Predicting tissue outcome in acute human stroke using temporally adaptive MRI-based algorithms

O. Wu[1], L. H. Schwamm[2], W. J. Koroshetz[2], T. Benner[1], W. A. Copen[3], R. G. Gonzalez[3], A. G. Sorensen[1]

[1]Athinoula A. Martinos Center, Massachusetts General Hospital, Charlestown, MA, United States, [2]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States, [3]Department of Radiology, Massachusetts General Hospital, Boston, MA, United States

Introduction: Viability of tissue after stroke depends on two main factors – the depth and the duration of the ischemic insult. Algorithms combining acute diffusion-weighted (DWI) and perfusion-weighted (PWI) MRI have been shown to accurately predict tissue infarction on a voxel-wise basis in both human and experimental cerebral ischemia by providing a measure of depth of ischemic injury. However, the duration of ischemia at the time of imaging, an important parameter for assessing tissue viability and therefore treatment strategy, has so far not been incorporated into models predicting tissue outcome in human stroke despite the fact that the temporal characteristics of DWI- and PWI-derived parameters are known to dramatically change over time. This temporal sensitivity has been implicitly taken into account by limiting predictive algorithms development and application to patients imaged within 12 h of stroke onset [1-4]. It has been recently shown that incorporating occlusion duration in models predicting tissue outcome in experimental animal models is feasible and can be used to provide insight into the spatio-temporal evolution of ischemic injury [5]. We apply these techniques to acute stroke patients first imaged < 12 h from onset who also received follow-up serial DWI/PWI at pre-defined time points, with a conventional follow-up (F/U) at least 5 days after initial imaging. We hypothesize that (1) predicting tissue outcome using models developed with only acute imaging data will underestimate tissue infarction for imaging studies performed at subacute time points and (2) models incorporating ischemic duration will perform significantly more accurately at subacute time points than models using only acute imaging data.

Materials/Methods: Eighteen patients were enrolled in a serial DWI and PWI study of the natural history of human cerebral ischemia. Volumetric findings have been reported for fourteen of these patients [6]. For this study, a subset of these patients with strokes involving the middle cerebral artery territory were retrospectively analyzed (n=6). Only imaging studies containing artifact free DWI and PWI data sets were used resulting in a total of 26 studies, performed at times ranging from 3.8 h to 178 d from stroke onset. None of the patients received thrombolytic or neuroprotective therapies. Using imaging protocols and techniques described previously [1], apparent diffusion coefficient (ADC), cerebral blood flow (rCBF), cerebral blood volume (rCBV), mean transit time (MTT), and tracer arrival time (DELAY) maps were calculated on a voxel-by-voxel basis. The low b-value (b=0 s/mm^2) and high b-value (b=1000 s/mm^2) images were used as the T2-weighted image (T2WI) and isotropic DWI (iDWI) maps respectively These seven images for each imaging time point along with follow-up studies were coregistered using a semi-automated image registration software package, (AIR 3.08, UCLA, CA) [7]. Images were normalized with respect to mean values in the contralateral hemisphere and then used as covariates in predictive algorithms based on a generalized linear model (GLM) whose output values are the probability of infarction [1]. Coefficients were calculated using bootstrapping and jackknifing [8]. Training regions consisted of infarcted tissue delineated by a neuroradiologist on the F/U T2 (at least 5-d from acute study) and non-infarcted tissue defined as all remaining ipsilateral hemisphere tissue.

GLM models were calculated using only acute MRI data (Model A), using MRI data from all time points (Model B) and using imaging time from stroke onset as a covariate and interaction term along with MRI data from all timepoints (Model C). These models were used to predict infarction risk at each imaging time point. Sensitivity and specificity of the GLM in predicting infarction were calculated. The Youden's J index [9], defined as sensitivity+specificity-1, was calculated to assess the overall performance of the algorithms at a single decision threshold of 50% for classifying infarcted tissue. Youden's J index=0 indicates no diagnostic value while J=1, indicates perfect diagnosis. 50% was used since all models were designed for the optimal operating point to be at this risk level. Average GLM-predicted infarction risk in tissue that infarcted was also calculated at each timepoint using the three models and compared. Subset analysis was performed dividing the data into acute (<48 h) (n=15) and chronic studies (> 48 h) (n=11). All statistical comparisons were performed using paired one-sided Wilcoxon signed rank tests.

