

Enhancement of automated blood flow estimates (ENABLE) from arterial spin-labeled MRI

Abstract

Purpose: To validate a multi-parametric automated algorithm – enhancement of automated blood flow estimates (ENABLE) – that identifies useful and poor arterial spin-labeled (ASL) difference images in multiple post labeling delay (PLD) acquisitions and thereby improve clinical ASL.

Materials and methods: ENABLE is a sort/check algorithm that uses a linear combination of ASL quality features. ENABLE uses simulations to determine quality weighting factors based on an unconstrained nonlinear optimization. We acquired a set of 6-PLD ASL images at 1.5T or 3.0T systems among 98 healthy elderly and adults with mild cognitive impairment or dementia. We contrasted signal to noise ratio (SNR) of CBF images obtained with ENABLE vs. conventional ASL analysis. In a subgroup, we validated our CBF estimates with single-photon emission computed tomography (SPECT) CBF images.

Results: ENABLE produced significantly increased SNR compared to a conventional ASL analysis (Wilcoxon signed-rank test, $p < 0.0001$). We also found the similarity between ASL and SPECT was greater when using ENABLE vs. conventional ASL analysis ($N=51$, Wilcoxon signed-rank test, $p < 0.0001$) and this similarity was strongly related to ASL SNR ($t=24$, $P < 0.0001$).

Conclusion: These findings suggest that ENABLE improves CBF image quality from multiple PLD ASL in dementia cohorts at either 1.5T or 3T, achieved by multi-parametric quality features that guided post-processing of dementia ASL.

Keywords: Multiple post labeling delay, cerebral blood flow, quality evaluation, dementia

Introduction

Arterial spin-labeled (ASL) magnetic resonance imaging (MRI) permits non-invasive quantification of cerebral blood flow (CBF) (1). Several dementia studies have shown regionally-specific chronic reduction of resting CBF, which suggests the utility of physiological imaging in dementia assessments (2,3). ASL is well suited for repeat scans and thus for screening, monitoring dementia progression and treatment responses. The CBF-weighted ASL difference (ASL-diff) images, however, have an intrinsically low signal to noise ratio (SNR) because the labeled blood water is only a small fraction of the total water in any given tissue voxel (1).

Prolonged arterial transit time (ATT) confounds CBF estimates in ASL; this tendency is more pronounced in adults with vascular disease. ATT describes the travel time of the blood from the labeling plane to tissue in the imaging plane. Prolonged ATT can compromise the CBF image due to the incomplete delivery of the ASL tracer and consequently hyperintense intravascular signal (1). Dai et al. found that accounting for ATT variability increased CBF values in the regions where ATT was longer than the post labeling delay (PLD) (4). Other studies have demonstrated multiple PLD (multi-PLD) ASL avoids intravascular ASL signal contamination during CBF quantification (1,4) and improves accuracy (5,6).

Liu et al. (7), Dai et al. (4) and Mutsaerts et al. (8) reported positive association between age and regional ATT in healthy adults and elderly with hypertension. Mak et al. found longer ATT in adults diagnosed with Alzheimer's disease compared to healthy elderly and an inverse correlation between ATT and cognitive performance (9). Hence, the use of multi-PLD ASL is of relevance to aging and dementia studies to account for the ATT variability.

Head motion between ASL control and label images can introduce spurious signals, posing a challenge among older and clinical cohorts (10,11). A recent ASL white paper recommends checking the individual ASL-diff images and evidence of head motion artifacts (1). Visual inspection is, however, time consuming and subject to inter-rater variability. Fortunately, there are post-processing algorithms available that automatically assess quality features in individual ASL-diff images by: 1) pre-defined thresholds on mean and standard deviation of ASL signal in tissue voxels (12), 2) maximizing ASL signal detectability (13) and 3) optimizing the correlation with the mean ASL-diff image (14,15).

Given the systematic differences in ASL contrast for different PLDs, it is not clear how to automatically evaluate quality of multi-PLD ASL images. For instance, short PLD images may be more or less susceptible to head motion than long PLD images. Therefore, a general solution to ASL image quality evaluation is of interest. A solution to this challenge should also strive to preserve as much ASL data as possible given the limited number of ASL-diff images.

In this study, we propose a multi-parametric quality evaluation procedure – enhancement of automated blood flow estimates (ENABLE) – to identify poor ASL-diff images in multi-PLD acquisitions. ENABLE used simulations – based on noise features of 1.5T and 3T clinical data – to determine quality criteria across any PLDs using several ASL quality features.

