<table>
<thead>
<tr>
<th>Title</th>
<th>Reliability and reproducibility of perfusion MRI in cognitively normal subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Jiang, L; Kim, M; Chodkowski, B; Donahue, MJ; Pekar, JJ; Van Zijl, PCM; Albert, M</td>
</tr>
<tr>
<td>Citation</td>
<td>Magnetic Resonance Imaging, 2010, v. 28 n. 9, p. 1283-1289</td>
</tr>
<tr>
<td>Issued Date</td>
<td>2010</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/123994">http://hdl.handle.net/10722/123994</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Magnetic Resonance Imaging. Copyright © Elsevier Inc.</td>
</tr>
</tbody>
</table>
Reliability and reproducibility of perfusion MRI in cognitively normal subjects☆

Li Jiang,a,b,c, Mina Kim,a,b,d, BettyAnn Chodkowskia, Manus J. Donahuea,b, James J. Pekara,b, Peter C.M. Van Zijla,b,⁎, Marilyn Alberete

aF. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD 21205, USA
bThe Russell H. Morgan Department of Radiology and Radiological Science, Division of MR Research, Johns Hopkins University, Baltimore, MD 21205, USA
cDepartment of Radiology, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510080, China
dDepartment of Diagnostic Radiology, The University of Hong Kong, Hong Kong
eDepartment of Neurocognitive Neurology, Johns Hopkins University, Baltimore, MD 21205, USA

Received 4 February 2010; revised 25 April 2010; accepted 8 May 2010

Abstract

Arterial spin labeling (ASL) magnetic resonance imaging (MRI) is becoming a popular method for measuring perfusion due to its ability of generating perfusion maps noninvasively. This allows for frequent repeat scanning, which is especially useful for follow-up studies. However, limited information is available regarding the reliability and reproducibility of ASL perfusion measurements. Here, the reliability and reproducibility of pulsed ASL was investigated in an elderly population to determine the variation in perfusion among cognitively normal individuals in different brain structures. Intraclass correlation coefficients (ICC) and within-subject variation coefficients (wsCV) were used to estimate reliability and reproducibility over a period of 1 year. Twelve cognitively normal subjects (75.5±5.3 years old, six male and six female) were scanned four times (at 0, 3, 6 and 12 months). No significant difference in cerebral blood flow (CBF) was found over this period. CBF values ranged from 46 to 53 ml/100 g per minute in the medial frontal gyrus (MFG) and from 40 to 44 ml/100 g per minute over all gray matter regions in the superior part of the brain. Data obtained from the first two scans were processed by two readers and showed high reliability (ICC = 0.97) and reproducibility (wsCV < 6%). However, over the total period of 1 year, reliability reduced to a moderate level (ICC = 0.63–0.74) with wsCVs of gray matter, left MFG, right MFG of 13.5%, 12.3%, and 15.4%, respectively. In conclusion, measurement of CBF with pulsed ASL provided good agreement between inter-raters. A moderate level of reliability was obtained over a 1-year period, which was attributed to variance in slice positioning and coregistration. As such pulsed ASL has the potential to be used for CBF comparison in longitudinal studies.

© 2010 Elsevier Inc. All rights reserved.

Keywords: Perfusion; Arterial spin labeling; Cerebral blood flow; Reliability; Reproducibility

1. Introduction

Cerebral blood flow is an important physiological parameter for the diagnosis and evaluation of neurological disorders as well as for the examination of brain function [1]. Arterial spin labeling (ASL) is a noninvasive magnetic resonance imaging (MRI) approach that uses magnetic labeling of intravascular blood water to generate images sensitive to cerebral blood flow (CBF) [2–4] and allows the quantification of CBF in physiological units (milliliters of blood per 100 g of tissue per minute). There are several ASL techniques, broadly classified into either continuous ASL (CASL) or pulsed ASL (PASL). Since ASL relies on tagging blood water spins magnetically using radiofrequency pulses instead of intravenous contrast agents, it is completely noninvasive and allows for frequent repeat scanning. ASL has been used to study CBF in several neurological diseases [3,5,6] but has been criticized for problems with quantification [7,8]. Recent studies in animals, however, have demonstrated a good correlation of local CBF measured by

