Shunting normal pressure hydrocephalus: the predictive value of combined clinical and CT data

Jan Vanneste, Paul Augustijn, Wee Fu Tan, Clemens Dirven

Abstract
The value of an ordinal global scale derived from combined clinical and CT data (clin/CT scale) to predict the clinical outcome in 112 patients shunted for presumed normal pressure hydrocephalus (NPH) was analysed. The clinical data were retrospectively collected, all CT scans were re-evaluated, and the clin/CT scale was determined blind to the results of further ancillary tests and to the post-surgical outcome. The scale ranked three classes of prediction: on the basis of clinical and CT characteristics, improvement after shunting was probable, possible, or improbable. The predictive value of the clin/CT scale for the subgroup of communicating NPH was established for two different strategies, depending on the strictness of selection criteria for shunting. In the subgroup of patients with presumed communicating NPH, the prevalence of shunt responsiveness was 29%; the best strategy was to shunt only patients with probable shunt-responsive NPH: the sensitivity was 0.54, the specificity 0.84, and the predictive accuracy 0.75, with a limited number of ineffective shunts (11%) and missed improvements (13%). The study illustrates its need to assess the pre-test probability of NPH based on combined clinical and CT data, before establishing the clinical usefulness of an ancillary test.

(4 Neurol Neurosurg Psychiatry 1993;56:251–256)

The syndrome of normal pressure hydrocephalus (NPH) is a rare and treatable cause of gait disorder, mental deterioration and urinary incontinence due to chronic hydrodynamic hydrocephalus. In spite of intensive research into tests to select appropriate candidates for a shunt, the success rate of CSF diversion ranges between 20% and 80%, depending on the aetiology of NPH and the severity of selection criteria. The clinical usefulness of ancillary tests remains controversial: some clinicians successfully shunted their patients on the basis of clinical and CT criteria, but others stated that these variables were unreliable and advocated further tests. Many clinicians still rely on the clinical signs, CT or MRI patterns, the clinical effect of one or more CSF taps, and sometimes cisternography to reach a therapeutic decision, because they are still reluctant to use more invasive or poorly available tests. Also, some promising results with CSF hydrodynamic tests have not been replicated in other studies. In addition, even recent articles describing the advantages of ancillary tests such as CSF hydrodynamic evaluation or MRI provide scant information on whether the pretest probability of NPH had been established on the basis of clinical and CT or MRI variables. As establishment of the pretest probability is essential to evaluate the additional predictive value of a test, we assessed the value of combined clinical and CT data to predict the post-surgical outcome in patients shunted for presumed NPH.

Methods

INCLUSION CRITERIA
The medical records of all patients shunted for presumed NPH between January 1980 and June 1989 in the four neurosurgical departments of Amsterdam were reviewed. In accordance with others, patients with both communicating and non-communicating hydrocephalus were included to consider a diagnosis of NPH. Although in some patients lack of improvement may have been due to shunt dysfunction, we used improvement after shunting as the gold standard for the diagnosis of NPH, designated shunt-responsive NPH (SR-NPH). The diagnosis of NPH was made on the basis of clinical criteria (gait disorder and mental deterioration with or without urinary urgency or incontinences), evidence of chronic hydrocephalus on CT scan, and normal CSF pressure at lumbar puncture. As this study was carried out blind to the results of ancillary tests such as cisternography, one or more CSF taps, or the results of pressure monitoring studies, no additional conditions were required for a tentative diagnosis of NPH. The evaluation of outcome after a shunt, including improvement, complications, residual deficit, and calculation of the benefit/harm ratio has been separately reported.

IMPROVEMENT AFTER SHUNTING
Improvement was ranked into 4 categories: no improvement; temporary improvement; when symptoms of NPH recurred after shunting without evidence of shunt malfunction (previously considered as unimproved); some improvement (I + 1): when gait distur-
bance, mental impairment, urinary incontinence or several ofthese signs improved only one point on pre-established ordinal scales, or when two independent assessors estimated that there was only slight improvement ofgait or mental function; marked improvement (I + 2): when one or several ofthese functions improved at least two points on the ordinal scales, or when at least two independent assessors agreed that the patient had substantially improved. The term “missed improvement” was applied to improved patients who would not have been shunted on the basis of combined clinical and CT data (false negative predictions). The term “ineffective shunt” referred to patients who would have been shunted on the basis of clinical and CT data, but did not improve after shunting (false positive predictions).

