Review Article

JCBFM

Journal of Cerebral Blood Flow & Metabolism
41(11) 1–18

DOI: 10.1177/0271678X211056702
journals.sagepub.com/home/jcbfm
SAGE

Reproducibility of cerebrovascular reactivity measurements: A systematic review of neuroimaging techniques*

Moss Y Zhao[1], Amanda Woodward[2], Audrey P Fan[3,4], Kevin T Chen[1], Yannan Yu[1], David Y Chen[5,6], Michael E Moseley[1] and Greg Zaharchuk[1]

Abstract

Cerebrovascular reactivity (CVR), the capacity of the brain to increase cerebral blood flow (CBF) to meet changes in physiological demand, is an important biomarker to evaluate brain health. Typically, this brain "stress test" is performed by using a medical imaging modality to measure the CBF change between two states: at baseline and after vasodilation. However, since there are many imaging modalities and many ways to augment CBF, a wide range of CVR values have been reported. An understanding of CVR reproducibility is critical to determine the most reliable methods to measure CVR as a clinical biomarker. This review focuses on CVR reproducibility studies using neuroimaging techniques in 32 articles comprising 427 total subjects. The literature search was performed in PubMed, Embase, and Scopus. The review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). We identified 5 factors of the experimental subjects (such as sex, blood characteristics, and smoking) and 9 factors of the measuring technique (such as the imaging modality, the type of the vasodilator, and the quantification method) that have strong effects on CVR reproducibility. Based on this review, we recommend several best practices to improve the reproducibility of CVR quantification in neuroimaging studies.

Keywords

Cerebrovascular reactivity, cerebrovascular reserve, reproducibility, repeatability, brain stress test, systematic review

Received 4 January 2021; Revised 25 August 2021; Accepted 3 October 2021

Introduction

Cerebrovascular reactivity (CVR) is an important measure of the ability of the brain vasculature to increase cerebral blood flow (CBF) in response to a vasoactive stimulus. Clinically, CVR is a crucial biomarker to determine the health of brain tissues as well as the future risk of cerebrovascular diseases.[1,2] CVR reduction has been linked with several risk factors including vessel stenosis and/or occlusion in sickle cell disease,[3] deposition of vasoconstrictors such as endothelin-1 and β-amyloid in Alzheimer's disease,[4,5] and small and large vessel damage in patients with diabetes.[6] Impaired CVR deleteriously affects the amount of vasodilation of the brain under stress conditions and can cause chronic neuropathology.[7] Specifically, without enough capacity or reserve, CBF cannot increase to meet higher physiological demand to sustain the function and metabolism of brain tissues. Diminished CVR

[1]Department of Radiology, Stanford University, Stanford, CA, USA
[2]Lane Medical Library, Stanford University, Stanford, CA, USA
[3]Department of Biomedical Engineering, University of California Davis, Davis, CA, USA
[4]Department of Neurology, University of California Davis, Davis, CA, USA
[5]Department of Medical Imaging, Taipei Medical University – Shuan-Ho Hospital, New Taipei City
[6]Department of Radiology, School of Medicine, Taipei Medical University, Taipei

*Research materials supporting this publication can be accessed at https://doi.org/10.25740/hd852bg4538.

Corresponding authors:
Moss Y Zhao, Department of Radiology, Stanford University, Stanford, CA, USA.
Email: mosszhao@stanford.edu
Greg Zaharchuk, Department of Radiology, Stanford University, Stanford, CA, USA.
Email: gregz@stanford.edu

levels have been linked with stroke, dementia, cognitive impairment, and hypertension.[3,8,9]

Two key components can affect the measurement and reproducibility of CVR: the CBF quantification technique and the method to achieve the vasoactive stimulus. A typical CVR exam involves two CBF measurements separated by the administration of the vasoactive stimulus. The CBF measurement techniques include different medical imaging modalities such as Transcranial Doppler (TCD) ultrasound, single-photon emission computerized tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI). The main advantage of TCD is that it is relatively easy to perform CBF or CVR measurements at the bedside, making it a fast and practical technique for screening patients at high risk of cerebrovascular diseases.[10] However, since TCD measures blood flood in the arteries, it lacks the spatial specificity of identifying abnormal CBF or CVR in different regions of the brain. The first CVR study was conducted using Xe-133 SPECT to measure the augmentation of CBF induced by CO_2 challenge in diabetes patients.[11] Results from this study showed that patients with a lower CVR experienced a higher risk of cerebrovascular disease because they were unable to meet the increasing demand for CBF. In addition to Xe-133, I-123 iodoamphetamine (IMP) is another commonly used radiotracer for SPECT-based CBF and CVR measurement.[12] Due to the long half-life of IMP (13.22 hours), however, the CVR experiments must be performed on multiple days to ensure that the tracer administered in the baseline condition does not affect the CBF measurement after vasodilation.

More recently, PET-based techniques have become the reference standard technique of CVR studies due to their high accuracy in CBF quantification both before and after the administration of vasodilators.[13] On the other hand, PET data is limited in spatial resolution due to a combination of factors including the physical size of the detector, positron range, sampling geometry, and statistical noise.[14] In a typical CVR study using PET, radiolabeled ^{15}O-water is commonly used as the tracer.[15] The relatively short half-life of ^{15}O-water water (2.04 minutes) allows the pre- and post-vasodilation CBF measurements to be obtained in a single imaging session. The tracer injected in the baseline condition can decay completely before the administration of vasodilators so that no remaining tracer is left to alter the CBF measurement after vasodilation. On the other hand, the short half-life of ^{15}O-water water requires that PET imaging and tracer production (cyclotron) facilities must be nearby, limiting the widespread use of PET for CVR experiments.

As a non-invasive and radiation-free modality, MRI has allowed both direct and indirect CVR measurements using such techniques as phase contrast (PC), blood oxygenation level dependent (BOLD) imaging, and arterial spin labeling (ASL). Compared with PET and SPECT, the advantages of MRI include higher spatial resolutions and that procedures can be repeated more easily without the additional supply of tracers or contrast agents. For example, the spatial resolution of the commonly used T1-weighted MRI can be in the order of $1 \times 1 \times 1\,mm^3$ and the spatial resolution of functional MRI scans such as BOLD is around $3.5 \times 3.5 \times 5\,mm^3$,[16] making MRI a preferred modality to assess subtle spatial variations. In PC-based CVR experiments, the blood flow volume (measured in ml/min) of the individual carotid and vertebral arteries can be used to measure CBF, but CVR cannot be assessed at the individual voxel level.[17] In BOLD-based CVR studies, the percentage change of BOLD signal intensity is considered to reflect a complex combination of arterial oxygenation, CBF, cerebral blood volume, and oxygen extraction fraction.[18] Arterial spin labeling (ASL) uses magnetically labeled blood water and enables the direct quantification of absolute CBF and CVR of the whole brain and regional brain territories.[19] ASL measurements, however, can be inaccurate in regions with very short or very long arterial transit time.

The choice of vasoactive stimulus can be broadly divided into intravenous drug administration and gas challenges, both of which act by increasing arterial partial pressure of CO_2, which causes robust cerebral vasodilation. Acetazolamide (ACZ), a carbonic anhydride inhibitor, is the most widely used vasodilation drug administered intravenously.[20] Gas challenge stimuli include having the subject control their breathing (primarily through hyperventilation or breath-holding approaches) to change blood CO_2 or having them inhale gas mixtures with alterations of the CO_2 concentration to cause hypocapnia and hypercapnia.[21,22] Factors affecting the selection of vasoactive stimuli used in CVR experiments include the desired extent of vasodilation, the number of vasodilation challenges performed per imaging session, availability of the vasoactive stimulus, and potential side effects. Thus, the key considerations for the implementation and interpretation of a CVR experiment are selecting a specific neuroimaging technique and vasoactive stimulus suitable to the population of interest.

As there are many combinations of imaging techniques and vasodilatory challenge approaches, it is not surprising that a wide range of CVR values have been reported in both patients and healthy volunteers. We define the factors (such as the type of imaging modality and vasodilator) associated with the experimental design and facilities as technological factors. In addition, CVR results may also depend on other factors,

such as the age and sex of the participants and the cognitive state during the experiments.[23–25] We define these intrinsic characteristics of the experimental subjects as physiological factors. The broad range of CVR variability between subjects may indicate that (1) the true CVR of each individual is indeed highly different from each other or (2) poor reproducibility of the current measuring techniques masks the genuine CVR variation of the population. It is important to distinguish these two cases, as reproducibility is critically important for making clinical decisions based on CVR information. This work is a systematic review focused on CVR reproducibility studies, using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) framework, with a focus on the impact of technological factors associated with the experimental design and physiological factors associated with the subject.

Methods

We followed the PRISMA checklist developed by Moher et al., throughout the systematic reviewing process.[26] Our search included subject headings and keywords for 4 key concepts: (1) cerebrovascular circulation including reserve and reactivity; (2) neuroimaging techniques; (3) vasoactive stimuli; (4) reproducibility of results. The search was developed in PubMed and then adapted for the additional information sources, and the exact keywords of the full search strategy for each information source are available in tables of search queries in Appendix 1. The relevant search query was performed in PubMed (https://pubmed.ncbi.nlm.nih.gov/), Embase (https://www.embase.com/#search), and Scopus (https://www.scopus.com/search/form.uri?display=basic) on April 21, 2020. No date limits were placed on the search, and the results were limited to English language studies only. Animal studies were excluded from the search results. All study types were included in the search results except conference papers and abstracts.

