Calibration of arterial spin labeling data—potential pitfalls in post-processing

Joana Pinto1 | Michael A. Chappell2,3 | Thomas W. Okell2 | Melvin Mezue2 | Andrew R. Segerdahl2 | Irene Tracey2,4 | Pedro Vilela5 | Patrícia Figueiredo1

1Institute for Systems and Robotics and Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
2Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
3Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom
4Nuffield Division of Anaesthetics, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, United Kingdom
5Imaging Department, Hospital da Luz, Lisbon, Portugal

Purpose: To assess the impact of the different post-processing options in the calibration of arterial spin labeling (ASL) data on perfusion quantification and its reproducibility.

Theory and Methods: Absolute quantification of perfusion measurements is one of the promises of ASL techniques. However, it is highly dependent on a calibration procedure that involves a complex processing pipeline for which no standardized procedure has been fully established. In this work, we systematically compare the main ASL calibration methods as well as various post-processing calibration options, using 2 data sets acquired with the most common sequences, pulsed ASL and pseudo-continuous ASL.

Results: Significant and sometimes large discrepancies in ASL perfusion quantification were obtained when using different post-processing calibration options. Nevertheless, when using a set of theoretically based and carefully chosen options, only small differences were observed for both reference tissue and voxelwise methods. The voxelwise and white matter reference tissue methods were less sensitive to post-processing options than the cerebrospinal fluid reference tissue method. However, white matter reference tissue calibration also produced poorer reproducibility results. Moreover, it may also not be an appropriate reference in case of white matter pathology.

Conclusion: Poor post-processing calibration options can lead to large errors in perfusion quantification, and a complete description of the calibration procedure should therefore be reported in ASL studies. Overall, our results further support the voxelwise calibration method proposed by the ASL white paper, particularly given the advantage of being relatively simple to implement and intrinsically correcting for the coil sensitivity profile.

KEYWORDS
ASL, calibration, cerebral blood flow, kinetic modeling, MRI
1 | INTRODUCTION

Arterial spin labeling (ASL) is a non-invasive MRI technique that provides quantitative images of tissue perfusion, by using magnetically labeled blood water protons as an endogenous blood flow tracer.\(^1\)\(^-\)\(^3\) ASL is acknowledged to have great potential as a completely non-invasive quantitative perfusion imaging technique, but its implementation has been challenging because of the intrinsically low SNR. This has motivated the development of a multitude of signal acquisition and processing strategies that aim to overcome this limitation as well as the publication of a ASL implementation consensus paper.\(^4\) Although pseudo-continuous ASL (pCASL) is the recommended labeling strategy, pulsed ASL (PASL) is still a commonly used technique.\(^5\)\(^-\)\(^9\)

In principle, cerebral blood flow (CBF) can be quantified based on a single time delay measurement—the post-labeling delay (PLD) in pCASL and the inversion time (TI) in PASL—provided that it is long enough relative to the transit time of the label bolus between the arteries and the capillaries, the so-called arterial transit time (ATT). However, if this condition cannot be assumed, which is the case in pathologies presenting delayed arterial transit times, multiple-PLD/TI measurements are necessary.\(^10\)\(^-\)\(^12\) Furthermore, if an appropriate kinetic model is fitted to the data, the multiple-PLD/TI strategy allows the assessment of not only CBF but also the publication of a ASL implementation consensus paper.\(^4\) For the practical implementation of the calibration pipeline, a number of subtler implementation choices must also be made, however, these are rarely discussed or even reported in ASL studies. Although some calibration method comparisons have been reported,\(^16\)\(^18\)\(^20\) the impact and potential pitfalls of the corresponding processing options have yet to be investigated.

Here, we systematically compare the impact of using different calibration post-processing pipelines on the quantification of perfusion and its reproducibility, using both PASL and pCASL acquisitions.

2 | THEORY

Calibration is required to obtain CBF (mL/100 g/min) in absolute units from ASL measurements, through normalization by \(M_{0x}\) (and \(\alpha\), the labeling/inversion efficiency)

\[
\text{CBF} = \frac{\text{CBF}_{\text{rel}}}{aM_{0x}} \times 6000. \tag{1}
\]

Calibration therefore entails estimation of \(M_{0x}\) and it is achieved in 2 main steps: (1) generation of \(M_{0x}\) map, and (2) derivation of \(M_{0x}\) from \(M_{0x}\) map.

2.1 | Generation of \(M_{0x}\) map

The method used for generating the \(M_{0x}\) map depends on the acquisition scheme and options chosen. There are 3 main strategies: long TR calibration scan (LongTR), ASL control averaging (CtrAvg), and control saturation recovery (SatRec).

2.1.1 | Long TR calibration scan (LongTR)

The LongTR approach is based on a separately acquired long TR scan, which approximates the equilibrium magnetization of tissue in each voxel \(i\), \(M_{0x}(i)\). This map should be corrected for the amount of \(T_1\) relaxation at the TR at each voxel, according to

\[
M_{0x,\text{corr}}(i) = \frac{M_{0x}(i)}{1 - e^{-\frac{TR_{\text{long}}}{T_1(i)}}} \tag{2}
\]

where \(T_{1,i}\) corresponds to the tissue-specific \(T_1\) in each voxel \(i\) and \(TR_{\text{long}}\) is the corresponding TR of the \(M_{0x}\) scan. This correction depends on a separate tissue segmentation procedure. Using a long TR minimizes the impact of the \(T_1\) correction, which may become negligible for sufficiently long TR values.