Grants Support: NIH-NCRR P41015241, Glaxo Smith Kline.
* Corresponding author. Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. Tel.: +1 410 443 923 9511; fax: +1 410 614 1948.
E-mail address: pvanzijl@mri.jhu.edu (P.C.M. Van Zijl).

Available online at www.sciencedirect.com
ASL compared with “gold-standard” invasive methods [9]. Furthermore, human studies of cerebral perfusion using ASL have shown that it can provide accurate and precise measurements of CBF over short periods of time [10]. There are also studies showing good correlation between ASL and contrast reagent based techniques and position emission tomography (PET) [3,7,11]. Three of the effects contributing to accuracy problems in ASL are magnetization transfer (MT), limited signal-to-noise ratio (SNR) and the presence of tagged blood water in the blood vessels when short label transfer times are used. The magnetization transfer insensitive labeling technique (TILT) ASL technique was specifically designed to be free of MT artifacts, and, when using an appropriately long wait time (TI) after labeling, is not impaired significantly by intravascular label [12,13].

Depending on the level of neuronal activity, CBF can fluctuate considerably. A reliable method for performing ASL would be important clinically for diagnosing and evaluating a range of pathologic disorders. Specifically, knowledge of reliability and reproducibility of ASL measurements as a function of time in healthy volunteers is critical for comparison with longitudinal studies on patients with slowly developing pathologies such as Alzheimer’s disease. Recently, such reproducibility studies were reported by several groups using different ASL schemes [2,14–16]. These studies mainly focused on short term inter-scan intervals, with the exception of Parkes et al. [2], who studied five subjects with intervals of several months and Hermes et al. [16] who used an interval of 7 weeks for validating continuous arterial spin labeling (CASL). In this study, we used the TILT-ASL approach [12,13] to evaluate the reliability and reproducibility of CBF quantification in cognitively normal elderly subjects over a duration of 1 year.

2. Methods

2.1. Subjects and MR acquisition

The Johns Hopkins Institutional Review Board approved this study, and all subjects provided written informed consent before participating. Twelve elderly subjects (six female and six male, mean age 75.5 years, S.D.=5.3 years, range=61–80 years), judged cognitively normal as defined by the Alzheimer’s Disease Neuroimaging Initiative [17] standards, were scanned four times. The time interval for all participants was 3 months between the first 3 scans and 6 months between the third and fourth scan. All subjects received a routine clinical MRI of the brain to determine if they were qualified for the study. Exclusion criteria included presence of other neurological diseases such as stroke, Parkinson’s disease, significant cerebrovascular disease, and evidence of cerebrovascular risk factors, including diabetes, arrhythmias, lacunar infarcts, and psychiatric disorders.