Materials and Methods

1. Simulation study

1.1. Overview of ENABLE

ENABLE is a sort/check algorithm to evaluate image quality for each PLD independently. We first align control and label images separately to the mean control image and calculate ASL-diff images. Second, we sort ASL-diff images based on their image quality with the goal of preserving highest quality volumes. Third, we compute a quality criterion and exclude ASL-diff images based on maximum of this quality criterion. To define a robust quality criterion, we used a combination of four ASL quality features. We determined the weightings of the quality criterion through simulations where ground-truth ASL-diff images were generated a priori.

1.2. Simulated ASL images

We used in-house tools based on FSL (18). We performed simulations using three sets of ground-truth images: CBF, ATT and large arterial blood volume maps obtained from previous empirical ASL analysis (Montreal neurological institute (MNI152; 2mm) template space, Figure 1A). We generated ASL image pairs – named ASL_{IDEAL} – using these ground-truth hemodynamic maps. This approach involved projecting the output images back to the original data by relying on the two-compartmental ASL kinetic model (19) given the following PLDs: 100, 500, 900, 1300, 1700 and 2100ms PLDs (i.e., the corresponding schedule for the empirical data acquired at 1.5T and 3.0T, described in the following section), shown in Figure 1B. To create simulated ASL noise, we extracted noise features from empirical data (N=98) and added them to ASL_{IDEAL} images. We simulated three noise sources. First, we calculated thermal noise features of each volume using non-brain voxels (13) and added Gaussian noise with the attained parameters to ASL_{IDEAL} . Second, we added voxel-wise physiological noise features (20) to the

simulated noisy ASL images. Third, we introduced twelve head motion parameters – including translations, rotations, scales and skews obtained from the motion correction process (MCFLIRT (21)) – using the FLIRT tool, effectively the reverse of the head motion correction process.

Consequently, we generated ninety-eight simulated ASL datasets (i.e., 98 cases with 24 ASL-diff images for each of the 6 PLDs) using noise features that resemble of 1.5T and 3T empirical ASL.

1.3. Sorting

The simulated control and label images underwent a sinc subtraction to generate twenty-four ASL-diff images per PLD (22). We computed the quality of each ASL-diff image using contrast to noise ratio (CNR). CNR is defined as the mean ASL difference signal in grey matter (GM) divided by the standard deviation of non-brain voxels. Thereafter, we sorted ASL-diff images based on their quality measures in descending order i.e., from the highest to the lowest CNR.

1.4. Quality criterion

We initially evaluated image quality from the six ASL-diff images with the highest CNR (23). We generated temporal mean and standard deviation ASL-diff images – named $ASL\text{-diff}_{\text{mean}}$ and $ASL\text{-diff}_{\text{std}}$, respectively – to calculate image quality features. Quality features included: 1) temporal CNR (tCNR: CNR of $ASL\text{-diff}_{\text{mean}}$ in GM), 2) detectability metric (Det: the proportion of $ASL\text{-diff}_{\text{mean}}$ voxels with significantly greater than zero signal in GM) (13), 3) coefficient of variation of $ASL\text{-diff}_{\text{mean}}$ (CoV: spatial standard deviation divided by spatial mean in GM) (24) and 4) temporal signal to noise ratio (tSNR: spatial mean of $ASL\text{-diff}_{\text{mean}}$ image divided by $ASL\text{-diff}_{\text{std}}$ image in GM). In addition, as an image similarity measure, we calculated a normalized mean-square difference metric between the $ASL\text{-diff}_{\text{mean}}$ and $ASL\text{-diff}_{\text{IDEAL}}$ in GM. Thereafter, we recalculated quality features and similarity measures with the addition of each successive

ASL-diff image from the quality sorted series. We obtained the GM mask from MNI152 template space tissue segmentation map using a probability threshold of 70%.

We assessed overall quality using a linear combination of quality features at any given PLD:

$$Quality\ criterion(n, m) = W1 \times tCNR(n, m) + W2 \times Det(n, m) + W3 \times CoV(n, m) + W4 \times tSNR(n, m) \quad (1)$$

$$Image\ similarity(n, m) = \sum_i (ASL\ diff_{mean}^i(n, m) - ASL\ diff_{IDEAL}^i)^2 / N(GM) \quad (2)$$

where n represents the number of ASL-diff images (range = 6-24), m represents a simulated ASL dataset (range = 1-98), i represents each voxel in GM and N(GM) is the number of GM voxels.

1.5. Training of ENABLE quality criterion

The goal was to use similarity measures to extract the quality weighting factors i.e., W1-W4. The optimization cost function was based on the distance between the similarity measure and the quality criterion:

$$Optimization\ cost\ function = \sum_{m,n} (Image\ similarity(n, m) - Quality\ criterion(n, m))^2 \quad (3)$$

We used an unconstrained nonlinear optimization tool (optimtool, MATLAB 2015) to find an optimum set of factors, i.e. W1-W4, that minimized the cost function (Equation 2). The 1.5T and 3T simulated data were treated separately.