Four steps were taken to select the relevant studies from the search results. (1) All duplicated search results were merged. (2) The title and abstract of each study were screened using the Covidence software (Covidence, Melbourne, Australia) and the inclusion and exclusion criteria shown in Table 1 by the authors of this work independently. The studies that did not satisfy the criteria were excluded. If 2 authors could not decide the relevance of a particular study based on the title and abstract information only, this study was retained for further screening in the next step. (3) The full text of the selected studies in Step 2 was screened using the same inclusion and exclusion criteria in Table 1. All conflicts of opinions were resolved to decide the relevance of each study, and the number of papers excluded was recorded for each exclusion criteria. (4) The studies selected from Step 3 formed the collection of studies for data extraction. Information extracted from the selected studies included the year of publication, location of the study, funding sources, subject information, sample size, experimental condition, imaging modality, region of interest (ROI), type of vasodilatory challenge, data analysis method, and time between repeated measurements.

Results

Summary of literature search

Applying the search queries to three databases, a total of 1625 papers were identified, 737 of which were duplicates, resulting in 888 unique articles for the initial title and abstract screening. Among the 888 studies, 92 were included in the full-text screening. After screening the full text of the 92 papers, 32 were retained for content extraction and quality assessment, leaving 60 studies excluded due to reasons such as not a reproducibility study, no quantitative CVR values, non-brain study, and unable to find the full text as shown in Figure 1. The 32 selected articles formed the basis for the rest of this review, as listed in Table 2.

Table 1. Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Peer reviewed journal article	Conference abstracts and conference papers
Human studies	Animal and phantom studies
English language	Non English language
Age of subjects greater than 18 years old	Age of subjects less than or equal to 18 years old
Vessel dilators included acetazolamide, gas challenge, and breath-hold	Free breathing
Reproducibility studies: at least two repeated measurements using the same protocol	Single session measurement or different protocols
Reported quantitative CVR values	No quantitative CVR measurement
Full text available	Full text unavailable

The 32 selected articles included reproducibility metrics on a total of 427 subjects (45% female in the 26 studies that reported sex distributions, age range: 18 and 83 years). Among these selected studies, the primary metrics of reproducibility were the within-subject coefficient of variation (wsCV) of repeated CVR measurement (19 out of 32 studies, 59%) and ICC (16 out of 32 studies, 50%); 8 studies (25%) used both of them. A wsCV less than 33% or ICC greater than 0.44 was considered the threshold for good reproducibility by 15 studies. Other metrics included the ratio between CVR differences and mean of the repeated sessions, Wilcoxon signed-rank test, and normalized standard deviation (SD) of repeated CVR measurement. Linear regression between CVR measurement of two sessions was used in 4 out of the 32 selected studies separately. In comparing the CVR levels between the selected studies, the data were grouped by the type of vasodilator used in the experiment as shown in Figure 2: acetazolamide (CVR unit: %); breath-hold (CVR unit %/mmHg); hypercapnia gas challenge (CVR unit: %/mmHg). CVR studies using hyperventilation were not included in this systematic review. After comparing the studies based on their experimental facilities, types of vasodilators, and the condition of the subjects before and during the experiments, we identified 14 factors affecting the measurement and reproducibility of CVR.

Physiological factors

Sex. The sex of the subjects was reported in 26 out of the 32 studies with a male to female ratio of 243 to 118 (male accounting for 67%), and the sex of the remaining 66 subjects was not reported. Kassner *et al.*, compared the CVR levels between male and female healthy volunteers using BOLD MRI and hypercapnia gas challenge as the vasodilator, and found that the mean grey matter (GM) and white matter (WM) CVR of men were both significantly higher than women (GM CVR = 0.181 vs 0.140%/mmHg; WM CVR = 0.096 vs 0.080 mm/Hg for male and female, respectively).[23] In a study including 166 healthy volunteers aged between 59 and 73 years, Spencer *et al.*, investigated the influence of sex on CVR values and reproducibility measured by TCD and hypercapnia gas challenge.[27] The group mean wsCV metric showed that CVR reproducibility of male subjects was marginally higher (wsCV = 11.9% for male vs 12.2% for female, no statistical tests reported). However, the results were the opposite when using the group mean ICC as the reproducibility metric (ICC = 0.86 for male vs 0.88 for female, no statistical tests reported).

Menopause. In a study that investigated the CVR induced by gas inhalation hypercapnia challenge using TCD, all female participants (N = 88, age = 65 ± 6 years) were post-menopause at the time of the experiment.[27] Comparing the CVR levels computed using the ratio between the flow velocity of MCA and the change in CO_2 concentration, the CVR of post-menopausal female subjects was significantly higher (17%) than male subjects. However, no significant differences in CVR reproducibility were observed between female and male subjects (mean wsCV = 12.0% and

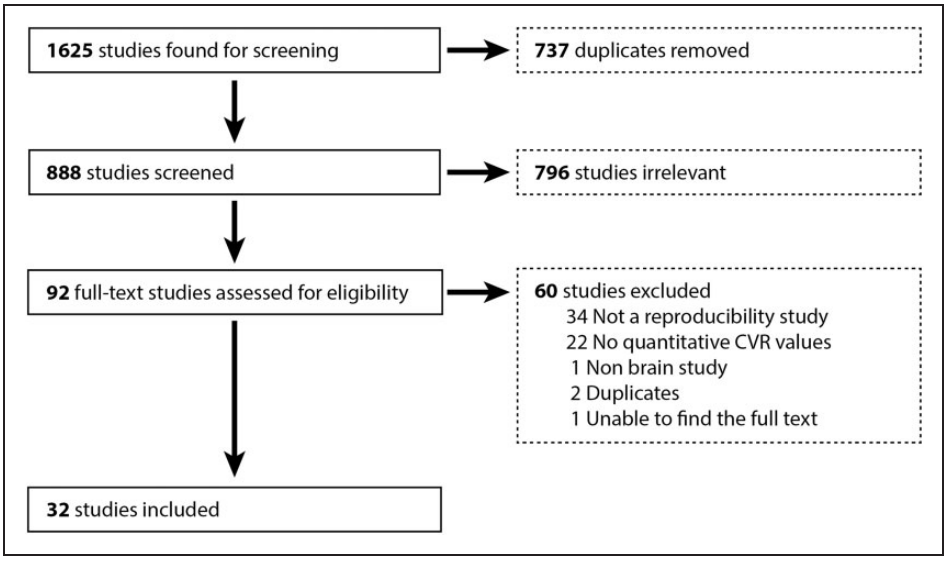

Figure 1. Search and screening results from three databases PubMed, Embase, and Scopus. After removing the duplicated papers, the title, abstract, and full text of each study was assessed using the inclusion and exclusion criteria.

Table 2. CVR studies included in this review.

Study	Population	Imaging Modality	Reproducibility Metric
Vasodilator: Acetazolamide			
Fülesdi et al.[74]	N = 10 no age or sex info	TCD	ICC
Spilt et al.[47]	N = 8 18–26 years; 8 males	PC MRI (1.5 T)	wsCV
Yen et al.[28]	N = 12 23–71 years; 9 males	ASL MRI (1.5 T)	Ratio between difference and mean CVR
Kimura et al.[75]	N = 42 58–62 years; 28 males	SPECT	
Puig et al.[29]	N = 10 21–28 years; 10 males	O15-water PET ASL MRI (3 T)	wsCV
Vasodilator: Breath hold			
Magon et al.[65]	N = 11 20–42 years; 5 males	BOLD MRI (3 T)	wsCV
Bright and Murphy[63]	N = 12 No age info; 7 males	BOLD MRI (3 T)	ICC
Lipp et al.[64]	N = 15 21–28 years; 8 males	BOLD MRI (3 T)	wsCV & ICC
Pinto et al.[48]	N = 10 no age info; 5 males	BOLD MRI (3 T)	wsCV & ICC
Cohen et al.[38]	N = 10 20–50 years; 6 males	ASL & BOLD MRI (3 T)	ICC
Peng et al.[30]	N = 9 21–29 years; 3 males	BOLD MRI (1.5 T & 3 T)	wsCV & ICC
Vasodilator: Hypercapnic gas inhalation			
Mayberg et al.[31]	N = 10 no age or sex info	TCD	
Smielewski et al.[42]	N = 10 44–83 years; no sex info	NIRS	Normalized SD
Totaro et al.[41]	N = 45 20–74 years; 21 males	TCD	ICC
Leontiev et al.[45]	N = 10 24–40 years; 5 males	ASL & BOLD MRI (3 T)	wsCV
Goode et al.[21]	N = 8 19–34 years; no sex info	BOLD MRI (1.5 T)	wsCV
van Beek et al.[32]	N = 11 no age info; 9 males	TCD	wsCV & Signed rank test
Kassner et al.[23]	N = 19 21–42 years; 10 males	BOLD MRI (1.5 T)	wsCV
Ances et al.[76]	N = 16 no age info; 12 males	ASL & BOLD MRI (3 T)	ICC
McDonnell et al.[37]	N = 10 20–46 years; no sex info	TCD	ICC
Sousa et al.[40]	N = 9 18–28 years; 4 males	BOLD MRI (3 T)	wsCV & ICC
Heijtel et al.[43]	N = 16 20–24 years; 9 males	O15-water PET ASL MRI (3 T)	wsCV
Strohm et al.[33]	N = 10 22–36 years; 6 males	TCD	ICC
Tancredi et al.[44]	N = 8 22–40 years; no sex info	ASL & BOLD MRI (3 T)	wsCV
Spencer et al.[27]	N = 166 no age info; 78 males	TCD	wsCV & ICC
Lajoie et al.[36]	N = 8 no age info; 4 males	ASL & BOLD MRI (3 T)	wsCV

(continued)

Table 2. Continued.

