Reliability of Assessing Percentage of Diffusion-Perfusion Mismatch

Shelagh B. Coutts, MBChB; Jessica E. Simon, MBChB; Anna I. Tomanek, MD; Philip A. Barber, MBChB; Jean Chan, MD; Mark E. Hudon, MD; J. Ross Mitchell, PhD; Richard Frayne, PhD; Michael Eliasziw, PhD; Alastair M. Buchan, MD; Andrew M. Demchuk, MD

Background and Purpose—Emergent neurovascular imaging holds promise in identifying new and optimum target populations for thrombolysis in stroke. Recent research has focused on patients with diffusion-weighted MRI (DWI)—perfusion-weighted MRI (PWI) mismatch as a marker of tissue at risk of infarction and a means to select the most suitable candidates for thrombolysis. The present study sought to estimate the reliability of assessing the percentage of DWI-PWI mismatch.

Methods—Thirteen patients with acute strokes had DWI and PWI within 7 hours of symptom onset. Six raters independently created relative mean transit time (rMTT) maps and then compared them with DWI images to assess the percentage of mismatch (PWI/DWI) in 10% increments. The MR scans were reassessed by 4 raters, tracing around the lesions to calculate the volume percentage of mismatch.

Results—Visual assessment had an interrater reliability of 0.68 (95% CI, 0.52 to 1.0; SEM = 21.6%) and an intrarater reliability of 0.80 (95% CI, 0.47 to 1.0; SEM = 16.9%). Hand-drawn assessment had an interrater reliability of 0.66 (95% CI, 0.45 to 1.0; SEM = 26.2%) and an intrarater reliability of 0.94 (95% CI, 0.81 to 1.0; SEM = 10.9%).

Conclusions—Results from the present study suggest that quantifying mismatch by the human eye is reproducible but not reliable among observers. This raises doubts about using mismatch for clinical decision making and clinical trial enrollment. (Stroke. 2003;34:1681-1685.)

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

Currently, the only effective treatment for ischemic stroke is intravenous tissue plasminogen activator (tPA) given within 3 hours of symptom onset.1 Unfortunately, most ischemic stroke patients do not reach the emergency department within this time window.2 Furthermore, not all patients treated within 3 hours benefit from thrombolytic therapy.3 Current research is focused on identifying imaging markers that may identify patients most likely to benefit from tPA.4 For example, identification of the ischemic penumbra4 may allow safer treatment with tPA by reducing the risk of intracerebral hemorrhage and may extend the time window for therapeutic intervention.5

Diffusion-weighted MRI (DWI) and perfusion-weighted MRI (PWI) are often used to identify salvageable tissue. DWI, which shows change in the local diffusion of water, allows visualization of early tissue damage, while PWI provides semiquantitative information on cerebral perfusion. When the acute PWI lesion volume is larger than the DWI lesion volume (referred to as mismatch), often the stroke evolves and grows.6–8 Tissue with hemodynamic change but without DWI ischemic lesions may be salvageable. Thus, mismatch has come to be used in patient selection for thrombolysis.9 Visual estimation of the 3-dimensional mismatch is difficult. The aim of the present study was to examine the interrater and intrarater reliability and variability of visually estimating the percentage of mismatch. The study parameters were chosen to mimic a clinical real-time assessment of mismatch that would be used to inform clinical decision making regarding thrombolysis.

Subjects and Methods

Patients
The present study was part of an institutionally approved protocol of emergency MRI of acute stroke patients performed from October 1999 to February 2002. Thirteen patients were selected retrospectively from the acute stroke MR database. An investigator who was...
not a rater selected study subjects. All patients were scanned within 7 hours of symptom onset, showed evidence of DWI (with restricted apparent diffusion coefficient) or PWI changes, and had optimal scan quality. Patients were chosen to reflect a variety of diffusion and perfusion lesion sizes, location of infarct, and clinical characteristics. The scans included examples of large, medium, and small mismatch (PWI>DWI), as well as examples of match (PWI=DWI) and reverse mismatch (PWI<DWI).