2.1.2 | Averaging of control images (CtrAvg)

If no background suppression is applied, it is also possible to estimate \(M_{0x}\) based on the control images. This can be carried out by averaging control images at a fixed TI (\(M_{\text{ctrl}}(i)\)), which should then be corrected for the amount of \(T_1\) relaxation...
during TI, at each voxel, to yield a corrected map of the tissue equilibrium magnetization

\[ M_{0,\text{corr}}(i) = \frac{M_{\text{ctrl}}(i)}{1 - A e^{-\frac{i}{T_1}}} \]  

(3)

Furthermore, in this approach, the inversion–saturation efficiency \((A)\) is also a variable. A value of 0.90 was chosen as it was the approximate average value of \(A\) estimated in SatRec.

For the generation of the \(M_{0t}\) map, one can argue that this method is less appropriate, mainly because the value of TI is not chosen for this purpose and therefore it is not sufficiently long to minimize the dependency on the values of \(T_1\) and \(A\).

### 2.1.3 Saturation recovery of control images (SatRec)

If no background suppression is applied, it is possible to estimate \(M_{0t}\) by fitting a saturation-recovery curve to the series of control images, \(M_{\text{ctrl}}(T_1)\), if multiple TIs are sampled and the acquisition sequence includes presaturation (which is the case of Q2TIPS PASL).\(^{17}\) This strategy allows estimation of \(M_{0t}\) as well as \(T_1\) and saturation efficiency \((A)\) (“asl_calib”, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BASIL/).

### 2.2 Derivation of \(M_{0t}\) from \(M_{0u}\) map

Once an \(M_{0u}\) map is generated, the next step is the derive \(M_{0t}\). Two types of methods are commonly used for that purpose: the reference tissue (RT) and the voxelwise (Voxel).

#### 2.2.1 Reference tissue methods (RT-CSF, RT-WM, RT-GM)

The reference tissue (RT) approach consists of computing a single \(M_{0t}\) value based on a homogeneous tissue region (GM, WM, or CSF). The first step is to correct the \(M_{0t}\) (as well as CBF) maps for the RF head coil sensitivity profile. In case the coil sensitivity profile is not available, an approximate correction may be carried out by normalizing the \(M_{0t}\) (and CBF) maps using a bias image estimated directly from the data using appropriate post-processing tools (e.g., FAST).\(^{21}\)

The next step is to define an appropriate RT mask across which an average \(M_{0t}\) value is computed, \(<M_{0t}>_{rt}\). This usually involves tissue segmentation based on a structural image (\(T_1\)-weighted image or \(T_1\), map in case this has been estimated) followed by registration of the RT mask with the \(M_{0t}\) map. Finally, the \(M_{0t}\) value is obtained by normalizing \(<M_{0t}>_{rt}\) with the respective \(\lambda\), while correcting for the difference in \(T_2^*\) relaxation between the reference tissue and arterial blood (associated with the \(M_{0t}\) measurement), yielding

\[ M_{0t} = \frac{<M_{0t}>_{rt} e^{\frac{TE}{T_2^*} - \frac{i}{T_2^*}}}{\lambda_{rt}} \]

(4)

where \(T_2^*\) is the transverse relaxation time of arterial blood, \(T_2^*\) is the transverse relaxation time of the reference tissue, and \(\lambda_{rt}\) is the blood–brain water partition coefficient of the corresponding reference tissue.

#### 2.2.2 Voxelwise method (Voxel)

In this method, an \(M_{0t}\) map is obtained through direct extrapolation from the \(M_{0t}\) value in each voxel, by normalizing with the respective tissue \(\lambda\), while correcting for the difference in \(T_2^*\) relaxation between the respective tissue and arterial blood, according to

\[ M_{0t}(i) = \frac{M_{0t}(i) e^{\frac{TE}{T_2^*} - \frac{i}{T_2^*}}}{\lambda_i} \]

(5)

where \(T_2^*\) and \(\lambda_i\) are the transverse relaxation time and the blood–brain water partition coefficient, respectively, of the tissue in each voxel. Regarding \(\lambda_i\), in principle, a different \(\lambda\) value exists in each voxel because \(\lambda\) differs as a function of tissue type and hence the voxel’s partial volume estimates (PVEs). For that reason, \(\lambda_i\) should be a PVE-weighted average of \(\lambda\). PVEs are usually obtained through tissue segmentation using appropriate tools (e.g., FAST).\(^{21}\) The \(M_{0t}\) map is then smoothed, to minimize noise contributions.

### 2.3 Simplifications

A number of assumptions and approximations to the theoretically ideal calibration just described are usually undertaken to overcome additional complexities in both data acquisition and processing. These are described here for each of the main calibration methods, and the respective post-processing options that are tested in this work are indicated, relative to the default options derived from the theoretical considerations above. All options are summarized in Table 1.

#### 2.3.1 Generation of \(M_{0t}\) map

**Long TR calibration scan (LongTR)**

The current recommendation indicates that the \(T_1\) relaxation correction should only be carried out if TR <5 s.\(^{4}\) However, the consensus paper also recommends the use of a voxelwise approach, in which case this may not have a significant impact (as we will show). We therefore tested the impact of not correcting the \(M_{0t}\) images for \(T_1\) relaxation (vs. correcting, the default).
Control averaging (CtrAvg)
One of the crucial points when using the CtrAvg approach is the value of the TI because of its strong interaction with the $T_1$ correction. We therefore compared using a longer (default; 2400 ms) and a shorter (800 ms) value, while correcting or not correcting for $T_1$.

