Reliability of CSF turbulence and choroid plexus visualization on fast-sequence MRI in pediatric hydrocephalus

Curtis J. Rozzelle, MD,¹ Casey Madura, MD, MPH,² and Ron W. Reeder, PhD³

¹Department of Neurosurgery, Division of Pediatric Neurosurgery, The University of Alabama at Birmingham and Children's of Alabama, Birmingham; ²Department of Neurosurgery, Michigan State University and Helen DeVos Children's Hospital, Grand Rapids, Michigan; and ³Department of Pediatrics, University of Utah, Salt Lake City, Utah

OBJECTIVE Endoscopic third ventriculostomy with choroid plexus cauterization for the treatment of neonatal and infant hydrocephalus has gained popularity in the past decade. Identifying treatment failure is critically important. Results of a pilot study of 2 novel imaging markers seen on fast-sequence T2-weighted axial MRI showed potential clinical utility. However, the reliability of multiple raters detecting these markers must be established before a multicenter validation study can be performed.

METHODS Two sets of de-identified single-shot T2-weighted turbo spin-echo axial images were prepared from scans of patients before and after they underwent endoscopic third ventriculostomy with choroid plexus cauterization between March 2013 and January 2016. The first set showed the lateral and third ventricles for visualization of turbulent CSF dynamics, and the second set showed the lateral ventricular atria for choroid plexus glomus detection. Three raters (Group 1) received written instructions before evaluating each image set once and then again 1 week later. Another 8 raters (Group 2) evaluated both image sets after oral instruction and group training on a pretest image set. Fleiss' kappa coefficients with 95% CIs were calculated for intrarater and interrater reliability in Group 1 and interrater reliability in Group 2.

RESULTS Intrarater reliability kappa coefficients for Group 1 were ≥ 0.74 for turbulence and ≥ 0.80 for choroid plexus; their interrater kappa coefficients at the initial assessment were 0.50 (95% CI 0.37–0.62) and 0.56 (95% CI 0.43–0.69), respectively. The Group 2 interrater kappa scores were 0.82 (95% CI 0.78–0.86) for turbulence and 0.62 (95% CI 0.58–0.66) for choroid plexus.

CONCLUSIONS With minimal training, intrarater reliability on visualization of turbulence and the choroid plexus was substantial, but interrater reliability was only moderate. After modestly increasing training, interrater reliability improved to near perfect and to substantial reliability for visualization of turbulence and choroid plexus, respectively. Given adequately trained observers, a multicenter investigation into the validity and potential clinical utility of the imaging markers seems feasible.

<https://thejns.org/doi/abs/10.3171/2017.7.PEDS17241>

KEY WORDS hydrocephalus; endoscopic third ventriculostomy; choroid plexus cauterization; magnetic resonance imaging; interrater reliability; intrarater reliability

ENDOSCOPIC third ventriculostomy with choroid plexus cauterization (ETV/CPC) has regained popularity for the treatment of hydrocephalus in very young children. Patient outcomes in this population after ETV/CPC remain under investigation, but patients who undergo this treatment have the potential for a shunt-free life if it is successful.

Treatment failure is the most common postoperative complication reported in the surgical management of hy-

drocephalus. Therefore, the identification of treatment failure (whether caused by ETV failure or shunt malfunction) is critically important. A recent report by Pindrik et al.³ described 2 novel radiographic markers associated with ETV/CPC failure. Both markers were identified using fast-sequence (FS) T2-weighted axial MRI sequences. The first marker is CSF turbulence, defined as the following: 1) the presence of an irregular, heterogeneous, or hypointense signal intensity in a pattern that suggests nonlinear

ABBREVIATIONS CP = choroid plexus; CPC = CP cauterization; ETV = endoscopic third ventriculostomy; FS = fast sequence.

SUBMITTED May 5, 2017. **ACCEPTED** July 13, 2017.

INCLUDE WHEN CITING Published online October 27, 2017; DOI: 10.3171/2017.7.PEDS17241.

CSF flow, 2) the presence of this signal intensity in at least 2 of 3 supratentorial ventricles, and 3) distinction of this signal intensity from imaging artifacts caused by patient movement.

