Reliability and Interobserver Variability of Evans’ Index and Disproportionately Enlarged Subarachnoid Space Hydrocephalus as Diagnostic Criteria for Idiopathic Normal Pressure Hydrocephalus

Abstract

Background: The image diagnosis of idiopathic normal-pressure hydrocephalus (iNPH) is based on the ventriculomegaly, whose criterion is an Evans’ Index (EI) >0.3. Recently, disproportionately enlarged subarachnoid space hydrocephalus (DESH) has been proposed as a morphological characteristic to iNPH. Several studies cast doubt on the reliability of these criteria in the diagnosis of iNPH. Furthermore, interobserver differences of these criteria have not yet been investigated. The objective of this study was to assess the diagnostic reliability and interobserver variability of EI and DESH. Materials and Methods: The preoperative magnetic resonance (MR) images of 84 definite iNPH patients were retrospectively evaluated by a neuroradiologist (NR) and physical therapist (PT). They independently assessed the EI and DESH. The MR images were evaluated preoperatively by a neurosurgeon (NS). The results were showed in mean (standard deviation). Results: The mean age was 78.4 (6.3) years (male:female = 49:35). The mean EI was 0.33 (0.04), 0.32 (0.04), and 0.31 (0.03) for NS, NR, and PT, respectively (P < 0.0001). The rate of accurate diagnosis of iNPH with EI >0.3 was 74%, 66%, and 61% for NS, NR, and PT, respectively, and there was a moderate level of agreement. By contrast, there was a substantially lower level of accuracy in assessment with DESH for all three evaluators as 50%, 44%, and 27% for NS, NR, and PT, respectively, again with a moderate level of agreement. However, the rates of patients fulfilling both EI >0.3 and DESH were remarkably lower than either of the two parameters individually at a mere 37%, 30%, and 16% for NS, NR, and PT, respectively, with a low level of agreement between the rates. Conclusion: This study suggests that DESH cannot be a diagnostic criterion for iNPH. If EI >0.3 and DESH were both necessary to diagnose iNPH, then more than 70% of patients would have been misdiagnosed and would have been deprived of the chance of treatment and its benefits. These results request a paradigm shift in the concepts of iNPH.

Keywords: Disproportionately enlarged subarachnoid space hydrocephalus, Evans’ Index, idiopathic normal pressure hydrocephalus, interobserver variability, ventriculocisternal shunt

Introduction

Idiopathic normal-pressure hydrocephalus (iNPH) is a syndrome mainly found in elderly people and comprises a triad of gait disturbance, dementia, and urinary incontinence that can be improved by cerebrospinal fluid (CSF) shunting.[1] In the image diagnosis of iNPH, ventriculomegaly is a mandatory condition that is defined as Evans’ Index (EI) >0.3 in both international and Japanese guidelines[2,3] until 2012 when the revised Japanese guidelines introduced disproportionately enlarged subarachnoid space hydrocephalus (DESH) as a necessary component in the image diagnosis of iNPH.[4]

EI in the diagnosis of iNPH is defined as a frontal horn ratio calculated at the maximum frontal horn ventricular width divided by the transverse inner diameter of the skull at the same horizontal plane.[4,5] As the shapes of the cranium and the lateral ventricle differ depending on the level of the horizontal plane, there is a possibility of error in the selection of the level for EI calculation. Furthermore, the judgment of DESH is subjective and therefore poses a potential risk of interobserver variability. Therefore, both parameters may vary from one evaluator to another (interobserver variability).