Imaging was performed on a Philips 3.0T Achieva MR scanner (Philips Medical Systems, Best, The Netherlands), using body coil transmission and sensitivity encoding (SENSE) head coil reception. Brain perfusion was measured with the TILT ASL approach [12] that labels inflowing blood water protons by applying two consecutive slice-selective 90° radiofrequency (RF) pulses inferior to the imaging slice. In the reference scan, the phase of the second 90° pulse is inverted relative to the first, effectively bringing blood water magnetization back to equilibrium. In this scheme, identical MT effects occur in the control and label images, which cancel when these images are subtracted. The tagging parameters, which were optimized in previous work [13], constituted an 80-mm-thick tagging volume located 12.5 mm inferior to the imaging slice to avoid overlap of the area excited by the labeling pulses with the bottom of the imaging slice [12,13,18]. TILT images were acquired using a gradient echo single-shot echo-planar imaging (EPI) sequence with TR/TE/TI of 2000/12/1600 ms, respectively, in order to minimize contamination from labeled spins in blood. Perfusion volumes were collected using nine 3-mm-thick transverse slices with a 1-mm-gap and were positioned parallel to the anterior–posterior comissure line. The superior-most slice reached the vertex of the brain (Fig. 1). The perfusion volumes had a field of view=240 mm, matrix=80×80 and SENSE-factor=2. The alternating labeled and non-labeled (control) ASL acquisitions were repeated 75 times and averaged to acquire sufficient SNR in about 5 scan minutes with the nominal (3×3) mm in-plane spatial resolution. The perfusion-weighted imaging data was derived by subtracting the labeled from the non-labeled imaging data. Co-planar inversion recovery (IR) images were obtained to assist in co-registering the ASL data and for gray matter segmentation. For IR experiments, a gradient echo EPI sequence was used with TR/TI/TE=3000/889/
Turbo spin-echo (TSE) (TR=3000ms, TE1/TE2=8/100ms, thick sections, 8° flip angle) and multislice double-echo 846/3.1 ms, 1×1-mm in-plane spatial resolution, 1.2-mm-resolution volumetric T1-weighted magnetization-prepared TILT except for the use of 22 slices for IR. 26 ms. The geometry of IR prescription was the same as for TILT except for the use of 22 slices for IR.

In addition to perfusion-weighted images, high-spatial-resolution volumetric T1-weighted magnetization-prepared rapid acquisition gradient-echo (MPRAGE) (TR/TI/TE=6.8/846/3.1 ms, 1×1-mm in-plane spatial resolution, 1.2-mm-thick sections, 8° flip angle) and multislice double-echo Turbo spin-echo (TSE) (TR=3000ms, TE1/TE2=8/100ms, 1×1-mm in-plane spatial resolution, 3-mm-thick sections, parallel to the anterior-posterior commissure line) anatomical images of the whole brain were acquired. The MPRAGE and TSE volumes were used for image registration and normalization into a standardized space. Double-echo TSE was also used to assess hyperintensities and other white matter abnormalities.

2.2. Data processing

The MPRAGE, double echo TSE, TILT, and the co-planar IR imaging data were transferred to off-line computer stations for further processing with FSL v3.3 (FMRIB Software Library, FMRIB, Oxford, UK). To optimize image subtraction, TILT images were aligned using a standard 3D rigid body motion correction tool (FSL/MCFLIRT) routine [19,20]. After coregistration, the labeled images were subtracted from the non-labeled images then averaged. A perfusion-weighted CBF volume was computed from the mean subtracted data using a single-compartment exchange model as described previously [13]. When transforming the perfusion-weighted data of each subject into the standard brain template produced by the Montréal Neurological Institute (MNI), the direct registration between the echo-planar imaging-based TILT data and MPRAGE data was not reliable because the geometric and signal intensity distortions in the echo-planar images were greater than those in the MPRAGE images. In contrast to MPRAGE imaging, the co-planar echo-planar IR imaging had the same distortions and spatial resolution but with a clearer differentiation between gray matter and white matter, while the TSE images had the same reconstruction spatial resolution as echo-planar IR imaging without visible distortion. Therefore, to register the perfusion-weighted imaging data to the MPRAGE images, a multistep registration algorithm was used as illustrated in Fig. 2. First, TILT data were registered to IR-EPI data using a six degrees of freedom (DOF) rigid body transformation. Then IR-EPI data were registered to TSE data, and TSE data were registered to MPRAGE data, effectively resulting in a coregistration of perfusion-weighted imaging data with MPRAGE data. Both IR-EPI to TSE and TSE to MPRAGE registrations were accomplished using 7-DOF global transformations. The MPRAGE imaging data were then registered to the standard MNI T1-weighted brain template image (2×2×2-mm spatial resolution) supplied by FSL v3.3 using 12-DOF affine transformations. Each registration step generated a transformation matrix, and all matrices were concatenated to create a final matrix that was used to transform the perfusion-weighted CBF map to MNI-space in a single transformation. The IR images were also transformed into MNI space, where they were used to differentiate gray matter and white matter according to image signal intensity after which the reader described a seed region manually.

