The Reproducibility of Cerebrovascular Reactivity Across MRI Scanners

Olivia Sobczyk[1,2], Ece Su Sayin[2,3], Kevin Sam[4], Julien Poublanc[1], James Duffin[2,3], Joseph A. Fisher[2,3,5]* and David J. Mikulis[1,5]

[1] Joint Department of Medical Imaging and the Functional Neuroimaging Laboratory, University Health Network, Toronto, ON, Canada, [2] Department of Anaesthesia and Pain Management, University Health Network, Toronto, ON, Canada, [3] Department of Physiology, University of Toronto, Toronto, ON, Canada, [4] Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [5] Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

Cerebrovascular reactivity (CVR) is defined as the ratio of the cerebral blood flow (CBF) response to an increase in a vasoactive stimulus. We used changes in blood oxygenation level-dependent (BOLD) MRI as surrogates for changes of CBF, and standardized quantitative changes in arterial partial pressure of carbon dioxide as the stimulus. Despite uniform stimulus and test conditions, differences in voxel-wise BOLD changes between testing sites may remain, attributable to physiologic and machine variability. We generated a reference atlas of normal CVR metrics (voxel-wise mean and SD) for each of two sites. We hypothesized that there would be no significant differences in CVR between the two atlases enabling each atlas to be used at any site. A total of 69 healthy subjects were tested to create site-specific atlases, with 20 of those individuals tested at both sites. 38 subjects were scanned at Site 1 (17F, 37.5 ± 16.8 y) and 51 subjects were tested at Site 2 (22F, 40.9 ± 17.4 y). MRI platforms were: Site 1, 3T Magnetom Skyra Siemens scanner with 20-channel head and neck coil; and Site 2, 3T HDx Signa GE scanner with 8-channel head coil. To construct the atlases, test results of individual subjects were co-registered into a standard space and voxel-wise mean and SD CVR metrics were calculated. Map comparisons of z scores found no significant differences between white matter or gray matter in the 20 subjects scanned at both sites when analyzed with either atlas. We conclude that individual CVR testing, and atlas generation are compatible across sites provided that standardized respiratory stimuli and BOLD MRI scan parameters are used. This enables the use of a single atlas to score the normality of CVR metrics across multiple sites.

Keywords: cerebrovascular reactivity, carbon dioxide, cerebral blood flow, MRI, reproducibility

INTRODUCTION

The magnitude of cerebral hemodynamic responses to alterations in arterial partial pressure of CO_2 ($PaCO_2$), referred to as cerebrovascular reactivity (CVR), has increased understanding of the determinants of cerebral perfusion in many neurovascular disorders. Presently, there are a number of techniques that have been developed for the assessment and mapping of CVR. Of these,

blood oxygenation level-dependent (BOLD) magnetic resonance imaging (MRI) signal intensity as a surrogate measure of cerebral blood flow (CBF) is in common use. However, the vasodilatory stimulus in most instances is uncontrolled, impeding test standardization, which is required for research and clinical use. Stimulus standardization requires precise, repeatable changes in $PaCO_2$ independent of subject, or patient, efforts. A standardized "brain stress test" would provide a uniform platform from which to assess and manage a wide variety of disorders affecting brain blood flow regulation, such as in steno-occlusive disease (Brawley et al., 1967; Brawley, 1968; Symon, 1969; Kleiser and Widder, 1992; Yonas et al., 1993; Webster et al., 1995; Molina et al., 1999; Markus and Cullinane, 2001; Ogasawara et al., 2002; Sasoh et al., 2003; Mandell et al., 2008, 2011; Fierstra et al., 2010; Silvestrini et al., 2011; Sobczyk et al., 2016), glioma (Fierstra et al., 2018a; Sebok et al., 2020), arteriovenous malformations (Fierstra et al., 2011), traumatic brain injury (Da Costa et al., 2016; Mutch et al., 2016) and remote stroke effects (Sebok et al., 2018; Hendrik Bas van Niftrik et al., 2020; Sebok et al., 2021). The advancement of the clinical and research applications of CVR has required, first, the standardization of the method (Sobczyk et al., 2014; Sobczyk et al., 2015), and second, the verification of the repeatability of the test (Sobczyk et al., 2015; Sobczyk et al., 2016). The third and final step, and the aim of this paper, is to validate the compatibility of data between sites and across MR platforms.

Test variability between sites may arise from disease activity, equipment variability, artifacts such as patient movement and from normal day-to-day physiological variability. Performing CVR studies using a standard stimulus and a single MR scanner with consistent sequence parameters optimizes stimulus and machine repeatability (Kassner et al., 2010; Sobczyk et al., 2016); therefore observed variation of CVR metrics outside of test variability is more attributable to changes in physiology and disease states. However, as much as the BOLD MR signal is dependent on the MR pulse parameters, there may also be hardware dependent variability and thus may differ between vendors and platforms despite the same scan parameters.

The first step for comparing data across scanners required the generation of an atlas consisting of merged CVR data from healthy subjects from which was calculated a voxel-wise mean and standard deviation of CVR. Single subject's CVR metrics can be compared to those of the atlas, analogous to the way blood clotting tests are compared to the normal range for *a particular* laboratory. Each voxel in an individual CVR scan would then be expressed in terms of standard deviations of the atlas mean (z score). Whereas absolute values for CVR may vary from center to center with the stimulus protocol and scan parameters, as long as the protocol for the CVR test and the atlas are the same, the z scores should be comparable across platforms. However, it would be very burdensome, expensive, and inefficient if an atlas had to be generated for every scanner. We hypothesized that if a repeatable and quantitative ventilatory protocol could be applied in conjunction with matching of critical MRI scan parameters, the variability in time-point to time-point and scanner-to-scanner variances could be minimized. A single healthy control atlas derived from one MRI system could then be used to assess single subjects across platforms. It would also enable the comparison of individuals and cohorts across institutions for use in multi-center trials.

The aim of this study was to evaluate whether the CVR atlases generated on different MRI platforms with the same, reproducible stimulus and the same scan parameters would have sufficient uniformity to be interchangeable across platforms for normalizing individual CVR data.

METHODS

Subjects and Ethical Approval

This study conformed to the standards set by the latest revision of the Declaration of Helsinki and were approved by the Research Ethics Board of the University Health Network and Health Canada. All subjects provided written and informed consent to participate in this study. We recruited 69 voluntary, healthy participants (31 females) between the ages of 18–80 years old by advertisement and word of mouth. Each subject was in good health with no history of neurological or cardiovascular disease, non-smokers taking no medication. The demographics for the cohort are represented in **Table 1**.

Experimental Protocol

CVR studies were performed on two separate MRI platforms. At Site 1, a 3T Magnetom SKYRAfit Siemens scanner with a 20-channel head and neck coil (bore diameter of 70 cm, slew rate 200 T/m/s, gradient peak 45 mT/m, software version Syngo MR D13B) was used, while at Site 2, a 3T HDx Signa GE scanner with an 8-channel head coil (bore diameter of 60 cm, slew rate of 150 T/m/s, gradient peak of 50 mT/m, software version HD16.0_V02_1131.a) was used.

All CVR studies reported in this study at both Site 1 and Site 2 consisted of matching MRI parameters and identical experimental CO_2 protocols. A total of 69 healthy subjects were scanned to create site-specific atlases (38 at Site 1 [17F, 37.5 ± 16.8 y] and 51 at Site 2 [22F, 40.9 ± 17.4 y]). From the 69 participants, 20 of those individuals completed CVRs at both sites within a 2-month time frame.

Hypercapnic Stimulus

Subjects breathed through a fitted facemask connected to a sequential gas delivery breathing circuit (Somogyi et al., 2005).

TABLE 1 | Summary of healthy subject demographics.

