Arterial spin labeling perfusion MRI: Inter-vendor reproducibility and clinical applicability
Mutsaerts, H.J.M.M.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Quantitative functional arterial spin labeling (fASL) MRI – sensitivity and reproducibility of regional CBF changes using pseudo-continuous ASL product sequences
Abstract

Arterial spin labeling (ASL) magnetic resonance imaging is increasingly used to quantify task-related brain activation. This study assessed functional ASL (fASL) using pseudo-continuous ASL (pCASL) product sequences from two vendors. By scanning healthy participants twice with each sequence while they performed a motor task, this study assessed functional ASL for 1) its sensitivity to detect task-related cerebral blood flow (CBF) changes, and 2) its reproducibility of resting CBF and absolute CBF changes (delta CBF) in the motor cortex. Whole-brain voxel-wise analyses showed that sensitivity for motor activation was sufficient with each sequence, and comparable between sequences. Reproducibility was assessed with within-subject coefficients of variation (wsCV) and intraclass correlation coefficients (ICC). Reproducibility of resting CBF was reasonably good within (wsCV: 12.1-14.9%; ICC: 0.68-0.80) and between sequences (wsCV: 14.0%; ICC: 0.71). Reproducibility of delta CBF was relatively low, both within (wsCV: 209-259%; ICC: 0.29-0.35) and between sequences (wsCV: 214%; ICC: 0.39), while inter-session variation was low. This may be due to delta CBF’s small mean effect (0.06-0.86 mL/100g gray matter/min). In conclusion, fASL seems sufficiently sensitive to detect task-related changes on a group level, with acceptable inter-sequence differences. Resting CBF may provide a consistent baseline to compare task-related activation to, but absolute regional CBF changes are more variable, and should be interpreted cautiously when acquired with two pCASL product sequences.
Introduction

Arterial spin labeling (ASL) perfusion magnetic resonance imaging (MRI) is being increasingly used for imaging of task-related brain activation. Such functional ASL (fASL) has been used to study the neural correlates of a multitude of cognitive domains, including attention, memory, language, visual, and sensorimotor processing, and is increasingly considered as an alternative to blood oxygen level-dependent (BOLD) functional MRI (fMRI), which has been predominantly used as a marker for neural activation during the last two decades.

ASL has several advantages over BOLD imaging with respect to acquisition and interpretation. First, ASL has better sensitivity in low frequency paradigms. The BOLD signal has been shown to be confounded by slow ‘drift’ effects in baseline signal, which are reduced in ASL imaging as a result of the pairwise subtraction of labeled and unlabeled images. Second, despite the intrinsically low signal-to-noise ratio (SNR) of ASL, spatial localization of neuronal activity seems more accurate when measured with ASL than with BOLD. The BOLD signal is affected by macrovascular venous effects whereas ASL is more sensitive to the microvasculature. The interpretation of the BOLD signal is more complex as it reflects a combination of cerebral blood flow (CBF), cerebral blood volume (CBV) and cerebral metabolic rate of oxygen consumption (CMRO2), whereas ASL provides a measure of CBF that is relatively less sensitive to other hemodynamic parameters. Furthermore, ASL provides an in principle quantitative measure of CBF, whereas the BOLD signal is relative. These advantages favor the application of fASL over fMRI BOLD for task-related brain imaging.

The quantitative aspect of ASL in particular could facilitate the comparison and exchange of CBF values across multiple sites and enable multicenter studies, for instance, to pool data. However, before fASL can be used as such, its variability needs to determined, not only within sessions and scanners, but also between product sequences of different vendors, as each vendor provides its own particular ASL implementation. Reproducibility of ASL in general is affected by intrinsic properties, such as low SNR and relative sensitivity to hemodynamics such as arterial transit time (ATT). In addition, although within-sequence reproducibility is sufficient for the commonly available labeling schemes, pseudo-continuous arterial spin labeling (pCASL) has been shown to be best reproducible within session, scanner, and vendor, being more stable and less variable than continuous ASL (CASL) and pulsed ASL (PASL). Even if vendors do employ the same labeling scheme, variability is introduced by differences between MR scanners, such as gradient design, radiofrequency transmit chain, receive-filters and reconstruction algorithms.
Nevertheless, not every user is aware of the potential impact of these factors and may assume that any ASL implementation will provide the same information, as can be expected from a quantitative technique. This may seem particularly appealing for the quantification of brain activation in functional imaging studies. The extent to which different vendor implementations affect these data is not known. We will therefore compare two pCASL product sequences of two different vendors, by limiting adjustment of sequence parameters to within the constraints imposed by the vendor-specific implementation.

Baseline or resting CBF values have been found to be well reproducible within sessions, within scanners, and between scanners of the same vendor on a whole-brain level, whereas on a regional level reproducibility was lower\textsuperscript{23, 24}. We previously assessed the reproducibility of whole-brain resting CBF within and between pCASL product sequences at 3T scanners of two different vendors\textsuperscript{25}. Mean global CBF did not differ between vendors, i.e. between product sequences, but voxel-by-voxel assessment revealed regional differences. Regional variability presents a challenge for fASL, where local effects are of particular interest. In addition to the variability in regional CBF changes, the variability in the detection of such CBF changes needs to be assessed. Sufficient and similar sensitivity to detect local task-induced CBF changes is a prerequisite for multicenter fASL implementations, and essential to good reproducibility.