After transforming all four CBF scans collected over 12 months into MNI space, voxels common to all four scans were used to define a mask volume. For gray matter perfusion measurements, ASL data were multiplied by this mask to exclude non-gray matter from the analysis. The regional perfusion was quantified in the bilateral middle frontal gyri (MFG) as defined by WFU_PickAtlas created by Wake Forest University [21].

2.3. Statistical analysis

The variability of mean CBF values for the same region over different sessions was investigated using analyses of variance (ANOVA) or the Student’s t test.

To examine the source of the variability in perfusion-weighted imaging data, a model with effects describing the influence of each subject (Sn, where n is the index for each subject) and each measurement (Tk, where k is the index for each measurement) on the perfusion signals was built to separate between-subject variations from within-subject variations and residual variations (ɛnk) according to the following equation:

\[ P_{nk} = \beta_n S_n + \beta_k T_k + \epsilon_{nk} \]  

(1)

\( \beta_n \) and \( \beta_k \) are the weights for subjects and measurements, respectively, in causing variations of perfusion (P). Assuming that the effects have random distributions without an interaction between subjects and measurements, the overall measurement variance for \( P_{nk} \) can be expressed as:

\[ \sigma_p^2 = \sigma_n^2 + \sigma_k^2 + \sigma_\epsilon^2 \]  

(2)

in which \( \sigma_n^2 \) is the variance between subjects, \( \sigma_k^2 \) the variance between measurements in a given subject and \( \sigma_\epsilon^2 \) the residual random variance. The reliability was computed as an intraclass correlation coefficient (ICC), as described by Shrout and Fleiss [22]:

\[ ICC = n (\sigma_n^2 - \sigma_\epsilon^2) / \sigma_{tot}^2 \]  

(3)
where

\[ \sigma_{\text{tot}}^2 = n(\sigma_n^2 - \sigma_e^2) + k(\sigma_k^2 - \sigma_e^2) + nk\sigma_e^2 \]  

(4)

An ICC close to 1.0 indicates high reliability; ICC values between 0.5 and 0.8 indicate a moderate level of reliability, whereas a value of 0.5 or lower indicates a randomness of results that has limited use in distinguishing subjects [15,23].

After determining that there were no statistically significant differences between the sessions, several reproducibility statistics were calculated. First, the within-subject standard deviation (SDw), which was estimated as the square root of the residual mean square using one-way ANOVA according to Bland and Altman [24,25], is a measure of the precision of the measurement error. The difference between a subject’s measurements and the true value is suspected to be less than 1.96×SDw for 95% observations. The repeatability coefficient, which describes the 95% confidence limits (α=5%) for the different repeated measurements, was estimated by the formula \( CL = \sqrt{2} \times 1.96 \times SDw \). Secondly, the within-subject coefficient of variation (wsCV), which quantifies measurements error relative to the size of the measurements, was calculated by dividing SDw by the global mean. When data were examined on the log10 scale, wsCV was approximated by wsCV=antilog(SDw)−1, which is appropriate when the within-subject standard deviation is not large compared with the level of measurement.

Reliability and reproducibility were determined separately for perfusion-weighted imaging data averaged over the whole gray matter (i.e., as covered by the TILT images; see Fig. 1) and bilateral MFG.

