Volumetric brain analysis in neurosurgery: Part 2. Brain and CSF volumes discriminate neurocognitive outcomes in hydrocephalus

Jason G. Mandell, MS, MD,1,2 Abhaya V. Kulkarni, MD, PhD,3 Benjamin C. Warf, MD,4,5 and Steven J. Schiff, MD, PhD1,6,7

1Center for Neural Engineering, Department of Engineering Science and Mechanics, and Departments of 2Bioengineering, 3Neurosurgery, and 4Physics, The Pennsylvania State University, University Park, Pennsylvania; 5Division of Neurosurgery, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6Department of Neurosurgery, Boston Children's Hospital, and Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts; 7CURE Children's Hospital of Uganda, Mbale, Uganda

OBJECT The evaluation of hydrocephalus remains focused on ventricular size, yet the goal of treatment is to allow for healthy brain development. It is likely that brain volume is more related to cognitive development than is fluid volume in children with hydrocephalus. This study tests this hypothesis by comparing brain and fluid volumes with neurocognitive outcome in pediatric patients with hydrocephalus.

METHODS Warf and colleagues previously acquired CT scans for pediatric patients in Uganda with myelomeningocele, measured frontal–occipital horn ratio (FOHR), and administered the modified Bayley Scales of Infant Development, third edition (BSID-III) to measure neurocognitive outcome that did not correlate with FOHR. In this present study, brain and fluid volumes were measured in 33 of these patients, 26 of whom required surgical treatment for hydrocephalus. Linear discrimination analysis (LDA) was used to test whether age-normalized brain and fluid volumes can discriminate neurocognitive outcome.

RESULTS Hydrocephalic patients show normal to small brain volumes and substantially larger fluid volumes compared with normal values. FOHR correlates highly with fluid volume (r = 0.84, p < 0.001) and substantially less with brain volume (r = −0.37, p = 0.03), while brain and fluid volumes do not correlate with each other (p = 0.99). Brain and CSF volumes correlated best with fine motor (p = 0.03, p = 0.01), cognitive (p = 0.05, p = 0.09), and expressive communicaton (p = 0.08, p = 0.08) scores. A combination of these 3 scores was used as a multivariate measure of neurocognitive outcome. Brain volume alone, unlike fluid volume, could discriminate high from low cognitive outcome (by t-test and ANOVA). It was shown that a combination of age-normalized brain and fluid volumes can discriminate neurocognitive outcome by 2-way LDA (p < 0.01) and 3-way LDA (p < 0.01). The multivariate LDA demonstrated the contribution of large fluid volume to a decrement in cognition.

CONCLUSIONS Hydrocephalus is treated by normalizing CSF, but normal brain development depends on brain growth. A combination of brain and CSF volumes appears to be significantly more powerful at predicting good versus poor neurocognitive outcomes in patients with hydrocephalus than either volume alone.

Clinical trial registration no.: NCT01936272 (clinicaltrials.gov)
http://thejns.org/doi/abs/10.3171/2014.9.PEDS12427

KEY WORDS hydrocephalus; myelomeningocele; outcome assessment; brain volume; developing country; Bayley Scales of Infant and Toddler Development

Hydrocephalus is accompanied by pathological changes in brain morphology, including thinning of the cortex, increases in water content, and loss of myelin in the periventricular white matter.2,3,7,10,11,14 However, treatment and evaluation remain focused on normalizing CSF volume, while measures of brain volume are not generally available. In a mouse model of hydrocephalus, we have previously shown that valuable information relevant to hydrocephalus can be obtained by measuring brain volume in addition to estimates of fluid...
volume and head circumference. Tracking brain volume is likely a critical measure in the management of hydrocephalus, as the ultimate goal of treatment is to allow for normal brain development, which is not reflected in measurements of fluid volume.

