Rapid Assessment of Perfusion–Diffusion Mismatch

Ken Butcher, MD, PhD; Mark Parsons, PhD, FRACP; Louise Allport, FRACP; Sang Bong Lee, PhD; P. Alan Barber, PhD, ZRACP; Brian Tress, FRACR; Geoffrey A. Donnan, MD, FRACP; Stephen M. Davis, MD, FRACP; for the EPITHET Investigators

Background and Purpose—For MR perfusion–diffusion (PWI-DWI) mismatch to become routine in thrombolysis patient selection, rapid and reliable assessment tools are required. We examined interrater variability in PWI/DWI volume measurements and developed a rapid assessment tool based on the Alberta Stroke Program Early CT Scores (ASPECTS) system.

Methods—DWI and PWI were performed in 35 patients with stroke <6 hours after symptom onset. DWI lesion and PWI (time to peak) volumes were measured with planimetric techniques by 4 raters and the 95% limits of agreement calculated. ASPECT scores were assessed separately by 4 investigators (2 experienced and 2 inexperienced) for DWI (MR DWI scores) and PWI (MR time to peak scores). MR mismatch scores were calculated as MR DWI-MR time to peak scores.

Results—Interobserver variability was much greater for PWI (95% limit of agreement = ±72.3 mL) than for DWI (95% limit of agreement = ±12.6 mL). A semiautomated PWI volume (time to peak = 2 s) was therefore used to calculate mismatch volume. MR mismatch scores ≥2 predicted 20% PWI-DWI mismatch by volume with mean 78% sensitivity (range, 72% to 84%) and 88% specificity (range, 83% to 90%). There was excellent agreement on mismatch classification using MR mismatch scores between experienced raters (weighted kappa scores of 0.94) with agreement in 34 of 35 cases. Agreement was less consistent between inexperienced raters (weighted kappa = 0.49, 28 of 35 cases).

Conclusions—Variability in planimetric mismatch measurements arises primarily from differences in PWI volume assessment. High specificity and interrater reliability may make MR mismatch scores an ideal rapid screening tool for potential thrombolysis patients. (Stroke. 2008;39:75-81.)

Key Words: brain imaging ■ cerebral blood flow ■ cerebral infarct ■ diffusion-weighted imaging ■ perfusion-weighed imaging

Mismatch between a larger perfusion-weighted imaging (PWI) abnormality and a smaller diffusion-weighted (DWI) lesion has been postulated to represent the ischemic penumbra.1,2 This tissue is at risk for infarction, but may also be potentially amenable to salvage with thrombolysis. Although the mismatch hypothesis remains unproven, it is being used with increasing frequency in acute stroke studies and even clinical practice.3,4 If PWI-DWI mismatch is to be incorporated into the routine selection process for thrombolysis, rapid assessment tools with good interrater reliability are required. There is presently no standardized method for rapid assessment of PWI-DWI mismatch.

It has previously been demonstrated that purely subjective assessments of PWI-DWI mismatch have poor interrater reliability.5 Conversely, interrater agreement of DWI volume has been shown to be excellent.6 We therefore hypothesized that the source of disagreement between raters assessing mismatch is related to interpretation of the perfusion images. Although planimetric volume measurement is accurate, it does require time intensive operator input, which may delay acute stroke therapy. Clinicians generally rely on qualitative assessments of images, sometimes in conjunction with rating systems. The Alberta Stroke Program Early CT Score (ASPECTS) is a validated semiquantitative scale useful for assessing the extent of ischemic changes within the middle cerebral artery (MCA) territory.7,8 This is a negative ordinal scale in which normal-appearing brains are scored as “10” and those with ischemic changes involving the entire MCA territory are rated “0.” The ASPECTS system has also been successfully applied to perfusion CT and MRI.9,10 By applying ASPECTS to DWI and PWI sequences, we have developed a novel tool known as the MR mismatch score.

