NIH Public Access
Author Manuscript
Radiology. Author manuscript; available in PMC 2007 April 11.

Published in final edited form as:
Radiology. 2005 March ; 234(3): 909–916.

Human Brain: Reliability and Reproducibility of Pulsed Arterial Spin-labeling Perfusion MR Imaging[1]

Geon-Ho Jahng, PhD, **Enmin Song, PhD**, **Xiao-Ping Zhu, MD, PhD**, **Gerald B. Matson, PhD**, **Michael W. Weiner, MD**, and **Norbert Schuff, PhD**
From the Departments of Radiology (G.H.J., E.S., X.P.Z., M.W.W., N.S.), Pharmaceutical Chemistry (G.B.M.), Neurology (M.W.W.), Medicine (M.W.W.), and Psychiatry (M.W.W.), University of California San Francisco, 4150 Clement St, Rm 114M, San Francisco, CA 94121.

Abstract

The Committee of Human Research of the University of California San Francisco approved this study, and all volunteers provided written informed consent. The goal of this study was to prospectively determine the global and regional reliability and reproducibility of noninvasive brain perfusion measurements obtained with different pulsed arterial spin-labeling (ASL) magnetic resonance (MR) imaging methods and to determine the extent to which within-subject variability and random noise limit reliability and reproducibility. Thirteen healthy volunteers were examined twice within 2 hours. The pulsed ASL methods compared in this study differ mainly with regard to magnetization transfer and eddy current effects. There were two main results: *(a)* Pulsed ASL MR imaging consistently had high measurement reliability (intraclass correlation coefficients greater than 0.75) and reproducibility (coefficients of variation less than 8.5%), and *(b)* random noise rather than within-subject variability limited reliability and reproducibility. It was concluded that low signal-to-noise ratios substantially limit the reliability and reproducibility of perfusion measurements.

Arterial spin-labeling (ASL) magnetic resonance (MR) imaging (1), at which blood water is magnetically labeled as an endogenous tracer of perfusion, is being used in studies of cerebral blood flow, in part because ASL is an entirely noninvasive technique that does not involve exposure to ionizing radiation or radioactive isotopes and thus facilitates improved patient safety. The results of comparative studies in which positron emission tomography (PET) or single photon emission computed tomography (SPECT) was performed in volunteers and patients have also demonstrated the accuracy of ASL MR imaging for measuring cerebral blood flow (2,3).

In contrast to volumes of brain structures, which have extremely little variability during short time intervals at structural MR imaging (4), rates of cerebral blood flow and perfusion can fluctuate considerably, depending on the rate of brain activity. For example, PET measurements of cerebral blood flow have varied owing to involuntary eye movements (5) and stimulation (6) and other stimuli that alter brain activity. Thus, physiologic variability in

[1]Supported in part by NIH/NIA grant AG012435, NIH/NIAAA grant AA11493, and a Research Enhancement Award Program (REAP) fund of the Department of Veterans Affairs.

Address correspondence to N.S. (e-mail: nschuff@mail.ucsf.edu)..

Author contributions:
Guarantors of integrity of entire study, G.H.J., N.S.; study concepts, M.W.W., N.S.; study design, G.H.J., M.W.W., N.S.; literature research, G.H.J., E.S., X.P.Z., G.B.M., N.S.; clinical studies, M.W.W.; data acquisition, G.H.J., G.B.M., E.S.; data analysis/interpretation, G.H.J., E.S., X.P.Z., M.W.W., N.S.; statistical analysis, G.H.J., N.S.; manuscript preparation, G.H.J., N.S.; manuscript definition of intellectual content, G.H.J., M.W.W., N.S.; manuscript editing, revision/review, and final version approval, all authors

Authors stated no financial relationship to disclose.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cerebral blood flow and perfusion may cause reduced measurement reliability— herein defined as the ability to consistently detect the same difference between subjects—and reduced reproducibility—herein defined as the ability to repeatedly detect the same value within subjects.

In addition to physiologic fluctuations, errors related to tissue magnetization transfer effects (7) and technical problems with transient magnetic fields (eddy currents) (8), which together with random noise may further limit reliability and reproducibility, also confound ASL MR imaging measurements. Recently, we proposed a scheme for performing pulsed ASL MR imaging, double inversion with proximal labeling of both tagging and control images (DIPLOMA) (9), to facilitate improved magnetization transfer and eddy current compensation without sacrificing labeling efficiency. The improvements that are possible with DIPLOMA were demonstrated to be comparable to those achievable with other often-used ASL MR imaging methods—namely, echo-planar imaging and signal targeting with alternating radiofrequency (EPISTAR) (10) and proximal inversion with a control for off-resonance effects (PICORE) (11).

The purposes of the study were to prospectively determine the global and regional reliability and reproducibility of three ASL MR imaging methods and to determine the extent to which within-subject variability and random noise limit measurement reliability and reproducibility.

Materials and Methods

Volunteers and Image Acquisitions

The Committee of Human Research of the University of California San Francisco approved this study, and all volunteers provided written informed consent before participating. Thirteen volunteers (nine women, four men; mean age, 45 years ± 14 [standard deviation]; age range, 29 – 64 years) were recruited from the community within a 3-month period, and after undergoing a screening interview to determine MR imaging suitability, they were enrolled in this study.

Brain perfusion was measured with the three pulsed ASL methods—DIPLOMA, PICORE, and EPISTAR–by using a 1.5-T MR imaging system (Siemens Vision; Siemens, Erlangen, Germany). To compare reliability and reproducibility between the different perfusion measurement methods, each subject was examined twice, during a test and a retest, with each pulsed ASL MR imaging method and with parameters optimized for each method. In each subject, the test-retest examinations were repeated within 2 hours; the patient table was moved and the instrumental parameters were readjusted between each examination.