Study	Population	Imaging Modality	Reproducibility Metric
Sobczyk et al.[16]	N = 15 20–53 years; 15 males	BOLD MRI (3 T)	wsCV
Ravi et al.[46]	N = 10 no age info; 4 males	BOLD MRI (3 T)	wsCV
Thrippleton et al.[39]	N = 15 20–50 years; 11 males	BOLD MRI (1.5 T)	wsCV
Dengel et al.[35]	N = 11 18–40 years; 5 males	BOLD MRI (3 T)	wsCV & ICC
Hou et al.[77]	N = 10 no age info, 3 males	BOLD MRI (3 T)	Linear regression of CVR Between sessions
Evanoff et al.[34]	N = 11 no age info, 5 males	BOLD MRI (3 T)	wsCV & ICC

12.2%; ICC = 0.89 and 0.87 for female and male subjects, respectively).[27]

Effects of caffeine. Among the 32 selected studies in Table 2, 13 restricted the subjects from consuming caffeine before conducting the experiment, accounting for 69% of the total subjects (293 out of 427). Eight studies (15% of all subjects) required participants to refrain from caffeine products at least 8 hours (maximum 12 hours) before the study[28–35] while the other 5 studied required less than 8 hours with a minimum of just 2 hours.[16,27,36–38] According to the correlation between the mean wsCV and the number of hours without consuming caffeine in the 13 studies that restricted caffeine consumption, a higher reproducibility (lower wsCV value) was associated with a longer period of time for subjects to refrain from caffeine ($r = -0.42$, $p < 0.05$). No significant correlation was obtained between the reproducibility measured by ICC and the time for the subject spent without consuming products with caffeine.

Effects of smoking. Five studies restricted smokers from participating in the study, and the number of non-smokers accounted for 50% (212 out of 427) of the total subjects.[16,27,31,32,37] The CVR reproducibility of these subjects was between 5.97% and 37.3% when measured by wsCV and between 0.429 and 0.891 when measured by ICC, both within the range of values of the other 27 studies that did not restrict smoking. However, no direct comparison was made to assess the CVR levels and the reproducibility between smokers and non-smokers.

Brain regions. A total of 19 studies investigated the regional CVR levels and reproducibility of different brain regions as well as the full brain, accounting for 50% of all subjects (212 out of 427). Four studies compared the GM and WM CVR measured by BOLD MRI (3 T) using a gas inhalation hypercapnia challenge,[16,21,23,39] and the CVR reproducibility of GM was significantly higher than WM in all subjects. For example, in a study that investigated the CVR of 19 healthy subjects using a controlled respiratory challenge, Kassner *et al.* demonstrated that CVR was more reproducible in GM than WM for both within-day (ICC = 0.92 and 0.88 for GM and WM, respectively) and between-day (ICC = 0.81 and 0.66 for GM and WM, respectively) comparisons.[23] Similar results have been reported in a study that measured the CVR of 8 healthy subjects in two sessions using a 1.5 T MRI and hypercapnic gas challenge. In this work, the CVR reproducibility in GM was higher than in WM for the BOLD data analyzed using both box-car and automated regression models.[21]

Two studies investigated CVR in cortical and subcortical regions using BOLD MRI and hypercapnia gas as the vasodilator. For example, in a study that investigated the CVR of 9 healthy volunteers in a paced deep breathing task, CVR measurements in the frontal lobe were found to be the most reproducible while the CVR in the cerebellum was the least reproducible (ICC = 0.90 vs 0.42 for frontal lobe and cerebellum, respectively).[40] Using BOLD MRI, Lajoie *et al.* demonstrated that the CVR in the anterior cingulate was found the be the most reproducible while the CVR in the precuneus was the least reproducible (wsCV = 24% and 35% for anterior cingulate and precuneus, respectively).[36]

Technological factors

Time between repeated experiments. The time between repeated CVR measurements of 14 of the 32 studies was shorter than or equal to a week while 12 were longer than a week; 4 studies included reproducibility measurements for both time intervals, while 2 did not provide any information. For all imaging techniques, a

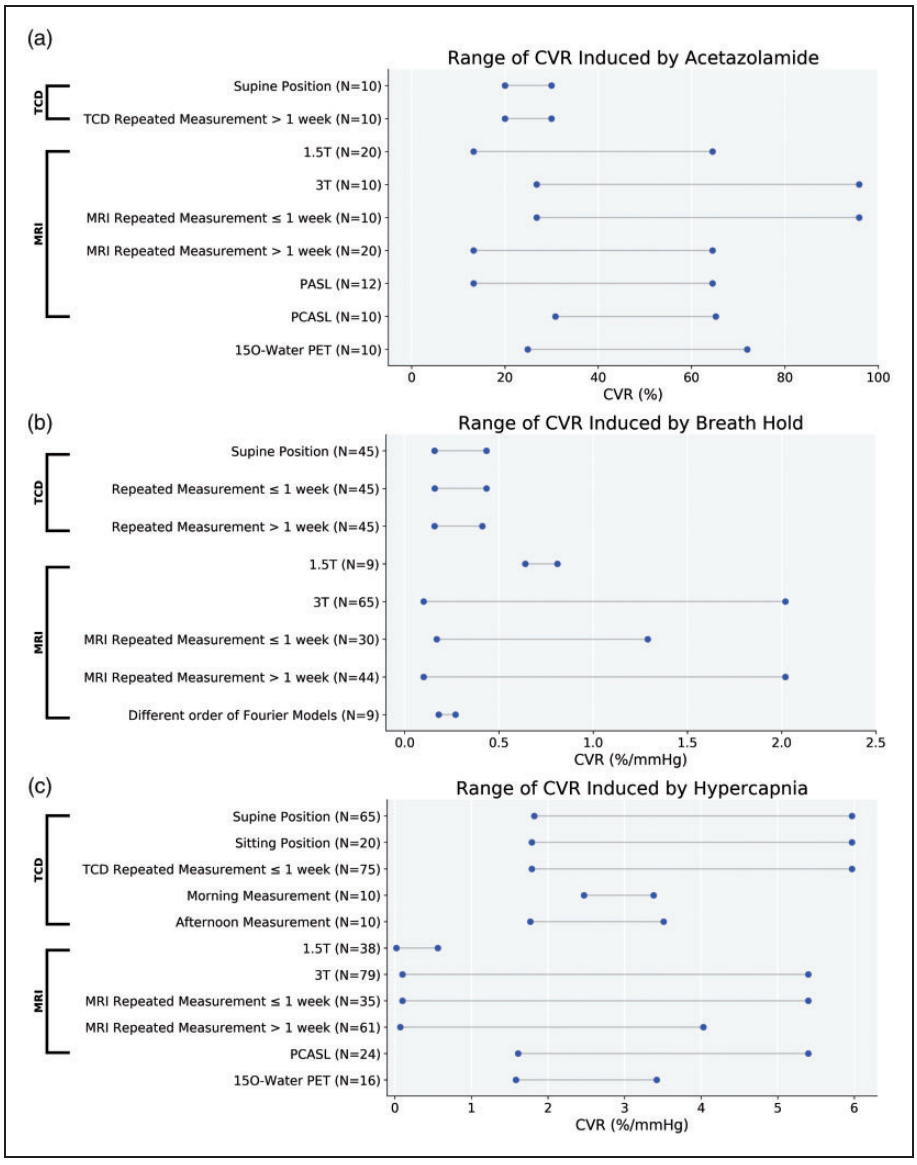

Figure 2. The range of CVR values using different modalities and vasodilators of all selected studies. (A) The range of CVR induced by ACZ was the largest among the three vasodilators. (B) The lowest wsCV was found in CVR induced by BH. (C) The range of CVR using PET and inhaling hypercapnia gas challenge was the lowest among all methods. Modalities that did not appear in any studies (N = 0) were excluded in the figure.

longer time between repeated measurements led to lower reproducibility. For example, comparing the reproducibility of CVR induced by breath-hold and hypercapnic gas challenge using TCD in 45 healthy volunteers (aged between 20 and 74 years), the ICC of CVR induced by BH was reduced by 59% (from 0.41 to 0.17) when the time interval increased from 1 hour to 24 hours.[41] A similar observation was found for the reproducibility of CVR induced by a hypercapnic gas challenge, in which the ICC reduced from 0.55 to 0.43 when the measurements were repeated after 24 hours.[41] In a study that compared the CVR reproducibility of 19 healthy volunteers using BOLD MRI at 5 and 15 days, the wsCV of both GM and WM CVR increased by 62% and 57%, respectively, at the longer time interval (Figure 3(c)).[23] Results from a CVR study on carotid artery occlusion patients also demonstrated that longer time between repeated sessions led to lower reproducibility as measured by normalized SD (reduced by 14.3% as the time between measurements increased from 10 min to 3 days).[42] These data indicated that the time between measurements is an important factor to assess reproducibility.

Time of day. One studied compared CVR of 10 healthy subjects measured at different times of the day using

TCD.[33] In this work, the CVR experiments were conducted in the mornings (8–10 am) and afternoons (5–7 pm) of two consecutive days using hypercapnic gas challenge for 2 to 4 minutes. For both sessions, CVR measured in the afternoon was significantly lower than in the morning (mean CVR = 3.30 and 2.65%/mmHg for the morning and afternoon of day 1; mean CVR = 3.05 and 2.54%/mmHg for the morning and afternoon of day 2; Figures 2(c) and 4(c)). There was, however, no change in reproducibility depending on the time of the day (ICC = 0.65 and 0.68 for morning and afternoon measurements respectively).

Imaging modality. Two studies compared the CVR of 26 healthy subjects measured by different imaging modalities using PET and MRI with ACZ and hypercapnia gas as the vasodilators.[29,43] In general, PET was the more reproducible measurement technique if the CVR was induced by the inhalation of hypercapnia gas (Figure 3(c), indicated by *) while it was less reproducible than MRI if the CVR was induced by ACZ (Figure 3(a)). Heijtel *et al.* compared the test-retest CVR of 16 healthy volunteers (age: 20–24 years, 9 males) using PET and ASL MRI with the subjects inhaling 5% CO_2 gas.[43] In two sessions separated by 34 days, the reproducibility of PET was higher than ASL MRI for both CVR and CBF measurement, as measured by the ICC. By contrast, using simultaneous PET/MRI that allows the PET and MRI data to be collected at the same time, Puig *et al.* demonstrated that ASL MRI had higher CVR reproducibility induced by ACZ than PET based on the wsCV of 10 healthy men (age: 21–28 years).[29] This study also showed that the CVR measured by ASL was significantly lower than the PET CVR by 3% to 12% in different regions of the brain.