Age range	Scan at site 1 only	Scan at site 2 only	Scan at both sites
18–28	8	8	7
29–38	4	4	7
39–54	3	11	2
55–76	3	8	4
Sex			
F	9	14	8
M	9	17	12
Total	18	31	20

The CO_2 stimulus consisted of clamping end-tidal partial pressure of CO_2 ($P_{ET}CO_2$) at the subjects resting value for 2 min, followed by a step increase by 10 mmHg for 2 min and returned to baseline for another 2 min. $P_{ET}CO_2$ was then reduced by 10 mmHg for 1 minute, followed by a steady rise in $P_{ET}CO_2$ to 15 mmHg above baseline over ∼5 min, and returned to baseline for 2 min. The stimulus pattern shown in **Figure 1** was programmed into a computer-controlled gas blender (RespirAct[TM], Thornhill Research, Toronto, Canada) that runs a prospective gas targeting algorithm (Slessarev et al., 2007), which controls the CO_2 stimulus independently of tidal volume and breathing frequency such that $P_{ET}CO_2$ is equivalent to $PaCO_2$ (Ito et al., 2008; Willie et al., 2012). The entire CO_2 stimulus was performed at normoxia ($P_{ET}O2$ ∼110 mmHg) that is also precisely controlled, independently of CO_2, by the RespirAct[TM] system.

Magnetic Resonance Imaging Protocol for Cerebrovascular Reactivity Map Generation

The protocol consisted of: **1.** 3D T1-weighted Inversion-Recovery Prepared Fast Spoiled Gradient Echo (IR-FSPGR) with TI = 450 ms, TR = 7.88 ms, TE = 3 ms, flip angle = 12°, voxel = 0.859 × 0.859 × 1 mm, matrix = 256 × 256, 146 slices, field of view = 24 × 24 cm, no interslice gap, acquisition time = 7.4 min. **2.** T2*-weighted gradient recalled echo sequence with echo-planar readout (EPI-GRE) with TR = 2400 ms, TE = 30 ms, flip angle = 70°, 41 slices, voxel = 3.5 mm^3, matrix = 64 × 64, number of volumes = 335, field of view = 24 × 24 cm, acquisition time = 13.4 min.

The acquired BOLD MRI scan and $P_{ET}CO_2$ data were analyzed using AFNI software (National Institutes of Health,

Bethesda, Maryland)[1] (Cox, 1996). BOLD images were first volume registered and slice-time corrected and coregistered to the IR-FSPGR. $P_{ET}CO_2$ data was then time-shifted to the point of maximum correlation with the whole brain average BOLD signal. A linear, least-squares fit of the BOLD signal data series to the $P_{ET}CO_2$ data series was then performed on a voxel-by-voxel basis, together with a linear trend regressor to calculate CVR values. This method has been described in greater detail elsewhere (Fierstra et al., 2010; McKetton et al., 2018a). CVR for each dataset was calculated for different portions of the stimulus including the step portion of the stimulus (which includes dynamic changes in the CVR values (Poublanc et al., 2015), and the linear portion of the ramp stimulus starting from the individuals resting CO_2 up to + 10 mmHg from resting (representing a steady state CVR). CVR was expressed as the percent change in BOLD signal per change in $P_{ET}CO_2$ (%/mmHg). Step and ramp CVR maps were color-coded based on a spectrum of colors denoting the magnitude of response in either the positive or negative direction (**Figure 2**).
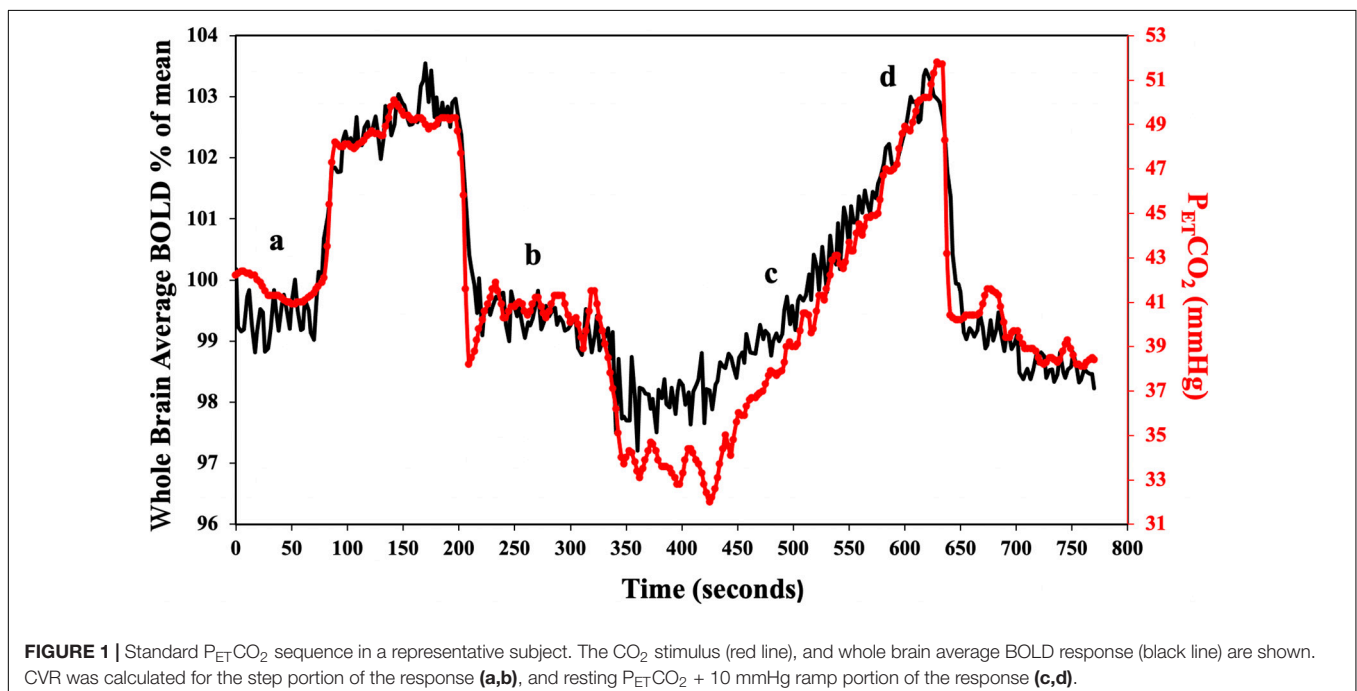
Analysis of Data
Construction of Cerebrovascular Reactivity Atlas

The normative data from all subjects across the two scanning sites were co-registered into Montreal Neurologic Institute (MNI) standard space using an analytical processing software (SPM8: Wellcome Department of Imaging Neuroscience, University of London, United Kingdom)[2]. A voxel-by-voxel mean CVR, and associated standard deviation (SD), was calculated for the step and ramp potions of the stimulus, to form site-specific atlases (AFNI software (Cox, 1996). Four sets of atlases were created

FIGURE 1 | Standard $P_{ET}CO_2$ sequence in a representative subject. The CO_2 stimulus (red line), and whole brain average BOLD response (black line) are shown. CVR was calculated for the step portion of the response **(a,b)**, and resting $P_{ET}CO_2$ + 10 mmHg ramp portion of the response **(c,d)**.

for each of CVR parameters: atlases were created for subjects scanned at each site (38-person atlas at Site 1, 51-person atlas at Site 2), and atlases that included only those individuals that were scanned at both sites (20 person atlases for both Site 1 and Site 2). Atlas generation methodology has been described in greater detail elsewhere (Poublanc et al., 2015; Sobczyk et al., 2015; McKetton et al., 2018b). The co-registered T1-weighted anatomical images for all subjects were segmented into gray and white matter (GM and WM) using the aforementioned SMP8 software. These masks were used to calculated mean GM and WM values for each vascular territory by hemisphere [left (L) and right (R) MCA (middle cerebral artery), ACA (anterior cerebral artery) and PCA (posterior cerebral artery)] for the CVR step and ramp portion of the stimulus in each subject.