1.6. Pass/fail

The final step of the quality evaluation procedure was to identify corrupted ASL-diff images. After obtaining the quality weightings, we computed the quality criterion for each ASL dataset. The maximum of this overall quality function corresponded to highest similarity between ASL-diff_{mean} and ASL-diff_{IDEAL} for each simulated ASL image. Therefore, we sought the maximum of a quality criterion and identified the corresponding ‘n’ – number of ASL-diff images – as the number of ASL-diff images retained by ENABLE. We classified ASL-diff images as ‘pass’ or

‘fail’ based on their involvement in the maximum of quality criterion. In other words, including ‘fail’ ASL-diff images decreased the overall quality metric and thus was discarded.

1.7. Evaluation of ENABLE on simulated data

We examined ENABLE performance against a conventional (CONV) multi-PLD ASL analysis, i.e. no image quality evaluation (4,25–27). We modeled multi-PLD images with an established ASL kinetic model tool (BASIL) (28) to generate CBF images. A free energy term provided a goodness of fit estimate such that lower free energy corresponds to a better fit (28). We calculated mean free energy in GM to compare CONV and ENABLE approaches. We also computed the similarity of the CBF_{CONV} and CBF_{ENABLE} with CBF_{IDEAL} ; this was based on the normalized mean-square difference metric in GM.

We performed statistical assessments in RStudio (R 3.2.2 GUI 1.66) with $p < 0.05$ observed as significant. Wilcoxon signed-rank test compared the similarity of CBF_{CONV} and CBF_{ENABLE} with CBF_{IDEAL} . In addition, we compared the mean free energy in GM between the CONV and ENABLE using Wilcoxon signed-rank test.

2. Empirical study

2.1. Image acquisition

The local Research Ethics Board approved the protocol of this study and all participants provided informed consent. Ninety-eight adults participated in this study and comprised three groups: healthy elderly ($n=15$, age range=59-86 years, female/male: 9/6), adults diagnosed with mild cognitive impairment (16) ($n=21$, age range=59-89 years, female/male: 11/10) and adults diagnosed with dementia (16) ($n=62$, age range=53-88 years, female/male: 35/27).

We performed MRI on a 1.5T GE Signa system (General Electric Medical Systems, Milwaukee, WI, USA) for sixty adults (mild cognitive impairment: $n=21$; dementia: $n=39$) and on a 3.0T GE

Discovery 750 system for thirty-eight adults (healthy elderly: n=15; dementia: n=23). The MRI protocol included a pseudo-continuous ASL acquisition (identical for 1.5T and 3.0T) for CBF estimation: 2D echo planar imaging, $64 \times 64 \times 17$ matrix, $3.5 \times 3.5 \times 5.6 \text{ mm}^3$, labeling duration = 1500ms, 6 PLDs of 100, 500, 900, 1300, 1700 and 2100ms and 12 control-label pairs for each PLD. TE/TR of 17ms/3.5-5s was used for a total ASL scan time under 10 minutes. A reference ASL acquisition for absolute CBF quantification was also acquired at the beginning of each PLD acquisition (17). Structural images were acquired for registration and tissue segmentation purposes. For 1.5T system, these acquisitions included: 1) T1-weighted axial 3D SPGR with TE/TR:5ms/35ms, flip angle 35° , 256×256 matrix, $0.9 \times 0.9 \text{ mm}^2$ and slice thickness of 1.2 to 1.4 mm depending on the head size and 2) interleaved Proton Density and T2-weighted axial 2D dual-echo spin echo: TE 30 and 80 ms, TR 3 s, 256×256 matrix, $0.78 \times 0.78 \text{ mm}^2$, slice thickness of 3mm. For 3.0T system, these acquisitions included: 1) T1-weighted axial 3D FSPGR with TE/TR/TI: 3.2/8.1/650 ms, flip angle 8° , 256×256 matrix, $0.86 \times 0.86 \text{ mm}^2$, slice thickness of 1 mm, 2) interleaved Proton Density and T2-weighted axial 2D fast spin echo: TE 11.1 and 90 ms, TR 2.5s, 256×256 matrix, $0.86 \times 0.86 \text{ mm}^2$, slice thickness of 3 mm and 3) T2-weighted Fluid Attenuated Inversion Recovery axial: TE/TR/TI: 140/9700/2200 ms, 256×256 matrix, $0.86 \times 0.86 \text{ mm}^2$, slice thickness 3 mm.