Received April 5, 2007; accepted May 1, 2007.
From the Department of Neurology (K.B., L.A., S.D.) and Radiology (B.T.), Royal Melbourne Hospital, University of Melbourne, Melbourne Australia; the Department of Neurology (K.B.), University of Alberta, Edmonton, Alberta, Canada; the Department of Neurology (S.B.L.), Catholic University of Korea, Seoul, South Korea; the Department of Neurology (P.A.B.), Auckland City Hospital, Auckland, New Zealand; the Department of Neurology (M.P.), John Hunter Hospital, Newcastle, Australia; and the Department of Neurology (G.D.), Austin Hospital, Melbourne, Australia.
Correspondence to Ken Butcher, MD, PhD, 2E3.13 WMC Health Sciences Centre, University of Alberta, Edmonton, Alberta, Canada. E-mail ken.b FTC55@ualberta.ca
© 2007 American Heart Association, Inc.
Stroke is available at http://stroke.ahajournals.org
DOI: 10.1161/STROKEAHA.107.490524
We had 2 aims in this study, the first of which was to identify the sources of error in mismatch assessment and develop a solution. The second was to apply ASPECTS scores to PWI and DWI and determine whether this novel tool could be used to identify patients with tissue at risk for infarction.

Methods

Patients

Forty patients with acute ischemic stroke were prospectively recruited from 9 centers participating in the Echoplanar Imaging Thrombolysis Evaluation Trial (EPITHET) and imaged with MRI 3 to 6 hours after symptom onset. Evaluation was restricted to the 35 of the 40 patients who had MCA territory infarction. Patients with anterior cerebral artery (n=1) and posterior circulation (n=3) infarcts were excluded because ASPECTS was designed for assessment of MCA territory infarcts. In addition, one patient with technically inadequate PWI data was also excluded. Informed consent was obtained from the patient/next of kin and local human research committees approved the protocol.

Imaging Protocol

Noncontrast CT scans were obtained before MRI. Patients with intracerebral hemorrhage or ischemic changes more than one third of the MCA territory were excluded as per the EPITHET protocol. MRI scans were obtained with 1.5-T EPI-equipped scanners (GE Signa/Siemens Vision/Symphony/Philips Intera). Perfusion-weighted images were obtained using a bolus of gadolinium diethylene triamine penta-acetic acid (0.2 mmol/kg), injected at 5 mL/s followed by 15 mL of saline. Twelve to 16 slices (32 to 50 time points) were obtained. Slice thickness was 5 to 6 mm ±1-mm gap, matrix sizes were 128×128×256×256, and field of view =40×40 cm. Diffusion-weighted images were obtained with single-shot spin-echo EPI sequences. Sixteen to 20 slices 5 to 6 mm ±1-mm gap were obtained. Matrix size was 128×128×256×256, field of view =40×40 cm, and TR/TE 6000/107 ms. Diffusion gradient strength was varied between 0 and 22 mT/m, resulting in b values of 0, 500, and 1000 s/mm.

Data Analysis

Postprocessing of raw perfusion images was performed centrally by a single investigator using the software package StrokeTool (DIS, Düsseldorf, Germany). This software was used to plot the change in MRI transverse relaxivity, which is linearly related to gadolinium diethylene triamine penta-acetic acid concentration, on a per-voxel basis over time. Time to peak of the impulse response curve (Tmax) maps was calculated using single value decomposition. This technique allows the impulse response curve to be calculated as a deconvolution of the raw perfusion images using an arterial input function. The arterial input function was selected from the MCA territory infarcts. In addition, one patient with technically inadequate PWI data was also excluded. Informed consent was obtained from the patient/next of kin and local human research committees approved the protocol.