The second marker is the presence or absence of a robust choroid plexus (CP) on the same images, which can serve as a marker of CPC extent, which correlates with treatment failure or success.² The criterion for this designation is either 1) clear visualization of the glomus in at least 1 ventricular atrium, which denotes the “presence” of CP, or 2) the lack, or diminutive size, of the glomus in both ventricular atria, which denotes the “absence” of CP.

The presence of either of these markers was associated with treatment failure in the pilot study. Combining the presence of both markers yielded a stronger association with treatment failure (sensitivity 77%, specificity 81%).³

For these findings to be valid and have clinical utility, they must be reliable, meaning that their observation must be consistent among multiple clinical observers. Therefore, we performed a study of the interrater and intrarater reliabilities of these 2 findings by using T2-weighted axial FS-MRI obtained in hydrocephalic infants and toddlers treated with ETV/CPC.

Methods

A retrospective review of MRI scans obtained previously for clinical purposes in hydrocephalic patients younger than 2 years of age was performed. The imaging modality used in this study was the shunt-function or FS-MRI incorporating single-shot T2-weighted turbo spin-echo sequences (axial, coronal, and sagittal). Patients included in this study underwent ETV/CPC between March 2013 and January 2016. Two separate sets of de-identified images were created for the initial review. The first set consisted of axial images of the lateral and third ventricles, and the second set consisted of axial images representing the region of the glomus of the CP. Each set was preceded by a slide using the criteria listed above. This survey methodology, which involves nonsensitive de-identified image data, qualified for a Category 2 exemption from institutional review board review.

These image sets were used for the evaluations described here. Each rater was blinded to patient identity, procedure/scan date, and clinical outcome. Three raters (1 board-certified pediatric neurosurgeon, 1 board-eligible pediatric neurosurgeon, and 1 board-certified neuroradiologist) evaluated each image set for the presence or absence of CSF turbulence and the CP. See Fig. 1 for examples from the image sets depicting each possible finding. The same 3 raters were asked to rate the same set of images, in a different random order, approximately 1 week later without any verbal prerating instruction being provided. These data were used to assess interrater and intrarater reliability. Intrarater reliability was assessed separately for each of these 3 raters. Interrater reliability was evaluated at the initial assessment and again 1 week later.

A second round of rating was conducted by 8 pediatric neurosurgeons (7 board certified, 1 board eligible), all of whom represented investigator members of the Hydrocephalus Clinical Research Network. In this instance, pre-

FIG. 1. Imaging obtained in a 6-month-old infant with hydrocephalus related to intraventricular hemorrhage in the setting of extreme prematurity. **A:** CP present. In this axial T2-weighted image taken before ETV/CPC, the CP, including the robust glomus in the right lateral ventricle (arrow), can easily be seen bilaterally. This image would be scored as showing the presence of the CP. Note that the presence of the glomus in only 1 ventricle is sufficient to score the CP as present. **B:** No CSF turbulence. In this axial T2-weighted image taken before ETV/CPC, representative of all images of the supratentorial ventricles, no distortion of the CSF in any location in either lateral ventricle can be seen. Even if the third ventricle showed distortion in the CSF that is not identified as motion/streak artifact, this image would be scored as showing no CSF turbulence, because the requirement is for 2 of the 3 supratentorial ventricles to show turbulence. **C:** CP absent. In this axial T2-weighted image taken after ETV/CPC, the CP is no longer obvious. The stubs of tissue noted along the inferior aspect of the thalamus (arrow) are not enough to score the CP as present in the rubric presented in the study. This image would be scored as showing the absence of the CP. **D:** CSF turbulence. In this axial T2-weighted image, obtained after ETV/CPC, obvious distortion of the CSF that is not caused by motion artifact can be seen and is present in both lateral ventricles (arrows). Therefore, regardless of the presence or absence of CSF turbulence in the third ventricle, this image would be scored as positive for CSF turbulence. It is important to note that not every image in the series showed this turbulence. The presence of turbulence on even 1 image is sufficient. CSF turbulence is also scored as present even if the distortion is visible in each ventricle on separate images.

rating education was provided with the same image sets used for Group 1. Prerating education consisted of in-person didactic instruction, a question-and-answer session, and group rating of some sample images. Only 1 round of rating was performed in this second stage, so only interrater reliability was assessed.

