


Automated Quality Evaluation Index for Arterial Spin Labeling Derived Cerebral Blood Flow Maps

Sudipto Dolui, PhD,^{1*}  Ze Wang, PhD,² Ronald L. Wolf, MD, PhD,^{1,3} Ali Nabavizadeh, MD,¹ Long Xie, PhD,⁴ Duygu Tosun, PhD,⁵ Ilya M. Nasrallah, MD, PhD,¹ David A. Wolk, MD,⁶ and John A. Detre, MD,^{1,6} for the Alzheimer's Disease Neuroimaging Initiative

Background: Arterial spin labeling (ASL) derived cerebral blood flow (CBF) maps are prone to artifacts and noise that can degrade image quality.

Purpose: To develop an automated and objective quality evaluation index (QEI) for ASL CBF maps.

Study Type: Retrospective.

Population: Data from N = 221 adults, including patients with Alzheimer's disease (AD), Parkinson's disease, and traumatic brain injury.

Field Strength/Sequence: Pulsed or pseudocontinuous ASL acquired at 3 T using non-background suppressed 2D gradient-echo echoplanar imaging or background suppressed 3D spiral spin-echo readouts.

Assessment: The QEI was developed using N = 101 2D CBF maps rated as unacceptable, poor, average, or excellent by two neuroradiologists and validated by 1) leave-one-out cross validation, 2) assessing if CBF reproducibility in N = 53 cognitively normal adults correlates inversely with QEI, 3) if iterative discarding of low QEI data improves the Cohen's d effect size for CBF differences between preclinical AD (N = 27) and controls (N = 53), 4) comparing the QEI with manual ratings for N = 50 3D CBF maps, and 5) comparing the QEI with another automated quality metric.

Statistical Tests: Inter-rater reliability and manual vs. automated QEI were quantified using Pearson's correlation. $P < 0.05$ was considered significant.

Results: The correlation between QEI and manual ratings ($R = 0.83$, CI: 0.76–0.88) was similar ($P = 0.56$) to inter-rater correlation ($R = 0.81$, CI: 0.73–0.87) for the 2D data. CBF reproducibility correlated negatively ($R = -0.74$, CI: -0.84 to -0.59) with QEI. The effect size comparing patients and controls improved ($R = 0.72$, CI: 0.59–0.82) as low QEI data was discarded iteratively. The correlation between QEI and manual ratings ($R = 0.86$, CI: 0.77–0.92) of 3D ASL was similar ($P = 0.09$) to inter-rater correlation ($R = 0.78$, CI: 0.64–0.87). The QEI correlated ($R = 0.87$, CI: 0.77–0.92) significantly better with manual ratings than did an existing approach ($R = 0.54$, CI: 0.30–0.72).

Data Conclusion: Automated QEI performed similarly to manual ratings and can provide scalable ASL quality control.

Evidence Level: 2

Technical Efficacy: Stage 1

J. MAGN. RESON. IMAGING 2024.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.29308

Received Aug 18, 2023, Accepted for publication Feb 6, 2024.

*Address reprint requests to: S.D., 3710 Hamilton Walk, Goddard Laboratories, Office 321, Philadelphia, PA 19104, USA.

E-mail: sudiptod@pennmedicine.upenn.edu

Some of the data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

From the ¹Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA; ²Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; ³Department of Neurosurgery, University of Pennsylvania, Philadelphia, Pennsylvania, USA; ⁴Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA;

⁵Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, California, USA; and ⁶Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Arterial spin labeled (ASL) perfusion MRI^{1,2} provides a noninvasive quantification of regional cerebral blood flow (CBF), which is an important physiological variable reflecting cerebrovascular health and a biomarker of regional brain function. ASL can be acquired as part of routine, non-contrast brain MRI and therefore has become the modality of choice to measure CBF in several large-scale multisite studies with measurements of brain structure and function.^{3–6} ASL-MRI has been validated against other established modalities for measuring CBF^{7–10} and has been extensively used to assess CBF changes in health and diseased conditions.^{3,6,10–14}

ASL MRI data acquisition involves magnetically labeling inflowing arterial blood water protons proximal to the brain, waiting for a brief period (post-labeling delay [PLD]) to allow the labeled blood to flow to the brain, acquiring a brain image (labeled image), and subtracting the labeled image from a control image that is acquired without labeling arterial blood. The control-labeled difference image is proportional to CBF and can be converted to a quantitative CBF map using a proton density image with appropriate models and assumptions.^{1,15} However, the control-label difference is a small percentage of the background signal which results in low signal to noise ratio in the CBF image. Additionally, subject motion, suboptimal choice of imaging parameters, and other non-idealities inherent to MRI scanners can lead to severe artifacts.^{4,16–20} Averaging multiple control-label pairs, using advanced signal processing strategies,^{4,16–18} and utilizing advanced ASL techniques such as background suppressed 3D acquisitions^{21–24} can improve ASL image quality, but considerable artifacts can still be present in resulting CBF maps particularly in non-compliant patient populations or when acquisition parameters are suboptimal.

Despite the advantages of ASL as a biomarker of cerebrovascular health and brain function, poor image quality can limit its sensitivity, potentially resulting in inconsistent or erroneous conclusions. Hence, it is ideal to exclude CBF maps of poor quality from analysis. However, manual quality control (QC) of ASL CBF maps based on visual inspection is subjective and susceptible to user bias, laborious to perform in large samples and requires considerable expertise. There is a need for an objective approach to evaluate the quality of CBF maps that is scalable to a large sample size and can potentially be applied to clinical scanning.

Prior work on ASL QC has mainly focused on identifying and discarding outlier control-labeled image pairs based on empirical methods,^{4,17,18,25} but not on assessing the quality of the final CBF map. One study has proposed computing the standard deviation of the perfusion weighted time series, dividing by the square root of the number of pairs to obtain a measure of standard error, averaging across the whole brain, normalizing by the whole brain averaged difference signal, and thereafter mapping it to a score between 1 and 4 to obtain the final quality index for the mean CBF map.²⁶

However, normalized standard error cannot assess systematic artifacts that are consistent in the time series such as those caused by short PLD, or in datasets that include only one output volume of the average control-labeled difference image rather than the control-labeled image time series (eg, product ASL on a GE MRI scanner). Moreover, temporal standard error quantifies the quality of the raw data and does not capture the quality of the final map which can sometimes be of better quality than the raw data through application of optimized signal processing strategies.^{4,17,18}

In this work, we aimed to develop a metric for automatic quantification of the quality of a CBF map, which we have referred to as an automated quality evaluation index (QEI). Incorporation of features of a corrupted CBF map, such as its dissimilarity with brain structure, large dispersion of CBF values within tissues and presence of voxels with negative CBF values, which are commonly used in visual evaluation of quality, can provide a realistic automated quality evaluation strategy that can be applied to the final CBF image.

Materials and Methods

MRI Data Acquisition and Processing

Deidentified MRI data from 221 participants obtained from different projects in the local institute and from the Alzheimer's Disease Neuroimaging Initiative (ADNI) were used to develop and validate the automated QEI. All data used in this study were acquired in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and written informed consents were obtained from all the subjects. Local institute studies were approved by the Institutional Review Board (IRB) of the institute. The ADNI study protocols were approved by the IRBs associated with each clinical site. All data were acquired on 3 T Siemens MRI scanners. ADNI data used for this study was acquired in ADNI-2, the third phase of ADNI (after ADNI and ADNI-GO). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to assess whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment, and early Alzheimer's disease (AD). More information can be obtained at <http://adni-info.org>.

