Arterial spin labeling perfusion MRI: Inter-vendor reproducibility and clinical applicability
Mutsaerts, H.J.M.M.
General discussion, conclusions and implications
Part I Inter-vendor reproducibility of arterial spin labeling

Throughout the lines of arterial spin labeling (ASL) developments, reliability assessments of ASL remained a central theme because of its intrinsically low signal-to-noise ratio (SNR). Part I of this thesis can be viewed as a sequel of the previous work of dr. Gevers et al. and dr. Heijtel et al. The first have established that the single-vendor intra- and inter-scanner reproducibility of ASL were sufficient and comparable. The latter have established that the reproducibility of ASL is comparable with H2O15-positron emission tomography (PET) when a two-compartment quantification model was used. These findings suggested that the measurement reproducibility of the state-of-the-art ASL sequence is sufficiently high, and that the variability of ASL may be dominated by physiological perfusion fluctuations rather than technical limitations. Since different magnetic resonance imaging (MRI) centers use scanners from different MRI vendors, the next logical step towards large clinical multi-center perfusion studies was to investigate the inter-vendor reproducibility of ASL.

The incentive to study the inter-vendor reproducibility of ASL (Chapters 2-4) was the possibility to pool ASL-based cerebral blood flow (CBF) results from two scanners from different MRI manufacturers. Although the labeling strategy implemented in the product sequence of both vendors was the same (pseudo-continuous ASL (PCASL)), the readout was fundamentally different, 3D spiral fast-spin echo (FSE) (General Electric (GE)) versus 2D gradient-echo echo-planar imaging (EPI) (Philips). The first visually appreciable difference between these sequences on both the individual subject level and on the group level was the difference in the spatial smoothing, leading to substantial mean CBF and inter-session variation differences between vendors on a voxel level. However, on a total gray matter (GM) level, there was excellent agreement between the mean CBF of both vendors. In addition, the intra-vendor inter-session variation of both vendors as well as the inter-vendor inter-session variation were comparable on a total GM level, and comparable with the between-weeks variation of several previous reproducibility studies. These findings led to the encouraging conclusion that baseline CBF values can be compared between fundamentally different sequences from different vendors on a total GM level (Chapter 2). This can be attributed to the fact that baseline long-term physiological perfusion fluctuations dominate sequence and hardware differences between vendors. The large differences on a voxel level were mainly attributed to the fundamental differences in readout modules implemented by both vendors. This is a clear obstacle for the comparison of ASL-based data between the different product sequences from the two vendors, which led to the initiation of the study in Chapter 5.
Similar to the baseline CBF comparison, the inter-session variation of a pharmacologically induced CBF decrease was comparable between sequences on a total GM level but not on a voxel level (Chapter 3). This again implies that on the total GM level, the inter-vendor variability is dominated by normal long-term perfusion fluctuations, which is in agreement with previous studies \(^1,9,11\). However, on a voxel level the readout differences between the product sequences dominated the perfusion variability. We showed that by additional smoothing of the 2D EPI data to achieve the same smoothness as the 3D spiral data, we were able to approach the inter-session variation that was observed for the 3D spiral readout. This is in agreement with the large difference in point spread function (PSF) between the sequences, as can be easily appreciated by visually comparing the CBF maps of both sequences \(^12\). In addition, it shows the potential benefit of additional spatial smoothing for the analysis of 2D EPI pharmacological ASL (phASL) data \(^13\). For the detection of small focal changes - such as in finger tapping functional ASL (fASL) (Chapter 4) - however, it remains questionable whether or not additional smoothing removes focal activation regions \(^14\). The residual variability differences between the 3D spiral data and smoothed 2D EPI data may be explained by the differences in background suppression efficiency. This was indicated in a previous study that studied differences in activation detectability between a 2D and two 3D readouts without background suppression, as well as the same two 3D readouts with background suppression \(^12\). This study showed that the detectability differences between 2D and 3D readouts themselves were small, but also that the detectability of the 3D readout increased significantly by the use of background suppression. Although we employed background suppression in both the 2D EPI and 3D spiral readouts in our studies, the efficiency of background suppression in a 2D readout steadily decreases with later acquired slices \(^15\).