For the first experiment—the test part of the test-retest—magnetization-prepared rapid gradient-echo scout anatomic imaging; fast low-angle shot imaging; and EPISTAR, DIPLOMA, and PICORE ASL perfusion-weighted imaging were performed. For the second experiment—the retest portion—after the patient table was moved out and while the patient rested on the table, the described imaging sequences were repeated in the same order in which they were performed in the first experiment. The patient was asked to rest on the table between the test and retest experiments while the table was moved in and out of the magnet. The 13 subjects were the same subjects examined in a previous study (9). Having imaged 13 subjects twice, we estimated that we would detect an 11% difference between the test and retest perfusion measurements at 95% confidence and with 80% power.

Details about pulsed ASL MR imaging sequences and acquisition parameters have been published previously (9); only a brief summary is given here. The PICORE, EPISTAR, and DIPLOMA pulsed ASL labeling methods, each involving the use of adiabatic hyperbolic secant

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

pulses to label blood water within a slab in the brain, are illustrated in Figure 1. The major differences between these three ASL methods are related to the compensation of effects from magnetization transfer and eddy currents. Because the pulse amplitudes for control (nonlabeled) and labeled ASL imaging are the same with the DIPLOMA method, in contrast to the different pulse amplitudes with the EPISTAR method, magnetization transfer effects are better compensated with DIPLOMA. Furthermore, because both control and labeled imaging involve the use of slab-selective gradients in DIPLOMA, in contrast to the gradients used in PICORE, eddy current compensation is better with DIPLOMA.

After pulsed ASL, the bolus of labeled blood water was sharpened by applying a series of short saturation pulses to the distal edge of the slab, as previously proposed for the Q2TIPS (quantitative imaging of perfusion using a single subtraction second version with thin-section periodic saturation after inversion time TI1) method (QUIPSS II with thin-section TI1 periodic saturation) (12). Five MR image sections were acquired by using 2500/15/1500 (repetition time msec/echo time msec/inversion time msec) with single-shot echo-planar imaging during alternating labeled and nonlabeled (control) ASL periods to derive per-fusion-weighted imaging data by subtracting the labeled from the nonlabeled imaging data (4.6 × 2.3-mm in-plane spatial resolution, five 8-mm-thick trans-verse oblique sections, 2-mm separation, and oriented 10° off the anteroposterior commissure line).

In addition to perfusion-weighted images, high-spatial-resolution volumetric T1-weighted magnetization-prepared rapid acquisition gradient-echo (13) (10/4/300, 1 × 1-mm in-plane spatial resolution, 1.5-mm-thick sections, 15° flip angle) and multiplanar intermediate-weighted fast low-angle shot (14) (195/6 [repetition time msec/echo time msec], 60° flip angle, 1.17 × 1.17-mm in-plane spatial resolution, 8-mm-thick sections, oriented 10° off the anteroposterior commissure line but covering the whole brain) images of the whole brain were acquired for image registration and normalization into a standardized space within the framework of the SPM99 statistical parametric mapping program (Wellcome Department of Cognitive Neurology, London, England).

Image Processing

The magnetization-prepared rapid acquisition gradient-echo, fast low-angle shot, labeled and nonlabeled echo-planar, and perfusion-weighted imaging data were transferred to off-line computer stations for further processing with SPM99. Perfusion-weighted image signals were corrected for variations in transmitter voltage and receiver gain to account for instrumental differences between per-fusion measurements.

Spatial normalization—To transform the perfusion-weighted imaging data of each subject into the standard brain template of SPM99, first, during a normalization step, T1-weighted magnetization-prepared rapid acquisition gradient-echo imaging data obtained during the test and retest acquisitions were coregistered with each other and averaged to eliminate potential bias toward systematic differences between the test or retest acquisitions. The direct registration between echo-planar imaging– based perfusion-weighted data and magnetization-prepared rapid acquisition gradient-echo data was not reliable because the geometric and signal intensity distortions at echo-planar imaging were greater than those at magnetization-prepared rapid acquisition gradient-echo imaging.

In contrast to the echo-planar images, the fast low-angle shot images had no visible distortions but the same spatial resolution. Thus, fast low-angle shot images that appeared intermediate between magnetization-prepared rapid acquisition gradient-echo images and echo-planar images in terms of geometric and signal intensity distortions were acquired. Therefore, to register the perfusion-weighted imaging data to the magnetization-prepared rapid acquisition gradient-echo images, a stepwise registration was used: First, echo-planar imaging data

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(reference data obtained at perfusion-weighted imaging) were registered to fast low-angle shot data. Then, fast low-angle shot data were registered to magnetization-prepared rapid acquisition gradient-echo data, effectively resulting in a coregistration of echo-planar imaging and perfusion-weighted imaging data with magnetization-prepared rapid acquisition gradient-echo images. The averaged magnetization-prepared rapid acquisition gradient-echo imaging data were then mapped onto the standard T1-weighted brain template image (Montreal Neurological Institute, Montreal, Quebec, Canada) of SPM99 by using 12 nonlinear parametric transformations. This process resulted in spatially normalized magnetization-prepared rapid acquisition gradient-echo images with 2 × 2 × 2-mm spatial resolution.

Finally, the echo-planar and perfusion-weighted imaging data were interpolated to the same spatial resolution as the magnetization-prepared rapid acquisition gradient-echo images and mapped into the standard T1 brain template space by using the same transformations that were determined for the magnetization-prepared rapid acquisition gradient-echo images.

Tissue segmentation—To differentiate between gray matter perfusion and white matter perfusion and to account for partial volumes of brain tissue and cerebro-spinal fluid in the perfusion-weighted imaging data, the normalized magnetization-prepared rapid acquisition gradient-echo images were first classified into probabilistic maps of cerebrospinal fluid, gray matter, and white matter within the framework of SPM99, which uses both image signal intensity and anatomic information to derive tissue classifications. Partial volume effects in perfusion-weighted imaging data were accounted for by filtering the coregistered perfusion-weighted imaging data with the probabilistic brain tissue masks.