The data used to develop the QEI included N = 101 ASL scans obtained using 2D pulsed ASL (PASL) or 2D pseudo continuous ASL (PCASL) and acquired using gradient-echo echoplanar imaging from young and older cognitively normal (CN) participants, patients with Mild Cognitive Impairment, Alzheimer's disease, Parkinson's Disease, and traumatic brain injury, acquired at the local institute and from the ADNI study. The scans were chosen to encompass a wide range of artifacts, but not considering any disease status. The acquisition details are provided in Table 1.

The data used to validate the QEI included ASL MRI data from the ADNI study in N = 53 CN older subjects (72.9 ± 7.4 years, 29 female) with low amyloid beta deposition (amyloid negative or A β –) acquired at baseline and at 3-months

TABLE 1. Details of ASL Data Used to Develop the Automated ASL QEI

Data Source	Diagnosis/Condition	Labeling Method	LT/PLD (Second)	Imaging Resolution (mm ³)	Number of Subjects
ADNI	Cognitively normal older subjects	PASL	0.7/1.9	4.0 × 4.0 × 4.0 (25% distance factor)	10
ADNI	Alzheimer's Disease	PASL	0.7/1.9	4.0 × 4.0 × 4.0 (25% distance factor)	11
Local institute	Cognitively normal older subjects	PCASL	1.5/1.5	3.4 × 3.4 × 5.0 (20% distance factor)	6
Local institute	Mild cognitive impairment	PCASL	1.5/1.5	3.4 × 3.4 × 5.0 (20% distance factor)	20
Local institute	Parkinson's Disease	PCASL	1.5/1.5	3.4 × 3.4 × 5.0 (20% distance factor)	17
Local institute	Traumatic Brain Injury	PCASL	1.5/1.5	3.4 × 3.4 × 5.0 (20% distance factor)	11
Local institute	Middle aged adults	PCASL	1.5/1.5	3.4 × 3.4 × 5.0 (20% distance factor)	26
Total					101

ADNI = Alzheimer's Disease Neuroimaging Initiative; LT = labeling time; PASL = pulsed ASL; PCASL = pseudo continuous ASL; PLD = post-labeling delay.

visit. We also considered the ADNI baseline ASL scans from $N = 27$ CN subjects (age: 79.5 ± 6.0 years, 14 female) who have elevated cerebral amyloid beta deposition levels (amyloid positive or A β +) suggesting preclinical Alzheimer's disease. Amyloid status (A β + or A β -) was determined based on Florbetapir positron emission tomography scans by computing the mean Florbetapir uptake from gray matter in a composite region of interest (ROI) relative to uptake in the whole cerebellum and dichotomized using a cut-off value of 1.11.^{27,28} Finally, we used ASL data from $N = 50$ additional participants acquired with a state-of-art ASL MRI sequence consisting of an unbalanced PCASL labeling with a labeling time of 1.8 seconds and PLD of 1.8 seconds and background suppressed 1-D accelerated 4-shot 3D spiral spin-echo image acquisition with 2.5 mm isotropic resolution.²⁹

CBF maps were generated using standard processing and quantification models as recommended by Alsop et al.¹ Briefly, the processing included motion correction of the control-label time series and application of the single-compartment CBF quantification model.¹ Mean CBF was obtained by discarding outlier volumes from the time series using a structural correlation based outlier rejection technique and a subsequent voxel-wise robust Bayesian based estimation procedure (combined method referred to as SCRUB: Structural Correlation and Robust Bayesian-based estimation).^{4,16} The CBF maps were smoothed using a Gaussian kernel with a full width at half max of 5 mm isotropic.

The development and validation of the QEI used structural information derived from T1-weighted MRI acquired in the same subjects. Structural images were probabilistically segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF)

using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>), and coregistered to the space of the corresponding CBF maps using the boundary-based registration implemented in FSL (<https://fsl.fmrib.ox.ac.uk/>). The probabilistic tissue segmentations assign each voxel a value between 0 and 1, which can be interpreted as the fraction of the specific tissue in that voxel. We also created binarized tissue specific masks by thresholding the tissue probabilistic maps to 0.9, which indicates more than 90 percent tissue in that voxel. These masks were used to obtain mean and variance of the CBF values in GM, WM and CSF that were used in both the QEI development and validation stages. Here "CBF values" indicates all the voxel values within the specific masks in the CBF image while "mean CBF" refers to the average of the CBF values within the mask.

Developing the Automated QEI

This step consists of manual rating of $N = 101$ CBF maps detailed in Table 1, extracting features from the CBF maps, modeling the relationships between the features and the ratings, and combining them into a single metric representative of the quality of the CBF map.

MANUAL RATING. Two neuroradiologists (RW with 23 years of experience and SAN with 12 years of experience in interpreting ASL images) rated the 101 CBF maps shown in Table 1 based on visual inspection. The maps were rated based on the level of image artifacts relative to plausible CBF information on a scale between 1 and 4, corresponding to unacceptable (1), poor (2), average (3), and excellent (4) quality. These manual ratings were used to develop and validate the ASL QEI as detailed below.

PROPOSED ASL QEI. The proposed ASL QEI was designed to be a numerical value between 0 and 1 with a higher number implying a higher quality CBF map. The QEI was designed to quantify features of CBF maps that are typically used during manual QC based on visual inspection. Specifically, the QEI quantifies three factors that characterize valid CBF maps: 1) the similarity between brain structure and CBF (structural similarity, since structure and function are normally correlated), 2) variability of the CBF values across voxels within each tissue (spatial variability, since high variability suggests noisy data and/or insufficiently long post-labeling delay), and 3) the presence of negative CBF values in GM (since GM CBF is expected to be higher than CBF in other regions and less likely to contain negative values due to noise). This definition of the QEI requires segmenting the structural images into GM, WM, and CSF and an accurate coregistration of the structural and the ASL image.

Structural similarity. To model the relationship between CBF values and brain structure, we constructed a structural pseudo-CBF (spCBF) map using a weighted sum of the GM and WM tissue probability maps with weights of 2.5 and 1 for GM and WM respectively since CBF in the GM is on an average about 2.5 times higher than white matter.^{11,30} Although partial volume effects and exchange of labeled spins between brain tissue and CSF compartments³¹ can contribute to very subtle ASL signal changes in CSF, we are ignoring these effects for the purpose of computing spCBF and assuming that the CSF is not perfused. The Pearson's correlation coefficient between spCBF and the actual CBF map for each subject (denoted by ρ_{ss} below where ρ is the correlation coefficient and ss is the structural similarity) was generated as one analytic metric used in the development of the QEI.

Spatial variability. Although CBF varies within tissue types, abnormally large spatial variability can occur because of extremely large or small (often negative) values resulting from motion artifacts (Fig. 1a). Use of a PLD significantly shorter than the time required for the labeled blood to travel to the voxel of interest (arterial transit time) can also lead to large signals in certain regions of the brains

due to retained label in arteries and apparent hypoperfusion in other regions of the brain (Fig. 1b). While such signal variability suggests long arterial transit time and has clinical significance,³² the resulting CBF map does not provide an accurate measure of cortical CBF and can bias regional or voxel-wise statistical analyses. To represent this artifact, we considered the index of dispersion or dispersion index (variance/mean, denoted by DI below) using the pooled variance of the CBF values in GM, WM, and the CSF masks and normalized by the mean GM CBF as one of the features that was used in the QEI metric. The pooled variance (V) was defined as

$$V = \frac{\sum_{k=1}^3 (N_k - 1) V_k}{\sum_{k=1}^3 (N_k - 1)},$$

where V_k is the variance of the CBF values in the k th tissue ($k = \text{GM, WM, CSF}$) and N_k is the number of voxels in that tissue. We did not use the coefficient of variation (CoV), defined as standard deviation/mean, as CoV is scale invariant, and thus CBF scaling will not affect the measure. In contrast, the dispersion index will penalize very large CBF values that can result from incorrect scaling of the data.

