Predicting Tissue Outcome in Acute Human Cerebral Ischemia Using Combined Diffusion- and Perfusion-Weighted MR Imaging

Ona Wu, MS; Walter J. Koroshetz, MD; Leif Østergaard, MD, PhD; Ferdinando S. Buonanno, MD; William A. Copen, MD; R. Gilberto Gonzalez, MD, PhD; Guy Rordorf, MD; Bruce R. Rosen, MD, PhD; Lee H. Schwamm, MD; Robert M. Weisskoff, PhD; A. Gregory Sorensen, MD

Background and Purpose—Tissue signatures from acute MR imaging of the brain may be able to categorize physiological status and thereby assist clinical decision making. We designed and analyzed statistical algorithms to evaluate the risk of infarction for each voxel of tissue using acute human functional MRI.

Methods—Diffusion-weighted MR images (DWI) and perfusion-weighted MR images (PWI) from acute stroke patients scanned within 12 hours of symptom onset were retrospectively studied and used to develop thresholding and generalized linear model (GLM) algorithms predicting tissue outcome as determined by follow-up MRI. The performances of the algorithms were evaluated for each patient by using receiver operating characteristic curves.

Results—At their optimal operating points, thresholding algorithms combining DWI and PWI provided 66% sensitivity and 83% specificity, and GLM algorithms combining DWI and PWI predicted with 66% sensitivity and 84% specificity voxels that proceeded to infarct. Thresholding algorithms that combined DWI and PWI provided significant improvement to algorithms that utilized DWI alone (P=0.02) but no significant improvement over algorithms utilizing PWI alone (P=0.21). GLM algorithms that combined DWI and PWI showed significant improvement over algorithms that used only DWI (P=0.02) or PWI (P=0.04). The performances of thresholding and GLM algorithms were comparable (P>0.2).

Conclusions—Algorithms that combine acute DWI and PWI can assess the risk of infarction with higher specificity and sensitivity than algorithms that use DWI or PWI individually. Methods for quantitatively assessing the risk of infarction on a voxel-by-voxel basis show promise as techniques for investigating the natural spatial evolution of ischemic damage in humans. (Stroke. 2001;32:933-942.)

Key Words: cerebral ischemia ■ magnetic resonance imaging, diffusion-weighted ■ magnetic resonance imaging, perfusion-weighted ■ stroke, acute

Efforts to limit infarction in acute stroke patients might gain significantly from an accurate means of identifying hypoperfused yet viable brain tissue. Diffusion-weighted MRI (DWI) and perfusion-weighted MRI (PWI) have been shown to be highly sensitive and specific in diagnosing acute human cerebral ischemia.1-8 These imaging techniques appear to provide superior early identification of regions likely to proceed to infarction compared with conventional MR or CT imaging.1-4,9 However, the prediction of tissue and clinical outcome from specific imaging characteristics remains challenging. Although studies have found correlations between acute DWI and PWI with patients’ clinical and follow-up imaging outcomes,10-15 the ability to predict clinical or tissue outcome in individual patients using a single modality still appears limited, perhaps due to the effects of stroke location and comorbid factors.

Attempts have been made to combine DWI and PWI by comparing lesion volumes identified by the 2 techniques. “Diffusion-perfusion mismatches,” in which the lesion volumes identified by one modality are larger than those by the other, have been reported by several groups.11-14,16 Many groups have reported larger lesion enlargement of the acute DWI lesion volume in cases where the acute PWI volume is larger13-21 than the DWI lesion. In cases where the acute DWI lesion was larger than the PWI lesion, total lesion growth was reduced.13-15,22 Based on these observations, many have hypothesized that these DWI-PWI mismatches may allow identification of salvageable tissue in individual patients.
These reported “mismatches” are of volumes of tissue rather than a voxel-by-voxel comparison. Heterogeneity in both ADC and flow values within acute ischemic tissue in humans have been well documented but have not been captured in these initial volumetric approaches. Therefore, volumetric approaches comparing gross differences in DWI and PWI lesion volumes may oversimplify the complex task of assessment of tissue viability in different regions within ischemic tissue. A voxel-by-voxel analysis, such as that developed by Welch and colleagues, may provide a more sensitive approach for identifying salvageable tissue. Their studies demonstrated that a combination of T2 and ADC information provided better prediction of cellular necrosis than algorithms that used them separately and that a voxel-by-voxel analysis may better demonstrate the underlying heterogeneity in the lesion.

A natural extension of these signature tissue algorithms is the inclusion of PWI. However, assessing the signatures’ significance becomes complicated, because each additional parameter leads to an exponential increase in the number of “signatures.” Furthermore, assuming only discrete states ignores the variances intrinsic to the data. A more complete algorithm may be one in which inputs are treated as random variables and the output is the probability of infarction for each given tissue voxel. In this study, we investigated a strategy that utilized statistical generalized linear model (GLM) algorithms, in which the output is not a map of stages of infarction but risk of future infarction.

Our hypotheses were therefore 2-fold. First, we sought to determine whether algorithms that combine diffusion and perfusion information provide more sensitive and specific predictors of tissue outcome than algorithms using only subsets of this information. Second, we examined whether a probabilistic algorithm provides an improved indicator of which tissue is at risk of infarction over thresholding-based approaches. We tested both hypotheses by retrospectively applying the different techniques to diffusion and perfusion indices acquired from acute stroke patients and comparing the algorithms’ voxel-by-voxel performances in predicting which tissue will proceed to infarction.