**Imaging Protocol**

MR images of the brain were obtained with a 3-T scanner (Signa; GE Medical Systems) equipped with high-performance gradients (40 mT/m, 184-μs rise time). All imaging was performed with the use of a standard quadrature head coil. The acute stroke imaging protocol included standard anatomic imaging (T2-weighted, fluid-attenuated inversion recovery [FLAIR], and MR angiography), DWI, and PWI. Only the latter 2 imaging sequences were evaluated in this study. DWI was performed with a single-shot, spin-echo, echo-planar imaging technique with a diffusion sensitivity of b=1000 s/mm², 7000 ms/96 ms (repetition time [TR]/echo time [TE]), 19 5-mm slices with 2-mm gap, 32×19-cm field of view, and a 192×192 acquisition matrix reconstructed to a 256×256 matrix. PWI used a single-shot, gradient-echo, echo-planar sequence with 2200 ms/25 ms. Ten 6-mm sections with a 3-mm gap, 32×19-cm field of view, and a 192×192 acquisition matrix were reconstructed to a 256×256 matrix. Five hundred ten images were collected over 112 seconds during the intravenous administration of a 20-mL bolus of gadopentetate dimeglumine (Magnevist; Berlex) injected at 5 mL/s. PWI acquisitions on our scanner were limited to collection of a maximum of 512 images. In this study we chose to collect 51 time points over 112 seconds (which necessitated a TR of 2200 ms) and to collect 10 slices at each time point. This long TR was chosen because in acute stroke the blood flow is often slow, and we need to acquire perfusion maps over the 90- to 120-second range. To increase the PWI coverage, the slice thickness and gap were increased compared with the DWI acquisition.

**Image Analysis**

The MR scans were independently assessed twice by 6 raters (2 neuroradiologists, 2 stroke neurologists, and 2 stroke fellows), blinded to all clinical information. All raters were experienced in assessing the acute changes of stroke on MRI. The DWI images were examined for evidence of restricted diffusion on a clinical workstation (Advantage Windows, General Electric Medical Systems). Relative mean transit time (rMTT) maps were created with the use of manufacturer-supplied software (Functool 2000, General Electric Medical Systems). Each rater was trained and used the same method. A lesion abnormality was considered present if there was any visible abnormality on the rMTT map when viewed on a gray scale. The DWI and PWI images were viewed on the MR workstation simultaneously. Windowing was adjusted by the rater, with no fixed levels designated. If DWI→PWI, the scan was rated as having reverse mismatch. If PWI→DWI, a matched deficit was considered present. If there was no evidence of a DWI lesion but there was a PWI lesion, then mismatch=100%. If PWI>DWI, the percentage of mismatch between DWI and the rMTT map was estimated {[rMTT volume−DWI volume]/rMTT volume}×100% to the nearest 10%. This led to an estimate of the 3-dimensional mismatch in lesion volumes. The raters repeated the aforementioned procedures after a period of at least 2 weeks to determine intrarater reliability.

Four raters manually drew lesion outlines around the DWI and rMTT lesion areas on all images twice, separated by a period of at least 1 week between each rating. We allowed displayed image contrast (window and level settings) to be varied to optimize visualization of the lesions. Lesion volumes were calculated for each image, and the percentage of mismatch was calculated, where volume=\[\text{area} \times \text{slicer number} \times (\text{slicethickness} + \text{interslicegap})\].

Estimates of interrater and intrarater reliability and SEM were calculated simultaneously from a 2-way random-effects ANOVA. One-sided, lower-limit 95% CIs were calculated about the estimates. Further analysis of reliability with the use of dichotomized cutoff points at 10% and 20% mismatch was conducted. Suggested benchmarks for reliability are as follows: slight, 0.00 to 0.20; fair, 0.21 to 0.40; moderate, 0.41 to 0.60; substantial, 0.61 to 0.80; and almost perfect, 0.81 to 1.00.

**Results**

The sample of patients included 6 women and 7 men. The median age was 65 years (range, 45 to 82 years), median time from symptom onset to scan was 152 minutes (range, 72 to 420 minutes), and median National Institutes of Health Stroke Scale score was 5 (range, 0 to 18). An example of a MR scan used in the present study is shown in the Figure.

For visual assessment representing the state of the art for real-time clinical decision making, the interrater reliability among 6 raters was 0.68 (95% CI, 0.52 to 1.0; SEM=21.6%). The intrarater reliability was 0.80 (95% CI, 0.47 to 1.0; SEM=16.9%). With the use of a dichotomized cutoff point of >10% versus ≤10% mismatch, the interrater reliability was 0.71, and intrarater reliability was 0.78. A dichotomized cutoff point of >20% versus ≤20% mismatch had lower reliability (intrarater=0.60, intrarater=0.72).

For PWI-DWI mismatch derived from volume calculations based on lesion tracing, the intrarater reliability among 4 raters was 0.66 (95% CI, 0.45 to 1.0; SEM=26.2%). The intrarater reliability was 0.94 (95% CI, 0.81 to 1.0; SEM=10.9%).

