Prediction of Surgical Outcomes in Normal Pressure Hydrocephalus by MR Elastography

Pragalv Karki, Matthew C. Murphy, Petrice M. Cogswell, Matthew L. Senjem, Jonathan Graff-Radford, Benjamin D. Elder, Avital Perry, Christopher S. Graffeo, Fredric B Meyer, Clifford R. Jack, Jr., Richard L. Ehman and John Huston III

AJNR Am J Neuroradiol 2024, 45 (3) 328-334
doi: https://doi.org/10.3174/ajnr.A8108
http://www.ajnr.org/content/45/3/328
Prediction of Surgical Outcomes in Normal Pressure Hydrocephalus by MR Elastography

Pragalv Karki, Matthew C. Murphy, Petrice M. Cogswell, Matthew L. Senjem, Jonathan Graff-Radford, Benjamin D. Elder, Avital Perry, Christopher S. Graffeo, Fredric B. Meyer, Clifford R. Jack, Jr., Richard L. Ehman, and John Huston III

ABSTRACT

BACKGROUND AND PURPOSE: Normal pressure hydrocephalus is a treatable cause of dementia associated with distinct mechanical property signatures in the brain as measured by MR elastography. In this study, we tested the hypothesis that specific anatomic features of normal pressure hydrocephalus are associated with unique mechanical property alterations. Then, we tested the hypothesis that summary measures of these mechanical signatures can be used to predict clinical outcomes.

MATERIALS AND METHODS: MR elastography and structural imaging were performed in 128 patients with suspected normal pressure hydrocephalus and 44 control participants. Patients were categorized into 4 subgroups based on their anatomic features. Surgery outcome was acquired for 68 patients. Voxelwise modeling was performed to detect regions with significantly different mechanical properties between each group. Mechanical signatures were summarized using pattern analysis and were used as features to train classification models and predict shunt outcomes for 2 sets of feature spaces: a limited 2D feature space that included the most common features found in normal pressure hydrocephalus and an expanded 20-dimensional (20D) feature space that included features from all 4 morphologic subgroups.

RESULTS: Both the 2D and 20D classifiers performed significantly better than chance for predicting clinical outcomes with estimated areas under the receiver operating characteristic curve of 0.66 and 0.77, respectively (P < .05, permutation test). The 20D classifier significantly improved the diagnostic OR and positive predictive value compared with the 2D classifier (P < .05, permutation test).

CONCLUSIONS: MR elastography provides further insight into mechanical alterations in the normal pressure hydrocephalus brain and is a promising, noninvasive method for predicting surgical outcomes in patients with normal pressure hydrocephalus.

ABBREVIATIONS: AUROC = area under the receiver operating curve; DESH = disproportionately enlarged subarachnoid hydrocephalus; DOR = diagnostic OR; FDR = false discovery rate; HCTS = high-convexity tight sulci; MRE = MR elastography; NPH = normal pressure hydrocephalus; NPV = negative predictive value; PPV = positive predictive value; SVM = support vector machine

Normal pressure hydrocephalus (NPH) is a CSF dynamics disorder1 with imaging features of enlarged ventricles and clinical symptoms of cognitive decline, gait disturbance, and urinary incontinence.2 NPH has an estimated prevalence of 2.1% for ages 65 and 70 and 8.9%3 for ages 80 and older. Overlapping symptoms with Alzheimer disease4 or Parkinson disease5 could lead to misdiagnosis of NPH. Contrary to these proteinopathies, NPH may be treated with ventriculoperitoneal shunt surgery6 with sustained improvement in about 80% of cases.7,8 Surgery can even reverse the symptoms of progressive dementia.9-11 However, due to the invasive nature of surgery, improving the predictability of outcomes is imperative.

A spinal tap test is commonly used to predict shunt outcome, with a high positive predictive value (PPV) of 92% but a low negative predictive value (NPV) of 37%.12 An improved area under the receiver operating curve (AUROC) and diagnostic OR (DOR) can be achieved by extended lumbar drainage and intracranial pressure measurements.13 These methods, however, have higher rates of infection and complications14-17 and are less widely available than the tap test. On the basis of a meta-analysis of several radiologic predictors, only callosal angle and periventricular white matter changes could significantly differentiate between shunt responders and nonresponders, though with low DOR values of 1.88 and 1.01, respectively.18 A machine learning
method developed on intracranial pressure and electrocardiogram features during the lumbar infusion test demonstrated excellent accuracy of 82% and an AUROC of 0.89. However, a noninvasive, safe, and practical alternative is still needed. A machine learning approach based on pattern analysis of MR elastography (MRE) data is a promising noninvasive, radiologic method for predicting the outcome of shunt surgeries.