For gray matter perfusion measurement, perfusion-weighted imaging data were multiplied by the mask image data, including data for at least 75% gray matter and less than 25% other brain tissue, such as white matter and cerebrospinal fluid. Similarly, for white matter perfusion measurement, the perfusion-weighted imaging data were multiplied by the mask image data, which included data for at least 75% white matter. Use of this segmentation step also eliminated signal contributions from large blood vessels, which were potentially misclassified in SPM99 as cerebro-spinal fluid.

Statistics

The spatially normalized and segmented perfusion-weighted imaging data were smoothed by using an isotropic Gaussian kernel filter of 8 mm full width at half maximum to meet the assumption of the Gaussian field theory for statistical parametric mapping. Paired *t* tests were used to assess voxel-to-voxel differences between test and retest measurements obtained with each pulsed ASL MR imaging method. A statistical threshold of $P = .001$, without correction for multiple comparisons, was used.

To examine the sources of the variability in perfusion-weighted imaging data obtained with each pulsed ASL method, a model with effects describing the influence of each subject (S_n, with n labeling each subject) and the influence of the test and retest examinations (with k labeling each test) on the perfusion signal (P_{nk}) was built to separate between-subject variations from within-subject variations (test-retest variations) and noise (ε) according to the following equation:

$$P_{nk} = \beta_n S_n + \beta_k T_k + \varepsilon_{nk}. \quad (1)$$

Here, β_n and β_k are the weights of each subject and each test or retest examination *(T)*, respectively, in variations of perfusion. With the assumption that the effects have random distributions without an interaction between subjects and tests, the overall measurement variance for P_{nk} in Equation (1) can be expressed as follows:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\sigma_P^2 = \sigma^2_{n,n'} + \sigma^2_{k,k'} + \sigma_\varepsilon^2. \quad (2)$$

Here, $\sigma^2_{n,n'}$ is the variance between subjects S_n and $S_{n'}$, $\sigma^2_{k,k'}$ is the variance due to test k and retest k′ in each subject, and σ_ε^2 is the variance due to random noise. The model was fit by using a random-effects analysis of variance design (SPLUS 6; Insightful, Seattle, Wash). For each pulsed ASL method, reliability was computed as an intraclass correlation coefficient (ICC), following the original concept of Shrout and Fleiss (15), according to the following equation:

$$\mathrm{ICC} = \frac{n\left(\sigma^2_{n,n'} - \sigma_\varepsilon^2\right)}{\sigma_{tot}^2}. \quad (3)$$

Here,

$$\sigma_{tot}^2 = n\left(\sigma^2_{n,n'} - \sigma_\varepsilon^2\right) + k\left(\sigma^2_{k,k'} - \sigma_\varepsilon^2\right) + nk\sigma_\varepsilon^2, \quad (4)$$

and *n* and *k* are the numbers of subjects and tests, respectively. An ICC of near unity indicates high reliability (an ICC of 1.0 indicates perfect reliability), whereas a value of 0.5 or lower indicates the randomness of results that have limited use in distinguishing subjects. The test-retest reproducibility of each pulsed ASL method was similarly computed as a within-subject variation coefficient (WSC) according to the following equation:

$$\mathrm{WSC} = \frac{k\left(\sigma^2_{k,k'} - \sigma_\varepsilon^2\right)}{\sigma_{tot}^2}, \quad (5)$$

with a value close to zero indicating high reproducibility. All other variations that were not explained by between- and within-subject effects were considered to be random noise *(N),* the coefficient of which was computed according to the following equation:

$$N = \frac{nk\sigma_\varepsilon^2}{\sigma_{tot}^2}. \quad (6)$$

Note that since in practice $\sigma_\varepsilon^2 / \sigma_{tot}^2$ is greater than zero, neither an ICC of 1 nor a WSC of 0 can ever be reached. Furthermore, as an alternative to the WSC, a coefficient of variation (CV, expressed as a percentage and equal to 100 · standard deviation/mean) was also computed to simplify comparisons with data in previous reports of reproducibility in perfusion imaging.

Reliability and reproducibility were determined separately for perfusion-weighted imaging data averaged over the global (ie, entire) brain, gray matter, and white matter. In addition, to determine regional effects on ICC and WSC, the reliability and reproducibility of perfusion measurements were determined for six regions of interest by using a template of normalized coordinates established by one author (G.H.J.); each region of interest was 24 × 24 × 24 mm and selected from the left and right frontal, parietal, and occipital lobes from standardized space in SPM99. Finally, to test the differences between the pulsed ASL methods in terms of reliability and reproducibility of perfusion measurements, a mixed-effect analysis of variance similar to the analysis performed by using Equation (1) was performed, with random effects describing the variations due to subjects and fixed effects describing the variations due to the ASL methods.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Results

In the top row of Figure 2, representative spatially normalized magnetization-prepared rapid acquisition gradient-echo, fast low-angle shot, and nonlabeled echo-planar images obtained in a 39-year-old volunteer are shown. Normalized perfusion-weighted images (derived from difference between labeled and nonlabeled echo-planar imaging data) obtained in the same volunteer during the first (test) and second (retest) examinations of the test-retest by using the three pulsed ASL methods are shown in the middle and bottom rows, respectively. Voxel-wise comparisons between the test and retest perfusion-weighted imaging data for the group of 13 subjects yielded no significant differences ($P > .001$, SPM99) among any of the three ASL methods.