Results: Models developed using data from only acute imaging (Model A) produced significantly ($p<.05$) lower predicted risk of infarction (.60±.2) in tissue that infarcted than models developed using data from multiple time points (Model B: .64±.2 and Model C: .64±.2). Subset analysis between model performance at acute and chronic time points showed that the estimated risk of infarction was significantly larger ($p=.04$) for the acute predictions for Model A (.66±.2 vs .52±.2). Although predicted infarction risk was reduced (.58±.2) at chronic time points compared to acute time points (.69±.2), the difference was not statistically significant ($p=.07$). No statistically significant difference was found for GLM-predicted risk for Model C between acute and chronic time points (.66±.2 and .62±.2). GLM-predicted risk for acute studies was greatest in maps generated by Model A ($p<.03$). For chronic studies, the situation was reversed with significantly smaller values produced by Model A than Model B ($p=.002$) or Model C ($p=.003$). Model C had a tendency to produce higher risk estimates than Model B ($p=.09$). Youden's J index was significantly smaller ($p<.02$) in Model A (.40±,2) at chronic time points than either Model B (.44±.2) or Model C (.46±.2), reflecting poorer performance. Model C also significantly outperformed Model B ($p=.03$) for chronic studies in terms of the Youden's J index. Acutely, no statistically significant difference was found between the different models (Model A: .58±.2, Model B: .60±.2, Model C: .55±.3). However, Model A performed significantly worse ($p=.03$) for chronic timepoints than acute timepoints. Model B had a tendency to perform poorer at later timepoints but this was not statistically significant ($p=.06$). Fig 1 shows examples of predicted infarction risk Models A-C along with DWI/PWI maps acquired over time along with conventional T2 FSE acquired at the same timepoints. Note, the DWI/PWI mismatches are reflected in the single risk maps as regions of high and low infarction risk for all models. Results from Model A show accurate prediction of the final infarct volume (180 d T2 FSE) for the acute studies, but poor characterization for the subacute ones. In comparison, the risk maps generated using Model C, which incorporates stroke onset time as a covariate, produced the best risk assessment.

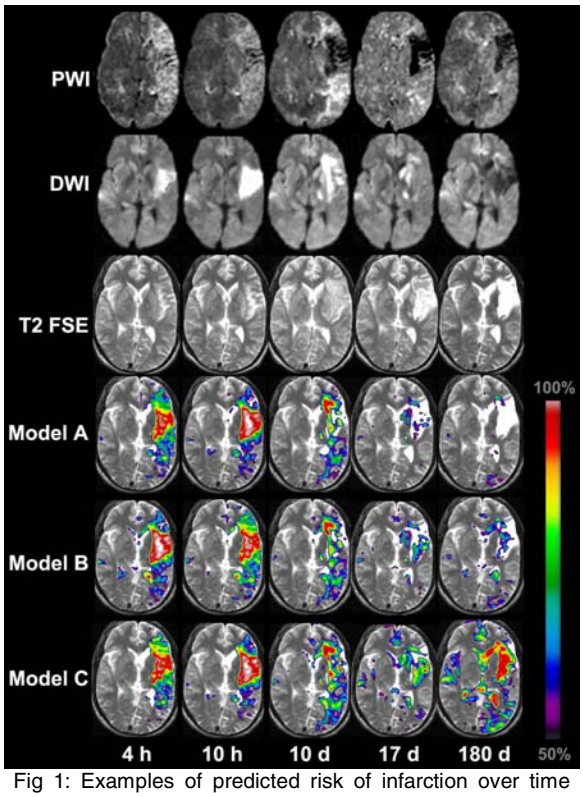

Fig 1: Examples of predicted risk of infarction over time using three different models along with DWI/PWI and T2 FSE data. The model which incorporates temporal information (Model C) provides over a wide range of times the best prediction of final infarct volume (T2 FSE at 180 d).

Discussion: Our results show that incorporation of ischemia duration improves prediction of infarction. We speculate that poorer performances of models developed using only acute imaging are due to the non-monotonic behavior of the model parameters, i.e. both increases and decreases of DWI is associated with infarcted tissue, with this dependency being temporally driven. Therefore, adding information regarding the duration of ischemia improved our algorithms. However, the performance of even these improved temporal models are somewhat reduced at chronic stages. We hypothesize that using non-linear models, rather than GLMs, would further improve prediction.

References: 1. Wu O, et al. *Stroke.* 2001; 32, 933-42. 2. Jacobs MA, et al. *Stroke.* 2001; 32, 943-9. 3. Jacobs MA, et al. *Stroke.* 2001; 32, 950-7. 4. Shen Q, et al. *J. Cereb. Blood Flow Metab.* 2004; 24, 887-97. 5. Wu O, et al; ISMRM 13th Scientific Meeting; 2005. 6. Schwamm LH, et al. *Stroke.* 1998; 29, 2268-76. 7. Woods RP, et al. *J. Comput. Assist. Tomogr.* 1992; 16, 620-33. 8. Efron B. *The jackknife, the bootstrap and other resampling plans.* 1982. 9. Youden WJ. *Cancer.* 1950; 3, 32–5.