Additionally, quality of life and associated neurocognitive performance are increasingly recognized as important measures of success in the management of hydrocephalus. We argue that in cases of pediatric hydrocephalus, brain volume is more important in determining neurocognitive performance than fluid volume alone. In this second of 3 companion papers, we build on previous work by exploring the relationship of brain and CSF volumes with neurocognitive outcome in Ugandan children with myelomeningocele treated for hydrocephalus.

### Methods

#### Study Cohort

Warf et al. previously acquired CT scans for pediatric patients in Uganda with myelomeningocele, measured frontal–occipital horn ratio (FOHR), and administered the modified Bayley Scales of Infant and Toddler Development, third edition (BSID-III) to measure neurocognitive outcome. Institutional review board oversight for data analysis was provided by CURE, Harvard University, and Penn State University (the last of which provided an exemption determination for de-identified images). Hydrocephalus was treated either by ventriculoperitoneal (VP) shunt placement or by endoscopic third ventriculostomy with choroid plexus catarization (ETV/CPC). The mean age of patients at the time of surgical treatment was 2.2 months (0.25–7.5 months) in the ETV/CPC group and 4.4 months (1.5–11 months) in the VP shunt group. The CT scans and BSID-III tests were administered postoperatively, with a mean age of the patients at the time of assessment of 15.6 months, with no significant difference in mean age among the groups (16.8 months for the untreated, 14.5 for the VP shunt, and 15.6 for the ETV/CPC group; p = 0.8). The BSID-III included measurements of fine motor skill, gross motor skill, expressive communication, receptive communication, cognitive ability, and social-emotional scores. In this present study, we analyze brain parenchymal volumes (including cerebral hemispheres, cerebellum, and brainstem) and intra-brain fluid volumes (including ventricular fluid and fluid contained within the boundary of the brain parenchyma) for 33 of these patients (ages 2 weeks to 72 months): 9 were treated for hydrocephalus with a VP shunt, 17 were treated with ETV/CPC, and 7 did not require treatment for hydrocephalus. The group not treated had stable ventriculomegaly, as reflected by large mean FOHR values, which were similar to values for both the treated groups.

### Normal Controls

Normative growth curves of male and female brain and intra-brain fluid volumes for children and adolescents from 0 to 18 years of age, presented in the companion paper of this series, were calculated using T2-weighted brain MRI data sets from the National Institutes of Health Pediatric MRI Data Repository provided by the Montreal Neurological Institute.

### Brain and Fluid Volume Analysis

A particle filter image segmentation algorithm was used to calculate brain and fluid volumes for all images. The goal of this method was to develop an intelligent segmentation tool that could work with both MR and CT images and lead to a more fully automated tool suitable for the calculation of brain volumes in clinical practice. Briefly, this method starts with a pixel classification step to create probability maps of brain and fluid based upon image intensity for each slice. Using the probability maps and a user-selected seed point, a particle filter is then used to extract the brain from surrounding tissues by tracking the outer edge of the brain. Finally, the initial probability maps of brain and fluid are used to segment the extracted brain into brain and fluid volumes.

Normative curves were used to estimate age-matched control volumes by a least-squares fit to a power law curve of the form \( y = Ax^b \), where \( 0 < b < 1 \). This was performed separately for male and female brain and fluid volumes of children and adolescents ages 0–18 years, and the coefficient of multiple determination, \( R^2 \), was used to test the effectiveness of each fit. \( R^2 \) is a measure of goodness-of-fit and is the ratio of the regression sum of squares to the total sum of squares. It quantifies how much of the variance is explained by the given model. These normative brain and fluid growth curves are essential to normalize the brain and fluid measures by age in the hydrocephalic patients.