When acquisition sequences include presaturation pulses (such as Q2TIPS PASL), quantification of the presaturation efficiency, $A$, might be an important factor. For this approach, we compared $A$ of 90% (default; the approximate average value of $A$ estimated in SatRec, see below) with $A$ of 100%.

Control saturation recovery (SatRec)
In the SatRec approach it is also possible to estimate $A$ as part of model fitting, which should be the ideal option. Therefore, we compared estimation of $A$ (default) with having a fixed $A$ of 100%.

### Table 1: Summary of calibration methods and post-processing options tested

| Method       | Options |
|--------------|---------|
| **Generation of $M_0$ map** |
| PASL-SatRec  | presaturation efficiency       | $A$ estimated (default) |
|              |                                   | $A = 100\%$ |
| PASL-CtrAvg  | TI value                          | long TI ~2400 ms (default) |
|              |                                    | short TI ~800 ms |
|              | $T_1$ correction                   | yes (default) |
|              |                                    | no |
|              | presaturation efficiency           | $A = 90\%$ (default) |
|              |                                    | $A = 100\%$ |
| pCASL-LongTR | $T_1$ correction                   | yes (default) |
|              |                                    | no |
| **Derivation of $M_0$ from $M_0$ map** |
| Reference tissue CSF | RT mask                          | restrictive (default) |
|              |                                   | intermediate |
|              |                                   | extensive |
|              | bias correction                    | yes (default) |
|              |                                    | no |
| Reference tissue WM | RT mask                           | restrictive (default) |
|              |                                   | extensive |
|              | bias correction                    | yes (default) |
|              |                                    | no |
| Reference tissue GM | RT mask                           | restrictive (default) |
|              |                                   | extensive |
|              | bias correction                    | yes (default) |
|              |                                    | no |
| Voxel       | smoothing                          | FWHM = 10.5 mm (default) |
|              |                                   | FWHM = 17.5 mm |
|              |                                    | no |
|              | $\lambda$ value                   | PVE-weighted average ($\lambda_{pve}$) (default) |
|              |                                   | brain average ($\lambda_{avg} = 0.9$) |
|              |                                    | tissue-specific ($\lambda_{tpec}$) |
|              |                                    | $\lambda_{pve} (i) = \lambda_{GM} PVE_{GM} (i) + \lambda_{WM} PVE_{WM} (i) + \lambda_{CSF} PVE_{CSF} (i)$ |
|              |                                    | $\lambda_{avg} (i) = \lambda_{GM} \times 0.5 + \lambda_{WM} \times 0.5$ |
|              |                                    | $\lambda_{tpec} (i) = \begin{cases} \lambda_{GM} & \text{if GM} \\ \lambda_{WM} & \text{if WM} \\ \lambda_{CSF} & \text{if CSF} \end{cases}$ |
2.3.2 Derivation of $M_{0a}$ from $M_{0o}$ map

Reference tissue methods (RT-CSF, RT-WM, RT-GM)
Whenever using the RT method, the choice of a tissue region of interest may range between being very restrictive to being more inclusive, resulting in considerably different PVEs and potentially very different calibrations. Here, we explored the impact of using a restrictive mask (default) with less restrictive masks. In particular, a restrictive tissue mask was obtained from the respective PVE maps in ASL space, through the application of stringent thresholds ($PVE_{\text{tissue}} \geq 0.9$). Additionally, the CSF mask was further intersected with a mask of the lateral ventricles derived from the MNI atlas.22 Furthermore, regarding coil sensitivity correction, we considered the bias field approach as default (see Discussion) and compared this with the option of no correction.

Voxelwise method (Voxel)
In the Voxel method, smoothing of the $M_{0o}$ map by a median filter with FWHM = 10.5 mm ($3 \times 3$) was considered default and was compared with using a median filter with FWHM = 17.5 mm ($5 \times 5$) as well as with the option of no smoothing. Regarding $\lambda$, the ideal PVE-weighted $\lambda$ (default) was compared with using a single brain value ($0.9$), the average of GM and WM $\lambda$ values, and a tissue-specific $\lambda$ (Table 1). For illustration, a summary of the Voxel method post-processing and options is displayed in Figure 1.

3 METHODS

The impact of the different calibration methods and respective post-processing options was evaluated in terms of average GM CBF quantification as well as its test–retest reproducibility.

3.1 Data acquisition
Test–retest multiple-PLD/TI pCASL and PASL data sets were previously acquired.15,23 The PASL study was approved by the Hospital da Luz Ethics Committee, and all subjects gave written informed consent in accordance with the Declaration of Helsinki. The pCASL study was carried out under an agreed technical development protocol approved by the Oxford University Clinical Trials and Research Governance office, in accordance with International Electrotechnical Commission and United Kingdom Health Protection Agency guidelines.

In the PASL study, 9 healthy volunteers ($22.9 \pm 5.6$ y, 4 males) were studied on a 3T Siemens Verio whole-body MRI system (Erlangen, Germany) using a 12-channel receive head RF coil, on 2 sessions separated by ~1 wk.15 In the pCASL study, 8 healthy volunteers ($28.3 \pm 2.5$ y, 6 males) were studied on a 3T Siemens Verio whole-body MRI system (Erlangen) using a 32-channel receive head RF coil, on 3 occasions separated by 1 wk and 1 mo.23 A reference image with no labeling or background suppression, TR = 6s, and all other parameters identical to the pCASL scan, was collected for calibration. In both data sets, $T_1$-weighted structural images were acquired from each subject for tissue segmentation and registration purposes. Acquisition details for both data sets are summarized in Table 2.

