NeuroImage 193 (2019) 157–166

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Test-retest reliability of cerebral blood flow for assessing brain function at rest and during a vigilance task

Fan Nils Yang [a,c], Sihua Xu [b,c], Andrea Spaeth [c,d], Olga Galli [d,e], Ke Zhao [c], Zhuo Fang [b,c], Mathias Basner [d], David F. Dinges [d], John A. Detre [c], Hengyi Rao [b,c,d,*]

[a] Department of Psychology, Sun Yat-sen University, Guangzhou, China
[b] Laboratory of Applied Brain and Cognitive Sciences, Shanghai International Studies University, Shanghai, China
[c] Center for Functional Neuroimaging, Department of Neurology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[d] Division of Sleep and Chronobiology, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[e] Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
ASL perfusion fMRI
Cerebral blood flow (CBF)
Reliability
Intraclass correlation coefficients (ICC)
Psychomotor vigilance test (PVT)

ABSTRACT

Arterial spin labeled (ASL) perfusion magnetic resonance imaging (MRI) is increasingly used to assess regional brain activity and cerebrovascular function in both healthy and clinical populations. ASL perfusion imaging provides a quantitative measure of regional brain activity by determining absolute cerebral blood flow (CBF) values at a resting state or during task performance. However, the comparative reliability of these ASL measures is not well characterized. It is also unclear whether the test-retest reliability of absolute CBF or task-induced CBF change measures would be comparable to the reliability of task performance. In this study, fifteen healthy participants were scanned three times in a strictly controlled in-laboratory study while at rest and during performing a simple and reliable psychomotor vigilance test (PVT). The reliability of absolute CBF and task-induced CBF changes was evaluated using the intraclass correlation coefficient (ICC) and compared to that of task performance. Absolute CBF showed excellent test-retest reliability across the three scans for both resting and PVT scans. The reliability of regional absolute CBF was comparable to that of behavioral measures of PVT performance, and was slightly higher during PVT scans as compared with resting scans. Task-induced regional CBF changes demonstrated only poor to moderate reliability across three scans. These findings suggest that absolute CBF measures are more reliable than task-induced CBF changes for characterizing regional brain function, especially for longitudinal and clinical studies.

1. Introduction

Functional magnetic resonance imaging (fMRI) offers a noninvasive approach to localize neural responses to various stimuli and tasks in the human brain, and has become the dominant imaging technique in clinical and cognitive neuroscience research. Many fMRI studies used the blood oxygenation level dependent (BOLD) contrast, which reflects a complex interaction among changes in cerebral blood flow (CBF), changes in cerebral blood volume (CBV), and changes in cerebral metabolic rate of oxygenation (Buxton et al., 2004; Detre and Wang, 2002). BOLD signal is also affected by a range of biophysical factors unrelated to brain physiology. Due to its indirect and multifactorial nature, the BOLD signal is primarily used in a qualitative manner, typically by providing information about which area in the brain is activated based on a group-level analysis of imaging data, which has limited usage in clinical studies (Pike, 2012). As more quantitative measurements are needed in basic and clinical science, some strategies have been developed to relate the BOLD signal to brain physiology in a quantitative manner (Pike, 2012; Shu et al., 2010).

In addition to quantitative BOLD fMRI, regional CBF has been increasingly used as a surrogate marker for brain function in both healthy and clinical populations (Detre et al., 2009), due to its tight coupling with regional brain activity, perfusion, and metabolism (Raichle, 1998). Using magnetically labeled blood water as an endogenous perfusion tracer,

* Corresponding author. Center for Functional Neuroimaging & Department of Neurology, University of Pennsylvania Perelman School of Medicine, Room D502, 3400 Spruce St, Philadelphia, PA, 19104, USA.
E-mail address: hengyi@pennmedicine.upenn.edu (H. Rao).

https://doi.org/10.1016/j.neuroimage.2019.03.016
Received 20 November 2018; Received in revised form 6 March 2019; Accepted 7 March 2019
Available online 17 March 2019
1053-8119/

arterial spin labeled (ASL) perfusion MRI can noninvasively quantify cerebral blood flow (in mL/100 g tissue per minute) both at rest and during the performance of cognitive or sensorimotor tasks. The close coupling between regional CBF and brain activity makes ASL particularly useful in longitudinal and clinical studies of brain function as it provides a direct measure of alterations in regional CBF in disorders of perfusion (Wolf and Detre, 2007). For instance, ASL perfusion imaging has been used for evaluating cerebrovascular disease, detecting and tracking disease progression, and measuring pharmacological effects and therapeutic responses for clinical trials (e.g. Alsop et al., 2015; J. J. Chen et al., 2011; Y. Chen et al., 2011; Franklin et al., 2012, 2011; Kim et al., 2012; Telischak et al., 2015).

Given the high potential of ASL in translational and clinical research, especially for longitudinal studies, its test-retest reliability needs to be carefully addressed. ASL provides absolute CBF measurements during rest or task performance, as well as task-induced relative CBF changes in different brain regions. In terms of resting-state CBF, high test-rest reliability has been well-demonstrated, with intra-class correlation coefficients (ICC) values typically greater than 0.60, falling in the good to excellent range (Y. Chen et al., 2011; Fazlollahi et al., 2015; Hodkinson et al., 2013; Jahng et al., 2005; Jiang et al., 2010; Klomp et al., 2012; Li et al., 2018; Mezue et al., 2014; Steketee et al., 2015; Tancredi et al., 2015; Wang et al., 2011; Wu et al., 2014; Xu et al., 2010; Zou et al., 2015). Test-rest reliability of resting-state CBF remains high across long test-retest intervals (Jiang et al., 2010; Mezue et al., 2014), and with data collected from different MRI centers (Gevers et al., 2011). However, the test-retest reliability of task-induced CBF changes remains controversial. For example, Steketee et al. (2015) found comparable reliability for absolute CBF at rest as well as CBF during task but much lower reliability of task-induced CBF changes in the primary motor area during a finger tapping task (ICC = 0.04–0.32). In contrast, Raoult et al. (2011) showed good to excellent reliability (ICC > 0.70) of proportional task-induced brain activity in the left primary motor cortex regardless of ASL or BOLD imaging protocol. This discrepancy may be due to different metrics of task-induced CBF changes. The present study used both metrics to calculate the task-induced CBF changes and compared their reliability.

Task-induced CBF changes have been increasingly used for gauging task-induced brain activation, and considered as an alternative to BOLD activation. However, the reliability of task-induced relative CBF measurements has not been well-established. Previous studies on BOLD fMRI reliability have suggested that the test-retest reliability of task-induced BOLD activation may be less stable when compared with task performance (Plichta et al., 2012; Upadhyay et al., 2015) and vary significantly with different statistical thresholds (Stevens et al., 2013). Moreover, few studies have examined the whole brain voxel-wise reliability, although it may be one of the most valuable metrics regarding fMRI reliability (Bennett and Miller, 2010). In addition, although the ability of ASL perfusion imaging to quantify CBF in the white matter (WM) is still controversial (Van Gelderen et al., 2008), recent evidence suggests that fMRI may be able to detect activation in the WM (Gawryluk et al., 2014; Mandl et al., 2008; Yarkoni et al., 2009) and CBF in the WM may be accurately quantified using ASL in a more time-efficient manner (Zhang et al., 2016). In this study, we evaluated the voxel-wise and region-of-interest (ROI) based reliability of absolute CBF measurements and task-induced CBF changes in the gray matter (GM), white matter, and task-related ROIs. CBF data both at resting-state and during performing a simple and reliable psychomotor vigilance test (PVT) (Drummond et al., 2005; Evranos et al., 2016; Lim and Dinges, 2008) were collected across three scanning sessions within 5 consecutive days. Reliability among task-based CBF, resting-state CBF and task-induced CBF changes, were compared quantitatively at voxel, regional, and network levels. In addition, we investigated whether the test-retest reliability of these absolute CBF measures or task-induced relative CBF change would be comparable to the reliability of task performance.

2. Methods

2.1. Participants

We analyzed data from fifteen healthy right-handed adults (6 females, mean age = 34.68 ± 9.03 yrs, years of education = 14.87 ± 2.25, BMI: 25.01 ± 4.15) who were recruited as control subjects for a strictly controlled in-laboratory sleep deprivation study (for more details, see Fang et al., 2015; Yang et al., 2018). Individuals reported normal sleep timing and duration and no habitual napping or sleep disturbances, assessed by questionnaire (Smith et al., 1989). Participants were screened for acute or chronic medical and psychological conditions, as well as drug and alcohol use by using questionnaires, physical examinations, and blood and urine tests. Use of caffeine, alcohol, tobacco and medications (except oral contraceptives) was not permitted in the week before and during the laboratory study, and compliance was verified by urine screenings.