We could not avoid biases inherent in retrospective studies, based on unstandardised medical recording. For example, the positive effect ofshunting may have been overestimated: we tried to counterbalance potential overoptimistic evaluations of clinicians by scrutinising the notes of the nursing staff, and, when necessary, by soliciting the opinions of the patient, the family and the general practitioner.

The Predictive Value of a Global Clinical Plus CT Scale (Clin/CT Scale)

Calculation of the predictive value of combined clinical and CT data has been summarised elsewhere.45 In brief, two clinical assessors evaluated all available pre-operative clinical data and CT scans without knowledge of the results of ancillary tests and to the outcome after shunting. The CT scans were re-evaluated and a number of measurements was carried out, according to a pre-established protocol. Thereafter, the clinical and CT variables were grouped together into an overall estimation and classified as probable, possible, or improbable SR-NPH.

This retrospective collection of data followed by a prospective cohort study, including observer “blinding” to the results of a therapeutic intervention, has been labelled as a “retroactive” study.46 Such studies may be particularly useful when results of prospective studies are not available. As we did not establish the intra- and interrater variabilities of predictions established with the clin/CT scale, additional investigations should also validate our results.

Variables Included in the Global Clin/CT Scale

1) Clinical data: aetiology, time ofonset, rate of progression and severity of gait disorder, mental impairment and bladder dysfunction; predominance of one of the three “classical” signs of NPH; the presence of signs not consistent with NPH; evidence of cerebrovascular disease; presence of other diseases explaining (at least partially) the clinical picture; coexisting diseases affecting the general condition.

2) CT scan. The following measurements were made: frontal horn index (FHI), being the ratio between the maximal width of the frontal horns and the width of the whole brain at the same level; roundness of the frontal horns (0: not; 1: slight; 2: marked); lateral ventricle index, being the ratio between the width of the lateral ventricle at the level of the cells media and the width of the whole brain at the same level; third ventricle index, being the maximal width of the third ventricle divided by the width of the temporal horns, expressed in grade 0: not seen; grade 1: slightly enlarged; grade 2: considerably enlarged. Convexity sulci (SU) were graded as not present (0), present (1), enlarged (2), considerably enlarged (3); the fourth ventricle, rated as normal or enlarged. Periventricular frontal or occipital lucencies (PVL) were noted as present or absent. The presence of white matter lucencies (WML) or leuko-araiosis36 was rated as follows: no lucencies (0); discrete and sporadic WML (1); moderate WML (2); pronounced and diffusely distributed WML (3); and third ventricle further enlarged (4).

Predictive Categories of the Clin/CT Scale

The clinical and CT criteria used to define the categories of probability were derived from major clinical investigations,47-71 12 17 18 39 40 and review articles.19 20 The three categories were defined as follows:

Probable SR-NPH The combination of all the following clinical and CT criteria was required:

A) Clinically 1) Predominance of gait disturbances; 2) Mental impairment ranging from not clinically present to moderate; 3) Absence of another disease which might explain the clinical symptoms.

Urinary incontinence, known aetiology, or rapid clinical progression increased the probability of NPH, but were not necessary to rank a patient into this category.

B) CT: 1) Rounded frontal horns; 2) At least moderate ventriculomegaly (FHI 0.40 or higher); 3) Absence of moderate or severe cortical atrophy, and 4) Absence of moderate or severe WML. Marked enlargement of temporal horns and third ventricle further increased the probability but was not a necessary feature.

Improbable SR-NPH One of the following clinical or CT characteristics was sufficient to consider SR-NPH as improbable:

A) Clinically Absence of gait disturbances; marked predominance of dementia; dementia characterised by predominance of cortical dysfunction such as aphasia or agnosia; presence of another disease explaining partial or total clinical picture (for example, SAE).

B) CT one of the following variables: only slightly dilated ventricles (FHI < 0.40); marked cortical atrophy; moderate or severe WML without concomitant ventricular enlargement out of proportion to the degree of white matter involvement.
Possible SR-NPH This category contained all patients who could not be ranked as probable or improbable SR-NPH. For example, although CT displays moderate or severe white matter involvement, it is not sufficient to explain a huge ventriculomegaly. Possible combinations of 2 diseases such as SAE plus NPH or Alzheimer’s disease plus NPH were ranked into this category.