Participants imaged on the 1.5T MRI system received a corresponding single-photon computed emission tomography (SPECT) CBF scan using a three-headed Philips system (Philips, Amsterdam, The Netherlands). SPECT and MRI sessions occurred within 30 days of each other. We obtained SPECT scans ninety minutes after administration of technetium-99m hexamethylpropyleneamine oxime ($^{99\text{m}}\text{Tc-HMPAO}$) or sixty minutes after administration of

^{99m}Tc ethyl cysteine dimer (^{99m}Tc -ECD) with high resolution fan beam collimators (voxel size: $3.56 \times 3.56 \times 1 \text{ mm}^3$).

2.2. Empirical evaluation

We analyzed empirical ASL data using both CONV and ENABLE approaches. ENABLE resembled the analysis steps described in the simulation except that attained quality weightings were used. We again used BASIL tool (28) for CBF estimation followed by spatial smoothing with a Gaussian filter with full width at half maximum of 6mm. We calculated SNR of the CBF_{CONV} and $\text{CBF}_{\text{ENABLE}}$ images. SNR is defined as mean CBF within the GM mask divided by standard deviation of CBF. We performed the tissue segmentation in subject specific space using semi-automated brain region extraction tool. Our tri-feature tissue segmentation for GM, white matter and cerebrospinal fluid, was first based on a Gaussian fitting algorithm using localized histograms from T1-weighted voxel intensities. This initial segmentation was further corrected for misclassification errors due to atrophy and subcortical vasculopathy using proton density, T2-weighted and/or T2-weighted Fluid Attenuated Inversion Recovery voxel intensities as previously described (29). We co-registered these tissue masks to ASL space using ASL reference images and the FLIRT tool (21). We compared CONV and ENABLE estimates as the difference in SNR, i.e. $\Delta\text{SNR}_{\text{GM}}$:

$$\Delta\text{SNR}_{\text{GM}} = (\text{SNR}_{\text{ENABLE,GM}} - \text{SNR}_{\text{CONV,GM}}) \times 2 / (\text{SNR}_{\text{ENABLE,GM}} + \text{SNR}_{\text{CONV,GM}}) \times 100 \quad (4)$$

A positive value for $\Delta\text{SNR}_{\text{GM}}$ favours ENABLE as a means of improving image quality, while a negative favours CONV.

An iterative algorithm (order subset expectation maximization) (30) reconstructed SPECT images followed by attenuation correction process and spatial smoothing (31). We co-registered SPECT CBF maps to the T1-weighted images and downsampled to 5mm isotropic resolution. We generated relative CBF maps by intensity normalization to mean CBF in GM. We performed a similar processing for ASL CBF_{CONV} and CBF_{ENABLE} to obtain intensity normalized ASL images. Thereafter, we assessed similarity between intensity normalized ASL (CONV and ENABLE) with SPECT using generalized Jaccard index (JI) defined as the following (32):

$$JI = \sum_i \min(ASL(i), SPECT(i)) / \sum_i \max(ASL(i), SPECT(i)) \quad (5)$$

where i represents each voxel in CBF images.

2.3. Statistical tests

We performed statistical assessments in RStudio (R 3.2.2 GUI 1.66) with $p < 0.05$ observed as significant. We compared SNR and free energy between the CONV and ENABLE using Wilcoxon signed-rank test. A general linear model examined the influence of field strength, diagnosis and mean head motion on ΔSNR_{GM} . We also assessed the effects of diagnosis, field strength and mean head motion on the number of retained ASL-diff images by ENABLE using a general linear model. For the SPECT subgroup, we assessed the similarity of intensity normalized ASL CBF_{CONV} and CBF_{ENABLE} with SPECT using Wilcoxon signed-rank test. Moreover, we used general linear models to evaluate the association between ASL-SPECT similarity and ASL SNR as well as number of retained ASL-diff images based on ENABLE.

Results

1. ENABLE training

Tables 1 and 2 show resultant quality weighting factors for 1.5T and 3.0T data, respectively. Large weights (i.e., numbers) in these quality matrices indicate greater influence for the corresponding quality feature. For instance, for PLD=100ms, *detectability* had the dominant influence on quality while *tSNR* had the dominant effect for PLD=900.

2. Evaluating ENABLE on simulated data

Simulations showed that CBF_{ENABLE} had greater similarity with CBF_{IDEAL} than did CBF_{CONV} ($p=0.03$). Mean free energy in GM improved with ENABLE compared to CONV ($p<0.001$).

3. Evaluating ENABLE on empirical data

Figure 2 shows quality features as a function of the amount of ASL data used. Figures 2A and 2B illustrate individual quality features and the quality criteria for three examples, respectively. Figure 2C shows mean ASL-diff images with different amounts of ASL data used. The last row shows the ASL-diff_{mean} images from ENABLE.