Statistical Analysis

Voxel-wise paired t-tests ($\alpha = 0.05$) were used to determine if significant differences existed between the 20 person atlases of each CVR parameter (step and ramp), which contained only subjects that were scanned at both sites. Permutation testing was then applied to the height of the maxima of the resulting statistic image, using the "randomize" permutation-based inference tool (Winkler et al., 2014) in FSL v.5.0.9 (FMRIB Library[3] that allowed for the maintenance of strong control over family-wise error. Additional statistical analyses were conducted on the mean GM, WM and hemispheric vascular territories for each sets of atlases using SigmaPlot 12.5 (Systat Software, Inc., San Jose California). To evaluate measurement reproducibility and investigate possible bias in the CVR data, Bland-Altman analysis was used to calculate the two sites limits of agreement between the 20 person atlases for GM and WM of each CVR parameter. A Two-Way repeated measures analysis of variance (rmANOVA) with the Holm-Sidak method of multiple comparison correction was performed on the 20 person atlases to determine if any significant differences existed between the groups ($\alpha = 0.05$). For the full site atlases, a Two-Way ANOVA was performed with the Holm-Sidak method of multiple comparison correction ($\alpha = 0.05$).

Construction of Z-Maps

Individualized z scores were calculated on a voxel-wise basis by co-registering the subject's CVR map to the same space as the reference atlases followed by scoring the subject's CVR values by assigning a z-score according to the mean and SD of the corresponding voxel of the atlas (Sobczyk et al., 2015). A healthy subject individual z-scores were calculated by a "leave one out"/"jackknife" procedure, which requires taking the healthy subject out from the atlases and z scoring that subject to atlases made up of the remaining subjects (Minoshima et al., 1995). The z values were then color coded and superimposed on the anatomical scans to form site-specific z-maps. Frequency distribution histograms (FDH) were created of the z-maps for each vascular territory. Absolute difference maps were calculated voxel-wise from the z-maps created from Site 1 and Site 2.
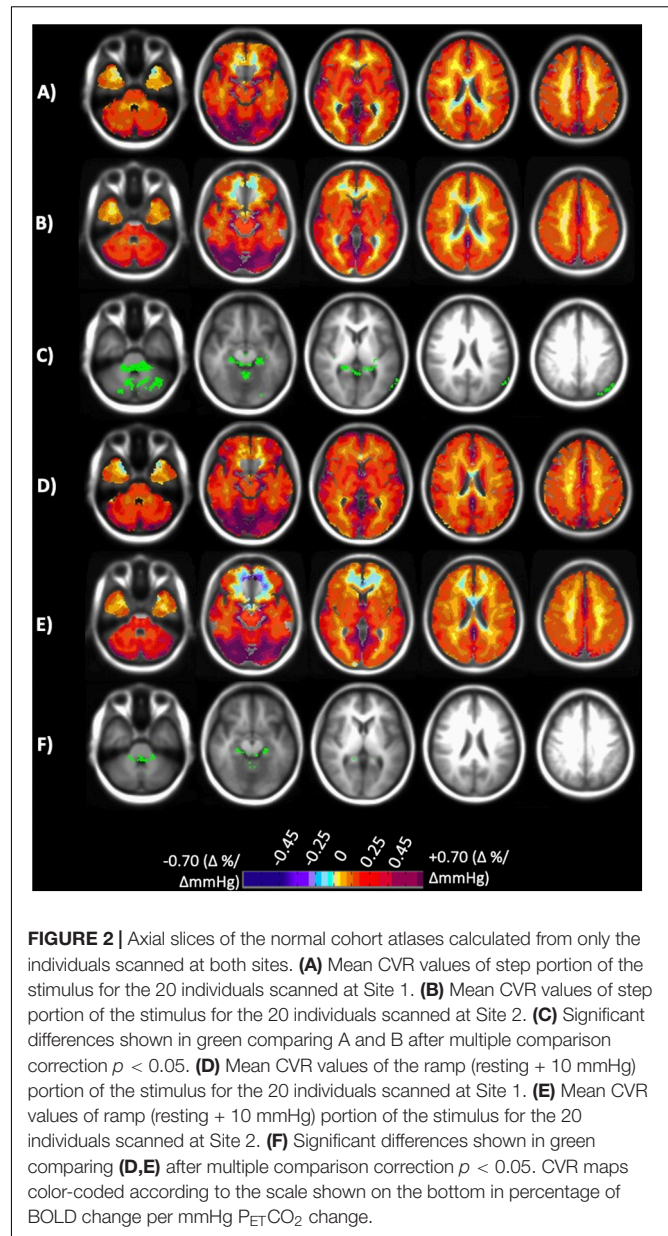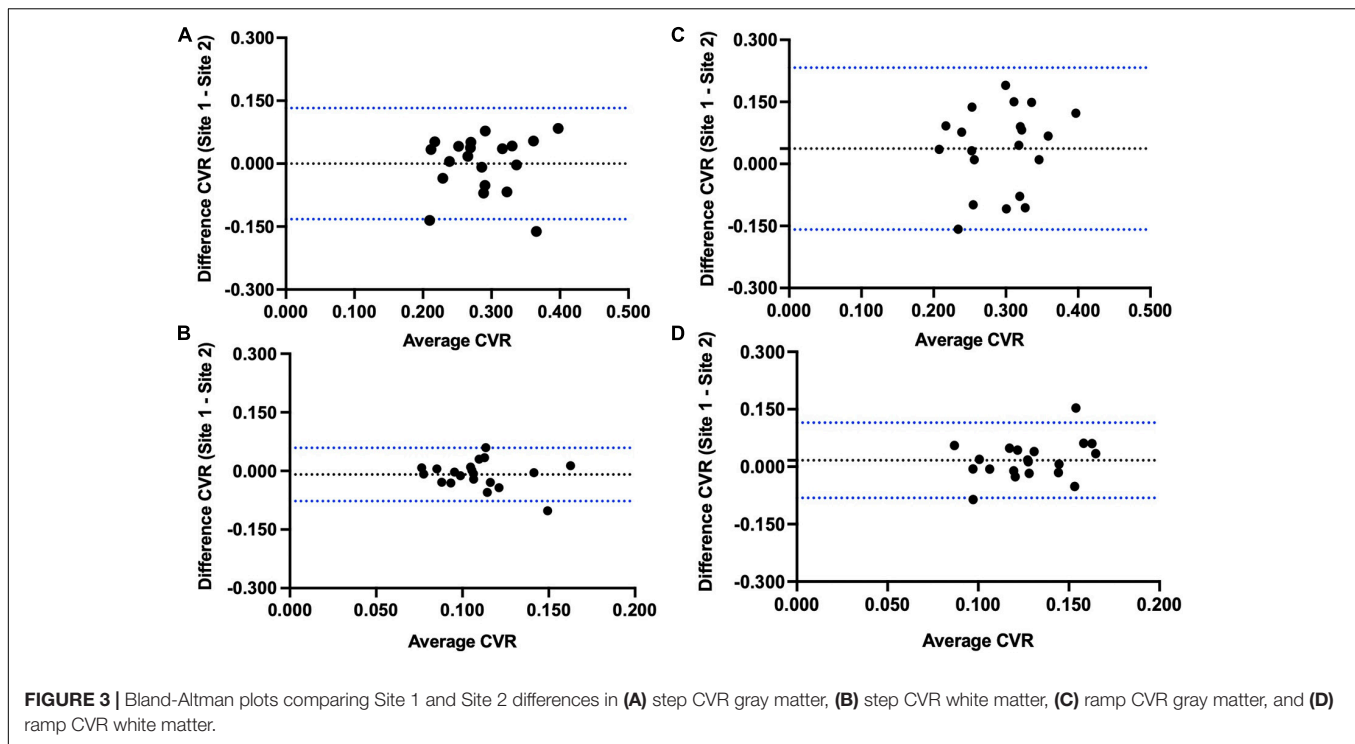
---

[3]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

FIGURE 2 | Axial slices of the normal cohort atlases calculated from only the individuals scanned at both sites. **(A)** Mean CVR values of step portion of the stimulus for the 20 individuals scanned at Site 1. **(B)** Mean CVR values of step portion of the stimulus for the 20 individuals scanned at Site 2. **(C)** Significant differences shown in green comparing A and B after multiple comparison correction $p < 0.05$. **(D)** Mean CVR values of the ramp (resting + 10 mmHg) portion of the stimulus for the 20 individuals scanned at Site 1. **(E)** Mean CVR values of ramp (resting + 10 mmHg) portion of the stimulus for the 20 individuals scanned at Site 2. **(F)** Significant differences shown in green comparing **(D,E)** after multiple comparison correction $p < 0.05$. CVR maps color-coded according to the scale shown on the bottom in percentage of BOLD change per mmHg $P_{ET}CO_2$ change.