As of yet, variability of quantitative fASL and variation of sensitivity for task-induced CBF changes between product sequences of different vendors have not been studied. Not only is this information essential for exchanging and comparing fASL data, but results generated by one product sequence can only be generalized to another if variability between them is known. The aim of the present study was to assess quantitative fASL by 1) assessing sensitivity to detect regional CBF changes in a voxel-wise whole-brain analysis, and 2) by investigating regional reproducibility of both resting CBF and task-induced CBF changes in the primary motor cortex, within and between pCASL product sequences from two major vendors. We investigated this by means of paced finger tapping, a simple behavioral paradigm that is known to elicit robust and consistent regional activation in the primary motor cortex in a multitude of activation studies using BOLD as well as fASL \textsuperscript{5, 7}. We employed this paradigm in healthy volunteers using vendor-supplied ASL sequences.
Methods

1.1 Participants
Twenty-two healthy volunteers, aged 18-40 years, were recruited as part of a larger study on ASL reproducibility. Participants were recruited through advertisement at the University of Amsterdam. Only participants with no history of neurological or psychiatric disease were included. Participants that used medication other than contraceptives, or had contraindications for MRI were excluded.

Participants were asked to limit their consumption of alcohol, nicotine and caffeine to a maximum of three units 12-24 hours prior to scanning, and to refrain from consuming alcohol, nicotine and caffeine 12 hours prior to scanning. The study was approved by the local medical research ethics committees of both sites: the Erasmus MC – University Medical Center Rotterdam and the Academic Medical Center, Amsterdam and was conducted according to the Declaration of Helsinki. All participants gave written informed consent and received financial compensation for participation.

1.2 Image acquisition
Imaging was performed on a 3T Intera (Philips Healthcare, Best, the Netherlands) and a 3T Discovery MR750 (GE Healthcare, WI, USA) scanner, using an 8 channel receive head coil. Participants were scanned twice on both scanners, i.e. four times in total, in no specific order. Scanning sessions were separated by at least one week, but no more than four weeks.

A high resolution 3D T1-weighted (T1w) scan for anatomical reference was acquired during one of the two sessions on each scanner. Perfusion data were acquired using pCASL product sequences that were provided by the vendor of each scanner. As we aimed to assess the reproducibility of the currently implemented (i.e. product) sequences, we chose to employ vendor-supplied sequences rather than to reprogram the sequences to make them match completely. Hence, parameters were only adjusted within the given limits of the clinical scanning sequences. Imaging data of the two sites were acquired by two different researchers. Strict agreements were made about the complete process of instructing and positioning participants in order to minimize differences between researchers and subsequently between sequences. Details of both pCASL sequences are listed in Table 1. Note that on the GE scanner a segmented 3D readout was employed, whereas on the Philips scanner single-shot multi-slice 2D imaging was combined with averaging to obtain a temporal resolution similar to the GE-sequence. Another difference between sequences was that on the GE scanner perfusion data are averaged during acquisition, whereas for Philips data is averaged after acquisition, after pair-wise subtraction of label and
control images during post-processing. As the GE sequence only provides three fixed post-labeling delays (1 025, 1 525 and 2 025 ms for respectively children, adults, and older adults or adults with cerebrovascular disease), a delay of 1 525 ms, considered most suitable for the current population, was selected for all ASL imaging.

<table>
<thead>
<tr>
<th>GE</th>
<th>Philips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readout sequence</td>
<td>3D FSE stack-of-spirals</td>
</tr>
<tr>
<td>Acquisition matrix</td>
<td>8 arms with 512 sampling points</td>
</tr>
<tr>
<td>Parallel imaging</td>
<td>No</td>
</tr>
<tr>
<td>Voxel size</td>
<td>3.75 * 3.75 * 4 mm</td>
</tr>
<tr>
<td>Field of view (FOV)</td>
<td>24 * 24 cm$^2$</td>
</tr>
<tr>
<td>Number of slices</td>
<td>36</td>
</tr>
<tr>
<td>Echo time</td>
<td>10.5 ms</td>
</tr>
<tr>
<td>Repetition time</td>
<td>4 600 ms</td>
</tr>
<tr>
<td>Flip angle</td>
<td>111°</td>
</tr>
<tr>
<td>Labeling duration</td>
<td>1 450 ms</td>
</tr>
<tr>
<td>Post-labeling delay</td>
<td>1 525 ms</td>
</tr>
<tr>
<td>Labeling plane (distance from AC-PC line in head-feet direction)</td>
<td>89 mm</td>
</tr>
<tr>
<td>Background suppression</td>
<td>Yes</td>
</tr>
<tr>
<td>Vascular crushers</td>
<td>No</td>
</tr>
<tr>
<td>Total scanning time</td>
<td>2:01 min</td>
</tr>
<tr>
<td>NEX; no. of excitations (GE)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Vendor-specific parameters of the pCASL product sequences. * anterior commissure – posterior commissure line
1.3 Motor activation paradigm

Eight ASL scans were acquired, during which participants performed a blocked motor activation task. Block length was equal to acquisition time of one scan, i.e. 2 minutes. Participants were instructed to tap the fingers of both hands to the thumbs in random order (finger tapping, FT) during the odd scans and to keep their hands still (rest) during even scans. FT was auditorily paced at a frequency of 1 Hz.