3.2 Data analysis
Image analysis was conducted using FSL (FSL5.0.1, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL) and MATLAB (2013a, http://mathworks.com). The multiple-PLD/TI ASL images in each data set were aligned with each other by motion correction using MCFLIRT.24 For each data set, at each PLD/TI, the control images and the pairwise differences between control and label images were averaged across repetitions, yielding

**FIGURE 1** Illustrative individual example of the voxel $M_{0a}$ derivation post-processing options
time series of mean control images \( M_{\text{ctrl}} \) (PLD/TI) and mean magnetization difference images \( \Delta M_{\text{diff}} \) (PLD/TI) as a function of PLD/TI, respectively. Additionally, for the PASL data set, off-resonance effects caused by imperfect inversion slice profile in 2D multi-slice imaging were corrected.\(^{11}\)

The structural images were segmented using FAST\(^{21}\) to estimate tissue masks and corresponding partial volume estimate (PVE) maps. Both types of image were co-registered to the ASL space of each subject and session using a linear transformation (FLIRT)\(^{24}\) (see Registration and Tissue Segmentation in the Supporting Information). For the PASL data sets, the \( T_1 \) map derived from the saturation recovery approach was used as reference image, instead of the ASL data, because it provided more tissue contrast. An extended kinetic model including a tissue contribution and an intravascular arterial compartment was fitted to \( \Delta M_{\text{diff}} \) (PLD / TI), yielding maps of relative CBF and aBV, and ATT\(^{14,25}\) (see Kinetic Modeling in the Supporting Information).

### 3.3 Calibration comparisons

Calibration of the data in each subject and session was carried out using the combinations of options and parameters listed in Tables 1 and 3. For the purpose of comparing calibration options, GM-average CBF values were obtained using a GM mask derived for each subject and session using the GM mask obtained from segmentation, co-registered to ASL space. Statistically significant differences between methods and options were tested using repeated-measures ANOVA. When significant effects were found, post-hoc analysis using pairwise \( t \) tests were carried out (\( P < 0.05 \)). The reproducibility of the GM average CBF measurements was also assessed, by computing the inter- and intra-subject coefficients of variation (CV\(_{\text{inter}}\) and CV\(_{\text{intra}}\)) for the median CBF values across GM (see Coefficients of Variation in the Supporting Information).\(^{15,23,26,27}\)

### 4 RESULTS

First, we present the quantification and reproducibility results obtained for the main calibration methods using the default post-processing options. Second, the impact of variations in the post-processing options is assessed.

#### 4.1 Comparison between calibration methods

Figure 2 shows illustrative examples of individual CBF maps obtained using the SatRec and CtrAvg \( M_{\text{ftr}} \) generation methods in PASL and the LongTR \( M_{\text{ftr}} \) generation method in pCASL and each of the \( M_{\text{ftr}} \) derivation methods (RT-CSF/RT-WM/RT-GM/Voxel) with the corresponding default post-processing options. Minor differences can be observed in the Voxel compared to the RT methods, particularly in the...
borders of CBF maps. This can be explained by the amplification of noise that occurs on the voxelwise division by the $M_{0t}$ map carried out in the Voxel method.

The group results for the average GM CBF values, obtained using the different calibration methods with the respective default options, for the first session of each data set, are presented in Figure 3. Small CBF differences across the different $M_{0a}$ derivation methods were observed although these were significantly different at times. In particular, for PASL SatRec, only RT-CSF and RT-WM methods were significantly different from the Voxel method. For PASL CtrAvg, CBF values were significantly different across methods except between RT-CSF and RT-WM methods. For pCASL LongTR approach, only CBF values from the RT-CSF method were significantly different from the other $M_{0a}$ derivation methods.

Finally, the reproducibility metrics of the average GM CBF values for the main calibration methods, using the default post-processing options, are presented in Figure 4. All methods were found to exhibit good reproducibility, with pCASL being superior to PASL, and the RT-WM method generally yielding the worst values. A significant main effect of $M_{0a}$ derivation method was found for each $M_{0a}$ map generation method. Subsequent post-hoc analysis yielded significant differences between several $M_{0a}$ derivation methods. The Voxel and RT-CSF methods yielded the lowest $CV_{intra}$ values, whereas the Voxel and RT-GM methods yielded the lowest $CV_{inter}$ values, with significant differences in several cases. The RT-WM method systematically carried out worst for both $CV_{intra}$ and $CV_{inter}$, often with significant differences.

### 4.2 Impact of calibration post-processing options

The group results for the GM average CBF values obtained when varying the calibration post-processing options are presented in Figure 5, for each of the main calibration methods tested on session 1 of each data set. For each method, there was a significant main effect of the post-processing option; subsequent pairwise comparisons revealed significant differences between some options, as indicated in the plots.

When applying the SatRec method for $M_{0a}$ generation, the RT-CSF approach yielded the highest CBF differences across methods, with the extensive mask producing significantly greater CBF values relative to the default option that
uses a much more restrictive mask. WM masking options resulted in a significant CBF decrease, but only of ~5% relative to the default value. Fixing $A = 100\%$ instead of allowing its estimation ($A \sim 90\%$) yielded significant CBF differences relative to default option for all methods except for RT-GM. Not correcting for field inhomogeneities using bias correction had small and non-significant impact on CBF quantification, although the impact was higher when using the RT-CSF approach compared with the other RT approaches. For the Voxel approach, an increase in kernel size for spatial
smoothing led to significant but small effects on CBF. Not applying any smoothing at all yielded non-significant differences. Finally, different options for the value of $\lambda$ also resulted in significant but small CBF differences when using an average $\lambda$, whereas non-significant differences were observed when using a tissue-specific $\lambda$.