### Data Analysis

The correlations between FOHR and brain volume, as well as FOHR and fluid volume, were measured with Pearson’s product-moment correlation coefficient, \( r \), which measures the strength of the linear association between two variables. For simple linear regression, the variable \( r \) is the square root of the coefficient of determination, \( R^2 \), described above. The correlation between raw and normalized brain and fluid volume was also measured. Additionally, the correlation of each of the six subtests of the BSID-III with normalized brain and normalized fluid volume was measured. Significance of all linear correlations were calculated with a t-test with \( n − 1 \) degrees of freedom, testing the hypothesis of no correlation, where \( n \) is the number of subjects. The \( p \) value is then the probability of getting a correlation as large as the observed value by random chance. Linear discrimination analysis (LDA) was used to test how well a multivariate measure of normalized brain and normalized fluid volume could predict neurocognitive outcome. For comparison, a t-test and ANOVA are used to measure significance of classification for univariate fluid and brain volumes alone. The ANOVA was followed by a post hoc Tukey test to measure the significance of the classification of each group.

LDA was originally developed as a method to classify data that had more than 1 measurement (multivariate) and that came from more than 1 group of items. LDA calculates the optimal way of combining the measurement variables together to separate and classify each group. We have refined this method to take into account modern numerical computer algorithms and have previously employed this refined version in the image analysis of an animal model of hydrocephalus.
LDA was used to test the hypothesis that normalized brain and fluid volumes will be able to classify neurocognitive outcome as measured by the BSID-III. By classification we mean the ability to separate outcomes into groups. We chose to incorporate in our classifier the three BSID-III measures that had the most substantial individual correlations with normalized brain and fluid volumes. Each normalized scale has a maximum of 20, with a mean of 10. Therefore the sum of the three cognitive scores has a maximum of 60 and a mean of 30. After adding these three scores together, the cognitive data were grouped by low (<15), medium (15–30), and high (>30) cognitive outcomes. The Wilks test statistic, $W$, was used to test for the significance of the classification. This likelihood ratio tests the hypothesis that each group subject mean, $\mu_k$, is equal:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$$

[Eq. 1]

The Wilks statistic is also used to test random combinations of regrouping the data with a bootstrapping method. The bootstrap method tests if the grouping defined prior to the LDA was likely to have occurred by chance. The $W$ statistics for each permutation of the data are compared with the $W$ statistics for the originally classified data. The bootstrap probability, $P_b$, is the probability that the original classification would occur randomly and is given by

$$P_b = \frac{N_{less} + 1}{N_{perm} + 1}$$

[Eq. 2]

where $N_{less}$ is the number of groupings with a statistic less than or equal to the original statistic, and $N_{perm}$ is the total number of permutations. This calculation includes the original grouping as one of the permutations in addition to those performed in the bootstrapping, and therefore 1 is added to $N_{less}$ and $N_{perm}$ in Eq. 2. To enable others to replicate our work, we have archived data samples and code used for the discrimination analysis, which are available for download as a Code Archive here.

Results

Normative growth curves of brain and fluid volumes are shown in Fig. 1, along with brain and fluid volumes measured from the study cohort. Patients in the study cohort show normal to small brain volumes and substantially larger fluid volumes than normal controls.

FOHR correlates substantially ($r = 0.84$) and significantly with CSF volume ($p < 0.00000001$), but does not correlate as substantially ($r = -0.37$) or significantly with brain volume ($p = 0.03$) (Fig. 2A and B). The lack of correlation between raw brain and fluid volume is shown directly in Fig. 2C ($r = 0.002$, $p = 0.99$). In contrast, although there is increased linear correlation shown with age-normalized brain volume versus normalized fluid volume ($r = -0.39$, $p = 0.02$), the data clearly do not fit a linear curve well (Fig. 2D). Because Warf et al. did not find a correlation between treatment type and cognitive outcome, all treatment types are analyzed together here, although they are distinguished by symbols in the figures for clarity.