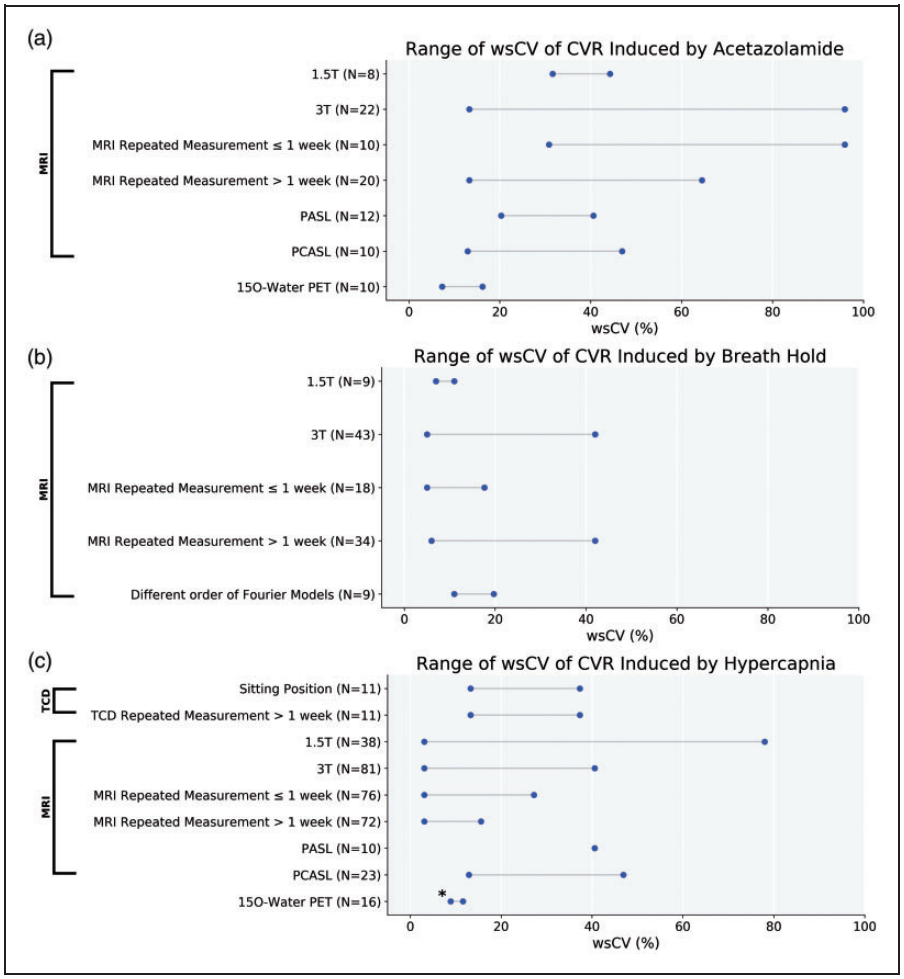

Figure 3. The range of wsCV using different modalities and vasodilators of all selected studies. (A) The range of wsCV of CVR induced by ACZ was the largest among the three vasodilators. (B) the wsCV of CVR induced by BH showed the lowest range. (C) The hypercapnia gas challenge technique was adopted by the highest number of studies among all methods. Overall, PET with hypercapnia gas challenge is the most reproducible technique (indicated by * in the subplot C). Modalities that did not appear in any studies (N = 0) were excluded in the figure.

Subject's posture in TCD. Among the 7 studies that used TCD for CVR measurement, 3 studies adopted the supine posture; 2 studies were performed using the sitting posture; 2 studies used both positions. As shown in Figure 2(a), the CVR ranged from 1.79% to 48% when CVR was induced by gas inhalation hypercapnia challenge and from 20% to 30% when ACZ was used. Overall, the reproducibility metrics of both wsCV and ICC (Figures 3(c), 4(a), and 4(c)) from these studies suggested that all subjects should remain in the same posture for a reliable CVR measurement using TCD and that the sitting position resulted in higher reproducibility than the supine position. CVR measured in the sitting position demonstrated significantly higher reproducibility than the value measured in the supine position (ICC = 0.822 vs 0.734 for sitting and supine positions, respectively [$p < 0.001$]) while no significant CVR differences were observed.[37] In a study that investigated the impact of changing positions during the repeated CVR experiments, different postures of the subjects could significantly affect the repeatability of TCD-based CVR measurement when subjects were examined in different positions (supine or sitting) in repeated sessions.[31]

Rater's bias for TCD. In measuring CVR using TCD, a well-trained rater (or observer) is needed to interpret the flow velocity data of the arteries of interest, making the bias of the rater a potential factor for reproducibility. McDonnell et al. investigated the influence of different raters in the CVR levels of the healthy control subjects examined in supine and sitting positions.[37] The results indicated that the inter-rater reproducibility assessed by ICC was significantly higher when the subjects were in the sitting position (ICC = 0.504 vs 0.081 for sitting and supine respectively), indicating that the CVR test should be conducted in the sitting position if the level of experience of the rater is low.

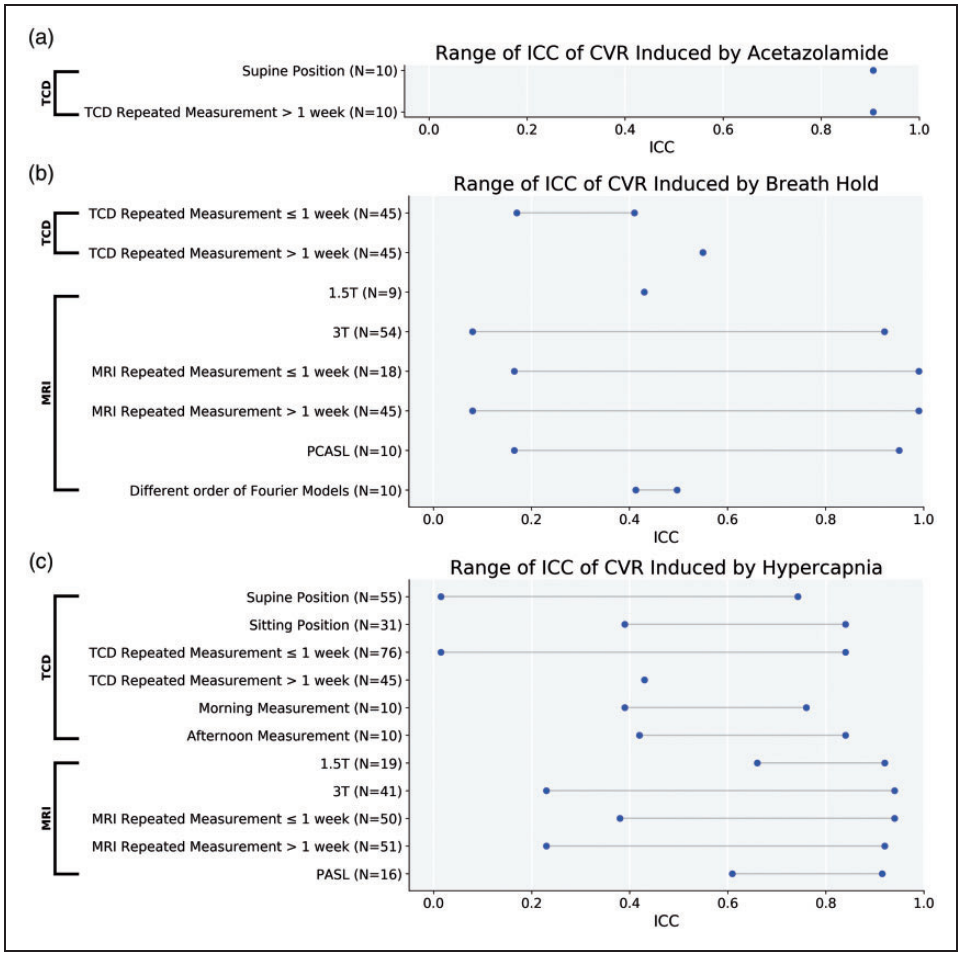

Figure 4. The range of ICC using different modalities and vasodilators of all selected studies. (A) Only one TCD study employed ICC as the metric for reproducibility. (B) The ICC of CVR was the highest using 3 T MRI with the measurement repeated within a week. (C) The range of ICC was the largest using inhaling hypercapnia gas as the vasodilator. Modalities that did not appear in any studies (N = 0) were excluded in the figure.