Both the mean phASL CBF decrease (Chapter 3) as well as the fASL CBF increase (Chapter 4) differed between the product sequences, although this difference was not significant for the latter. Interestingly in this respect, was that the main voxel-wise CBF differences between sequences in Chapter 2 and 3 were found in the posterior watershed area. These regions are known for having the longest transit times (TTs) \(^4,16\). Because the 3D sequence readout has a single post-labeling delay (PLD) whereas the PLD of the 2D EPI sequence increases along an inferior-superior gradient, both sequences differ much in their TT sensitivity. It is most likely that the abovementioned cases of CBF disagreement between the product sequences can be explained by the differences in TT sensitivity between both readout modules. This is in agreement with a previous comparison between the performance of a 3D spiral readout with either a PLD of 1500 or 2500 ms \(^17\). Regions with long TT showed hypoperfusion when the 1500 ms PLD was used, but not with the 2500 ms PLD. In addition, reproducibility was higher for 2500 ms PLD, suggesting that TT contributes in variability when a short PLD is chosen. From these observations, we
may conclude that TT can influence both the accuracy and the precision of ASL measurements, and that its influence depends upon the employed readout module. This strongly encourages the development of sequences that simultaneously measure CBF and TT without significant penalties in SNR for the CBF measurement\textsuperscript{18,19}.

Using previously proposed methods\textsuperscript{20}, we provided sample size calculations for future phASL studies in Chapter 3. We underlined that sample size requirements for the detection of any effect may be small, but that the requirements for detecting differences in an effect are quadratically larger. Although this observation is easily deduced from sample size equations, we believe that it is important to acknowledge this observation when planning future phASL or fASL studies. The same effect was observed for the within-subject coefficient of variation (wsCV) in Chapters 3 and 4. Whereas the inter-session variation of the absolute CBF decrease was not very different from baseline inter-session CBF variation in both cases, the small effect size of phASL (~20%) and fASL (~4%) increased the wsCV fivefold and 25-fold respectively. As a consequence, the finger tapping fASL activation (Chapter 4) was barely statistically detectable. It should be noted, however, that this low detectability should not discourage the use of ASL to detect task-based CBF changes\textsuperscript{12,21}. Whereas the temporal resolution of ASL for the 2D EPI sequence was reasonable (8 seconds), it was very low for the 3D spiral sequence (135 seconds). For a fair comparison, the 2D EPI data was averaged to obtain the same temporal resolution. Although the SNR of an individual 2D EPI ASL subtraction is very low, a higher temporal resolution may still benefit statistical modeling. Moreover, current efforts are aimed at increasing the temporal resolution of 3D ASL sequences while retaining its benefits in terms of higher SNR\textsuperscript{22}.

Chapter 5 concerned a triple vendor comparison. Whereas the previous chapters aimed to investigate the penalty of using fundamentally dissimilar standard ASL implementations of different vendors, the study in this chapter aimed to investigate the inter-vendor reproducibility when sequences are optimized to be as identical as possible between the scanners. The first and most apparent result when the near-identical sequences (Chapter 5) are compared with the dissimilar sequences (Chapter 2), is that both the mean CBF and inter-session variation maps were visually much more comparable between scanners from different vendors for near-identical sequences than for the dissimilar sequences. Likewise, the GM-white matter (WM) CBF ratios were significantly different between dissimilar sequences whereas these ratios were very comparable with near-identical sequences. This was in agreement with our hypothesis that ASL results are much more comparable between vendors when the same readout and similar sequence parameters are used\textsuperscript{12}.
A main limitation of this study was that scanner hardware differences were too large to create a single identical sequence for the three scanners of the different vendors. The reason for these differences in hardware specifications was that MRI hardware development is ongoing, and one scanner was older than the two others. The solution was to split the triple vendor comparison into two dual vendor comparison studies. For scanners with similar hardware specifications it would have been possible to create the same near-identical sequence for the scanners from all three vendors. However, this main limitation led to an important serendipitous finding. Because the hardware specifications of one scanner were flexible, it participated in both comparisons. Consequently, the results of these slightly different sequences could be compared on the same vendor. Whereas the total GM inter-session variation of study 2 was comparable with the inter-session variation in Chapter 2, and with other previous between-weeks reproducibility studies, the inter-session variation of study 1 was twice as high as in study 2. In other words, the precision was more comparable between identical sequences employed on scanners from different vendors than between slightly less identical sequences on the same scanner. The same was true for the accuracy. The quantitative agreement seemed much less affected by vendor effects (differences in scanner hardware and software) than by sequence effects, which were mainly differences in echo time and inter-pulse time during labeling. These results demonstrate the importance of keeping sequence parameters identical for future multi-center ASL studies. In addition, the same message is very important for clinical single-site ASL studies. Scanner hardware and software are often updated halfway in an ongoing clinical study, and these results provide strong evidence that it is very important to keep sequence parameters as identical as possible within the limitations created by such scanner updates.