The reliability, expressed as ICCs; reproducibility, expressed as WSCs; and random noise of perfusion measurements obtained with each ASL method in the global brain, gray matter, and white matter are summarized in Table 1. Also listed in Table 1, as another measure of reproducibility, are CVs. Measurement reliability was highest with DIPLOMA: The ICC reached 0.81 in the global brain and about 0.80 in both the gray matter and the white matter. In comparison, reliability was lower overall with the EPISTAR and PICORE methods. The lowest ICC was 0.75, which was obtained with EPISTAR in gray matter. Noise values were higher than WSCs overall, indicating that at a given ICC, noise rather than WSC is the factor limiting reliability. The lowest noise value, 0.11, was achieved with DIPLOMA in gray matter, whereas the highest noise value, 0.24, was obtained with EPISTAR, also in gray matter. In comparison, the lowest WSC, 0.01, which indicated the best measurement reproducibility, was achieved with the EPISTAR and PICORE methods in the global brain, whereas the highest WSC, 0.09, was achieved with DIPLOMA in gray matter. Similar to the WSC, the CV, an alternative measure of reproducibility, was lowest, 6.11% (indicating best reproducibility), with EPISTAR in the global brain and highest, 8.42%, with PICORE in gray matter.

The differences in reliability between the ASL methods represented only a trend. Analysis of the differences in ICC, WSC, and noise between the ASL methods at mixed-effects analysis of variance yielded F test scores of 2.9 ($df = 1, 76$; $P = .09$) for the global brain, 3.4 ($df = 1, 76$; $P = .07$) for gray matter, and 1.9 ($df = 1, 76$; $P = .17$) for white matter.

To study how the results of an examination that yields a higher signal-to-noise ratio might compare with pulsed ASL results, the reliability (ICC), reproducibility (WSC), and random noise of labeled echo-planar imaging data, which reflect measurements in static brain structures, as well as those in structures with fluctuating perfusion, are listed in Table 2. Labeled echo-planar imaging also yields a much higher signal-to-noise ratio than perfusion-weighted imaging. Labeled echo-planar imaging yielded markedly higher ICCs than did perfusion-weighted imaging: Values ranged from 0.94 with PICORE in gray matter to 0.99 with the DIPLOMA and EPISTAR methods in white matter. Furthermore, the WSC and noise at labeled echo-planar imaging were 0.01– 0.06 and often were similar in magnitude, as compared with these values at perfusion-weighted imaging, with which noise always exceeded WSC.

Regional maps of reliability (ICC), reproducibility (WSC), and random noise constructed on a voxel-by-voxel basis are shown in Figure 3. Consistent with the results in Table 1, the ICC maps showed overall higher reliability with DIPLOMA than with EPISTAR, which yielded reliability losses primarily in regions close to the brain surface. The PICORE and DIPLOMA methods yielded comparable ICC maps, which also showed generally higher reliability in gray matter than in white matter. The WSC maps showed generally low within-subject variability (high reproducibility) throughout the brain; furthermore, they exhibited markedly lower values than the noise maps.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finally, the correlation plots shown in Figure 4 are graphic representations of the reliability of measurements obtained in selected regions of interest in the frontal, middle, and posterior parts of the brain with each pulsed ASL method. The effects of region on reliability were significant, with lower reliability in the frontal regions of interest than in the posterior regions with all ASL methods ($F = 159.3$; $df = 1, 465$; $P < .001$).

Discussion

The main finding in our study was that random noise made significantly greater contributions to fluctuations in perfusion signal than did within-subject variations, regardless of the ASL method used. This finding implies that the primary factor limiting the reliability of per-fusion measurements with ASL MR imaging is poor signal-to-noise ratio—due to the fact that ASL MR images are generated from the subtraction of labeled from nonlabeled echo-planar imaging data—and not biologic fluctuations in perfusion within subjects. An important conclusion based on this result is that the reliability of ASL MR imaging should increase with improved signal-to-noise ratios.

There are a number of ways to improve the signal-to-noise ratio at ASL MR imaging. The use of higher magnetic field strengths results in increased signal-to-noise ratios at ASL MR imaging for two reasons: Spin polarization is increased and the T1 relaxation of arterial blood is extended; these phenomena together yield a larger perfusion signal, although some of the gain in signal-to-noise ratio may be lost owing to increased physiologic noise (16). Improved magnetization transfer cancellation at ASL MR imaging represents another possible mechanism through which the signal-to-noise ratio is increased. Approaches for magnetization transfer cancellation in ASL MR imaging include null pulse techniques (17), which generate no net spin excitation on non-labeled (control) echo-planar images, and dual-coil techniques (18), which involve the use of separate transmit and receive coils for labeling and imaging perfusion. Finally, other promising methods to improve the signal-to-noise ratio at ASL MR imaging include image reconstruction methods, with which a priori information from structural MR imaging data is used to gain spatial information for ASL MR imaging and thus effectively increase the acquisition efficiency per time (19).

We did not observe a significant increase in perfusion measurement reliability or reproducibility with DIPLOMA compared with the reliability and reproducibility observed with the EPISTAR and PICORE ASL methods. However, the DIPLOMA method yielded generally smaller noise levels than did the other two methods; this result is consistent with earlier findings of improved magnetization transfer and eddy current compensation with DIPLOMA (9). One explanation for this finding is that the magnetization transfer and eddy current cancellations with DIPLOMA still may not be perfect, and, therefore, the reduction in instrumental noise was insignifi-cant compared with the variations due to biologic noise.

Another explanation is that with improved ASL techniques, within-subject variability may become a more dominant factor of reliability and reproducibility than noise. This theory is further supported by our observations that decreased noise was accompanied by increased within-subject variability (WSC) with the DIPLOMA method, as compared with the relationship between noise and within-subject variability observed with the other ASL methods. Therefore, in future developments of improved ASL, within-subject variability should be considered as a potential factor that might limit the reliability of perfusion measurements.

An interesting result of the region-of-interest analysis was the lower reliability of perfusion measurements in frontal brain regions compared with the measurement reliability in posterior regions. Higher magnetic susceptibility due to intracranial cavities, which cause both signal intensity variations and geometric distortions at perfusion-weighted imaging, is the most likely

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

reason for the lower reliability in frontal brain regions. In contrast, increased within-subject variability in frontal brain regions is an unlikely reason for limited reliability: In this study, WSCs, as compared with noise values, remained small in frontal regions. Since susceptibility effects generally increase with higher magnetic field strength, susceptibility may be a major challenge to the reliability of higher-field-strength ASL MR imaging.