MRI field strength. Among 23 MRI studies, 6 studies were performed using 1.5 T MRI systems; 18 studies used 3 T; 1 study used both 1.5 T and 3 T. Using a gas inhalation hypercapnia challenge at 1.5 T, CVR values in 3 studies ranged between 0.02%/mmHg and 0.56%/mmHg (Figure 2(c)).[23,39,44] The range was between 0.1%/mmHg and 5.4%mmHg in 7 studies performed using 3 T MRI systems and a hypercapnia challenge for BOLD and ASL MRI (Figure 2 (c)).[16,34,40,43–46] Using ACZ as the vasodilator, the spectrum of the CVR values measured at 1.5 T from 20 healthy subjects in 2 studies was between 13.3–64.5%;[28,47] at 3 T, the range was 26.8–95.9% for 10 subjects.[29] According to the healthy cohort data (N = 9) from a study that performed breath-hold, the CVR reproducibility measured using a 3 T MRI system was 29% higher than that measured using a 1.5 T MRI for short term measurements within 30 minutes, while the reproducibility was essentially the same (wsCV = 5% at 3 T vs 7% at 1.5 T).[30] The same study demonstrated that reproducibility for both field strengths was worse when the time between repeated scans was longer (68–92 days), but was not significantly different at the 2 field strengths (wsCV = 11% at 3 T vs 12% at 1.5 T, $p > 0.05$; ICC = 0.79 at 3 T vs 0.71 at 1.5 T, $p > 0.05$).[30]

ASL techniques. Among the studies that used MRI in CVR quantification, pulsed arterial spin labeling (PASL) techniques were used in 2 studies and pseudo-continuous arterial spin labeling (PCASL) in 5 studies. The range of CVR measured by PCASL (N = 24) was 1.71%/mmHg and 5.4%/mmHg using gas inhalation hypercapnia challenge (Figure 2 (c)).[43,44] For ACZ-induced CVR values, the range was 13.3% and 64.5% for PASL and 30.8% and 65.2% for PCASL.[28,29] Although the range of wsCV of repeated CVR measurements differed between the two ASL techniques (wsCV between 12.9% and 46.9% for PCASL and between 23.0% and 40.6% for PASL),[28,29,36,43,45] no direct comparison of CVR reproducibility between PASL and PCASL was reported in the selected studies. One study measured CVR using percentage of signal change induced by vasodilator, instead of percentage of absolute CBF change.[38] The result (78.3 ± 27.4% signal increase) of this work was excluded from Figure 2 due to inconsistent units.

BOLD MRI techniques. Among 18 studies that used BOLD MRI, one study investigated the impact of BOLD MRI sequence design on the CVR quantification and reproducibility of 10 healthy volunteers (age 21–37 years, 4 males).[46] In this work, three BOLD protocols were compared using different combinations of scanning parameters (TR/TE = 1500/30 ms vs 1500/21 ms vs 800/22.5 ms; FA = 60°, 89°, 72°, respectively) on the CVR induced by inhaling hypercapnia gas for 50 seconds. The results showed that the BOLD technique with the longest TR/TE (1500/30 ms) and lowest FA (60°) produced the highest CVR values, with a mean CVR greater than 0.2%/mmHg and significantly higher than the CVR measured by the other two sequences. However, the reproducibility of the three sequences was not significantly different when assessed using wsCV of the repeated measurements.

BOLD analysis models. The impact of using different models to quantify BOLD signal change on CVR quantification and reproducibility was investigated by 2 studies in 19 healthy subjects.[21,48] Among the various BOLD data analysis techniques, the general linear model with different Fourier orders has been widely applied to estimate BOLD response.[49–51] Pinto et al. compared the impact of using different orders of the Fourier model (from 0th to the 4th order) on GM CVR induced by BH for 20 seconds.[48] CVR increased significantly with the order of the Fourier model (from 0.18 to 0.27%/mmHg, Figure 2(b)). CVR reproducibility, however, only improved up to the 3rd order of the Fourier model (ICC increased from 0.54 to 0.75, Figure 4(b)) but deteriorated when using the 4th order of the Fourier model (ICC = 0.70). In a study that investigated the CVR of 10 subjects using hypercapnia gas challenge and BOLD data collected from a 1.5 T MRI system, Goode *et al.* demonstrated that CVR derived using percent signal intensity changes of the BOLD data showed good reproducibility (mean wsCV = 7.4%, Figure 3(c)).[21] Comparing between the CVR measured by BOLD and TCD, we have identified 4 studied that applied the computer-controlled system and the CVR measurement modalities included TCD and BOLD MRI.[16,33,34,36] The range of ICC of the repeated CVR measurements in these 4 studies was between 0.30 and 0.84 while the range of ICC of the other studies was between 0.015 and 0.92. The source of data in Figures 2 to 4 can be found in Tables 3 to 5 respectively.

Discussion

In this work, we reviewed multiple studies that investigated the reproducibility of CVR using different imaging modalities and various vasodilators. This systematic review compared the reproducibility of different CVR techniques, including TCD, MRI, PET, and SPECT. We identified 32 articles from 3 databases with a total of 427 subjects. Among the factors affecting the reproducibility of CVR, the physiological factors associated with the experimental subject included sex, consuming caffeine before CVR studies, and

smoking; the technological factors associated with the design and implementation of the CVR test included: type of imaging modality, subject's posture during TCD experiments, and data analysis method. The primary findings of this paper were: (1) consuming caffeine before had a significant impact on CVR and its reproducibility and at least 12 hours of caffeine restriction is recommended; (2) the sitting posture should be used for CVR studies using TCD and data should be interpreted by the same rater; (3) the scanning parameters and analysis method must be taken into account when evaluating CVR studies using BOLD MRI.

Physiological factors

The CVR difference between male and female subjects was investigated in two studies that used BOLD MRI and TCD with hypercapnia gas challenge as the vasodilator.[23,27] The between-gender differences may be partially explained by the previous observation that the different levels of total hemoglobin in male and female are linked to a 20% greater BOLD signal change associated with functional activation in

Table 3. Studies referenced in Figure 2 of the main manuscript.

Conditions	References
(A) CVR induced by acetazolamide	
TCD supine position	74
TCD repeated measurement >1 week	74
1.5 T	28,47
3 T	29
MRI repeated measurement ≤1 week	29
MRI repeated measurement >1 week	28,47
PASL	28
PCASL	29
15 O-water PET	29
(B) CVR induced by breath hold	
TCD supine position	41
TCD repeated measurement ≤1 week	41
TCD repeated measurement >1 week	41
1.5 T	30
3 T	30,38, 48,63–65
MRI repeated measurement ≤1 week	30,48,63
MRI repeated measurement >1 week	30, 38,64,65
Different order of Fourier models	48
(C) CVR induced by hypercapnia	
TCD supine position	31,37,41
TCD sitting position	33,37
TCD repeated measurement ≤1 week	31,33,37,41
Morning measurement	33
Afternoon measurement	33
1.5 T	23,39,44
3 T	16,34,40, 43–46
MRI repeated measurement ≤1 week	40,44,46
MRI repeated measurement >1 week	16,23,34,43
PCASL	43,44
15 O-Water PET	43

Table 4. Studies referenced in Figure 3 of the main manuscript.

Conditions	References
(A) wsCV of CVR induced by acetazolamide	
1.5T	47
3 T	28,29
MRI repeated measurement ≤1 week	29
MRI repeated measurement >1 week	28,47
PASL	28
PCASL	29
15 O-water PET	29
(B) wsCV of CVR induced by breath hold	
1.5T	30,
3 T	30,48,64,65
MRI repeated measurement ≤1 week	30,48,
MRI repeated measurement >1 week	30,64,65
Different order of Fourier models	48
(C) wsCV of CVR induced by hypercapnia	
TCD sitting position	32
TCD repeated measurement >1 week	32
1.5 T	21,23,39
3 T	16,34,35,43–46
MRI repeated measurement ≤1 week	23,34,35,40,44,46
MRI repeated measurement >1 week	16,23,34,35,43
PASL	45
PCASL	36
15 O-water PET	43

Table 5. Studies referenced in Figure 4 of the main manuscript.

Conditions	References
(A) ICC of CVR induced by acetazolamide	
TCD supine position	74
TCD repeated measurement >1 week	74
(B) ICC of CVR induced by breath hold	
TCD repeated measurement ≤1 week	41
TCD repeated measurement >1 week	41
1.5 T	30
3 T	30,38,48,64,63
MRI repeated measurement ≤1 week	30
MRI repeated measurement >1 week	30,38,63,64
PCASL	38
Different order of Fourier models	48
(C) ICC of CVR induced by hypercapnia	
TCD supine position	37,41
TCD sitting position	32,33,37
TCD repeated measurement ≤1 week	32,33,37,41
TCD repeated measurement >1 week	41
TCD morning measurement	33
TCD afternoon measurement	33
1.5 T	23
3 T	34,35,40,76
MRI repeated measurement ≤1 week	23,34,35,40
MRI repeated measurement >1 week	34,35,76
PASL	76

males.[52] Such CVR differences may also help identify factors that lead to a higher risk of migraine and lower incidence of stroke in females.[53] Although one study showed that CVR reproducibility of men was slightly lower than women, no studies have shown statistically significant differences. Therefore, while CVR levels may differ between the sexes, there does not appear to be an effect on reproducibility.[27]

Since caffeine and smoking are both vasoactive substances, we examined their impact on CVR quantification and reproducibility based on the duration that the subject was asked to refrain from caffeine and smoking. Higher reproducibility was correlated with a longer period of caffeine restriction (between 2 and 8 hours in 5 studies), implying that it is essential for experimental participants not to consume caffeine for the same or longer period of time before CVR studies. Another important consideration is the different impact on CBF and CVR due to acute and regular caffeine consumption. It has been shown that the short-term effect of caffeine on blood flow was more pronounced in individuals consuming a low dose (less than 200 mg/day) of caffeine.[54] Since caffeine has been shown to downregulate vascular adenosine receptors, CBF tends to normalize in regular caffeine users.[55] According to a study that applied PET imaging to measure CBF, the baseline CBF prior to caffeine consumption had a strong influence on the magnitude of the caffeine-induced variation.[56] In studies that applied BOLD MRI for CBF and CVR measurements, BOLD signal was significantly correlated with caffeine consumption and the BOLD signal change in visual cortex was significantly greater in subjects with high caffeine consumption (greater than 300 mg/day).[57] Therefore, acute caffeine consumption of large quantities should be strictly restricted in CVR measurements. The influence of smoking on CVR was challenging to assess directly due to the lack of a systematic comparison of CVR between smokers and non-smokers. Thus, future work is needed to evaluate the role of smoking and other lifestyle factors on CVR measurement.

