Characterizing physiological heterogeneity of infarction risk in acute human ischaemic stroke using MRI

Ona Wu,1,2 Søren Christensen,1 Niels Hjort,1 Rick M. Dijkhuizen,2 Thomas Kucinski,3 Jens Fiehler,3 Götz Thomalla,4 Joachim Röther4 and Leif Østergaard1

1Center for Functionally Integrative Neuroscience, Department of Neuroradiology, Århus University Hospital, Århus C, Denmark, 2Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands, 3Departments of Neuroradiology and 4Departments of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Correspondence to: Ona Wu, PhD, Center for Functionally Integrative Neuroscience, Århus University Hospital, Building 30, Nørrebrogade 44, 8000 Århus C, Denmark
E-mail: ona@pet.auh.dk

Viable tissues at risk of infarction in acute stroke patients have been hypothesized to be detectable as volumetric mismatches between lesions on perfusion-weighted (PWI) and diffusion-weighted magnetic resonance imaging (DWI). Because tissue response to ischaemic injury and to therapeutic intervention is tissue- and patient-dependent, changes in infarct progression due to treatment may be better detected with voxel-based methods than with volumetric mismatches. Acute DWI and PWI were combined using a generalized linear model (GLM) to predict infarction risk on a voxel-wise basis for patients treated either with non-thrombolytic (Group 1; n = 11) or with thrombolytic therapy (Group 2; n = 27). Predicted infarction risk for both groups was evaluated in four ipsilateral regions of interest: tissue acutely abnormal on DWI (Core), tissue acutely abnormal on PWI but normal on DWI that either infarcts (Recruited) or does not (Salvaged), and tissue normal on both DWI and PWI that does not infarct (Normal) by follow-up imaging > 5 days. The performance of the models was significantly reduced for the thrombolysed group compared with the group receiving standard treatment, suggesting an alteration in natural progression of the ischaemic cascade. Average GLM-predicted infarction risk values in the four regions were different from one another for both groups. GLM-predicted infarction risk in Salvaged tissue was significantly higher (P = 0.02) for thrombolysed patients than for non-thrombolysed patients, suggesting that thrombolysis rescued tissue with higher infarction risk than typically measured in tissue that spontaneously recovered. The observed spatial heterogeneity of GLM-predicted infarction risk values probably reflects the varying degrees of tissue injury and salvageability that exist after stroke. MRI-based algorithms may therefore provide a more sensitive means for monitoring therapeutic effects on a voxel-wise basis.

Keywords: mathematical modelling; cerebral ischaemia; magnetic resonance imaging; thrombolytic therapy; outcome measures

Abbreviations: ADC = apparent diffusion coefficient; AUC = area under the ROC curve; CBF = cerebral blood flow; CBV = cerebral blood volume; DELAY = tracer arrival delay; DWI = diffusion-weighted MRI; GLM = generalized linear model; iDWI = isotropic DWI; IQR = interquartile range; MLV = measured lesion volume; MTT = mean transit time; NIHSSS = National Institutes of Health Stroke Scale Score; PI = prediction interval; PLV = predicted lesion volume; PWI = perfusion-weighted MRI; ROC = receiver operating characteristic; rt-PA = recombinant tissue plasminogen activator; T2 EPI = T2-weighted image; TIMI = thrombolysis in myocardial infarction; WM = white matter

