Repeatability of Perfusion Measurements in Gliomas Using Pulsed and Pseudo-continuous Arterial Spin Labelling (ASL) MRI

Amirah F Alsaedi[1]; David L Thomas [2,3]; Enrico De Vita [4,7]; Jasmina Panovska-Griffiths[5,6]; Sotirios Bisdas[2,7]; Xavier Golay [2,7]

Author Affiliations:

1. Department of Radiology Technology, Taibah University, Medina, KSA.
2. Department of Brain Repair & Rehabilitation, UCL Queen Square Institute of Neurology, University College London, London, UK
3. Leonard Wolfson Experimental Neurology Centre, UCL Queen Square Institute of Neurology, University College London, London, UK
4. Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
5. The Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford
6. The Queen's College, University of Oxford.
7. Lysholm Department of Neuroradiology, The National Hospital for Neurology and Neurosurgery, University College Hospitals NHS Trust, London, UK

Corresponding Author Info:

Amirah Faisal Alsaedi

Email: afsaedi@taibahu.edu.sa

Keywords: magnetic resonance imaging, arterial spin labelling, repeatability, cerebral blood flow, tumour blood flow

Running title: ASL Repeatability in Adult Gliomas

Acknowledgments:

AFA is funded by Taibah University from the Saudi Arabia Government. DLT was supported by the UCL Leonard Wolfson Experimental Neurology Centre (PR/ylr/18575), and the Wellcome Trust (Centre award 539208). EDV is funded by the Wellcome Engineering and Physical Sciences Research Council (EPSRC) Centre for

Medical Engineering at Kings College London [WT 203148/Z/16/Z].SB is funded by National Institute for Health Research to UCLH Biomedical research centre (Grant: BRC399/NS/RB/101410). XG received funding by the European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 667510. This work received funding from the UCLH NIHR Biomedical Research Centre (DLT, SB, XG).

Abstract

Object: To investigate the repeatability of perfusion measures in gliomas using pulsed- and pseudo-continuous- arterial spin labelling (PASL, PCASL) techniques, and evaluate different regions-of-interest (ROIs) for relative tumour blood flow (rTBF) normalisation.

Materials and methods: Repeatability of cerebral blood flow (CBF) was measured in the Contralateral Normal Appearing Hemisphere (CNAH) and in brain tumours (aTBF). rTBF was normalised using both large/small ROIs from the CNAH. Repeatability was evaluated with intra-class-correlation-coefficient (ICC), Within-Coefficient-of-Variation (WCoV) and Coefficient-of-Repeatability (CR).

Results: PASL and PCASL demonstrated high reliability (ICC > 0.9) for CNAH-CBF, aTBF and rTBF. PCASL demonstrated a more stable signal-to-noise ratio (SNR) with a lower WCoV of the SNR than that of PASL (10.9%-42.5% vs. 12.3%-29.2%). PASL and PCASL showed higher WCoV in aTBF and rTBF than in CNAH CBF in WM and GM but not in the caudate, and higher WCoV for rTBF than for aTBF when normalised using a small ROI (PASL 8.1% vs. 4.7%, PCASL 10.9% vs. 7.9%, respectively). The lowest CR was observed for rTBF normalised with a large ROI.

Discussion: PASL and PCASL showed similar repeatability for the assessment of perfusion parameters in patients with primary brain tumours as previous studies based on volunteers. Both methods displayed reasonable WCoV in the tumour area and CNAH. PCASL's more stable SNR in small areas (caudate) is likely to be due to the longer post-labelling delays.

1 Introduction

Cerebral blood flow (CBF) quantification using arterial spin labelling (ASL) has been suggested to provide an imaging biomarker that can be used in both glioma grading [1,2] and for evaluation of tumour progressions following radiotherapy [3–5]. However, fundamental differences between ASL implementations (e.g. different acquisition sequences and spin labelling approaches) are known to affect the sensitivity of CBF estimation due to variation in measurement precision. Other potential sources of errors can also include scanner instability and short-term physiological fluctuations in tumour perfusion. Repeatability and reproducibility analysis are useful tools to validate the precision of the quantitative measurements in clinical trials and enable exploitation of their benefits and integration into clinical practice [6].

A previous study investigating the reproducibility of quantitative STAR labelling of arterial region (QUASAR) [7] reported within-subject standard deviations (WS-SD) due to scanner instability of 3.1 ml/100g/min, increasing to 4.3 ml/100g/min after repositioning, and to 5.3 ml/100g/min including long term physiological variations over separate measurements. The scanner instability was thus considered as the main cause of quantification error, since each further increase in WS-SD is smaller than the initial value. This result was confirmed by another study that compared the test-retest reproducibility of pulsed-ASL (PASL), continuous-ASL (CASL) and pseudo-continuous-ASL (pCASL) approaches [8]. However, one of the main issues which limits generalizability of these studies is that they have only reported ASL reproducibility in healthy volunteers [7,8]. A recent study which evaluated the intra-session reliability of PCASL in 6 patients with glioblastoma (GBM), showed high reliability in both the

normal-appearing grey matter (NAGM) (ICC>0.90, WCoV 3.40%-7.12%) and tumour (ICC 0.98; WCoV 4.91%) [9].

ASL measurements in brain tumours studies are usually expressed as absolute tumour blood flow (aTBF) or relative (or normalised) tumour blood flow (rTBF). rTBF has been shown to be more reliable than aTBF when distinguishing between high- from low-grade gliomas (HGGs and LGGs, respectively) [1,10,11]. Using a large ROI as the reference region for normalization provides excellent reliability, but can introduce inaccuracies due to the variability of the bolus arrival time (BAT) throughout the ROI [12]. Thus, the choice of tissue type to act as an internal reference and the size of the ROI used for this purpose is likely to add bias and affect the variability and repeatability of rTBF.

Our study aims to assess the repeatability of both pulsed ASL (PASL) and pseudo-continuous ASL (PCASL) with 3D GRASE and to decide which internal reference, both in terms of the ROI size and region of placement, is better for normalising the rTBF.

2 Methods

2.1 Patients

Forty adult glioma patients (WHO histological glioma grades II-IV; aged 43±14 years) were recruited in this study. All participants provided signed informed consent, and the study was approved by the local institutional research board and the ethics committee. Sixteen subjects were scanned using PASL (9 HGGs (grade III/IV), 7 LGGs (grade II)); this data was retrospectively re-analysed from a previous study comparing perfusion measurements derived from ASL, dynamic susceptibility contrast (DSC) and dynamic contrast-enhanced (DCE) in clinical neuro-oncology [13]. An additional 24 subjects

were prospectively identified and scanned using PCASL (9 HGGs (grade III/IV), 15 LGGs (grade II)). Both groups of patients were scanned on clinical 3T MR scanners (Siemens Healthineers, Erlangen, Germany).

2.2 ASL acquisitions

PASL data were acquired with a flow-sensitive alternating inversion recovery (FAIR) ASL labelling approach using a segmented 3D GRAdient- and Spin-Echo (GRASE) readout (8 shots; for other sequence parameters, see Table 1). For the calculation of CBF maps in absolute biological units (ml/100 g/min), images were acquired using the same acquisition scheme without background suppression and with three saturation recovery times (1s, 2s, and 4s). M_0 and T_1 were estimated by fitting these saturation recovery images using NiftyFit, an open-source software developed by the Centre for Medical Image Computing at University College London [14].

PCASL data was similarly acquired using a segmented GRASE readout, except with four shots (Table 1). For CBF quantification, a proton density (PD) image was acquired as the first volume of the series, with the same imaging readout but without labelling or background suppression.