Technological factors

TCD is an ultrasound technique that provides a global CVR measurement by quantifying blood flow velocities in the intracranial vessels. TCD is widely available and inexpensive but provides limited information about brain parenchymal perfusion.[58] An important assumption of TCD measurements is that the arterial blood vessel diameter remains unchanged during the challenge.[59] Since flow is equal to velocity (measured) times area (not measured), changes in CBF due to alterations of vessel cross sectional area may make it difficult to compare TCD CVR measurements with modalities that focus on CBF. The administration of hypercapnia gas has in fact been shown to increase vessel diameter in several studies, which potentially violates this assumption.[60–62] One TCD study demonstrated better CVR reproducibility in the sitting position.[37] In the same study, the data also suggested that inter-rater bias was also the lowest for the sitting position. Therefore, the most reproducible CVR procedure using TCD should be conducted in the sitting position and the data examined by a single rater, assuming this is feasible.

BOLD is a commonly used MRI technique to measure CVR due to its relatively high sensitivity to cerebral hemodynamic MRI signals induced by a vasodilator. The CVR measurement of BOLD MRI depends on both the implementation of the BOLD sequence and the data analysis technique. Among the studies that applied BOLD MRI with a single-echo, the TE ranged from 21 to 60 ms with values between 30 and 40 ms being the mostly widely used.[16,23,35,63–65] It should be noted that the TE values chosen are related to the T2* of gray matter, which depends on the field strength of the static magnetic field (1.5 T and 3 T). In the study that compared the CVR using different BOLD scanning parameters, the sequence with a long TR/TE and low FA achieved the highest CVR levels induced by inhaling hypercapnia gas.[46] The reproducibility of the different implementations of the BOLD sequence, however, did not show significant differences. These results suggested that when evaluating the reported CVR values measured by BOLD it is important to assess them in the context of the imaging parameters used, particularly TR, TE, and FA. The analysis method can also introduce variations in BOLD CVR values when different models are used. Although the general linear model is the most widely applied technique to detect the signal changes in BOLD data, the model can be implemented with different degrees of complexity.[51] In a study that investigated the modeling techniques for BOLD responses to BH tasks in CVR studies, the data showed that Fourier series were suitable for modeling BOLD signal response and that higher degrees of the Fourier model to analyze BOLD MRI data were found to improve the CVR reproducibility, but only up to the 3rd order of the model.[48] Beyond that threshold, more complex models tend to overfit the data, leading to reduced robustness. The inclusion of sinusoidal components in the Fourier model was found to better explain the BOLD data than the convolution between task timing and hemodynamic response function.[66] It also showed advantages to accommodate incorrect task performance in non-compliant subjects, thus making it a more robust model for processing BOLD data of patient.[63] The Fourier model approach should be

equally applicable to other types of experimental designs (such as end-expiration BH tasks) of CVR experiments using BOLD MRI together with a well-designed experiment and patient's compliance to ensure highly reproducible CVR results. For CVR studies using BOLD MRI, we can conclude that the analysis model is the dominant factor affecting the CVR reproducibility for data collected using the same technique, while the specific implementation of the BOLD acquisition had a higher influence on the actual CVR values than the reproducibility of the measurement.

ASL is a quantitative MRI technique that enables the measurement of CBF using magnetically labeled blood water as the endogenous tracer. The primary difference between the various ASL techniques is how the spins of the arterial blood are inverted. Different implementations such as pulsed or pseudo-continuous ASL will result in different SNR. Although no direct comparison of CVR has been made using the two ASL techniques, the range of wsCV values showed a high overlap among the five studies (12.9–46.9% for PCASL and 23.0–40.6% for PASL).[28,29,36,43,45] Although PCASL is the recommended approach for CBF measurement at a resting condition, an open question is whether this technique is also the most reproducible for CVR measurement that requires CBF quantification at both rest and stressed conditions.

Simultaneous PET/MRI enables the direct comparison of CVR measured by PET and MRI at the same time. According to a study that measured CVR induced by ACZ using PET/MRI, ASL MRI was the more reproducible modality for the CVR quantification of healthy males, although the CBF agreement before and after the administration of ACZ was poor between the two modalities.[29] CVR measured by PET may have been more variable than MRI due to its sensitivity to the attenuation bias and the choice of analysis models, while the ASL MRI data were collected using an advanced sequence that accounted for the potential errors due to arterial transit time. The attenuation correction was achieved using the structural data from a separate CT scan that was previously developed only for a specific group of patients instead of a more general data-driven approach, which potentially reduced the accuracy of the CVR measurement using PET.[67,68] Therefore, the data of this study implied that despite ^{15}O-water PET being the gold standard for CBF quantification, ASL MRI may be the more reproducible technique for CVR measurement.

In both healthy subjects and patients, a higher reproducibility is associated with a shorter gap between repeated measurement for studies using TCD and MRI.[23,41,42] This is consistent with the notion that many physiological markers remain relatively stable during a short period of time in a normal external environment. In this review, we set a week as the threshold for the short and long time between repeated experiments based on the distribution of the time in the selected literature. A longer gap between repeated studies may lead to reduced reproducibility due to subtle changes in the response to the vasodilator for different physiological conditions, which depends on multiple factors including cognition, heart rate, and blood pressure.[69] In the selected studies, however, these data were unavailable and the investigation on the exact response to different vasodilators was also limited. There is data indicating no significant differences between the reproducibility of CVR measured in the morning and afternoon.[33] But this same study showed differences in CVR itself based on time of day. Therefore, a rule of thumb for selecting the time of CVR study is that the time between repeated measurements should be as consistent as possible for all subjects in the same experimental condition while the actual time of the day to perform the CVR measurement is less important as long as it is consistent. Of course, this consistency is nearly impossible to ensure in clinical studies, so ranges of CVR based on time of day may be required to develop normal ranges in order to detect significant CVR changes, either between groups or within individuals in longitudinal studies.

Among the reviewed studies, three types of vasodilators were applied to induce the change in CBF: ACZ, inhaling hypercapnic gas, and BH. In considering the impact of different breath-hold duration (9, 15, and 21 seconds) on CVR measurement and reproducibility, the results revealed that a long BH was associated with both high CVR levels and reproducibility.[65] This can be explained by the higher arterial CO_2 concentrations during a longer BH that increased the BOLD signal and the resulting CVR. Another important observation was that a longer BH duration reduced the errors in CVR computation caused by the early negative CBF change. Without a long enough BH duration, such a subtle CBF decrease could mask the actual autoregulation response induced by the accumulation of CO_2 concentration. However, one concern about long BH is that many patients cannot comply with BH instructions.[22] As the mostly widely used vasodilation mechanism, hypercapnia gas inhalation challenge was applied in 21 out of 32 studies. The application of the gas-inhalation techniques differed in the composition of the gas mixture, concentration of CO_2 in the gas mixture, and delivery method. In terms of the concentration of CO_2 in the hypercapnia condition, it ranged between 4% and 10% with 5% being the mostly frequently applied in 11 studies. For the studies using 4% to 7% concentration of CO_2, we observed that a higher

concentration led to larger CVR measurements and lower wsCV values. However, this trend discontinued at 10% concentration of CO_2, where the wsCV of the whole brain was 25.8% higher (or less reproducible) than the wsCV (17.5%) of the study that used 7% CO_2.[21] The delivery method of the gas mixture was another factor that caused different CVR measurements and reproducibility, and the primary considerations was the duration of the delivery of hypercapnia gas, which ranged from 30 seconds to 5 minutes in the studies that we reviewed. Considering the concentration and the delivery method together, we found that keeping the concentration of the CO2 lower than 7% and delivering for between 2 and 4 minutes yielded the highest CVR reproducibility. The facility of the gas delivery may also affect the outcome of CVR studies due to the compatibility with the imaging modality and comfort with the different positions of the experimental subjects. A previous technical review outlined 3 commonly used gas delivery apparatus.[70] However, no data were reported that compared the different gas delivery facilities in relation to the reproducibility of the CVR measurements. Comparing the CVR measurements induced by hypercapnic-gas challenge for 2 minutes and BH for as long as the subject can tolerate, the reproducibility of using hypercapnic-gas challenge was higher than BH in both same day and different day (ICC = 0.55 vs 0.41 for same day; ICC = 0.43 vs 0.17 for different days).[41] It should be noted that this study was conducted using TCD with 5% CO_2 concentration and the longest gap between the repeated measurements was only 24 hours. Thus, we can conclude that using TCD with hypercapnic-gas challenge should be favored for short-term reproducibility of CVR measurements. In terms of the use of ACZ in CVR measurement, although the exact amount of ACZ administered varied based on subject body weight, no studies were found that investigated the impact of different doses of ACZ on CVR in the same session, likely because it takes a long time for cerebral hemodynamics to return to the resting state after the administration of ACZ.

Future considerations

The absence of reference CVR values of healthy subjects across the age and sex spectra is a major challenge for selecting the most appropriate imaging modality for CVR studies. Among the various CVR methods reviewed in this work, the reliability of recently developed gas-free (breath-hold) methods is worth further investigation in different populations and compared with the conventional acetazolamide- or CO_2-based methods. A set of practical considerations have been outlined in a methodological guide including the duration of the task, training and instructions for the experimental subject, and data analysis methods.[71] These findings will provide insights on the potential of using breath-hold as the endogenous vasodilator in future CVR studies. To reduce the impact of variations, consideration should be given to measuring each participant's hemoglobin levels as part of the CVR measurements using MRI modalities. For PET based CVR studies, it is also important to evaluate different scanning and reconstruction methods as well as the how the availability of PET scanners and cyclotrons might affect the widespread application of PET based CVR measurements. Furthermore, a consensus on the standard CVR data analysis methods should be established to enhance the reproducibility of future prospective studies. Since CVR has the potential to predict the risk of stroke, future studies should also focus on the change in CVR of patients with a high risk of stroke after treatment. The results from these studies could establish reference values to evaluate the effects of treatment on vascular hemodynamics as well as outcomes in the population being studied. The large variability of CVR measurements is one important factor that has hindered the clinical application of CVR for different cerebrovascular diseases. Therefore, more systematic studies investigating the normal range of CVR for different population using different modalities are desired to enable CVR to become an effective biomarker for cerebrovascular diseases.