Spatial heterogeneity of cerebral infarction risk

Introduction

Early identification of tissue at risk of infarction after acute stroke has been postulated to be a potential aid to therapeutic decision-making, thereby improving patient outcome (Kidwell et al., 2003; Lees et al., 2003; Schellinger et al., 2003; Levine, 2004). As such, there has been increasing interest for assisting clinical decision-making with neuroimaging techniques such as diffusion-weighted MRI (DWI), which is highly sensitive to acute tissue damage, and perfusion-weighted MRI (PWI), which is sensitive to haemodynamic disturbance (Fisher et al., 2001; Hermier et al., 2003; Nighoghossian et al., 2003; Schellinger et al., 2003; Levine, 2004; Warach and Baron, 2004; Hjort et al., 2005). Simple mismatches between larger lesion volumes manifested in acute PWI than in acute DWI have been speculated to be useful for identifying, on an individual patient basis, viable tissue that is at risk of infarction without therapeutic intervention (Baron and Warach, 2005; Davis et al., 2005; Hjort et al., 2005). However, studies have shown that these mismatches, based on dichotomized PWI/DWI parameters, can underestimate the amount of salvageable tissue (Fiehler et al., 2002a; Guadagno et al., 2004) or overestimate the amount of tissue at risk of infarction (Sorensen et al., 1999; Coutts et al., 2003; Sobesky et al., 2004). This may be due to heterogeneity of tissue response to ischaemic injury and to therapeutic intervention (Lo et al., 2005). Multiparametric algorithms that combine MRI modalities on a voxel-wise basis have been shown to more accurately predict risk of infarction in acute human cerebral ischaemia than when these MRI methods are used separately (Jacobs et al., 2003; Herment et al., 2003; Boyle et al., 2004). As such, there has been increasing interest for assisting clinical decision-making with neuroimaging techniques such as diffusion-weighted MRI (DWI), which is highly sensitive to acute tissue damage, and perfusion-weighted MRI (PWI), which is sensitive to haemodynamic disturbance (Fisher et al., 2002a, b; 2004a, b). Patients who developed type 2 parenchymal haematomas, associated with altering the clinical course of ischaemic stroke (Fiorelli et al., 1999), were excluded. Patients were treated either with non-interventional standard medical treatment (i.e. no thrombolysis) (Group 1, n = 11) or with intravenous recombinant tissue plasminogen activator (rt-PA) therapy (Group 2, n = 27). Thrombolysis was performed ≤3 h according to ECASS II criteria (Hacke et al., 1998). For patients where MRI was the primary and only imaging modality, modified criteria were used, excluding patients from thrombolysis with signs of intracerebral haemorrhage on MRI and those with DWI lesions >50% of the MCA territory. Thrombolysis beyond 3 h was performed if the PWI lesion volume exceeded the DWI lesion by ≥20%, the DWI lesion was smaller than one-third of the middle cerebral artery territory, and informed consent was obtained. The National Institutes of Health Stroke Scale Score (NIHSSS) was assessed by a stroke neurologist at each imaging time point. Reperfusion was determined on the basis of PWI and MR angiography studies acquired on Day 1 and F/U using modified thrombolysis in myocardial infarction (TIMI) criteria (Fiehler et al., 2004a). Patients were classified as exhibiting no, minimal, incomplete or complete reperfusion (TIMI = 0, 1, 2 and 3, respectively). Early reperfusers were defined as patients with complete reperfusion (TIMI = 3) on Day 1. Information on reperfusion by Day 1 was not available for one patient in Group 2.

MRI studies

Details of the DWI and PWI imaging protocol have been reported previously (Fiehler et al., 2002a, b; 2004a, b). Apparent diffusion coefficient (ADC) maps were calculated from a 3-point fit (b-values = 0, 500 and 1000 s/mm$^2$) (Fiehler et al., 2002a). The b-value = 0 and b-value = 1000 DWI images were used as the T2-weighted image (T2 EPI) and isotropic DWI (iDWI) maps, respectively. Cerebral blood flow (CBF), cerebral blood volume (CBV) and mean transit time (MTT) maps were calculated from signal curves converted to concentration changes over time curves and deconvolved with an arterial input function selected from the ipsilateral hemisphere (Ostergaard et al., 1996). Tracer arrival delay (DELAY) (Wu et al., 2003) was determined as the time of the peak of the residue function.