### 3. Results

The data from eight subjects for the first two scans were processed by two separate readers (raters 1 and 2) using the same post-processing methods separately and blindly. Table 1 summarizes the inter-rater reliability (expressed in terms of ICCs) and reproducibility of these data. There was high reliability for assessment of the bilateral MFG between the two raters, with ICCs being greater than 0.97. The wsCVs were less than 6%, indicating almost complete agreement between the two raters. The correlation plots in Fig. 3 are a graphic representation of the reproducibility between two raters for measurements obtained in bilateral MFG in the first two scans. The high correlation coefficients \( R^2 \geq 0.94 \) between CBF values measured by the two raters support the measured reproducibility parameters.

Table 2 shows the mean CBF for 12 subjects, as obtained over the four measurements of 3-6 months apart. Even though there seems to be an apparent increase in the CBF values over the year, there was no statistically significant difference between the mean CBF values acquired over a period of 1 year (Fig. 4) for all three regions of interest studied: left MFG \( [df=3, 44]; F=0.33, P=0.80 \), right MFG \( [df=3, 44]; F=0.59, P=0.62 \) and gray matter \( [df=3, 44]; F=0.50, P=0.68 \). The mean differences between measurements ranged from 0.8 to 5.0 ml/100g per minute. There was also no significant difference between the CBF value of left MFG and right MFG at each of the four time points \( (P \text{ values were .58, .96, .88 and .64, respectively}). \)

The reliability and reproducibility of CBF measurements for the four scans over a 1-year period are summarized in Table 3. ICC decreased to 0.63–0.74, representing a moderate level of reliability. The within-subject standard deviation varied between 5.02 and 6.30 ml/100g per minute. The 95% CL varied between 13.9 and 17.5 ml/100 g per minute. wsCV is 12.3% in left MFG, 15.4% in right MFG and 13.5% in gray matter.

### 4. Discussion

The knowledge of reliability and reproducibility of ASL measurements in healthy volunteers is important for diagnosing and evaluating slowly developing pathologic disorders on longitudinal studies. The aim of the present study was to examine the inter-rater reproducibility and the reliability and reproducibility of the PASL-TILT approach for regional CBF measurement over a period of 1 year.

The mean CBF value of MFG in elderly control subjects ranged from 46±11 to 53±10 ml/100 g per minute and the mean CBF value of the gray matter in the superior part of the brain ranged from 40 to 44 ml/100 g per minute over a 1-year period. This is consistent with results reported from PET, commonly showing gray matter CBF values in the range of 40–55 ml/100 g per minute [26]. The CBF values we measured are slightly lower compared to the values reported by other ASL studies [2,10]. In addition to methodological differences (no interference of MT effects and removal of blood signal contributions through the use of a long post-labeling wait time), the reason for this reduced value may be that we studied frontal gray matter. It has been reported that CBF is different between lobes, with the greatest values found in the occipital lobe and the lowest in the frontal lobe [16]. Another possibility is that we studied elderly subjects (age greater than 70 years in 11 out of 12 subjects), while the age of subjects in most other studies was less than 60 years. In support of this, Parkes et al [2,27] reported that gray matter perfusion decreases by 0.45% per year for a
population of age ranging from 20 to 67 years. It is common to find nonspecific hyperintensity in the subcortical white matter in elderly people, which generally corresponds to chronic small vessel disease. It is well known that normal aging is accompanied by brain atrophy [28]. Thinning of cortex results in increasing contamination of gray matter voxel by white matter and, especially, cerebrospinal fluid (CSF), which artificially reduces the measured gray matter CBF values via partial-voluming effects. For ASL measurements, it is not possible to separate out gray matter from CSF due to the large voxel size used (3×3×3 mm³, nominally), and the partial volume effect will reduce the value of gray matter perfusion.

