Measuring vascular reactivity with breath-holds after stroke: a method to aid interpretation of group-level BOLD signal changes in longitudinal fMRI studies

Fatemeh Geranmayeh¹, Richard J.S. Wise¹, Robert Leech¹, and Kevin Murphy²

¹ Computational Cognitive and Clinical Neuroimaging Laboratory, Imperial College, Hammersmith Hospital, London, W12 0NN, United Kingdom
² CUBRIC, School of Psychology, Cardiff University, Cardiff, CF10 3AT, United Kingdom

Correspondence to:
Kevin Murphy,
CUBRIC, School of Psychology,
Park Place,
Cardiff,
CF10 3AT
United Kingdom
murphyk2@cardiff.ac.uk
Abstract

Blood oxygenation level dependent (BOLD) contrast fMRI is a widely used technique to map brain function, and to monitor its recovery after stroke. Since stroke has a vascular etiology, the neurovascular coupling between cerebral blood flow and neural activity may be altered, resulting in uncertainties when interpreting longitudinal BOLD signal changes. The purpose of this study was to demonstrate the feasibility of using a recently validated breath-hold task in patients with stroke, both to assess group level changes in cerebrovascular reactivity (CVR) and to determine if alterations in regional CVR over time will adversely affect interpretation of task-related BOLD signal changes. Three methods of analyzing the breath-hold data were evaluated. The CVR measures were compared over healthy tissue, infarcted tissue, and the peri-infarct tissue, both sub-acute (two weeks) and chronically (four months). In this cohort, a lack of CVR differences in healthy tissue between the patients and controls indicates that any group level BOLD signal change observed in these regions over time is unlikely to be related to vascular alterations. CVR was reduced in the peri-infarct tissue but remained unchanged over time. Therefore, although a lack of activation in this region compared to the controls may be confounded by a reduced CVR, longitudinal group-level BOLD changes may be more confidently attributed to neural activity changes in this cohort. By including this breath-hold based CVR assessment protocol in future studies of stroke recovery, researchers can be more assured that longitudinal changes in BOLD signal reflect true alterations in neural activity.

Keywords: vascular reactivity, stroke, fMRI, breath-hold
**Introduction**

Over recent decades, functional magnetic resonance imaging (fMRI) has become a useful tool when studying the recovery of brain function following a stroke or other focal brain injury. Interpretations of fMRI findings should take into account the limitations of this technique, and factors that influence the fMRI derived Blood Oxygen Level Dependent (BOLD) signal (Logothetis, 2008). BOLD contrast fMRI is not only dependent on increased oxygen metabolism due to increased neural activity since activation-related differences also depend on changes in cerebral blood flow (CBF) and cerebral blood volume (Buxton, 2012; Davis et al., 1998; Hoge et al., 1999). Since stroke has a vascular etiology, this may alter the normal neurovascular coupling between CBF and oxygen consumption (taken as representative of neural activity), with resulting uncertainty when interpreting BOLD signal changes.

Reductions in cerebrovascular reactivity (CVR), the response of cerebral blood vessels to a vasodilatory stimulus, are associated with factors that predispose to stroke (Gupta et al., 2012; Markus, 2001). Acute vascular changes as a result of stroke may further reduce CVR. Reductions in CVR will lead to alterations in neurovascular coupling; that is, a given change in neural activity will no longer be represented by the same BOLD signal change. Regional changes in CVR and BOLD signal have been observed in adults after stroke (Krainik et al., 2005). Combining BOLD fMRI with a vasoactive stimulus such as CO₂ provides a robust means of characterizing CVR (Shiino et al., 2003). In many studies of clinical populations, an increase in arterial CO₂ levels is achieved using administration of a vasodilatory gas through a face mask during scanning (Heyn et al., 2010). In patients with stroke, this may not always be desirable or feasible due to the discomfort of the experimental setup.

A simple method to measure BOLD CVR in patient groups using a breath-hold task has been developed (Bright and Murphy, 2013; Murphy et al., 2011). Breath-holding offers an alternative to gas administration, and provides a similar BOLD CVR measure (Kastrup et al., 2001). Until recently it had been assumed that the degree of cooperation to perform a breath-holding task would vary too greatly across a patient population to provide a reliable measure of CVR (Spano et al., 2013). However, Bright and Murphy (2013) have shown that by monitoring end-tidal CO₂ levels during a breath-hold task, robust and repeatable measures of CVR can be obtained, even if the breath-hold task is poorly performed. This opens the way to using such a task in a stroke population to account for CVR differences. These measures can, in turn, be used to adjust the BOLD signal to more accurately reflect the neural responses to stimuli in a task-related fMRI study (Murphy et al., 2011).