Table 1 Imaging parameters of the investigated 3D ASL protocols		
	PASL (Siemens product)	PCASL (Siemens WIP)
Scanner	Siemens 3T Skyra	Siemens 3T Prisma
Number of segments	8	4
TR/TE (ms)	3500/12	4600/13.28
Bolus duration (ms)	1000 (estimated)*	1800
Post-labelling delay time (PLD/TI) (ms)	1800	1800
FOV (mm)	240 × 240	220 x 220
Matrix size	64 × 52	64 x 60
Voxel size (mm^3)	3.75 × 4.63 × 5.25	3.44 x 3.67 x 4
Slice thickness (mm)	5.25	4
Slices per slab	20	28
Acquisition Time	3min 52s	5min 15s
Refocusing flip angle (°)	130	130
Turbo / EPI factors	15/13	28/15
Background suppression scheme (BS)	2 Inversion pulses with timings calculated to null static tissue magnetization with T1=700ms and 1400ms 100ms before acquisition	same as for PASL
Number of repeats/averages	4	8
Echo spacing	0.54 ms	0.51 ms

* Note that due to a technical issue, the bolus duration was estimated for the PASL datasets.

2.3 CBF map calculation

After applying motion correction to the raw data using MCFLIRT [15] (from FMRIB Software Library (FSL)), ASL difference images (control – label) (ΔASL) were calculated. This resulted in four ΔASL for PASL, and six ΔASL for PCASL, because the first two measurements were removed from all PCASL data sets due to the presence of CSF artefacts. CBF maps are typically calculated by averaging over several repeats to increase SNR. Here, rather than measuring the repeatability between the CBF maps calculated from each measurement (single ΔASL), CBF maps

were generated by averaging different combinations of single ΔASL measurements (Figure 1), to establish a number of measurements for each patient (known as moving block bootstrapping [16]). Repeated measurements from the same subject present as a finite time series of correlated data. Typical bootstrapping resamples measurements assuming they are independent and random, however, this assumption is violated in our study. Instead, moving block bootstrapping [16] overcome this problem as it resamples the observations as blocks instead of individual observation. This creates replicated samples and increase the accuracy of sampling distribution estimate without altering the dataset.

Consequentially, each combination (or block) involved half of the acquired ΔASL measurements. The 4 ΔASL PASL images thus produced six combinations and the 6 ΔASL PCASL images produced 20 combinations. Each of these generated combinations was then averaged, and CBF maps were calculated using the simplified general kinetic model proposed in the ISMRM (International Society for Magnetic Resonance in Medicine) Perfusion Study Group consensus paper[17] (eq.1 and 2):

For FAIR-PASL:

$$\mathrm{CBF} = \frac{6000.\lambda.(\Delta\mathrm{ASL}).e^{\frac{\mathrm{TI}}{\mathrm{T1a}}}}{2.\alpha.\mathrm{TI1}.\mathrm{M0}}\ \textit{[ml/100 g/min]} \qquad \text{Equation 1}$$

For PCASL:

$$\mathrm{CBF} = \frac{6000.\lambda.(\Delta\mathrm{ASL}).e^{\frac{\mathrm{PLD}}{\mathrm{T1a}}}}{2.\alpha.\mathrm{T1}a.PD.(1-e^{\frac{-\tau}{\mathrm{T1}a}})}\ \textit{[ml/100 g/min]} \qquad \text{Equation 2}$$

Each equation contained assumed constant values, measured values for each subject, and fixed timing parameters for each acquisition. The constant values were a factor of 6000, used to convert the CBF unit of ml/g/s to ml/100 g/min; λ, the water blood/brain partition coefficient, defined to be 0.9 ml/g; α, the labelling efficiency, dependent on the labelling approach used (in this case, for PCASL = 0.85 and for PASL = 0.98); and T1a, the blood relaxation time, which was approximately equal to 1650ms at 3T [17]. Please note that in this study, the reduction in labelling efficiency due to the use of BS pulses was not considered, as it was kept constant for each method throughout the project and as such, would not affect any calculated coefficient of variation. The measured PD (PCASL) or fitted M_0 (PASL) value was used as a scaling image for the ΔASL image, enabling the absolute quantification of CBF. The fixed duration values were the labelling bolus duration (LD) (τ/TI1 for PCASL/PASL, respectively), and the post labelling delay (PLD)/inflow time (TI) for PCASL/PASL, respectively.

The PD for PCASL was multiplied by an appropriate scaling factor, as recommended by Alsop et al. [17] (eq.3) since it was acquired with TR less than 5s.

$$\mathrm{SI\_}M0 = \mathrm{SI\_}acq\,.\frac{1}{(1 - e^{\left(\frac{-TR}{T1,tissue}\right)})} \qquad \text{Equation 3}$$

The TR was 4.6s and the T1 of the grey matter (~1.3s) [18], resulting in a multiplying factor of 1.03.

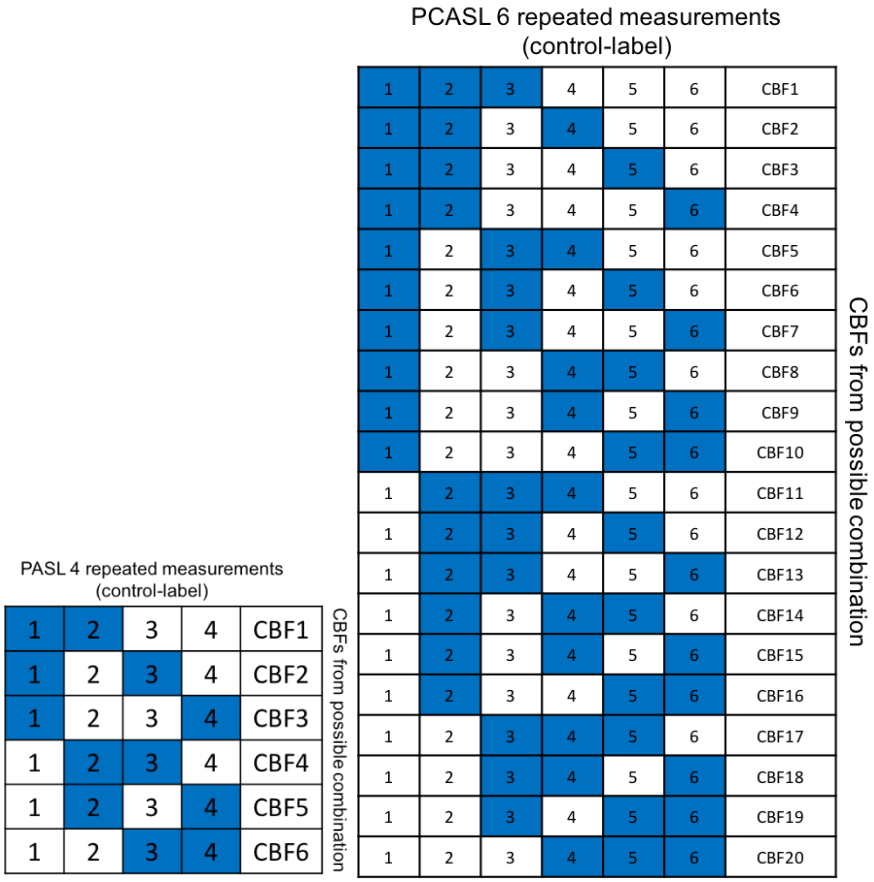


Fig 1 The permutation matrices illustrate all the possible combinations without repetition of k-element subsets (half of the repeated measurements) from a set of an element (repeated measurements, single ΔASL). The PASL has 2-elements subsets of a 4-elements set while the PCASL have 3-elements subsets of a 6-elements set. Thus, PASL dataset produces 6 combinations and the PCASL one 20 combinations of CBF maps.

2.4 Image analysis

All MR images were converted to NIfTI format using MRIconvert, (version 2.0.x, https://lcni.uoregon.edu/downloads/mriconvert/mriconvert-and-mcverter).