Regional PWI and DWI image analysis was performed using the Analyze software package (Biomedical Imaging Resource, Rochester, NY). Isotropic DWI hyperintense regions were outlined visually by 2 stroke neurologists and 2 stroke fellows. The 2 stroke neurologists were considered experienced on the basis that each had >4 years performing volumetric analysis. The 2 stroke fellows had less than 6 months’ experience and were considered to be inexperienced raters. Investigators were free to vary intensity window level and width settings. Tmax volumes were outlined in the same manner. Investigators individually calculated mean Tmax in contralateral homologous regions. The latter were mirror images of the ipsilateral regions of interest reflected on a 180° axis. Each investigator then applied a threshold to Tmax maps based on mean contralateral values +2 s and calculated a second volume. Mismatch was calculated as the difference in volumes between each observer’s Tmax, Tmax +2 s, and DWI measurements. Each investigator required approximately 20 minutes to complete all planimetric measurements for each patient.

A semiautomated Tmax volume was calculated separately using a +2-s threshold relative to the start of the impulse response. This Tmax +2 volume was used as the objective reference to which all semiquantitative mismatch assessments were compared. Standardized mismatch volumes were calculated as the difference between this semiautomated Tmax +2 s volume and the mean DWI volume measurement for all raters. A mismatch pattern was considered to be present if the standardized mismatch volume exceeded DWI volume by at least 20%.

MR ASPECT Scores

ASPECT scores of DWI (MR DWI scores) and Tmax (MR Tmax scores) images were then recorded independently >2 weeks after the planimetric volume measurements. The 2 ASPECTS slices at and immediately superior to the basal ganglia were first identified on the structural T2-weighted images on which DWI and PWI sequences are based. MR DWI and MR Tmax scores were then rated by each investigator. Hyperintensity on DWI or Tmax prolongation within an ASPECTS region resulted in a deduction of 1 point on each score. MR mismatch scores were then calculated by subtracting MR Tmax scores from MR DWI scores.

Statistical Analysis

Statistical analysis was performed using Stata (Statacorp). The 95% limits of agreement (mean difference±2 SDs) for DWI and PWI regions of interest were calculated and interrater differences illustrated using Bland-Altman plots. The ability of MR mismatch scores to predict PWI-DWI volume mismatch was assessed with receiver-operator characteristic curves for each rater. Interrater receiver-operator characteristic differences were tested using a χ² test of the area under each curve. Interrater reliability of MR DWI scores, MR Tmax scores and MR mismatch scores was assessed with a weighted kappa analysis. Kappa scores were weighted to penalize differences of >1 as described previously.

Results

The 35 patients (27 men; median age, 73 years; range, 39 to 87 years) were imaged with MRI at a median of 4.5 hours (range, 2.7 to 5.6 hours) after symptom onset. Median acute National Institutes of Health Stroke Scale score was 11 (range, 4 to 25).

Planimetric Diffusion-Weighted Imaging, Perfusion-Weighted Imaging, and Mismatch Volumes

Interrater differences in DWI lesion volume measurements for each patient assessed by the 4 investigators are illustrated in Figure 1. The reference DWI volume of each individual patient was the mean measurement of all 4 observers as previously described. The mean DWI reference volume of the entire sample of 35 patients was 51.5±52.4 mL. The mean difference between individual raters and the reference volumes ranged from a minimum of −1.8 mL to a maximum of +2.4 mL (Figure 1). The 95% limits of agreement, for absolute volumes, between all 4 observers were ±12.6 mL. Experienced observers measured slightly larger volumes on average relative to the inexperienced observers, but overall differences were not significant (Figure 1). Thus, DWI lesion volumes varied very little between observers.

Interrater disagreement was much greater for PWI measures. The mean Tmax abnormality reference volume of the sample was 163.4±87.2 mL. The mean difference in Tmax volumes, relative to the reference value, ranged from −47.1
to +22.9 mL and the 95% limits of agreement were ±72.3 mL (Figure 1). Inexperienced observers tended to draw smaller regions of interest on average. Mismatch volumes also varied widely between observers. The mean mismatch volume was 111.8 ± 81.5 mL. The mean difference in mismatch volumes, relative to the reference value, was similar to that of the Tmax volume differences, ranging from -45.4 to +20.5 mL and the 95% limits of agreement were ±70.2 mL (Figure 1).