Negative GM CBF. As a physiological quantity, CBF should be strictly positive in intact brain parenchyma. However, artifacts in CBF can lead to non-physiological negative CBF values, most commonly in a group of spatially neighboring voxels. Therefore, quantification of voxels with negative CBF can provide an indication of presence of artifacts and hence we considered proportion of voxels in GM with negative CBF (denoted as P_{nGMCBF} below, P is the probability or proportion, and nGMCBF is negative GM CBF) as the final metric that was used in the QEI development. We did not consider voxels in white matter as they have lower CBF and therefore a higher probability of having negative voxels due to thermal or physiological noise.

Figure 2 shows the scatter plots between the three metrics and the average ratings from the two raters, where the ratings were normalized in [0,1] range, to have the resulting QEI in the same range.

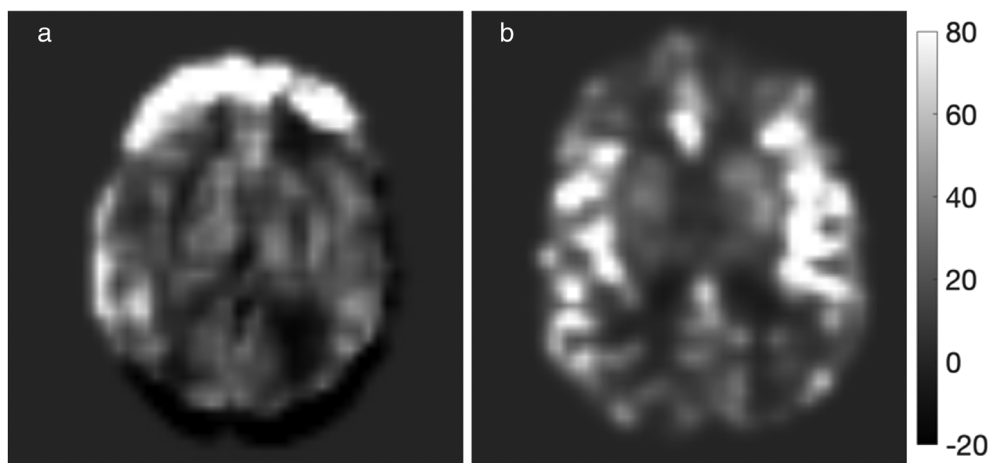


FIGURE 1: Examples of large spatial variability in arterial spin labeling derived cerebral blood flow (a) due to motion or (b) the post-labeling delay (150 msec) being significantly shorter than the arterial transit time resulting in labeled signal retained in the arteries instead of the tissue parenchyma while imaging.

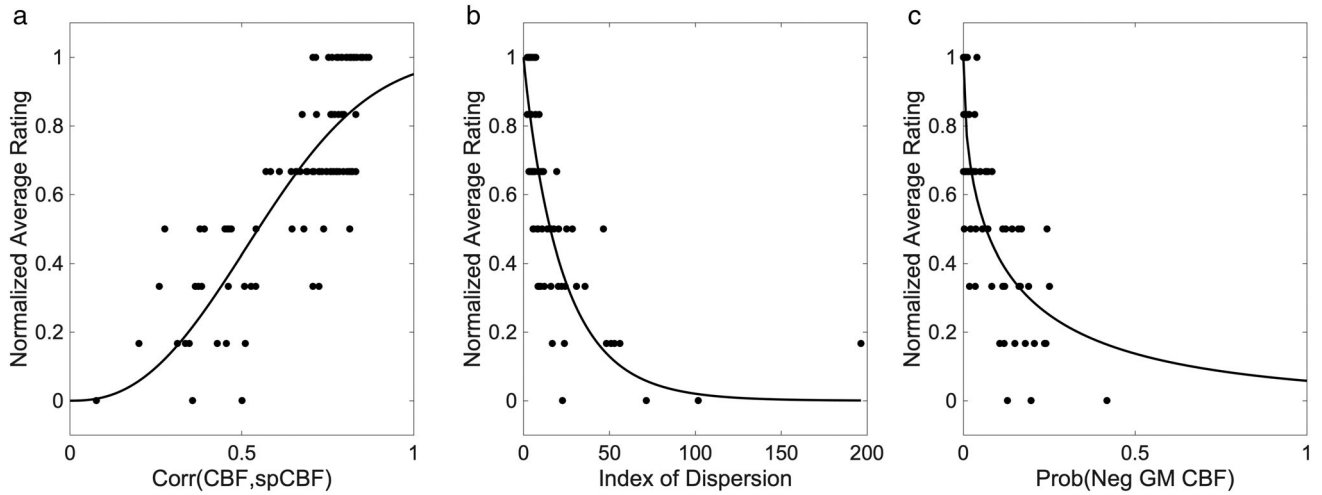


FIGURE 2: Scatter plots of normalized average rating of the two raters with (a) correlation between cerebral blood flow (CBF) and structural pseudo CBF, (b) index of dispersion or dispersion index pooled across the different tissues, and (c) probability of negative CBF values in gray matter. The fits correspond to (a) $1 - \exp(-3.0x^{2.4})$, (b) $\exp(-0.1x^{0.9})$, and (c) $\exp(-2.8x^{0.5})$ respectively, where the values were obtained using a nonlinear least square fit.

The ratings approximately follow a $(1 - \exp(-\alpha x^\beta))$ relationship with structural similarity and an $\exp(-\alpha x^\beta)$ relationship with the other two metrics, where x stands for the various metrics. Consequently, we fitted models of these forms to the data to compute α 's and β 's for each criterion (Fig. 2). The choice of the exponential functions was to limit each component of the QEI in $[0, 1]$ range. The final QEI was computed as the geometric mean of the three metrics. Specifically,

$$QEI = \sqrt[3]{(1 - e^{-3p_{ss}^{2.4}})e^{-(0.1DI^{0.9} + 2.8p_{nGMCBF}^{0.5})}}.$$

Although other averaging process such as arithmetic or harmonic mean could be used, we specifically selected the geometric mean because it is in between the two measures in sensitivity to the lowest of the components and penalizing the automated QEI if one of the metrics is small. Additionally, instead of fitting the ratings with all the features at the same time, we fitted them separately to identify the contributing factor of the low QEI score. A schematic diagram of the proposed QEI is shown in Fig. 3.

Validation of the Proposed QEI

We followed several direct and indirect strategies to validate and demonstrate the usage of the proposed QEI.

LEAVE-ONE-OUT CROSS VALIDATION. We used a leave-one-out cross validation strategy in which we trained the algorithm using all but one sample from the $N = 101$ scans used to develop the QEI (detailed in Table 1), estimated the parameters and computed the QEI for the left-out cases. Thereafter we computed the Pearson's correlation coefficient (R) between these computed QEIs of the left-out cases with the average of the corresponding visual ratings across the folds and compared that to the inter-rater correlation using the method of Steiger³³ (details in Statistical Analysis section). Additionally, we obtained the receiver operating characteristics (ROC) curve for QEI to identify maps judged as average and

excellent by both the raters, computed the area under the ROC curve (AUC) of this classification and obtained a QEI threshold from the curve.