Conventional MR images (T1-weighted (T1-W), post-contrast T1-weighted (T1-C), T2-weighted (T2-W) and Fluid-attenuated inversion recovery (FLAIR)) were used as anatomical templates. ITK-SNAP[19] (version 3.6.0; www.itksnap.org) was used to segment the whole tumour volume manually on the T2-weighted images for PASL and the FLAIR images for PCASL, according to availability. 3D high resolution anatomical

T1-W MPRAGE images were used for automatic segmentation of the grey matter (GM) and white matter (WM) volumes in the contralateral normal-appearing hemisphere (CNAH) with the FAST tool from the FSL library [20]. Subsequently, these CNAH GM and WM volumes were used as "large ROIs" for rTBF normalisation. The caudate nucleus (caud) in the CNAH was segmented automatically from the MPRAGE using the FIRST tool from the FSL library [20], and was utilised as an example of "small ROI" for rTBF normalisation. Care was taken to avoid the tumour if it crossed the midline to the contralateral hemisphere. In short, the overlapping area between the CNAH binary mask and the tumour binary mask was subtracted from the CNAH binary mask, to generate a tumour free binarized CNAH mask. Rigid followed by affine registration in NiftyReg [21] was used to resample all segmented ROIs to the M_0/PD images for PASL/PCASL, respectively. This is because M_0/PD images have the same resolution as the CBF maps and possess similar anatomical features to structural images, thus facilitating co-registration (see Figure 2).

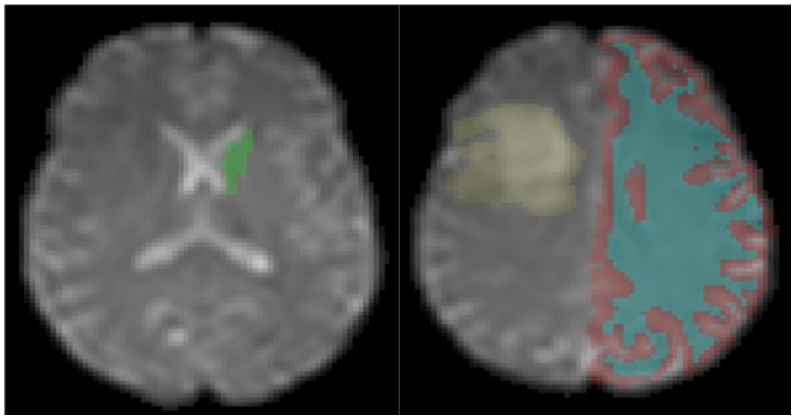


Fig 2 Automatically segmented ROIs in the contralateral healthy hemisphere in a patient with right frontal glioma, which was manually segmented. Green, caudate; red, GM; blue, WM; yellow, tumour.

Values were extracted from the maps as follows: i) tumour: mean aTBF and 95th-percentile (aTBF-95per); ii) CNAH: mean CBF in WM, GM, and caudate (CBF-WM, CBF-GM, and CBF-caud, respectively). The normalised tumour values relative to WM, GM, and caud (rTBF-WM, rTBF-GM and rTBF-caud, respectively) and their 95th-percentiles (rTBF-WM-95-per, rTBF-GM-95-per and rTBF-caud-95-per, respectively) were calculated. Note that the 95 percentile value of perfusion within the tumour is an important biomarker that has been shown to be relevant for diagnosis purposes in primary gliomas [1]. SNR was calculated for the ΔASL images and the calculated CBF maps in the CNAH ROIs according to the NEMA methodology [22] (eq.4):

$$SNR = \sqrt{2}\frac{\text{mean (ROI from (image1))}}{SD\ \text{(ROI from (image1} - \text{image2))}} \qquad \text{Equation 4}$$

With image1 and image2 representing either two consecutive ΔASL images or calculated CBF maps, respectively. The SNR was only calculated in the CNAH ROIs in order to avoid any variation which might result from abnormal tumour angiogenesis. Thus, the SNR from the ΔASL in WM, GM and caud were CBF_WM_SNR, CBF_GM_SNR and CBF_caud_SNR while from CBF maps were ΔASL_WM_SNR ΔASL_GM_SNR, and ΔASL_caud_SNR, respectively.

2.5 Statistical analysis

As this study sought to determine the extent to which the PASL and PCASL can be considered repeatable in terms of measuring aTBF/rTBF in tumour and CBF in the CNAH, repeatability was assessed separately for each ASL method. Test-retest repeatability was assessed based on the repeatability indices according to QIBA

(Quantitative Imaging Biomarkers Alliance[1]), namely the intra-class correlation coefficient (ICC), the within-coefficient-of-variation (WCoV in %) and the Coefficient of Repeatability (CR).

The ICC (two-way mixed effect, absolute agreement) was calculated for each of the extracted values from the CBF maps, grouped within each patient. This ICC model was selected as repeated measurements cannot be assumed to be a random sample[23], and repeated measurements are meaningless without the presence of agreement. The ICC helps to determine the ratio of the true variance to the observed variance. ICC values should be interpreted as follows: <0.5 shows "poor reliability, 0.5 to 0.75 shows moderate reliability, 0.75 to 0.9 demonstrates good reliability, and >0.9 offers excellent reliability.

The within-subject coefficient of variation (WCoV in [%]) is another repeatability index that represents the within-patient variation and is given in (eq.5):

$$WCoV\ [\%] = \left(\frac{WS\text{-}SD}{mean\ of\ the\ measurements}\right) \cdot 100 \qquad \text{Equation 5}$$

Where WS-SD is the within-subject standard deviation. WCoV reflects the variability within the data due to random error. Less variability implies higher reliability. The WCoV was calculated among the repeated measurements for each patient, as the within-subject variation is not expected to be uniform among tumour patients. Applying the root mean square approach [24], the WCoVs were squared for each patient to find

[1] http://qibawiki.rsna.org/images/8/8c/FMRITechnicalPerformanceIndices041613.pdf, accessed on June 2018.

the overall mean and 95% CI, after which the square roots of the mean and the 95% CI were calculated.

A further repeatability index is the Coefficient of Repeatability (CR), also known as the smallest real difference (SRD) [25,26]; it allows quantification of the variation (error) in the same unit as the measured value (ml/100 g/min in this case) with 95% probability. Lower CR corresponds to a more sensitive/responsive measurement. Changes less than the CR are not detectable and deemed to be due to measurement variation, meaning that the method is insensitive to changes smaller than the CR.

To estimate the CR, first, the within-subject standard deviation (WS-SD) was calculated among the repeated measurements for each patient, as it is not expected to be uniform among tumour patients. The same procedure was used as for the WCoV and the CR was calculated according to (eq.6):

$$CR = 1.96\sqrt{2(WS\text{-}SD)^2} \qquad \text{Equation 6}$$

For both PASL and PCASL acquisitions, all extracted values from both PASL and PCASL were normally distributed according to Shapiro's test of normality ($p > 0.5$), the $(WCoV)^2$ of each tumour perfusion value was compared between normal CBF and aTBF, between normal CBF and rTBF, and between aTBF and rTBF using paired t-tests. Also, the $(WCoV)^2$ of each tumour perfusion value was compared between HGGs and LGGs using heteroschedastic t-tests, as the variance could not be estimated to be equal *a priori*.

On the other hand, as different acquisition methods resulting in potential bias in the perfusion distributions were used in PASL and PCASL groups, the normality assumption could not be guaranteed, and the $(WCoV)^2$ of aTBF, rTBF, normal CBF and SNR were thus compared between both acquisitions using Mann-Whitney U tests. The IBM SPSS version 24 (*SPSS Inc, Chicago, IL*) statistical package was used for the statistical analysis.

3 Results

3.1 Intraclass correlation coefficient (ICC)

Generally, all perfusion metrics obtained from both PASL and PCASL demonstrated high reliability (ICC > 0.9) as illustrated in Table 2. Nevertheless, the between-patient SD of the extracted values was lower for PCASL than for PASL (Table 2). The tumour perfusion values using PASL demonstrated higher between-patients SD.