Data were collected and analyzed by our statistician coauthor (R.W.R.). Intrarater and interrater reliabilities were assessed using Fleiss' kappa statistic. Fleiss' kappa is a generalization of Scott's pi statistic, which is used in place of Cohen's kappa to measure intrarater and interrater reliabilities in experiments with more than 2 raters.¹

TABLE 1. Intrarater reliability

Assessment	Fleiss' Kappa (95% CI)
Turbulence	
Rater 1	0.82 (0.60–1.00)
Rater 2	0.81 (0.59–1.00)
Rater 3	0.74 (0.52–0.96)
CP	
Rater 1	0.80 (0.58–1.00)
Rater 2	0.90 (0.68–1.00)
Rater 3	0.85 (0.63–1.00)

Each rater reviewed all subjects twice (3 raters, 79 subjects).

Approximate 95% CIs were computed using the estimated standard error for kappa. Analysis was performed using the MAGREE macro in SAS 9.4 (SAS Institute).

Results

In the first phase of this study, in which only 3 raters were involved, a total of 79 axial images showing the lateral and third ventricles were compiled in random order for review. In the case of the CSF turbulence assessment, some axial images encompassed both the lateral and third ventricles together, and others required separate images for them. In the case of the CP, the image that showed the “most” CP was used.

Assessment of intrarater reliability was performed by simply randomizing the presentation order of images and performing the assessment a second time approximately 1 week later. Intrarater reliability (Table 1) for the presence or absence of the CP was near perfect ($\kappa = 0.8–0.9$), and for CSF turbulence, near-perfect agreement ($\kappa = 0.74–0.82$) was also found.

Assessment of the interrater reliability was completed in 2 steps. The first stage, which involved 3 raters (2 neurosurgeons and 1 radiologist) (Table 2) yielded mixed results. The interrater reliability at first assessment for the presence or absence of CSF turbulence and for CP revealed only moderate agreement ($\kappa = 0.50$ and 0.56 , respectively).

The high intrarater reliability and modest interrater reliability suggest that at least some of the disagreement could be attributable to differences in interpretation of the criteria. Therefore, a second round of testing was initiated to investigate the effect of didactic instruction and demonstration before the raters evaluated the CSF turbulence image set only. The instructions and demonstration were given in person by the first author (C.J.R.) during a semiannual meeting of Hydrocephalus Clinical Research Network investigators. A total of 8 raters were then polled, and the results were aggregated (Table 3). Interrater reliability for the presence of the CP was 0.62 , which indicates substantial agreement. For CSF turbulence, $\kappa = 0.82$, which indicates almost perfect agreement.

Discussion

ETV/CPC represents the marriage of 2 old procedures with recent increased application in neonates and infants

TABLE 2. Interrater reliability at first assessment

Assessment	Fleiss' Kappa (95% CI)
Turbulence	
Test 1	0.50 (0.37–0.62)
Test 2	0.46 (0.33–0.58)
CP	
Test 1	0.56 (0.43–0.69)
Test 2	0.58 (0.45–0.71)

Each rater reviewed all subjects twice (3 raters, 79 subjects).

with hydrocephalus in the United States. Repopularized by the work of Dr. Ben Warf in Uganda,⁴ the outcomes of these patients in the United States remain under investigation.

Although there is no substitute for clinical measures, such as head circumference, it might be possible to predict likely failures based on radiographic findings alone. The Pindrik et al. pilot study³ found the potential utility of the presence or absence of CSF turbulence and CP after ETV/CPC as 2 such markers. In this study, we built on that work by demonstrating the reliability among raters in determining the presence or absence of these findings.

As with all new techniques, some education is required. Our results clearly indicate that even modest education regarding what represents CSF turbulence and the presence or absence of the CP greatly improved reliability.