The reliability and reproducibility of brain perfusion examinations performed with different imaging modalities have been reported previously. However, comparisons with this study are limited because in most other studies (20-22), dayto-day variations—rather than the hour-to-hour variations measured in this study—were assessed. Furthermore, in these previous reports the variability between and within subjects was not separated from noise, and, therefore, the effect of different test-retest examinations on within-subject variability is difficult to interpret in this comparison.

The reliability of contrast material–enhanced computed tomography (CT) performed twice within 24 hours in eight patients with glioma reached a Pearson correlation coefficient of 0.88 (20). However, reliability was determined by using Pearson correlation coefficients, which are generally higher values of reliability than are ICCs (23). In most other perfusion studies, reproducibility was reported in terms of CVs. In a pulsed ASL study involving flow-sensitive alternating inversion-recovery MR imaging (24) performed in healthy subjects, the investigators reported CVs of 18% in gray matter and 14% in white matter for day-to-day variability in cerebral blood flow (21), as compared with CVs of 6.11%–8.42% reported in this study.

In an H_2>O^{15} PET study with healthy volunteers, investigators reported CVs of between 8% in white matter and 10% in gray matter for 2-day intervals (22). In a xenon 133 SPECT inhalation study with healthy subjects, investigators reported a CV of 14% for 1-day intervals (25). In the H_2O^{15} PET study (22), reproducibility was markedly poorer with contrast-enhanced perfusion MR imaging, as reflected by CVs of only 40% in gray matter and 30% in white matter for 2-day intervals.

A limitation of this study was that the reliability and reproducibility of cerebral blood flow measurements, which are more relevant than perfusion values in measuring brain function, were not determined. Quantitative computations of cerebral blood flow from perfusion data require additional information, such as the T1 values of tissue and blood, the arterial transit time, and other parameters that depend on the dynamics of brain perfusion and were not measured in this study because of the prohibitively long acquisition times required.

Another limitation of this study was the exclusion of perfusion-weighted imaging data from patients in the determination of reliability. Because cerebral per-fusion may be lower overall in patients than in healthy subjects, with a resulting reduction in signal-to-noise ratio, we may have overestimated the reliability of ASL in patients. The reason for excluding patients in this study was based on our experience that perfusion-weighted imaging data obtained in patients often indicate head movement and have increased disease-related variability, which would have complicated our studies of the reliability of the different ASL methods.

Another limitation was that no statistical voxel-by-voxel analysis of the test-retest data was performed because of complications in modeling correlations of adjacent voxels (lack of independence). Finally, in the absence of a reference standard for per-fusion measurement, systematic errors were not evaluated. However, perfusion measured by using DIPLOMA ASL varied by less than 1.3%, as compared with that measured by using the EPISTAR and PICORE methods. Therefore, if a systematic error was introduced with DIPLOMA, it was of a magnitude similar to that of the systematic errors introduced by the EPISTAR and PICORE methods.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In summary, the reliability and reproducibility of perfusion examinations performed by using pulsed ASL MR imaging are limited primarily by low signal-to-noise ratios rather than by biologic perfusion fluctuations within subjects.

Acknowledgments

We thank John Ash-burner, PhD, of the University College London for extremely valuable comments and advice regarding the use of SPM99 and for his generosity in providing a MatLab program for data analysis. We also thank Colin Studholme, PhD, of the University of California San Francisco for helpful discussions regarding image registration.

Abbreviations

ASL, arterial spin labeling; CV, coefficient of variation; DIPLOMA, double inversion with proximal labeling of both tagging and control images; EPISTAR, echo-planar imaging and signal targeting with alternating radiofrequency; ICC, intraclass correlation coefficient; PICORE, proximal inversion with a control for off-resonance effects; WSC, within-subject variation coefficient.

References

1. Williams DS, Detre JA, Leigh JS, Koretsky AP. Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proc Natl Acad SciUSA 1992;89:212–216.
2. Gaa J, Warach S, Wen P, Thangaraj V, Wielopolski P, Edelman RR. Noninvasive perfusion imaging of human brain tumors with EPISTAR. Eur Radiol 1996;6:518–522. [PubMed: 8798035]
3. Ye FQ, Berman KF, Ellmore T, et al. H(2)(15)O PET validation of steady-state arterial spin tagging cerebral blood flow measurements in humans. Magn Reson Med 2000;44:450–456. [PubMed: 10975898]
4. Wei X, Warfield SK, Zou KH, et al. Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. J Magn Reson Imaging 2002;15:203–209. [PubMed: 11836778]
5. Wenzel R, Bartenstein P, Dieterich M, et al. Deactivation of human visual cortex during involuntary ocular oscillations: a PET activation study. Brain 1996;119(pt 1):101–110. [PubMed: 8624674]
6. Andersson G, Lyttkens L, Hirvela C, Furmark T, Tillfors M, Fredrikson M. Regional cerebral blood flow during tinnitus: a PET case study with lidocaine and auditory stimulation. Acta Otolaryngol 2000;120:967–972. [PubMed: 11200593]
7. Pekar J, Jezzard P, Roberts DA, Leigh JS Jr, Frank JA, McLaughlin AC. Perfusion imaging with compensation for asymmetric magnetization transfer effects. Magn Reson Med 1996;35:70–79. [PubMed: 8771024]
8. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. Magn Reson Med 1995;34:65–73. [PubMed: 7674900]
9. Jahng GH, Zhu XP, Matson GB, Weiner MW, Schuff N. Improved perfusion-weighted MRI by a novel double inversion with proximal labeling of both tagged and control acquisitions. Magn Reson Med 2003;49:307–314. [PubMed: 12541251]
10. Edelman RR, Chen Q. EPISTAR MRI: multislice mapping of cerebral blood flow. Magn Reson Med 1998;40:800–805. [PubMed: 9840822]
11. Wong EC, Buxton RB, Frank LR. Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. NMR Biomed 1997;10:237–249. [PubMed: 9430354]
12. Luh WM, Wong EC, Bandettini PA, Hyde JS. QUIPSS II with thin-slice TI1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. Magn Reson Med 1999;41:1246–1254. [PubMed: 10371458]
13. Mugler JP 3rd, Brookeman JR. Rapid three-dimensional T1-weighted MR imaging with the MP-RAGE sequence. J Magn Reson Imaging 1991;1:561–567. [PubMed: 1790381]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