Part II Clinical applicability of arterial spin labeling

Despite its potential relevance as a micro-vascular biomarker, ASL-based WM CBF measurements were previously not deemed feasible for two main reasons. Firstly, because of too low SNR due to longer TT than in GM, and secondly because of too low spatial resolution to acquire pure WM signal, i.e. which is not contaminated by signal from the GM. Recently, SNR improvements rendered the detection of WM perfusion signal possible in the majority of WM voxels. However, WM voxels with the highest and most reliably measured perfusion signal are situated in the juxtacortical peripheral WM, whereas signal in deep central WM voxels are not measurable even after 10 minutes of scanning. These findings raise the question to what extent they are due to a real perfusion gradient from the peripheral to central WM, or rather to the spatial extent of signal contamination of GM CBF. Signal contamination can be expected to result from a combination of acquisition PSF, motion and...
interpolation in the post-processing stage \(^{12, 14, 29}\). Moreover, as the SNR of ASL within the WM remains relatively poor \(^{30}\), it is important to balance between including as many voxels as possible into a WM region of interest (ROI) to increase the statistical power of WM CBF on the one hand, and excluding as many peripheral WM voxels to exclude GM signal contamination on the other hand \(^{31}\).

In Chapter 6, perfusion signal of WM voxels medial to the cortical GM (inward contamination) were compared with perfusion signal of voxels lateral to the cortical GM (outward contamination) within a single slice. This analysis is based on the assumption that the extent of inward and outward GM signal contamination is similar. Our analyses showed that significant outward GM signal contamination stretched across 3 voxels. Therefore, the conclusion was that a typical WM mask should be eroded three-fold to create a WM ROI that holds as many relatively uncontaminated WM voxels as possible. However, it should be acknowledged that the spatial extent of signal contamination may differ much between studies or even between subjects, because of differences in acquisition, subject motion and post-processing methods \(^{12, 14, 29}\). Therefore, the proposed three-voxel erosion guideline should be customized to the data under investigation. Nevertheless, after three-voxel erosion the WM signal was still significantly larger than the signal outside the GM (i.e. cerebrospinal fluid (CSF), bone and air). This observation confirms previous research \(^{28, 30}\) and shows the possibility to significantly detect uncontaminated perfusion signal in the WM with ASL. However, this signal could still be too weak to be able to make statistical inferences in group analyses.

There is accumulating evidence that TT measurements may both significantly improve the quantification of CBF \(^7, 32\) and provide additional information that could even be more valuable than CBF in some clinical cases \(^{33-36}\). One way to estimate TT is to calculate the ratio of ASL sequences with and without vascular crushing, which is known as flow encoding arterial spin tagging (FEAST) \(^{37}\). In Chapters 7 and 8, FEAST was applied with slightly adapted parameters, to measure TT as well as both crushed and non-crushed CBF. The application of vascular crushing removes some of the macro-vascular contribution to CBF, which may be less of interest than the pure micro-vascular tissue perfusion signal \(^{38}\). On the other hand, when macro-vascular signal is removed this will also reduce the available SNR \(^{37}\). On the individual subject level it is important to retain as most SNR as possible, especially in elderly with neurodegenerative or cerebrovascular disease. Therefore, crushing is not recommendable for clinical applications of ASL on the single subject level \(^7\). For group analyses, it may be a different case, as we showed in Chapter 7. In this study, correlations with age or gender were stronger for crushed CBF-values than without crushing. It should be noted that the applied crushing was moderate (i.e. velocity encoding cutoff 5 cm/s), removing only signal in the largest vessels with minor SNR penalty. When
vascular crushing is applied to completely remove macro-vascular perfusion signal, a lower velocity cutoff is used (e.g. velocity cutoff 1-3 cm/s) and the resulting SNR penalty may be too severe for applications in elderly patients. Therefore, if any crushing is applied, the consensus paper recommended to apply moderate crushing (4 cm/s) which is close to the velocity cutoff that was used in these chapters. The observation that slight crushing increased statistical power, suggests that it can be advantageous to decrease the influence of macro-vascular perfusion fluctuations on the reproducibility of the ASL measurement - even at the expense of some SNR.