The study was approved by the Institutional Review Board of the University of Pennsylvania. All participants provided written informed consent before enrollment, which was in accordance with the Declaration of Helsinki. Participants were compensated for participating in the study.

2.2. Experimental design

Participants were screened on two occasions before joining the study and participated in a mock MRI scanner training during the second interview session. During the study, participants remained in the laboratory at the Clinical Translational Research Center of Hospital of the University of Pennsylvania for 5 consecutive days (4 nights). On day 1, participants received about 9 h time-in-bed (2130 h-0630 h) to adjust to the unfamiliar sleep environment in the laboratory. On days 2–5, participants were permitted 8 h of sleep opportunity (2230 h - 0630 h). Individuals were continuously monitored by research and hospital staff to ensure adherence to the protocol. Each participant completed an MRI scan section three times at the same time-of-day (between 0700 h and 1000 h) during the morning of days 2, 3, and 5. During each scan section, each participant completed a 10-min PVT during the ASL scan to acquire the task-based CBF data. Two additional ASL scans of 4-min resting-state CBF were also acquired prior to (REST1) and following (REST2) the PVT scan, respectively. During the resting-state scan, participants were instructed to keep eyes open and look at a cross fixation displayed on the screen. During the PVT scan, subjects were asked to perform a sustained vigilant attention task (see below).

2.3. Psychomotor vigilance test (PVT)

The PVT is a reaction-time-based measure of sustained attention that has long been used as a measure of vigilant attention (Drummond et al., 2005; Evranos et al., 2016; Lim and Dinges, 2008). Participants are instructed to press a button as fast as they can in response to a yellow millisecond counter inside a red rectangular area in the center of a dark screen, which is presented at a uniform random inter-stimulus interval (2–10s) for a total of 10 min (Lim and Dinges, 2008). Upon a response, the millisecond counter stops and remains on the screen for another second to allow participants to view their reaction time (RT). Button presses when the millisecond counter is not presented are counted as false alarms. Reaction times exceeding 500 ms are counted as lapses of attention. The following indexes were calculated as a measure of overall level of PVT performance: mean RT, median RT, standard deviation (SD) of RT, and mean of the fastest 10% of RT (FRT) (Gui et al., 2015; Lim and Dinges, 2010, 2008).

2.4. Imaging data acquisition and analyses

Imaging data were collected using a Siemens 3.0 T Trio whole-body scanner (Siemens AG, Erlangen, Germany). Structural T1-weighted 3D

magnetization-prepared rapid gradient-echo (MPRAGE) images were obtained using a sequence with the following parameters: TR = 1.62 s, TE = 3.09 ms, FOV = 187 * 250 mm^2, matrix size = 192 * 256, slice thickness = 5 mm, inter-slice gap = 1 mm. ASL perfusion images were acquired by using a pseudo-continuous ASL (pCASL) sequence with a 2D gradient echo planar imaging (EPI) readout with the following parameters: TR = 4 s, TE = 18 ms, FOV = 220 * 220 mm^2, matrix size = 64 * 64, number of slices = 20, slice thickness = 5 mm, inter-slice gap = 1 mm, labeling time = 1.5s, post-labeling delay time = 1.0 s.

The quantitative CBF measurements were preprocessed using fMRI Grocer toolbox (https://www.nitrc.org/projects/fmri_grocer/). The pipeline consisted of correction of head motion, voxel-wise CBF quantification, spatial normalization with resampling to $2 \times 2 \times 2\,mm^3$ and smoothing with an 8 mm full-width half maximum (FWHM) kernel, similar to a previous test-retest study (Jann et al., 2015). We also tried the same analyses with a 6 mm FWHM smoothing kernel and the results did not change (data not shown). The mean CBF map was obtained by averaging preprocessed CBF images across the time series. In order to eliminate noise from processing, we created a mask for CBF-mean higher than 5 (ml/100 g/min) and lower than 150 (ml/100 g/min) in all absolute CBF-mean maps. Based on this mask, images of task-induced CBF change (ΔCBF) between task- and resting-state were calculated in two separate ways: (1) absolute CBF changes ($\Delta CBF_{absolute}$): the result of subtracting task-based CBF from the average CBF of the two resting-states [$\Delta CBF_{absolute} = CBF_{task} - (CBF_{rest1} + CBF_{rest2})/2$], (2) proportional CBF changes ($\Delta CBF_{proportional}$): the result of dividing $\Delta CBF_{absolute}$ by the average of the two resting-states [$(\Delta CBF_{proportional} = \Delta CBF_{absolute}/(CBF_{rest1} + CBF_{rest2})/2$].

To assess CBF reliability in regions related to the PVT, regions of interest (ROIs) were defined a priori based on the activation map of an independent sample of 43 adults who completed the resting and PVT scans as a part of the larger study at baseline (Yang et al., 2018). Voxel-wise differences between PVT performance and the two rest states were assessed using SPM8 and limited at $p < 0.001$ without correction for multiple comparison (see Fig. 2, row six for un-thresholded t-value map). Clusters larger than 50 voxels were binarized to create exclusive ROI masks (see Table S1). GM and WM masks were generated from gray.nii and white.nii with a threshold of 10% and 90% probability, respectively. Absolute CBF was also calculated for the 12 resting-state networks defined from Zhu et al. (2013), which include: Default mode network (DMN), limbic system network, auditory network, dorsal medial prefrontal network, executive network, left attentional network, ventral medial prefrontal network, primary visual network, right attentional network, salience network, second visual network, and sensorimotor network.

Test-retest reliability of CBF data and PVT performance was assessed using the interclass correlation coefficient (ICC) as defined in the previous study (McGraw and Wong, 1996), where BMS and JMS are the between- and within-subjects errors respectively, EMS is the mean residual error, k is the number of repeated sessions (k = 3) and n is the sample size (n = 15).

$$ICC(A-1) = \frac{BMS - EMS}{BMS + (k-1)EMS + (JMS - EMS)\frac{k}{n}}$$

There are multiple methods to calculate ICC (McGraw and Wong, 1996). Here we used ICC(A-1) rather than the more widely used ICC(C-1) because ICC(A-1) assesses absolute agreement between measurements and takes session effects into account while ICC(C-1) assesses the agreement of ranks between measurements and does not taken session effects into consideration. Since the current study focused more on the absolute CBF values rather than the rank of CBF measurements, it was more appropriate to use ICC(A-1) instead of ICC(C-1). ICC values were characterized as follows: 0–0.39 (poor), 0.40–0.59 (fair), 0.6–0.74 (good), and 0.75–1.0 (excellent) (Hodkinson et al., 2013). All ICC values and their corresponding p values and 95% confidence intervals (CI) were calculated based on MATLAB 8.0 (The Mathworks Inc.) and the fMRI Grocer toolbox (https://www.nitrc. org/projects/fmri_grocer/).

Histograms of voxel-wise ICC values were calculated to qualitatively compare CBF images. Paired Wilcoxon's signed-rank tests (Yan et al., 2013) were used to examine the differences between distributions of ICC values between CBF images. The median ICC (MedICC), an ROI reliability measured by the median of the ICC distributions within a region, was used in further voxel-wise analyses due to its stability under various conditions of smoothing and cluster size (Caceres et al., 2009). The 95% confidence interval of the MedICC for each ROI was calculated from the binomial distribution according to Bland (2000). Since the medians have large number of observations, assuming they are normally distributed, we performed two sample t-tests between ROIs to evaluate which regions have higher stability. We also performed the power calculation. Assuming a difference of 0.2 for between ICC distributions and 0.2 for the ICC standard deviation (Raoult et al., 2011; Steketee et al., 2015), the power calculation (using a Java Application available at https://homepage.divms.uiowa.edu/~rlenth/Power/) indicated that the null hypothesis (i.e., the ICC reliability did not differ between absolute CBF measurements and task-induced CBF changes) could be rejected with a power of 0.95 with a sample size of 15 subjects.

3. Results

3.1. Test-retest reliability of behavioral measurements

Outcomes for the primary measures of PVT performance are shown in Table 1. The mean RT, median RT, SD of RT and mean FRT demonstrated excellent reliability (ICC>0.8) across the three scans.

3.2. Test-retest reliability of absolute CBF measurements

For each absolute CBF measure, group mean maps were highly similar across the three scans (Fig. 1 Panel A, B, & C). Consistent with previous test-retest reliability studies (Hodkinson et al., 2013; Zou et al., 2015), ICC values ranged from good to excellent (ICC > 0.6) for nearly all brain voxels for the mean absolute CBF maps of PVT, REST 1, and REST 2 scans (Fig. 2). Voxel-wise-level test-retest reliability of absolute CBF measurements was evaluated using median ICCs and ICC distributions. Consistent with a previous study (Zou et al., 2015), median ICC value of REST 1 in gray matter was slightly higher than that in white matter (see Table 2). In contrast, median ICCs of PVT and REST 2 scans in gray matter were significantly lower than those in white matter ($p < 0.001$).