3.2 aTBF/rTBF versus normal CBF

Generally among the acquired maps from PASL and PCASL, the estimated WCoV of the aTBF_95per, rTBF and rTBF-95per (normalised to caudate) were significantly higher than the CBF WCoV in WM and GM. The estimated WCoV of the aTBF, rTBF and rTBF-95per (normalised to WM or GM) were not significantly different from the WCoV of CBF in GM and WM, whereas the WCoV of the aTBF and rTBF (normalised to WM or GM) were both significantly lower than the estimated WCoV of CBF in the caudate (Table 1 in the supplementary material).

3.3 aTBF versus rTBF

From both PASL and PCASL datasets, the WCoV of rTBF and rTBF_95per that were normalised to a "small ROI" were significantly higher than the WCoV of aTBF (Tables

2 and Figure 3, and Table 1 in the supplementary material) whereas no significant increase in the WCoV of aTBF was found as compared to rTBF and rTBF_95per when the latter was normalised to "large ROI". Normalisation of the tumour perfusion metrics to the whole WM or GM of the CNAH showed its ability to filter out some of the signal variations, while using the "small ROI" raised the signal variation level, where the lowest observed CR values were coming from the rTBF normalised to large ROIs. All the estimated CR of the measured values are illustrated in Table 2 and Figure 4.

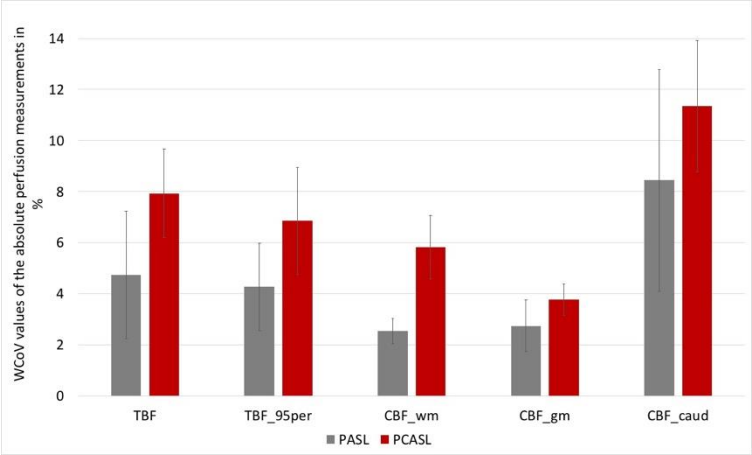
3.4 HGGs versus LGGs

Among all the acquired maps from PASL and PCASL, there were no significant differences ($p > 0.05$) between the HGGs and LGGs regarding the estimated WCoV of the tumour perfusion values (aTBF and rTBF).

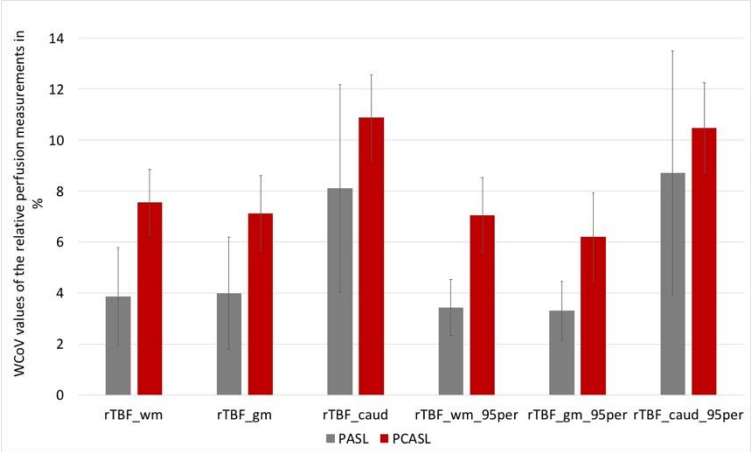
3.5 PASL versus PCASL

In the PASL CBF maps, macrovascular artefacts were observed, coming from accumulation of labelled spins in large vessels; Figure 5 shows examples of CBF maps for both PASL and PCASL. With the imaging parameters used within this study, the mean CBF and SNR values from PCASL were lower than those from PASL (Table 2, Figure 6). Furthermore, as expected, the SNR showed significant negative correlation with the WCoV (WM, $r = -0.632$, $p < 0.001$; GM, $r = -0.602$, $p < 0.001$; caudate, $r = -0.463$, $p = 0.003$). The WCoV of the perfusion metrics were significantly higher ($p < 0.05$) for PCASL than PASL (Table 2 in the supplementary material). In contrast, the WCoV of the SNR was significantly lower for PCASL than PASL in the caudate, and showed the same trend, though non-significant in the WM and the GM.

a.



b.



c.

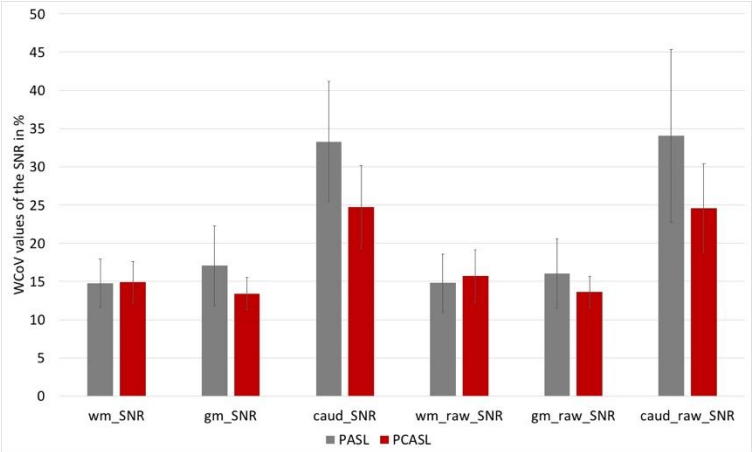
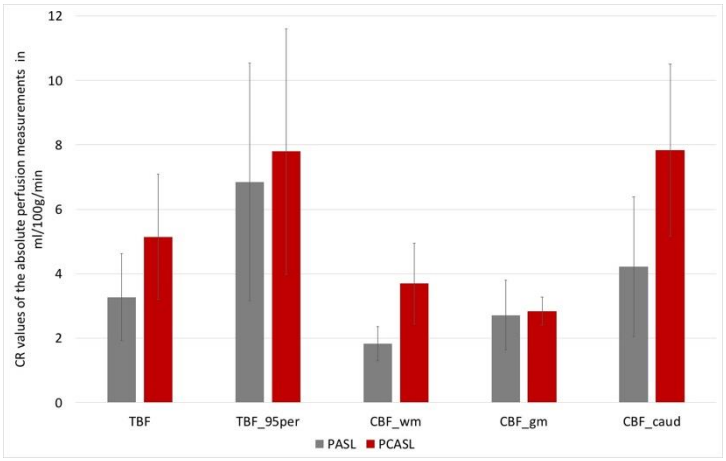
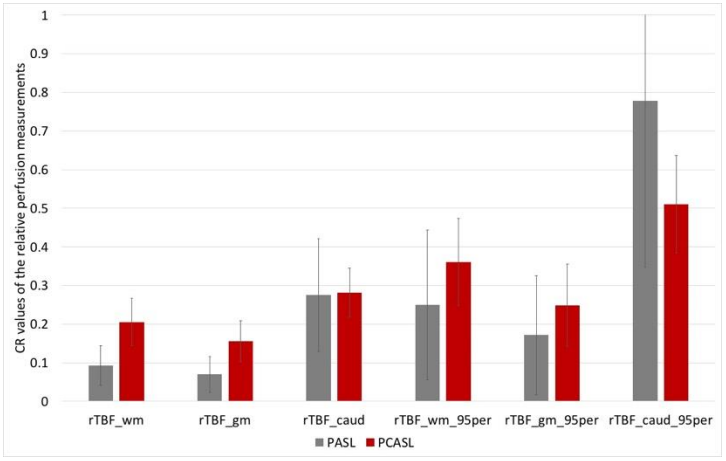


Fig 3 %WCoV from PASL and PCASL of (a) the absolute perfusion measurements, (b) the relative perfusion measurements and (c) the SNR from the CNAH.

a.



b.



c.