Perfusion-Weighted Imaging Threshold Application Effects

Application of a PWI threshold, relative to the contralateral hemisphere, was associated with a substantial decrease in interrater variability of perfusion deficit volume measurements. The mean Tmax +2 s abnormality volume was 93.8 ± 62.8 mL. The mean difference in Tmax +2 s measured volumes, relative to the reference semiautomated value, ranged from -15.1 to +15.7 mL and the 95% limits of agreement narrowed to ±31.5 mL (Figure 2). The semiautomated reference Tmax +2 s volume was very similar to the mean volume calculated by all 4 raters. Calculation of mismatch using Tmax +2 s volumes was also associated with reduced interobserver variability. The mean mismatch volume was 42.2 ± 60.7 mL. The mean difference in mismatch volume, relative to the reference value, ranged from -13.4 to +14.7 mL and the 95% limits of agreement were ±33.1 mL.

Figure 1. Plots of raw measurements (top) and Bland-Altman plots (bottom) illustrate interrater differences in planimetric volume assessments of DWI lesions, PWI (Tmax) deficits, and mismatch volumes (calculated as Tmax-DWI volume). Bland-Altman plots indicate the difference between each individual rater, for each patient, and the reference volume. The reference volume is calculated as the mean volume for all 4 raters. Solid horizontal lines represent the maximum and minimum mean differences between individual raters and the reference volumes. The 95% limits of agreement, for absolute volumes, between all 4 observers are represented by the dashed horizontal lines. The plots indicate that the source of interrater disagreement with respect to mismatch volume is largely related to PWI measurements. Experienced rater volumes are indicated by circles and inexperienced rater measurements with triangles.

Figure 2. Demonstration of the effect of threshold application to PWI volumes. Areas of disagreement were most commonly in the periphery of the PWI deficit, where Tmax delay is intermediate and heterogeneous (arrows). These areas of disagreement were largely removed after application of the +2 s threshold to the Tmax maps (bottom). The raw volumes of all 4 raters are much more similar and the Bland-Altman plots confirm improved 95% limits of agreement (see Figure 1). The blue line represents an automatic threshold of +2 s based on the arterial input function. This standard volume was used to calculate a standardized mismatch volume.
MR Mismatch Scores

Examples of MR mismatch score assessment are shown in Figure 3. The median MR DWI scores, MR Tmax scores, and MR mismatch scores were 7, 3, and 2, respectively. MR DWI scores were inversely correlated with planimetric DWI volumes ($\rho = -0.75$, $P<0.001$; Figure 4). MR Tmax scores were also inversely correlated with planimetric Tmax + 2 s volumes ($\rho = -0.64$, $P<0.001$; Figure 4). MR mismatch scores (MR DWI-MR Tmax) correlated with planimetric mismatch, calculated as $\text{Tmax} + 2 \text{s} - \text{DWI volume}$ ($\rho = 0.67$, $P<0.001$; Figure 4).

A total of 26 patients had a standardized mismatch pattern. The ability of the MR mismatch score to predict this definition of mismatch is illustrated with the receiver-operator characteristic curve in Figure 5. Receiver-operator characteristic curve analysis indicated an MR mismatch score of $\geq 2$ provided optimal sensitivity and specificity for prediction of mismatch by volume. An MR mismatch score of $\geq 2$ predicted $>20\%$ mismatch by volume with a mean sensitivity of 78\% (interrater range 72\% to 84\%) and specificity of 88\% (interrater range 83\% to 90\%). The mean correct classification rate was 83\% (interrater range 77\% to 90\%). Although the area under the receiver-operator characteristic curve of one of the inexperienced raters was slightly smaller than the other investigators, the differences were not significant ($\chi^2 = 4.68$, $P = 0.20$). The optimal cut point MR mismatch score of $\geq 2$ was the same for all 4 users regardless of experience (Figure 5).