Subjects and Methods

Patient Selection

Diffusion- and perfusion-weighted images of patients with hyperacute cerebral ischemia acquired within 12 hours of symptom onset between the years August 1994 and August 1997 were examined retrospectively (n=94). To avoid potential confounds due to different types of ischemic damage and to obtain a relatively homogenous population of stroke patients, this study was limited to patients with clinical signs suggestive of a major cerebral artery occlusion. This study therefore included only Trial of ORG 10172 in Acute Stroke (TOAST) classification subtypes of large-artery atherosclerosis, cardioembolism, stroke of other determined etiology, and stroke of undetermined etiology. Subtypes of small-vessel occlusion and noncerebral artery occlusions were excluded (n=32). Other exclusion criteria included treatment with thrombolitics or neuroprotective agents (n=10) and nonavailability of acute diffusion or perfusion studies due to motion- or equipment-induced artifact (n=4). Patients were excluded if a follow-up axial T2-weighted fast spin-echo (FSE) imaging study 5 days or later was not available to confirm extent of lesion volume (n=34). A total of 14 patients satisfied these inclusion criteria. Diffusion and perfusion findings in 9 of 14 of these patients have been reported previously. Table 1 summarizes their demographics and stroke subtype classifications.

Image Acquisition

Imaging was performed on a 1.5-T General Electric Signa MR instrument with 5.4.2 software (General Electric Medical Systems) and retrofitted with echo-planar imaging (EPI) capabilities via an Advanced NMR Systems hardware upgrade that included the “catch and hold” modification. Table 1 summarizes the MR acquisition parameters for the patients. Multislice axial DWI were acquired by either sampling 3 orthogonal directions at b values of 1010 s/mm² (n=3) or sampling the full diffusion tensor at b values of 1221 s/mm² (n=11) with single-shot pulsed field gradient spin-echo EPI using imaging parameters described in previously published reports. The isotropic DWI was formed from the geometric mean of the high b value single-shot images. The ADC image was calculated from the slope of the linear regression fit of the log of the high and low b-value images versus their b values. PWI were acquired from dynamic susceptibility contrast images by using either spin-echo (n=10) or gradient-echo (n=4) EPI pulse sequences. Images were acquired during the first pass of a bolus of 0.1 mmol/kg (gradient-echo) or 0.2 mmol/kg (spin-echo) of body weight of gadopentetate dimeglumine contrast agent (Magnavist; Berlex Laboratories) injected with an MRI-compatible power injector (Medrad). For both the diffusion and perfusion studies, the FOV was 400×200 mm² with an acquisition matrix of 256×128 acquired with a slice thickness of 6 mm and a 1-mm interslice gap. Relative regional cerebral blood volume (CBV), relative cerebral blood flow (CBF) and mean transit time maps were calculated using techniques described in previously published reports. Each patient was also imaged with conventional sequences following the acute stroke protocol previously described in published reports.

Coregistration

The volumetric diffusion, perfusion, and follow-up data were spatially coregistered with an automated image registration software package, AIR 3.08 (University of California at Los Angeles). The initial low b value T2-weighted EPI, ADC, DWI, and follow-up T2-weighted FSE images were coregistered to the same dimensions (128×128×11 or 128×128×10 voxels), orientation, and coordinates as the perfusion images using an affine, 12-parameter transformation model and trilinear interpolation. Voxels from “normal”-appearing gray matter in the unaffected, contralateral hemisphere from the coregistered initial T2 images were outlined before generation of the predictive maps. For all 6 acute-stage images, voxel values were normalized by dividing by the mean of these outlined regions to produce “relative” values (rT2, rADC, rDWI, rCBF, rCBV, rMTT).

Development of Generalized Linear Model Algorithms

In our GLM algorithms, tissue outcome was modeled as a binary variable (infarcted/noninfarcted) P, where the value 1 represented infarcted tissue and value 0 noninfarcted tissue. In a GLM, for a binary variable, the probability of tissue infarcting can be represented by the logistic function

\[ P = \frac{1}{1 + e^{-\eta(x)}} \]

where \( \eta(x) \) is a linear function of its input parameters, x, \( x \in \mathbb{R}^d \),

\[ \eta(x) = \beta^T x + \alpha \]

and \( \beta \) is the vector of calculated coefficients and \( \alpha \) the bias or intercept term for the GLM. The \( \alpha \) term provides the base value for P if all of the input parameters x are zero. The \( \beta \) coefficients can be interpreted as the multiplicative effects on P due to changes in the input parameters. A supervised approach was used to calculate the coefficients in the GLM algorithms. Using commercial image processing software...
TABLE 1. Patient Demographics, Vascular Territory, Presenting Symptoms, Stroke Subtype, Imaging Times, and Follow-Up Lesion Volume