1.4 Data processing

The imaging data were processed according to the methods described in full in Bron et al., 2014.

1.4.1 Tissue segmentation

The unified tissue segmentation method (Ashburner and Friston, 2005) of SPM8 (Statistical Parametric Mapping 8, Wellcome Trust Centre for Neuroimaging, University College London, UK) was used to obtain gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) probability maps from the T1w image.

1.4.2 ASL post-processing

For Philips data, label and control pCASL images were pair-wise subtracted (without motion correction) and averaged to obtain perfusion weighted images. For GE data, the perfusion-weighted images, as provided by the scanner, were used. For each participant, the perfusion-weighted image and the GM probability map were rigidly registered (Elastix registration software). The results of the registration were visually inspected.

1.4.3 Quantification

To quantify the perfusion-weighted maps of both pCASL sequences as cerebral blood flow (CBF) maps, a single-compartment model was used:

\[
CBF (ml/100g/min) = \frac{6000 \lambda \Delta M e^{PLD/T_{1a}}}{2a T_{1a} M_0 a (1-e^{-T_{1a}/T_{1a}})}
\]

Parameters used in this model and their values are summarized in Table 2. Differences in effective post-labeling delay for different slices resulting from the 2D multi-slice readout were accounted for in the Philips data (Table 2).
Chapter 4  
Sensitivity and reproducibility of regional CBF changes using PCASL product sequences

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>blood-brain partition coefficient for gray matter</td>
<td>0.9 mL/g</td>
</tr>
<tr>
<td>$\Delta M$</td>
<td>perfusion-weighted image for arterial blood (50 ms) during the 17 ms echo time (TE) by $e^{\frac{TE}{T2^*}}$ (St Lawrence and Wang, 2005))</td>
<td>Philips: corrected for transversal magnetization decay time ($T_2^*$)</td>
</tr>
<tr>
<td>$M_{0a}$</td>
<td>magnetization of arterial blood</td>
<td>GE: obtained by individual proton density maps, adjusted for T1 equilibrium decay time of gray matter tissue ($T1_{GM}$, 1.2 s) during saturation recovery time ($t_{sat}$, 2 s) by $1 - e^{-\frac{t_{sat}}{T1_{GM}}}$ Philips: scanner average (3.7*10^6 a.u.) from previous study (Heijtel et al., 2014)</td>
</tr>
<tr>
<td>PLD</td>
<td>post-labeling delay</td>
<td>1 525 ms</td>
</tr>
<tr>
<td>$T_{1a}$</td>
<td>longitudinal relaxation time of arterial blood</td>
<td>1 650 ms (Lu et al., 2004)</td>
</tr>
<tr>
<td>$a$</td>
<td>labeling efficiency</td>
<td>GE: $a * 0.75$ Philips: $a * 0.83$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>labeling duration</td>
<td>GE: 1 450 ms Philips: 1 650 ms</td>
</tr>
</tbody>
</table>

Table 2. variables of the single-compartment model used for quantification (based Alsop et al., 2014)

1.5 Whole-brain voxel-wise preprocessing and activation sensitivity

Registered T1w images and CBF maps were transformed to a common template space based on the T1w images of all participants; CBF maps were smoothed using an isotropic 8 mm full width at half maximum (FWHM) kernel.

Voxel-wise differences within and between sequences in relation to finger tapping were assessed using SPM8. Averaged CBF maps per block of finger tapping (FT) and rest were convolved with the hemodynamic response function and modeled on an individual level using a General Linear Model (GLM), yielding parameter estimates for the main effects of FT and rest and the contrast [FT > rest]
which were subsequently used in group analyses. Task-induced changes were assessed per session per sequence by pair-wise comparison of main effects of FT and rest for each participant during each session. As participants were scanned in a random order, we chose to assess inter-sequence differences by means of a repeated measures ANOVA on the contrast [FT > rest] for GE (session) 1 compared to Philips 2 and GE 2 to Philips 1 (n=44). By comparing sessions this way, we can assume that temporal physiological variation affected intra-sequence and inter-sequence reproducibility to a similar extent. All voxel-wise results were thresholded at p<0.001 without correction for multiple comparisons, to be maximally sensitive to intra- and inter-sequence differences in the detection of CBF changes.