The criteria of EI >0.3 were derived from the previous definition of iNPH. EI in the diagnosis of iNPH is defined as a frontal horn ratio calculated at the maximum frontal horn ventricular width divided by the transverse inner diameter of the skull at the same horizontal plane.[4,5] As the shapes of the cranium and the lateral ventricle differ depending on the level of the horizontal plane, there is a possibility of error in the selection of the level for EI calculation. Furthermore, the judgment of DESH is subjective and therefore poses a potential risk of interobserver variability. Therefore, both parameters may vary from one evaluator to another (interobserver variability).
ventriculomegaly for child hydrocephalus studied by means of pneumoencephalogram. It was not a result of a direct comparison between iNPH patients and healthy controls. DESH was proposed as a unique morphological configuration to iNPH, and Hashimoto et al. confirmed its diagnostic value for iNPH. However, many elderly people are known to show asymptomatic ventriculomegaly with imaging features of iNPH on magnetic resonance imaging (MRI) (AVIM) without presenting symptoms of iNPH. In addition, it is known that there are many iNPH patients who do not show the expected DESH findings, but whose symptoms improved by CSF shunting. As for EI, about 20%–30% of the healthy elderly people are known to have EI >0.3. These studies suggest that neither EI nor DESH might be appropriate image diagnostic criteria for iNPH. A very recent study showed that DESH and other morphologic MRI markers including sharp callosal angle are not proper diagnostic image criteria.

Before the publication of the guidelines for diagnosis and management of iNPH, there had not been a clear definition of ventriculomegaly for iNPH. One of the authors (KT, who is also the neurosurgeon in this study: NS) started ventriculoatrial (VA) shunt for iNPH before the publication of Japanese guidelines in 2004. He placed VA shunt only for the patients with a positive lumbar tap test whose brain computed tomography or MRI showed anterior horn rounding, third ventricular ballooning, sylvian fissure enlargement, or convexity subarachnoid space enlargement, which are not commonly observed in normal adults. He found that some of these cases did not meet the criteria of either EI >0.3 or DESH, but he nevertheless observed satisfactory postoperative outcomes for these iNPH patients.

After the publication of the guidelines, neurologists and neurosurgeons again came to recognize iNPH as a possible cause of gait disturbance, dementia, and urinary incontinence. Once the image diagnostic criterion of ventriculomegaly is defined as EI >0.3, the patients who do not fulfill this criterion would not be diagnosed as iNPH and be left untreated even though they show the typical triad of the disease without other responsible disorders. Since 2011, DESH has been accepted as a conjunct image indicator of iNPH along with EI in Japan. This may further narrow the window for the chance to receive shunt surgery for the possibly treatable iNPH patients.

The purpose of this study was to assess the reliability of EI and DESH in diagnosing iNPH by estimating retrospectively the MR images of definite iNPH patients. The second objective of the study was to examine the variability and agreement between the evaluators for their judgments of EI and DESH criteria.

**Materials and Methods**

From April 2007 to December 2014, we had 288 definite iNPH patients out of 330 VA shunts for probable iNPH at Kashiwatanaka Hospital NPH Center. All the shunted patients were installed with a programmable valve. Of these definite iNPH patients, 84 patients had clear preoperative axial and coronal brain MRI scans as well as follow-up period of over 1 year. These were the definite iNPH patients included in this study because their scans could be used to assess both EI and DESH.

The NS who performed the VA shunts had evaluated EI and DESH before the surgery. All patients except one received a lumbar tap test with a positive response. One patient had an unruptured cerebral aneurysm and the lumbar tap test was not performed. However, his MRI findings fulfilled the revised Japanese guideline. A neuroradiologist (NR) and a physical therapist (PT) who were blind to the patients’ clinical information independently calculated EI and evaluated DESH for these cases. EI was computed using an axial section of MRI T1 images according to the calculation method described in the guideline not by the original method. DESH was judged based on the original paper by Kitagaki et al.

The mean EIs evaluated by three raters were compared by means of one-way ANOVA. The rate of a positive finding for iNPH was calculated for the three raters. Student’s test was used to compare the means of baseline variables. Paired sample test was used to compare scores before shunt surgery and at 3 months. Fleiss kappa test was used to assess the agreement between all the three raters, and Cohen’s kappa was used to assess the significance of difference/disagreement between the raters. $P < 0.05$ was set as statistical significance. The results were showed in mean (standard deviation).
Ethical considerations

The data collection was approved by the Institutional Review Board of Kashiwa Takana Hospital and informed consent was obtained in a written form.

Results

There were 49 males and 35 females. The overall mean age was 78.4 (6.3) years, while the mean age for males was 78.1 (6.8) years, and for females, it was 78.7 (5.6) years. There was no gender difference in the mean age ($P = 0.66$) [Table 1].