It could be argued that with optimal sequence timing - i.e. labeling duration and PLD - macro-vascular contribution is minimal and crushing may not be required. In this respect, the mean PLD applied in these chapters was relatively low (1800 ms vs. 2000 ms as recommended in elderly patients). However, TT can be expected to vary more in elderly patients than in healthy volunteers, because of differences in vessel tortuosity, vascular resistance and cerebral autoregulation deficiency. Therefore, it may remain difficult - if not impossible - to exactly confine ASL timing parameters to the subject that is being scanned. The physiological correlations with the estimated TT were even stronger than with the crushed or non-crushed CBF measurements - not only for the correlations with age and gender (Chapter 7) but also with white matter lesion (WML) volume (Chapter 8). There could be good physiological reasons for this observation, including the fact that the preferential location of WML is in regions with the longest TT, such as watershed areas. However, to which extent these stronger correlations with crushed CBF and TT had physiological origins, or were merely due to a decrease in the effect of macro-vascular perfusion fluctuations on the measurement variability, cannot be differentiated from our data.

Interestingly, a recent study found lower measurement variability for TT than for CBF, using a multi-PLD PCASL sequence. This suggests that the stronger correlations with TT are at least partly due to higher stability of TT compared to CBF measurements. It is noteworthy that new ASL techniques are currently under development that can simultaneously measure both CBF and TT - some even without sacrificing SNR for the single time point CBF map. Moreover, these techniques enable a more absolute quantification of TT, whereas the quantification of FEAST-based TT depends much on PLD and CBF velocity differences between subjects and brain regions. These new ASL techniques are probably preferable over FEAST for clinical research applications, and can be expected to enhance the value of ASL even further.

Despite these limitations of FEAST in terms of absolute quantification, we were able to construct TT maps that resemble the known anatomy of the cerebral vasculature. Therefore, the spatial accuracy of FEAST is at least sufficient to create a population-based TT gradient. These maps can be helpful in
future ASL research for the definition of vascular ROIs, if hypothesized perfusion effects are restricted to certain flow territories only, or if spatial averaging is required when an anatomical structure is too small considering ASL limitations in terms of SNR or spatial smoothing. Potential applications include the investigation of a vascular component to degeneration of the dementia-relevant regions precuneus and hippocampus, which are supplied by the distal anterior and posterior cerebral arteries. Alternatively, if no TT measurements are available, our ROI-based TT values can be used to improve CBF quantification for regional TT variation in the elderly.

In Chapters 8 and 9, we investigated correlations between CBF measured within GM, normal appearing WM (NAWM) and WML ROIs and WML volume, as well as with other clinical and hematological parameters. These measurements were performed in elderly with hypertension (Chapter 8) and children with sickle cell disease (SCD) (Chapter 9). It has been suggested that the WML in both populations share a similar cerebrovascular etiology. In Chapter 8, GM CBF was higher than NAWM CBF, which was higher than WML CBF. Interestingly, the ratios between GM, NAWM and WML CBF were comparable to ratios that were previously reported with contrast-based perfusion MRI. Although SNR may be too low for sufficient precision of ASL-based WM CBF measurements in the elderly, the accuracy of ASL as compared to contrast-based perfusion MRI appears sufficient within the WM. The fact that WM CBF was correlated with hematological parameters in children with SCD (Chapter 9) shows that WM CBF measurements can be sufficiently reliable for clinical correlations. This can be explained by the shorter TT and higher CBF in children with SCD, leading to higher SNR for ASL-based WM CBF measurements. Furthermore, physiological perfusion fluctuations are lower for WM than for GM CBF, which may also increase statistical power. These findings show potential for ASL-based perfusion measurements within the NAWM and WML ROIs in children with SCD.