**Discussion**

Results from the present study suggest that quantifying mismatch by the human eye is reproducible but not reliable among observers. The margin of error between raters was large. For example, if one rater were to estimate the percentage mismatch as 10% and another were to estimate it at 40%, then the difference of 30% would still be within the error of the measurement (±21.6%).

The prognostic significance of mismatch in which the PWI lesion is defined by the rMTT remains unclear, but this method has made its way into clinical decision making and trial design. For example, the Desmoteplase in Acute Stroke (DIAS) trial tests the hypothesis that patients with ≥20% DWI-PWI mismatch will benefit from use of intravenous desmoteplase in a 3- to 9-hour time window. We
conclude from the results of the present study that real-time determination of percent mismatch may lead to inaccurate clinical characterization for the individual patient. Additionally, the failure to account for this error may lead to underpowered clinical trials. However, the level of intrarater reliability may be sufficient for central review of images by a single individual.

This study was performed at 3.0 T. Imaging at this field strength had a number of advantages and disadvantages compared with 1.5 T. Within the context of this study, 3.0-T imaging would intrinsically have had a higher signal-to-noise ratio, which was used to acquire higher than typical resolution DWI and PWI data. DWI contrast is independent of field strength, whereas PWI contrast in T2*-weighted imaging increases with field strength. In this study the TE at 3 T (25 ms) was reduced to compensate for the additional T2*-weighting.

Advances leading to rapid creation of quantified perfusion maps and computer-assisted volume measurements may obviate the need for visual estimates of mismatch in the future. In the interim, we would caution the use of visual inspection of mismatch as the deciding factor in treatment or trial enrollment.

Acknowledgments

Dr Barber was supported by Heart and Stroke Foundation of Canada, Canadian Institutes for Health Research, and Alberta Heritage Foundation Fellowship. Dr Buchan was supported by the Heart and Stroke Foundation of Canada, Canadian Institutes for Health Research, and the Canadian Stroke Network. Dr Coutts was supported by Heart and Stroke Foundation of Canada fellowship and Alberta Heritage Foundation for Medical Research Fellowship. Dr Demchuk was supported by the Alberta Heritage Foundation for Medical Research and Canadian Institutes for Health Research. Dr Eliasziw was supported by the Natural Sciences and Engineering Research Council of Canada. Dr Frayne was supported by the Alberta Heritage Foundation for Medical Research and the Heart and Stroke Foundation of Canada. Dr Mitchell was supported by the Alberta Heritage Foundation for Medical Research and the Multiple Sclerosis Society of Canada. Dr Simon was supported by a Canadian Stroke Network Fellowship and Alberta Heritage Foundation for Medical Research Fellowship. Dr Tomanek was supported by an Alberta Heritage Foundation for Medical Research Fellowship. The 3.0-T MR scanner used in this study was funded in part by the Canada Foundation for Innovation. The acute stroke imaging was paid for by the Alberta Foundation for Health Research. We would like to acknowledge Dr Michael D. Hill for his help in editing the manuscript.

References


Editorial Comment

Mismatch or Misconception?

The clinical utility of diffusion weighted MRI for assessing acute stroke patients has been known for some time. The combination of diffusion- and perfusion-weighted imaging (using dynamic MRI to follow an intravenous bolus injection of contrast agent) has been shown to be valuable for predicting infarct growth by defining the mismatch between the lesion volume defined by diffusion imaging and that defined by abnormal blood flow or volume (the so-called diffusion-perfusion mismatch). There have been numerous published studies, both in human patients and in various animal stroke models, supporting the use of diffusion- and perfusion-weighted MRI for characterizing acute stroke lesions and predicting lesion progression and eventual stroke outcome. As a result, diffusion/perfusion MRI has already become a part of the clinical decision-making process for acute stroke at many sites, and has also made its way into clinical trial design. With this in mind, it is rather surprising that little work has been done to measure the reliability of these techniques in the clinical setting. This article by Coutts et al addresses this issue by comparing estimates of the diffusion/perfusion mismatch, as determined by 6 individuals, in a group of acute stroke patients.
Coutts et al sound an important note of caution when interpreting such MRI data. They have shown that operator-defined measurements of the diffusion/perfusion mismatch in acute stroke patients are quite reproducible for any given individual but that interobserver reliability is rather poor. As they point out, this has serious implications when such measurements are determining the application of a potentially dangerous therapy or evaluating the efficacy of a new drug. Their study was performed using a commercially available 3T MRI scanner and imaging methods also available as standard on the system. For analysis they have used manufacturer-supplied software to calculate maps of the diffusion and perfusion parameters and used a commercially available analysis workstation for viewing and measuring them. As such, this approach is perhaps the simplest and easiest to implement, but it is also that used by many clinical sites and is the approach perhaps most likely to be adopted by clinically oriented departments wanting to begin diffusion/perfusion stroke imaging.