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Sex</th>
<th>Vascular Territory</th>
<th>Symptoms</th>
<th>TOAST Classification</th>
<th>Initial DWI/PWI, Days, Volume,</th>
<th>Follow-Up T2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>78</td>
<td>F</td>
<td>LMCA R</td>
<td>Aphasia</td>
<td>Cardioembolism</td>
<td>4 5 86.7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>77</td>
<td>F</td>
<td>LPCA L</td>
<td>Anamnestic syndrome</td>
<td>Cardioembolism</td>
<td>4 6 0.41</td>
<td></td>
</tr>
<tr>
<td>*3</td>
<td>58</td>
<td>M</td>
<td>LMCA R</td>
<td>R hand weakness and slurred speech</td>
<td>Cardioembolism</td>
<td>6 5 5.3</td>
<td></td>
</tr>
<tr>
<td>*14</td>
<td>64</td>
<td>F</td>
<td>RACA L</td>
<td>L hand and L leg weakness</td>
<td>UC</td>
<td>5.5 7 2.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>69</td>
<td>M</td>
<td>LPCA L</td>
<td>R homonymous hemianopia and pure alexia without agraphia</td>
<td>Cardioembolism</td>
<td>3 5 22.9</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>79</td>
<td>M</td>
<td>LMCA R</td>
<td>R hemiplegia and mutism</td>
<td>LAA</td>
<td>7 8 130.1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>33</td>
<td>M</td>
<td>LMCA R</td>
<td>R hemiplegia and aphasia</td>
<td>LAA</td>
<td>4 178 83.9</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>M</td>
<td>LMCA R</td>
<td>R hemiplegia and mutism</td>
<td>UC</td>
<td>5 44 58.7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>65</td>
<td>M</td>
<td>RMCA R</td>
<td>L weakness and hemineglect</td>
<td>LAA</td>
<td>10 379 96.6</td>
<td></td>
</tr>
<tr>
<td>*10</td>
<td>61</td>
<td>M</td>
<td>RMCA R</td>
<td>L hemiparesis</td>
<td>Cardioembolism</td>
<td>3 11 1.2</td>
<td></td>
</tr>
<tr>
<td>‡11</td>
<td>72</td>
<td>M</td>
<td>RMCA R</td>
<td>Aphasia</td>
<td>Cardioembolism</td>
<td>6 54 21.7</td>
<td></td>
</tr>
<tr>
<td>‡12</td>
<td>80</td>
<td>F</td>
<td>RMCA L</td>
<td>L facial drop and pronator drift</td>
<td>OIC (balloon occlusion)</td>
<td>11 7 32.8</td>
<td></td>
</tr>
<tr>
<td>‡13</td>
<td>45</td>
<td>M</td>
<td>RMCA R</td>
<td>L hemiparesis and plegia of leg</td>
<td>OIC (dissection)</td>
<td>4 57 18.9</td>
<td></td>
</tr>
<tr>
<td>‡14</td>
<td>45</td>
<td>M</td>
<td>LMCA R</td>
<td>R homonymous hemianopia, r hemiparesis</td>
<td>OIC (dissection)</td>
<td>7 58 92.2</td>
<td></td>
</tr>
</tbody>
</table>

Unless otherwise noted, all diffusion sequences were acquired axially at TR=6000, b-value=1221 s/mm² up to 20 slices and perfusion sequences with TR/TE=1500/75 ms, 11 slices, and 46 time points. LAA indicates large-artery atherosclerosis; OIC, other identified cause; and UC, undetermined cause.

DWI sampled in 3 directions, b-value=1010 s/mm²; PWI consists of 10 slices and 51 time points.

Gradient echo EPI sequences used for PWI at TR/TE=1500/50 ms.

Early spontaneous reperfusion (<15 hours) as determined by serial MR perfusion studies.

(Alice, Hayden Image Processing Solutions), training regions were selected by outlining brain tissue volumes that were clearly infarcted or noninfarcted in the ipsilateral hemisphere in the coregistered follow-up axial T2 FSE images by a neuroradiologist blinded to the predictive map results. Care was taken to avoid including regions demonstrating chronic changes on T2, such as old stroke lesions or periventricular white matter abnormalities. Selection of “normal” voxels was also limited to the ipsilateral hemisphere in slices that showed evidence of infarction. Combinations of initial rT2 EPI, rADC, rDWI, rCBF, rCBV, and rMTT values from these outlined training regions were used as the input vector x in the training stage. Because GLM algorithms assume independent observations, only training regions were used as the input vector x in the training stage. The coefficients for the GLMs, β, were calculated using an iterative reweighted least-squares algorithm in S-PLUS 3.4 (StatSci). Selection of covariates was on the basis of the Akaike Information Criterion (AIC), whereby terms were included if their addition resulted in reductions in prediction error values that were a function of both training error and complexity.34 The AIC therefore provided an objective means to evaluate the trade-off between minimizing residual training error and complexity.34 The algorithm with the minimum AIC is therefore one with the minimum number of parameters and minimum training error. Automatic parameter selection was not utilized because all the input parameters were not independent with MTT=CBF/CBV and DWI=T2 exp(b ADC). Therefore, in selecting covariates, independent parameters rT2, rADC, rCBF, and rCBV were considered first for inclusion, followed by the higher-order covariates of rDWI and rMTT. For purposes of comparing the 2 techniques, combinations of DWI and PWI identical to those created for the thresholding algorithms were generated for the GLM algorithms.