MRE allows noninvasive evaluation of tissue mechanical properties using acoustic waves. Previous studies have demonstrated that the mechanical properties of the brain are altered by NPH. These alterations occur in specific patterns, the presence or absence of which can distinguish patients with NPH from healthy controls and those with Alzheimer disease. Past MRE studies evaluated mechanical changes in the brain due to NPH with cases considered as a single group. However, patients with NPH have different morphologic phenotypes that can be assessed with MR imaging.

In this study, we first tested the hypothesis that the different morphologic phenotypes of NPH are associated with unique mechanical signatures. Then, we tested the hypothesis that those mechanical features could improve prediction of the clinical response to shunt surgery compared with the use of mechanical features derived from consideration of patients with NPH as a single group.

MATERIALS AND METHODS

Patient Recruitment
We retrospectively identified 137 patients who underwent 3T MR imaging for suspected NPH from April 2014 to December 2022. Nine cases were excluded because of comorbidities, including contusions and meningiomas, or technical failure during tissue segmentation caused by exceptionally large ventricles (particularly in congenital cases). From the remaining 128 suspected cases, 68 participants who had normal opening pressure (<25 cm CSF) during lumbar puncture and gait improvement with the spinal tap test underwent treatment with ventriculoperitoneal shunt placement. Of these patients, 54 were shunt responders, and 14 were nonresponders. Shunt responders were defined as patients who had improvement in gait, cognition, or urinary incontinence at >1 month after shunt placement per neurology or neurosurgery clinical notes. The clinician’s assessment of improvement was based on patient reports, gait examination/analysis, and/or mental status examinations.

Data from a group of cognitively healthy controls were included from a previously published study. These participants were recruited from the Mayo Clinic Study of Aging and had previously undergone Pittsburgh Compound B-PET imaging to determine that they were free of a significant amyloid load.

Image Acquisition
Participants were scanned once on either a GE Signa HDx or GE Discovery MR750W or a Siemens Magnetom Prisma scanner. The acquisitions were comparable among the scanners and included MRE and structural imaging. MRE was performed using a flow-compensated, spin-echo, echo-planar imaging pulse sequence. Shear waves were introduced via a pneumatic actuator at a frequency of 60 Hz. Structural imaging included a whole-brain T1-weighted MPRAGE or 3D inversion-recovery spoiled gradient-recalled acquisition.

The acquisition parameters for the patients with NPH on the 3T GE Healthcare scanner were the following: T1-weighted MPRAGE: TR/TE/TI = 6.3/2.6/900 ms, flip angle = 8°, FOV = 260 × 260 mm, matrix = 256 × 256, section thickness = 1.2 mm; and MRE: TR/TE = 3601.2/57.3 ms, FOV = 240 × 240 mm, matrix = 72 × 72, section thickness = 3 mm. The acquisition parameters for the control participants were the following: T1-weighted 3D inversion-recovery spoiled gradient-recalled: TR/TE = 6.3/2.8 ms, flip angle = 11°, FOV = 270 × 270 mm, matrix = 256 × 256, section thickness = 1.2 mm; and MRE: TR/TE = 4001.3/59.3 ms, FOV = 240 × 240 mm, matrix = 80 × 80, section thickness = 3 mm.

The acquisition parameters on the 3T Siemens scanner were the following: T1-weighted MPRAGE: TR/TE/TI = 2300/3.1/945 ms, flip angle = 9°, FOV = 240 × 256 mm, matrix = 320 × 300, section thickness = 0.8 mm; and MRE: TR/TE = 4800/54 ms, FOV = 240 × 240 mm, matrix = 80 × 80, section thickness = 3 mm.

Evaluation of Morphologic Features
A neuroradiologist classified patients with suspected NPH into 4 subgroups based on their morphologic features assessed on structural imaging. These 4 groups were the following: 1) high-convexity tight sulci (HCTS); 2) congenital hydrocephalus (Congenital); 3) ventriculomegaly alone (Ventric), and 4) neither ventriculomegaly nor HCTS (Neither). HCTS was defined as focal narrowing or effacement of the sulci at the midline/vertex. Most of the cases of HCTS also had enlarged Sylvian fissures and ventriculomegaly, imaging features of disproportionately enlarged subarachnoid hydrocephalus (DESH). The patients with HCTS alone and DESH were considered as 1 group because no significant differences in mechanical properties were detected. Ventriculomegaly was defined by Evans Index > 0.3, which measures the ratio of frontal horn width to internal skull width and the absence of HCTS. Congenital was defined as ventriculomegaly, diffusely narrowed cerebral sulci, and features of impaired aqueductal flow, including aqueductal web, aqueductal stenosis, or triventriculomegaly with a normal fourth ventricle. The Neither group had neither ventriculomegaly nor HCTS.