14. Crawley AP, Wood ML, Henkelman RM. Elimination of transverse coherences in FLASH MRI. Magn Reson Med 1988;8:248–260. [PubMed: 3205155]
15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
16. Kruger G, Glover GH. Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Reson Med 2001;46:631–637. [PubMed: 11590638]
17. Matson, GB.; Schleich, T. Design of null pulses for use in pulsed arterial spin labeling by the principle of time reversal (abstr). Proceedings of the Ninth Meeting of the International Society for Magnetic Resonance in Medicine; International Society for Magnetic Resonance in Medicine; Glasgow, Scotland. 2001; p. 687
18. Zhang W, Silva AC, Williams DS, Koretsky AP. NMR measurement of perfusion using arterial spin labeling without saturation of macromolecular spins. Magn Reson Med 1995;33:370–376. [PubMed: 7760704]
19. Webb AG, Liang ZP, Magin RL, Lauterbur PC. Applications of reduced-encoding MR imaging with generalized-series reconstruction (RIGR). J Magn Reson Imaging 1993;3:925–928. [PubMed: 8280985]
20. Gillard JH, Antoun NM, Burnet NG, Pickard JD. Reproducibility of quantitative CT perfusion imaging. Br J Radiol 2001;74:552–555. [PubMed: 11459735]
21. Yen YF, Field AS, Martin EM, et al. Test-retest reproducibility of quantitative CBF measurements using FAIR perfusion MRI and acetazolamide challenge. Magn Reson Med 2002;47:921–928. [PubMed: 11979571]
22. Carroll TJ, Teneggi V, Jobin M, et al. Absolute quantification of cerebral blood flow with magnetic resonance, reproducibility of the method, and comparison with H2(15)O positron emission tomography. J Cereb Blood Flow Metab 2002;22:1149–1156. [PubMed: 12218421]
23. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30:1–15. [PubMed: 10907753]
24. Kim SG. Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. Magn Reson Med 1995;34:293–301. [PubMed: 7500865]
25. Blauenstein UW, Halsey JH Jr, Wilson EM, Wills EL, Risberg J. 133Xenon inhalation method: analysis of reproducibility—some of its physiological implications. Stroke 1977;8:92–102. [PubMed: 835162]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Figure 1.
Pulse diagram of the three pulsed ASL methods used in this study: DIPLOMA, PICORE, and EPISTAR. Schemes for spin-labeled *(Label)* and nonlabeled *(Control)* ASL MR image acquisitions are shown. Adiabatic inversion pulses (radiofrequency *[RF]*, curved profiles) with flip angles applied on-resonance (π) and off-resonance (π_{off}) and magnetic field gradients (*G*, squared profiles) are illustrated. Amplitudes of the inversion pulses indicate the relative magnitude of radiofrequency power.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