Changes in CVR over time may potentially be problematic, particularly in longitudinal fMRI studies investigating recovery after the onset of stroke (Saur, 2006; Ward et al., 2003). In the longitudinal study by Saur and colleagues (2006), investigating language recovery following a left hemisphere stroke, there was no activity in the left hemisphere language areas, within the same vascular territory as the infarct, during the first few days after the stroke. Although this was interpreted as a lack of language-related neural activity caused by the stroke, an alternative explanation is that the absence of BOLD signal was the result of altered cerebral haemodynamics rather than a decline in neural activity *per se*. For example, it has been shown that the morphology of the haemodynamic response function (HRF) in the left perisylvian area is altered in some aphasic patients following a stroke (Bonakdarpour et al., 2007). In another study, the characteristics of the HRF response to an auditory
comprehension task changed as patients progressed from the acute to the sub-acute phase of stroke (Altamura et al., 2009). In addition, major arterial stenoses result in a compensatory reduction in distal cerebrovascular resistance, which in turn will alter CVR. Thus, it has been demonstrated that the HRF ipsilateral to an internal carotid artery stenosis has different temporal characteristics to that in the contralateral hemisphere (Carusone and Srinivasan, 2002). Furthermore, large vessel cervical or cerebral artery steno-occlusive disease can induce competitive intra-cerebral redistribution of flow from territories with low vasodilatory reserve to those with high reserve (Sobczyk et al., 2014). Thus without taking into account acute and chronic changes in normal CVR, task-related BOLD signal changes over time, particularly in longitudinal studies of stroke, are prone to misinterpretation.

One fMRI study investigating aphasic stroke recovery has used a simple breath-hold task to account for differences in haemodynamic responsiveness in stroke patients (van Oers et al., 2010). One advantage of using the Bright and Murphy (2013) approach, compared to the method used in that study, is that it quantifies the end-tidal CO2 changes to the breath-holds. This compensates for variability in task performance across the patients, providing more reliable quantification of the CVR. The end-tidal CO2 measure provides an accurate model of the expected BOLD response for an individual patient, allowing for the different delays in the breath-hold response across varying brain regions. In addition, the use of paced breathing between breath-hold challenges in the Bright and Murphy (2013) method leads to a more consistent baseline condition, and thus a less variable measure (Scouten and Schwarzbauer, 2008). Finally, the Bright and Murphy (2013) approach uses end-expiration measures, avoiding the biphasic BOLD response that is observed in end-inspiration challenges (Thomason and Glover, 2008). This leads to a more repeatable starting point for the breath-hold challenge, resulting in improved reproducibility (Scouten and Schwarzbauer, 2008).

The purpose of the current study is to demonstrate the feasibility of this breath-hold technique in patients with stroke, and to determine whether, in this cohort, alterations in CVR in the peri-infarct tissue over time might result in misinterpretation of group-level task-related BOLD signal changes. Three methods of analyzing the data were investigated: GlobOpt, in which a single model of the BOLD response was derived from the end-tidal CO2 trace; VoxOpt, when the delay between the end-tidal CO2 and fMRI time series was allowed to vary on a voxel-wise basis to account for slower regional haemodynamic responses; and RHsig, a method by which a BOLD model was derived from the contralesional right hemisphere and does not require measurement of the end-tidal CO2. The CVR measures were compared in healthy, peri-infarct tissue, and infarcted tissue at two time points after the ictus, in the sub-acute and chronic phase. The peri-infarct tissue was specifically included in the analysis as some studies have attributed the level of BOLD signal change in this region to stroke recovery (Heiss et al., 1999; Heiss et al., 2006; Rosen et al., 2000; Saur et al., 2006). The results demonstrate the utility of this method to assess CVR in a cohort of patients with stroke, and suggest that inclusion of this protocol in longitudinal BOLD studies of stroke recovery will allow more confident interpretation of group results.
Methods

Participants

Forty-six stroke patients (referred to as the PT group) with left lateralized cerebral infarction were scanned (30 male, average age 61.2 years, range 26-79 years). Table 1 shows their demographic data, including cerebrovascular risk factors, details of the size and site of the stroke lesion, the mode and outcome of vascular imaging, and the timing of the fMRI scans. The majority of patients had embolic stroke that was either from a cardio-embolic source (e.g. atrial fibrillation, or patent foramen ovale), or artery-to-artery embolism (e.g. from the carotid tree). There were 18 patients with occlusive arterial disease either in the intracranial vessels (e.g. due to stenosis or fresh thrombus) or dissection in the carotid tree.

Twenty-four patients were scanned in the sub-acute phase post-stroke (SPPS, also referred to as Sess1) and 42 in the chronic phase post-stroke (CPPS, also referred to as Sess2) (Fig. 1a). Mean time after stroke for SPPS was 14 days (range 5-28 days) and for CPPS was 114 days (range 84-200 days). Additionally, 26 healthy volunteers (referred to as the HV group) were scanned (9 male, average age 56.5 years, range 37-78 years), 17 of whom were scanned again after approximately 100 days. There was no between-group differences in age (t-test \( P = 0.15 \)) or sex (Fisher's exact test \( P = 0.5 \)). The National Research Ethics Service Committee (West London) approved the study.