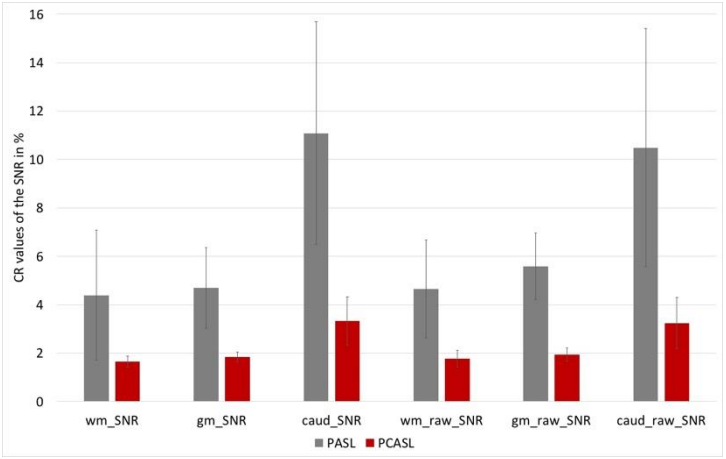


Fig 4 Quantified CR from PASL and PCASL of (a) the absolute perfusion measurements in [ml/100g/min], (b) the relative perfusion measurements and (c) the SNR from the CNAH.

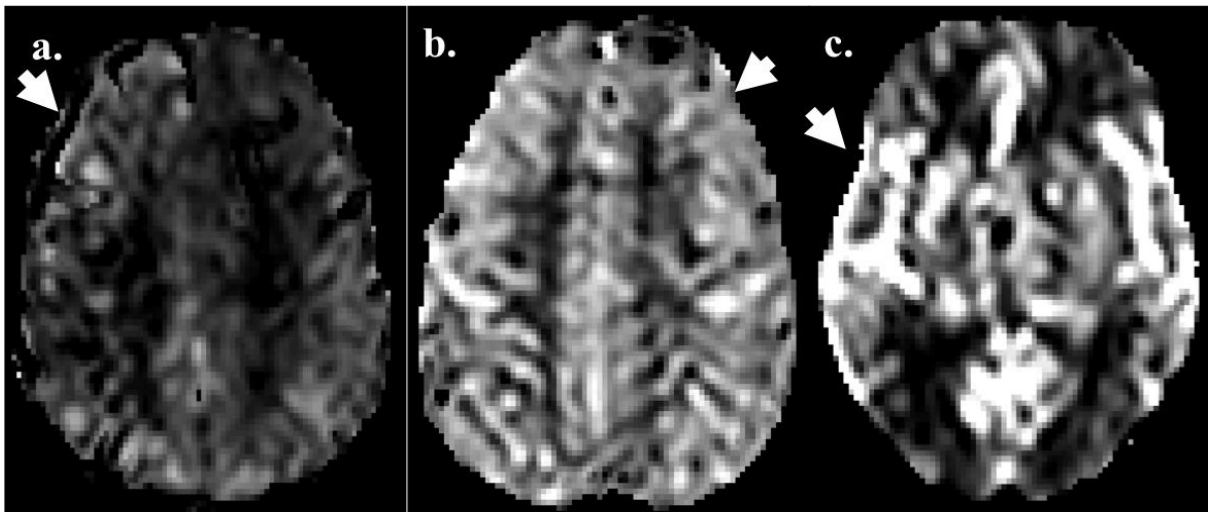
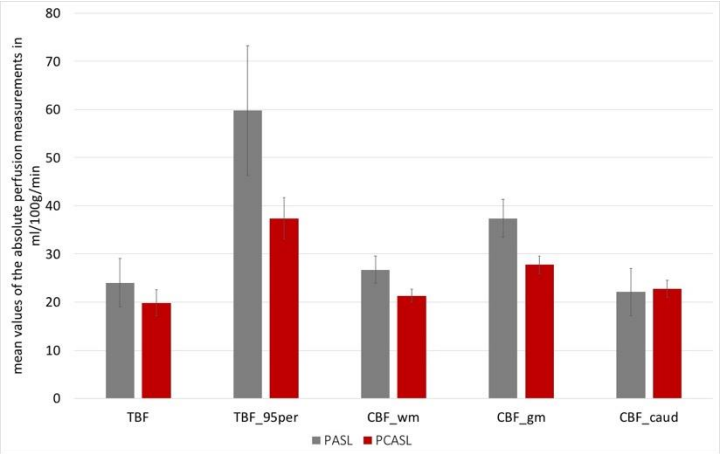
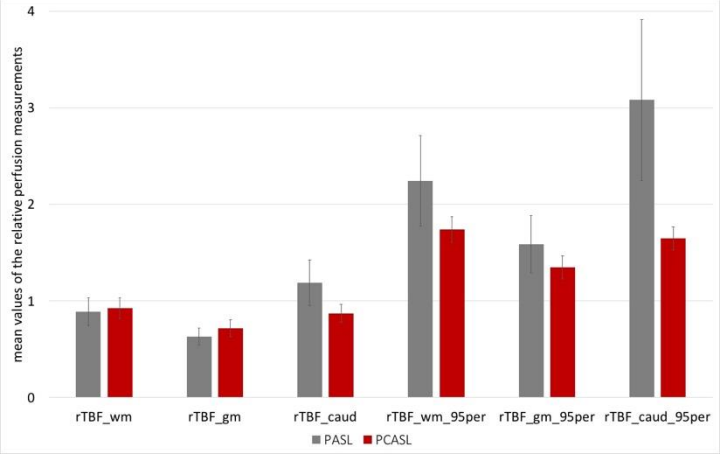


Fig 5 An example illustrating the relationship between the WCoV and the ASL-SNR specifically in the CNAH (indicated by arrows). (a.) PCASL CBF-map from for 56-year-old glioma patient, the tumour is located in the left hemisphere while the right hemisphere is the CNAH (arrow). The WCoV values were WM= 7.2%; GM= 5.1%; caudate= 16.9%; the high WCoV values from the CNAH presented with low perfusion SNR. (b.) PCASL CBF-map from a 33-year-old glioma patient, the tumour was located at the right hemisphere whereas the left hemisphere is the CNAH (arrow). The WCoV values are WM= 3%; GM= 1.3%; caudate= 9.2%; the low WCoV from the CNAH correlated with high perfusion SNR. (c.) PASL CBF-map in a 55-year-old glioma patient, the tumour was located in the left hemisphere while the right hemisphere is the CNAH (arrow). The WCoV values are WM= 3.9%; GM= 4.9%; caudate= 10.12%; even though the CNAH showed small WCoV that is due to the high SNR, this is an artificial increase in SNR mostly due to accumulated labelled spins inside the macrovasculature. Note that the images are shown in grayscale to enable visual differentiation between true perfusion (as in a. and b.) and macro-vascular artefact (as in c.).

a.



b.



c.