Limitations

In this work, we identified 3 major types of vasodilators, including hypercapnic gas inhalation, ACZ injection, and BH, based on their effects on expanding blood vessels; each has its advantages and disadvantages. The administration of vasodilators may cause common side effects such as nausea and numbness. For example, one study reported that 44% of the subjects experienced the side effects of numbness and motor weakness in a CVR study using 1 g of ACZ and SPECT imaging.[72] However, the side effects of the vasodilators on CVR reproducibility have not been systematically investigated, in particular on how the side effects influence the successful completion rate of the test-retest CVR experiments. In this context, it is worth documenting the actual success rate of the repeated CVR measurement for each vasodilator used in different imaging modalities. A questionnaire may be employed to evaluate the physical and emotional comfort of the patient after each CVR study, allowing the optimal use of vasodilators for the most comfortable CVR study. It is also valuable to consider the technical requirement of applying different types of

vasodilators, which may affect the repeatability of CVR measurements. For instance, breath-hold is highly dependent on the tolerance of individuals and may not be able to induce the same level of vasodilation as CO_2 inhalation. Similarly, the inter-subject variability in vasodilation induced by ACZ injection remains an open question, which may be addressed by a systematic study to investigate the response to a fixed amount of ACZ (such as 1 g) in normal subjects. Another limitation is the confounding effects from the association between imaging and non-imaging data. Among studies reviewed, no attempts were made to investigate the sensitivity of the CVR measurement to such confounds, which may lead to a systematic bias on the true CVR levels. A possible solution to this challenge is modeling the potential impacts on the confounds using the general linear model employed to resolve similar issues in the UK Biobank data.[73] Future CVR studies could thus be planned based on the potential impact of these confounds.

Conclusions

This review critically assesses the factors affecting CVR reproducibility using different imaging modalities and vasodilators. We identified multiple aspects of both the measurement technique and subject-related factors that can have an influence on both CVR values and the reproducibility of the measurement. Based on the physiological factors reviewed, CVR exams should be conducted at approximately the same time of the day to minimize the fluctuation of physiological factors and subjects should be restricted from consuming caffeine at least 12 hours before the study to ensure a high reproducibility of the CVR measurements. Although it was inconclusive to identify the perfect CVR experiment, we have outlined several recommendations for future CVR studies to improve reproducibility. As for the different modalities, TCD should be performed in sitting position and assessed by the same rater. Long BH duration is preferred to ensure a more accurate CVR measurement and reproducibility. A blood test should be included as part of the routine CVR exam to account for the impact due to variations in hemoglobin and hematocrit levels. Future CVR studies should be cognizant of these issues and designed to minimize their impact. More study is required to evaluate rigorously other potential confounds, including patient age, which has not been extensively studied.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the American Heart Association (Grant: 826254) and National Institutes of Health (Grant R01EB025220-02, 4R00NS102884-03 and 1K99-AG068310-01A1).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

Moss Zhao: Methodology, investigation, writing, funding acquisition
Amanda Woodward: Methodology, data curation.
Audrey Fan: Investigation.
Kevin Chen: Investigation.
David Chen: Investigation.
Yannan Yu: Investigation.
Michael Moseley: Supervision.
Greg Zaharchuk: Conceptualization, supervision, funding acquisition.

ORCID iDs

Moss Y Zhao https://orcid.org/0000-0002-0210-7739
Amanda Woodward https://orcid.org/0000-0002-8815-1369
Kevin T Chen https://orcid.org/0000-0001-8556-5307

Supplemental material

Supplemental material for this article is available online.

References

1. Leung J, Duffin J, Fisher JA, et al. MRI-based cerebrovascular reactivity using transfer function analysis reveals temporal group differences between patients with sickle cell disease and healthy controls. *NeuroImage Clin* 2016; 12: 624–630.
2. Juttukonda MR and Donahue MJ. Neuroimaging of vascular reserve in patients with cerebrovascular diseases. *NeuroImage* 2019; 187: 192–208.
3. DeBaun MR and Kirkham FJ. Central nervous system complications and management in sickle cell disease. *Blood* 2016; 127: 829–838.
4. Zlokovic BV. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci* 2011; 12: 723–738.
5. Kawanabe Y and Nauli SM. Endothelin. *Cell Mol Life Sci CMLS* 2011; 68: 195–203.
6. Murray AD, Staff RT, Shenkin SD, et al. Brain white matter hyperintensities: relative importance of vascular risk factors in nondemented elderly people. *Radiology* 2005; 237: 251–257.
7. Sam K, Peltenburg B, Conklin J, et al. Cerebrovascular reactivity and white matter integrity. *Neurology* 2016; 87: 2333–2339.
8. Silvestrini M, Pasqualetti P, Baruffaldi R, et al. Cerebrovascular reactivity and cognitive decline in patients with Alzheimer disease. *Stroke* 2006; 37: 1010–1015.

9. Barnes JN, Harvey RE, Miller KB, et al. Cerebrovascular reactivity and vascular activation in postmenopausal women with histories of preeclampsia. *Hypertension* 2018; 71: 110–117.
10. Macedo-Campos R de S, Adegoke SA, Figueiredo MS, et al. Cerebral vasoreactivity in children with sickle cell disease: a transcranial doppler study. *J Stroke Cerebrovasc Dis* 2018; 27: 2703–2706.
11. Dandona P, James IM, Newbury PA, et al. Cerebral blood flow in diabetes mellitus: evidence of abnormal cerebrovascular reactivity. *Br Med J* 1978; 2: 325–326.
12. Ogasawara K, Ogawa A, Terasaki K, et al. Use of cerebrovascular reactivity in patients with symptomatic major cerebral artery occlusion to predict 5-year outcome: comparison of xenon-133 and iodine-123-IMP single-photon emission computed tomography. *J Cereb Blood Flow Metab* 2002; 22: 1142–1148.
13. Fan AP, Jahanian H, Holdsworth SJ, et al. Comparison of cerebral blood flow measurement with [15O]-water positron emission tomography and arterial spin labeling magnetic resonance imaging: a systematic review. *J Cereb Blood Flow Metab* 2016; 36: 842–861.
14. Moses WW. Fundamental limits of spatial resolution in PET. *Nucl Instrum Methods Phys Res Sect Accel Spectrometers Detect Assoc Equip* 2011; 648: S236–S240.
15. Zhao MY, Fan AP, Chen DY-T, et al. Cerebrovascular reactivity measurements using simultaneous 15O-water PET and ASL MRI: impacts of arterial transit time, labeling efficiency, and hematocrit. *NeuroImage* 2021; 233: 117955.
16. Sobczyk O, Crawley AP, Poublanc J, et al. Identifying significant changes in cerebrovascular reactivity to carbon dioxide. *Am J Neuroradiol* 2016; 37: 818–824.
17. de Boorder MJ, Hendrikse J and van der Grond J. Phase-contrast magnetic resonance imaging measurements of cerebral autoregulation with a breath-hold challenge: a feasibility study. *Stroke* 2004; 35: 1350–1354.
18. Yezhuvath US, Lewis-Amezcua K, Varghese R, et al. On the assessment of cerebrovascular reactivity using hypercapnia BOLD MRI. *NMR Biomed* 2009; 22: 779–786.
19. Zhao MY, Václavů L, Petersen ET, et al. Quantification of cerebral perfusion and cerebrovascular reserve using turbo-QUASAR arterial spin labeling MRI. *Magn Reson Med* 2020; 83: 731–748.
20. Bonte FJ, Devous MD and Reisch JS. The effect of acetazolamide on regional cerebral blood flow in normal human subjects as measured by single-photon emission computed tomography. *Invest Radiol* 1988; 23: 564–568.
21. Goode SD, Krishan S, Alexakis C, et al. Precision of cerebrovascular reactivity assessment with use of different quantification methods for hypercapnia functional MR imaging. *AJNR Am J Neuroradiol* 2009; 30: 972–977.
22. Jahanian H, Christen T, Moseley ME, et al. Measuring vascular reactivity with resting-state blood oxygenation level-dependent (BOLD) signal fluctuations: a potential alternative to the breath-holding challenge? *J Cereb Blood Flow Metab* 2017; 37: 2526–2538.
23. Kassner A, Winter JD, Poublanc J, et al. Blood-oxygen level dependent MRI measures of cerebrovascular reactivity using a controlled respiratory challenge: reproducibility and gender differences. *J Magn Reson Imaging JMRI* 2010; 31: 298–304.
24. Leung J, Kim JA and Kassner A. Reproducibility of cerebrovascular reactivity measures in children using BOLD MRI. *J Magn Reson Imaging JMRI* 2016; 43: 1191–1195.
25. Catchlove SJ, Parrish TB, Chen Y, et al. Regional cerebrovascular reactivity and cognitive performance in healthy aging. *J Exp Neurosci* 2018; 12: 117906951 8785151.
26. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med* 2009; 6: e1000097.
27. Spencer MD, Tyndall AV, Davenport MH, et al. Cerebrovascular responsiveness to hypercapnia is stable over six months in older adults. *Plos One* 2015; 10: e0143059.
28. Yen YF, Field AS, Martin EM, et al. Test-retest reproducibility of quantitative CBF measurements using FAIR perfusion MRI and acetazolamide challenge. *Magn Reson Med* 2002; 47: 921–928.
29. Puig O, Henriksen OM, Vestergaard MB, et al. Comparison of simultaneous arterial spin labeling MRI and 15O-H2O PET measurements of regional cerebral blood flow in rest and altered perfusion states. *J Cereb Blood Flow Metab* 2019; 40: 1621–1633.
30. Peng S-L, Yang H-C, Chen C-M, et al. Short- and long-term reproducibility of BOLD signal change induced by breath-holding at 1.5 and 3 T. *NMR Biomed* 2020; 33: e4195.
31. Mayberg TS, Lam AM, Matta BF, et al. The variability of cerebrovascular reactivity with posture and time. *J Neurosurg Anesthesiol* 1996; 8: 268–272.
32. Beek AH, van Wit HM, de Rikkert MGO, et al. Incorrect performance of the breath hold method in the old underestimates cerebrovascular reactivity and goes unnoticed without concomitant blood pressure and end-tidal CO2 registration. *J Neuroimaging* 2011; 21: 340–347.
33. Strohm J, Duffin J and Fisher JA. Circadian cerebrovascular reactivity to CO2. *Respir Physiol Neurobiol* 2014; 197: 15–18.
34. Evanoff NG, Mueller BA, Marlatt KL, et al. Reproducibility of a ramping protocol to measure cerebral vascular reactivity using functional magnetic resonance imaging. *Clin Physiol Funct Imaging* 2020; 40: 183–189.
35. Dengel DR, Evanoff NG, Marlatt KL, et al. Reproducibility of blood oxygen level-dependent signal changes with end-tidal carbon dioxide alterations. *Clin Physiol Funct Imaging* 2017; 37: 794–798.
36. Lajoie I, Tancredi FB and Hoge RD. Regional reproducibility of BOLD calibration parameter M, OEF and resting-state CMRO2 measurements with QUO2 MRI. *Plos One* 2016; 11: e0163071.
37. McDonnell MN, Berry NM, Cutting MA, et al. Transcranial doppler ultrasound to assess cerebrovascular reactivity: reliability, reproducibility and effect of posture. *PeerJ* 2013; 1: e65.