1.6 ROI preprocessing and regional reproducibility analysis

1.6.1 ROI labeling and selection

Individual CBF maps were transformed to individual T1w image space for region of interest (ROI) analysis. ROIs for each participant were defined using a multi-atlas approach by registering thirty labeled T1w images, each containing 83 ROIs,29,30 to the participants’ T1w images, using a rigid, affine, and non-rigid model consecutively. For the current study, we focused on the bilateral primary motor cortex (precentral gyri). Analysis of CBF in the primary motor cortex was performed in GM only.

1.6.2 CBF post-processing in the primary motor cortex

For every pCASL scan, mean GM CBF values were obtained from the left and right precentral gyrus (primary motor cortex). CBF values were averaged per session over the four FT blocks and over the four rest blocks, and then over the primary motor cortex bilaterally, such that for every session we obtained one mean GM CBF value in the bilateral primary motor cortex for the FT condition and one for the rest condition.

1.6.3 Task-induced regional CBF changes within sessions

To assess task-induced CBF changes in the primary motor cortex, mean FT CBF (CBF_{FT}) and resting CBF (CBF_{rest}) within sessions were compared with paired t-tests (p<.05). Absolute CBF changes as a result of finger tapping are referred to as delta CBF: \( \Delta \text{CBF}_{\text{FT}} - \Delta \text{CBF}_{\text{rest}} \). Delta CBF within sessions was normalized by expressing it as a percentage of resting CBF: \( \frac{\Delta \text{CBF}_{\text{FT}} - \Delta \text{CBF}_{\text{rest}}}{\text{CBF}_{\text{rest}}} \).
1.6.4. Reproducibility of regional resting CBF and regional task-induced CBF changes

Intra- and inter-sequence reproducibility were assessed by the following measures:

1. Within-subject coefficients of variation (wsCV) were calculated as the ratio of the standard deviation of the CBF difference (SD\text{diff}) between sessions to the mean CBF value of those sessions: \( wsCV = 100\% \times (SD\text{diff}/\text{mean value}) \). The SD\text{diff}, rather than the SD of the mean, was used to reflect the extent of variability in differences in relation to the mean. Mean CBF values over sessions, mean CBF differences between sessions, SD\text{diff}, and wsCVs and their 95% confidence intervals (CIs) are reported for CBF\text{rest}, CBF\text{PT}, and delta CBF.

Intra-sequence measures were calculated between the two sessions per sequence. Inter-sequence measures were calculated by comparing GE (session) 1 to Philips 2 and GE 2 to Philips 1 (\( n=44 \)).

2. Intraclass correlation coefficients (ICC) and 95% CIs were calculated for CBF\text{rest}, CBF\text{PT}, and delta CBF. A two-way-random model and absolute agreement were employed to allow for generalization of the results and to take into account systematic variability between sequences, respectively. ICCs were defined as function of ANOVA mean squares using the following formula: \(^{31}\)

\[
\frac{BMS-EMS}{BMS+(k-1)EMS+k/n(JMS-EMS)}
\]  

[2]

in which BMS refers to the between-targets mean square (i.e. variance between participants), JMS refers to the between-judges mean square (i.e. variance between intra- or inter-sequence sessions) and EMS to the residual mean square (i.e. residual sources of variance), in a two-way ANOVA with \( n = 22 \) (intra-sequence) or 44 (inter-sequence) targets and \( k = 2 \) judges.

Intravendor ICCs were calculated between the two sessions per sequence by comparing GE (session) 1 to Philips 2 and GE 2 to Philips 1 (\( n=44 \)).

3. Bland-Altman plots and 95% limits of agreement (mean difference \( \pm 1.96 \times \text{SD\text{diff}} \)) were created for CBF\text{rest} and delta CBF to visualize agreement between sessions and between sequences respectively.

Statistical analyses were carried out in IBM SPSS Statistics, version 20.0 (New York, USA).
Results

2.1 Participant characteristics
Nine male and 13 female volunteers with a mean age of 22.1 ± 2.1 years (range: 19-27 years) participated in the study. It should be noted that one participant had CBF values that were 2 – 3 standard deviations higher than the group mean, but this participant was retained in the analysis as data were normally distributed (Kolmogorov-Smirnov tests did not detect significant deviations from normality in any session, p>.05). Scan sessions were separated by 2.8 ± 1.0 weeks on the GE and 2.6 ± 0.9 weeks on the Philips scanner (not significant (n.s.)). Inter-sequence sessions were separated by 3.1 ± 1.1 weeks (GE 1 – Philips 2) and 2.6 ± 1.9 weeks (GE 2 – Philips 1), n.s.. Sessions on the GE scanner took place at an earlier time of day than sessions on the Philips scanner: 3:26pm ± 4h00min and 3:55pm ± 3h34min versus 8:16pm ± 2h06min and 7h47pm ± 2h38min respectively, p<.05.