To validate the performance of the GLMs, a jacknifing approach was followed wherein the coefficients for each patient’s algorithms were calculated using the other patients in the study as training data.33 Jacknifing was used to avoid bias that would otherwise occur if the algorithm’s performance were evaluated on the same data that was used to train the algorithm. Using the calculated coefficients, the risk of a voxel of tissue going on to infarction was calculated with Equations 1 and 2. The 95% confidence intervals for the computed coefficients obtained using a data set containing data from all patients. The average of the coefficients of the GLM algorithms obtained from the 14 training data subsets was also compared with the coefficients of the aggregate GLM algorithm. Two-tailed Z tests were used for the statistical comparisons.

Thresholding Algorithms

For the thresholding algorithms, a strategy similar to that reported by Welch et al33 was followed. Tissue was classified as abnormal if the initial diffusion or perfusion values were greater than a specified number of standard deviations (SDs) from the mean value measured in the contralateral noninfarcted gray matter regions. We generated tissue signature maps by using images calculated from the diffusion study (T2+ADC+DWI), images calculated from the perfusion study (CBF+CBV+MTT), and combinations of images from both studies. For the combined study, we generated signature maps using combinations of T2 and ADC with 1 perfusion parameter (CBF, CBV or MTT) and all 6 parameters (T2+ADC+DWI+CBF+CBV+MTT). The combinations of the parameters used for the thresholding algorithms were selected to be identical to the combination of parameters used in the GLM algorithms for the purpose of comparing the 2 techniques. For creating signature maps, a threshold of 2 SDs from the mean of the contralateral values was used. Each of the resulting signatures was taken to represent a different “state” of
infarction. Voxels not meeting any of the threshold criteria were given a “normal” signature. For the thresholding algorithms, which are based on an unsupervised approach not requiring training data from other subjects, the nonnormalized data sets were used.

**Evaluation of Algorithm Performance**

To evaluate the accuracy of the thresholding and GLM algorithms, the same infarcted and noninfarcted regions used in the training of the GLM algorithms were used. The performance of each of the algorithms was evaluated on its ability to accurately discriminate the infarcted from noninfarcted regions in the ipsilateral hemisphere. By comparing the predicted maps with lesions demonstrated on follow-up conventional MR images, the number of voxels predicted to infarct that actually did infarct (true positives [TP]), and the number that did not infarct (false-positives [FP]) were tabulated. In addition, we tracked the number of voxels predicted not to infarct that remained noninfarcted (true negatives [TN]) as well as those that became infarcted (false-negatives [FN]). From these counts, the algorithm’s sensitivity or true positive ratio, TPR = TP/(TP + FN), and specificity or true negative ratio, TNR = TN/(TN + FP), were calculated. Receiver operating characteristic (ROC) curves were then generated for each algorithm by plotting TPR (sensitivity) against the false-positive ratio (FPR) (1-specificity). For thresholding algorithms, the number of SDs was varied from −5 to 5 in 0.1 increments for all parameters except MTT. For MTT, cutoff values ranged from −10 to 10 SDs in 0.2 increments due to its larger range of values. For the GLM algorithms, the probability cutoffs for classifying tissue to be infarcted were varied from 0 to 1 in 0.01 increments. The area under the ROC curves (AUC) has been shown to represent the probability that an image will be correctly ranked above the normal or abnormal and therefore used to assess the performance of diagnostic systems.36 We calculated the AUC for the ROC curves for the estimates of each parameter. The coefficients and intercepts for the aggregate GLM algorithm (Table 2) were not significantly different (P = 0.5) from the mean of the coefficients across the 14 subjects. However, the coefficients and intercepts for some parameters in individual patients showed significant differences (P ≤ 0.05) from the aggregate GLM algorithm (Table 2), demonstrating the potential effects of training data on algorithm development.

**ROC Analysis**

Figure 1 shows the ROC curves of the pooled results from the thresholding and GLM methods across all 14 patients for the multivariate and univariate GLM algorithms. For both approaches, the multivariate GLM algorithms performed better than the univariate GLM algorithms, as measured by higher ROC curves. Furthermore, GLM algorithms that combined diffusion and perfusion data performed better than the rT2 + rADC + rDWI or rCBF + rCBV + rMTT GLM algorithms, as shown by the higher ROC curves. (In the interest of clarity, only the ROC curve of the single diffusion-perfusion combinations with the highest curve is shown. This was the model combining T2, ADC, and MTT.) The full 6-parameter “combined algorithm” has a higher ROC curve than GLM algorithms using only rT2 + rADC + rMTT or rT2 + rADC + rCBF + rCBV parameters, consistent with the AIC results. For the diffusion- and perfusion-based GLM algorithms, the multivariate algorithms provided the best performance in terms of ROC curves, and therefore the univariate diffusion and perfusion studies are not discussed in further detail in this study. Of the combined algorithms, the algorithm using all 6 parameters provided the best performance, and therefore the other combined algorithms are also not discussed in the remainder of this study.

**Results**

**Generalized Linear Model Algorithms**

Based on the AIC, GLM algorithms using different combinations of the possible 6 input parameters were evaluated and compared, with data from all 14 patients. Of the independent parameters, rCBF resulted in the greatest reduction in the AIC, followed by rT2, then rADC, and finally rCBV. Adding the higher-order terms, rDWI and rMTT resulted in a further reduction of AIC and were therefore included in the combined diffusion and perfusion GLMs. Therefore, the optimal GLM algorithm by the AIC requires all 6 parameters. However, for the purpose of comparison, the GLM coefficients for all possible 63 GLMs were calculated. A subset of these 63 GLMs that performed best in terms of sensitivity and specificity is shown in Table 2, along with the standard errors for the estimates of each parameter. The coefficients and intercepts for the aggregate GLM algorithm (Table 2) were not significantly different (P = 0.5) from the mean of the coefficients across the 14 subjects. However, the coefficients and intercepts for some parameters in individual patients showed significant differences (P ≤ 0.05) from the aggregate GLM algorithm (Table 2), demonstrating the potential effects of training data on algorithm development.