Generalized linear model analysis

Calculation of the parameters for a generalized linear model (GLM) has been described previously (Wu et al., 2001). In brief, the outcome status of tissue can be modelled as a binary variable, $P$, where the value 1 represents infarcted tissue and the value 0, non-infarcted tissue. The probability of tissue infarction was represented by the logistic function

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

(1)

where $\eta(x)$, the predictor, is a linear function of its input parameters, $x$, $\eta(x) = b'x + \alpha$. (2)

and $\beta$ is the vector of calculated coefficients and $\alpha$ is the bias or intercept term for the GLM. In this study, the input vector, $x$,
consisted of acute T2 EPI, ADC, iDWI, CBF, CBV, MTT and DELAY maps. All images were co-registered using semi-automated image registration software [MNI Autoreg (Collins et al., 1994)] to one another as well as to a probabilistic brain atlas (Mazziotta et al., 2001). All images were then normalized with respect to mean values measured in normal contralateral white matter (WM). All input maps, except DELAY, were normalized by dividing by the mean. Relative DELAY maps were calculated by subtracting the mean of contralateral WM values. Training regions were defined as the union of hyperintensities on the follow-up T2 EPI and iDWI and non-infarcted tissue as remaining ipsilateral hemisphere tissue. Coefficients of the GLM were calculated and then used to estimate risk of infarction on an individual voxel-wise basis for data from a new subject. Prediction intervals (PIs) were also produced on a voxel-wise basis (MathSoft, 2000). All risk maps were median-filtered to reduce false-positives due to noise.

To evaluate the performance of the GLMs, a jack-knifing or leave-one-out approach was followed for Group 1 (Efron, 1982) to avoid bias that would otherwise occur if a model’s performance was evaluated on the same data that trained the model. For Group 2, an aggregate model trained with data from all Group 1 patients was used. The coefficients for the jack-knifed models were compared with the coefficients of the aggregate model (one-tailed Z-tests). For all models, the bootstrap estimate (Efron, 1982) of the mean of the GLM coefficients was used, with care taken that an equal number of infarcted and non-infarcted samples were used to compensate for imbalanced training data (Japkowicz and Stephen, 2002).

**Statistical analysis**

Two-tailed Wilcoxon rank-sum tests were used for unpaired data, and two-tailed Wilcoxon signed rank tests were performed for pairwise comparisons. Correlation analyses were performed using Pearson product-moment correlation coefficient. For evaluating the accuracy of the GLM predictions, sensitivity and specificity of the model for identifying abnormal tissue on F/U were calculated along with receiver operating characteristic (ROC) curves by varying the probability threshold between 0 and 100% for classifying tissue as infarcted. The area under the ROC curves (AUC) that represents the probability that an image will be correctly classified normal or abnormal was calculated and compared (Hanley and McNeil, 1982). To assess accuracy of the predictions, the root mean square error, RMSE = 1/N \sqrt{\sum (y_i - \hat{\pi}_i)^2}, where y_i is the true outcome for the tissue (0 or 1, i.e. not-infarcted or infarcted), \hat{\pi}_i is the predicted GLM risk value and N is the total number of voxels, was calculated and compared across groups.

The predicted lesion volume (PLV) was defined as tissue where GLM-predicted infarction risk was >50%. Fifty per cent was chosen as the threshold since the models were trained to produce the optimal operating point at this cut-off by using an equal number of infarcted and non-infarcted voxels. Measured lesion volumes (MLV) were defined as described above for training the GLM. Initial ischaemic lesions (Core) were defined as hyperintense regions on acute DWI. Acute PWI lesions were defined as regions of hyperintensities on acute MTT maps. Acute lesions were initially demarcated as tissue with values >2 SDs from mean contralateral values, and then manually adjusted by a neurorologist blinded to the GLM-predicted results to correct for errors in the automatic outlines due to imaging artefacts. Lesion growth (Recruited) was defined as infarcted tissue outlined on follow-up not initially present on the acute DWI. Tissue at risk of infarction that was presumably salvaged (Salvaged) was defined as areas initially abnormal on the PWI but normal on follow-up MRT. Normal tissue was defined as tissue with no apparent abnormalities on acute and follow-up imaging. To minimize errors due to poor co-registration, only patients with at least 5 cm^3 in the Recruited or Salvaged regions were used for ROI analysis. Differences in MLV, PLV and GLM-predicted values between early complete reperfusers (TIMI = 3 by Day 1) and the other patients were also compared.