With all patients being definitive iNPH patients, an improvement was observed in the clinical rating scales of the patients after the shunt surgery. Table 2 briefly demonstrates the changes in clinical scores before shunt procedure and the best scores after surgery for the Modified Rankin Scale, mini-mental state examination, and Japanese idiopathic normal-pressure hydrocephalus Grading Scale.\[10\]

The mean EI was 0.33 (0.04), 0.32 (0.04), and 0.31 (0.03) for NS, NR, and PT respectively ($P < 0.0001$). Table 3 shows the numbers and percentages of the cases correctly diagnosed as iNPH patients (EI >0.3) by three evaluators. The percentage of the cases correctly diagnosed as iNPH with EI >0.3 criteria was between 61% and 74%, and the level of agreement was moderate ($\kappa = 0.59$).

Table 4 shows the numbers and percentage of cases correctly diagnosed as iNPH with DESH evaluated by the same three evaluators. The percentage of cases correctly as iNPH using the DESH criteria was 50% or less, with a moderate level of agreement between them ($\kappa = 0.522$).

Table 5 shows the numbers and percentages of cases by their EI and DESH status as determined by the NR. The percentage of the patients fulfilling both the diagnostic criteria proposed in Japanese guidelines was lower than 30%. When we compared the clinical rating scales of these cases before shunt and the best scores over 3-month follow-up after VA shunt, they were very similar to the overall mean differences and significance level, indicating that this group of patients was not different from the total group.

Interobserver agreement was assessed between the pairs of raters. In all evaluations, NS had the highest rate of correctly identifying iNPH. Therefore, using NS as the reference point, the NS-NR agreement and NS-PT agreement for EI scores were moderate ($\kappa = 0.58$ and 0.66, respectively), while the disagreement was significant between NS and PT ($P = 0.003$). For DESH, the NS-NR agreement was good ($\kappa = 0.79$), but the NS-PT agreement was poor ($\kappa = 0.41$), and disagreement was again significant between NS-PT ($P < 0.001$). When the combined score was examined, NS-NR agreement was again moderate ($\kappa = 0.68$), but NS-PT agreement further worsened ($\kappa = 0.30$) and disagreement remained significant ($P < 0.001$).

Discussion

This study suggests that DESH cannot be a reliable diagnostic image criterion for iNPH because over 50% of definite iNPH patients did not show DESH [Table 4].
Furthermore, it also indicates that over 30% of definite iNPH patients can be overlooked even by applying EI >0.3 for iNPH diagnosis [Table 3]. We also disclosed that there were not negligible interobserver differences for both calculating EI and evaluating DESH.

Recent studies from outside Japan support our findings on DESH.\textsuperscript{15,17,18} However, many recent studies from Japan support the diagnostic validity of DESH.\textsuperscript{10,19-21} Since the publication of the guidelines in 2004,\textsuperscript{31} Japanese neurosurgeons have operated on almost exclusively the iNPH patients with DESH. The patients with iNPH triad without DESH had only a small chance to be operated on or even to receive lumbar tap test for the diagnostic examination. This biased situation in Japan may explain the discrepancy between the conclusions on the validity of DESH from outside Japan and from Japan. The diagnostic value of DESH for iNPH has been derived from comparing the images of only 11 iNPH cases and the images of the corresponding number of patients with Alzheimer’s disease and vascular dementia.\textsuperscript{17} The MRI images of the healthy elderly were not included. Iseki et al. disclosed that there are many asymptomatic cases with DESH.\textsuperscript{19} DESH is a unique morphological configuration. Therefore, apart from the diagnostic standpoint, it is necessary to investigate the underlying mechanisms that create this image finding.