Fig. 3 shows the number of voxels against ICC values. Marked by negative skew, the profiles for PVT, REST 1, and REST 2 scans are similar. Wilcoxon signed-rank tests revealed that ICC values were significantly higher for the PVT scans than for REST 1 and REST 2 scans within gray matter, white matter and all ROIs combined (all $p < 0.001$). Median ICC values of PVT, REST 1 and REST 2 scans within gray matter, white matter and 8 task-related ROIs are shown in Fig. 4. Median ICC for the PVT scans was significantly higher than that of REST 1 ($p < 0.001$) in all ROIs except for right lateral occipital gyrus (rLOG) and left lateral Occipital gyrus (lLOG). Similarly, CBF during the PVT scan was more stable than REST 2 ($p < 0.001$) in all ROIs except for right lateral occipital gyrus (rLOG), left lateral occipital gyrus (lLOG), right middle frontal gyrus (rMFG), and orbital part of right superior frontal gyrus (ORBsup).

Table 1
Reliability measures for participants' behavioral performances on PVT.

	Mean RT	Median RT	SD RT	Mean FRT
Scan 1	301.9 (14.6)	280.7 (10.0)	102.9 (27.7)	221.9 (7.3)
Scan 2	322.9 (28.4)	286.3 (12.1)	138.7 (62.6)	225.4 (7.4)
Scan 3	297.0 (22.0)	274.3 (10.3)	104.5 (51.1)	217.9 (6.2)
ICC	0.812	0.820	0.823	0.830
95% CI	0.622–0.926	0.637–0.929	0.643–0.930	0.655–0.933

Note. Standard errors are reported in parentheses.

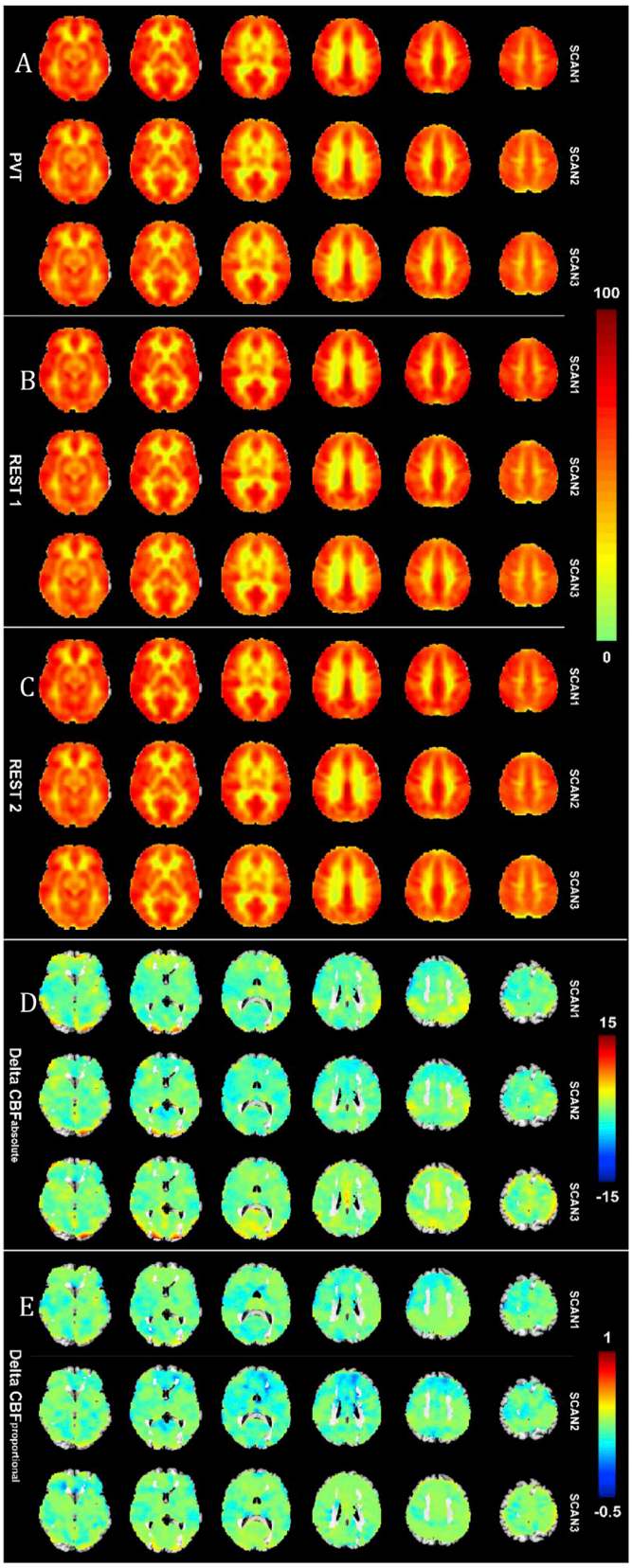

Fig. 1. Group mean maps of five metrics across three scans. Absolute CBF mean map across threes scans during PVT (panel A), REST 1 (panel B) and REST 2 (panel C). For delta CBF changes (Panel D and E), additional mask was applied to avoid noise (see method 2.5). Panel D denotes delta $CBF_{absolute}$, while Panel E illustrates delta $CBF_{propotional}$.

Of the 12 networks generated by in our previous study (Zhu et al., 2013), the default mode network (DMN), primary and secondary visual networks demonstrated the highest reliability during both task and resting states (median ICC > 0.80) (Fig. 5). In general, CBF during PVT performance was more stable than during REST 1 and REST 2 scans in a majority of these networks, especially in auditory and limbic networks.

Consistent with previous findings (Y. Chen et al., 2011; Hodkinson et al., 2013; Zou et al., 2015), ROI-wise analyses further revealed that both task- and resting-state CBF demonstrated good to excellent reliability in both gray matter and white matter (ICC: 0.73–0.88; Table 3). Absolute CBF in the right inferior occipital (rIOG) and left inferior occipital (lIOG) gyri demonstrated the greatest reliability (ICC: 0.89–0.91) among the eight a priori defined ROIs. While CBF reliability for PVT, REST1 and REST2 scans was comparable in the rIOG and lIOG (ICCs > 0.87), reliability during PVT performance was higher as compared with either resting scan in the other six ROIs.

2.3. Test-retest reliability of task-induced relative CBF changes

For each task-induced relative CBF change ($\Delta CBF_{absolute}$ and $\Delta CBF_{proportional}$), group mean maps across the three scans are shown in Fig. 1 Panel D & E. The reliability for task-induced CBF changes was much lower than for task- and resting-state absolute CBF measurements (Fig. 2).

For voxel-wise analyses, the reliability of task-induced CBF changes was poor in gray matter, white matter, and all eight ROIs (ICC < 0.4; Fig. 6). The superior temporal gyrus (rSTG) and lIOG yielded the highest ICC values, but the reliability remained poor (ICC < 0.4). Notably, $\Delta CBF_{proportional}$ was more stable than $\Delta CBF_{absolute}$ in lLOG ($p < 0.01$), while in other task-related ROIs, including GM and WM, $\Delta CBF_{absolute}$ had a slightly higher median ICC value. However, these ICC values are too low to be meaningful and will be excluded from further discussion.

For ROI-based analyses, the reliability of $\Delta CBF_{absolute}$ and $\Delta CBF_{proportional}$ was also poor with confidence intervals almost double those of absolute CBF measurements (see Table 3). Only three out of the eight a-priori ROIs reached statistical significance (right superior temporal gyrus, right inferior occipital gyrus, and left inferior occipital gyrus) but reliability remained poor (ICC < 0.4) with the exception that $\Delta CBF_{proportional}$ in lIOG showed fair reliability across scans (ICC = 0.45).

3. Discussion

Using both voxel-wise and ROI-based analyses, we compared the test-retest reliability of task-based and resting-state absolute CBF measurements, as well as task-induced relative CBF changes during the PVT task. Consistent with Jewett et al. (1999), the behavioral performance of PVT, as measured by four different indices, were highly stable across three scans.

The reliability of resting-state CBF across scans was slightly but significantly lower than that of PVT task-based CBF in gray matter and white matter. Consistent with Zou et al. (2015), gray matter regions had higher reliability than white matter regions during a pre-task resting state. Surprisingly, during the post-task resting state and the task, reliability was higher in whiter matter regions, despite the fact that white matter has a much lower temporal signal to noise ratio as compared with gray matter. Although there have been reports of task-induced activation in white matter (Gawryluk et al., 2014; Mandl et al., 2008; Yarkoni et al., 2009), few studies have investigated the reliability of white matter regions. While our data indicate stability of white matter CBF across scans, it is possible that the relative lack of activation in white matter in general driving the stability of findings. Further research incorporating multi-modal analysis is needed to clarify the role of white matter activation and inter-scan reliability during task scans and resting-states.