This study was limited by the image set originating from just 1 institution. The sample size, therefore, was limited. The effect of technical imaging differences across institutions and the difference in timing of postoperative imaging could not be assessed in this study. Because Pindrik et al.³ found that these 2 markers have only limited sensitivity and specificity, relying on them in isolation of clinical findings, such as head circumference, is unjustified.

Despite these limitations, the results of this study, in combination with those of the pilot study by Pindrik et al.,³ justify additional investigation of these markers in a prospective study. Should these 2 radiographic markers be validated in a prospective fashion, it might be possible to use them to identify likely failures much earlier than do increased head circumference or ventriculomegaly. Rather than relying on physical measurements to avail themselves, necessarily coinciding with a significant increase in head circumference, it might become possible to identify these children before their head grows significantly. Ideally, this would enable earlier intervention and thereby prevent continued abnormal head growth, thinning of the cortical mantle, and brainstem dysfunction.

TABLE 3. Interrater reliability at second assessment

Assessment	Fleiss' Kappa (95% CI)
Turbulence	0.82 (0.78–0.86)
CP	0.62 (0.58–0.66)

Eight raters reviewed 80 subjects.

Conclusions

The assessment of failure of ETV/CPC in neonates, infants, and toddlers is currently limited to acknowledgment of increased head growth and/or ventriculomegaly. The presence or absence of CSF turbulence and the CP after ETV/CPC might represent radiographic markers that could facilitate earlier detection of procedural failures. Previous research found a rationale for using these markers, and our study built on those results by evaluating the interrater and intrarater reliabilities of such markers and finding better results when the raters were guided through example cases first. Because these markers can be seen with minimally invasive FS-MRI studies, the risks of excessive radiation (CT scans) and/or the need for sedation (formal MRI studies) are obviated. A multicenter validation study is planned to assess the clinical utility of these markers in a larger population of neonates, infants, and toddlers with hydrocephalus.

Acknowledgments

We appreciate the raters' time and efforts, without which the study could not have been completed. The Group 1 (nonauthor) raters were Saurabh Guleria, MD, and Brandon Rocque, MD, and the Group 2 raters, courtesy of the Hydrocephalus Clinical Research Network (chairman John Kestle, MD), were Todd Hankinson, MD, Abhaya Kulkarni, MD, PhD, David Limbrick, MD, PhD, Patrick McDonald, MD, Jonathan Pindrik, MD, Jay Riva-Cambrin, MD, Mandeep Tamber, MD, PhD, John Wellons, MD, PhD, and William Whitehead, MD.

References

1. Fleiss J: Measuring nominal scale agreement among many raters. *Psychol Bull* 76:378–382, 1971

2. Kulkarni AV, Riva-Cambrin J, Browd SR, Drake JM, Holubkov R, Kestle JR, et al: Endoscopic third ventriculostomy and choroid plexus cauterization in infants with hydrocephalus: a retrospective Hydrocephalus Clinical Research Network study. *J Neurosurg Pediatr* 14:224–229, 2014
3. Pindrik J, Rocque BG, Arynchyna AA, Johnston JM, Rozzelle CJ: Radiographic markers of clinical outcomes after endoscopic third ventriculostomy with choroid plexus cauterization: cerebrospinal fluid turbulence and choroid plexus visualization. *J Neurosurg Pediatr* 18:287–295, 2016
4. Warf BC, Rocque BG, Arynchyna AA, Johnston JM, Rozzelle CJ: Comparison of endoscopic third ventriculostomy alone and combined with choroid plexus cauterization in infants younger than 1 year of age: a prospective study in 550 African children. *J Neurosurg* 103 (6 Suppl):475–481, 2005

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Rozzelle. Acquisition of data: Rozzelle, Madura. Analysis and interpretation of data: Rozzelle, Madura. Drafting the article: Madura. Critically revising the article: Rozzelle, Reeder. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Rozzelle. Statistical analysis: Reeder. Study supervision: Rozzelle.

Correspondence

Curtis J. Rozzelle, Children's of Alabama, Department of Neurosurgery, 400 Jarman F. Lowder Bldg., 1600 7th Ave. South, Birmingham, AL 35233. email: curtis.rozzelle@childrensal.org.