RESULTS

Atlas Results for Only Subjects Scanned at Both Sites

CVR step and ramp atlas maps were created for the 20 individuals that were scanned at both sites and compared to one another to discern site-related differences. **Figure 2** displays the atlas maps for CVR step and ramp at each site. A voxel-wise paired t-test for each of the CVR parameter maps between Site 1 and Site 2 was preformed, and significant differences ($p < 0.05$) are displayed in **Figure 2**. Minimal significant differences were found between Site 1 and Site 2 for both CVR parameters.

The results of the Bland-Altman analysis for between site reproducibility for the 20 individuals that were scanned at both

FIGURE 3 | Bland-Altman plots comparing Site 1 and Site 2 differences in **(A)** step CVR gray matter, **(B)** step CVR white matter, **(C)** ramp CVR gray matter, and **(D)** ramp CVR white matter.

sites are illustrated in **Figure 3**. Mean site difference for CVR step (mean $\pm$ 1.95 SD) was $-1.15 \times 10^{-5} \pm 0.068$%/mmHg in GM and -0.009 ± 0.035%/mmHg in WM. Mean site difference for CVR ramp was 0.037 ± 0.099%/mmHg in GM and 0.017 ± 0.050%/mmHg in WM.

Mean step and ramp CVR values of the 20 subjects scanned at Site 1 and Site 2 for GM, WM and each hemispheric vascular territory, were calculated. GM and WM values are presented in **Table 2** along with the associated SD. Mean CVR values for each hemispheric vascular territory are presented in **Supplementary Table 1**. On average, CVR measures were similar in both GM and WM between sites (*P*-values presented in **Table 2**, $\alpha = 0.05$). The Two-Way rmANOVA found a statistically significant interaction between site and region of interest (ROI) and found significant differences between sites for step CVR in the RACA GM and WM, and LPCA GM. Significant differences between sites for ramp CVR were found in the L and R MCA GM, and the LACA GM and WM (*P*-values presented in **Supplementary Table 1**, $\alpha = 0.05$). An interpretation of these results with respect to the implications of the interaction between site and ROI factors is found in the discussion section.

Atlas Results for All Subjects Scanned at Either Site

CVR step and ramp atlas maps were created for all subjects scanned at either site, with 38 subjects in total scanned at Site 1 and 51 subjects scanned at Site 2. **Figure 4** displays the atlas maps for CVR step and ramp at Site 1 and Site 2.

A Two-Way ANOVA was performed for each of the CVR parameter maps comparing all subjects scanned at each site for GM, WM, and each hemispheric vascular territory. The

overall means, SD, and associated *p*-values for GM and WM are presented in **Table 3**. Mean CVR values for each hemispheric vascular territory are presented in **Supplementary Table 2**. No significant difference between the CVR step and ramp for GM and WM were found between the two sites (*P*-values presented in **Table 3**, $\alpha = 0.05$). For the hemispheric ROIs, a statistically significant interaction between site and region of interest (ROI) and significant differences between sites were found for step CVR in the L and R MCA GM, L and R ACA GM, LPCA GM, and L and R PCA WM. Significant differences between sites for ramp CVR were found in the LMCA GM, L and R ACA GM, L and R ACA WM, LPCA GM, and the L and R PCA WM (*P*-values presented in **Supplementary Table 2**, $\alpha = 0.05$). An interpretation of these results with respect to the implications of the interaction between site and ROI factors is found in the discussion section.

Z-map Normalization Examples

Figure 5 presents the CVR and accompanying z-maps and absolute differences of the z scores from a healthy subject that was

TABLE 2 | Mean, standard deviation (SD) and *p*-values for gray (GM) and white matter (WM) regions of each CVR parameter for the 20 healthy subjects that were scanned at both sites.

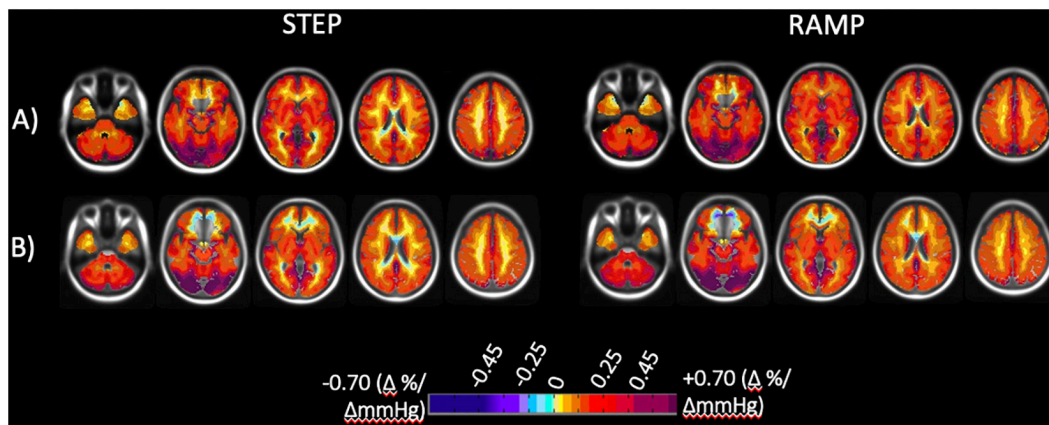|		Site 1		Site 2		
		Mean (%/mmHg)	SD	Mean (%/mmHg)	SD	*P*-value
CVR GM	Step	0.287	0.0649	0.287	0.0620	0.999
	Ramp	0.312	0.0781	0.275	0.0641	0.114
CVR WM	Step	0.104	0.0253	0.113	0.0310	0.255
	Ramp	0.136	0.0396	0.120	0.0280	0.154

FIGURE 4 | Axial slices of the normal cohort atlases calculated from all individuals for both the step and ramp (resting + 10 mmHg) portions of the stimulus. **(A)** Mean CVR values for all individuals scanned at Site 1 (38 individuals total). **(B)** Mean CVR values for all individuals scanned at Site 2 (51 individuals total). CVR maps color-coded according to the scale shown on the bottom in percentage of BOLD change per mmHg $P_{ET}CO_2$ change.

scanned at both sites and was removed from the atlases prior to calculating the z scores (known as a "jackknife" procedure what is used when taking a subject from a reference atlas and comparing them back to the atlas to perform z-scoring (Minoshima et al., 1995). The FDH of the step and ramp z scores for each vascular territory and an additional normal example is presented in **Supplementary Material**.

Figure 6 shows the application and comparison of the CVR atlases for each parameter to a patient with steno-occlusive disease whose CVR was collected at Site 2 only, 3 years after the atlas data was collected. Included are z scoring and absolute voxel-wise differences between the z-maps. The FDH of the step and ramp z scores for each vascular territory and an additional patient example with steno-occlusive disease is presented in **Supplementary Material**.

DISCUSSION

The main finding of this study is that voxel-wise CVR test results are not significantly different between two sites using different MR scanners when a precisely repeatable hypercapnic stimulus is applied and consistent MR scan parameters are specified. CVR test results of 20 subjects, scanned at both sites, were used to create atlases of CVR variability for the two locations.

Voxel-wise paired t-tests applied to the atlases showed minimal differences (**Figures 2C,F**), and Bland-Altman analysis plots (**Figure 3**) displayed no proportional difference. Consequently, normalization of individual CVR values via z scores may be made using either atlas.

With respect to further atlas comparison using two-way rmANOVA, although such analysis of the step CVR atlases found significant differences for most of the GM and WM ACA territories (areas prone to susceptibility artifacts), these differences were not interpretable with respect to site difference since there was a statistically significant interaction between factors Site and ROI ($P \leq 0.001$). The ramp CVR atlases which use only a few points of the full ramp stimulus (resting $P_{ET}CO_2$ + 10 mmHg) and are consequently more variable than the step CVR measurements, also showed several ROIs that were significantly different between sites, with the majority again in the ACA territories. However, interaction between site and ROI was also found and so these findings are not interpretable with respect to site difference.