Sixteen out of ninety-eight datasets were removed from the empirical analysis due to unavailability of anatomical images or poor tissue segmentation/co-registration. ΔSNR_{GM} was significantly greater than zero, i.e., SNR was higher for CBF_{ENABLE} compared to CBF_{CONV} (Wilcoxon signed-rank test, $p<0.0001$). There was a significant effect of mean head motion on ΔSNR_{GM} ($N=82$, $t=4.3$, $p<0.001$) but no effect of field strength or diagnosis ($p>0.2$). Figure 3A shows the scatter plot of ΔSNR_{GM} and mean head motion of ASL images. Figure 3B illustrates CBF_{ENABLE} and CBF_{CONV} examples for visualization purposes. The mean of CBF_{ENABLE} in GM was higher than CBF_{CONV} (Wilcoxon signed-rank test, $p<0.001$). Model fit index also improved in ENABLE compared to CONV (Wilcoxon signed-rank test, $p<0.001$). The number of retained

ASL-diff images by ENABLE ranged from 79 to 143 out of 144 volumes (123 ± 12.5). We found a negative association between the number of retained ASL-diff images with mean head motion ($N=82$, $t=-3.9$, $p<0.001$) but no effect of field strength or diagnosis ($p>0.15$).

ASL-SPECT similarity was higher for CBF_{ENABLE} compared to CBF_{CONV} (Wilcoxon signed-rank test, $p<0.0001$). In addition, ASL-SPECT similarity was strongly associated with ASL SNR as shown in Figure 4A ($N=51$, $t=21$, $p<0.0001$). ASL-SPECT similarity was not related to the number of retained ASL-diff images based on ENABLE ($p>0.2$). Figure 4B illustrates ASL CBF examples of moderate and high similarity with SPECT.

Discussion

This study shows that multi-parametric ASL quality features can be used for automated image processing and produce higher quality CBF images. This work adds to the growing ASL literature focused on maximizing the use of ASL in dementia imaging (1). With ENABLE, we strived to identify corrupted ASL-diff images while preserving highest quality volumes. This approach was deemed successful on the basis of simulations and validation with SPECT CBF.

Compared to previous ASL image processing procedures that focused on single quality features (12–15), ENABLE used an optimized combination of four quality features. Each of these parameters represent a different ASL quality feature: tCNR shows the ASL signal changes in GM compared to non-brain voxels as the number of included images increases, tSNR and CoV demonstrate temporal and spatial variations of ASL signal in GM, respectively and detectability is a measure of proportion of voxels with meaningful ASL signal in GM. This combination had the effect of robust quality criterion that varied depending on the PLD. Negative weightings for CoV and tSNR indicated these features had an inverse relationship with quality criterion, whereas tCNR and detectability had a direct effect on quality criterion. In general, detectability

and tSNR had the greatest influence by virtue of their higher weightings compared to CoV and tCNR.

We validated ENABLE by two approaches: 1) simulations with the ground-truth CBF and 2) empirically by direct comparison to SPECT CBF. Simulations revealed that CBF_{ENABLE} had higher similarity with the ground-truth. Empirical ASL also supported ENABLE since CBF_{ENABLE} had a higher agreement with SPECT compared to CBF_{CONV} . These results argue that conventional ASL processing will yield inferior CBF images compared to more sophisticated image processing steps. Although all ENABLE steps, including simulations, were fully automated and can be executed for different protocols, the reported weighting factors can be used for similar ASL sequences. For acquisitions with crusher gradients or long label/long delay sequences, weighting factors may be optimized by the simulations and reported for future studies. A caveat of ENABLE would be heavily segmented 3D ASL where few if any ASL-diff images are available.

This study addresses some of the challenges of clinical ASL, where the trade-off between scan duration and image quality is paramount. This is particularly challenging with multi-PLD ASL since fewer pairs are collected at each PLD, of which we started with 6 to be consistent with a previous report on minimum number of required ASL-diff images (23). Our study showed increased SNR of CBF images after quality evaluation indicating that ENABLE can properly distinguish corrupted volumes from reliable ASL images. SNR improvement and number of retained ASL-diff images was related to mean head motion meaning that people who tend to move more gain more from the quality evaluation procedure. Since it is known that excessive head motion is prevalent in older adults and clinical populations (10), incorporating ENABLE maybe a valuable component of clinical ASL processing. It is, however, worth mentioning that

post-processing improvements may not always be sufficient to achieve high quality CBF maps. Figure 3B, representative 1 and Figure 4B, representative 3 are examples of poor acquisitions that could not be resolved retrospectively and hence blood flow estimates could not be enhanced during post-processing.