**Results**

Median time to MRI for both groups was 3 h [interquartile range (IQR) = 2.5–3.3 for Group 1 and 2.6–3.6 for Group 2]. Median age in years for Groups 1 and 2 was 60 (Group 1: IQR = 57–68; Group 2: IQR = 51–68). There was a significant difference between both groups in the acute NIHSS with more patients presenting with greater severity (P = 0.002) in the rt-PA treatment group (median = 15, IQR = 12–19) than in Group 1 (median = 9, IQR = 4–10). In Group 1, five patients had TIMI = 3 by Day 1 with 8 by F/U. In Group 2, five patients had TIMI = 3 by Day 1 with 16 by F/U. No significant difference in initial Core lesion volumes (Group 1: 34 ± 42 cm^3; Group 2: 21 ± 28 cm^3) was found; however, MTT lesion volumes were significantly smaller (P = 0.008) for Group 1 (85 ± 88 cm^3) than Group 2 (165 ± 81 cm^3) owing to patient selection criteria. No significant difference was found between lesion volumes measured on follow-up (Group 1: 60 ± 80 cm^3; Group 2: 63 ± 71 cm^3). Significant differences were found between the seven acute DWI- and PWI-derived parameters (Fig. 1) for Core, Recruited and Salvaged ROIs for patients with Recruited and Salvaged volumes >5 cm^3 for Group 1 (n = 7) and Group 2 (n = 22).

The GLM coefficients of the aggregate model derived from all Group 1 training data were -0.7 ± 0.2, 1.2 ± 0.2, 6.4 ± 0.2, 0.2 ± 0.03, -0.1 ± 0.03, 1.1 ± 0.06 and 0.2 ± 0.01 for relative T2 EPI, ADC, iDWI, CBF, CBV, MTT and DELAY, respectively, with a bias term of ~9.5 ± 0.3. Three of the jack-knifed models had significantly different coefficients from the aggregate model. Figure 2 shows examples of GLM-predicted risk of infarction for patients from both groups involving DWI < PWI who experienced partial reperfusion (TIMI = 2) by Day 1. The GLM-calculated risk map encompasses territory that is abnormal on MTT as well as DWI. For Fig. 2A, the amount of tissue predicted to infarct corresponds well with the follow-up infarct, while for the rt-PA-treated patient (Fig. 2B) the amount of tissue is clearly overestimated. Parts of the initial DWI lesion also recover (arrow), even though probability of infarction is very high. Here, lower and upper 95% PIs correspond well with regions of abnormality on the DWI and MTT maps.

ROC curves of the pooled results across all patients for Group 1 and Group 2 are shown in Fig. 3. Using a threshold of 50%, sensitivity and specificity were 77 and 91%, respectively, for Group 1 and 77 and 80% for Group 2. The AUC for the rt-PA-treatment patients (0.85 ± 0.06) was less (P = 0.07)
than that for the non-rt-PA-treated patients (0.90 ± 0.05). The overall RMSE was significantly greater ($P = 0.0009$) for Group 2 (0.39 ± 0.06) than for Group 1 (0.30 ± 0.07). A significant positive correlation with initial NIHSS score was found for the average GLM-predicted value in tissue predicted to infarct ($r = 0.37; P = 0.03$). The volume of the PLV was also significantly correlated with initial NIHSSS ($r = 0.43, P = 0.01$). The PLV for both groups was significantly correlated with respect to the follow-up lesions (Group 1: $r = 0.90, P = 0.0002$; Group 2: $r = 0.68, P < 0.0001$) though the

Fig. 1 Acute DWI and PWI parameter values (mean ± SD) in Core, Recruited and Salvaged ROIs in patients with Recruited and Salvaged volumes > 5 cm$^3$ for Group 1 ($n = 7$) and Group 2 ($n = 22$) patients. All values were normalized by dividing by mean values in normal ipsilateral hemisphere except for rDELAY, which was calculated by subtracting mean normal values. *$P < 0.05$ Core versus Recruited; †$P < 0.05$ Core versus Salvaged; ‡$P < 0.05$ Recruited versus Salvaged; §$P < 0.05$ versus Normal; ¶$P < 0.1$ Group 1 versus Group 2.