Task-based absolute CBF measurement demonstrated the highest reliability in gray matter, particularly in several task-related ROIs including the right superior temporal gyrus, right hippocampus, right

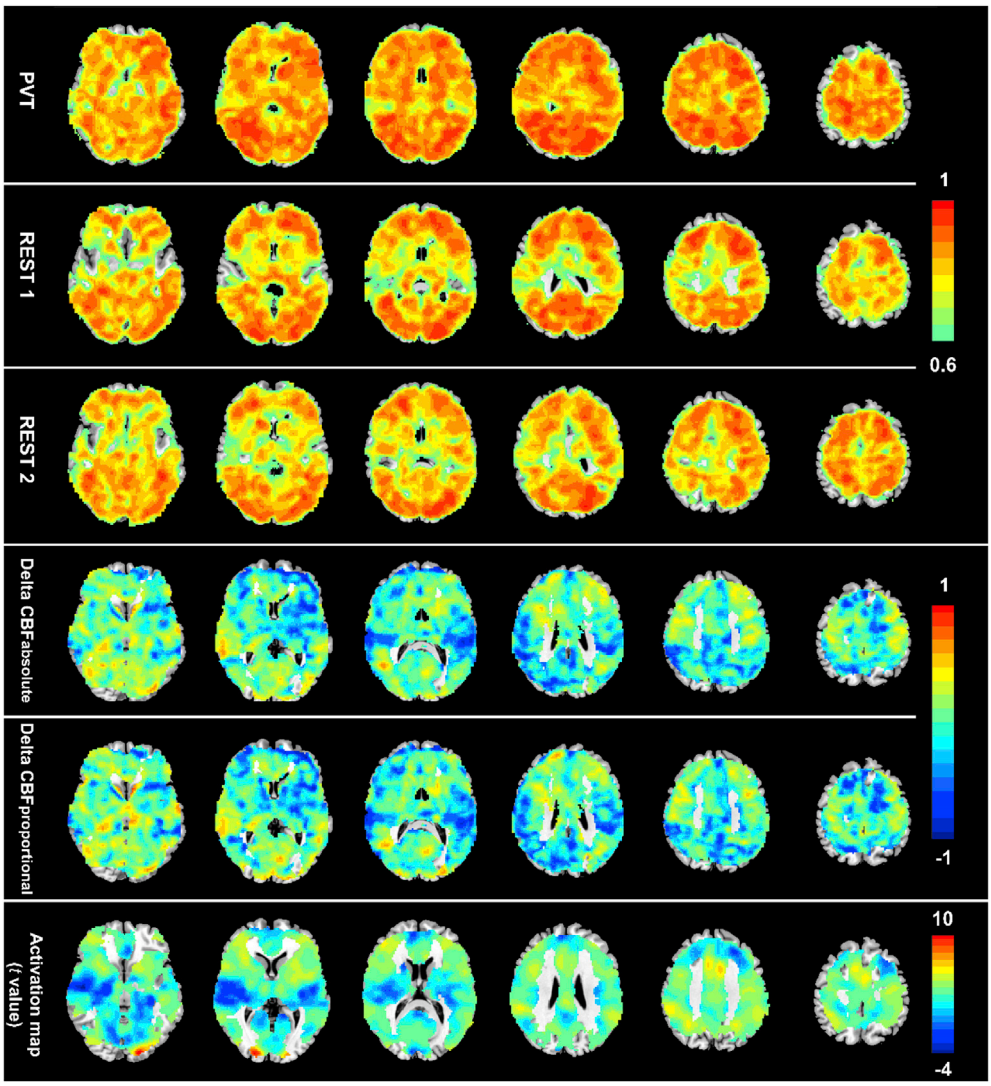

Fig. 2. ICC metrics and activation map. First three rows denote ICC map of PVT, REST 1 and REST 2. For illustration purposes, ICC values were thresholded at 0.6 (ICC>0.6 represents good to excellent reliability). Fourth and fifth row illustrate ICC value of two task-induced CBF changes. ICC values ranges from −1 to 1, although negative ICC values are meaningless. Sixth row represents activation map of PVT (t-value).

Table 2
Median ICCs of absolute CBF measurements in GM and WM.

	REST 1	PVT	REST 2
median ICCs in GM	0.7527	0.7995	0.7484
median ICCs in WM	0.7420	0.8430	0.7839
t-value (GM-WM)	−4.52***	48.15***	20.99***

Note. ***$p < 0.001$.

median cingulate gyrus, and right supramarginal gyrus. This pattern of high reliability in task-based CBF parallels the high reliability of behavioral measures of PVT performance.

To further explore the difference between reliability of task-based and resting-state CBF, we conducted voxel-wise comparisons within the 12 networks defined by Zhu et al. (2013). In all networks, test-retest reliability during PVT was greater than that of the pre- and post-task resting states. The largest difference among the three states were observed in the limbic network, which has been previously associated with PVT performance (Drummond et al., 2005). This finding not only provides found that the reliability was comparable between absolute CBF and task performance, but also suggests that studies dependent on reliability of limbic activation across sessions may benefit from using task-based CBF as opposed to resting-state CBF.

Our findings are also consistent with a recent study investigating the reliability of CBF under four different resting-state/task conditions (Li et al., 2018). Using a PVT paradigm with a much lower task demand (about 16 stimuli in 8 min PVT in that study compared to about 80 stimuli in 10 min PVT in the current study), significantly higher global CBF reliability was also found for the PVT condition than the fixation condition. Notably, the observed ICC values of CBF during the PVT and resting-state scans from our study were higher than those from Li et al. (2018) study, which may be due to the much more strictly controlled study protocol, laboratory environment, and sleep history, activity, and food intake of subjects in the current study.

Task-induced changes in regional CBF during PVT performance demonstrated poor to moderate reliability across the 8 a-priori defined ROIs using both voxel-wise and ROI-wise analyses. Several studies investigated test-retest reliability of task-induced CBF changes with mixed results. While an earlier study reported good to excellent reliability of proportional task-induced brain activity in the primary motor cortex (Raoult et al., 2011), a more recent study showed poor reliability in absolute task-induced CBF changes in a similar brain region (Steketee et al., 2015). It should be noted that the findings of good to excellent reliability were obtained using a “maximum amplitude flexion-extension

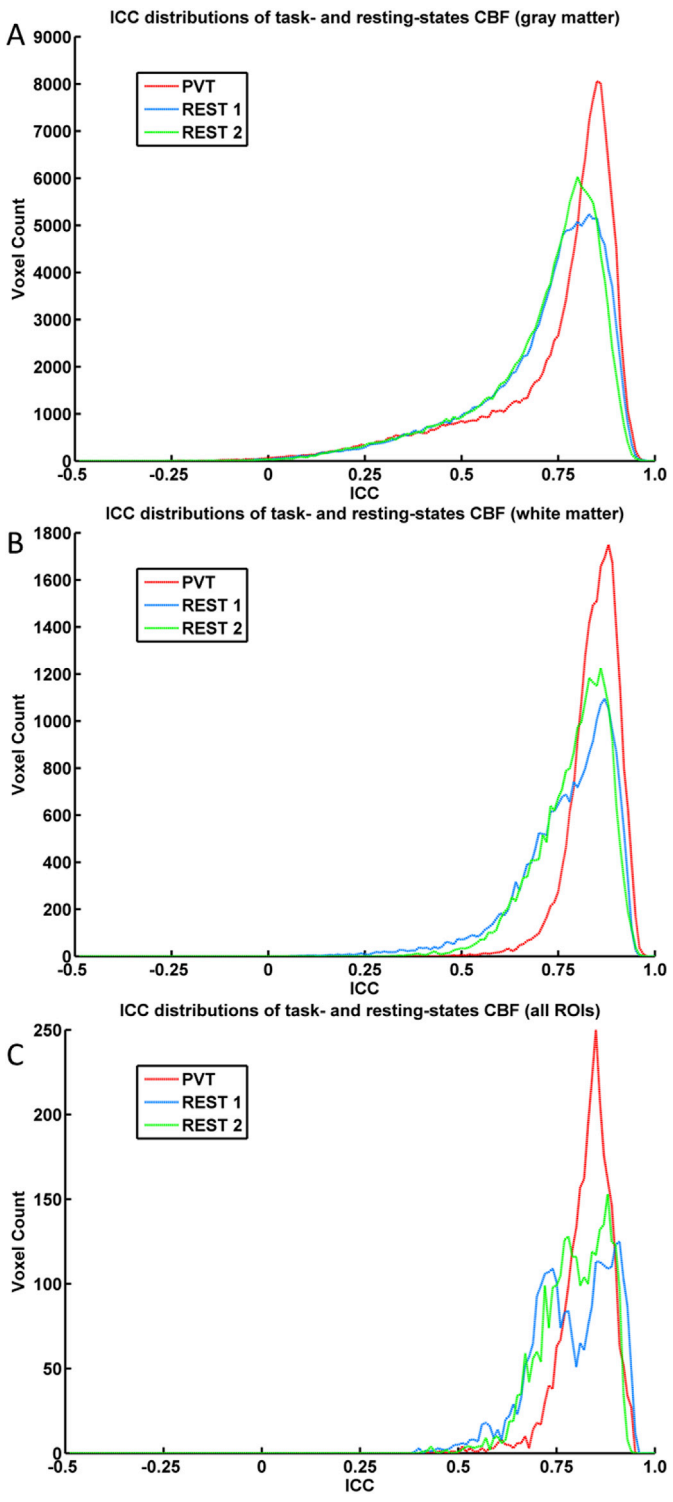

Fig. 3. The ICC distributions of task and two resting-state CBF within GM (panel a), white matter (panel b) and all ROIs combined network (panel c). PVT has higher ICC distributions than REST 1 and REST 2 in all three conditions ($p < 0.001$).