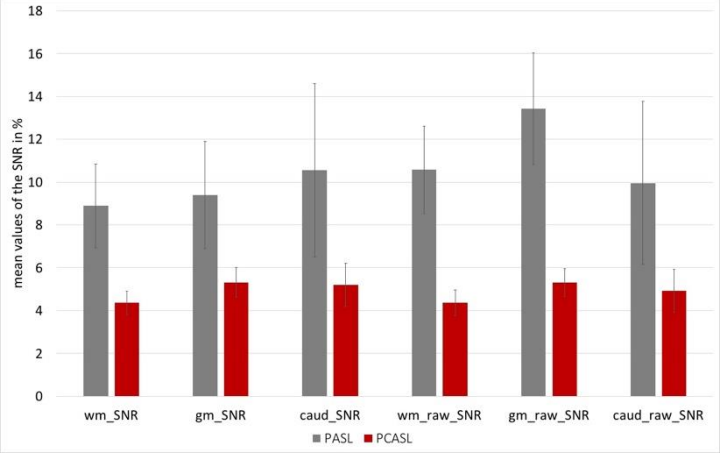


Fig 6 Mean values from the PASL and PCASL of (a) the absolute perfusion measurements in [ml/100g/min], (b) the relative perfusion measurements and (c) the SNR from the CNAH.

Table 2 between patients mean, SD and 95%CI; within patients ICC, WCoV and CR estimates in the extracted measurements from the CBF maps								
	PASL				PCASL			
	Mean ± SD (95%CI)	ICC (95%CI)	WCoV (95%CI)	CR (95%CI)	Mean±SD (95%CI)	ICC (95%CI)	WCoV (95%CI)	CR (95%CI)
aTBF	24.04 ± 9.98 (18.7, 29.4) [ml/100g/min]	0.998 (0.995, 0.999)	4.7 (2.2, 6.3)	3.3 (1.92, 5.017) [ml/100g/min]	19.9 ± 6.56 (16.9, 22.8) [ml/100g/min]	0.996 (0.994, 0.998)	7.9 (6.2 to 9.4)	5.1 (3.2, 6.5) [ml/100g/min]
aTBF_95per	59.8 ± 26.97 (45.4, 74.1) [ml/100g/min]	0.999 (0.997, 0.999)	4.3 (2.6, 5.5)	6.9 (3.166, 9.16) [ml/100g/min]	37.4 ± 10.49 (32.9, 41.8) [ml/100g/min]	0.996 (0.994, 0.998)	6.8 (4.7 to 8.4)	7.8 (3.9, 10.3) [ml/100g/min]
CBF_wm	26.7 ± 5.62 (23.7, 29.7) [ml/100g/min]	0.998 (0.995, 0.999)	2.5 (2, 2.9)	1.8 (1.303, 2.24) [ml/100g/min]	21.3 ± 3.34 (19.9, 22.8) [ml/100g/min]	0.992 (0.987, 0.996)	5.8 (4.6 to 6.9)	3.7 (2.4, 4.6) [ml/100g/min]
CBF_gm	37.4 ± 7.89 (33.2, 41.6) [ml/100g/min]	0.997 (0.995, 0.999)	2.7 (1.7, 3.5)	2.7 (1.64, 3.472) [ml/100g/min]	27.8 ± 4.49 (25.9, 29.7) [ml/100g/min]	0.997 (0.996, 0.999)	3.8 (3.2 to 4.3)	2.8 (2.4, 3.2) [ml/100g/min]
CBF_caud	22.1 ± 9.92 (16.8, 27.4) [ml/100g/min]	0.996 (0.992, 0.998)	8.4 (4.1, 11.2)	4.2 (2.057, 5.605) [ml/100g/min]	22.8 ± 4.43 (20.9, 24.7) [ml/100g/min]	0.980 (0.966, 0.990)	11.3 (8.8 to 13.4)	7.8 (5.2, 9.8) [ml/100g/min]
rTBF_wm	0.89 ± 0.29 (0.73, 1.04)	0.998 (0.996, 0.999)	3.86 (1.9, 5.1)	0.092 (0.040, 0.124)	0.92 ± 0.26 (0.81, 1.03)	0.996 (0.993, 0.998)	7.5 (6.3 to 8.7)	0.205 (0.144, 0.252)
rTBF_gm	0.63 ± 0.18 (0.53, 0.72)	0.997 (0.993, 0.999)	3.98 (1.7, 5.3)	0.07 (0.023, 0.096)	0.71 ± 0.22 (0.62, 0.81)	0.997 (0.995, 0.998)	7.1 (5.7 to 8.3)	0.156 (0.1025, 0.195)

rTBF_caud	1.19 ± 0.47 (0.94, 1.4)	0.993 (0.986, 0.997)	8.1 (4, 10.7)	0.27 (0.128, 0.366)	0.87 ± 0.23 (0.77, 0.97)	0.990 (0.983, 0.995)	10.9 (9.2 to 12.3)	0.281 (0.218, 0.333)
rTBF_wm_95per	2.24 ± 0.94 (1.7, 2.7)	0.998 (0.997, 0.999)	3.4 (2.3, 4.3)	0.24 (0.055, 0.36)	1.74 ± 0.32 (1.61, 1.88)	0.992 (0.986, 0.996)	7.06 (5.58 to 8.3)	0.361 (0.248, 0.445)
rTBF_gm_95per	1.59 ± 0.60 (1.3, 1.9)	0.998 (0.997, 0.999)	3.3 (2.17, 4.15)	0.17 (0.0175, 0.24)	1.35 ± 0.29 (1.22, 1.47)	0.995 (0.992, 0.998)	6.19 (4.47 to 7.5)	0.249 (0.142, 0.322)
rTBF_caud_95per	3.08 ± 1.67 (2.2, 3.9)	0.995 (0.991, 0.998)	8.7 (3.9, 11.7)	0.77 (0.347, 1.045)	1.65 ± 0.29 (1.52, 1.77)	0.979 (0.964, 0.989)	10.5 (8.7 to 11.9)	0.51 (0.38, 0.611)
CBF_wm_SNR	8.9 ± 3.94 (6.8, 10.9)	0.968 (0.936, 0.987)	14.8 (11.7, 17.4)	4.4 (1.69, 5.97)	4.36 ± 1.36 (3.79, 4.9)	0.985 (0.975, 0.993)	14.9 (12.2 to 17.2)	1.65 (1.43, 1.84)
CBF_gm_SNR	9.4 ± 4.98 (6.7, 12)	0.978 (0.965, 0.991)	17.1 (11.8, 21.01)	4.7 (3.036, 5.91)	5.32 ± 1.70 (4.6, 6.04)	0.989 (0.982, 0.995)	13.41 (11.29 to 15.24)	1.84 (1.63, 2.028)
CBF_caud_SNR	10.6 ± 8.08 (6.2, 14.9)	0.955 (0.909, 0.982)	33. 3 (25.4, 39.6)	11.1 (6.481, 14.28)	5.2 ± 2.47 (4.15, 6.24)	0.954 (0.923, 0.977)	24.7 (19.291 to 29.15)	3.3 (2.33, 4.097)
ΔASL_wm _SNR	10.6 ± 4.09 (8.4, 12.7)	0.969 (0.938, 0.988)	14.8 (10.9, 17.8)	4.6 (2.619, 6.031)	4.4 ± 1,44 (3.8, 4.99)	0.983 (0.972, 0.992)	15.7 (12.27 to 18.54)	1.8 (1.43, 2.055)
ΔASL_gm_SNR	13.4 ± 5.22 (10.6, 16.2)	0.971 (0.941, 0.988)	16. 1 (11.5, 19.5)	5.6 (4.2, 6.69)	5.3 ± 1.59 (4.6, 5.98)	0.986 (0.976, 0.993)	13.6 (11.6 to 15.4)	1.9 (1.67, 2.183)
ΔASL_caud_SNR	9.9 ± 7.61 (5.9, 14)	0.956 (0.910, 0.982)	34.1 (22.7, 42.5)	10.5 (5.6, 13.74)	4.9 ± 2.47 (3.88, 5.96)	0.957 (0.928, 0.978)	24.6 (18.8 to 29.2)	3.2 (2.187, 4.025)