TEST-RETEST RELIABILITY. Here, we considered the baseline and 3-months ASL MRI data from the ADNI study from $N = 53$ A β - cognitively normal (CN) older subjects. CBF values derived from ASL images in CN people acquired 3 months apart would not be expected to change beyond physiological variability unless driven by artifacts in the images. Hence, we hypothesized that the test-retest reliability would be higher if the CBF map in both the acquisitions were of high quality, and it would degrade if at least one of the acquisitions was of poor quality, where the quality was quantified by the QEI. Mean CBF values in 116 ROIs from the Automated Anatomical Labeling (AAL) atlas³⁴ were extracted from ASL data of each subject in the two sessions. For each ROI and each subject, the level of agreement between CBF values was assessed using the CoV. The CoV values for the different ROIs for the specific subject were combined by computing their root mean square to obtain a three-months reliability measure of the two CBF maps of the specific subject. As opposed to computing a CoV based on a single measure of global CBF, this approach provides quantification of reliability across all gray matter regions. The computed CoV for each subject was log transformed because of its right skewed distribution and was analyzed as a function of the poorer QEI between the two sessions using Pearson's correlation coefficient as a summary metric for the relationship. We also divided the data into two groups based on the QEI threshold derived in the previous section and computed the Pearson's correlation coefficient for each group. Note that the AAL atlas provides segmentation of the GM; we did not consider any WM ROI as the ADNI PASL data has a relatively low signal to noise ratio that which is expected to be even lower in the WM.

GROUP DIFFERENCE BETWEEN A β - AND A β + COGNITIVELY NORMAL SUBJECTS. We statistically compared the mean GM CBF of the $N = 53$ A β - with the $N = 27$ A β + CN subjects using 2-sample t test after covarying for age and sex. We

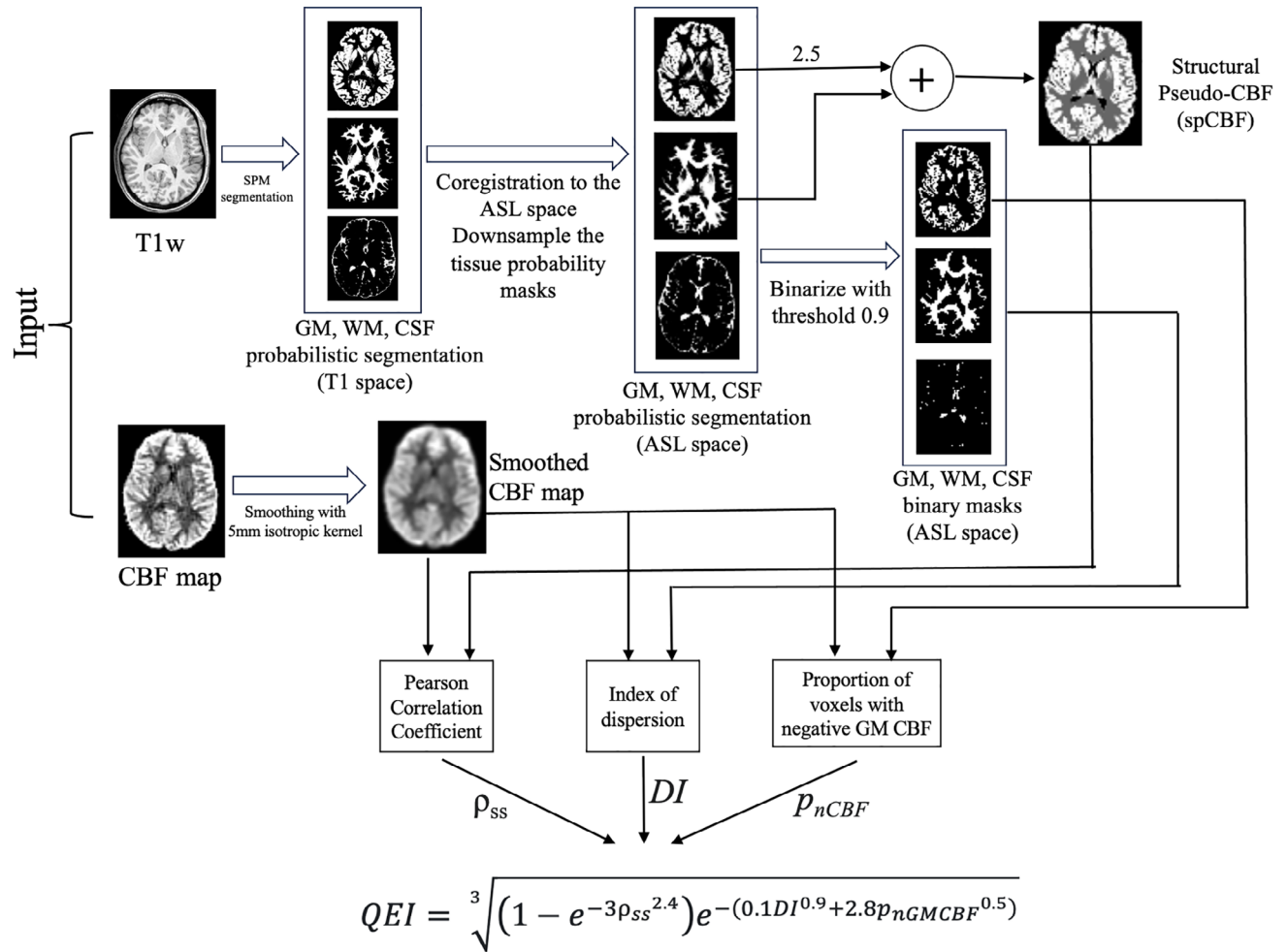


FIGURE 3: A schematic diagram of the proposed quality evaluation index (QEI).

also computed the Cohen's d effect size of the comparison, which was defined as the difference between mean GM CBF averaged across subjects of the two groups divided by the pooled standard deviation of the two groups. Next, we assessed whether discarding data with poor quality could increase effect size. For this, we iteratively discarded CBF maps (one subject in each iteration) based on lowest QEI values and recomputed the effect sizes for GM CBF differences between the two groups. This was made possible since the automated QEI provided a ranking allowing choosing the data with worst quality that was discarded. Pearson's correlation between the iteration number and the effect size was considered as a summary statistic.

ASSESSING THE QEI IN BS 3D ACQUISITIONS. To assess the utility of QEI for data acquired with a state-of-art ASL sequence, the two neuroradiologist also rated $N = 50$ CBF maps of variable quality acquired with the 3D BS PCASL sequence,²⁹ using the same rating scale of 1–4. We computed the Pearson's correlation coefficient between the automated QEI and the average of the two raters and compared that to the inter-rater correlation using the method of Steiger.³³

COMPARISON WITH EXISTING AUTOMATED ASL QUALITY INDEX. Finally, we compared our QEI to the method

included in ASL-MRICloud,²⁶ which is the only other available automated metric that assigns a scalar value to the quality of the CBF image. We referred the reference method as ASL-MRICloud quality index (ASL-MC QI). We computed ASL-MC QI for the BS 3D ASL data and computed the correlation between that and the average of the manual ratings. The ASL-MC QI assigns a value of 1 (excellent) to 4 (poor) to an ASL image. To be consistent with the paper where higher value represents a better CBF map, we modified the definition of the ASL-MC QI by reversing its sign, which resulted in a positive correlation with the manual ratings. We compared this correlation with that obtained between the manual ratings and the proposed method using the method of Steiger.³³

Statistical Analysis

All statistical analyses were performed in MATLAB Version: 9.14.0.2206163 (R2023a), Natick, Massachusetts, USA: The MathWorks Inc.; 2023. The comparisons of the correlation coefficients were performed using the method of Steiger³³ as implemented using the online software developed by Lee and Preacher.³⁵ The method converts each correlation coefficient into a z -score using Fisher's transformation, computes the asymptotic covariance of the estimates and subsequently uses the quantities in an asymptotic z -test. The details of the other statistical analyses have been listed

above under each subsection. $P < 0.05$ was considered statistically significant.