The Pearson product-moment correlation coefficients ($r$) and $p$ values for all correlations for all BSID-III tests are summarized in Table 1. The plots of correlations of the different BSID-III subtests that correlated most strongly with normalized brain and fluid volumes are shown in Fig. 3. Fine motor score (Fig. 3A and B) shows a significant linear correlation with both normalized brain volume ($r = 0.40$, $p = 0.03$) and normalized fluid volume ($r = -0.45$, $p = 0.01$). Gross motor score does not show a correlation with either brain or fluid volume. However, all of these patients scored poorly on the gross motor subtest because of their paraplegia and thus their scores on this subtest would not be expected to correlate as well with brain and CSF volumes. Expressive language (Fig. 3C and D) showed moderate correlations with normalized brain volume ($r = 0.33$, $p = 0.08$) and normalized CSF volume ($r = -0.32$, $p = 0.08$). Receptive language showed poor correlation with brain volume, but more substantial correlation with normalized fluid volume ($r = -0.25$, $p = 0.18$). The cognitive subtest (Fig. 3E and F) showed stronger correlation with normalized brain ($r = 0.36$, $p = 0.05$) but less with normalized fluid volumes ($r = -0.31$, $p = 0.18$). The social-emotional score showed moderate correlation with normalized brain volume ($r = 0.25$, $p = 0.20$), but virtually no correlation with normalized fluid volume. We call attention to these correlations that failed to reach significance as univariate measures, because combining the most sub-
stantial features might well demonstrate significance as a multivariate measure.

We selected the 3 Bayley tests that most strongly correlated with brain and fluid volume \((r > |0.3|)\). LDA showed that a combination of normalized brain and fluid volumes discriminates neurocognitive outcome as determined by the cumulative fine motor, expressive language, and cognitive scores. A 2-way analysis significantly discriminates patients with scores above the mean of 30 from those with scores at or below the mean \((W = 0.72, p < 0.01)\) (Fig. 4). A projection of the data onto each axis is also shown. A t-test shows that the data are not significantly separated by fluid volume alone \((p = 0.09)\). The data are significantly classified by brain volume alone \((p < 0.01)\), but the contribution of large fluid volumes to affecting cognition (the slanted discrimination line) is lost using the univariate t-test. The classification of data by neurocognitive score was checked by randomly shuffling the group assignments with a bootstrap method. The bootstrap statistic is repeated for 1000 random permutations of the data. The results are shown as a histogram of the frequency of the Wilks statistic for each permutation, and \(W\) values of the original data are marked with an asterisk. In the 2-way analysis there were only 5 other ways to regroup the data to obtain a better \(W\) statistic, leading to a 0.6% probability that the prior classification was obtained by chance (Fig. 4). We also performed these analyses with a 3-way analysis, comparing patients with a score above 30 (the mean), between 15 and 30, and below 15. This discrimination is also significant \((W = 0.64, p < 0.01)\), and the bootstrap method showed a 1.7% probability that the prior classification was obtained by chance (Fig. 5). A projection of the data is also shown in Fig. 5. As measured by ANOVA, fluid volume alone is not able to separate the data \((p = 0.10)\); indeed a post hoc Tukey test shows that no groups have means that are significantly different from other groups. Separation using brain volume alone does show significance \((p = 0.01)\) by ANOVA; here, the post hoc Tukey test shows that the group with scores below 15 and the group with scores above 30 have means that are significantly different \((p = 0.01)\). However, the middle group with scores between 15 and 30 is not significantly different from either the group with scores less than 15 \((p = 0.43)\) or the group with scores

![FIG. 2. A and B: FOHR is plotted against brain and fluid volumes. The Pearson correlation coefficient for FOHR and brain volume is moderate \((r = -0.37)\) but shows significance \((p = 0.03)\), while a correlation between FOHR and fluid volume is highly substantial \((r = 0.84)\) and significant \((p < 0.00000001)\). C and D: Raw fluid volume does not have a significant correlation with raw brain volume \((r = 0.002, p = 0.99)\). Normalized fluid volume does have a significant correlation with normalized brain volume \((p = 0.02)\), although the correlation is moderate \((r = -0.39)\). ETV = endoscopic third ventriculostomy; MMO = myelomeningocele only (no VPS or ETV); VPS = ventriculoperitoneal shunt.]