In contrast with the SatRec approach, the CtrAvg approach revealed great sensitivity to some post-processing options, particularly the $T_1$ correction, and more so when using the RT-CSF method. Significant increases in CBF as high as ~100% were found when using the RT-CSF method if no $T_1$ correction was carried out. When using $TI = 800$ ms, significant CBF differences were found for RT-WM and RT-GM methods compared to the default option.

When applying the pCASL LongTR approach for $M_0$ generation, differences in CBF values were obtained with the different options, particularly when using the RT-CSF method. When not performing $T_1$ correction, the RT-CSF method yielded a significantly greater difference compared with the default options, with the Voxel method yielding a significant but much smaller difference. Consistently with the PASL SatRec results, less restrictive masks led to important and significant CBF differences, except for the RT-GM method, particularly in the case of the extensive CSF mask. Also, similarly with PASL SatRec but to a higher extent, bias correction had greater and always significant impact on all RT methods, particularly on RT-CSF and RT-GM. When using the Voxel method, the different spatial smoothing options yielded non-significant differences relative to the default option. Non-significant differences were also observed when using a tissue-specific $\lambda$ but, in contrast, average $\lambda$ yielded a statistically significant difference.

5 | DISCUSSION

We systematically compared the main ASL calibration methods as well as the associated post-processing options, in terms of CBF values and corresponding test–retest reproducibility.

5.1 | Comparison between calibration methods

When comparing the different calibration methods using the default post-processing options, minimal differences were observed in CBF within each data set (Figure 2). When focusing on the GM average CBF values (Figure 3), relatively small differences were found despite being significant in some cases.

Overall, the RT-CSF method yielded the lowest CBF values across $M_0$ estimation methods. Although, for the PASL SatRec approach, only RT-CSF and RT-WM methods were significantly different from the Voxel method, for the pCASL LongTR approach, significantly lower CBF values were obtained when using the CSF as a reference tissue, compared with other strategies for deriving $M_{0a}$ from $M_0$. This finding may be explained by the fact that our pCASL data set, obtained using a 32-channel head coil, suffered from $B_1$ field inhomogeneities that were not fully accounted for by bias field correction. In fact, this interpretation is corroborated by the observation that, in the case of pCASL, when using less restrictive CSF masks (including the whole lateral ventricles rather than just a few central voxels where coil sensitivity is much lower), the CBF values increased considerably and became comparable to the ones obtained using the Voxel approach.

It should be noted that the bias field correction approach used here is not the ideal method for accounting for the coil sensitivity profile. Nevertheless, this strategy serves the purpose of evaluating the relative differences between calibration methods and post-processing options. In fact, it clearly emphasizes the impact of sensitivity correction on CBF, particularly when using the RT-CSF approach. Furthermore, the CBF values obtained with RT methods, when corrected for the bias field, become closer to the ones obtained with the Voxel method, which intrinsically corrects for this issue. Nevertheless, further work is necessary to fully estimate the impact of this strategy on calibration methods. Methods for the estimation of the coil sensitivity profile have been proposed, but these require the acquisition of additional data using imaging sequences that are not widely available. Alternatively, a proton density image acquired with a body coil can also be used to obtained an approximate sensitivity map. However, this method is limited by the fact that it only corrects for non-uniformity in the receive field, not the transmit field, which is also not spatially uniform at field strengths of 3T or higher, and this correction may not account for other factors (e.g., prescan transmit settings).

It is possible to circumvent RF field inhomogeneity issues altogether by using the voxelwise method. In this case, $M_{0a}$ is computed at each voxel based on the measured $M_0$ image and it is therefore affected by the same RF coil sensitivity as the ASL measurements themselves. Normalization of the estimated CBF map by the $M_0$ map therefore intrinsically corrects for coil sensitivity variations. This well-known effect has been explicitly reported, with observed signal loss in anterior regions on both proton-density and perfusion-weighted ASL images being successfully restored in the final perfusion maps by calibration using a Voxel approach. Nevertheless, the voxelwise division intrinsic to this approach might lead to increased variability in CBF quantification because of noise amplification and also to edge and/or partial volume effects.
5.2 | Impact of calibration post-processing options

We observed that some of the processing options had great impact on the final results whereas others were relatively unimportant. Interestingly, very few studies have investigated the impact of different calibration strategies, and critically, they are not in agreement. One study compared the 3 $M_0$ derivation schemes that we also tested here (RT-CSF, RT-WM, and Voxel) using data collected with 3 single-TI (1400 ms) PASL sequences (including PICORE-Q2TIPS) and found ~35% higher CBF values for the RT-CSF relative to the Voxel method. Another study also observed a discrepancy in CBF quantification when using different calibration methods based on single-PLD pCASL data, with RT-WM yielding the lowest CBF values and Voxel the highest (~20% difference). A third study compared different calibration methods on CASL and pCASL data and found only minor differences. However, they reported slightly lower CBF values using RT-CSF compared with RT-WM possibly because of not correcting for the uneven sensitivity profile of the head coil.