Results

Figure 4 shows representative CBF maps rated as excellent, average, poor, and unacceptable by both raters and the corresponding automated QEIs. Figure 1a shows a representative slice from an ASL CBF map data that was contaminated with motion, resulting in a markedly reduced QEI of 0.3. Figure 1b shows a representative slice from an ASL CBF map that was erroneously acquired with a PLD of 150 msec instead of 1500 msec, resulting in a reduced QEI of 0.25.

Leave-One-Out Cross Validation

In the leave-one-out-cross validation analysis, the correlation between the automated QEI and the mean of the two sets of ratings was 0.83 (CI: 0.76–0.88), which was similar ($P = 0.56$) to the correlation between the two sets of human ratings (0.81, CI: 0.73–0.87). Figure 5a shows the ROC curve corresponding to distinguishing average and excellent CBF maps from the rest (AUC = 0.96). The dotted line in the plot shows the point corresponding to a 99% sensitivity and 79% specificity, and a QEI value of 0.53.

Test–Retest Reliability

Figure 5b shows a scatter plot of the logarithm of subject specific CoV, representing reliability of the two ASL scans from individual subjects acquired 3-months apart, vs. the poorer QEI of the two scans. There was an overall negative correlation between the quantities with a correlation coefficient of -0.74 (CI: -0.84 to -0.59). When the data was divided based on the QEI threshold of 0.53 derived from the ROC curve, the correlation was higher with a Pearson's correlation coefficient of $R = -0.79$ (CI: -0.90 to -0.61) for QEI < 0.53 , while the correlation was both attenuated and no longer statistically significant for QEI ≥ 0.53 ($R = -0.34$, $P = 0.13$).

Group Difference Between $A\beta^-$ and $A\beta^+$ Cognitively Normal Subjects

Figure 6a shows the histogram of the QEI values for each of the two groups of 27 $A\beta^+$ and 53 $A\beta^-$ ADNI participants expressed as a percentage in each group. Considering all participants, there was no significant difference in GM CBF between the two groups with a small effect size of 0.18; here a positive effect size implies higher CBF in the $A\beta^+$ group. Figure 6b shows the results of the group comparisons at each

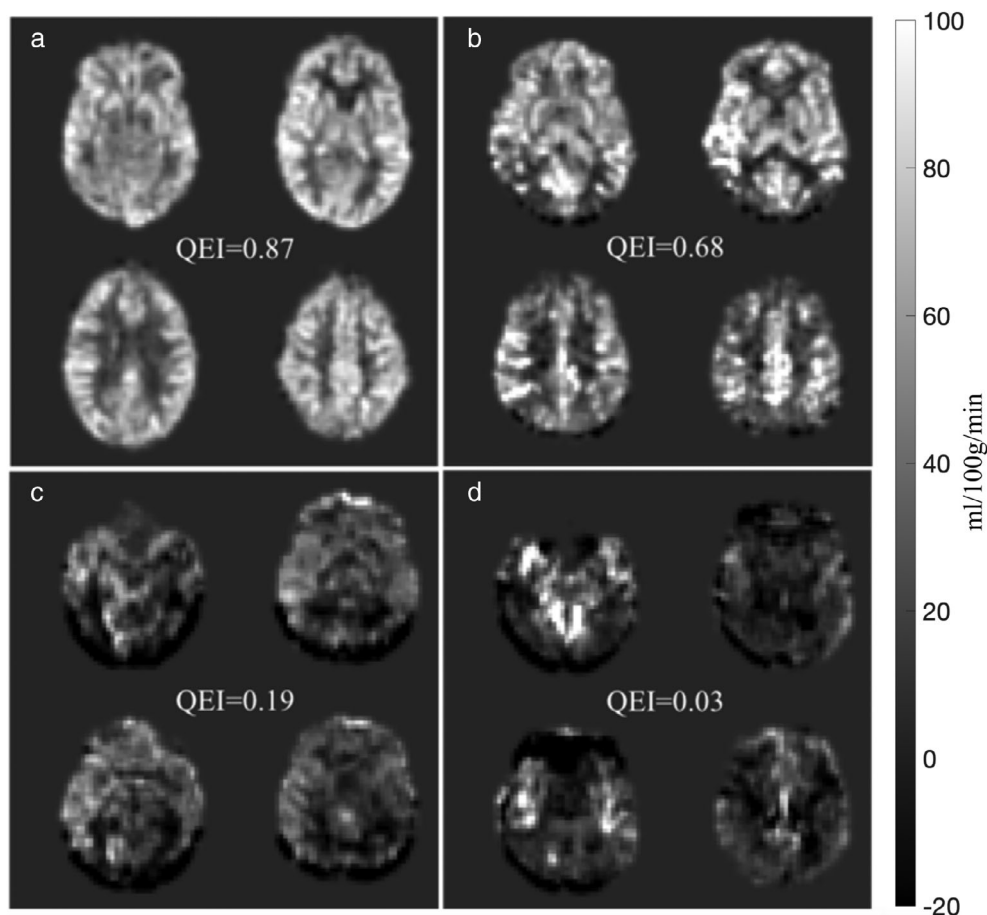


FIGURE 4: Examples of (a) excellent (QEI = 0.87), (b) average (QEI = 0.68), (c) poor (QEI = 0.19), and (d) unacceptable (QEI = 0.03) cerebral blood flow maps obtained with non-background suppressed 2D ASL data, where the two raters provided the same ratings.

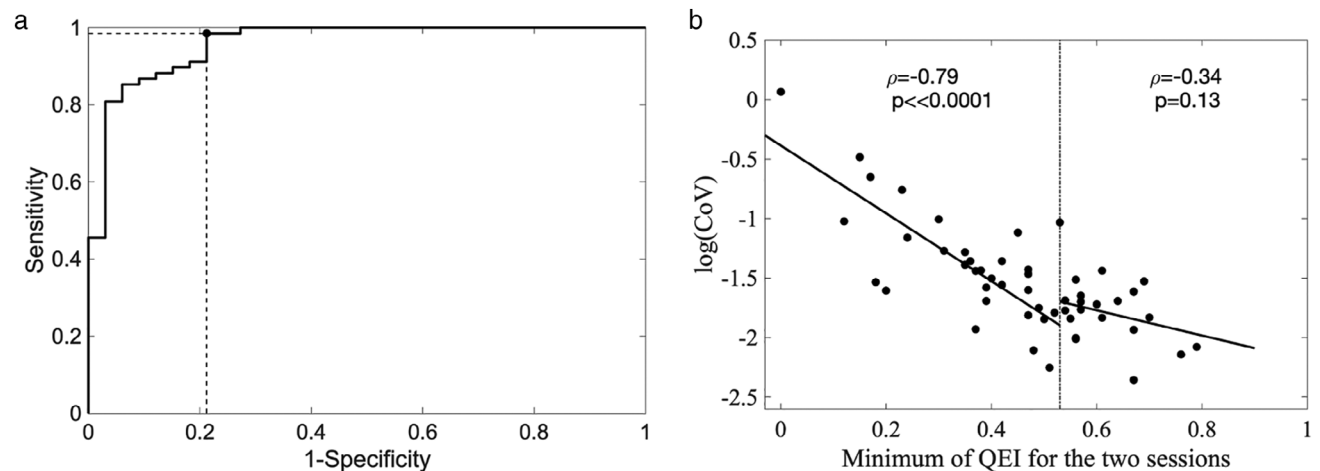


FIGURE 5: (a) Receiver operating characteristic (ROC) curve corresponding to distinguishing Average and Excellent CBF maps from the rest as judged by both the raters. The intersection of the dotted lines shows the operating point and corresponds to a sensitivity of 0.99, specificity of 0.79 and QEI = 0.53. (b) Scatter plot of minimum of QEI for the two sessions (worse of the two CBF maps) vs. logarithm of Coefficient of Variation (CoV). Correlation coefficient between QEI and $\log(\text{CoV})$ was -0.74 ($P < 0.0001$). The plot also shows two fitted regression lines in subsets of data divided based on a QEI threshold of 0.53.