Our results suggest that accounting for ATT variability (i.e., multi-PLD acquisition) and removing spurious ASL-diff images results in moderate to high similarity between 1.5T ASL and SPECT – measured by generalized Jaccard index ranging from 0 (completely dissimilar) to 1 (identical) (32). This is in line with previous 3T ASL reports. For instance, moderate to high similarity between 3.0T ASL and ^{15}O positron emission tomography CBF images was reported in vascular and dementia cohorts (27,33). Wang et al. (26) and Qiu et al. (34) reported moderate similarity between multi-PLD ASL and CT in patients with delayed blood delivery i.e., long ATT. Liu et al. compared 3.0T ASL and SPECT in young healthy adults and reported a high similarity between CBF images (35). In addition, Sigurdsson et al. recently investigated the feasibility of 1.5T ASL on healthy elderly and reported a moderate to high levels of repeatability and reliability – comparable to younger populations (36). Altogether, these results show the feasibility of conducting ASL at 1.5T for dementia assessments when a 3T system is not available.

This study has some limitations. First, the quality weightings obtained from simulations were data driven and potentially subject to change depending on the sample i.e., age, diagnosis, etc. To examine this effect, we calculated quality weightings for random subgroups (size=20) and did not observe significant differences in weighting factors (data not shown). This suggests that the weightings are robust to the participant sample. A second limitation was the absence of SPECT images for participants imaged at 3.0T MRI system due to clinical workflow at our site at the

time of this study. Therefore, we were unable to investigate SPECT agreement at the more desirable ASL field strength of 3T. In the 1.5T ASL-SPECT comparison, we observed a strong correlation between ASL SNR and ASL-SPECT similarity. This accentuates the significance of PLD schedule design since participants with prolonged ATT may have intravascular artifacts decreasing the similarity between CBF modalities. Intravascular ASL artifact in CBF images results in low SNR as recently shown by Mutsaerts et al (24). In hindsight, a maximum PLD value of 2100ms in the current study may still be too short for some patients. Therefore, optimizing the multiple PLD ASL sequence parameters can decrease intravascular artifact and thus increase SNR, presumably leading to higher similarity with clinical CBF images. Further research is needed to optimize scan time by decreasing the number of ASL pairs in some PLDs (37) as well as optimizing PLD schedule in people with presumed delayed blood delivery.

In conclusion, we developed an image quality evaluation algorithm that works seamlessly within an ASL image processing pipeline. Executing ENABLE on empirical multi-PLD ASL images – acquired for clinical cohorts at 1.5T and 3.0T MRI systems – produced SNR improvements compared to conventional ASL analysis. In addition, ENABLE improved the agreement between ASL and SPECT CBF images. Our findings support the use of ASL at either 1.5T or 3T with multi-PLD acquisition and multi-parametric quality evaluation particularly in clinical populations where head motion and prolonged ATT are prevalent.

References

1. 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(1):102–16.
2. Nelson AR, Sweeney MD, Sagare AP, Zlokovic B V. Neurovascular dysfunction and neurodegeneration in dementia and Alzheimer’s disease. *Biochim Biophys Acta* 2015; 1862(5):887-900.
3. Iturria-Medina Y, Sotero RC, Toussaint PJ, Mateos-Perez JM, Evans AC, Alzheimer's Disease Neuroimaging Initiative. Early role of vascular dysregulation on late-onset Alzheimer’s disease based on multifactorial data-driven analysis. *Nat Commun* 2016;7:11934. doi: 10.1038/ncomms11934.
4. Dai W, Fong T, Jones RN, et al. Effects of Arterial Transit Delay on Cerebral Blood Flow Quantification Using Arterial Spin Labeling in an Elderly Cohort. *J Magn Reson Imaging* 2017;45:472–481.
5. MacIntosh BJ, Lindsay a C, Kylintireas I, et al. Multiple inflow pulsed arterial spin-labeling reveals delays in the arterial arrival time in minor stroke and transient ischemic attack. *AJNR Am J Neuroradiol* 2010;31(10):1892–4.
6. MacIntosh BJ, Marquardt L, Schulz UG, Jezzard P, Rothwell PM. Hemodynamic alterations in vertebrobasilar large artery disease assessed by arterial spin-labeling MR imaging. *AJNR Am J Neuroradiol* 2012;33(10):1939–44.
7. Liu Y, Zhu X, Feinberg D, et al. Arterial spin labeling MRI study of age and gender