3.6 Discussion

In this study, PASL and PCASL repeatability in gliomas was evaluated using three indices, namely ICC, WCoV and CR. As both patient groups were not the same across the ASL labelling approaches, the evaluation was undertaken separately. Our findings demonstrated here that aTBF WCoV was not different than rTBF WCoV when a large ROI was used as an internal reference. This suggest that when normalizing TBF values, a large ROI is better than a small one to avoid artefactual increased in noise. In addition, the CR (or SRD) was lowest for the rTBF when normalised to the large ROIs, where a small CR indicates high sensitivity to changes [26]. In a recent study by Zhou et al. the measured WCoV was lower for large ROIs than for small ROIs, for example WCoV from temporal lobe was 4.95% while from the putamen was 7.12%. This is supported by the inverse relationship between the within-subject variation and the ROI size seen in ASL and reported in several studies [27,28]. Previous studies reported the high diagnostic accuracy for glioma grading using rTBF, and also the role of the rTBF in the improvement of the detection of small differences of perfusion in the case of moderate SNR [1,29]. The lack of difference in WCoV in this study between both indices might be due to the relatively artificial way of estimating repeatability indices, based on split measurements. Generally however, in clinical studies where an internal reference is chosen, it is important to select it carefully, preferably a large ROI based on GM within the CNAH. These results are highly significant in demonstrating the value of ASL as an important tool for both prediction of the grade of the tumour [1,2], as well as its role in the follow-up of patients post treatment, for which pseudo-progression, real progression and radiotherapy-induced necrosis are sometimes difficult to differentiate [5].

The reliability for both PASL and PCASL was excellent (ICC>0.9) in concordance with previous studies, where it ranged from good to excellent [8,9,30]. We believe that the high-reliability in our measurements was due to the short period between the repeated measurements, as previous ASL reliability studies showed the adverse effect of long time interval among the scan sessions [8,30]. As such, one of the potential shortcomings of this study relates to the fact that it is really a simple repeatability study and not a reproducibility one, in which times between scans are usually much larger. The slight superiority of PASL over PCASL in reliability in this study is primarily due to the difference in acquisition parameters. In particular, the voxel size between both methods was ~1.7 times larger for our PASL implementation than for the PCASL one. In addition, presence within the PASL data of a slightly higher between patient SD might also contribute to the ICC increases, as is generally the case when between patient SD is greater than the within patient SD [31].

The WCoV of CBF in CNAH was found to be comparable with previous studies [7,8,27]. Higher WCoV in the perfusion metrics (both for normal CBF and tumour) was observed in PCASL compared to PASL. Probable reasons are the same as before, including intrinsic differences in the two patient groups, the acquisition parameters, and in particular the voxel size. Besides, the difference in labelling efficiency between both sequences, with PASL having a generally slightly higher efficiency than PCASL, might have also possibly contributed [32].

The higher WCoV of PCASL-based estimation of CBF does not necessarily mean the superiority of PASL over PCASL, as both labelling methods employed were applied to different patient groups. The quantified means from the CNAH (CBF-WM, CBF-GM

and CBF-caud) were higher in PASL than in PCASL. In addition, PCASL has been reported to display higher SNR and reproducibility than PASL when applied to the same patient group [8]. Furthermore, in our study, the WCoV of the SNR was lower for PCASL for small regions, and showed a trend towards reduced WCoV in the rest of the brain, which indicates a potential higher stability in comparison to PASL.

It is worth noting that there is an inverse relationship between the amplitude of the ASL signal and its variation (WCoV) across time [8]. As a matter of fact, even an artefactual increase in the SNR from the accumulated labelling spins in the macro-vessels will reduce the WCoV, if it can be measured reproducibly. Therefore, the decrease in the WCoV of the PASL perfusion metrics might be a consequence of the higher ASL signal due to labelled spins that have not yet reached the microcirculation. This phenomenon is also known as a macro-vascular artefact or Arterial Transit Artefact (ATA) (see Figure 5) and is mainly due to the delayed arrival of the labelled blood to the tissue.

This underscores the significant role of the PLD selection and its impact on the robustness of the CBF quantification. This is most critical in tumours with abnormal transit times due to the newly formed tortuous vessels. A recent ASL study suggested using phase-contrast MRI to determine the blood velocity profile and inform the selection of the PLD [33]; this might help to ensure complete delivery of the labelled bolus into the tissue of interest, however at the cost of scan time, and thus might not be practically implementable in the clinics. Another possible solution for addressing the issues of ATA is the use of more advanced ASL methods, such as multi-delay ASL [34].

The main limitation of our study is its small sample size of the repeated measurements, which has been addressed using moving block bootstrapping [16]. In this study, we focused on comparing the short-term repeatability of the different ASL acquisition techniques, and did not address within-subject variation due to patient repositioning and physiological tissue perfusion perturbations and changes, which is of particular importance for longitudinal tumour studies. However, based on the findings of the large QUASAR study [7], an estimate of these additional effects could be established as a guiding rule regarding sample size calculation for future clinical trials. Most importantly however, was the recognition that repeatability was not lower in primary brain tumour patients than in healthy volunteers, as assessed in previous studies. This is particularly remarkable due to the use of split acquisition in blocks, resulting in a generally lower average scan time than used clinically.

3.7 Conclusion

This repeatability study highlights the excellent reliability of both PASL and PCASL for neuro-oncological studies, in line with previous volunteer studies. However, aTBF/rTBF estimates tend to maintain high within-subject variability that may obscure the detection of between-subject variation with small changes in tumour perfusion. When estimating rTBF, large ROIs need to be used to filter out systemic misleading within-subject variation and thus provide more robust perfusion measurements. Finally, in this study, no large difference between repeatabilities were observed between PASL and PCASL, which might have been attributed primarily to the difference in acquisition parameters, and both methods can therefore be recommended to assess brain perfusion, with the caveat that PCASL allows for longer

PLD than PASL, which therefore minimises the macro-vascular artefacts, and provide better estimates of patient perfusion.

Author contributions statement: All authors contributed to the study conception and design. Study conception and design were fulfilled by Xavier Golay, Sotirios Bisdas and Amirah F Alsaedi. Data collection were performed by David L Thomas and Enrico De Vita. Analysing and interpretation of the data were performed by Amirah F Alsaedi and Jasmina Panovska-Griffiths. The first draft of the manuscript was written by Amirah F Alsaedi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

XG is founder, shareholder and CEO of Gold Standard Phantoms, a UCL spinout company producing, among others, ASL phantoms.

3.8 References

1. Alsaedi A, Doniselli F, Jäger HR, Panovska-Griffiths J, Rojas-Garcia A, Golay X, Bisdas S. The value of arterial spin labelling in adults glioma grading: systematic review and meta-analysis. Oncotarget [Internet]. 2019; 10: 1589–601. doi: 10.18632/oncotarget.26674.

2. Kong L, Chen H, Yang Y, Chen L. A meta-analysis of arterial spin labelling perfusion values for the prediction of glioma grade. Clin Radiol [Internet]. The Royal College of Radiologists; 2017; 72: 255–61. doi: 10.1016/j.crad.2016.10.016.

3. Weber MA, Thilmann C, Lichy MP, Günther M, Delorme S, Zuna I, Bongers A, Schad LR, Debus J, Kauczor HU, Essig M, Schlemmer HP. Assessment of Irradiated Brain Metastases by Means of Arterial Spin-Labeling and Dynamic Susceptibility-Weighted Contrast-Enhanced Perfusion MRI: Initial Results. Invest Radiol. 2004; 39: 277–87. doi: 10.1097/01.rli.0000119195.50515.04.

4. Yun TJ, Cho HR, Choi SH, Kim H, Won JK, Park SW, Kim JH, Sohn CH, Han MH. Antiangiogenic effect of bevacizumab: Application of arterial spin-labeling perfusion MR imaging in a rat glioblastoma model. Am J Neuroradiol [Internet]. 2016; 37: 1650–6. doi: 10.3174/ajnr.A4800.