Thus, an atlas of CVR in healthy people generated in one location can provide the same population normalization for CVR (Sobczyk et al., 2015) and interval differences in CVR (Sobczyk et al., 2016) regardless of where the individual measurements were made. Two patient examples (**Figure 6** and **Supplementary Figure 2**) support this conclusion as well as the z score FDHs provided in **Supplementary Material**.

If the goal of using a single atlas to assess normality regardless of CVR testing site is to be achieved, every effort is required to eliminate non-physiologic variability. The two sources of CVR variability, stimulus and the measurement of its effect are the barriers to site-independent assessment of CVR. In studies where the vasoactive stimulus cannot be quantitated, much less standardized [e.g., acetazolamide injection, breath hold duration (Fierstra et al., 2013), and inhalation of a fixed CO_2 concentration (Fisher, 2016)], it produces an unknowable variability in the CVR.

The CVR stimulus must be standardized, not only with respect to the sequence of CO_2 changes used in terms of baseline, but also

TABLE 3 | Mean, standard deviation (SD) and p-values for gray (GM) and white matter (WM) regions of each CVR parameter for all subjects that were scanned at each site.

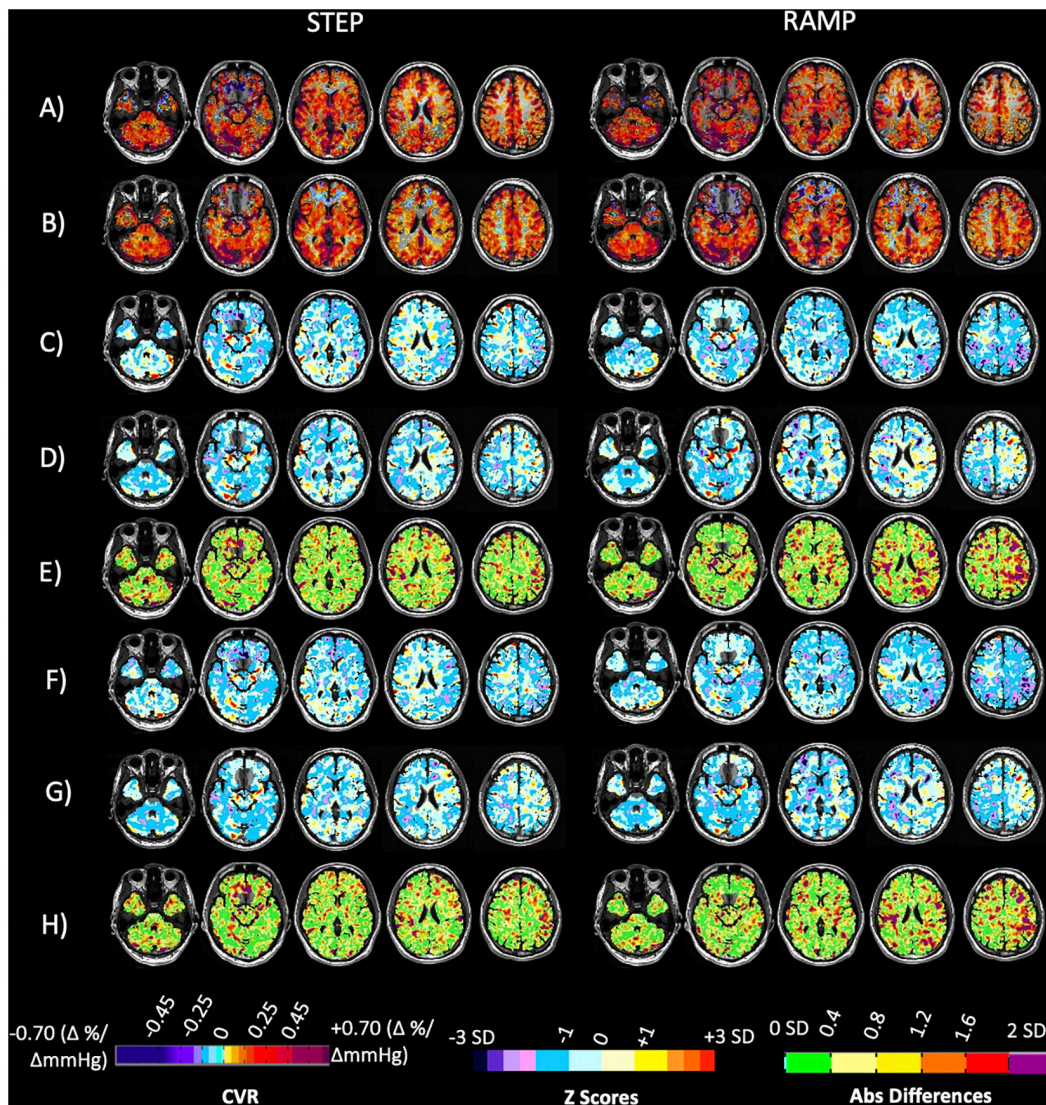|		Site 1		Site 2		
		Mean (%/mmHg)	SD	Mean (%/mmHg)	SD	P-value
CVR GM	Step	0.300	0.0567	0.291	0.063	0.532
	Ramp	0.317	0.0669	0.303	0.071	0.337
CVR WM	Step	0.114	0.0278	0.116	0.032	0.901
	Ramp	0.143	0.0358	0.140	0.033	0.689

FIGURE 5 | Case illustration of a normal subject. Displayed are axial slices for **(A)** the subject CVR maps for the step and ramp (resting $P_{ET}CO_2$ + 10 mmHg) portion of the stimulus obtained at Site 1. **(B)** The subject CVR maps for the step and ramp (resting $P_{ET}CO_2$ + 10 mmHg) portion of the stimulus obtained at Site 2. **(C)** Z-maps of the subject calculated by comparing their CVR parameter maps from Site 1 **(A)** to the corresponding 20 person CVR parameter atlases collected at Site 1. **(D)** Z-maps of the subject calculated by comparing their CVR parameter maps obtained at Site 2 **(B)** to the corresponding 20 person CVR parameter atlases collected at Site 2. **(E)** Voxel-wise absolute difference maps calculated from the z-maps in **(C,D)**. **(F)** Z-maps of the subject calculated by comparing their CVR parameter maps obtained from Site 1 **(A)** to the corresponding full 38 person CVR parameter atlases collected at Site 1. **(G)** Z-maps of the subject calculated by comparing their CVR parameter maps obtained from Site 2 **(B)** to the corresponding full 51 person CVR parameter atlases collected at Site 2. **(H)** Voxel-wise absolute difference maps calculated from the z-maps in F and G. Associated color scales found at bottom of figure. The CVR color scale denotes areas of positive and negative response in percentage of BOLD change per mmHg $P_{ET}CO_2$ change. The z scores and absolute difference scales provide a perspective of the statistically normal differences in CVR in standard deviation (SD).

to the direction of change, amplitude, and duration of the true independent variable, the arterial PCO_2. Giving a standard CVR stimulus that is comparable between patients, institutions, and over time, allows for the monitoring of a disease progress, and the results of therapeutic interventions to be assessed (Sobczyk et al., 2016) assuming inconsequential variability in the output of the MR scanner.