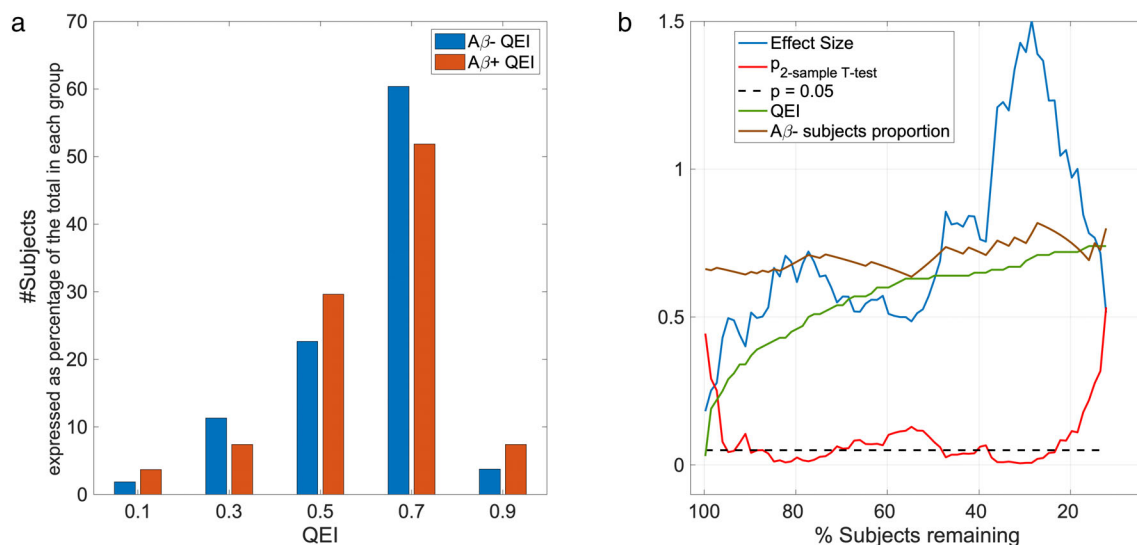


FIGURE 6: (a) Histogram of the QEI values of the amyloid negative and positive participants from ADNI expressed as a percentage of the total in each group. (b) Effect size and P value (along with $P = 0.05$) corresponding to two sample t test in differentiating the two groups as data of poor quality are iteratively discarded. The plot also shows the QEI value of the excluded participants and the proportion of amyloid negative participants in the analysis.

iteration when the participants with the lowest QEI were iteratively excluded. The effect size of the group comparisons is shown in blue and there was an overall trend of an increase in effect size when data of poor quality was iteratively excluded until the point where loss of sample size begins to dominate. The Pearson's correlation coefficient between the iteration number and the effect size was 0.72 (CI: 0.59–0.82). The QEI of the discarded subjects at each iteration is shown in green and the proportion of retained $A\beta^-$ subjects is shown in brown. The uncorrected P values of the comparisons for each step of discarding subjects are shown in red. The group difference, as represented by the P value was not statistically significant with all subjects but became significant

when the data of worst quality were discarded; it remained or tended toward statistical significance until the point when the sample size decreased by more than 75%. The proportion of the amyloid negative subjects remained consistent (range: 0.64–0.82) as participants were excluded based on lowest QEI.

Assessing the QEI in BS 3D Acquisitions

The correlation of the ratings of the two raters for the $N = 50$ BS 3D ASL scans was 0.78 (CI: 0.64–0.87). In comparison, the correlation between the automated QEI with the average of the two sets of ratings was 0.86 (CI: 0.77–0.92), which was statistically similar ($P = 0.09$) to the inter-rater

correlation. Figure 7 shows examples of automated QEI values for different rating categories where the two raters provided the same ratings.

Comparison with Existing Automated ASL Quality Index

We discarded the data of two subjects, one for having a single control-label pair for which the computation of ASL-MC QI was not possible, and the other for having overall negative mean CBF due to extensive artifacts leading to out of bound ASL-MC QI. For the remaining 48 subjects, the correlation between ASL-MC QI and the average human rating was 0.54 (CI: 0.30–0.72) while that between the QEI and the average human rating was 0.87 (CI: 0.77–0.92), the latter being significantly higher than that obtained with ASL-MC QI.

Discussion

This work proposes an automated QEI for ASL-derived CBF maps using a simple algorithm reflecting basic features of high-quality CBF maps and trained using manual ratings from neuroradiologists. The QEI showed excellent agreement with observer ratings with a correlation similar to inter-rater agreement. Using an indirect validation method, CBF maps

with higher quality as determined by the automated QEI showed better test–retest reliability with ADNI data acquired 3 months apart. We also showed the usefulness of the metric in discarding data from statistical analysis to compare preclinical Alzheimer's subjects with older CN participants. Finally, we demonstrated the applicability of the QEI to a state-of-the-art BS 3D data and its superiority over a reference method. Data quality is increasingly recognized as one of the most important confounders in brain imaging research,³⁶ and the proposed QEI, being an objective measure, is expected to improve the rigor and reproducibility of ASL MRI research.

Manual QC is common in studies involving ASL but is often subjective. Raters can differ in their assessments of acceptable artifacts, which might also depend on the method of visualization. As expected, the correlation between the ratings of the two raters was not perfect. Although not explicitly tested as a part of this study, the intra-rater agreement is also not expected to be perfect. Additionally, the agreement can be lower with raters new to the field who have limited experience with ASL CBF maps. The automated rating, being an objective measure, has the advantage of perfect reproducibility thus increasing scientific rigor, though there can potentially be minor differences because of differences in software