**TABLE 2. Coefficients of GLM Algorithms for All 14 Subjects**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$\alpha$</th>
<th>rT2</th>
<th>rADC</th>
<th>rDWI</th>
<th>rCBF</th>
<th>rCBV</th>
<th>rMTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>rT2 + rADC + rDWI</td>
<td>−10.0 ± 0.2</td>
<td>−2.9 ± 0.2</td>
<td>4.9 ± 0.2</td>
<td>6.7 ± 0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rCBF + rCBV + rMTT</td>
<td>−1.2 ± 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rT2 + rADC + rMTT</td>
<td>−3.6 ± 0.06</td>
<td>4.4 ± 0.08</td>
<td>−3.5 ± 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rT2 + rADC + rCBF + rCBV</td>
<td>−1.6 ± 0.05</td>
<td>4.4 ± 0.08</td>
<td>−3.3 ± 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined algorithm</td>
<td>−11.7 ± 0.2</td>
<td>−3.0 ± 0.2</td>
<td>5.9 ± 0.2</td>
<td>7.1 ± 0.2</td>
<td>−1.2 ± 0.1</td>
<td>0.5 ± 0.06</td>
<td>0.7 ± 0.03</td>
</tr>
</tbody>
</table>

The columns labeled rT2, rADC, rDWI, rCBF, rCBV, and rMTT represent the mean and SE of the weighting coefficient for each respective parameter when utilizing all 14 patients for the training data set. The column labeled $\alpha$ is the bias or intercept term. Ellipses indicate that the parameter was not used for a particular multivariate algorithm. Each row represents the coefficients for the different GLM algorithms investigated.

By assuming equal prevalence of infarcted (0.5) and noninfarcted (0.5) voxels and equal costs of false-positives and false-negatives, the OOPs for the ROC curves were determined numerically by finding the points on the ROC curves where the slope was equal to 1.

**Results**

**Generalized Linear Model Algorithms**

Based on the AIC, GLM algorithms using different combinations of the possible 6 input parameters were evaluated and compared, with data from all 14 patients. Of the independent parameters, rCBF resulted in the greatest reduction in the AIC, followed by rT2, then rADC, and finally rCBV. Adding the higher-order terms, rDWI and rMTT resulted in a further reduction of AIC and were therefore included in the combined diffusion and perfusion GLMs. Therefore, the optimal GLM algorithm by the AIC requires all 6 parameters. However, for the purpose of comparison, the GLM coefficients for all possible 63 GLMs were calculated. A subset of these 63 GLMs that performed best in terms of sensitivity and specificity is shown in Table 2, along with the standard errors for the estimates of each parameter. The coefficients and intercepts for the aggregate GLM algorithm (Table 2) were not significantly different (P = 0.5) from the mean of the coefficients across the 14 subjects. However, the coefficients and intercepts for some parameters in individual patients showed significant differences (P ≤ 0.05) from the aggregate GLM algorithm (Table 2), demonstrating the potential effects of training data on algorithm development.

**ROC Analysis**

Figure 1 shows the ROC curves of the pooled results from the thresholding and GLM methods across all 14 patients for the multivariate and univariate GLM algorithms. For both approaches, the multivariate GLM algorithms performed better than the univariate GLM algorithms, as measured by higher ROC curves. Furthermore, GLM algorithms that combined diffusion and perfusion data performed better than the rT2 + rADC + rDWI or rCBF + rCBV + rMTT GLM algorithms, as shown by the higher ROC curves. (In the interest of clarity, only the ROC curve of the single diffusion-perfusion combinations with the highest curve is shown. This was the model combining T2, ADC, and MTT.) The full 6-parameter “combined algorithm” has a higher ROC curve than GLM algorithms using only rT2 + rADC + rMTT or rT2 + rADC + rCBF + rCBV parameters, consistent with the AIC results. For the diffusion- and perfusion-based GLM algorithms, the multivariate algorithms provided the best performance in terms of ROC curves, and therefore the univariate diffusion and perfusion studies are not discussed in further detail in this study. Of the combined algorithms, the algorithm using all 6 parameters provided the best performance, and therefore the other combined algorithms are also not discussed in the remainder of this study.

Algorithms that use only perfusion imaging appear to have greater sensitivity in regions of low specificity (FPR > 0.3). For algorithms that use only diffusion imaging, the reverse appear true; that is, the diffusion-based algorithm had greater sensitivity in regions of low specificity (FPR > 0.9).
sensitivity than perfusion-based algorithms in ranges of high specificity (FPR < 0.3). When we combine perfusion and diffusion information concurrently, we obtain an overall increase in sensitivity. Table 3 shows the specificities associated with the OOPs for both thresholding and GLM algorithms, along with their corresponding sensitivities. The OOPs are comparable for both thresholding and GLM algorithms. For both algorithms, from the ROC curves shown in Figure 1, the combined algorithms have the greatest sensitivities at each of the specificities listed in Table 3.

