcelona, Spain. email: 26382app@comb.es.