2.2 Whole-brain voxel-wise activation sensitivity of pCASL sequences
Voxel-wise CBF changes in relation to FT are illustrated with t-statistic maps in Figure 1. Both with GE (fig. 1A) and Philips (fig. 1B) CBF increases were observed in the bilateral primary motor cortex in both sessions. Additional activation was observed in the supplementary motor area and the left cerebellum in GE session 2; and in the thalamus, and supplementary motor area in both Philips sessions. As can be appreciated visually, Philips (fig. 1B) seems to be more sensitive to detect activation than GE (fig. 1A). Upon formal assessment with repeated measures ANOVA (fig. 1C), differences between pCASL sequences were found in the right primary motor cortex, left precuneus, right posterior cingulate, and in the bilateral thalamus.
2.3 Task-induced regional CBF changes within sessions

Absolute CBF values differed systematically between pCASL sequences, with mean CBF_{rest} values being 3.1 mL/100g GM/min higher as measured with GE than with Philips. The increase in CBF in the motor cortex as a result of FT was significant for both sequences and all sessions (Table 3). Normalized CBF changes showed a larger increase in CBF for the two GE sessions (4.65 and 3.31%) than for the two Philips sessions (2.46 and 2.65%).
Sensitivity and reproducibility of regional CBF changes using PCASL product sequences

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>CBF (mL/100g GM/min)</th>
<th>p-value</th>
<th>Normalized delta CBF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 1</td>
<td>FT</td>
<td>68.2</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>65.2</td>
<td>13.7</td>
</tr>
<tr>
<td>Session 2</td>
<td>FT</td>
<td>68.2</td>
<td>9.24</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>66.0</td>
<td>10.3</td>
</tr>
<tr>
<td>Philips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 1</td>
<td>FT</td>
<td>71.9</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>70.1</td>
<td>13.1</td>
</tr>
<tr>
<td>Session 2</td>
<td>FT</td>
<td>69.0</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>67.2</td>
<td>14.1</td>
</tr>
</tbody>
</table>

Table 3. Mean CBF (mL/100g GM/min) and standard deviations (SD) in the motor cortex during finger tapping and rest, and respective p-values and normalized delta CBF, per sequence per session. CBF: cerebral blood flow, SD: standard deviation, FT: finger tapping, R: rest.

### 2.4 Reproducibility of regional resting CBF and regional task-induced CBF changes

WsCV, SD diff, mean CBF values and ICCs and 95% CIs are reported for CBF<sub>rest</sub>, CBF<sub>FT</sub>, and delta CBF in Table 4. WsCVs were comparable within and between pCASL sequences for CBF<sub>rest</sub> and CBF<sub>FT</sub>. WsCVs for delta CBF were much higher as the mean effect of delta CBF is much smaller than that of CBF<sub>rest</sub> or CBF<sub>FT</sub>.

Reproducibility of CBF<sub>rest</sub> and CBF<sub>FT</sub> in terms of ICCs was moderate to good for both sequences, with ICCs of .68 and .66 for GE and .80 and .77 for Philips, respectively. Reproducibility of absolute delta CBF was poor for both sequences with ICCs of .35 (GE) and .29 (Philips), and CIs being 1.5 – 2.4 times larger than for CBF<sub>rest</sub> and CBF<sub>FT</sub>. Between sequences, reproducibility was reasonable for CBF<sub>rest</sub> (.71) and CBF<sub>FT</sub> (.65) and poor for absolute delta CBF (ICC: .39). Inter-sequence CIs for delta CBF were approximately 1.5 times larger than for CBF<sub>rest</sub> and CBF<sub>FT</sub>.
### Table 4

Mean CBF measurements and reproducibility estimates between sessions and sequences for resting CBF (CBF<sub>rest</sub>), finger tapping CBF (CBF<sub>FT</sub>) and delta CBF in the primary motor cortex. CBF: cerebral blood flow, SD: standard deviation, wsCV: within subject coefficient of variation, ICC: intraclass correlation coefficient, CI: confidence interval.