38. Cohen AD and Wang Y. Improving the assessment of breath-holding induced cerebral vascular reactivity using a multiband multi-echo ASL/BOLD sequence. *Sci Rep* 2019; 9: 5079.
39. Thrippleton MJ, Shi Y, Blair G, et al. Cerebrovascular reactivity measurement in cerebral small vessel disease: rationale and reproducibility of a protocol for MRI acquisition and image processing. *Int J Stroke* 2017; 13: 195–206.
40. Sousa I, Vilela P and Figueiredo P. Reproducibility of hypocapnic cerebrovascular reactivity measurements using BOLD fMRI in combination with a paced deep breathing task. *NeuroImage* 2014; 98: 31–41.
41. Totaro R, Marini C, Baldassarre M, et al. Cerebrovascular reactivity evaluated by transcranial doppler: reproducibility of different methods. *Cerebrovasc Dis* 1999; 9: 142–145.
42. Smielewski Peter Czosnyka M, Pickard JD and Kirkpatrick P. Clinical evaluation of near-Infrared spectroscopy for testing cerebrovascular reactivity in patients with carotid artery disease. *Stroke* 1997; 28: 331–338.
43. Heijtel DFR, Mutsaerts HJMM, Bakker E, et al. Accuracy and precision of pseudo-continuous arterial spin labeling perfusion during baseline and hypercapnia: a head-to-head comparison with 15O H2O positron emission tomography. *NeuroImage* 2014; 92: 182–192.
44. Tancredi FB, Lajoie I and Hoge RD. 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* 2015; 42: 1144–1157.
45. Leontiev O and Buxton RB. Reproducibility of BOLD, perfusion, and CMRO2 measurements with calibrated-BOLD fMRI. *NeuroImage* 2007; 35: 175–184.
46. Ravi H, Thomas BP, Peng S-L, et al. On the optimization of imaging protocol for the mapping of cerebrovascular reactivity. *J Magn Reson Imaging* 2016; 43: 661–668.
47. Spilt A, Boom RV, den Kamper AM, et al. MR assessment of cerebral vascular response: a comparison of two methods. *J Magn Reson Imaging* 2002; 16: 610–616.
48. Pinto J, Jorge J, Sousa I, et al. Fourier modeling of the BOLD response to a breath-hold task: optimization and reproducibility. *NeuroImage* 2016; 135: 223–231.
49. Woolrich MW, Ripley BD, Brady M, et al. Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* 2001; 14: 1370–1386.
50. Woolrich MW, Behrens TEJ and Smith SM. Constrained linear basis sets for HRF modelling using variational bayes. *NeuroImage* 2004; 21: 1748–1761.
51. Woolrich MW, Behrens TEJ, Beckmann CF, et al. Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage* 2004; 21: 1732–1747.
52. Levin JM, Frederick B de B, Ross MH, et al. Influence of baseline hematocrit and hemodilution on BOLD fMRI activation. *Magn Reson Imaging* 2001; 19: 1055–1062.
53. Sacco S, Ornello R, Ripa P, et al. Migraine and hemorrhagic stroke. *Stroke* 2013; 44: 3032–3038.
54. Addicott MA, Yang LL, Peiffer AM, et al. The effect of daily caffeine use on cerebral blood flow: how much caffeine can we tolerate? *Hum Brain Mapp* 2009; 30: 3102–3114.
55. Sigmon SC, Herning RI, Better W, et al. Caffeine withdrawal, acute effects, tolerance, and absence of net beneficial effects of chronic administration: cerebral blood flow velocity, quantitative EEG and subjective effects. *Psychopharmacology (Berl)* 2009; 204: 573–585.
56. Cameron OG, Modell JG and Hariharan M. Caffeine and human cerebral blood flow: a positron emission tomography study. *Life Sci* 1990; 47: 1141–1146.
57. Laurienti PJ, Field AS, Burdette JH, et al. Dietary caffeine consumption modulates fMRI measures. *NeuroImage* 2002; 17: 751–757.
58. Gupta A, Chazen JL, Hartman M, et al. Cerebrovascular reserve and stroke risk in patients with carotid stenosis or occlusion: a systematic review and meta-analysis. *Stroke* 2012; 43: 2884–2891.
59. Herrera Campos CR, Beltramini GC, Avelar WM, et al. Cerebral vasomotor reactivity assessment using transcranial doppler and MRI with apnea test. *Braz J Med Biol Res* 2016; 49: e5437.
60. Shirai M, Sada K and Ninomiya I. Analysis of diameter and flow velocity changes in small pulmonary vessels during regional alveolar hypercapnia. In: Manabe H, Zweifach BW and Messmer K (eds) *Microcirculation in circulatory disorders*. Japan, Tokyo: Springer, 1988, pp.451–456.
61. Ward ME. Interaction between hypoxia and hypercapnia in regulating canine diaphragm arteriolar diameter. *J Appl Physiol Bethesda Md 1985* 1996; 80: 802–809.
62. de Matthaeis A, Greco A, Dagostino MP, et al. Effects of hypercapnia on peripheral vascular reactivity in elderly patients with acute exacerbation of chronic obstructive pulmonary disease. *Clin Interv Aging* 2014; 9: 871–878.
63. Bright MG and Murphy K. Reliable quantification of BOLD fMRI cerebrovascular reactivity despite poor breath-hold performance. *NeuroImage* 2013; 83: 559–568.
64. Lipp I, Murphy K, Caseras X, et al. Agreement and repeatability of vascular reactivity estimates based on a breath-hold task and a resting state scan. *NeuroImage* 2015; 113: 387–396.
65. Magon S, Basso G, Farace P, et al. Reproducibility of BOLD signal change induced by breath holding. *NeuroImage* 2009; 45: 702–712.
66. Murphy K, Harris AD and Wise RG. Robustly measuring vascular reactivity differences with breath-hold: normalising stimulus-evoked and resting state BOLD fMRI data. *NeuroImage* 2011; 54: 369–379.
67. Chen LT, Vu A, Xu J, et al. Evaluation of highly accelerated simultaneous multi-slice EPI for fMRI. *NeuroImage* 2015; 104: 452–459.
68. Andersen FL, Ladefoged CN, Beyer T, et al. Combined PET/MR imaging in neurology: MR-based attenuation

correction implies a strong spatial bias when ignoring bone. *NeuroImage* 2014; 84: 206–216.
69. Parati G, Revera M, Giuliano A, et al. Effects of acetazolamide on Central blood pressure, peripheral blood pressure, and arterial distensibility at acute high altitude exposure. *Eur Heart J* 2013; 34: 759–766.
70. Liu PB, De Vis J and Lu H. Cerebrovascular reactivity (CVR) MRI with CO2 challenge: a technical review. *NeuroImage* 2018; 187: 104–115.
71. Pinto J, Bright MG, Bulte DP, et al. Cerebrovascular reactivity mapping without gas challenges: a methodological guide. *Front Physiol* 2021; 11doi:10.3389/fphys.2020.608475.
72. Saito H, Ogasawara K, Suzuki T, et al. Adverse effects of intravenous acetazolamide administration for evaluation of cerebrovascular reactivity using brain perfusion Single-Photon emission computed tomography in patients with major cerebral artery steno-occlusive diseases. *Neurol Med Chir (Tokyo)* 2011; 51: 479–483.
73. Alfaro-Almagro F, McCarthy P, Afyouni S, et al. Confound modelling in UK biobank brain imaging. *NeuroImage* 2020; 224: 117002.
74. Fülesdi B, Limburg M, Bereczki D, et al. Impairment of cerebrovascular reactivity in long-term type 1 diabetes. *Diabetes* 1997; 46: 1840–1845.
75. Tsujikawa T, Kimura H, Matsuda T, et al. Arterial transit time mapping obtained by pulsed continuous 3D ASL imaging with multiple post-label delay acquisitions: comparative study with PET-CBF in patients with chronic occlusive cerebrovascular disease. *PLoS One* 2016; 11: e0156005.
76. Ances B, Vaida F, Ellis R, et al. Test–retest stability of calibrated BOLD-fMRI in HIV− and HIV+ subjects. *NeuroImage* 2011; 54: 2156–2162.
77. Hou X, Liu P, Li Y, et al. The association between BOLD-based cerebrovascular reactivity (CVR) and end-tidal CO2 in healthy subjects. *NeuroImage* 2020; 207: 116365.