The clinical question in Chapter 8 was whether WML volume is associated with either isolated focal or also with more general perfusion deficits, using the methodologies that were introduced in Chapters 6 and 7. Both hypotheses are deemed plausible but not yet fully elucidated. Hence, it is hitherto unknown which of these two hypotheses should be given priority in terms of further research or treatment development. In our study, WML volume was correlated with WML CBF but not with NAWM CBF or GM CBF. As abovementioned, the fact that CBF was lower within WML regions than in regions with NAWM is in agreement with previous literature based on MRI perfusion measurements with exogenous contrast agents. The correlation between WML CBF and WML volume can - at least partly - be attributed to various degrees of ASL signal contamination from the NAWM into the WML, especially considering the large spatial extent of signal contamination as discussed in Chapter 6.
To what extent the WML CBF is truly correlated to WML volume cannot be differentiated with these data.

The absence of a correlation between GM CBF and WML volume in Chapter 8 disagrees with some previous studies $^{26, 51, 52}$. Possible explanations include that the strength of this correlation depends upon the study population, techniques used for the detection of WML and measurement of CBF, the presence of perfusion confounders and the type of statistical analysis. Future research should aim at disentangling effects generated by the multitude of diseases that are associated with aging, such as cardiovascular disease, large- and small vessel cerebrovascular disease and neurodegeneration. Although the studied cohort was restricted to community dwelling elderly with hypertension without any serious medical conditions, the abovementioned diseases can still be expected to be present in such a population. Another question that is left to be answered is whether the correlation between GM CBF and WML volume found in other populations is of a causal nature or that the two are simply synchronous consequences of a worsening condition of the cerebral vasculature.

In Chapter 9, these WML are discussed in the context of children with SCD. Similar to WML in the elderly, two hypotheses exist that could explain the development of WML. The first relates to endothelial dysfunction $^{61, 62}$. In SCD, sickled red blood cells and other inflammatory mediators induce the activation of the endothelium, leading to impaired vasodilation and increased adhesiveness of the endothelium $^{63}$. Markers of endothelial dysfunction, such as ADAMTS13, have been shown to be correlated with WML in SCD $^{64}$. Our findings show that this marker is also correlated to CBF. The second pathophysiological hypothesis is related to the loss of cerebrovascular reactivity $^{53}$: the severe chronic anemia in these children can apparently be mostly compensated by increased CBF due to increased cardiac output and cerebral vasodilation $^{54, 55}$. However, this means that CBF cannot be increased much more in periods of increased demand for oxygen and nutrients to sustain cerebral metabolism. In other words, the cerebrovascular reactivity is reduced in these children, and moments of transient ischemia may lead to the development of WML $^{53, 56}$. Fetal hemoglobin (HbF) is a type of hemoglobin that does not polymerize and reduces the concentration of pathological hemoglobin S (HbS) $^{58, 59}$. Our results showed that children with lower fetal hemoglobin (HbF) levels had both higher CBF and higher WML volume. These findings could suggest that children with lower HbF levels require a higher compensatory CBF increase to sustain their cerebral metabolism. In addition, they suggest that low HbF levels and high CBF lead to more WMLs development, possibly as a result of a more impaired cerebrovascular reactivity. Therefore, although we did not find a direct correlation between CBF and WML volume, these findings are in agreement with the cerebrovascular reactivity hypothesis $^{57}$. It should
be noted that part of the investigated population was treated with hydroxyurea therapy. This therapy increases the level of HbF, which is known to reduce several symptoms of this disease with less side effects than chronic blood transfusion therapy. Our finding that CBF is inversely correlated with HbF implicates that less compensatory CBF increase is required with higher HbF levels. Hence, cerebrovascular reactivity can potentially be restored by increasing HbF levels. From these findings we can hypothesize that patients with high baseline CBF benefit the most by increasing HbF levels with hydroxyurea therapy. In this respect, ASL could potentially develop as a non-invasive biomarker.
Conclusions

Part I  Inter-vendor reproducibility of ASL
1 On a voxel level, CBF measurements from different vendors are only comparable when the same readout modules are used (Chapters 2-5).
2 Differences between 2D and 3D ASL sequences can be mainly explained by differences in TT sensitivity, PSF and efficiency of background suppression (Chapters 2-4).
3 Differences in ASL sequence parameters should be avoided, since even slight differences in ASL sequence parameters can have a large effect upon the reproducibility (Chapter 5)