- effects on brain perfusion hemodynamics. *Magn Reson Med* 2012;68(3):912–22.
8. Mutsaerts HJMM, van Dalen JW, Heijtel DFR, et al. Cerebral Perfusion Measurements in Elderly with Hypertension Using Arterial Spin Labeling. *PLoS One* 2015;10(8):e0133717.
 9. Mak HKF, Chan Q, Zhang Z, et al. Quantitative assessment of cerebral hemodynamic parameters by QUASAR arterial spin labeling in Alzheimer’s disease and cognitively normal Elderly adults at 3-tesla. *J Alzheimers Dis* 2012;31(1):33–44.
 10. Seto E, Sela G, McIlroy WE, et al. Quantifying head motion associated with motor tasks used in fMRI. *Neuroimage* 2001;14(2):284–97.
 11. Van Dijk KRA, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* 2012;59(1):431–8.
 12. Tan H, Maldjian JA, Pollock JM, et al. A fast, effective filtering method for improving clinical pulsed arterial spin labeling MRI. *J Magn Reson Imaging* 2009;29(5):1134–9.
 13. Shirzadi Z, Crane DE, Robertson AD, et al. Automated removal of spurious intermediate cerebral blood flow volumes improves image quality among older patients: A clinical arterial spin labeling investigation. *J Magn Reson Imaging* 2015;42(5):1377–85.
 14. Wang Z, Das SR, Xie SX, Arnold SE, Detre JA, Wolk DA. Arterial spin labeled MRI in prodromal Alzheimer’s disease: A multi-site study. *NeuroImage Clin* 2013;2:630–6.
 15. Dolui S, Wang Z, Shinohara RT, Wolk DA, Detre JA. Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series. *J Magn Reson Imaging* 2016;1–12.
 16. McKhann G, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging-

- Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7(3):263–9.
17. Chaudhary S, Scouten A, Schwindt G, et al. Hemodynamic effects of cholinesterase inhibition in mild Alzheimer's disease. *J Magn Reson Imaging* 2013;38(1):26–35.
 18. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage* 2012;62(2):782–90.
 19. Chappell MA, MacIntosh BJ, Donahue MJ, Gunther M, Jezzard P, Woolrich MW. Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI. *Magn Reson Med* 2010;63(5):1357–65.
 20. Makedonov I, Black SE, MacIntosh BJ. BOLD fMRI in the White Matter as a Marker of Aging and Small Vessel Disease. *PLoS One* 2013;8(7):1–9.
 21. 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(2):825–41.
 22. Liu TT, Wong EC. A signal processing model for arterial spin labeling functional MRI. *Neuroimage*. 2005;24(1):207–15.
 23. Fernández-Seara MA, Edlow BL, Hoang A, Wang J, Feinberg DA, Detre JA. Minimizing acquisition time of arterial spin labeling at 3T. *Magn Reson Med* 2008;59(6):1467–71.
 24. Mutsaerts HJ, Petr J, Václavů L, et al. The spatial coefficient of variation in arterial spin labeling cerebral blood flow images. *J Cereb Blood Flow Metab* 2017;271678X16683690.
 25. Liu Y, Zhu X, Feinberg D, et al. Arterial spin labeling MRI study of age and gender

- effects on brain perfusion hemodynamics. *Magn Reson Med* 2012;68(3):912–22.
26. Wang R, Yu S, Alger JR, et al. Multi-delay arterial spin labeling perfusion MRI in moyamoya disease--comparison with CT perfusion imaging. *Eur Radiol* 2014;24(5):1135–44.
 27. Tsujikawa T, Kimura H, Matsuda T, Fujiwara Y. Arterial Transit Time Mapping Obtained by Pulsed Continuous 3D ASL Imaging with Multiple Post-Label Delay Acquisitions : Comparative Study with PET-CBF in Patients with Chronic Occlusive Cerebrovascular Disease. *PloS One* 2016; 11(7): e0159894.
 28. Chappell MA, Groves AR, Whitcher B, Woolrich MW. Variational Bayesian Inference for a Nonlinear Forward Model. *IEEE Trans Signal Process* 2009;57(1):223–36.
 29. Dade LA, Gao FQ, Kovacevic N, et al. Semiautomatic brain region extraction: a method of parcellating brain regions from structural magnetic resonance images. *Neuroimage* 2004;22(4):1492–502.
 30. Bruyant PP. Analytic and iterative reconstruction algorithms in SPECT. *J Nucl Med* 2002;43(10):1343–58.
 31. Nakamura Y, Tomiguchi S. Usefulness of the Chang attenuation correction method with use of a CT-based μ map by FBP reconstruction in (201)Tl SPECT-MPI. *Ann Nucl Med* 2015;29(5):467–73.
 32. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging* 2006;25(11):1451–61.
 33. Kamano H, Yoshiura T, Hiwatashi A, et al. Arterial spin labeling in patients with chronic cerebral artery steno-occlusive disease: correlation with (15)O-PET. *Acta Radiol*