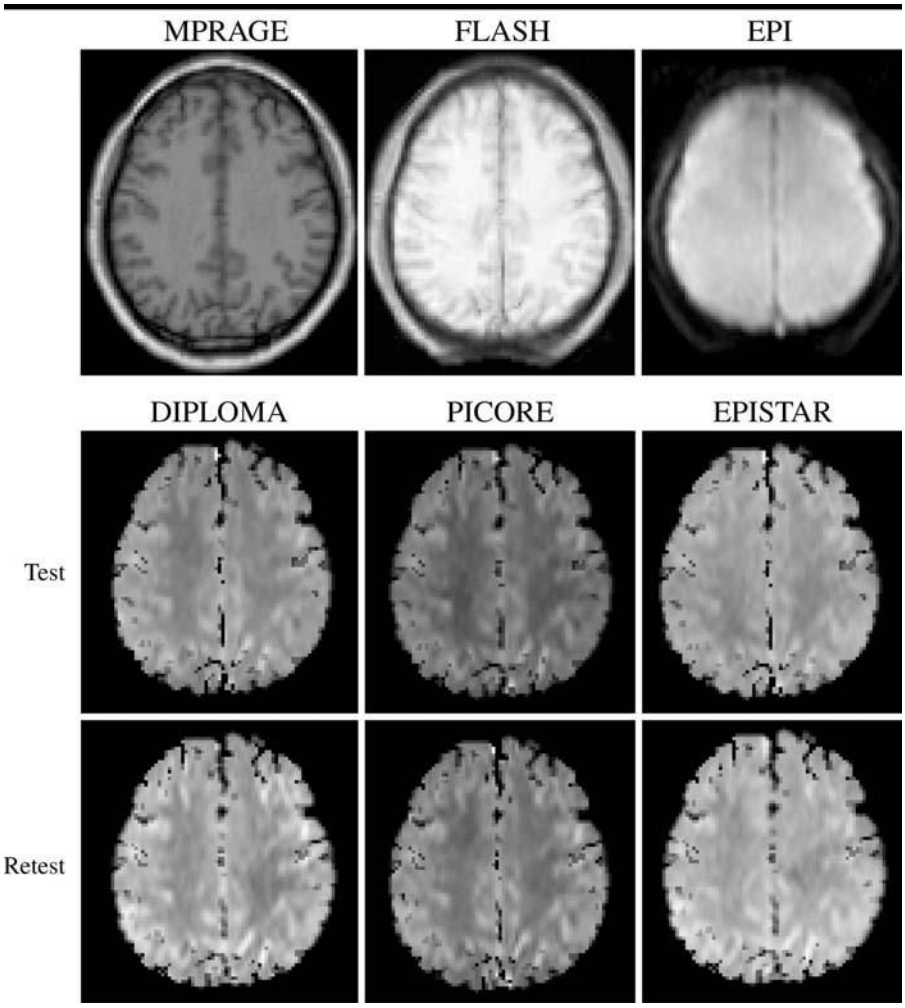

Figure 2.
Top row: representative spatially normalized transverse T1-weighted magnetization-prepared rapid gradient-echo *(MPRAGE)* (10/4/300), intermediate-weighted fast low-angle shot *(FLASH)* (195/6), and control echo-planar *(EPI)* (2500/15/1500) images obtained in 39-year-old male volunteer. Middle row: perfusion-weighted images obtained, by subtracting nonlabeled from labeled echo-planar imaging data, during first (test) part of test-retest examination in the same volunteer by using three pulsed ASL methods. Bottom row: perfusion-weighted images obtained during second (retest) part of test-retest examination in the same volunteer about 1 hour later. Images in middle and bottom rows show the quality of registration and spatial normalization on perfusion-weighted MR images obtained during test-retest examinations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

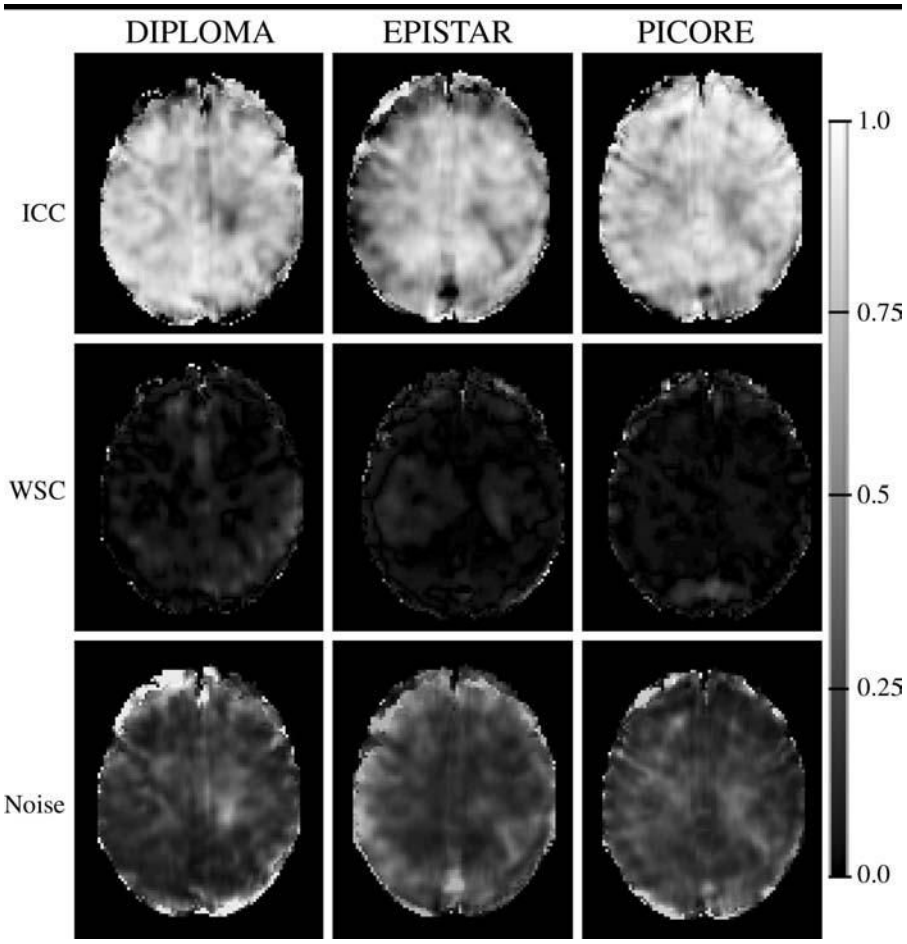

Figure 3.

Statistical maps derived from spatially normalized perfusion-weighted imaging data from 13 healthy volunteers show regional variations in computed reliability (ICC, top row), within-subject variability (WSC, middle row), and noise (bottom row) with the different pulsed ASL perfusion methods. The maps show a relatively smooth distribution of high reliability throughout the brain and noise exceeding within-subject variability with all three ASL methods. Gray-level scale on right indicates measurement reliability, expressed as an ICC: The darker the shading, the lower the reliability; black indicates no reliability (ICC = 0). The lighter the shading, the higher the reliability; white indicates perfect reliability (ICC = 1).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