There are several open questions arising from Coutts et al’s study, perhaps the most important is the reason why the intraobserver reliability is poor. Although acute stroke lesions are generally bright and well defined on diffusion-weighted images, the calculated maps of relative blood flow, volume, time-to-peak, or mean transit time are often noisy, which, combined with variable and graded blood flow around stroke lesions, often makes it difficult to accurately judge the extent of the perfusion lesion. Thus different observers may well come to different conclusions when visually interpreting these images. Furthermore, in the present study there was no “standard” perfusion measurement with which to compare the MRI data. Such a standard may not really exist for humans, although several groups have used the more established PET or SPECT technology to validate MRI perfusion data. A good correlation has been shown between MRI and SPECT-derived perfusion lesions in acute stroke patients, while a good correspondence between blood flow and volume measured with MRI and PET was found in a pig stroke model. However, a study in human volunteers comparing quantitative MRI and PET perfusion in repeated measurements on the same subjects showed that PET was still much more reliable. These other studies used the more accurate arterial deconvolution approach for processing the MRI perfusion data (which still involves some human intervention). Interestingly, reliability of the MRI perfusion measurements was improved (although still inferior to PET) by removing artifacts from large vessels. The great majority of clinical PWI studies have used the gradient-echo echo-planar imaging method to follow the contrast bolus injection. This has the greatest sensitivity to the magnetic susceptibility effects of the contrast agent, but larger vessels often show up very bright. In principle, spin-echo EPI has greater sensitivity to microvascular perfusion and can avoid large vessel artifacts; however, this approach usually requires a double dose of contrast agent, and a good empirical comparison between gradient-and spin-echo PWI in stroke has, to my knowledge, yet to be published.

One additional consideration in Coutts et al’s study is their use of a 3T MRI scanner, compared with the more standard 1.5T field strength. Although this probably does not affect the reliability of their measurements, there has been some debate about the possible advantages of 3T in the acute stroke setting. One might expect, very roughly, at least a 2-fold increase in signal-to-noise (SNR) at 3T and this has driven interest in 3T for functional MRI, anatomical imaging, etc. However, a clear advantage for 3T in acute stroke MRI has yet to be demonstrated due to several confounding factors, including increased tissue T1 (reducing SNR for PWI), increased susceptibility artifacts in EPI, and increased RF power deposition (SAR).

Although improved perfusion MRI data processing algorithms and new MR contrast agents can yield better perfusion images, advances in automated or semi-automated definition of the lesion, and approaches combining all available MRI data will allow a more accurate and reproducible definition of the extent of the stroke lesion and estimates of the eventual stroke outcome. However, in the meantime, while we wait for advances in imaging methodology and automated lesion classification to make their way into commercially available (and clinically useful) tools, perhaps the best approach is to use a small group of carefully trained individuals for measuring all the perfusion and diffusion maps at a given site, or for a given trial, and to consider constraining the viewing conditions to reduce interobserver variations.

In short, the message from this study is to be cautious when using diffusion/perfusion data to guide acute stroke therapy. In my view, this should be combined with optimism that various methodological improvements could significantly increase the reliability of this MRI approach. Most recently it was shown, using 2 human operators to manually define lesions on perfusion diffusion and FLAIR images, that acute PWI data could accurately differentiate core and “penumbral” brain tissue prior to recanalization. This again demonstrates the power of perfusion/diffusion MRI and its potential to eventually become perhaps the most valuable determinant of acute stroke therapy options.

Alex de Crespigny, PhD, Guest Editor
Department of Radiology
Massachusetts General Hospital
Harvard Medical School
Charlestown, Massachusetts

References


Reliability of Assessing Percentage of Diffusion-Perfusion Mismatch
Shelagh B. Coutts, Jessica E. Simon, Anna I. Tomanek, Philip A. Barber, Jean Chan, Mark E. Hudon, J. Ross Mitchell, Richard Frayne, Michael Eliasziw, Alastair M. Buchan and Andrew M. Demchuk

*Stroke*. 2003;34:1681-1683; originally published online June 12, 2003;
doi: 10.1161/01.STR.0000078840.96473.20

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2003 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/34/7/1681

**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/