- 2013;54(1):99–106.
34. Qiu D, Straka M, Zun Z, Bammer R, Moseley ME, Zaharchuk G. CBF measurements using multidelay pseudocontinuous and velocity-selective arterial spin labeling in patients with long arterial transit delays: Comparison with xenon CT CBF. *J Magn Reson Imaging* 2012;36(1):110–9.
 35. Liu P, Uh J, Devous MD, Adinoff B, Lu H. Comparison of relative cerebral blood flow maps using pseudo-continuous arterial spin labeling and single photon emission computed tomography. *NMR Biomed* 2012;25(5):779–86.
 36. Sigurdsson S, Forsberg L, Aspelund T, Geest RJ Van Der. Feasibility of Using Pseudo-Continuous Arterial Spin Labeling Perfusion in a Geriatric Population at 1.5 Tesla. *PloS One* 2015;10(12):e0144743.
 37. Kramme J, Gregori J, Diehl V, Madai VI, von Samson-Himmelstjerna FC, Lentschig M, et al. Improving perfusion quantification in arterial spin labeling for delayed arrival times by using optimized acquisition schemes. *Z Med Phys Elsevier B.V.* 2015;25(3):221–9.

Table 1: Quality weightings from the optimization process for 1.5T based simulated data.

<i>Quality features</i>	<i>tCNR</i>	<i>Det</i>	<i>CoV</i>	<i>tSNR</i>
<i>PLD=100ms</i>	0.5	1.3	-0.5	-1.2
<i>PLD=500ms</i>	0.7	1.2	-0.7	-1.0
<i>PLD=900ms</i>	0.7	1.0	-0.3	-1.1
<i>PLD=1300ms</i>	0.8	0.8	-0.4	-1.0
<i>PLD=1700ms</i>	0.2	1.5	-0.6	-1.0
<i>PLD=2100ms</i>	0.1	1.6	-0.7	-1.0

tCNR: temporal contrast to noise ratio, Det: detectability, CoV: coefficient of variation, tSNR: temporal signal to noise ratio.

Table 2: Quality weightings from the optimization process for 3.0T based simulated data.

<i>Quality features</i>	<i>tCNR</i>	<i>Det</i>	<i>CoV</i>	<i>tSNR</i>
<i>PLD=100ms</i>	0.4	1.6	-0.9	-1.1
<i>PLD=500ms</i>	0.3	1.7	-0.9	-1.0
<i>PLD=900ms</i>	0.7	1.0	-0.3	-1.1
<i>PLD=1300ms</i>	0.1	1.8	-1.0	-1.0
<i>PLD=1700ms</i>	0.5	1.4	-0.8	-1.0
<i>PLD=2100ms</i>	0.3	1.8	-0.6	-0.8

tCNR: temporal contrast to noise ratio, Det: detectability, CoV: coefficient of variation, tSNR: temporal signal to noise ratio.

Figure legends:

Figure 1: Ideal “noise-free” ASL images (MNI152; 2mm template space). The ground-truth cerebral hemodynamic images (A) generated ideal “noise-free” ASL difference images (B) based on a two compartmental ASL kinetic model. These images served as reference images in the simulations. The numbers in each column show post labeling delay values, i.e., 100, 500, 900, 1300, 1700 and 2100ms. We simulated noisy ASL datasets by adding noise features from empirical data.

Figure 2: Quality criteria examples for ASL acquired at PLD=1700ms. A: Quality features with respect to the number of included ASL-diff images. (tCNR: temporal contrast to noise ratio, Det: detectability, CoV: coefficient of variation, tSNR: temporal signal to noise ratio) B: weighted summation of quality features using weightings obtained from the optimization process. C: mean ASL-diff images when 6, 12, 18 or 24 ASL-diff images were considered. The last row shows the ENABLE result along with the corresponding number of included ASL-diff images. ASL signal increase is visible in the last row (ENABLE) compared the second last row, which uses all pairs and is defined as the conventional approach (CONV). In addition, head motion related rim enhancement decreases in ENABLE compared to CONV.

Figure 3: Empirical comparison of CONV and ENABLE. A: SNR improvement of CBF images with respect to mean head motion. $\Delta\text{SNR}_{\text{GM}}$ was significantly greater than zero and related to mean head motion. B: Three examples of CBF_{CONV} and $\text{CBF}_{\text{ENABLE}}$ images. The numbers of retained ASL differences images based on ENABLE were 92, 113 and 127 out of 144 volumes for representatives 1, 2 and 3, respectively. Representative 1 was an outlier removed from panel A; MCFLIRT estimated mean head motion of 1.1 mm for this participant. CBF increase is

visible in CBF_{ENABLE} compared to CBF_{CONV} for representatives 1 and 2. Head motion related rim enhancement is diminished for representative 2.

Figure 4: ASL-SPECT comparison. A: The similarity measure (Jaccard Index) is strongly correlated with the ASL SNR. B: Three examples of intensity normalized ASL and SPECT CBF images. The number of retained ASL difference images based on ENABLE were 130, 133 and 116 out of 144 volumes for representatives 1, 2 and 3, respectively. The intravascular artifact is visible in ASL CBF images with lower ASL-SPECT similarity e.g., representative 3.