The Predictive Value of a Global Clin/CT Scale in Communicating NPH (CoNPH). As most problems of decision making are encountered in patients with CoNPH, we calculated the predictive value of a global Clin/CT scale in this subgroup. Two strategic scenarios were considered: 1) Strict selection, consisting of shunting only patients with probable SR-CoNPH; 2) Liberal shunting policy, consisting of shunting all patients with probable and possible SR-CoNPH.

The Predictive Value of the Clin/CT Scale Compared with That of a Hydrodynamic Test
Measurement of the conductance to CSF outflow (Cout) obtained during lumbo-ventricular perfusion tests in patients with presumed NPH has emerged as a highly predictive test by some investigators. We therefore calculated the additional value of this test in our patients with idiopathic probable and possible SR-CoNPH. This additional predictive value of Cout was calculated with the assumption that the same high sensitivity and specificity would have been achieved as it was by Børgesen et al.

Results
The pre-operative clinical data of 166 consecutive patients shunted for presumed NPH were reviewed. Four patients were excluded because they had an accompanying diagnosis which might have interfered with signs of NPH and 10 were excluded because their medical records were lost or did not contain sufficient information on post-surgical outcome. A further 40 patients were withheld from this study because of insufficient preoperative clinical data (N = 8), unavailable preoperative CT scans (N = 30), or both (N = 2).

Thus 112 patients plus their CT scans remained for further analysis. The aetiology of hydrocephalus in this group was idiopathic in 91; secondary and communicating in 9; non-communicating in 14, and in 13 was associated with aqueduct stenosis. Table 1 correlates the degrees of improvement with the predictive categories of the Clin/CT scale. Improvement after shunting was 35% in the whole group, and 29% in patients with communicating hydrocephalus.

Predictive Value of the Clin/CT Scale in CoNPH

Probable SR-CoNPH. Twenty six patients (27% of the total CoNPH group) were allocated to this category. The median age was 72 years, range 45–85; 11 were males. Nine had idiopathic NPH with a known cause and 17 were idiopathic. On CT the median FH index was 0.46, range 0.40–0.56. In 22 patients the sulci were absent (0) or visible (1), in 18 the temporal horns were considerably enlarged; WML grade 1 were seen in 10 patients. In this group, overall permanent improvement was 58% (95% CI 37–77%), with no significant difference between idiopathic (65%, 95% CI 38–86%) and secondary (44%, 95% CI 14–79%) CoNPH.

Possible SR-CoNPH was considered in 33 patients, with a median age of 76 years, range 62–84; 23 were males. All had idiopathic NPH. On CT the median FH index was 0.45 (range 0.39–0.57). Cortical sulci were absent or normal in 21/33, slightly enlarged in 10/33 and considerably enlarged in 2/33. WML were absent in 18, grade 1 in 8 and grades 2 or 3 in 7 patients. Shunting all patients with “possible SR-CoNPH” would have led to substantial improvement in 15% (95% CI 5–32%).

Improvable SR-NPH was diagnosed in 39 patients. All had idiopathic CoNPH. The median age was 74 years, range 45–84 years. On CT, the median FH was 0.42 (range 0.37–0.50). The main reasons for rejecting SR-NPH were: marked predominance of dementia (12); dementia including substantial cortical dysfunction consistent with senile dementia of Alzheimer’s type (7); clinical and CT features consistent with SAE (29); CT showed marked atrophy (3). Many patients had more than one reason to be categorised as improvable SR-NPH. Shunting patients of this category would have led to substantial amelioration in 6% (95% CI 0–9%) and slight improvement in 13% (95% CI 4–27%).