### TABLE 1. Summary of Pearson product moment coefficients \((r)\) and \(p\) values for BSID-III tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Normalized Brain Volume</th>
<th>Normalized CSF Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(r)</td>
<td>(p) Value</td>
</tr>
<tr>
<td>Fine motor</td>
<td>0.40</td>
<td>0.03</td>
</tr>
<tr>
<td>Gross motor</td>
<td>0.14</td>
<td>0.45</td>
</tr>
<tr>
<td>Expressive comm.</td>
<td>0.33</td>
<td>0.08</td>
</tr>
<tr>
<td>Receptive comm.</td>
<td>0.06</td>
<td>0.77</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>Social-emotional</td>
<td>0.25</td>
<td>0.20</td>
</tr>
</tbody>
</table>
greater than 30 \( (p = 0.08) \), demonstrating the advantage of the multivariate measure. The LDA shows that the combined measurements of both normalized brain volume and normalized fluid volume are potentially important predictors of neurocognitive outcome in pediatric hydrocephalus.

**Discussion**

In this study, we measured brain and fluid volumes for 33 patients in Uganda with treated hydrocephalus or stable ventriculomegaly. We are able to show that in these children, brain and fluid volumes can develop independently of each other, and that a multivariate measure of both brain and fluid volumes can discriminate neurocognitive outcome.

Fluid volumes ranged in these children from 3 to 100 times normal fluid volume. Brain volumes of the study patients ranged from small to normal compared with normative data (Fig. 1). It is important to note that the normative curve is calculated for North American child development, which may be different from normative growth of the rural Ugandan child. There are presently no more specific normative data on brain growth in African children nor for children with myelomeningocele without hydrocephalus.

Previously, Warf et al. found no significant association between FOHR and any BSID scale score, regardless of treatment in this same cohort. We show here that when the values for fluid volume are normalized, the results indicate that patients who have a severe accumulation (40–
FIG. 4. A 2-way LDA is used to separate neurocognitive scores at or below the mean of 30 (solid circles) and above the mean (open circles) by normalized fluid volume and normalized brain volume. The squares are the mean values for each group. The dashed line is the discrimination line. The discrimination is significant ($W = 0.72, p < 0.01$). Also shown is a projection of the data onto each axis with the discrimination lines for brain and fluid volumes individually. A t-test shows that the data are not significantly separated by fluid volume alone ($p = 0.09$). Brain volume, however, does significantly classify the data ($p < 0.01$). The bootstrap method showing a histogram of the $W$ statistic for random regroupings of the data are significant ($p < 0.006$). The $W$ values for the original data are marked with an asterisk. prob = probability.

FIG. 5. A 3-way LDA is used to separate neurocognitive scores below 15 (solid black circles), between 15 and 30 (solid gray circles), and above 30 (open circles), by normalized fluid volume and normalized brain volume. The squares indicate the mean values for each group. Dashed lines are the discrimination lines. The discrimination is significant ($W = 0.64, p < 0.01$). A projection of the data onto each axis is also shown here. The bootstrap method showing a histogram of the $W$ statistic for random regroupings of the data is significant ($p < 0.017$). The $W$ values for the original data are marked with an asterisk. An ANOVA shows that the data are not significantly classified by fluid volume alone ($p = 0.10$). Classification by brain volume shows significance ($p = 0.01$) by ANOVA, with a post hoc Tukey test showing significant differences only between the group with scores less than 15 and the group with scores above 30. However, the middle group with scores between 15 and 30 is not significantly different from either the group with scores less than 15 ($p = 0.43$) or the group with scores greater than 30 ($p = 0.08$), demonstrating the advantage of the multivariate measure.
100 × normal) do cognitively worse than patients who do not (Figs. 4 and 5). Below this level of CSF accumulation, it seems that, as Warf et al. state, normal ventricle size is not necessarily crucial for normal development. Indeed, our results suggest that in pediatric hydrocephalus a normal brain volume is more important in neurocognitive development than fluid volume.