From Figure 1, we see that both thresholding and GLM methods produce similar ROC curves when results were pooled across the 14 subjects. ROC curves were also generated on an individual patient basis and the area under the curves (AUC) calculated. The differences between the multivariate algorithms’ AUCs were calculated for the thresholding and GLM algorithms. For the thresholding algorithm, the combined algorithm had significantly higher AUCs than the diffusion-based algorithm (T2+ADC+DWI) (P=0.02), indicating better overall performance of the combined threshold algorithm over the initially proposed diffusion-only thresholding algorithm.23–26 The difference between the combined algorithm and CBF+CBV+MTT threshold algorithms were not significant (P=0.21). No significant difference was found between the performances of threshold algorithms based purely on diffusion (T2+ADC+DWI) and those based purely on perfusion (CBF+CBV+MTT) (P=0.52). For the GLM algorithms, the combined algorithm showed a significant improvement over diffusion-based algorithms (rT2+rADC+rDWI) (P=0.02) and perfusion-based algorithms (rCBF+rCBV+rMTT) (P=0.04). There was no significant difference between multivariate diffusion and multivariate perfusion GLM algorithms (P=0.50). The lack of difference between the diffusion and perfusion algorithms for both GLM and thresholding algorithms is most likely because diffusion algorithms have lower sensitivity at low specificity than perfusion algorithms but higher sensitivity at high specificity, which may in turn translate into equivalent AUCs.

Differences between the AUCs for the GLM algorithms and their corresponding threshold algorithm counterparts were calculated and compared. The GLM and thresholding algorithms that used diffusion data (P=0.33), perfusion data (P=0.64), or combined algorithms (P=0.27) performed comparably.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Cutoff Value</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thresholding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2+ADC+DWI</td>
<td>2.2</td>
<td>0.87</td>
<td>0.54</td>
</tr>
<tr>
<td>CBF+CBV+MTT</td>
<td>1.6</td>
<td>0.64</td>
<td>0.72</td>
</tr>
<tr>
<td>Combined algorithm</td>
<td>2.7</td>
<td>0.83</td>
<td>0.66</td>
</tr>
<tr>
<td>GLM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2+ADC+DWI</td>
<td>34</td>
<td>0.90</td>
<td>0.50</td>
</tr>
<tr>
<td>CBF+CBV+MTT</td>
<td>28</td>
<td>0.65</td>
<td>0.71</td>
</tr>
<tr>
<td>Combined algorithm</td>
<td>32</td>
<td>0.84</td>
<td>0.66</td>
</tr>
</tbody>
</table>

The optimal operating points were determined for each of the evaluated algorithms. The cutoff values used for classification of infarcted and noninfarcted voxels that are associated with the OOPs are also shown. The cutoff values are in number of SDs for the thresholding algorithm for all parameters with the exception of MTT. The cutoff threshold of MTT was twice the SDs of the other 5 parameters. The cutoff values for the GLM algorithm are in percent risk of infarction. The third column and fourth represents the specificities and sensitivities at the OOPs for each of the algorithms.

Table 3. Optimal Operating Points for Thresholding and GLM Algorithms

![Figure 1](http://stroke.ahajournals.org/content/1/10/937/F1.large.jpg)
**Example Cases**

Figure 2 shows the acute imaging studies and thresholding maps for patient 14. The tissue signature maps are the results of using only hyperacute diffusion data ($T2_{ADC}^{1}DWI$), hyperacute perfusion data ($CBF+CBV+MTT$), and combining all 6 input parameters (combined algorithm). The diffusion-based algorithm, though identifying a smaller region at risk of infarction in the ipsilateral hemisphere than either the perfusion-based algorithm or combined algorithm, also demonstrates an abnormal signature in the contralateral hemisphere. Abnormal tissue signatures in the perfusion-based algorithm are predominantly limited to the ipsilateral hemisphere, although they encompass an area much greater than the follow-up infarct volume. Because misclassifications are cumulative in the thresholding algorithms, the results in the combined diffusion and perfusion algorithms have similarly high sensitivity but poor specificity as that shown for the perfusion-based algorithms. However, a greater number of tissue states exist in the combined algorithm, which results in greater heterogeneity than those based on algorithms incorporating only diffusion or perfusion information.

Figure 3 shows the results of the GLM algorithms using the same imaging data as shown in Figure 2. We again observe that algorithms using diffusion alone ($rT2+rADC+rDWI$) underestimate the follow-up infarct volume. Maps that use only perfusion information ($rCBF+rCBV+rMTT$) overestimate the follow-up infarct volume. The combined algorithm, however, predicts an area at high risk of infarction, as evidenced by the red-yellow region, that correlates well with the follow-up lesion areas, as demonstrated on the 2-month follow-up T2 FSE image shown in Figure 2. In addition, for all algorithms, the regions predicted to be at high risk of infarction are predominantly localized to the ipsilateral hemisphere compared with the results of the thresholding algorithm.