Part II  Clinical applicability of ASL
4 Uncontaminated WM perfusion signal can be detected with ASL, if tissue masks are carefully eroded to minimize GM contamination (Chapter 6).
5 The statistical power of ASL may be increased by moderate vascular crushing, suggesting that removing macro-vascular perfusion variability can be more important than holding on to sufficient SNR (Chapter 7).
6 Information from TT measurements may carry important diagnostic value and simultaneous measurements of CBF and TT should be carried out when possible (Chapters 7-8).
7 WML volume is correlated with WML CBF but not with normal appearing WM or GM CBF in elderly with hypertension, suggesting that WML development is the result of local rather than systemic perfusion disturbances (Chapter 8).
8 WM CBF is correlated with clinically relevant hematological parameters in children with SCD, alluding to the possible development of ASL-based CBF measurements as a non-invasive biomarker in this population (Chapter 9).
General discussion, conclusions and implications

Implications

The conclusions of the inter-vendor reproducibility studies (Part I) have the following implications for the development of ASL for clinical studies. We have demonstrated the importance of keeping ASL sequence parameters identical, in order to be able to pool ASL-based CBF data (Chapters 2-5). An important practical single-center example is the update of scanner hardware or software in the middle of a cross-sectional or longitudinal clinical trial, which may lead to or require small changes in ASL sequence parameters. These results encourage investigators to carefully tweak sequence parameters to keep them as identical as possible between the scans before and after the scanner update. The same should be attempted when multi-center studies are being planned. When this is not possible - e.g. when pooling data from existing multi-center ASL studies where product sequences have been used - investigators are encouraged to acknowledge the effects of using different sequences. Main examples that result from different readout modules are discussed in this thesis, including TT sensitivity (Chapters 2-3), PSF (Chapter 2) and background suppression efficiency (Chapters 2-4). For clinical practice, absolute quantitative agreement between ASL sequences may not be required. Therefore, slight sequence differences may be less of an issue here and we recommend instead to only focus on using the same readout when visually comparing CBF maps between different clinical centers. Alternatively, it may be helpful to remove data smoothness differences (Chapter 3) that result from the differences in PSF between product sequences. This could be easily implemented as post-processing option in medical imaging software.

The conclusions of the clinical applicability studies (Part II) have the following implications for the development of ASL for clinical studies. We have demonstrated the possibility to estimate the effective spatial extent of GM signal contamination within acquired data (Chapter 6). We showed the possibility to measure uncontaminated WM signal, but also highlighted the importance to acknowledge this spatial extent of signal contamination. The findings of Part I imply that, as ASL technology advances, the reliability of ASL will at some point be limited by physiological perfusion fluctuations rather than measurement precision. These findings are reinforced by the findings in elderly with hypertension in Chapter 7, implying that it can be more important to reduce physiological perfusion fluctuations than to hold on to sufficient SNR in large clinical studies. Although it has been shown for individual clinical cases, the results of Chapters 7 and 8 are the first to imply that TT measurements can be more important than CBF measurements in large clinical studies. This finding encourages research physicians to focus on more perfusion parameters than CBF only and stimulate the current development of sequences that can simultaneously measure CBF and TT. Chapters 8 and 9 provide examples on how
ASL perfusion measurements can bridge the research gap between clinical or hematological parameters and accumulated cerebral pathology such as WML.

When the study subjects and findings of this thesis are compared with previous ASL theses, we can conclude that ASL steadily advances towards a more stable position within clinical research. In this respect, the limitations of ASL are shifting from fundamental sequence development limitations to sequence choices by MRI vendors, physiological perfusion fluctuations and the confounding effect of TT. Current developments aimed at WM perfusion measurements, the simultaneous measurement of CBF and TT and the reduction of the influence of physiological perfusion fluctuations are expected to increase the clinical value of ASL even further. Future efforts aimed at the creation of normal and pathological perfusion templates may eventually enable the development of ASL as clinical and psychopharmacological biomarker.

Reference list


33. MacIntosh BJ, Graham SJ. Magnetic resonance imaging to visualize stroke and characterize stroke recovery: a review. *Front Neurol* 2013; **4**:60.