Fig. 2 Examples of GLM-predicted risk of infarction along with 95% PI in cases of DWI and PWI mismatch. For clarity, only GLM-predicted values > 50% are shown overlaid on F/U. (A) Output for non-rt-PA-treated 58-year-old male imaged at 2.5 h, initial NIHSSS = 13, TIMI = 2 at Day 1 and TIMI = 2 by the 7-day F/U. (B) Output for rt-PA-treated 52-year-old male imaged at 2.8 h, initial NIHSSS = 20, TIMI = 2 on Day 1 and TIMI = 3 by 7-day F/U. For A, the predicted lesion corresponds well with tissue that is infarcted (regions of hyperintensity) by F/U, while for B the predicted lesion is much larger than the amount of tissue that actually infarcts. Areas initially abnormal on DWI may also recover (arrowheads). Regions of the lower 95% PI correspond with acute DWI abnormality, while regions of the upper 95% PI correspond with acute MTT lesions.
correlation was reduced for the rt-PA-treated group ($P = 0.06$). The difference between the PLVs and MLVs for Group 1 ($37 \pm 35 \text{ cm}^3$) was significantly ($P = 0.01$) smaller than that measured for Group 2 ($83 \pm 55 \text{ cm}^3$).

Average GLM-predicted values (Group 1: $0.66 \pm 0.20$, Group 2: $0.72 \pm 0.11$) were significantly larger (Group 1: $P = 0.001$; Group 2: $P < 0.0001$) for tissue that infarcted than for ipsilateral tissue that did not infarct (Group 1: $0.22 \pm 0.07$; Group 2: $0.30 \pm 0.07$). For tissue that did not infarct, GLM-predicted values were significantly higher ($P = 0.009$) in rt-PA-treated patients than those for Group 1.

Table 1 summarizes differences between early complete reperfusers (TIMI $= 3$ by Day 1) and patients who did not completely reperfuse within 24 h (TIMI $\neq 3$ by Day 1). Owing to limited number of patients, data from Group 1 and 2 were combined since no significant differences were found between Groups 1 and 2 for these three parameters for each subgroup (early versus non-early). Figure 3 shows the ROC curves for the pooled results from patients demonstrating early complete reperfusion and those who did not. There was a significant difference in AUC between Groups 1 and 2 for the non-early reperfusers ($P = 0.01$) and therefore Group 1 and 2 ROC results were not combined. For the early reperfusers, sensitivity and specificity for Group 1 were 74 and 91%, respectively, whereas for the others they were 79 and 91%, respectively. For Group 2, sensitivity and specificity were 67 and 81% for the early reperfusers and 77 and 80% for the rest. Figure 5A and B show examples of predicted infarction risk for patients in both groups demonstrating early reperfusion. In the case of Fig. 5A, the patient did not receive rt-PA since he had exhibited reperfusion signs on the acute MRI and is an example where DWI $> $ PWI. For Fig. 5B, the patient demonstrated a large PWI $> $ DWI and was treated with rt-PA. The predicted risk in tissue that infarcted (yellow arrowheads) is much higher than that in tissue that was salvaged. Furthermore, the infarct area corresponds well with tissue abnormal on the lower 95% PI. For a patient who failed to reperfuse despite rt-PA treatment (Fig. 6), large GLM-predicted risk values were measured in the PLV.