<table>
<thead>
<tr>
<th></th>
<th>GE</th>
<th>95% CI</th>
<th>Philips</th>
<th>95% CI</th>
<th>Inter-sequence</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean CBF</strong></td>
<td>65.6</td>
<td>60.7 – 70.5</td>
<td>68.7</td>
<td>62.9 – 74.4</td>
<td>67.1</td>
<td>63.5 – 70.7</td>
</tr>
<tr>
<td><strong>Mean CBF difference</strong></td>
<td>-0.80</td>
<td>-5.14 – 3.54</td>
<td>2.94</td>
<td>-0.74 – 6.61</td>
<td>-3.11</td>
<td>-5.96 – 0.25</td>
</tr>
<tr>
<td><strong>SD difference</strong></td>
<td>9.79</td>
<td>6.65 – 12.9</td>
<td>8.28</td>
<td>5.63 – 10.9</td>
<td>9.38</td>
<td>7.33 – 11.42</td>
</tr>
<tr>
<td>wsCV (%)</td>
<td>14.9</td>
<td>9.10 – 20.8</td>
<td>12.1</td>
<td>5.72 – 18.4</td>
<td>14.0</td>
<td>9.81 – 18.13</td>
</tr>
<tr>
<td><strong>ICC</strong></td>
<td>0.68</td>
<td>0.37 – 0.86</td>
<td>0.80</td>
<td>0.58 – 0.91</td>
<td>0.71</td>
<td>0.53 – 0.84</td>
</tr>
<tr>
<td><strong>Mean CBF</strong></td>
<td>68.2</td>
<td>64.1 – 72.3</td>
<td>70.4</td>
<td>65.5 – 75.3</td>
<td>69.3</td>
<td>66.3 – 72.3</td>
</tr>
<tr>
<td><strong>Mean CBF difference</strong></td>
<td>0.04</td>
<td>-3.72 – 3.81</td>
<td>2.88</td>
<td>-0.57 – 6.34</td>
<td>-2.25</td>
<td>-5.01 – 0.50</td>
</tr>
<tr>
<td><strong>SD difference</strong></td>
<td>8.48</td>
<td>5.76 – 11.2</td>
<td>7.80</td>
<td>5.29 – 10.3</td>
<td>9.04</td>
<td>7.07 – 11.02</td>
</tr>
<tr>
<td>wsCV (%)</td>
<td>12.4</td>
<td>7.53 – 17.4</td>
<td>11.1</td>
<td>5.55 – 16.6</td>
<td>13.1</td>
<td>9.44 – 16.7</td>
</tr>
<tr>
<td><strong>ICC</strong></td>
<td>0.66</td>
<td>0.34 – 0.85</td>
<td>0.77</td>
<td>0.52 – 0.90</td>
<td>0.65</td>
<td>0.44 – 0.79</td>
</tr>
<tr>
<td><strong>Mean CBF</strong></td>
<td>2.61</td>
<td>0.88 – 4.33</td>
<td>1.75</td>
<td>0.42 – 3.09</td>
<td>2.18</td>
<td>1.10 – 3.26</td>
</tr>
<tr>
<td><strong>Mean CBF</strong></td>
<td>0.85</td>
<td>-1.57 – 3.26</td>
<td>-0.06</td>
<td>-2.06 – 1.95</td>
<td>0.86</td>
<td>-0.56 – 2.28</td>
</tr>
<tr>
<td><strong>SD difference</strong></td>
<td>5.45</td>
<td>3.70 – 7.20</td>
<td>4.53</td>
<td>3.08 – 5.98</td>
<td>4.67</td>
<td>3.65 – 5.68</td>
</tr>
<tr>
<td>wsCV (%)</td>
<td>209</td>
<td>207 – 212</td>
<td>259</td>
<td>257 – 261</td>
<td>214</td>
<td>213 – 216</td>
</tr>
<tr>
<td><strong>ICC</strong></td>
<td>0.35</td>
<td>-0.08 – 0.67</td>
<td>0.29</td>
<td>-0.16 – 0.63</td>
<td>0.39</td>
<td>0.12 – 0.62</td>
</tr>
</tbody>
</table>

**Delta CBF**

<table>
<thead>
<tr>
<th></th>
<th>GE</th>
<th>95% CI</th>
<th>Philips</th>
<th>95% CI</th>
<th>Inter-sequence</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean CBF</strong></td>
<td>2.61</td>
<td>0.88 – 4.33</td>
<td>1.75</td>
<td>0.42 – 3.09</td>
<td>2.18</td>
<td>1.10 – 3.26</td>
</tr>
<tr>
<td><strong>Mean CBF</strong></td>
<td>0.85</td>
<td>-1.57 – 3.26</td>
<td>-0.06</td>
<td>-2.06 – 1.95</td>
<td>0.86</td>
<td>-0.56 – 2.28</td>
</tr>
<tr>
<td><strong>SD difference</strong></td>
<td>5.45</td>
<td>3.70 – 7.20</td>
<td>4.53</td>
<td>3.08 – 5.98</td>
<td>4.67</td>
<td>3.65 – 5.68</td>
</tr>
<tr>
<td>wsCV (%)</td>
<td>209</td>
<td>207 – 212</td>
<td>259</td>
<td>257 – 261</td>
<td>214</td>
<td>213 – 216</td>
</tr>
<tr>
<td><strong>ICC</strong></td>
<td>0.35</td>
<td>-0.08 – 0.67</td>
<td>0.29</td>
<td>-0.16 – 0.63</td>
<td>0.39</td>
<td>0.12 – 0.62</td>
</tr>
</tbody>
</table>
Figure 2-I and II show the agreement between sessions and sequences for CBF\textsubscript{rest} and absolute delta CBF respectively. For both CBF\textsubscript{rest} and delta CBF, differences between sessions were somewhat larger for measurements performed with GE (fig. 2A-I and II) than with Philips (fig. 2B-I and II). The low intra-sequence reproducibility of delta CBF (fig. 2A-II, 3B-II) is illustrated by the variability relative to the mean effect being higher for delta CBF than for CBF\textsubscript{rest}, as the mean effect of delta CBF is much smaller than that of CBF\textsubscript{rest} (fig. 2A-I, 2B-I). Both CBF\textsubscript{rest} and delta CBF show a comparable spread in differences within and between sequences. Figure 2C-I and II show the agreement between sequences for CBF\textsubscript{rest} and delta CBF respectively, and illustrate that although the spread in differences is higher for CBF\textsubscript{rest} (2C-I), the variability relative to the mean effect is twice as large for delta CBF (fig. 2C-II), as the mean effect is small.
Discussion