The current evaluation of hydrocephalus remains focused on estimates of fluid volume even though the ultimate goal of treatment is to allow suitable brain development for normal cognitive function. Although it seems self-evident that measurement of total brain volume is vitally important in the evaluation of hydrocephalus, volume remains difficult to estimate by visual inspection of brain images in the hydrocephalic infant. Since our currently employed metrics—head circumference, cortical mantle thickness, and FOHR—all fail to estimate brain volume well, our findings argue for the incorporation of direct volumetric measurements in the evaluation of hydrocephalus outcomes. With further development of more automated volumetric methods, building upon the technology described in the companion study, the incorporation of brain and fluid volumes will become more feasible in the clinical setting.

Limitations
As previously discussed, treatment of the children in this study was not randomized. In most of the patients receiving a VP shunt (7 of 9), the shunt was placed following a failed or abandoned ETV/CPC procedure. As children with more complications tend to have worse cognitive outcomes, it is possible that cognitive performance was affected by complication rate. Additionally, although imaging and test administration were performed as close in time as possible, the average delay between imaging and test administration was 5.25 months, with an equal number of cases in which imaging occurred before testing as after. Very young children can show rapid brain development and it is possible that the correlations between brain volume and testing will have been affected by this gap in time.

Lastly, our sample size was small and limited to the case records still accessible from the 2009 study by Warf et al. Validating our findings prospectively, with larger numbers of patients and including patients without spina bifida, is important to determine in what form our results would generalize. In terms of cognitive scoring, the most robust selection of Bayley tests for patients with other forms of hydrocephalus will almost certainly be different from the selection of tests for our study population, for whom tests such as gross motor were unsuitable for inclusion in our outcome metrics.

These limitations of treatment bias, imaging and test delay, and sample size are addressable. At the time of this writing, a Phase III randomized controlled surgical trial in Africa of VP shunt treatment versus ETV/CPC is being carried out with National Institutes of Health support (ClinicalTrials.gov registration number NCT01936272), and the prospective data applying these volumetric methods to hydrocephalic patients based upon intention to treat and neurocognitive outcome, addressing the above limitations, should be available in several years upon trial completion.

Conclusions
We treat hydrocephalic brains by treating the accumulation of CSF; yet it is the growth of the brain that ultimately determines cognitive development and quality of life. CSF volume alone may not even be a factor until a critical maximum is reached, at which point it likely interferes with parenchymal brain function. A combination of brain volume and CSF volume appears to be significantly more powerful at discriminating good versus poor neurocognitive outcomes in hydrocephalus than either volume alone—or indirect measures of fluid volume such as FOHR. Although this study focused on measurements at a single point in time, brain development and hydrocephalus treatment are dynamic processes. Future management may benefit from predictive treatment by finding the optimal trajectory of brain and fluid volume for an individual patient, based upon normal growth curves, with the goal of optimizing brain and cognitive development.

References
14. Raimondi AJ, Bailey OT, McLone DG, Lawson RF, Echeverry A: The pathophysiology and morphology of murine hy-
doroccephalus in Hy-3 and Ch mutants. Surg Neurol 1:50–55, 1973

Supplemental Information
Previous Presentation
Portions of this work were presented as an oral presentation at: 39th Annual Meeting of the AANS/CNS Section on Pediatric Neurological Surgery, December 2, 2010, in Cleveland, OH.

Companion Papers

Code Archive

Correspondence
Steven J. Schiff, W311 Millennium Science Complex, Pollock Rd., University Park, PA 16802-2131. email: sschiff@psu.edu.