Table 1 Differences between early reperfusers (TIMI $= 3$ on Day 1) and non-early reperfusers (TIMI $= 0, 1, 2$ on Day 1)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Early reperfusers ($n = 10$) (mean $\pm$ SD)</th>
<th>Non-early reperfusers ($n = 27$) (mean $\pm$ SD)</th>
<th>$P$-value (early versus non-early)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLV (cm$^3$)</td>
<td>$27 \pm 53$</td>
<td>$75 \pm 77$</td>
<td>$0.01$</td>
</tr>
<tr>
<td>PLV (cm$^3$)</td>
<td>$95 \pm 59$</td>
<td>$145 \pm 75$</td>
<td>$0.11$</td>
</tr>
<tr>
<td>GLM in PLV (a.u.)</td>
<td>$0.73 \pm 0.03$</td>
<td>$0.78 \pm 0.06$</td>
<td>$0.04$</td>
</tr>
</tbody>
</table>
Discussion
Our results show that infarction risk after stroke is spatially heterogeneous at the hyperacute stage, consistent with the notion of an ischaemic core surrounded by a penumbra of salvageable tissue (Astrup et al., 1981; Hossmann, 1994; Lo et al., 2005), probably reflecting varying degrees of tissue injury and salvageability. We observed that tissue presenting abnormal acute perfusion values that recovered compared with that which infarcted exhibited significantly lower GLM-predicted infarction risk values in agreement with findings from earlier reports (Furlan et al., 1996; Schaefer et al., 2003; Shimosegawa et al., 2005). We found that with rt-PA treatment tissue that was more severely injured and hence at greater risk of infarction, as reflected by significantly larger GLM-predicted infarction risk values, was salvaged compared with tissue that spontaneously recovered in the non-rt-PA-treated group, consistent with findings reported by others (Fiehler et al., 2002a; Butcher et al., 2005; Loh et al., 2005).

Because of intrallesional heterogeneity, effects of therapeutic intervention may not be readily detected with simple volumetric approaches. Previous attempts to use imaging as a surrogate end-point for evaluating therapeutic efficacy, on either a volumetric [Clark et al., 1999; The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study Group, 2000; Rother et al., 2002; Lees et al., 2003] or change from initial baseline infarct volume basis (Warach et al., 2000), have met with limited success. Voxel-based approaches, in contrast, may be able to more sensitively monitor treatment effects by comparing each
Brain et al. they synthesize the different patterns possible with DWI/techniques that combine multiple MRI modalities is that risk as captured in acute imaging is inherently incorporated however, was still able to detect a beneficial effect of rt-PA compared with the actual MLVs. Our voxel-based approach, thrombolysis, as supported by the significantly larger PLVs than they would have been if the patients had not received strokes initially, which led to larger lesion volumes than those found for the non-rt-PA-treated patients. Even though rt-PA-treated patients had follow-up lesion sizes on a volumetric basis comparable with the non-thrombolysed patients, it is likely that these lesion volumes were smaller than they would have been if the patients had not received thrombolysis, as supported by the significantly larger PLVs compared with the actual MLVs. Our voxel-based approach, however, was still able to detect a beneficial effect of rt-PA even in this small patient cohort since baseline infarction risk as captured in acute imaging is inherently incorporated into these models.

Perhaps one of the biggest benefits of voxel-based techniques that combine multiple MRI modalities is that they synthesize the different patterns possible with DWI/PWI mismatch within a single image (Jacobs et al., 2001; Rose et al., 2001; Wu et al., 2001). As has been previously shown by these studies and confirmed in this one, combinations of DWI and PWI can predict tissue outcome more accurately than using the parameters separately. The spatial distribution of GLM-predicted values may provide insight on tissue salvageability, leading us to speculate that in addition to using a visual mismatch between DWI/PWI for guiding therapeutic intervention (Hacke et al., 2005; Hjort et al., 2005), mismatch between tissue at high and intermediate risk may augment selecting patients most likely to respond to therapies by providing a relatively easy-to-interpret synopsis of tissue injury in a single image. Furthermore, the PIs appear to give further insight into degree of salvageability.

Although development of these algorithms can be fairly complex, involving jack-knifing and bootstrapping techniques, nonetheless, once the model coefficients have been derived, generation of infarction risk maps are relatively straightforward and can be performed feasibly within a few minutes, depending on computing platform. The steps would be as follows:

(i) Generation of parametric ADC, CBF, CBV, MTT and DELAY maps.
(ii) Co-registration of DWI and PWI parameters to one another and to the probabilistic brain atlas.
(iii) Normalization of input parameters with mean measured values in contralateral WM.
(iv) Application of Equations (1) and (2) to the normalized and co-registered images.
(v) Median filtering of the predicted risk maps.