If the stimulus can be standardized the remaining source of technical variability is the variation in scanner settings. In the

case of the atlases developed in this study, the main difference between the two platforms were the head coils. The Siemens scanner had a 20-channel coil whereas the GE scanner only had an 8-channel coil; however, the scan settings were the same. While this difference will affect the signal to noise ratio (SNR) between machines, the z-map images will be otherwise the same. The variability between subjects is expected to be greater than the experimental noise within a subject even with a lower SNR coil. The effect of using different scan parameters such as voxel
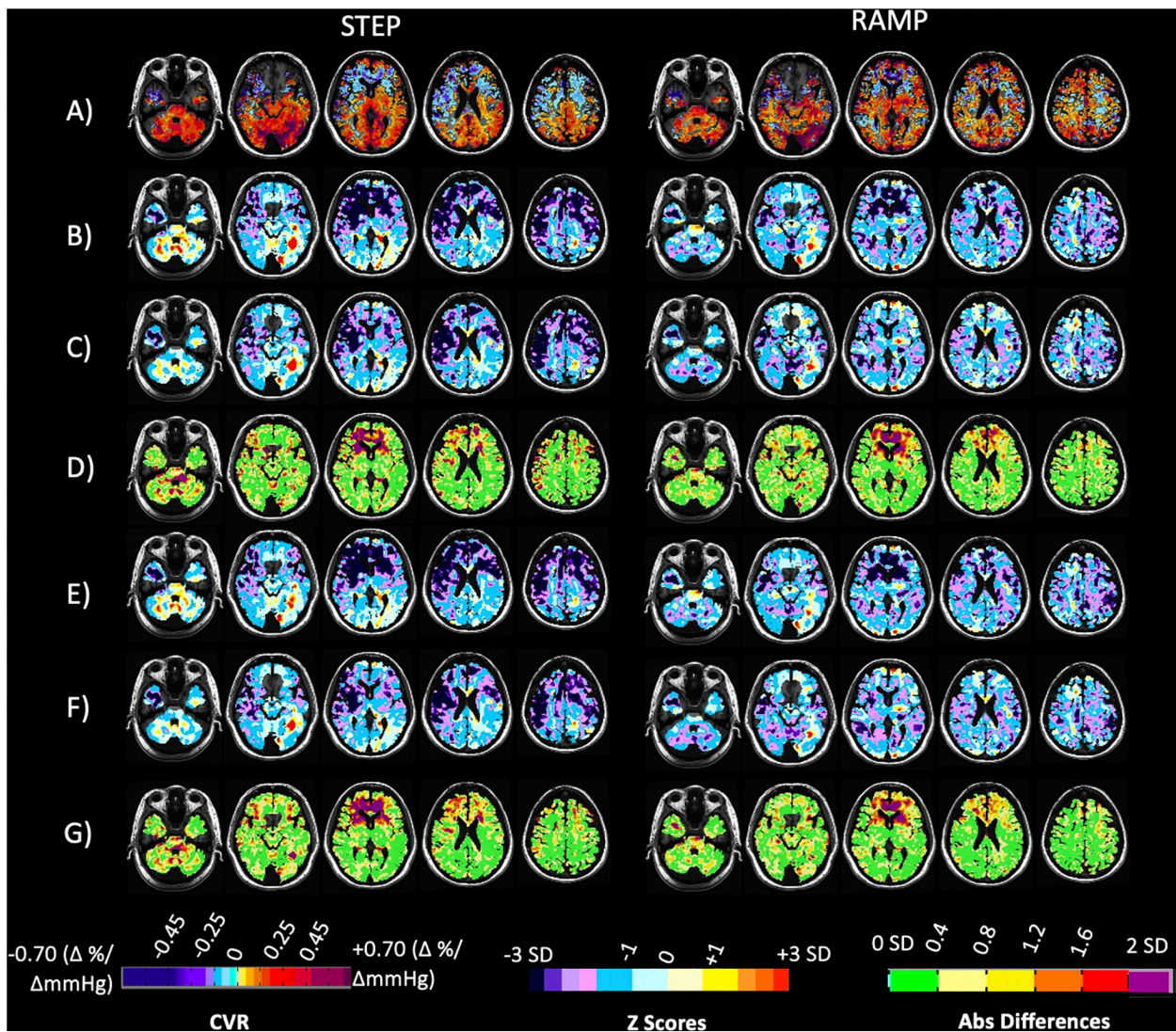
FIGURE 6 | Case illustration of a 40-year-old patient with transient ischemic attacks and bilateral narrowing of the internal carotid arteries, scanned at Site 2 only. Displayed are axial slices for **(A)** the subject's CVR maps for the step and ramp (resting $P_{ET}CO_2$ + 10 mmHg) portion of the stimulus. **(B)** Z-maps of the patient calculated by comparing their CVR parameter maps **(A)** to the corresponding 20 person CVR parameter atlases collected at Site 1. **(C)** Z-maps of the patient calculated by comparing their CVR parameter maps **(A)** to the corresponding 20 person CVR parameter atlases collected at Site 2. **(D)** Voxel-wise absolute difference maps calculated from the z-maps in **(B,C)**. **(E)** Z-maps of the patient calculated by comparing their CVR parameter maps in **(A)** to the corresponding full 38 person CVR parameter atlases collected at Site 1. **(F)** Z-maps of the patient calculated by comparing their CVR parameter maps in **(A)** to the corresponding full 51 person CVR parameter atlases collected at Site 2. **(G)** Voxel-wise absolute difference maps calculated from the z-maps in **(E,F)**. The CVR color scale denotes areas of positive and negative response in percentage of BOLD change per mmHg $P_{ET}CO_2$ change.

size on the individual scans and comparing them to the atlases was not examined.

If these two technical issues are standardized, the remaining sources of variability between scans must arise from differences in physiological factors affecting the MR signals. These include the physiological expression of the variety of day-to-day experiences such as amount of sleep the previous night, foods eaten, tobacco use, emotional state, and ingested medications. However, the effects of these experiences may be dampened or cancel each other out. In any event, their effect size will be small compared to those of the technical conditions. Indeed test-retest differences

were measured and quantified in a single scanner over weeks to months as mean ± SD for each voxel by Sobczyk et al. (2016). In the current study we show that CVR of normal individual voxels on 2 separate scanners fall within the mean and less than 2 SD of the CVR measured with either scanner.

Clinical Implications

When z scoring of scans was first contemplated, it was considered that each institution would need to use their own standardized stimulus pattern for patients and repeat that pattern on a control group to generate an atlas for each scanner. Our findings suggest

that for a standardized stimulus and MRI sequence, the variation in CVR is sufficiently small that a single atlas would be suitable across platforms. In addition, specialized atlases such as those controlling for sex, age, or medication, can be distributed across scanner locations. These considerations imply that all CVR data generated by a standard stimulus and scanner settings would be comparable, enabling a standard CVR test.

Research Implications

For all intents and purposes, the CVR atlas functions as the "control" cohort for all CVR studies, thereby requiring smaller test cohorts. The reduced test variability would also enable the detection of smaller effect sizes with smaller cohorts. Data from various centers would be directly comparable, as is the case with other standardized tests such as the clotting test prothrombin time. The atlas can be made to have greater specificity for a given condition by narrowing the variability of the cohort to control for specific factors. For example, for examining CVR in patients with multiple sclerosis, the atlas can be restricted to subjects of younger age, and if examining females, then restricted by eliminating data from males.

Limitations

Accurate measurements of CBF in advanced cerebrovascular disease is difficult to achieve. The limitations of BOLD as a surrogate measure of relative CBF changes are well known (Sobczyk et al., 2014; Duffin et al., 2015; Tancredi et al., 2015), nevertheless BOLD has been shown to correlate well with other modalities, including FAIR arterial spin-labeling (Hoge et al., 1999; Mandell et al., 2008; Chen and Pike, 2010) and more recently with (^{15}O–)H$_2$O positron emission tomography (PET) (Fierstra et al., 2018b; Hauser et al., 2019). With respect to the CO$_2$ stimulus, although a general linear model (GLM) is assumed for matching BOLD to P$_{ET}$CO$_2$, this relationship is very complex and not necessarily well approximated by assuming a GLM (Fisher et al., 2017; Duffin et al., 2018).

Atlases were also constructed and compared between sites using all subjects tested so that they included CVR data from the 20 subjects studied in both locations. While this comparison is like that of the matched 20 atlases, nevertheless, we feel that these atlases were worth examining because they contained many subjects unique to a particular test site, and even when used to examine a cohort, the data from the overlapped subjects were gathered at the alternate location.