Weighted kappa scores indicated excellent interrater agreement between the experienced users for MR DWI, MR Tmax, and MR mismatch scores (Table). Inexperienced raters, however, had only a fair interrater agreement rate. Experienced raters agreed on mismatch classification, using MR mismatch scores, in 34 of 35 cases (weighted kappa=0.94). The agreement rate decreased to 28 of 35 cases between inexperienced raters (weighted kappa=0.49).

Discussion

Although the mismatch hypothesis has not yet been tested conclusively, penumbral selection is being used to select patients for acute thrombolytic therapy.3,4 Rapid and reliable assessment of PWI-DWI mismatch is therefore an important goal. In this study, we have shown that ASPECTS scores of PWI and DWI data can be used to derive a semiquantitative ordinal scale, which predicts PWI-DWI mismatch by volume with high sensitivity and specificity. This MR mismatch score may be calculated without the time-intensive investigator input required for planimetric analysis so that access to acute stroke therapies need not be delayed. This study has also confirmed that interobserver variability of mismatch volume assessment is significant, even when planimetric measurement techniques are used. The source of disagreement is primarily related to interpretation of the PWI data. Agreement is improved by application of a perfusion threshold, which serves to standardize the data in an objective manner.

Planimetric Assessment of Mismatch

Visual delineation of DWI lesions and PWI deficits is commonly used in acute stroke MRI research studies. We have shown that planimetric DWI volumes are quite consistent between observers, even those with relatively little experience. This is consistent with previous reports that interobserver variation of DWI volumes is $<5\%$.16–18 A recent systematic evaluation indicates that the mean absolute difference in DWI planimetric volume measurements made by 2 raters is $2.4\pm 4.7$ mL.19 Conversely, these authors observed larger interrater differences in measured PWI volumes similar to our own ($19.4\pm 34.6$ mL).19 The borders of perfusion deficits are subject to greater disagreement primarily due to the fact that the periphery of the PWI abnormality often contains tissue with heterogeneous oligemia, much of which is unlikely to be at risk of infarction (Figure 2).1 This can lead to significant variation in measured volumes and therefore mismatch assessment.

Despite interobserver disagreement, planimetric techniques are the most accurate method available for measuring volumes and should remain the standard used in reporting the results of MRI-based research studies. In the absence of validated quantitative perfusion measures, we suggest that a minimal threshold, such as $+2\text{s}$, should be used to standardize all PWI time domain maps. Tmax $+2\text{s}$ was used as a standardized PWI measure in a recent observational study of...
MRI profiles and thrombolysis. This parameter will also be used to define mismatch in the primary EPITHET analysis.

Semiquantitative Assessment of Mismatch: MR Mismatch Scores

Planimetric measurement techniques are presently reserved for in-depth “offline” analyses. Fully automated planimetric assessment tools may one day allow accurate volume calculations in the hyperacute setting, but this is not possible with standard clinical MRI software currently in use. Instead, most clinicians and diagnosticians make qualitative assessments of lesion size, which are potentially prone to error. It has been shown that semiquantitative assessment scales can improve recognition of significant patterns in acute stroke imaging.