Our results highlight that one of the processing options with greatest impact was the correction of the $M_0$ map for incomplete $T_1$ relaxation in each tissue based on the TR of the respective image acquisition, particularly when using the RT-CSF method. In the current guidelines, this correction is only recommended when TR < 5 s. However, this recommendation is coupled with the use of the Voxel method for extrapolating $M_0$, in which case we do not observe large effects of $T_1$ correction. In fact, the impact of $T_1$ correction is much greater if the RT-CSF derivation method is used, because CSF has a much longer $T_1$ than GM or WM. In this case, a TR longer than 16 s would be required for an almost full recovery of the longitudinal magnetization. Therefore, we believe that this correction should be applied even if TR is longer than 5 s, when the RT-CSF calibration method is used. A related option with relatively high impact is the value of PLD/TI used to extract $M_0$. In the case of the CtrAvg approach in PASL, using a lower TI leads to greater sensitivity in CBF quantification. For PASL, the presaturation efficiency value was also important, because this is sequence-specific and therefore should be chosen accordingly. The impact of the bias correction approach is highly dependent, not only on the $M_0$ derivation method, but also on the type of ASL data set and the RF coil sensitivity of the acquisition. In contrast with the 12-channel RF coil used for the PASL data set, the 32-channel RF coil used for the pCASL data set displayed a relatively heterogeneous sensitivity profile, leading to significant impact of bias correction, particularly when using RT-CSF and RT-GM methods.

Overall, our results indicate that when acquiring multiple-TI PASL data, the SatRec approach is preferable to the CtrAvg approach because the latter strongly depends on the choice of PLD/TI value and on $T_1$ correction. In terms of $M_0$ derivation, the RT-CSF approach was the most sensitive to variations in the processing options, making this approach more prone to CBF discrepancies. In addition, variability in the CBF differences across subjects was also greater than with other $M_0$ derivation methods. The RT-CSF method is highly dependent, not only on $T_1$ correction as discussed before, but also on the tissue mask and whether or not bias correction is carried out. The RT-WM and RT-GM methods
are also dependent on these parameters but to a lesser extent. Notwithstanding, the RT-WM approach might not be feasible whenever WM lesions are present, such as in multiple sclerosis or small vessel disease. On the other hand, the RT-GM approach is highly influenced by PVE, rendering the use of restrictive GM masks mandatory. Nevertheless, approaches for PVE correction most often target only the relative CBF images and not $M_0$ measures. Therefore, we believe that this correction should not affect our results concerning the comparison of calibration methods. Similarly to the RT-WM and RT-GM methods, the Voxel method yielded relatively small CBF differences across different processing options. In fact, some options had negligible impact on CBF quantification, namely the degree of smoothing. As expected, the use of an averaged, fixed $\lambda$ underestimates GM CBF. A summary scheme of recommended calibration procedures, as well as the corresponding options that should be applied and/or reported, is depicted in Figure 6.

5.3 CBF values, spatial distributions and reproducibility

The maps obtained in our study exhibit spatial distributions that are consistent with the expected variations across tissues and brain regions. All average GM CBF values found in our study were within the wide range of values reported in the literature. However, the PASL data set systematically produced lower CBF values than the pCASL data set. Surprisingly, in a study using 5 different commonly used ASL sequences, the opposite pattern was seen between single-PLD pCASL and multiple-TI PASL using the RT-CSF method (25/40 mL/100 g/min for pCASL/PASL). In another study, similar values were found for both PASL and pCASL single-PLD/TI sequences (62/60 mL/100 g/min for PASL/pCASL), using different calibrations strategies (extra calibration scan for PASL/control averaging for pCASL). The differences between pCASL and PASL observed here are of unknown cause and are likely explained partially by specific implementation issues.

The inter- and intra-subject coefficients of variation obtained in our study for GM average CBF are all within intervals of good reproducibility. For PASL, $CV_{\text{intra}} \sim 10$–12% is in agreement with the literature. A multiple-TI QUASAR ASL study using a model-free approach reported within-week $CV_{\text{intra}} \sim 10\%$, whereas a study from our group achieved within-week $CV_{\text{intra}}$ values in the range of 10–21%. A single-TI PASL study, using a sequence similar to the 1 used in our study, found within-week $CV_{\text{intra}} \sim 6\%$. Regarding pCASL, our results of ~6–7% are also in accordance with the literature. A single-PLD pCASL study using 2 different PLD values yielded $CV_{\text{intra}} \sim 5$–10%, which increased with PLD. Consistently, a multiple-PLD pCASL study obtained $CV_{\text{intra}} \sim 5\%$ for average week-repeat GM CBF value.

5.4 Limitations and future work

Although our approach aimed to systematically and comprehensively test all relevant processing options in the calibration methods used, options regarding relatively less important parameters could also be tested, in particular the tissue-specific $T_2^*$ used to correct for $T_2^*$ decay when deriving the $M_0$ maps. Another option that could be tested is the use of subject-specific or voxelwise $T_1$ values, which could be estimated directly when using the SatRec approach in PASL or acquired separately. Importantly, some of the calibration methods may not be appropriate to use in certain clinical populations. For example, in patients with WM abnormalities, such as in multiple sclerosis or small vessel disease, the RT-CSF or RT-GM methods may be preferable to the RT-WM.

Furthermore, we acknowledge that some options could not be tested because of the lack of data availability. In particular, the $LongTR$ approach should be also compared using PASL, to determine the best method for $M_0$ generation. Additionally, a more precise approach for correction of the coil sensitivity profile should be carried out. Furthermore, we also acknowledge that the EPI images suffer from susceptibility-induced geometric distortions that might affect image co-registration. This can be circumvented by acquiring a $B_0$ field map and using this to correct such distortions. Finally, a prospective study assessing the impact of calibration strategies by acquiring both PASL and pCASL on the same subjects should be conducted to further validate our results.