Small differences were apparent in both the atlas maps and the z score maps. The two scanner specific atlases derived from the same group of 20 individuals tested on both 3T scanners from different vendors were virtually identical but exhibited small differences (**Figure 2**, green in 2C and 2F). These were primarily within the cerebral spinal fluid of the ambient cisterns and cerebellum, which we attribute to susceptibility effects, and in venous structures. Similarly, when these atlases were derived from larger groups of different individuals, and included 20 of the same individuals, the atlases were virtually identical (**Figure 4**). However, there were differences in anterior inferior frontal lobe tissues. The BOLD signal is unreliable in areas where the applied static magnetic field is inhomogeneous. This typically occurs at the skull base and most often affects the anteromedial and inferior frontal lobes (gyrus rectus and orbito-frontal gyrus). CVR values in these regions are typically unreliable and can even be false positive for steal physiology.

Comparing CVR z score maps (**Figures 6D,G**) of a single patient calculated from control atlases generated from the same 20 healthy controls on each of the two scanners did find small differences after excluding those regions previously discussed where differences were presumably due to susceptibility effects. These differences were mostly within 0.4 SD, with a few regions showing 0.8 SD differences. Thus, a voxel-wise variance less than 1 SD is achievable when comparing a single individual against control atlases generated on different vendor scanners.

Variability within the atlases may be attributed to differences in age and sex, although we have not found differences in CVR across age ranges (McKetton et al., 2018a). Subjects for our reference CVR atlases were not selected for these or any other characteristics. As a result, the atlas contains a high variability in the normal cohort, which reduced the significance of the variability in output between platforms. However, the types of vascular disease typically studied including moyamoya, steno-occlusive disease, small vessel disease, and traumatic brain injury, result in large CVR changes that exceed those of machine variability, and thus a single atlas is suitable for these conditions. Should very subtle CVR changes be sought, the method may be made more sensitive by generating a local atlas with a cohort that contains the characteristics and medical conditions, or medications, for which one wishes to control; for example, assembling a cohort of women between the ages of 20 and 30 in an atlas for investigation for subtle CVR signs in multiple sclerosis.

CONCLUSION

This study has demonstrated the feasibility of using a single atlas across platforms. A single atlas would be useful for normalizing clinical studies of individual patients as well as pooling compatible data in multicenter trials.

DATA AVAILABILITY STATEMENT

Anonymized data will be shared by request from any qualified investigator for purposes such as replicating procedures and results presented in the article provided that data transfer is in agreement with the research institution on the general data protection regulation. Requests to access the datasets should be directed to JAF, joe.fisher@utoronto.ca

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University Health Network. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained

from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

OS and KS were involved with data collection. OS, ES, JD, and JAF were involved in the data analysis and drafted the manuscript. All authors participated in the feedback and

writing process following the initial drafting of the manuscript, contributed to the design and conceptualization of the study.

SUPPLEMENTARY MATERIAL

REFERENCES

Brawley, B. W. (1968). The pathophysiology of intracerebral steal following carbon dioxide inhalation, an experimental study. *Scand. J. Clin. Lab. Invest. Suppl.* 102, 13:B.

Brawley, B. W., Strandness, D. E. Jr., and Kelly, W. A. (1967). The physiologic response to therapy in experimental cerebral ischemia. *Arch. Neurol.* 17, 180–187. doi: 10.1001/archneur.1967.00470260070008

Chen, J. J., and Pike, G. B. (2010). MRI measurement of the BOLD-specific flow-volume relationship during hypercapnia and hypocapnia in humans. *Neuroimage* 53, 383–391. doi: 10.1016/j.neuroimage.2010.07.003

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Da Costa, L., Van Niftrik, C. B., Crane, D., Fierstra, J., and Bethune, A. (2016). Temporal profile of cerebrovascular reactivity impairment, grey matter volumes and persistent symptoms after mild traumatic head injury. *Front. Neurol.* 7:70. doi: 10.3389/fneur.2016.00070

Duffin, J., Sobczyk, O., Crawley, A. P., Poublanc, J., Mikulis, D. J., and Fisher, J. A. (2015). The dynamics of cerebrovascular reactivity shown with transfer function analysis. *NeuroImage* 114, 207–216. doi: 10.1016/j.neuroimage.2015.04.029

Duffin, J., Sobczyk, O., Mcketton, L., Crawley, A., Poublanc, J., Venkatraghavan, L., et al. (2018). Cerebrovascular resistance: the basis of cerebrovascular reactivity. *Front. Neurosci.* 12:409. doi: 10.3389/fnins.2018.00409

Fierstra, J., Conklin, J., Krings, T., Slessarev, M., Han, J. S., Fisher, J. A., et al. (2011). Impaired peri-nidal cerebrovascular reserve in seizure patients with brain arteriovenous malformations. *Brain* 134, 100–109. doi: 10.1093/brain/awq286

Fierstra, J., Poublanc, J., Han, J. S., Silver, F., Tymianski, M., Crawley, A. P., et al. (2010). Steal physiology is spatially associated with cortical thinning. *J. Neurol. Neurosurg. Psychiatry* 81, 290–293. doi: 10.1136/jnnp.2009.188078

Fierstra, J., Sobczyk, O., Battisti-Charbonney, A., Mandell, D. M., Poublanc, J., Crawley, A. P., et al. (2013). Measuring cerebral vascular reactivity: what stimulus to use? *J. Physiol.* 591, 5809–5821. doi: 10.1113/jphysiol.2013.259150

Fierstra, J., Van Niftrik, C., Piccirelli, M., Bozinov, O., Pangalu, A., Krayenbuhl, N., et al. (2018a). Diffuse gliomas exhibit whole brain impaired cerebrovascular reactivity. *Magn. Reson. Imaging* 45, 78–83. doi: 10.1016/j.mri.2017.09.017

Fierstra, J., Van Niftrik, C., Warnock, G., Wegener, S., Piccirelli, M., Pangalu, A., et al. (2018b). Staging hemodynamic failure with blood oxygen-level–dependent functional magnetic resonance imaging cerebrovascular reactivity. *Stroke* 49, 621–629. doi: 10.1161/strokeaha.117.020010

Fisher, J. A. (2016). The CO2 stimulus for cerebrovascular reactivity: fixing inspired concentrations vs. targeting end-tidal partial pressures. *J. Cereb. Blood Flow Metab.* 36, 1004–1011. doi: 10.1177/0271678x16639326

Fisher, J. A., Sobczyk, O., Crawley, A., Poublanc, J., Dufort, P., Venkatraghavan, L., et al. (2017). Assessing cerebrovascular reactivity by the pattern of response to progressive hypercapnia. *Hum. Brain Mapp.* 38, 3415–3427.

Hauser, T.-K., Seeger, A., Bender, B., Klose, U., Thurow, J., Ernemann, U., et al. (2019). Hypercapnic BOLD MRI compared to H215O PET/CT for the hemodynamic evaluation of patients with Moyamoya Disease. *Neuroimage Clin.* 22:101713. doi: 10.1016/j.nicl.2019.101713

Hendrik Bas van Niftrik, C., Sebok, M., Muscas, G., Piccirelli, M., Serra, C., Krayenbuhl, N., et al. (2020). Characterizing ipsilateral thalamic diaschisis in

symptomatic cerebrovascular steno-occlusive patients. *J. Cereb. Blood Flow Metab.* 40, 563–573. doi: 10.1177/0271678x19830532

Hoge, R. D., Atkinson, J., Gill, B., Crelier, G. R., Marrett, S., and Pike, G. B. (1999). Investigation of BOLD signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model. *Magn. Reson. Med.* 42, 849–863. doi: 10.1002/(sici)1522-2594(199911)42:5<849::aid-mrm4>3.0.co;2-z

Ito, S., Mardimae, A., Han, J., Duffin, J., Wells, G., Fedorko, L., et al. (2008). Non-invasive prospective targeting of arterial PCO2 in subjects at rest. *J. Physiol.* 586, 3675–3682. doi: 10.1113/jphysiol.2008.154716

Kassner, A., Winter, J. D., Poublanc, J., Mikulis, D. J., and Crawley, A. P. (2010). Blood-oxygen level dependent MRI measures of cerebrovascular reactivity using a controlled respiratory challenge: reproducibility and gender differences. *J. Magn. Reson. Imaging* 31, 298–304. doi: 10.1002/jmri.22044