EI >0.3 has been a universally accepted definition of ventriculomegaly and the most frequently used diagnostic criterion.\textsuperscript{2-5} However, not all the iNPH patients with EI >0.3 improved by shunt surgery.\textsuperscript{8,15,22} Agerskov et al. reported only 68.5% improvement rate for the iNPH patients with EI >0.3.\textsuperscript{15}

The definition of ventriculomegaly of EI >0.3 is derived from the study of pediatric hydrocephalus.\textsuperscript{6} It was modified for CT or MRI to define the ventriculomegaly of iNPH.\textsuperscript{2-4} Again, we had no direct comparisons between iNPH patients and healthy controls. About 20%-30% of the healthy elderly people are known to have EI >0.3.\textsuperscript{12,13}

As we stated in the materials and methods part, we had 288 definite iNPH cases of 330 probable iNPH cases received VA shunt. Although not all of them did not fulfill the criterion of EI >0.3, the improvement rate was 87.3%. Although Agerskov et al. did not state the type of surgery in their article,\textsuperscript{15} ventriculoperitoneal (VP) shunt was used as surgical treatment. We can expect this from other European studies.\textsuperscript{23-25} Our improvement rate was significantly better than that reported by Agerskov et al.,\textsuperscript{15} (P < 0.0001, Fisher’s exact test) despite that our study included the patients with EI ≤0.3. Hashimoto et al.\textsuperscript{16} reported an improvement rate of 80% for VP shunt in iNPH fulfilling both EI >0.3 and DESH. Kazui et al. reported 75% improvement for lumboperitoneal (LP) shunt\textsuperscript{22} for iNPH fulfilling both EI >0.3 and DESH. Although it is easy to just list the improvement rate of the reported studies, it is not fair to compare the improvement rates since the study designs were different. However, our study was quite different from others which contained definite iNPH without EI >0.3, and we applied VA shunt. Liu et al. reported 78% improvement rate for VA shunt (EI >0.3).\textsuperscript{26} Since they did not mention to DESH, they operated on the patients without DESH. VA shunt may be a better treatment modality than VP or LP shunt for iNPH.

Taking these facts together into consideration, ventriculomegaly defined as EI >0.3 cannot be a mandatory condition to diagnose iNPH. However, since it covered about 70% of the definite iNPH, it can still have some diagnostic value for iNPH.

To our knowledge, this is the first study to investigate the interobserver variability for both calculating EI and evaluating DESH. We have several image diagnostic criteria and clinical rating scales in the neurosurgical field. Rating scales commonly used in neurosurgery for subarachnoid hemorrhage (SAH) are coma scales\textsuperscript{27,28} and image criteria.\textsuperscript{29} These scales and criteria are used to predict the outcome of the disease or to make a treatment plan. We do not make the diagnosis of SAH by means of these scales and criteria. In addition, a high level of accuracy and consistency among evaluators is confirmed for these scales and diagnostic criteria.\textsuperscript{30-32}

Quite different from these scales and criteria, EI >0.3 and DESH are diagnostic image criteria for iNPH. Being directly involved in the diagnostic procedure, these scales need to be highly accurate and consistent individually and in combination (as per the current Japanese guidelines) in making the diagnosis of iNPH. However, we observed the contrary. To calculate EI, it is necessary to select a proper slice of CT or MRI. Each slice has different sizes and shapes of the cranial circumference, and it is possible to induce the Delboeuf illusion.\textsuperscript{33} It is also necessary to set boundaries CSF and brain parenchyma. The boundary between the ventricle and brain parenchyma is not always clear cut. These facts may be the cause of interobserver variability of EI. Evaluating DESH is subjective, and the possibility of interobserver variability is inevitably involved. Interobserver variability can be avoided by introducing computer-assisted automated measurement system.\textsuperscript{34} However, the reliability of EI and DESH to diagnose iNPH is a quite another problem. Our study indicates that the reliability of DESH and EI may not improve.

**Conclusion**

In conclusion, the results of this study suggest that the paradigm shift is inevitable in the diagnosis and treatment of iNPH. The current image diagnostic criteria of the Japanese guidelines are not reliable in diagnosing iNPH patients, and there is an urgent need to revise them. Using the current guidelines, the vast majority of treatable iNPH patients will remain untreated and suffer...
unnecessarily even though they clearly have the typical INPH triad. The results of this study also request the randomized controlled trial to determine the best surgical modality for INPH.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

29. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm...