Table 2 shows the predictive value of the first strategy, that is, shunting only patients with probable SR-CoNPH. It would have led to a predictive accuracy of 75% (95% CI
Table 3  Strategy 2: liberal shunting criteria

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Substantial</th>
<th>Some</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shunt all probable and possible SR-CoNPH</td>
<td>17</td>
<td>6</td>
<td>36</td>
<td>59</td>
</tr>
<tr>
<td>Do not shunt improbable SR-CoNPH</td>
<td>0</td>
<td>5</td>
<td>34</td>
<td>39</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>11</td>
<td>70</td>
<td>98</td>
</tr>
</tbody>
</table>

Sensitivity for substantial improvement: 1-0 (17/17)
Specificity: 0-68 (23/28)
Positive predictive value: 39% (23/59)
Negative predictive value: 87% (34/39)
Accurate predictive rate: 58% (57/98).

Table 4  Predictive value of \( C_{out} \) in NPH

<table>
<thead>
<tr>
<th>( C_{out} ) (mmHg/ml/Min)</th>
<th>N = 64</th>
<th>Improved</th>
<th>Trans-improved</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( &lt; 0.8 )</td>
<td>45</td>
<td>6</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>( &gt; 0.08 )</td>
<td>0</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>19</td>
<td>64</td>
<td></td>
</tr>
</tbody>
</table>

Prevalence of SR-NPH: 70%
Sensitivity 1-0; specificity 0-68
Positive predictive value: 88%
Negative predictive value: 100%
Accurate predictive value: 91%

Sensitivity and specificity
With strategy 1 (strict selection), the sensitivity of shunting only probable SR-CoNPH was 0-71 (12/17, 95% CI 0-44-0-90) for substantial improvement, 0-54 (15/28, 95% CI 0-34-0-72) for overall improvement, and the specificity was 0-73 (8/11, 95% CI 0-39-0-94).

With strategy 2 (shunting both probable and possible SR-CoNPH) the sensitivity would have increased to 1 (95% CI 0-80-1-0) for substantial improvement and 0-82 (95% CI 0-63-0-80) for overall improvement; the specificity would have been lowered to 0-49 (95% CI 0-36-0-61).

Table 4 shows the correlation of the values of \( C_{out} \) and outcome after shunting as described by Bergersen et al. A sensitivity of 100% (95% CI 92–100%) and specificity of 68% (95% CI 43–87%) was found, provided that the cut off for shunting was determined at \( C_{out} \leq 0.08 \), and that temporary improvement would have been considered as no improvement. This led to a positive predictive value of 88% and an accurate predictive rate of 91% (95% CI 81–96%) in a patient group with a pretest probability of shunt-responsiveness of 70%.

Table 5 lists the predictive value of a Clin/CT scale in our patients with idiopathic probable and possible SR-CoNPH. In this subgroup with a pre-test probability of shunt-responsiveness of 38% (19/50), the accurate predictive rate would have been 72% (36/50).

Table 6 shows that the application of \( C_{out} \) in the same subgroup would have led to an accurate predictive rate of 80% (95% CI 67–89).

Discussion
To our knowledge, this is the first study on NPH in which the predictive values, the sensitivity, and the specificity of combined clinical and CT criteria with observers “blinded” to the results of further tests and to the post-surgical outcome have been calculated. We established the predictive value of such a Clin/CT scale to assess whether clinical and CT criteria are as unreliable as asserted by investigators who reported on the clinical usefulness of ancillary tests.

As we studied only patients who had been shunted, we inevitably introduced a selection bias. In spite of this preselection, the three predictive classes of the Clin/CT scale were equally distributed (table 1) and the prevalence of shunt responsiveness in the group of CoNPH was about 29%, which probably reflects common practice. An unresolvable problem was that, in spite of a rather high rate of shunt revisions, failure to improve in

Table 5  Predictive value of the best strategy with the Clin/CT scale in idiopathic probable and possible SR-CoNPH

<table>
<thead>
<tr>
<th>Improvement</th>
<th>substantial</th>
<th>Some</th>
<th>No improvement</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shunt probable SR-CoNPH</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Do not shunt possible SR-CoNPH</td>
<td>5</td>
<td>3</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>6</td>
<td>31</td>
<td>50</td>
</tr>
</tbody>
</table>

Prevalence of SR-NPH: 70%
Sensitivity 1-0; specificity 0-68
Positive predictive value: 86% (19/23)
Negative predictive value: 100% (21/21)
Accurate predictive rate: 80% (40/50)
some patients may have been due to inadequate shunts rather than to a misdiagnosis of NPH.\(^7\)

Shunting only patients with probably SR-CoNPH led to the highest predictive accuracy. With this policy (table 2), neurologists would have referred less than one third of the shunted patients to the neurosurgeon, with limited erroneous decisions: only 5% of the whole group with presumed SR-CoNPH would have been erroneously withheld from substantial improvement and 13% of the whole group would have had ineffective surgery.