The results of applying the statistical algorithms to a patient with early reperfusion, as defined by follow-up perfusion studies, are shown in Figure 4. The acute MRI studies for patient 11 appear normal, with the exception of decreased CBF and increased MTT in the left temporoparietal lobe. The imaging study 8 hours later shows a slight diffusion abnormality in the area shown abnormal in the initial perfusion study. However, the remaining perfusion defects appear to have resolved as demonstrated by the CBF and MTT maps, suggesting the occurrence of spontaneous reperfusion. Both the thresholding- and the GLM-based risk maps overpredict the follow-up infarct volume in the 2-month follow-up T2 FSE. The resolution of much of the abnormalities in the follow-up imaging study was consistent with the patient’s improved clinical outcome.

**Discussion**

Our data demonstrate the feasibility of generating, on a voxel-by-voxel basis, quantitative predictive maps of tissue outcome with use of acute MRI images. Our results are consistent with earlier studies which show that tissue predictive algorithms can combine on a voxel-by-voxel basis multiple image modalities successfully into a single map of...
tissue status. Our results extend these earlier algorithms in 2 ways. First, we extend the algorithms based on thresholding to include perfusion information. Second, we extend tissue signature algorithms to provide a voxel-by-voxel prediction of the viability of brain tissue in “risk maps” using a generalized statistical approach.

**Combined Diffusion and Perfusion Algorithms**

By extending tissue signature algorithms based on thresholding to include perfusion information, our results show that such inclusion improves the predictive power of signature maps. While only a trend toward improved performance was demonstrated in the case of the combined thresholding algorithm over the perfusion-based algorithm ($P=0.21$), we believe that further optimization of the threshold approach and refinement in interpretation of results would improve its utility. Although statistical significance was not found in the AUCs for the multivariate diffusion-based algorithm compared with the multivariate perfusion-based algorithm using either thresholding or GLM, the ROC curves demonstrate the difference between the diffusion-based and perfusion-based algorithms in their tradeoffs between sensitivity and specificity, a characteristic not evident in simple AUC indices. We observe that diffusion-based algorithms have higher sensitivity in regions of high specificity or low FPR, whereas perfusion-based algorithms have higher sensitivity in regions of low specificity. Combined algorithms appear to provide the best trade-off in terms of maintaining high sensitivity at high specificity.

The high specificity of diffusion-based algorithms is not unexpected because of the association between high risk of infarction and changes in diffusion parameters, which are believed to detect tissue with altered cellular water homeostasis caused by severe energy depletion and breakdown in Na$^+$-K$^+$ pump activity. The level of sensitivity of diffusion-based algorithms is time dependent: less sensitive at very early imaging times before DWI reaches its maximum, and more sensitive hours later when DWI lesion size approaches the “final” infarct size. However, a simple reduction
of ADC may not be a marker for irreversibly injured tissue, and indeed, a set “threshold” for irreversible ADC reductions may be difficult to determine, because the threshold varies as a function of depth and duration of ischemia. The perfusion parameters, on the other hand, presumably reflect the state of nutritive flow to the voxel of tissue. The lack of specificity but high sensitivity in perfusion-based algorithms may be attributed to the presence of metabolically viable hypoperfused tissue at flow levels below the threshold for electrical neuronal failure. The likelihood for tissue to infarct is a combined function of the degree and the duration of blood flow reduction, which have been shown to vary spatially and temporally. Therefore, perfusion-based algorithms may also have a similar level of time-varying sensitivity and specificity that varies on a voxel-by-voxel basis.

GLM Algorithms
Of the 2 techniques examined in this study for combining diffusion and perfusion information, the GLM method may provide results that are straightforward to interpret as additional parameters are included in the algorithm. In initial thresholding algorithms, a key feature was the ability to assign each tissue signature based on imaging to a possible physiological state of the tissue. However, with the addition of multiple parameters, each additional term exponentially complicates output interpretation, since the signature maps create additional states whose biological significance is not necessarily clear. Nevertheless, thresholding algorithms may provide unique insight regarding heterogeneity of the ischemic lesion at any single point in time. Further investigations correlating evolution of these signatures with histology may provide insight into the pathophysiologic significance of the different signatures. GLM algorithms, on the other hand, provide the risk of the tissue infarcting as a continuous variable that ranges between 0 and 1, and therefore, as stroke evolves, the risk of individual voxels of tissue can be monitored quantitatively by a single variable. The recruitment of voxels in the presumed “ischemic penumbra” might therefore be quantified as the change in risk in the peripheral areas from low probability to high probability over time.

Our algorithms have been trained on data from patients who did not receive thrombolytic or neuroprotective therapy. The 2