6 CONCLUSION

In conclusion, we found that considerable discrepancies in CBF values can be obtained when using poor choice simplifications in the post-processing calibration of pCASL and PASL data. Nevertheless, strategies based on reference tissue or voxelwise methods both perform well when carefully implemented in terms of CBF quantification and reproducibility. In general, the greatest sensitivity was found for correction for incomplete $T_1$ relaxation when using proton density reference images for calibration. Correction for RF field inhomogeneities also had great impact, as did the value of presaturation efficiency, whereas the values of brain–blood water partition coefficient had moderate impact. In contrast, the degree of spatial smoothing applied to the calibration images or the mask used for the reference tissue had minor effects. The voxelwise approach, as proposed in the ASL white paper, has the advantage of simplicity, while the reference tissue methods are more vulnerable to sensitivity and relaxation correction errors, in particular when using CSF. Nevertheless, our results emphasize the need for consistent post-processing options across studies as well as the need for a complete description of the various post-processing

$CV_{\text{intra}}$ within-week CBF values in the range of 10–21%. A single-TI PASL study, using a sequence similar to the 1 used in our study, found within-week $CV_{\text{intra}} \sim 6\%$. Regarding pCASL, our results of ~6–7% are also in accordance with the literature. A single-PLD pCASL study using 2 different PLD values yielded $CV_{\text{intra}} \sim 5$–10%, which increased with PLD. Consistently, a multiple-PLD pCASL study obtained $CV_{\text{intra}} \sim 5\%$ for average week-repeat GM CBF value.
calibration options so that absolute CBF quantification is effectively achieved. Failure to take the impact of these fine calibration options into account would seriously compromise the use and applicability of ASL perfusion imaging.

**ACKNOWLEDGMENTS**

This work was supported by the Portuguese Foundation for Science and Technology (FCT) grants PTDC/BBB-IMG/2137/2012, PD/BD/135114/2017 and UID/EEA/50009/2019; the Engineering and Physical Sciences Research Council UK (EP/P012361/1) and the Wellcome Trust (203139/Z/16/Z). The authors wish to express their gratitude to the European Cooperation in Science and Technology-Arterial Spin Labelling Initiative in Dementia (COST-AID Action BM1103). We are also grateful for the insightful comments and recommendations of the Referees.

**DATA AVAILABILITY STATEMENT**

The underlying data associated with figures in this study are available from the Oxford University Research Archive (ORA-Data) (DOI: 10.5287/bodleian:xQx8Vnv5e). For enquiries regarding the imaging data contact: ARM for the pCASL data (andrew.segerdahl@ndcn.ox.ac.uk) or PF for the PASL data (patricia.figueiredo@tecnicoulisboa.pt).

**ORCID**

Joana Pinto [https://orcid.org/0000-0002-6863-1883](https://orcid.org/0000-0002-6863-1883)

Michael A. Chappell [https://orcid.org/0000-0003-1802-4214](https://orcid.org/0000-0003-1802-4214)

Thomas W. Okell [https://orcid.org/0000-0001-8258-0659](https://orcid.org/0000-0001-8258-0659)

Patrícia Figueiredo [https://orcid.org/0000-0002-0743-0869](https://orcid.org/0000-0002-0743-0869)