task", which is known to yield high blood flow changes in robust somatotopy of activated brain areas (59% as compared to about 1–15% in the current study). This suggests that inter-scan reliability of task-induced CBF changes may be proportional to the magnitude of task-induced activation in target brain regions. In the PVT task that induced smaller changes of CBF activation, the task-induced CBF differences (calculated by subtracting resting-state CBF from task-based CBF) only reflected a small proportion (−3.7%–13.8%, see Table 3) of absolute CBF, and therefore may contain greater noise and show reduced signal-to-noise ratio (SNR) and decreased reliability across sessions. Indeed, a range of previous studies have consistently suggested that pre-task resting brain activity may contain valuable task-related information and be able to predict task performance (e.g., Baldassarre et al., 2012; Gui et al., 2015; Lim et al., 2010; Sala-Llonch et al., 2012). It is not surprising that we observed task-induced CBF activation to be less stable than task-based absolute CBF since the task-related information in the resting CBF was largely abandoned by calculating the differences between resting state and task performance.

To our knowledge, this is the first study examining the test-retest reliability of task-induced brain function changes using both whole brain voxels-wise, ROI- and network-based analyses. The present data demonstrate comparable median voxel-wise and ROI-wise reliability of CBF changes. Median ICC values of proportional CBF changes were higher than absolute CBF changes in lIOG. The same pattern was also found in the ROI-wise analysis, which supports that proportional CBF changes may be more precise and reliable than absolute CBF changes, likely due to the fact that proportional CBF changes control for effects of basal perfusion variations (Steketee et al., 2015).

The current study has some limitations. First, the test-retest intervals among three scan sessions were 1 or 2 days in the current study, which are relatively short time periods when compared to many other test-retest reliability ASL studies (e.g. Chen et al., 2011; Hodkinson et al., 2013; Jiang et al., 2010; Klomp et al., 2012; Li et al., 2018; Mezue et al., 2014; Steketee et al., 2015; Wang et al., 2011; Wu et al., 2014; Zou et al., 2015). The test-retest reliability of CBF across days may be higher than the reliability across longer test-retest intervals. Further studies are necessary to compare the test-retest reliability of task-based and resting-state absolute CBF measurements as well as task-induced CBF changes with much longer intervals across weeks, months, or even years.

Second, we used a 2D EPI based pCASL sequence to acquire CBF data with a 1 s post-labeling delay time, which is quite short compared to typical ASL imaging parameters recommendation (Alsop et al., 2015). The 1.0 s delay time was chosen to be consistent with our previous pCASL study which obtained high quality of mean CBF maps with concurrent BOLD signal available for functional connectivity analysis (Zhu et al., 2013). Such short post-labeling delay time, however, may limit the accuracy of absolute quantification of regional CBF, especially for white matter. Future studies using a longer post-labeling delay time, such as 1.8 s, should be able to quantify white matter CBF more accurately, since it is close to the 1.7 s arterial transit time of white matter measured by a sensitive single voxel spectroscopic approach (Zhang et al., 2016).

Third, the task-induced CBF reliability assessment in this study was related to average CBF value changes across the complete task-block and included many time-points, and thus might involve different cognitive, motor and affective processes during the 10-min task performance. However, the reliability of PVT behavior performance was also based on performance measures (i.e., mean RT, median RT, SD RT, and mean FRT) averaged across the complete PVT-block, therefore also involving different cognitive, motor and affective processes during the 10-min task performance. Thus, we cannot draw a conclusion from our findings that a single and simple motor outcome may be more reliable than the temporally varying cognitive and affective processes before, after and during such a response. Unfortunately, the SNR of CBF signal from a single time-point of the current pCASL sequence was not good enough and we could not conduct the CBF analysis examining the relevant time frames of each motor response during the PVT. Future studies are needed to address this important question with significantly improved SNR for CBF signal from a single time-point, for example by using the 3D background-suppressed pCASL techniques.

Fourth, although participants' activity and the experimental protocol were strictly controlled before and during the three scan sessions when they stayed in the laboratory across the five days and four nights in this study, we were not able to acquire physiological signals such as heart rate

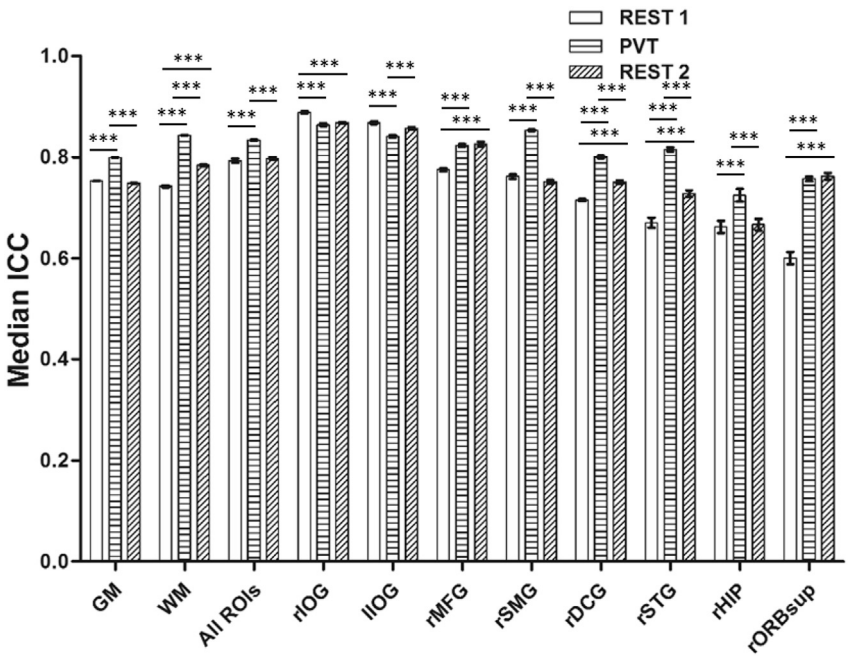

Fig. 4. Median ICCs of task- and resting-state CBF. Median ICC value of PVT was significantly higher than that of REST 1 in six out of eight task-related ROIs but not in the rLOG and lLOG. Similarly, PVT were more stable than REST 2 in the rSMG, rSMG, rSTG and rHIP, but not in the rLOG, lLOG, rMFG and rORBsup. *Note*. Error bar denotes standard error. *** denotes $p < 0.001$.

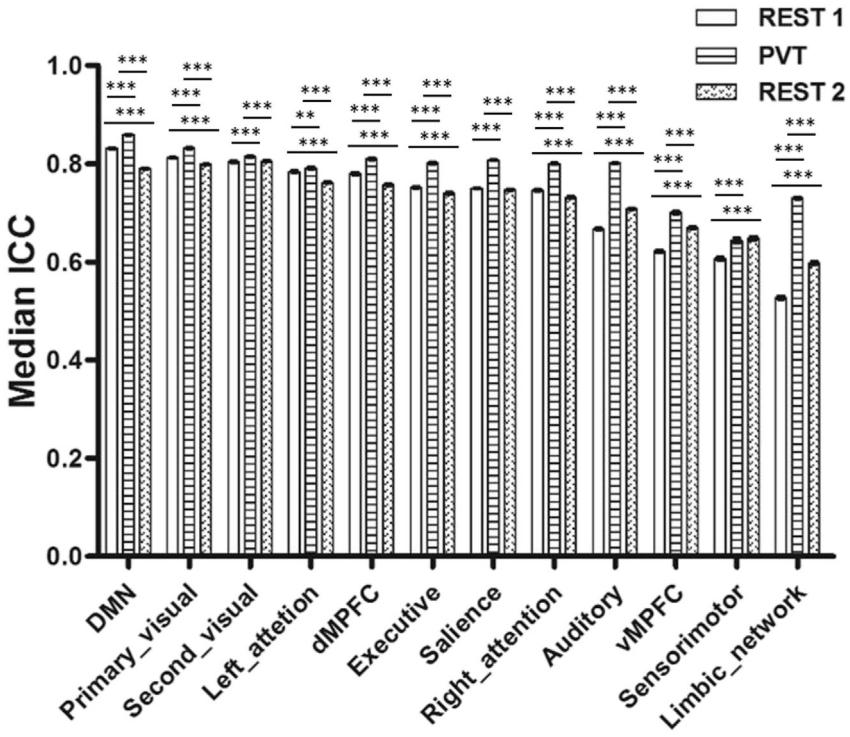

Fig. 5. Median ICCs of task-based and resting-state CBF in 12 brain networks. PVT was more stable than both REST 1 and REST 2 in majority of those 12 networks, especially in auditory and limbic network. *Note*. Error bar denotes standard error. ***, ** denotes $p < 0.001$, $p < 0.01$, respectively.

and respiration data during the fMRI scans. Future studies may include physiological data collection during the ASL scans for noise correction which may further increase the test-retest reliability of CBF measurements.