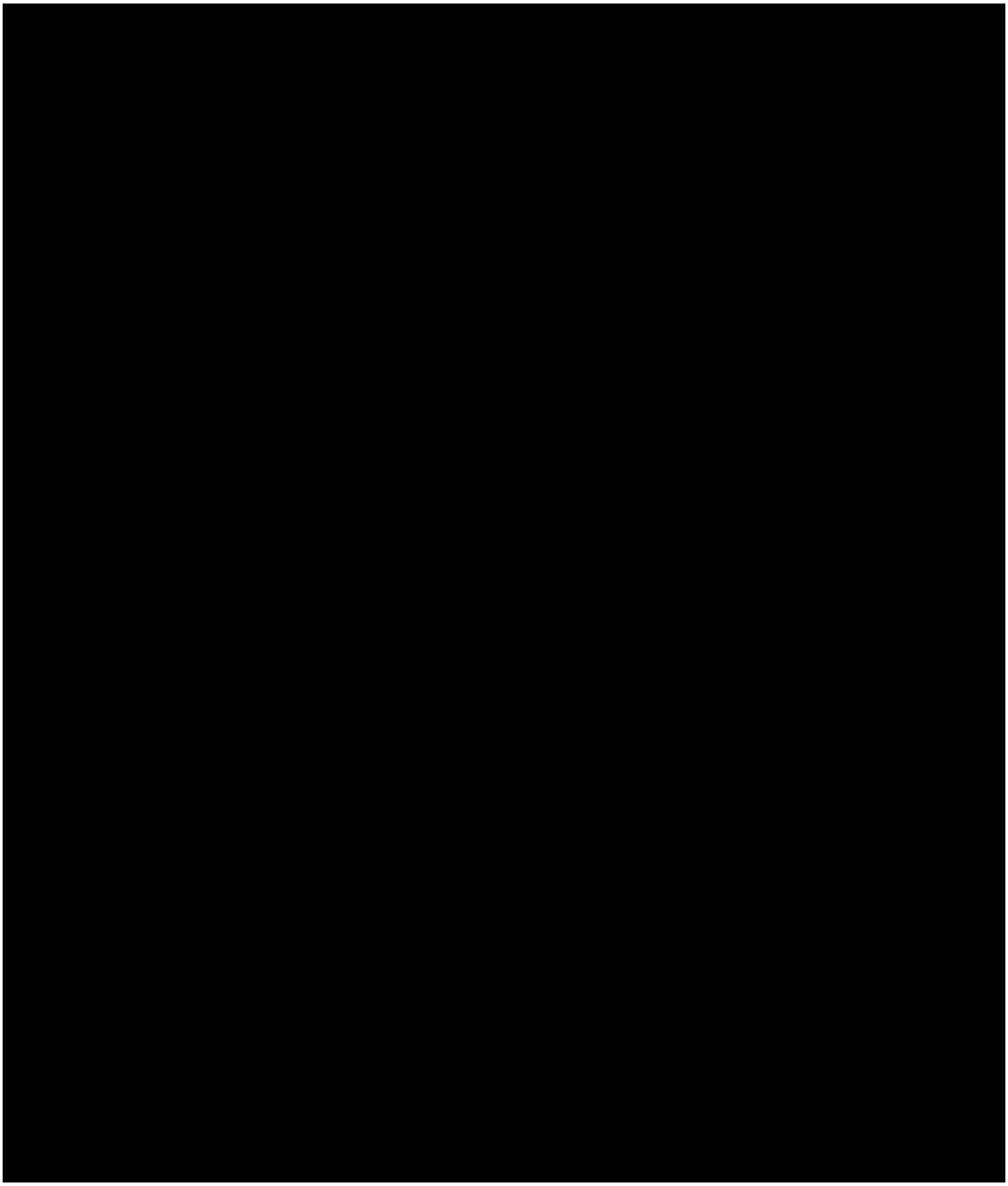

Figure 4.
Correlation plots of test-retest perfusion measurements in frontal and parietal regions of interest *(ROI)* in 13 volunteers, derived from spatially normalized perfusion-weighted imaging *(PWI)* data. Plots show that correlations between the test and retest examinations were slightly better in parietal than in frontal brain regions, regardless of the ASL method used. Perfusion was measured in arbitrary units *(a.u.)*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 1

Reliability, Reproducibility, and Random Noise of Perfusion-weighted Imaging Measurements Obtained with Different Pulsed ASL Methods

Brain Region and Measurement Parameter*	DIPLOMA	EPISTAR	PICORE
Global brain			
ICC †	0.81 (0.61, 0.96)	0.78 (0.41, 0.93)	0.78 (0.40, 0.92)
WSC	0.06	0.01	0.01
Noise	0.13	0.21	0.21
CV (%)	7.07	6.11	8.26
Gray matter			
ICC †	0.80 (0.63, 0.96)	0.75 (0.34, 0.92)	0.76 (0.33, 0.92)
WSC	0.09	0.01	0.01
Noise	0.11	0.24	0.23
CV (%)	6.77	6.13	8.42
White matter			
ICC †	0.80 (0.50, 0.94)	0.79 (0.44, 0.93)	0.79 (0.40, 0.93)
WSC	0.03	0.01	0.02
Noise	0.17	0.20	0.19
CV (%)	7.74	6.31	8.38

* ICC, a measure of reliability; WSC, a measure of reproducibility; and noise are indexes of total variability and vary between 0 and 1. CV, an alternative measure of reproducibility, is a percentage of the mean value.

† Numbers in parentheses are 95% confidence intervals. For the other measures, the confidence intervals (not listed) scale with the ratio of their mean values to the mean value of the ICC.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 2

Reliability, Reproducibility, and Random Noise of Labeled Echo-planar MR Imaging Measurements Obtained with Different Pulsed ASL Methods

Brain Region and Measurement Parameter*	DIPLOMA	EPISTAR	PICORE
Global brain			
ICC †	0.98 (0.94, 0.99)	0.98 (0.93, 0.99)	0.96 (0.88, 0.99)
WSC	<0.01	<0.01	<0.01
Noise	<0.02	<0.02	<0.04
Gray matter			
ICC †	0.96 (0.88, 0.99)	0.97 (0.90, 0.99)	0.94 (0.80, 0.98)
WSC	<0.01	<0.01	<0.01
Noise	<0.03	<0.03	<0.06
White matter			
ICC †	0.99 (0.97, 0.99)	0.99 (0.96, 0.99)	0.98 (0.93, 0.99)
WSC	<0.01	<0.01	<0.01
Noise	<0.01	<0.01	<0.02

*Perfusion-weighted images were obtained by subtracting the labeled from the nonlabeled echo-planar imaging data. ICC, a measure of reliability; WSC, a measure of reproducibility; and random noise are indexes of total variability and vary between 0 and 1.

†Numbers in parentheses are 95% confidence intervals.