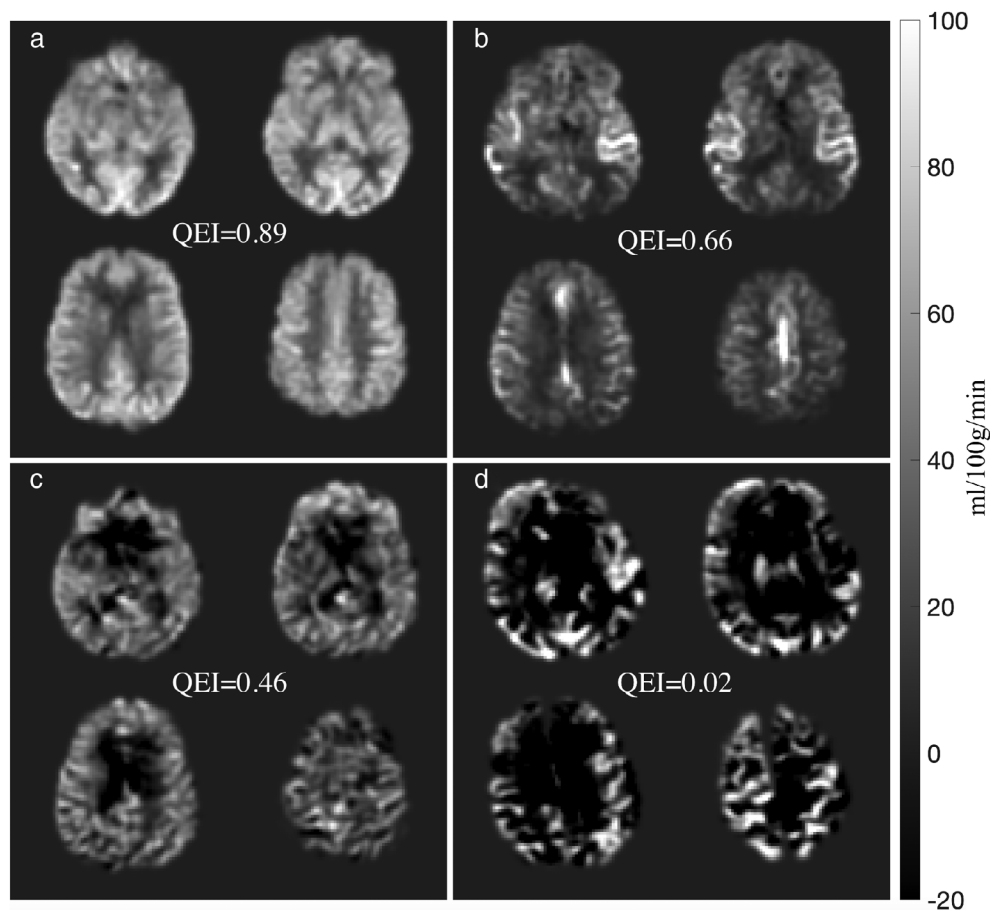


FIGURE 7: Examples of (a) excellent (QEI = 0.89), (b) average (QEI = 0.66), (c) poor (QEI = 0.46), and (d) unacceptable (QEI = 0.02) CBF maps obtained with background suppressed 3D ASL data, where the two raters provided the same ratings.

versions across users. The QEI can also be useful for detecting errors in registration of structural and ASL CBF images in automated image processing pipelines as such misregistration is expected to assign low QEI to the CBF maps because of the dependence of the QEI on structural similarity and the use of masks derived from structural images to derive mean CBF in different tissue types. This is important as such misregistration will result in incorrect extraction of CBF values from different ROIs that will affect subsequent statistical analyses. The QEI can also potentially identify systematic errors in acquisition, such as motion artifacts or incorrect sequence parameters in specific sites of a multi-site study. Systematic low QEI in specific sites can help guide protocol modifications that can include use of additional measures to restrict head motion or modification of imaging parameters, which can preserve data for analysis.

We performed indirect validation of the QEI by assessing if the test–retest reliability from data acquired 3-months apart decreases with the presence of image artifacts in at least one of the scans. The negative correlation between the automated QEI and the CoV, which quantified test–retest reliability, shows the efficacy of the QEI in detecting artifacts. When the data was divided based on a QEI threshold, we observed that the correlation was more pronounced for lower QEI compared to higher QEI data. This shows that the reliability was worse when one of the scans was considerably poor, while the improvement in reliability was subtle when the data achieved a certain quality.

The usefulness of the automated QEI as a quality control measure was shown by comparing the mean CBF in GM between subjects with preclinical AD and cognitively normal participants without evidence of early AD pathologic change from the ADNI study. Prior studies showed hyperperfusion and hypermetabolism^{37–39} in A β + CN subjects compared to A β – controls and we aimed to replicate the results in the ADNI cohort. The automated QEI provided a ranking of the quality that would otherwise have required a manual quality control. Although, there was no significant difference between the two groups with a weak effect size when data from all the participants were considered, there was a trend of increase in effect size with higher CBF in preclinical AD when data of poor quality were iteratively excluded. Although the successive tests were not corrected for multiple comparisons, the findings illustrate the importance of excluding poor data from statistical analyses. Most experimental designs rely on increasing sample sizes to increase sensitivity for group effects; however, we showed here that discarding data of poor quality can increase effect size in statistical comparisons. Using the approach in other ASL datasets may improve sensitivity for testing hypotheses about group differences. Although the results showed that the analysis benefitted by discarding a large proportion of ADNI PASL data which has relatively low signal to noise

ratio, a much lower percentage of ASL data gets excluded from any analysis in general.^{3,6,40}

The QEI was trained on non-BS 2D data which is expected to have higher level of artifacts compared to background-suppressed ASL data. However, we also demonstrated its applicability to BS 3D data in a small sample where the agreement between the QEI with manual rating was similar to inter-rater agreement. In addition, the QEI has been applied to studies with 3D background suppressed ASL data acquired both in the local institute and in multisite studies such as the Multi Ethnic Study of Atherosclerosis (MESA) and the Epidemiology of Diabetes Interventions and Complications (EDIC) study (data processing, QC and analysis in progress) to identify corrupted data and the agreement of the CBF maps with the automated QEIs have been manually verified. Nevertheless, future work is required to identify artifacts specific to 3D acquisition.

When comparing the proposed QEI with the automated ASL quality index included in ASL-MRICloud, we obtained significant improvement in agreement with manual ratings using our method. This is expected since the reference method only considers the temporal variability of the data while our method considers more intrinsic characteristics of a CBF map.

Limitations

First, since simple smoothing of the CBF maps results in reduced noise and image artifacts that improves values of the metrics used in the QEI, the proposed QEI is sensitive to smoothness of the data with higher smoothness receiving a higher QEI. Therefore, the QEI does not necessarily assign higher scores to CBF maps with greater spatial resolution. So, although the QEI can rate the quality of any CBF map, it cannot necessarily be used to directly compare data quality across protocols and analytic pipelines. Matching of imaging protocol and processing, particularly the CBF map smoothness, should result in more comparable QEI values across studies. Currently our recommendation is to smooth the CBF maps by a 5 mm isotropic Gaussian kernel before computing the QEI as that was used to derive the QEI parameters and the cut-off value. Although the QEI distribution and cut-off can vary based on the acquisition/smoothing of the dataset, a cut-off value of 0.5 has worked reliably for a wide variety ASL protocols in multiple studies. Note that a lower cut-off value can increase the sample size and hence can be useful for data of poor quality to attain a certain sample size. On the other hand, a higher cut-off can obtain a finer tuning for CBF maps of better quality. Second, the definition of QEI requires the availability of a high-resolution structural image. Although this is common in research settings that are the major application of the QEI, this constitutes a limitation of the technique. Third, the QEI value relies on the quality of the coregistration between the structural and the ASL

images and hence can assign a lower QEI to good CBF map with poor registration with the structural image. However, as discussed earlier, since the automated analysis pipelines typically also require accurate coregistration between the structural and the ASL images, this is a potential benefit as assignment of lower QEI and subsequent flagging of the CBF maps will help in identifying and correcting the error. Fourth, the QEI can potentially result in low values in the presence of large structural lesions. However, we do not expect a drastic reduction of the QEI in this situation as the artifactual factors that the QEI penalizes generally results in larger intensity changes than that resulting from structural lesions. Fifth, the QEI provides a global quality measure of the CBF map; there can be local artifacts that can still bias statistical analyses if mean CBF in the artifactual region is considered in the analysis and the QEI is above the chosen cut-off value for discarding CBF maps. Sixth, the QEI was developed based on a relatively small sample size of training data and manual ratings. Finally, the QEI reported here was developed based on training with data from Siemens scanners and a limited range of imaging protocols. In future work, we aim to extend this method by incorporating more extensive types of data from multiple scanner platforms.

Conclusions

We designed, optimized, and validated an automated QEI for ASL MRI that is sensitive to typical artifacts in CBF maps and performs comparably to manual quality assessments, providing an objective means of rapidly assessing quality of CBF maps.