**REFERENCES**

1. Detre JA, Leigh JS, Williams DS, Koretsky AP. Perfusion imaging. *Magn Reson Med.* 1992;23:37–45.
2. Williams DS, Detre JA, Leigh JS, Koretsky AP. Magnetic resonance imaging of perfusion using spin inversion of arterial water. *Proc Natl Acad Sci U S A.* 1992;89:212–216.
3. Alsop DC, Detre JA. Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. *J Cereb Blood Flow Metab.* 1996;16:1236–1249.
4. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med.* 2015;73:102–116.
5. Wang Y, Saykin AJ, Pfeuffer J, et al. Regional reproducibility of pulsed arterial spin labeling perfusion imaging at 3T. *NeuroImage.* 2011;54:1188–1195.
6. Gevers S, van Osch MJ, Bokkers RPH, et al. Intra- and multicenter reproducibility of pulsed, continuous and pseudo-continuous arterial spin labeling methods for measuring cerebral perfusion. *J Cereb Blood Flow Metab.* 2011;31:1706–1715.
7. Mutsaerts HJMM, Steketee RME, Heijtel DF, et al. Reproducibility of pharmacological ASL using sequences from different vendors: implications for multicenter drug studies. *MAGMA.* 2015;28:427–436.
8. Gil-Gouveia R, Pinto J, Figueiredo P, Vilela PF, Martins IP. An arterial spin labeling MRI perfusion study of migraine without aura attacks. *Front Neurol.* 2017;8:280.
9. Pimentel M, Vilela P, Sousa I, Figueiredo P. Localization of the hand motor area by arterial spin labeling and blood oxygen level-dependent functional magnetic resonance imaging. *Hum Brain Mapp.* 2013;34:96–108.
10. Machnitsch BJ, Lindsay AC, Kylintireas I, et al. Multiple inflow pulsed arterial spin-labeling reveals delays in the arterial arrival time in minor stroke and transient ischemic attack. *Am J Neuroradiol.* 2010;31:1892–1894.
11. Figueiredo P, Clare S, Jezzard P. Quantitative perfusion measurements using pulsed arterial spin labeling: effects of large region-of-interest analysis. *J Magn Reson Imaging.* 2005;21:676–682.
12. Gonzalez-At JB, Alsop DC, Detre JA. Cerebral perfusion and arterial transit time changes during task activation determined with continuous arterial spin labeling. *Magn Reson Med.* 2000;43:739–746.
13. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magn Reson Med.* 1998;40:383–396.
14. Chappell MA, MacIntosh BJ, Donahue MJ, Günther M, Jezzard P, Woolrich MW. Separation of macrovascular signal in multi-inversion time arterial spin labeling MRI. *Magn Reson Med.* 2010;63:1357–1365.
15. Sousa I, Vilela P, Figueiredo P. Reproducibility of the quantification of arterial and tissue contributions in multiple postlabeling delay arterial spin labeling. *J Magn Reson Imaging.* 2014;40:1453–1462.
16. Wu WCC, St Lawrence KS, Lich DJ, Wang DI. Quantification issues in arterial spin labeling perfusion magnetic resonance imaging. *Top Magn Reson Imaging.* 2010;21:65–73.
17. Luh WM, Wong EC, Bandettini PA, Hyde JS. QUIPSS II with thin-slice T1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. *Magn Reson Med.* 1999;41:1246–1254.
18. Cävusoğlu M, Pfeuffer J, Ügurbil K, Uluadž K. Comparison of pulsed arterial spin labeling encoding schemes and absolute perfusion quantification. *Magn Reson Imaging.* 2009;27:1039–1045.
19. Chen Y, Wang Z, Detre JA. Impact of equilibrium magnetization of blood on ASL quantification. In *Proceeding of the 19th Annual Meeting of ISMRM*, Montreal, Canada, 2011. p. 300.
20. Fazlollahi A, Bourgeat P, Liang X, et al. Reproducibility of multiphase pseudo-continuous arterial spin labeling and the effect of post-processing analysis methods. *NeuroImage.* 2015;117:191–201.
21. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging.* 2001;20:45–57.
22. Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 3D statistical neuroanatomical models from 305 MRI volumes. In *Proceedings of the IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, San Francisco, CA, 1993. p. 1813–1817.
23. Mezue M, Segerdahl AR, Okell TW, Chappell MA, Kelly ME, Tracey I. Optimization and reliability of multiple postlabeling
delay pseudo-continuous arterial spin labeling during rest and stimulus-induced functional task activation. *J Cereb Blood Flow Metab.* 2014;34:1919–1927.

24. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage.* 2002;17:825–841.

25. Chappell MA, Woolrich MW, Kazan S, Jezzard P, Payne SJ, MacIntosh BJ. Modeling dispersion in arterial spin labeling: validation using dynamic angiographic measurements. *Magn Reson Med.* 2013;69:563–570.

26. Pinto J, Jorge J, Sousa I, Vilela P, Figueiredo P. Fourier modeling of the BOLD response to a breath-hold task: optimization and reproducibility. *NeuroImage.* 2016;135:223–231.

27. Tjandra T, Brooks J, Figueiredo P, Wise RG, Matthews PM, Tracey I. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *NeuroImage.* 2005;27:393–401.

28. Dai W, Robson PM, Shankaranarayanan A, Alsop DC. Sensitivity calibration with a uniform magnetization image to improve arterial spin labeling perfusion quantification. *Magn Reson Med.* 2011;66:1590–1600.

29. Ahlgren A, Wirestam R, Knutsson L, Petersen ET. Improved calculation of the equilibrium magnetization of arterial blood in arterial spin labeling. *Magn Reson Med.* 2018;80:2223–2231.

30. Okell TW, Chappell MA, Kelly ME, Jezzard P. Cerebral blood flow quantification using vessel-encoded arterial spin labeling. *J Cereb Blood Flow Metab.* 2013;33:1716–1724.

31. Chappell MA, Groves AR, MacIntosh BJ, Donahue MJ, Jezzard P, Woolrich MW. Partial volume correction of multiple inversion time arterial spin labeling MRI data. *Magn Reson Med.* 2011;65:1173–1183.

32. Asllani I, Borogovac A, Brown TR. Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. *Magn Reson Med.* 2008;60:1362–1371.

33. Mutsaerts HJ, van Osch MJ, Zelaya FO, et al. Multi-vendor reliability of arterial spin labeling perfusion MRI using a near-identical sequence: implications for multi-center studies. *NeuroImage.* 2015;113:143–152.

34. Chen Y, Wang D, Detre JA. Test-retest reliability of arterial spin labeling with common labeling strategies. *J Magn Reson Imaging.* 2011;33:940–949.

35. Petersen ET, Mouridsen K, Golay X. The QUASAR reproducibility study, part II: results from a multi-center arterial spin labeling test-retest study. *NeuroImage.* 2010;49:104–113.

36. Wu B, Lou X, Wu X, Ma L. Intra- and interscanner reliability and reproducibility of 3D whole-brain pseudo-continuous arterial spin-labeling MR perfusion at 3T. *J Magn Reson Imaging.* 2014;39:402–409.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

FIGURE S1 Illustrative examples of MPRAGE images registered to ASL space (bottom) and respective CSF (blue) and GM (yellow) masks retrieved from MPRAGE image segmentation (top), for the PASL and pCASL datasets

How to cite this article: Pinto J, Chappell MA, Okell TW, et al. Calibration of arterial spin labeling data—potential pitfalls in post-processing. *Magn Reson Med.* 2019;00:1–13. [https://doi.org/10.1002/mrm.28000](https://doi.org/10.1002/mrm.28000)