Finally, the presented study only examined the test-retest reliability of CBF signals quantified by ASL perfusion imaging. A number of other quantitative MRI techniques have also been developed to assess the hemodynamic and metabolic responses to modulations in brain function and activity (for a review, see Pike, 2012), such as the calibrated BOLD, quantitative BOLD, imaging of cerebral metabolic rate of oxygen consumption (CMRO2), and vascular space occupancy (VASO) imaging of cerebral blood volume (CBV). It will be very interesting to examine whether the test-retest reliability of these quantitative MRI methods will similarly show advantages over task-induced hemodynamic and metabolic

Table 3
ROI-wise ICC values.

	GM	WM	All ROIs	rSTG	rHIP	rIOG	lIOG	rDCG	rORBsup	rSMG	rMFG
CBF_{PVT}											
Scan 1	57.7(3.8)	22.1(1.9)	62.7(4.6)	63.5(3.2)	53.7(2.7)	64.2(6.9)	62.3(5.9)	59.6(3.5)	46.6(3.2)	68.6(4.4)	68.5(5.1)
Scan 2	53.9 (3.0)	20.7(1.8)	58.3(3.7)	61.8(2.5)	53.4(2.8)	57.7 (5.7)	56.6 (5.4)	56.1(2.9)	45.1 (2.9)	64.5 (3.2)	64.4 (4.2)
Scan 3	53.8(2.9)	20.0(1.6)	57.4(3.7)	58.5(2.9)	55.9(3.0)	57.6(5.7)	57.8(5.0)	54.6(3.0)	41.6(3.7)	61.8(3.5)	61.7(3.8)
ICC	0.88***	0.91***	0.90***	0.82***	0.77***	0.90***	0.88***	0.82***	0.78***	0.88***	0.84***
95%CI	0.73–0.96	0.78–0.97	0.75–0.97	0.61–0.93	0.56–0.91	0.76–0.96	0.74–0.95	0.62–0.93	0.57–0.91	0.65–0.96	0.65–0.94
$CBF_{REST\ 1}$											
Scan 1	58.5(4.0)	23.3(1.9)	60.5(4.4)	67.6(4.3)	56.7(2.9)	59.2(6.2)	56.5(5.7)	60.2 (3.7)	45.9(3.8)	66.5 (4.2)	68.5(4.8)
Scan 2	54.1 (3.4)	20.0(1.5)	56.0(3.8)	64.5(2.5)	55.6(2.9)	54.1(6.0)	51.8(5.5)	55.2(2.7)	43.9(2.5)	62.6(3.4)	61.4(4.3)
Scan 3	53.6(3.2)	19.5(1.8)	54.1(3.8)	59.8(2.9)	53.0(3.1)	52.8(5.7)	53.0(4.9)	51.4(2.6)	41.8(3.3)	59.6(3.9)	61.2 (4.1)
ICC	0.87***	0.82***	0.89***	0.66***	0.72***	0.91***	0.90***	0.75***	0.65***	0.78***	0.80***
95%CI	0.70–0.95	0.55–0.94	0.70–0.96	0.37–0.86	0.48–0.89	0.80–0.97	0.78–0.96	0.38–0.91	0.38–0.85	0.55–0.91	0.57–0.92
$CBF_{REST\ 2}$											
Scan 1	57.8(4.1)	21.7(2.3)	61.4(4.9)	65.5(3.8)	56.2(3.1)	62.1(7.3)	59.3(6.5)	59.5(3.6)	48.1(3.2)	65.9(4.9)	67.4(5.1)
Scan 2	54.7(3.1)	21.62(1.6)	57.2(3.6)	61.3 (3.0)	53.9(2.5)	55.9 (5.7)	53.8(5.5)	56.8(2.5)	44.8(2.7)	63.2(3.1)	64.5(3.8)
Scan 3	52.7(3.5)	21.1(1.7)	54.5(3.9)	60.3(2.8)	55.9(3.2)	51.1(5.8)	50.6(5.4)	54.3(3.0)	43.1(4.0)	61.1(3.3)	63.7(4.3)
ICC	0.88***	0.89***	0.89***	0.75***	0.70***	0.89***	0.88***	0.78***	0.79***	0.79***	0.84***
95%CI	0.72–0.96	0.76–0.96	0.69–0.96	0.51–0.90	0.44–0.88	0.67–0.96	0.71–0.96	0.56–0.91	0.58–0.92	0.59–0.92	0.67–0.94
ΔCBF_{abs}											
Scan 1	−0.4(0.6)	–	1.9(0.9)	−3.1(1.2)	−2.7(1.2)	3.6(1.8)	4.5(1.4)	−0.3(1.3)	−0.4(1.4)	2.4(1.3)	0.5(2.3)
Scan 2	−0.5(0.6)	–	1.7(0.6)	−1.1(1.1)	−1.3(2.1)	2.7(1.0)	3.8(1.1)	0.1(0.9)	0.8(1.2)	1.6(1.3)	1.4(1.5)
Scan 3	0.6(0.9)	–	3.1(0.7)	−1.6(1.4)	1.5(1.3)	5.7(1.0)	6.0(1.2)	1.8(1.6)	−0.9(1.1)	1.4(1.3)	−0.7(1.8)
ICC	0.26	–	0.32*	0.30*	0.01	0.34*	0.34*	−0.07	0.11	−0.28	−0.15
95%CI	−0.04–0.61	–	0.02–0.65	−0.01–0.64	−0.22–0.37	0.04–0.66	0.03–0.66	−0.29–0.30	−0.17–0.49	−0.42–0.02	−0.35–0.20
ΔCBF_{pro}											
Scan 1	-.003(.01)	–	.036(.02)	-.037(.02)	-.044(.02)	.069(.03)	.102(.03)	.001(.02)	.004(.03)	.042(.02)	.016(.04)
Scan 2	-.005(.01)	–	.031(.01)	-.016(.02)	-.018(.04)	.062(.02)	.090(.03)	-.001(.02)	.016(.02)	.030(.02)	.026(.03)
Scan 3	.020(.02)	–	.064(.01)	-.026(.02)	.029(.02)	.128(.03)	.138(.03)	.037(.03)	-.021(.04)	.029(.02)	-.001(.03)
ICC	0.12	–	0.28*	0.28*	0.07	040**	0.45**	−0.12	−0.02	−0.28	−0.14
95%CI	−0.15–0.48	–	−0.01–0.61	−0.03–0.63	−0.17–0.43	0.10–0.70	0.13–0.73	−0.32–0.24	−0.27–0.36	−0.42–0.02	−0.35–0.22

Note. Means of CBF of each scan are reported with standard errors in parentheses.
*$p < 0.05$ **$p < 0.01$ ***$p < 0.00$.

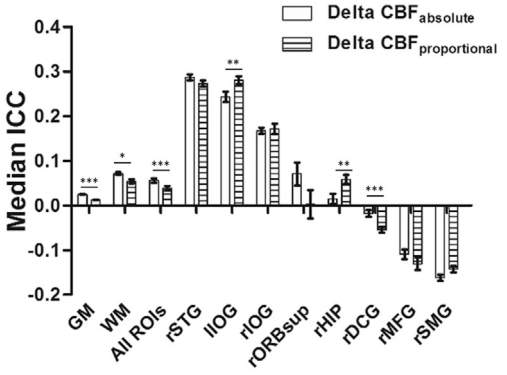

Fig. 6. Median ICCs of task-induced CBF changes. The highest ICC value was found in the right superior temporal gyrus (rSTG) and lIOG, with ICC values around 0.3. In particular, delta $CBF_{proportional}$ were more stable than delta $CBF_{absolute}$ in the lLOG. *Note*. Error bars denote standard errors.

changes, which may provide a potentially more valuable set of tools for exploring pathophysiological conditions in future clinical applications.