Thus in centres with limited technological facilities, shunting only patients with probable CoNPH on the basis of a global clinical/CT scale would be justified, especially in view of the serious surgical risks and the very low rate of marked improvement in the categories of possible and improbable SR-CoNPH. The major drawback is that these patients and their families must accept a 42% chance of ineffective surgery and a 5% chance of being erroneously withheld a highly beneficial procedure.

In patients with possible SR-CoNPH—the category with the highest diagnostic uncertainty—the likelihood of improvement after a shunt is estimated at 30–40%.\(^{11,14}\) This contrasts with our experience, as shunting all patients with possible SR-CoNPH would have resulted in a disappointingly low percentage of substantial benefit (15%). As serious shunt-related residual morbidity approximates 7–8%,\(^9\) a very reserved therapeutic attitude towards patients with possible SR-CoNPH would be justified, as the benefits would scarcely outweigh the risks of serious post-surgical harm. It is possible that our criteria to classify hydrocephalic patients into this category were too liberal, but it is also possible that some investigators included improvement of limited clinical importance to calculate the benefits of a shunt.

Shunting patients with improbable SR-CoNPH appeared as a desperate attempt to reverse an irreversible process of mental deterioration, most probably due to degenerative brain disease or to SAE. Our large retrospective study\(^8\) revealed that many neurologists and neurosurgeons submitted patients with improbable NPH to a neurosurgical "last-chance" procedure. As very few patients of this group fare only slightly better after a shunt, submitting them to a shunting procedure is unjustified as the benefit/harm ratio is unacceptably low.

With a combined clin/CT scale the most accurate predictive rate in the whole group of CoNPH was 75% (strategy 1). This predictive accuracy is higher than commonly assumed, but was mainly due to the high predictive value of a negative decision, that is, not to shunt patients with possible or improbable SR-CoNPH. Even with strict selection, the positive predictive value of a combined Clin/CT scale remained low (58%). Thus by combining clinical and CT criteria for a global predictive scale, we did not succeed in eliminating the need for non-invasive, easily accessible tests with reproducible results, assuming, of course, that their sensitivity, specificity and predictive values would be meaningfully higher than those obtained with combined clinical and CT criteria. Continuous intracranial CSF pressure recording\(^6\) and calculation of \(C_{\text{out}}\) have been described as the best predictive tests. Unfortunately, at least in the Netherlands, these techniques are inaccessible to most clinicians.\(^{23}\) In the best hands, the accuracy of \(C_{\text{out}}\) was very high, reaching levels of 90% or even higher, but this was obtained in a group of patients in whom a large majority may have presented with a typical SR-NPH, as 70% improved after a shunt.\(^6\) Other reports have illustrated that CSF infusion tests may be less accurate.\(^{26-28}\)

In our 50 patients with idiopathic communicating NPH there was a large spectrum of probabilities indicating SR-NPH, explaining that the pretest probability of SR-NPH was only 39%: this approximates to the percentage encountered in common neurological practice. In this group, the most difficult regarding decision making, the predictive accuracy of \(C_{\text{out}}\) would have been 80%, which is only 8% better than that obtained with the best strategy based on clinical and CT criteria. The main advantage of assessing the \(C_{\text{out}}\) would have been to avoid missed substantial improvements in 10% of the patients.

In conclusion, our study suggests that further studies on NPH should include assessment of the pretest probability based on combined clinical and CT variables. Only then can one assess the usefulness of an ancillary test to assist clinicians in their decisions to shunt or not to shunt patients with presumed NPH.

We acknowledge the useful suggestions of Professors H van Crel, M Vermeulen and Drs A Hijdra, GAG Davies and RP Golding.

3 Jacobs L, Conti D, Kinkel WR, Manning EJ. "Normal-pressure" hydrocephalus. Relationship of clinical and radiographic findings to improvement following shunt surgery. JAMA 1976;235:10–12.
11 Petersen RC, Mokri B, Laws ER Jr. Surgical treatment of...