Using pCASL product sequences as supplied by two MR vendors, we found that sensitivity to detect primary motor cortex activation was sufficient and comparable for both sequences. Secondly, we found intra- and inter-sequence reproducibility of resting CBF in the motor cortex to be reasonably good, as was reproducibility of CBF during finger tapping. More important in the context of fASL however, we found absolute CBF changes to be only moderately reproducible, both within and between sequences, despite a consistent task-induced CBF increase within sessions at the group level.

This study adopted a pragmatic and clinically applicable approach in employing pCASL product sequences as supplied by the vendors. Parameters were adjusted to match each other as much as possible, but within the constraints of the provided sequence. This reflects the de facto situation of commercially supplied sequences in which parameters can only be adjusted to a certain extent. As CBF measured with ASL is claimed to be and promoted as a quantitative measure, one would assume that measurements are independent of sequence parameters. Here we show to what extent standard implementation of the sequences gives rise to substantial differences and thus limited reproducibility for functional ASL.

The consistent increase in CBF in the motor cortex during finger tapping is in line with previous cross-sectional fASL studies employing finger tapping paradigms. Our results replicate these findings between sessions for both sequences. Although we found the extent and spatial patterns of activation to differ between sequences, they both detected activation in the primary motor cortex as a result of finger tapping, and formal comparison of activation patterns between sequences demonstrated only minor difference in motor areas. This supports the notion that on the group level, fASL is sufficiently sensitive to detect activation in the primary motor cortex, and that differences in sensitivity between pCASL product sequences are acceptable.

We found reproducibility of resting CBF in the motor cortex to be reasonably good within and good between pCASL sequences. Our intra-sequence results are in line with previous whole-brain resting CBF studies that have shown reasonable to good reproducibility within and between sessions on the same scanner, as well as between scanners of the same vendor. We recently compared resting gray matter CBF between product sequences of two different vendors and also found it to be well reproducible on the global level. The current study focused on regional resting gray matter CBF and showed that, when compared to global CBF, reproducibility was slightly lower. Previous studies found that smaller regions
are subject to higher variability and thus may yield lower reproducibility estimates. In addition, regional reproducibility has been found to be lower for CBF measured at an interval of 2-4 weeks than for measurements within one day, implying that temporal physiological differences dominate between-weeks reproducibility. Other sources of physiological variation may also be present, such as the one participant with CBF values that were consistently higher than the group mean.

Apart from temporal dynamics in physiology, regional variability could also be affected by the difference in effective post-labeling delay (PLD) between pCASL sequences. Although the same initial PLD (1 525 ms) was applied for both sequences in the current study, the 2D multi-slice acquisition employed by Philips allows labeled blood more time to reach superior slices, including regions with longer arrival times, than does the single time-point 3D acquisition employed by GE. This difference in effective PLD between sequences may have contributed to decreased reproducibility between their respective measurements. Moreover, the level of reproducibility of different regions seems to vary with PLD, with a PLD of 2 500 ms yielding better reproducibility than a PLD of 1 500 ms when using a 3D single time-point sequence. Therefore, effective PLD differences may have affected our results, as the precentral gyrus is located at the superior aspect of the brain, and therefore exhibits longer transit delays. This may have led to an underestimation of CBF values due to incomplete inflow of label, and to higher variability due to difference in arterial arrival times.

Differences between sequences may not only affect resting CBF measures, but also those of motor activation. Studies that compared fASL data obtained with 2D and 3D sequences found that activated clusters are generally larger when using 3D sequences, while 2D sequences yield larger effect sizes in terms of relative CBF changes. Our results on the other hand showed larger activated clusters with the 2D sequence, and larger relative CBF changes with the 3D sequence. The larger effective PLD of Philips may have allowed more labeled blood to reach the primary motor cortex during finger tapping than GE, yielding larger activation clusters. In addition, the 3D sequence is more susceptible to spatial blurring, which obscures the gray matter to white matter contrast, and may attenuate signal from the gray matter. The larger relative signal change measured with GE on the other hand may be explained by decreasing arterial transit time (ATT) as a result of finger tapping. Shorter ATT after finger tapping may have caused the relative signal change as measured by GE to be higher because of its shorter effective PLD, particularly in the superior regions.