Kleiser, B., and Widder, B. (1992). Course of carotid artery occlusions with impaired cerebrovascular reactivity. *Stroke* 23, 171–174. doi: 10.1161/01.str.23.2.171

Mandell, D. M., Han, J. S., Poublanc, J., Crawley, A. P., Fierstra, J., Tymianski, M., et al. (2011). Quantitative measurement of cerebrovascular reactivity by blood oxygen level-dependent MR imaging in patients with intracranial stenosis: preoperative cerebrovascular reactivity predicts the effect of extracranial-intracranial bypass surgery. *AJNR Am. J. Neuroradiol.* 32, 721–727. doi: 10.3174/ajnr.a2365

Mandell, D. M., Han, J. S., Poublanc, J., Crawley, A. P., Stainsby, J. A., Fisher, J. A., et al. (2008). Mapping cerebrovascular reactivity using blood oxygen level-dependent MRI in Patients with arterial steno-occlusive disease: comparison with arterial spin labeling MRI. *Stroke* 39, 2021–2028. doi: 10.1161/strokeaha.107.506709

Markus, H., and Cullinane, M. (2001). Severely impaired cerebrovascular reactivity predicts stroke and TIA risk in patients with carotid artery stenosis and occlusion. *Brain* 124, 457–467. doi: 10.1093/brain/124.3.457

McKetton, L., Sobczyk, O., Duffin, J., Poublanc, J., Sam, K., Crawley, A. P., et al. (2018a). The aging brain and cerebrovascular reactivity. *Neuroimage* 181, 132–141. doi: 10.1016/j.neuroimage.2018.07.007

McKetton, L., Venkatraghavan, L., Poublanc, J., Sobczyk, O., Crawley, A. P., Rosen, C., et al. (2018b). Importance of collateralization in patients with large artery intracranial occlusive disease: long-term longitudinal assessment of cerebral hemodynamic function. *Front. Neurol.* 9:226. doi: 10.3389/fneur.2018.00226

Minoshima, S., Frey, K. A., Koeppe, R. A., Foster, N. L., and Kuhl, D. E. (1995). A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. *J. Nucl. Med.* 36, 1238–1248.

Molina, C., Sabin, J. A., Montaner, J., Rovira, A., Abilleira, S., and Codina, A. (1999). Impaired cerebrovascular reactivity as a risk marker for first-ever lacunar infarction: a case-control study. *Stroke* 30, 2296–2301. doi: 10.1161/01.str.30.11.2296

Mutch, W. a. C, Ellis, M. J., Ryner, L. N., Graham, M. R., Dufault, B., Gregson, B., et al. (2016). Brain magnetic resonance imaging CO2 stress testing in adolescent postconcussion syndrome. *J. Neurosurg.* 125, 648–660. doi: 10.3171/2015.6.jns15972

Ogasawara, K., Ogawa, A., and Yoshimoto, T. (2002). Cerebrovascular reactivity to acetazolamide and outcome in patients with symptomatic internal carotid or middle cerebral artery occlusion: a xenon-133 single-photon emission computed tomography study. *Stroke* 33, 1857–1862. doi: 10.1161/01.str.0000019511.81583.a8

Poublanc, J., Crawley, A. P., Sobczyk, O., Montandon, G., Sam, K., Mandell, D. M., et al. (2015). Measuring cerebrovascular reactivity: the dynamic response to a step hypercapnic stimulus. *J. Cereb. Blood Flow Metab.* 35, 1746–1756. doi: 10.1038/jcbfm.2015.114

Sasoh, M., Ogasawara, K., Kuroda, K., Okuguchi, T., Terasaki, K., Yamadate, K., et al. (2003). Effects of EC-IC bypass surgery on cognitive impairment in patients with hemodynamic cerebral ischemia. *Surg. Neurol.* 59, 455–460; discussion460–453.

Sebok, M., Van Niftrik, C. H. B., Halter, M., Hiller, A., Seystahl, K., Pangalu, A., et al. (2020). Crossed cerebellar diaschisis in patients with diffuse glioma is associated with impaired supratentorial cerebrovascular reactivity and worse clinical outcome. *Cerebellum* 19, 824–832. doi: 10.1007/s12311-020-01174-y

Sebok, M., Van Niftrik, C. H. B., Piccirelli, M., Bozinov, O., Wegener, S., Esposito, G., et al. (2018). BOLD cerebrovascular reactivity as a novel marker for crossed cerebellar diaschisis. *Neurology* 91, e1328–e1337.

Sebok, M., Van Niftrik, C. H. B., Piccirelli, M., Muscas, G., Pangalu, A., Wegener, S., et al. (2021). Crossed cerebellar diaschisis in patients with symptomatic unilateral anterior circulation stroke is associated with hemodynamic impairment in the ipsilateral MCA territory. *J. Magn. Reson. Imaging* 53, 1190–1197. doi: 10.1002/jmri.27410

Silvestrini, M., Viticchi, G., Falsetti, L., Balucani, C., Vernieri, F., Cerqua, R., et al. (2011). The role of carotid atherosclerosis in Alzheimer's disease progression. *J. Alzheimers Dis.* 25, 719–726.

Slessarev, M., Han, J., Mardimae, A., Prisman, E., Preiss, D., Volgyesi, G., et al. (2007). Prospective targeting and control of end-tidal CO2 and O2 concentrations. *J. Physiol.* 581, 1207–1219. doi: 10.1113/jphysiol.2007.129395

Sobczyk, O., Battisti-Charbonney, A., Fierstra, J., Mandell, D. M., Poublanc, J., Crawley, A. P., et al. (2014). A conceptual model for CO2-induced redistribution of cerebral blood flow with experimental confirmation using BOLD MRI. *Neuroimage* 92, 56–68. doi: 10.1016/j.neuroimage.2014.01.051

Sobczyk, O., Battisti-Charbonney, A., Poublanc, J., Crawley, A. P., Sam, K., Fierstra, J., et al. (2015). Assessing cerebrovascular reactivity abnormality by comparison to a reference atlas. *J. Cereb. Blood Flow Metab.* 35, 213–220. doi: 10.1038/jcbfm.2014.184

Sobczyk, O., Crawley, A. P., Poublanc, J., Sam, K., Mandell, D. M., Mikulis, D. J., et al. (2016). Identifying significant changes in cerebrovascular reactivity to carbon dioxide. *AJNR Am. J. Neuroradiol.* 37, 818–824. doi: 10.3174/ajnr.a4679

Somogyi, R. B., Vesely, A. E., Preiss, D., Prisman, E., Volgyesi, G., Azami, T., et al. (2005). Precise control of end-tidal carbon dioxide levels using

sequential rebreathing circuits. *Anaesth. Intensive Care* 33, 726–732. doi: 10.1177/0310057x0503300604

Symon, L. (1969). The concept of intracerebral steal. *Int. Anesthesiol. Clin.* 7, 597–615. doi: 10.1097/00004311-196907030-00009

Tancredi, F. B., Lajoie, I., and 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. doi: 10.1002/jmri.24878

Webster, M. W., Makaroun, M. S., Steed, D. L., Smith, H. A., Johnson, D. W., and Yonas, H. (1995). Compromised cerebral blood flow reactivity is a predictor of stroke in patients with symptomatic carotid artery occlusive disease. *J. Vasc. Surg.* 21, 338–344; discussion344–335.

Willie, C. K., Macleod, D. B., Shaw, A. D., Smith, K. J., Tzeng, Y. C., Eves, N. D., et al. (2012). Regional brain blood flow in man during acute changes in arterial blood gases. *J. Physiol.* 590, 3261–3275. doi: 10.1113/jphysiol.2012.228551

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi: 10.1016/j.neuroimage.2014.01.060

Yonas, H., Smith, H. A., Durham, S. R., Pentheny, S. L., and Johnson, D. W. (1993). Increased stroke risk predicted by compromised cerebral blood flow reactivity. *J. Neurosurg.* 79, 483–489. doi: 10.3171/jns.1993.79.4.0483