In summary, here we show that absolute CBF measures have excellent reliability across three neuroimaging scans both at resting states and during performance of a simple vigilance task, which was comparable to the reliability of behavioral measures of task performance. In contrast, the reliability of task-induced CBF changes was much lower, demonstrating poor to moderate test-retest reliability. These findings suggest that absolute CBF measures, particularly during task performance, are a more reliable biomarker than task-induced CBF changes for characterizing brain function for future longitudinal and clinical studies.

Acknowledgments

This research was supported in part by the grants from National Institutes of Health (R01 HL102119, R01 MH107571, and R21 AG051981), Shanghai International Studies University Major Research Project (Grant number 20171140020), and the Program for Professors of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (TP2016020). The funders had no role in the study design, data collection and analysis, data interpretation, writing of the manuscript, or the decision to submit the article for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.03.016.

References

Alsop, D.C., Detre, J.A., Golay, X., Günther, M., Hendrikse, J., Hernandez-Garcia, L., Lu, H., Macintosh, B.J., Parkes, L.M., Smits, M., Van Osch, M.J.P., Wang, D.J.J., Wong, E.C., Zaharchuk, G., 2015. 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. 73, 102–116. https://doi.org/10.1002/mrm.25197.

Baldassarre, A., Lewis, C.M., Committeri, G., Snyder, A.Z., Romani, G.L., Corbetta, M., 2012. Individual variability in functional connectivity predicts performance of a perceptual task. Proc. Natl. Acad. Sci. Unit. States Am. 109, 3516–3521. https://doi.org/10.1073/pnas.1113148109.

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. 1191, 133–155. https://doi.org/10.1111/j.1749-6632.2010.05446.x.

Buxton, R.B., Uludağ, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. In: NeuroImage. https://doi.org/10.1016/j.neuroimage.2004.07.013.

Caceres, A., Hall, D.L., Zelaya, F.O., Williams, S.C.R., Mehta, M.A., 2009. Measuring fMRI reliability with the intra-class correlation coefficient. Neuroimage 45, 758–768. https://doi.org/10.1016/j.neuroimage.2008.12.035.

Chen, J.J., Rosas, H.D., Salat, D.H., 2011. Age-associated reductions in cerebral blood flow are independent from regional atrophy. Neuroimage 55, 468–478. https://doi.org/10.1016/j.neuroimage.2010.12.032.

Chen, Y., Wang, D.J.J., Detre, J.A., 2011. Test-retest reliability of arterial spin labeling with common labeling strategies. J. Magn. Reson. Imaging 33, 940–949. https://doi.org/10.1002/jmri.22345.

Detre, J.A., Wang, J., 2002. Technical aspects and utility of fMRI using BOLD and ASL. Clin. Neurophysiol. 113, 621–634. https://doi.org/10.1016/S1388-2457(02)00038-X.

Detre, J.A., Wang, J., Wang, Z., Rao, H., 2009. Arterial spin-labeled perfusion MRI in basic and clinical neuroscience. Curr. Opin. Neurol. 348–355. https://doi.org/10.1097/WCO.0b013e32832d9505.

Drummond, S.P.A., Bischoff-Grethe, A., Dinges, D.F., Ayalon, L., Mednick, S.C., Meloy, M.J., 2005. The neural basis of the psychomotor vigilance task. Sleep 28, 1059–1068. https://doi.org/10.1093/sleep/28.9.1059.

Evranos, B., Kocyigit, D., Gurses, K.M., Yalcin, M.U., Sahiner, M.L., Kaya, E.B., Ozer, N., Aytemir, K., 2016. Increased left atrial pressure predicts recurrence following successful cryoablation for atrial fibrillation with second-generation cryoballoon. J. Interv. Card Electrophysiol. 46, 145–151. https://doi.org/10.1007/s10840-016-0107-8.

Fang, Z., Spaeth, A.M., Ma, N., Zhu, S., Hu, S., Goel, N., Detre, J.A., Dinges, D.F., Rao, H., 2015. Altered salience network connectivity predicts macronutrient intake after sleep deprivation. Sci. Rep. 5. https://doi.org/10.1038/srep08215.

Fazlollahi, A., Bourgeat, P., Liang, X., Meriaudeau, F., Connelly, A., Salvado, O., Calamante, F., 2015. Reproducibility of multiphase pseudo-continuous arterial spin labeling and the effect of post-processing analysis methods. Neuroimage 117, 191–201. https://doi.org/10.1016/j.neuroimage.2015.05.048.

Franklin, T.R., Shin, J., Jagannathan, K., Suh, J.J., Detre, J.A., O'Brien, C.P., Childress, A.R., 2012. Acute baclofen diminishes resting baseline blood flow to limbic structures: a perfusion fMRI study. Drug Alcohol Depend. 125, 60–66. https://doi.org/10.1016/j.drugalcdep.2012.03.016.

Franklin, T.R., Wang, Z., Sciortino, N., Harper, D., Li, Y., Hakun, J., Kildea, S., Kampman, K., Ehrman, R., Detre, J.A., O'Brien, C.P., Childress, A.R., 2011. Modulation of resting brain cerebral blood flow by the GABA B agonist, baclofen: a longitudinal perfusion fMRI study. Drug Alcohol Depend. 117, 176–183. https://doi.org/10.1016/j.drugalcdep.2011.01.015.

Gawryluk, J.R., Mazerolle, E.L., D'Arcy, R.C.N., 2014. Does functional MRI detect activation in white matter? A review of emerging evidence, issues, and future directions. Front. Neurosci. 8, 839. https://doi.org/10.3389/fnins.2014.00239.

Gevers, S., Van Osch, M.J., Bokkers, R.P.H., Kies, D.A., Teeuwisse, W.M., Majoie, C.B., Hendrikse, J., Nederveen, A.J., 2011. Intra-and multicenter reproducibility of pulsed, continuous and pseudo-continuous arterial spin labeling methods for measuring cerebral perfusion. J. Cereb. Blood Flow Metab. 31, 1706–1715. https://doi.org/10.1038/jcbfm.2011.10.

Gui, D., Xu, S., Zhu, S., Fang, Z., Spaeth, A.M., Xin, Y., Feng, T., Rao, H., 2015. Resting spontaneous activity in the default mode network predicts performance decline during prolonged attention workload. Neuroimage 120, 323–330. https://doi.org/10.1016/j.neuroimage.2015.07.030.

Hodkinson, D.J., Krause, K., Khawaja, N., Renton, T.F., Huggins, J.P., Vennart, W., Thacker, M.A., Mehta, M.A., Zelaya, F.O., Williams, S.C.R., Howard, M.A., 2013. Quantifying the test-retest reliability of cerebral blood flow measurements in a clinical model of on-going post-surgical pain: a study using pseudo-continuous arterial spin labelling. NeuroImage Clin 3, 301–310. https://doi.org/10.1016/j.nicl.2013.09.004.

Jahng, G.-H., Song, E., Zhu, X.-P., Matson, G.B., Weiner, M.W., Schuff, N., 2005. Human brain: reliability and reproducibility of pulsed arterial spin-labeling perfusion MR imaging. Radiology 234, 909–916. https://doi.org/10.1148/radiol.2343031499.

Jann, K., Gee, D.G., Kilroy, E., Schwab, S., Smith, R.X., Cannon, T.D., Wang, D.J.J., 2015. Functional connectivity in BOLD and CBF data: SIMILARITY and reliability of resting brain networks. Neuroimage 106, 111–122. https://doi.org/10.1016/j.neuroimage.2014.11.028.

Jewett, M.E., Dijk, D.J., Kronauer, R.E., Dinges, D.F., 1999. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 22, 171–179. https://doi.org/10.1093/sleep/22.2.171.

Jiang, L., Kim, M., Chodkowski, B.A., Donahue, M.J., Pekar, J.J., Van Zijl, P.C.M., Albert, M., 2010. Reliability and reproducibility of perfusion MRI in cognitively normal subjects. Magn. Reson. Imaging 28, 1283–1289. https://doi.org/10.1016/j.mri.2010.05.002.

Kim, J., Whyte, J., Patel, S., Europa, E., Slattery, J., Coslett, H.B., Detre, J.A., 2012. A perfusion fMRI study of the neural correlates of sustained-attention and working-memory deficits in chronic traumatic brain injury. Neurorehabilitation Neural Repair 26, 870–880. https://doi.org/10.1177/1545968311434553.

Klomp, A., Caan, M.W.A., Denys, D., Nederveen, A.J., Reneman, L., 2012. Feasibility of ASL-based phMRI with a single dose of oral citalopram for repeated assessment of serotonin function. Neuroimage 63, 1695–1700. https://doi.org/10.1016/j.neuroimage.2012.07.038.