Stiffness and Damping Ratio Map Calculation
Stiffness and damping ratio maps were computed using neural network inversion as previously described. After mechanical property estimation, maps were warped into template space for analyses. These methods are further described in the Online Supplemental Data.

Mapping of Phenotypic Effects on Mechanical Properties
To identify significant differences in the mean stiffness and damping ratios between the groups, we fit a linear model at each voxel with predictors including age, sex, scanner system, and a set

of categoric variables for group assignment by one-hot encoding. Difference maps and corresponding t-statistics were calculated for stiffness and the damping ratio between HCTS and the other groups. A false discovery rate (FDR) corrected $Q_{0.05}$ as computed by the Storey method was considered significant.

**Pattern Analysis**

In this study, we used a previously described pattern analysis method. This method summarizes each person’s MRE result by measuring its spatial correlation with the expected mechanical pattern, which is obtained by contrasting 2 groups of interest, while controlling for effects of no interest (ie, age, sex, scanner). Flow charts explaining the procedure are shown in Fig 1. By considering only a single contrast of interest (HCTS versus controls), a 2D feature space is computed (1 pattern score for each of stiffness and damping ratio). By considering all possible contrasts that arise from subtyping the NPH participants, we computed a 20D feature space.

**Machine Learning Classification Model for Shunt Prediction**

Support vector machine (SVM) classifiers were trained to predict a successful surgical outcome by using leave-one-out cross-validation to estimate out-of-sample accuracy. Separate SVMs were trained using either the 2D or 20D feature spaces. We compared the 2 models using the following performance metrics: the AUROC, accuracy, DOR, PPV, and NPV. We first conducted a permutation test to assess whether the AUROC of each model was significantly greater than a random classifier. We then conducted a permutation test to assess whether the performance metrics of the 20D feature space offered improvement compared with the 2D space.

**RESULTS**

There were 172 participants in this study, 44 controls, and 128 with suspected NPH. Of the 128 patients with suspected NPH, 91 had morphologic features of HCTS, 12 had congenital hydrocephalus, 20 had ventriculomegaly only, and 5 had neither ventriculomegaly nor HCTS.

A group-wise boxplot of the mean shear stiffness is shown in Fig 2A. There was a statistically significant difference in the mean stiffness of the whole brain between the HCTS and control groups ($P < .05$, t test). Axial slices of the averaged stiffness maps for each group are shown in Fig 2B. The HCTS group was characterized by stiffening at the midline vertex and softening around the periventricular region. Ventric and congenital groups showed patterns similar to those of HCTS, but the stiffening at the vertex was shifted toward the frontal region of the brain. In the Neither...
group, softening was evident around the periventricular region, but without stiffening at the vertex.

Figure 3 shows a group-wise boxplot of the mean damping ratio in panel A and damping ratio maps in panel B. The mean damping ratio showed a stepwise decrease as groups exhibited an increasing number of anatomic features. Significant differences in the \( t \) tests between the groups are labeled in the boxplot. In Fig 3, damping ratio patterns demonstrated an overall decline in values in NPH phenotypes compared with controls, with greater differences toward the cranial direction.

The difference maps for the stiffness between HCTS and other groups are shown in Fig 4. A gray-scale map of voxelwise differences is overlaid with a \( t \)-statistic map thresholded for statistical significance (FDR corrected with \( Q < 0.05 \)). There were 93,254
voxels that were significantly different between HCTS and controls. The HCTS group had a cluster of voxels with higher stiffness at the midline vertex compared with the Ventric and Neither groups. HCTS had 6931 voxels with a significant difference in comparison with Neither and 11,364 voxels in comparison with Ventric. The HCTS and Congenital groups differed significantly in fewer voxels (735), without any discernible pattern.

**Figure 5** illustrates the difference maps for damping ratios of HCTS versus other groups. According to the thresholded t-statistic maps, damping ratio values were lower overall for the HCTS group. There were significant differences in 144,233 voxels between HCTS and control, 40,019 voxels between HCTS and Neither, 43,984 voxels between HCTS and Ventric, and none between HCTS and Congenital. A globally lower damping ratio of the HCTS group is consistent with the findings in the boxplot of Fig 3A.