Acknowledgments

This study was supported by National Institutes of Health Grants P30AG072979, P41EB015893, R01AG040271, R01MH080729, R01NS111115, R03AG063213, R21AG080518. Some of the data used for this study were collected and shared by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research &

Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

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):102-116.
2. Detre JA, Leigh JS, Williams DS, Koretsky AP. Perfusion imaging. *Magn Reson Med* 1992;23(1):37-45.
3. Dolui S, Wang Z, Wang DJJ, et al. Comparison of non-invasive MRI measurements of cerebral blood flow in a large multisite cohort. *J Cereb Blood Flow Metab* 2016;36(7):1244-1256.
4. Dolui S, Wang Z, Shinohara RT, Wolk DA, Detre JA, Alzheimer's Disease Neuroimaging Initiative. Structural Correlation-based Outlier Rejection (SCORE) algorithm for arterial spin labeling time series. *J Magn Reson Imaging* 2017;45(6):1786-1797.
5. MacIntosh BJ, Shirzadi Z, Atwi S, et al. Metabolic and vascular risk factors are associated with reduced cerebral blood flow and poorer mid-life memory performance. *Hum Brain Mapp* 2020;41(4):855-864.
6. Dolui S, Detre JA, Gaussoin SA, et al. Association of intensive vs standard blood pressure control with cerebral blood flow: Secondary analysis of the SPRINT MIND randomized clinical trial. *JAMA Neurol* 2022;79(4):380-389.
7. Ewing JR, Cao Y, Knight RA, Fenstermacher JD. Arterial spin labeling: Validity testing and comparison studies. *J Magn Reson Imaging* 2005;22(6):737-740.
8. Heijtel DF, Mutsaerts HJ, Bakker E, et al. Accuracy and precision of pseudo-continuous arterial spin labeling perfusion during baseline and hypercapnia: A head-to-head comparison with $(1)(5)\text{O}$ $\text{H}(2)\text{O}$ positron emission tomography. *Neuroimage* 2014;92:182-192.
9. Ye FQ, Berman KF, Ellmore T, et al. $\text{H}(2)(15)\text{O}$ PET validation of steady-state arterial spin tagging cerebral blood flow measurements in humans. *Magn Reson Med* 2000;44(3):450-456.
10. Dolui S, Fan AP, Zhao MY, Nasrallah IM, Zaharchuk G, Detre JA. Reliability of arterial spin labeling derived cerebral blood flow in periventricular white matter. *Neuroimage Rep* 2021;1(4):100063.
11. Dolui S, Vidorreta M, Wang Z, et al. Comparison of PASL, PCASL, and background-suppressed 3D PCASL in mild cognitive impairment. *Hum Brain Mapp* 2017;38(10):5260-5273.
12. Ware JB, Dolui S, Duda J, et al. Relationship of cerebral blood flow to cognitive function and recovery in early chronic traumatic brain injury. *J Neurotrauma* 2020;37(20):2180-2187.
13. Moonen JEF, Nasrallah IM, Detre JA, et al. Race, sex, and mid-life changes in brain health: Cardia MRI substudy. *Alzheimers Dement* 2022;18(12):2428-2437.
14. Lindner T, Bolar DS, Achten E, et al. Current state and guidance on arterial spin labeling perfusion MRI in clinical neuroimaging. *Magn Reson Med* 2023;89(5):2024-2047.

15. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magn Reson Med* 1998;40(3):383-396.
16. Dolui S, Wolk DA, Detre JA. SCRUB: A structural correlation and empirical robust Bayesian method for ASL data. *Proceedings of the International Society of Magnetic Resonance in Medicine*. Singapore: ISMRM; 2016.
17. Li Y, Dolui S, Xie DF, Wang Z. Priors-guided slice-wise adaptive outlier cleaning for arterial spin labeling perfusion MRI. *J Neurosci Methods* 2018;307:248-253.
18. Shirzadi Z, Stefanovic B, Chappell MA, et al. Enhancement of automated blood flow estimates (ENABLE) from arterial spin-labeled MRI. *J Magn Reson Imaging* 2018;47(3):647-655.
19. Wang Z, Aguirre GK, Rao H, et al. Empirical optimization of ASL data analysis using an ASL data processing toolbox: ASLtbx. *Magn Reson Imaging* 2008;26(2):261-269.
20. Wang Z. Improving cerebral blood flow quantification for arterial spin labeled perfusion MRI by removing residual motion artifacts and global signal fluctuations. *Magn Reson Imaging* 2012;30(10):1409-1415.
21. Fernandez-Seara MA, Wang Z, Wang J, et al. Continuous arterial spin labeling perfusion measurements using single shot 3D GRASE at 3 T. *Magn Reson Med* 2005;54(5):1241-1247.
22. Maleki N, Dai W, Alsop DC. Optimization of background suppression for arterial spin labeling perfusion imaging. *Magn Reson Mater Phys Biol Med* 2012;25(2):127-133.
23. Ye FQ, Frank JA, Weinberger DR, McLaughlin AC. Noise reduction in 3D perfusion imaging by attenuating the static signal in arterial spin tagging (ASSIST). *Magn Reson Med* 2000;44(1):92-100.
24. Dolui S, Tisdall D, Vidorreta M, et al. Characterizing a perfusion-based periventricular small vessel region of interest. *NeuroImage Clin* 2019; 23:101897.
25. Wang Z, Das SR, Xie SX, et al. Arterial spin labeled MRI in prodromal Alzheimer's disease: A multi-site study. *NeuroImage Clin* 2013;2: 630-636.
26. Li Y, Liu P, Li Y, et al. ASL-MRICloud: An online tool for the processing of ASL MRI data. *NMR Biomed* 2019;32(2):e4051.
27. Tosun D, Veitch D, Aisen P, et al. Detection of beta-amyloid positivity in Alzheimer's disease neuroimaging initiative participants with demographics, cognition, MRI and plasma biomarkers. *Brain Commun* 2021; 3(2):fcab008.
28. Landau SM, Breault C, Joshi AD, et al. Amyloid-beta imaging with Pittsburgh compound B and florbetapir: Comparing radiotracers and quantification methods. *J Nucl Med* 2013;54(1):70-77.
29. Vidorreta M, Wang Z, Chang YV, Wolk DA, Fernandez-Seara MA, Detre JA. Whole-brain background-suppressed pCASL MRI with 1D-accelerated 3D RARE stack-of-spirals readout. *PloS One* 2017;12(8): e0183762.
30. Johnson NA, Jahng GH, Weiner MW, et al. Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: Initial experience. *Radiology* 2005;234(3):851-859.
31. Petitclerc L, Schmid S, Hirschler L, van Osch MJP. Combining T(2) measurements and crusher gradients into a single ASL sequence for comparison of the measurement of water transport across the blood-brain barrier. *Magn Reson Med* 2021;85(5):2649-2660.
32. Mutsaerts HJ, Petr J, Vaclav L, et al. The spatial coefficient of variation in arterial spin labeling cerebral blood flow images. *J Cereb Blood Flow Metab* 2017;37(9):3184-3192.
33. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980;87(2):245-251.
34. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 2002; 15(1):273-289.
35. Lee IA, Preacher KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. <http://quantpsy.org> 2013.
36. Rosen AFG, Roalf DR, Ruparel K, et al. Quantitative assessment of structural image quality. *Neuroimage* 2018;169:407-418.
37. Fazlollahi A, Calamante F, Liang X, et al. Increased cerebral blood flow with increased amyloid burden in the preclinical phase of alzheimer's disease. *J Magn Reson Imaging* 2020;51(2):505-513.
38. Oh H, Habeck C, Madison C, Jagust W. Covarying alterations in Abeta deposition, glucose metabolism, and gray matter volume in cognitively normal elderly. *Hum Brain Mapp* 2014;35(1):297-308.
39. Johnson SC, Christian BT, Okonkwo OC, et al. Amyloid burden and neural function in people at risk for Alzheimer's disease. *Neurobiol Aging* 2014;35(3):576-584.
40. Dolui S, Li Z, Nasrallah IM, Detre JA, Wolk DA. Arterial spin labeling versus 18F-FDG-PET to identify mild cognitive impairment. *NeuroImage Clin* 2020;25:102146.