Li, Z., Vidorreta, M., Katchmar, N., Alsop, D.C., Wolf, D.H., Detre, J.A., 2018. Effects of resting state condition on reliability, trait specificity, and network connectivity of brain function measured with arterial spin labeled perfusion MRI. Neuroimage 173, 165–175. https://doi.org/10.1016/j.neuroimage.2018.02.028.

Lim, J., Dinges, D.F., 2010. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychol. Bull. 136, 375–389. https://doi.org/10.1037/a0018883.

Lim, J., Dinges, D.F., 2008. Sleep deprivation and vigilant attention. In: Annals of the New York Academy of Sciences, pp. 305–322. https://doi.org/10.1196/annals.1417.002.

Lim, J., Wu, W. chau, Wang, J., Detre, J.A., Dinges, D.F., Rao, H., 2010. Imaging brain fatigue from sustained mental workload: an ASL perfusion study of the time-on-task effect. Neuroimage 49, 3426–3435. https://doi.org/10.1016/j.neuroimage.2009.11.020.

Mandl, R.C.W., Schnack, H.G., Zwiers, M.P., van der Schaaf, A., Kahn, R.S., Hulshoff Pol, H.E., 2008. Functional diffusion tensor imaging: measuring task-related fractional anisotropy changes in the human brain along white matter tracts. PLoS One 3. https://doi.org/10.1371/journal.pone.0003631.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1, 30–46. https://doi.org/10.1037/1082-989X.1.1.30.

Mezue, M., Segerdahl, A.R., Okell, T.W., Chappell, M.A., Kelly, M.E., Tracey, I., 2014. Optimization and reliability of multiple postlabeling delay pseudo-continuous arterial spin labeling during rest and stimulus-induced functional task activation. J. Cereb. Blood Flow Metab. 34, 1919–1927. https://doi.org/10.1038/jcbfm.2014.163.

Pike, G.B., 2012. Quantitative functional MRI: concepts, issues and future challenges. Neuroimage 62, 1234–1240. https://doi.org/10.1016/j.neuroimage.2011.10.046.

Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A.B.M., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., Meyer-Lindenberg, A., 2012. Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. Neuroimage 60, 1746–1758. https://doi.org/10.1016/j.neuroimage.2012.01.129.

Raichle, M.E., 1998. Behind the scenes of functional brain imaging: a historical and physiological perspective. Proc. Natl. Acad. Sci. Unit. States Am. 95, 765–772. https://doi.org/10.1073/pnas.95.3.765.

Raoult, H., Petr, J., Bannier, E., Stamm, A., Gauvrit, J.Y., Barillot, C., Ferré, J.C., 2011. Arterial spin labeling for motor activation mapping at 3T with a 32-channel coil: reproducibility and spatial accuracy in comparison with BOLD fMRI. Neuroimage 58, 157–167. https://doi.org/10.1016/j.neuroimage.2011.06.011.

Sala-Llonch, R., Peña-Gómez, C., Arenaza-Urquijo, E.M., Vidal-Piñeiro, D., Bargalló, N., Junqué, C., Bartrés-Faz, D., 2012. Brain connectivity during resting state and subsequent working memory task predicts behavioural performance. Cortex 48, 1187–1196. https://doi.org/10.1016/j.cortex.2011.07.006.

Shu, C.J., Campbell, D.O., Lee, J.T., Tran, A.Q., Wengrod, J.C., Witte, O.N., Phelps, M.E., Satyamurthy, N., Czernin, J., Radu, C.G., 2010. Novel PET probes specific for deoxycytidine kinase. J. Nucl. Med. 51, 1092–1098. https://doi.org/10.2967/jnumed.109.073361.

Smith, C.S., Reilly, C., Midkiff, K., 1989. Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. J. Appl. Psychol. 74, 728–738. https://doi.org/10.1037/0021-9010.74.5.728.

Steketee, R.M.E., Mutsaerts, H.J.M.M., Bron, E.E., Van Osch, M.J.P., Majoie, C.B.L.M., Van Der Lugt, A., Nederveen, A.J., Smits, M., 2015. Quantitative functional Arterial Spin Labeling (fASL) MRI - sensitivity and reproducibility of regional CBF changes using pseudo-continuous ASL product sequences. PLoS One 10 e0132929. https://doi.org/10.1371/journal.pone.0132929.

Stevens, M.T.R., D'Arcy, R.C.N., Stroink, G., Clarke, D.B., Beyea, S.D., 2013. Thresholds in fMRI studies: reliable for single subjects? J. Neurosci. Methods 219, 312–323. https://doi.org/10.1016/j.jneumeth.2013.08.005.

Tancredi, F.B., Lajoie, I., Hoge, R.D., 2015. Test-retest reliability of cerebral blood flow and blood oxygenation level-dependent responses to hypercapnia and hyperoxia using dual-echo pseudo-continuous arterial spin labeling and step changes in the fractional composition of inspired gases. J. Magn. Reson. Imaging 42, 1144–1157. https://doi.org/10.1002/jmri.24878.

Telischak, N.A., Detre, J.A., Zaharchuk, G., 2015. Arterial spin labeling MRI: clinical applications in the brain. J. Magn. Reson. Imaging 41, 1165–1180. https://doi.org/10.1002/jmri.24751.

Upadhyay, J., Lemme, J., Anderson, J., Bleakman, D., Large, T., Evelhoch, J.L., Hargreaves, R., Borsook, D., Becerra, L., 2015. Test-retest reliability of evoked heat stimulation BOLD fMRI. J. Neurosci. Methods 253, 38–46. https://doi.org/10.1016/j.jneumeth.2015.06.001.

Van Gelderen, P., De Zwart, J.A., Duyn, J.H., 2008. Pittfalls of MRI measurement of white matter perfusion based on arterial spin labeling. Magn. Reson. Med. 59, 788–795. https://doi.org/10.1002/mrm.21515.

Wang, Y., Saykin, A.J., Pfeuffer, J., Lin, C., Mosier, K.M., Shen, L., Kim, S., Hutchins, G.D., 2011. Regional reproducibility of pulsed arterial spin labeling perfusion imaging at 3T. Neuroimage 54, 1188–1195. https://doi.org/10.1016/j.neuroimage.2010.08.043.

Wolf, R.L., Detre, J.A., 2007. Clinical neuroimaging using arterial spin-labeled perfusion magnetic resonance imaging. Neurotherapeutics 4, 346–359. https://doi.org/10.1016/j.nurt.2007.04.005.

Wu, B., Lou, X., Wu, X., Ma, L., 2014. Intra- and interscanner reliability and reproducibility of 3D whole-brain pseudo-continuous arterial spin-labeling MR perfusion at 3T. J. Magn. Reson. Imaging 39, 402–409. https://doi.org/10.1002/jmri.24175.

Xu, G., Rowley, H.A., Wu, G., Alsop, D.C., Shankaranarayanan, A., Dowling, M., Christian, B.T., Oakes, T.R., Johnson, S.C., 2010. Reliability and precision of pseudo-continuous arterial spin labeling perfusion MRI on 3.0 T and comparison with15O-water PET in elderly subjects at risk for Alzheimer's disease. NMR Biomed. 23, 286–293. https://doi.org/10.1002/nbm.1462.

Yan, C.G., Craddock, R.C., Zuo, X.N., Zang, Y.F., Milham, M.P., 2013. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. Neuroimage 80, 246–262. https://doi.org/10.1016/j.neuroimage.2013.04.081.

Yang, F.N., Xu, S., Chai, Y., Basner, M., Dinges, D.F., Rao, H., 2018. Sleep deprivation enhances inter-stimulus interval effect on vigilant attention performance. Sleep 41 zsy189. https://doi.org/10.1093/sleep/zsy189.

Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., Braver, T.S., 2009. BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4. https://doi.org/10.1371/journal.pone.0004257.

Zhang, X., Ronen, I., Kan, H.E., Teeuwisse, W.M., van Osch, M.J.P., 2016. Time-efficient measurement of multi-phase arterial spin labeling MR signal in white matter. NMR Biomed. 29, 1519–1525. https://doi.org/10.1002/nbm.3603.

Zhu, S., Fang, Z., Hu, S., Wang, Z., Rao, H., 2013. Resting state brain function analysis using concurrent BOLD in ASL perfusion fMRI. PLoS One 8. https://doi.org/10.1371/journal.pone.0065884.

Zou, Q., Miao, X., Liu, D., Wang, D.J.J., Zhuo, Y., Gao, J.H., 2015. Reliability comparison of spontaneous brain activities between BOLD and CBF contrasts in eyes-open and eyes-closed resting states. Neuroimage 121, 91–105. https://doi.org/10.1016/j.neuroimage.2015.07.044.