5. Wang L, Wei L, Wang J, Li N, Gao Y, Ma H, Qu X, Zhang M. Evaluation of perfusion MRI value for tumor progression assessment after glioma radiotherapy: A systematic review and meta-analysis. Medicine (Baltimore). 2020; 99: e23766. doi: 10.1097/MD.0000000000023766.

6. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, Boellaard R, Bohndiek SE, Brady M, Brown G, Buckley DL, Chenevert TL, Clarke LP, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol [Internet]. Nature Publishing Group; 2017; 14: 169–86. doi: 10.1038/nrclinonc.2016.162.

7. Petersen ET, Mouridsen K, Golay X. The QUASAR reproducibility study, Part II: Results from a multi-center Arterial Spin Labeling test-retest study. Neuroimage [Internet]. Elsevier Inc.; 2010; 49: 104–13. doi: 10.1016/j.neuroimage.2009.07.068.

8. Chen Y, Wang DJJ, Detre JA. Test-retest reliability of arterial spin labeling with common labeling strategies. J Magn Reson Imaging [Internet]. 2011; 33: 940–9. doi: 10.1002/jmri.22345.

9. Zhou L, Wang Y, Pinho MC, Pan E, Xi Y, Maldjian JA, Madhuranthakam AJ. Intrasession Reliability of Arterial Spin-Labeled MRI–Measured Noncontrast Perfusion in Glioblastoma at 3 T. Tomography [Internet]. 2020; 6: 139–47. doi: 10.18383/j.tom.2020.00010.

10. Warmuth C, Günther M, Zimmer C. Quantification of blood flow in brain tumors: Comparison of arterial spin labeling and dynamic susceptibility-weighted contrast-enhanced MR imaging. Radiology. 2003; 228: 523–32. doi: 10.1148/radiol.2282020409.

11. Cebeci H, Aydin O, Ozturk-Isik E, Gumus C, Inecikli F, Bekar A, Kocaeli H, Hakyemez B. Assesment of perfusion in glial tumors with arterial spin labeling; Comparison with dynamic susceptibility contrast method. Eur J Radiol [Internet]. Elsevier Ireland Ltd; 2014; 83: 1914–9. doi: 10.1016/j.ejrad.2014.07.002.

12. Figueiredo PM, Clare S, Jezzard P. Quantitative perfusion measurements using pulsed arterial spin labeling: Effects of large region-of-interest analysis. J Magn Reson Imaging [Internet]. 2005; 21: 676–82. doi: 10.1002/jmri.20329.

13. Fallatah S. Perfusion Imaging in Brain Tumours. UCL (University College London); 2016.

14. Melbourne A, Toussaint N, Owen D, Simpson I, Anthopoulos T, De Vita E, Atkinson D, Ourselin S. NiftyFit: a Software Package for Multi-parametric Model-Fitting of 4D Magnetic Resonance Imaging Data. Neuroinformatics [Internet]. Neuroinformatics; 2016; 14: 319–37. doi: 10.1007/s12021-016-9297-6.

15. Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage [Internet]. 2002; 17: 825–41. doi: 10.1006/nimg.2002.1132.

16. Mignani S, Rosa R. The moving block bootstrap to assess the accuracy of statistical estimates in Ising model simulations. Comput Phys Commun. 1995; 92: 203–13. doi: 10.1016/0010-4655(95)00114-7.

17. Alsop DC, Detre JA, Golay X, Günther M, Hendrikse J, Hernandez-Garcia L, Lu H, Macintosh BJ, Parkes LM, Smits M, Van Osch MJP, Wang DJJ, Wong EC, 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: 102–16. doi: 10.1002/mrm.25197.

18. Wansapura JP, Holland SK, Dunn RS, Ball WS. NMR relaxation times in the human brain at 3.0 Tesla. J Magn Reson Imaging. 1999; 9: 531–8. doi: 10.1002/(SICI)1522-2586(199904)9:4<531::AID-JMRI4>3.0.CO;2-L.

19. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage. 2006; 31: 1116–28. doi: 10.1016/j.neuroimage.2006.01.015.

20. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage [Internet]. 2011; 56: 907–22. doi: 10.1016/j.neuroimage.2011.02.046.

21. Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S. Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed. 2010; 98: 278–84. doi: 10.1016/j.cmpb.2009.09.002.

22. NEMA. NEMA Standards Publication MS 1-2001: Determination of Signal-to-Noise

Ratio (SNR) in Diagnostic Magnetic Resonance Imaging. Natl Electr Manuf Assoc [Internet]. 2014; 2008: 21. Available from https://www.nema.org/Standards/Pages/Determination-of-Signal-to-Noise-Ratio-in-Diagnostic-Magnetic-Resonance-Imaging.aspx

23. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med [Internet]. Elsevier B.V.; 2016; 15: 155–63. doi: 10.1016/j.jcm.2016.02.012.

24. Bland M. How do I calculate a within-subject coefficient of variation [Internet]. 2006. Available from https://www-users.york.ac.uk/~mb55/meas/cv.htm

25. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The Case for Using the Repeatability Coefficient When Calculating Test-Retest Reliability. PLoS One. 2013; 8: 1–7. doi: 10.1371/journal.pone.0073990.

26. Schuck P, Zwingmann C. The 'smallest real difference' as a measure of sensitivity to change: a critical analysis. Int J Rehabil Res. 2003; 26: 85–91. doi: 10.1097/01.mrr.0000070759.63544.65.

27. Wang Y, Saykin AJ, Pfeuffer J, Lin C, Mosier KM, Shen L, Kim S, Hutchins GD. Regional reproducibility of pulsed arterial spin labeling perfusion imaging at 3T. Neuroimage [Internet]. Elsevier Inc.; 2011; 54: 1188–95. doi: 10.1016/j.neuroimage.2010.08.043.

28. Mutsaerts HJMM, Steketee RME, Heijtel DFR, Kuijer JPA, Van Osch MJP, Majoie CBLM, Smits M, Nederveen AJ. Inter-vendor reproducibility of pseudo-continuous arterial spin labeling at 3 Tesla. PLoS One. 2014; 9. doi: 10.1371/journal.pone.0104108.

29. Aslan S, Lu H. On the sensitivity of ASL MRI in detecting regional differences in cerebral blood flow. Magn Reson Imaging [Internet]. Elsevier Inc.; 2010; 28: 928–35. doi: 10.1016/j.mri.2010.03.037.

30. 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: 1919–27. doi: 10.1038/jcbfm.2014.163.

31. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M, Zahlmann G, Kondratovich M V, O'Donnell K, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Stat Methods Med Res [Internet]. 2015; 24: 27–67. doi: 10.1177/0962280214537344.

32. Sokolska M, Bainbridge A, Rojas-Villabona A, Golay X, Thomas DL. Effect of labelling plane angulation and position on labelling efficiency and cerebral blood flow quantification in pseudo-continuous arterial spin labelling. Magn Reson Imaging [Internet]. Elsevier; 2019; 59: 61–7. doi: 10.1016/j.mri.2019.02.007.

33. Robertson AD, Matta G, Basile VS, Black SE, Macgowan CK, Detre JA, MacIntosh BJ. Temporal and spatial variances in arterial spin-labeling are inversely related to large-artery blood velocity. Am J Neuroradiol. 2017; 38: 1555–61. doi: 10.3174/ajnr.A5257.

34. Yang XS, Zhao B, Wang G, Xiang J, Xu S, Liu Y, Zhao P, Pfeuffer J, Qian T. Improving the grading accuracy of astrocytic neoplasms noninvasively by combining timing information with cerebral blood flow: A multi-Ti arterial spin-labeling MR imaging study. Am J Neuroradiol [Internet]. 2016; 37: 2209–16. doi: 10.3174/ajnr.A4907.