In Fig 6, scatterplots of damping ratio and stiffness pattern scores are shown for 4 different contrasts. The HCTS versus control contrast (Fig 6A), HCTS and control cases form distinct clusters, demonstrating the separability of these groups based on pattern scores. The remaining NPH subgroup cases were distributed among these clusters with intermediate pattern scores. In the HCTS-versus-Ventric contrast (Fig 6B), Ventric and HCTS cases were further separated than mentioned above because the reference map in this contrast excluded the feature of ventriculomegaly. Figure 6C shows the pattern scores for HCTS-versus-Neither contrast, and Fig 6D shows the pattern scores for HCTS-versus-Congenital contrast. The reference features extracted in these latter contrasts are labeled in Fig 1B. The scatterplots for the remaining 6 contrasts are included in the Online Supplemental Data.

**Figure 7** shows the receiver operating characteristic curves with 5 performance metrics for the SVM classification models trained with the limited (2D) or expanded (20D) feature space. The AUROC was 0.66 for the 2D feature space (greater than a random classifier with $P < .05$, permutation test) in comparison with 0.77 using the 20D feature space ($P < .01$).

The accuracy of the 20D feature space was 72% compared with 66% for the 2D feature space. The DOR was 6.50 compared with 1.06. PPV was 0.91 compared with 0.80, and NPV was 0.40 compared with 0.21. Though all metrics performed better using the 20D feature space, the difference between the AUROC and NPV was not statistically significant on the basis of a permutation test. The differences in the DOR and the PPV were statistically significant with $P < .05$, and the difference in accuracy approached the level of significance with $P = .06$.

**DISCUSSION**

This study reproduced the previous finding that HCTS/DESH is associated with characteristic patterns of stiffness and damping ratio alterations. By subclassifying the patients with NPH according to the presence or absence of specific anatomic features, we reported significant differences in brain mechanical properties associated with each phenotype. Furthermore, we showed that the pattern scores computed to summarize these findings at the individual level perform significantly better than chance at predicting the surgical outcomes. Finally, the 20D feature space improved predictions compared with the 2D feature space, indicating that a more detailed summary of the MRE result contains clinically useful information and merits further investigation.
Predicting the outcome of shunt surgery is a challenging task. Spinal tap tests are commonly used for prediction with high PPVs but low NPVs. Due to the invasive nature of surgery and the potential for complications, the ability to predict negative outcomes is critical. DESH, which is HCTS along with enlarged Sylvian fissures, is an imaging feature used in the diagnosis of NPH under the widely accepted Japanese criteria. DESH and HCTS have been found to predict clinical improvement after shunt placement in several studies. However, studies have indicated that relying only on tight high convexity to predict shunt outcome would exclude patients with other NPH phenotypes who could also benefit from surgery, given that HCTS and DESH have even lower performance metrics than the spinal tap test. In this study, we present a noninvasive machine learning approach based on MRE for predicting surgical outcomes in NPH that considers the spectrum of NPH imaging phenotypes and not just DESH.

The results of this study are limited primarily by the number of cases in each NPH subgroup and the number of cases undergoing shunt placement with clinical follow-up. The sample size impacts both the pattern score estimation and the classifier training. Nonetheless, this is the largest MRE study on NPH, to our knowledge. Another limitation of this study is that some of the features in the expanded feature set are likely counterproductive to the classification model. However, we did not use any feature selection because the sample size was not sufficient to add this layer of model tuning. Thus, the presented approach should provide a conservative estimate of model performance and further improvement is expected with additional data.

CONCLUSIONS

In addition to the clinical importance of predicting shunt efficacy, it is also vital to establish biomarker-derived features for various morphologic phenotypes of NPH to better understand its pathophysiology. The morphologic phenotypes of NPH exhibit distinct mechanical signatures using MRE. Pattern analysis based on MRE presents a promising method for improving diagnosis and prediction of shunt outcomes. In addition, this methodology could be relevant in distinguishing NPH from other neurologic disorders that may have overlapping imaging and/or clinical presentations mimicking NPH, such as Parkinson disease, Alzheimer disease, or progressive supranuclear palsy. The study provides motivation for further research on the underlying mechanical biomarkers of different phenotypes of NPH, in addition to collecting more clinical follow-up after shunt surgery to improve prediction abilities.

Disclosure forms provided by the authors are available with the full text and PDF of this article at www.ajnr.org.

REFERENCES

4. Borzage M, Saunders A, Hughes J, et al. The first examination of diagnostic performance of automated measurement of the callosal angle in 1856 elderly patients and volunteers indicates that 12.4%