Our findings indicated that the MR mismatch scoring system predicted mismatch by volume with greater specificity than sensitivity. As many as 28% of patients with mismatch by volume (<20%) were assessed as having a nonmismatch pattern by raters using MR mismatch scores. This may not represent a disadvantage of using an ASPECTS-based system to assess acute stroke MRI images.

| Table. Weighted Kappa Scores of Experienced and Inexperienced Raters for MR DWI, MR Tmax, and MR Mismatch Scores |
|-----------------|----------|-----------|---------------|
| Raters          | Score    | Weighted Kappa | Agreement   |
| Experienced     | MR DWI   | 0.95       | Excellent    |
|                 | MR Tmax  | 0.98       | Excellent    |
|                 | MR Mismatch | 0.94     | Excellent    |
| Inexperienced   | MR DWI   | 0.74       | Good         |
|                 | MR Tmax  | 0.67       | Good         |
|                 | MR Mismatch | 0.49     | Fair         |
On the contrary, because the optimal criteria for definition of significant mismatch have not been established, it is possible that the qualitative score provides a more accurate estimation of salvageable tissue than the arbitrary 20% volumetric definition. An MR mismatch score of 2 indicates that 2 MCA regions are hypoperfused, but not yet compromised. Although the planimetric measurements indicated that some patients had at least 20% mismatch by volume in the presence of MR mismatch scores of less than 2, these may not be ideal thrombolysis candidates, because the hypoperfused regions already had evidence of some tissue compromise. This is also consistent with another proposed advantage of the ASPECTS system, specifically that functionally important subcortical regions with smaller volumes are given the same weight as the larger cortical regions.8,22 It must be emphasized that the true significance of a 20% mismatch has not yet been established. Accordingly, although we have shown an MR mismatch score of ≥2 predicts 20% mismatch, this does not in itself imply these represent the optimal thrombolysis candidates.

The primary advantage of applying semiquantitative ordinal scales to stroke image analysis is improved interrater agreement over subjective binary assessments. ASPECT scores have previously been shown to standardize acute stroke noncontrast CT assessment7,15 and have been applied to CT perfusion and CT angiographic source images with good interrater reliability.10,23 The present investigation indicates that semiquantitative assessments of mismatch can reliably be made by different observers, but, like with CT, this improves with experience.24,25 Nonetheless, variability between our inexperienced raters was still superior to a previous report of purely qualitative mismatch assessments.5

In contrast to the relatively large interobserver variability of planimetric Tmax measurements, interrater agreement appears to be very similar for MR Tmax and MR DWI scores. This likely reflects the fact that a regional analysis does not necessitate absolute agreement. The source of disagreement between investigators is generally at the periphery of the PWI deficit; however, in the majority of cases, these areas are smaller than an entire ASPECTS region. Thus, although investigators will measure different planimetric volumes, they may often record the same MR mismatch scores.

The chief limitation of this study is the use of the same investigators in the volumetric and MR mismatch scores portions of the investigation. Furthermore, volume measurements and MR mismatch scoring were completed in a sequential and nonrandomized fashion. We attempted to minimize the effect of prior experience with a substantial time interval between assessments as has been reported in previous studies.5 In addition, we have applied the mismatch scoring system to only one set of PWI maps (Tmax). It remains to be determined if the system is as effective with other PWI parameters, including time to peak, mean transit time, and relative CBF maps. Finally, this study lacks outcome data to assess the ability of MR mismatch scores to predict final infarction. This will be performed at the completion of the EPITHET study, an ongoing randomized, controlled trial of tissue plasminogen activator versus placebo in the 3- to 6-hour time window.

Sources of Funding

The Investigator-led EPITHET trial is supported by the National Health and Medical Research Council of Australia. Study drugs were provided by Boehringer Ingelheim.

Disclosures

None.

References

17. Ritzl A, Meisel S, Wittsack HJ, Fink GR,Siebler M, Modder U, Seitz RJ. Development of brain infarct volume as assessed by magnetic resonance


Rapid Assessment of Perfusion–Diffusion Mismatch
Ken Butcher, Mark Parsons, Louise Allport, Sang Bong Lee, P. Alan Barber, Brian Tress, Geoffrey A. Donnan and Stephen M. Davis
for the EPITHET Investigators

Stroke. 2008;39:75-81; originally published online December 6, 2007;
doi: 10.1161/STROKEAHA.107.490524
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 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/39/1/75

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/