The reproducibility of delta CBF, i.e. the CBF difference observed between finger tapping and rest, is less straightforward to interpret. Despite the smaller inter-session variation of delta CBF differences (as
indicated by smaller standard deviations of differences) when compared to that of resting CBF and finger tapping CBF, wsCVs are extremely high because of the modest mean effect size of delta CBF. ICCs indicated poor intra-sequence reproducibility as well. Previous fASL studies reported wsCVs of 10-11%, 38, 39 and ICCs up to 0.74, 40 between sessions that were a week apart. However, these studies assessed the reproducibility of relative CBF changes, instead of absolute CBF changes. It has been suggested that relative CBF changes are more accurate and robust than absolute CBF changes, as they may reduce potential effects of basal perfusion variations on measures of neuronal activation 40. Relative CBF changes may therefore generate higher ICCs than absolute CBF changes. Nevertheless, as absolute quantification is a specific advantage of fASL, it seems more appropriate to investigate the reproducibility of absolute CBF changes.

Slight variation in signal change as a result of finger tapping has been observed, but shown to be similar between sessions that took place on the same day or on different days 39. Raoult et al. 40 found similar levels of variation using a finger flexion-extension paradigm. Moreover, they found task-induced CBF to be higher, albeit not significantly, with shorter sequence lengths, and concluded that a sequence duration of 4 minutes – i.e. a motor paradigm with 4 blocks of 30s on/off activation – is optimal for clinical practice 38. Longer sequence durations are considered to induce habituation and thus decreased activation. Our paradigm consisted of 2 minute blocks of on/off activation because of the limited temporal resolution of the GE sequence. These relatively long blocks thus may have attenuated activation, and thus the effect of finger tapping as compared to other studies, which may have reduced reproducibility.

On the other hand, one of the major reasons to use ASL for functional imaging is its suitability for low frequency designs, as it is much less sensitive to drift effects over time than BOLD fMRI 41. Wang et al. 5 even demonstrated that fASL shows constant sensitivity across different task frequencies corresponding to blocks lengths ranging from 0.5 – 5 minutes, with ASL outperforming BOLD contrast at a block length of 4 minutes. Some higher cognitive functions, such as sustained attention 1 depend on an experimental design with even longer blocks to detect slow, low-frequency signal changes of interest, for which fASL is particularly well suited. Despite ASL’s appropriateness for such cognitive paradigms, we purposely chose a simple behavioral paradigm known to elicit robust and consistent regional activation, before moving on to more complex processes and paradigms. We find that even this simple motor activation paradigm gives rise to substantial variability, which warrants caution with respect to more complex and less robust designs.
To our knowledge, no studies exist on the reproducibility of task-induced CBF changes using pCASL product sequences from two different vendors. In the current study, we found reproducibility of task-induced CBF to be comparable within and between sequences, both in terms of wsCV and ICC. Nevertheless, the findings indicate that absolute CBF changes in the motor cortex still vary considerably, and this variation needs to be taken into account when comparing regional quantitative CBF changes, particularly between sequences. Therefore, absolute fASL data should not be simply pooled between vendors, i.e. between product sequences.

This study has some limitations. First, we did not collect information on motor behavior. Although variations in frequency were avoided by externally pacing the finger tapping, we may have missed individual variations in tapping, which may have added to the variability. Next, as time of acquisition differed between sequences, with GE data collected earlier on the day, diurnal fluctuations in CBF may have added to variability between sequences. This potentially affected reproducibility of resting CBF more than that of delta CBF, as the latter is based on a subtractive measure. Still, inter-sequence reproducibility of resting CBF was found to be reasonably good. Furthermore, subsequent analysis of the Philips data was performed in a similar manner to maximize comparability with the GE data, i.e. by averaging over rest and activation periods, whereas one would normally choose to exploit the higher temporal resolution in a more formal manner within the design matrix. Finally, due to practical constraints we studied the product sequences of only two out of the three major vendors on the market. Although assessment of variability between the three vendors would have been more comprehensive, the current study was conducted as a proof-of-principle, and demonstrated as such that substantial variability is already evident when product sequences of two vendors are compared. Future work should be directed at optimizing ASL sequences for functional imaging, and at assessing sensitivity and reproducibility of fASL in single-subject designs, as longitudinal studies and clinical application of fASL will eventually need to be aimed at repeated measurements within individuals.

In conclusion, in a voxel-wise whole-brain analysis, fASL shows sufficient sensitivity to detect regional CBF changes on a group level, both within and between pCASL product sequences of two different vendors. The between sequence reproducibility of fASL is comparable with within sequence reproducibility, although inter-sequence differences in readout should be taken into account. Although reproducibility of regional resting CBF is affected by differences in sequence implementation, particularly in the readout, resting CBF in the motor cortex may provide a reasonably consistent baseline to compare task-induced CBF to. The relatively low reproducibility of task-induced CBF changes in the primary motor cortex, however, should be taken into consideration when comparing fASL data between
sessions and particularly between pCASL product sequences as implemented by different vendors. Its interpretation should be performed with caution in repeated measurements and multicenter designs, as current vendor-specific implementations do not allow for simple pooling of functional ASL data.

Reference list


98


Figure 2. Agreement within and between pCASL sequences for resting CBF and delta CBF. Bland Altman plots of intra-sequence (A, B) and inter-sequence (C) agreement for resting CBF (I) and delta CBF (II) in the primary motor cortex. The solid line indicates the mean difference between sessions, dotted lines the 95% limits of agreement.