Figure 4. A 72-year-old man (patient 11) imaged within 6 hours of presenting with symptoms of conduction aphasia. His acute diffusion studies appear normal. However, in the perfusion maps, areas of decreased CBF and increased MTT are observed in the left temporoparietal lobe. The imaging study 8 hours later show a slight diffusion abnormality in the area shown abnormal in the initial perfusion study. However, the remaining perfusion defects appear to have resolved as demonstrated by the CBF and MTT maps. This normalization of MTT on the early follow-up study suggests that the patient had spontaneously reperfused soon after the initial insult. Both the thresholding and the GLM-based risk map overpredict the follow-up infarct volume, as reflected by a 2-month follow-up T2 FSE.
patients with spontaneous reperfusion were specifically not excluded from the training set since their inclusion was believed to be a better reflection of the naturally occurring ischemic stroke patient population in which spontaneous reperfusion has been detected within 24 hours after symptom onset in 24% of patients with transcranial Doppler ultrasound. Therefore, our algorithms’ predictions seem likely to be based on the natural evolution of ischemic tissue undergoing infarction. However, our training set is small, and therefore does not yet capture the full range and frequency of stroke evolution possibilities. For example, if in a new patient an event occurs to interrupt the progression of ischemic damage as quantified from the training patient data, the probability of infarction of individual tissue regions may change greatly. This was apparent in the case of patient 11, who exhibited spontaneous reperfusion (Figure 4). For such circumstances, progression of infarct lesion size has been shown to be diminished. A similar change in probabilities might be seen after successful therapeutic reperfusion or after administration of an effective neuroprotective agent. This method, therefore, appears to provide a technique that might be used to monitor this change in risk quantitatively. Were this approach to be validated, the GLM approach could become a useful statistical method for evaluating the efficacy of novel therapies and possibly even develop into a tool to help guide the choice of appropriate therapy for individual patients.

Future Investigation

Our data demonstrate in a preliminary fashion the feasibility of combining diffusion and perfusion information into a single index of tissue risk. Although collection of additional patient data will make the specific algorithm parameters more robust, this would not necessarily change the methodology we have developed for analyzing and quantifying this natural history data. On the other hand, we believe there are still many avenues of investigation for improving these algorithms. Clearly, the retrospective aspect of this study limited our models. As demonstrated by the large variance in lesion volumes and etiologies across patients in this report, prospective studies involving a greater cohort of patients with standardized MR acquisition parameters and follow-ups at set intervals are needed to further test the validity of the algorithmic approaches described here. For example, an overestimation of “final” lesion volumes in some patients may have occurred because of the possible presence of vasogenic edema at 5 days after ictus, resulting in the use of wrongly classified voxels in the training and evaluation of our algorithms. In addition, inaccuracies in the coregistration may have introduced errors in both algorithm development and evaluation. Although intrasubject studies have shown the average misregistration size to be <1 mm, less than our voxel dimensions, the maximum misalignment has been reported to be as large as 3.8 mm. This suggests that our algorithms’ results may be inaccurate for cases involving small infarct volumes. The addition of acute clinical variables as covariates may also improve our models’ performances, as has been demonstrated by another study that predicted clinical outcome by combining imaging data with initial clinical variables. A priori assumptions in algorithm design, principally that the risk of infarction changes linearly with the covariates, may also have negatively impacted the performance of both thresholding and GLM algorithms. Several studies have shown that the risk of infarction does not change linearly for some of the algorithm variables. For example, ADC has been well documented to first decrease in acute cerebral ischemia before pseudonormalizing and increasing in the chronic stage. This nonlinear behavior may also hold true for perfusion metrics even in the hyperacute stage. Recent studies have found both increased and decreased CBV in acutely imaged lesions (<12 hours) that become infarcted, as shown by follow-up MR studies. The GLM algorithms we used in this study assume linear behavior. This suggests that additional investigations of algorithms that take into consideration the nonlinear behavior of covariates may provide improved performance.

Finally, there are a few additional technical limitations to our approach. Our models are almost certainly limited because they do not account for the intrinsic anatomic variations in both normal and pathophysiologic conditions. For instance, white matter may be misclassified as territory at risk of infarction because its normal flow values fall within the ischemic range for gray matter. Expert models that can differentiate white from gray matter and apply the appropriate tissue specific model to obtain an assessment of infarction risk can potentially compensate for this limitation.

Conclusion

Despite some limitations, we have shown that algorithms combining diffusion and perfusion information can assess the risk of infarction at the acute stage with greater sensitivity and specificity than algorithms using diffusion and perfusion information separately. Of the combined algorithms studied, the generalized linear model algorithm may provide the preferred approach owing to its potentially greater ease of interpretation with its single index of risk. Although further investigation and algorithm refinement is necessary, this method for quantitatively assessing the risk of infarction on a voxel-by-voxel basis shows promise as a technique for not only gaining insight into the natural spatial evolution of ischemic damage in humans but also evaluating the effects that novel therapies may have on this process.

Acknowledgments

This work was supported in part by PHS R01NS8477 (Ms Wu and Dr Sorensen) and NIH grants RR13213 (Dr Gonzalez), NS34626 (Dr Gonzalez), R01-1H39810 (Drs Sorensen and Rosen), 5P41-RR14075 (Dr Rosen), and CA83159 (Dr Gonzalez). The authors would like to thank Dr Elkan Halpern, Dr Emery Brown, and Dr Rick M. Dijkhuizen for manuscript review and feedback. Massachusetts General Hospital has filed a provisional patent application on the use of the statistical/ GLM approach for the evaluation of novel pharmacological agents; this application lists Drs Wu and Sorensen as inventors.

References

3. Barber PA, Darby DG, Desmond PM, Gerraty RP, Yang Q, Li T, Jolley D, Donnan GA, Tress BM, Davis SM. Identification of major ischemic...


Predicting Tissue Outcome in Acute Human Cerebral Ischemia Using Combined Diffusion- and Perfusion-Weighted MR Imaging


*Stroke*. 2001;32:933-942
doi: 10.1161/01.STR.32.4.933

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/32/4/933

**Permissions**: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

**Reprints**: Information about reprints can be found online at:
http://www.lww.com/reprints

**Subscriptions**: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/