The first step would also be necessary in simpler volumetric analysis, and tools to generate these parametric maps are available directly from the console for several MR systems. Co-registration software is also available and automatic (Collins et al., 1994) and would require only a few minutes since the acute images are already closely aligned. Because co-registration to the probabilistic brain atlas does not require high accuracy since the results are only used to generate WM normalization masks, this step can be performed rapidly. The normalization step is semi-automated, with the only user interaction consisting of the specification of contralateral hemisphere. Steps (iv) and (v) do not require user interaction and can be performed very quickly, on the order of only a few minutes.

Whether the coefficients derived in this study, or any other study, can be applied to other centres needs to be further investigated. The coefficients that we report here differ from those reported in another study (Wu et al., 2001) owing to differences in training data, which, in that earlier study, were normalized to user-defined normal grey matter; involved spin-echo and gradient-echo PWI maps; were imbalanced, that is, different amounts of infarcted and non-infarcted tissue; and involved patients imaged at later time points (<12 h). One drawback of these models is that they depend on the training data used to develop them. We eliminated some of the variability of the earlier model by switching to objectively determined and pre-defined normalization regions, only gradient-echo PWI maps and balanced training sets. However, a few persisting confounds remain, such as the extent of grey matter and WM in the lesions used in training the models. Recent studies have shown that tissue recovery and response to thrombolytic therapy may be dependent on tissue type (Bristow et al., 2005; Koga et al., 2005; Arakawa et al., 2006). We are currently investigating models that explicitly take into account tissue type (Wu et al., 2004), which would probably improve predictive performance and reproducibility.

Another factor influencing the reproducibility of the training coefficients is the homogeneity of disease aetiology and pathophysiology of the training data. We found significant differences between the jack-knifed coefficients and the aggregate model in three patients who demonstrated early partial or complete reperfusion. With increased number of patients, the influence of one patient will have less effect on the generated coefficients. The number of patients needed to stabilize risk prediction will depend on the homogeneity of the training data. The greater the
variability of the patient data used in the training cohort, the larger the number of patients will be required to reduce sensitivity of the model to addition of one patient. Clearly, 11 patients are not enough to fully characterize variations that can be seen in human stroke populations. For example, one contributing factor to our variability is probably the early time point involved in this study (<6 h), which has been previously shown to be linked to greater variability in clinical and imaging outcomes (Schellinger et al., 2001). Additional studies are needed, involving larger number of patients to determine the sample size necessary to stabilize the prediction. Ideally, the patients should be stratified according to reperfusion to increase the homogeneity of the training data. Since prospectively one does not know if a patient will reperfuse or not, one could make two sets of risk maps to predict tissue outcome under scenarios of reperfusion or no reperfusion. One could also stratify models on treatment to predict outcome assuming different treatment strategies. We are currently investigating models that take into account therapy as a factor, which we speculate will also improve prediction of tissue outcome in treated patients (Wu et al., 2005a).

Another limitation of these algorithms, as well as of other approaches relying on acute imaging, is that their prognosis is based on extent of tissue injury present at scan time. Injury due to secondary events over the course of patient care, for example, additional damage due to hypotensive events or fractionally lysed clots occluding distal arteries (Helgason, 1992; Dijkhuizen et al., 2001), would not be predictable by our data-driven models. Indeed the key to the detection of effects due to therapeutic intervention with our proposed technique is the reliance that an effective treatment alters disease progression, as we saw in the cases of patients who reperfused owing to rt-PA administration. The course of disease could also be altered naturally, such as through spontaneous reperfusion (Furlan et al., 1996; Butcher et al., 2005). However, if the number of cases with altered outcomes is significantly larger in the treatment arm compared with the placebo arm, one could conclude that the new treatment had a significant biological effect.