When estimating the reliability and reproducibility between inter-raters, the high ICCs and low wsCVs showed almost complete agreement between the two raters, indicating that the post-processing methods were effective and reliable. The reliability in the left MFG was slightly better than the reliability in the right MFG. Though prior studies showed differences between the CBF values for each hemisphere [2], this small measurement discrepancy may result from different raters. For the post processing, the raters manually selected a seed region to perform the gray matter segmentation. Different readers (raters) will choose the different seed region, thus resulting in different gray matter segmentation for each rater. Note that the result

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Mean CBFa,b for four measurements during 1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Left MFG</td>
<td>49±11</td>
</tr>
<tr>
<td>Right MFG</td>
<td>46±11</td>
</tr>
<tr>
<td>Gray matter</td>
<td>40±10</td>
</tr>
</tbody>
</table>

a In ml/100 g per min.
b ±1 S.D.
of coregistration was the same between raters as it was done automatically.

For the reliability among four scans over a 1-year period, the ICCs decreased to 0.63–0.74, indicating a moderate level of reliability. Jahng [15] has reported ICC values of 0.8–0.9 for different PASL sequences, but the test–retest interval was only 2 h, a brief period over which physiological variability is not expected to be a factor. However, over a period of 1 year, the possible effects from physiological variability cannot be omitted. Another possible variability may be due to slice positioning. We scanned up to the vertex of the brain. It is known that, for EPI, the ghosting artifact is more obvious near the vertex of brain, making it difficult to get good co-registration between the upper slices. In our study, we did not omit the most superior slices which may decrease the reliability.

To examine the reproducibility of the TILT measurements, we calculated the within-subject standard deviation as well as the within-subject coefficient of variation. These analyses indicated that even after 1 year, we can be 95% confident that the differences of CBF values between repeated measurements of gray matter will be 17 ml/100 g per minute or smaller, thus well comparable with the reproducibility analysis of Hermes [16] and similar to the other reliabilities reported based on 1-h or 1-week intervals [2,14]. The wsCV values calculated for gray matter and MFG ranged from 12.3% to 15.4%, which is similar to data by Floyd [10], who reported a wsCV of 13% for whole-brain CBF, while smaller vascular regions demonstrated a mean wsCV of 14±2%. This suggests that, even after 1 year, the reproducibility of TILT measurements is not significantly diminished. Though the smaller the wsCV, the better the reproducibility, a number of factors can influence the reproducibility of MRI examinations. These included subject setup and localization of slice positions such that the same anatomical part is imaged on sequential examinations. In attempt to minimize variability due to slice position, the same overlapping voxels in the four scans were used in the analysis. Another factor that can influence measurement variability is intrinsic physiological change in the tissue itself. As there was a long-time interval between examinations, variations in physiological conditions over the long inter-scan period should be considered. In addition, precision in CBF measurements using ASL perfusion MRI is also dependent on noise related to the scanning technique, including scanner condition and experimental environment.

### 5. Conclusion

ASL can be used for direct measurement of cerebral blood flow, and has been used in the studies of various cerebrovascular disorders in the brain [3,6,17]. We found good agreement between raters and a moderate level of reliability when using pulsed arterial spin labeling to determine absolute CBF over one year. Since ASL perfusion MRI is less invasive and less expensive than other approaches using radioactive tracers or paramagnetic contrast agent, it has the potential advantage in frequent...
follow-up studies to monitor the longitudinal changes of cerebral blood flow in clinical. As a reproducible and reliable quantitative CBF measurement, TILT PASL is also sensitive enough to track such changes and has potential for the study of monitoring disease progression of degenerative diseases such as Alzheimer’s disease.

Acknowledgments

This work was supported by grants from Glaxo Smith Kline and NIH (NCRR P41015241). The authors are grateful to Terri Brawner, Kathleen Kahl, Ivana Kusevic and Joe Gillen for experimental assistance. Dr. Kim is supported by a grant from Philips Medical Systems to Kennedy Krieger Research Institute. One of the supporting agencies, the National Center for Research Resources (NCRR, P41 RR015241), is a component of the National Institutes of Health (NIH). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. Dr. van Zijl is a paid lecturer for Philips Medical Systems. This arrangement has been approved by Johns Hopkins University in accordance with its conflict of interest policies.

References