For patients with early complete reperfusion, we found that the follow-up infarction volumes were smaller than those for those who did not demonstrate early reperfusion, consistent with previous studies associating early reperfusion with reduced lesion volumes (Jansen et al., 1999; Parsons et al., 2002; Neumann-Haefelin et al., 2004; Butcher et al., 2005). The reduced accuracy of these models in patients who demonstrated early reperfusion is consistent with previous reports showing poorer correlation of baseline DWI and PWI lesion volumes with outcome lesion volumes after recanalization (Schellinger et al., 2001). We also found that the average GLM-predicted values for these patients at the acute stage were also smaller than those in patients who did not completely reperfuse. Since the average GLM-predicted values are significantly correlated with respect to initial NIHSS score, we speculate that the level of pre-treatment GLM-predicted infarction risk reflects severity of the initial ischaemic injury and therefore provides insight on which patients are more likely to reperfuse or will be more responsive to intervention. Our tools may therefore provide a means to evaluate potential for successful thrombolytic treatment on an individual basis before treatment. This is speculation at this point since owing to the imbalance of the two patient groups additional analysis needs to be performed on data collected as part of a randomized, placebo-controlled, double-blind clinical trial of novel therapies to validate the hypothesis that predictive algorithms can be used to reduce the number of patients needed to test the efficacy of a new treatment. One way to do this is to perform GLM analysis on randomly sampled balanced patient subsets of a clinical trial powered to detect benefit by traditional end-points and determine if similar effects can be detected using smaller sample sizes.

Furthermore, high values of predicted infarction risk using only DWI and PWI should not preclude rt-PA treatment with the current model since it is not specific for irreversible tissue injury and indeed performs poorly in cases of reversible ADC reductions. Advanced MRI techniques, such as pH-weighted MRI (Zhou et al., 2003) or MRI-measured oxygen extraction fraction (Lee et al., 2003), may provide additional information to allow better discrimination of oligemic, and therefore potentially viable tissue, from irreversibly infarcted tissue, which will not recover even with successful reperfusion. Moreover, the inclusion of non-imaging covariates that have been implicated in the success of therapeutic intervention, such as onset time to start of treatment (Hacke et al., 2004), site of arterial occlusion (Derex et al., 2004; Fiehler et al., 2005), blood glucose levels, haematocrit, age and initial NIHSS score, to name a few, into these algorithms may further improve predictive performance. Without having to increase the complexity of the produced risk maps, the approach we have presented here can readily incorporate these additional imaging and non-imaging parameters to improve predictions of tissue outcome. Studies are under way to incorporate non-imaging covariates (Wu et al., 2005b).

**Conclusion**

In conclusion, we have shown that MRI-based algorithms that predict the risk of infarction in ischaemic brain tissue can be used to assess the degree of tissue injury on a voxel-wise basis, which in turn can be used to provide greater insight into stroke pathophysiology and to improve diagnosis, prognosis and management of stroke patients. We speculate that treatment-induced tissue salvage can be detected as discrepancies between pre-treatment infarction risk predictions and actual tissue outcome. Thus, if the post-treatment measured tissue outcome is significantly improved from pre-treatment predicted destiny, it can be inferred that the altered tissue ‘fate’ was due to altered disease
progression, whether spontaneously or therapeutically induced. Our results suggest that these algorithms are promising metrics for evaluating the effects of novel treatments owing to increased sensitivity gained by comparing pre-treatment predicted outcome with post-treatment measured outcome.

Acknowledgements
This work was supported in part by grants from the Royal Dutch Academy of Arts and Sciences, the Danish Medical Research Council, Danish National Research Foundation and the German Kompetenznetzwerk Schlaganfall sponsored by the Bundesministerium für Bildung und Forschung (B5; No. 01GI9902/4). We would like to acknowledge Kim Mouridsen for statistical analysis advice and Dr Anders Rodell for data analysis assistance.

References
Spatial heterogeneity of cerebral infarction risk

Levine SR. Optimizing an individual's treatment in acute stroke: is a magnetic resonance map leading us towards the holy grail? J Neurol Sci 2004; 225: 1–2.