Scanning parameters

MRI data were obtained on a Siemens Magnetom Trio 3 Tesla scanner. Whole brain, dual-echo, BOLD GE-EPI images (TR=2s, TE1=13ms, TE2=31ms, voxel-size 3.5x3.5x3mm\(^3\), 36 slices) were collected during a breath-hold task (see below). Only the second echo was analyzed in this study. Quadratic shim gradients were used to correct for magnetic field inhomogeneities within the brain, and residual field artefacts were compensated after acquiring fieldmap images. A high-resolution 1mm\(^3\) isotropic T\(_1\)-weighted whole-brain structural image was obtained.

In addition, diagnostic Diffusion Weighted MRI (DWI) and apparent diffusion co-efficient (ADC) imaging obtained by the clinical stroke services at the time of each patient’s admission to hospital was available to assist with determination of the boundary of regional infarction (see below). DWI and ADC images were performed in a Siemens Verio 3 Tesla scanner via diffusion-weighted echoplanar imaging (EPI) sequences. Nineteen DWI axial slices were acquired at \( b = 0, 500, \) and 1000 s/mm\(^2\) covering the entire brain. Other imaging parameters were: TR/TE 4600/90 ms, field of view 267 x 267, matrix 192 x 192, slice thickness 5 mm, with 6.5 mm spacing. The ADC maps were acquired with identical parameters with \( b = 1000 \) s/mm\(^2\).

Breath-hold paradigm and end-tidal CO\(_2\) pre-processing

Each participant performed a breath-hold task during the BOLD imaging to measure vascular reactivity (Bright and Murphy, 2013; Murphy et al., 2011). The task consisted of 14s of natural breathing, followed by 16s paced breathing, then a 15s end-expiration breath-hold down to an unforced depth. After the breath-hold, a quick exhalation of residual air was
performed prior to a return to natural breathing, which allowed the measurement of end-tidal CO$_2$ increases as a result of the breath-hold. This cycle was repeated six times over the course of the scan. The participants were trained outside the scanner and instructed to proceed to each stage by a series of visual pictorial and written cues presented at the centre of a screen.

End-tidal CO$_2$ traces were recorded throughout the experiment at a sampling frequency of 200 Hz, via a nasal cannula attached to a Medrad Veris capnograph. The recorded end-tidal CO$_2$ was converted to units of mmHg. A home-grown peak-detection algorithm was used to determine end-tidal CO$_2$ values by determining the partial pressure of CO$_2$ at the end of each breath. A linear interpolation was made between the end-tidal CO$_2$ measure of the final breath before the breath-hold and the quick exhalation CO$_2$ measure after the breath-hold. After this interpolation, a scale invariant convolution with a HRF was performed so that the scaling of the end-tidal CO$_2$ regressors remains the same. Convolution with a standard double gamma variate HRF (Friston et al., 1998) was performed because it has been demonstrated that more variance can be explained in the BOLD breath-hold data (Murphy et al., 2011). In this way, the amplitude of the regressors reflects the true increase in end-tidal CO$_2$ so that a quantitative measure of CVR could be made.

**fMRI pre-processing**

The BOLD breath-hold data were corrected for motion using AFNI’s 3dvolreg program by performing a rigid body alignment of each volume to the middle volume (http://afni.nimh.nih.gov/afni/). Each voxel’s time series was converted to a %BOLD signal time series by dividing by the mean across time points. A global %BOLD time series was calculated for each participant by averaging over all voxels in the brain. Linear drifts were removed from the data. No smoothing was performed to avoid contamination of CVR values between different regions.

**Analyses of breath-hold data - GlobOpt, VoxOpt and RHsig**

With the aim of determining the best way to analyze the breath-hold data, three analyses were performed: Global Optimization (GlobOpt), Voxel-wise Optimization (VoxOpt) and Right Hemisphere Signal (RHsig) (see Fig. 1c). First, to remove delays between the BOLD data and the end-tidal CO$_2$ trace that are caused both by physiological processes (such as the time for alveolar diffusion of CO$_2$ in the lung, and the time for blood to travel between the lungs and brain) and by the experimental apparatus (delays along the long sampling line between the participant in the MR scanner and the capnograph), the delay was optimized to the corresponding global BOLD signal (see top panel in figure 1c). In practice, this was performed by temporally shifting the end-tidal CO$_2$ trace between -15s and +15s in 0.1s steps, to find the delay that explained the most variance in the global BOLD signal. This globally-optimized CO$_2$ trace then formed the regressor for the GlobOpt analysis; that is, it was fitted to each voxel’s time series with a general linear model (GLM) using AFNI’s 3dDeconvolve function. The beta weight from the regression for each voxel gives the breath-hold response in that voxel in units of %BOLD signal change per mmHg change in end-tidal CO$_2$. 
The VoxOpt analysis differed from the GlobOpt analysis by allowing the CO\textsubscript{2} delay to be optimized on a voxel-wise basis. In practice, a similar analysis was performed where the CO\textsubscript{2} trace was included as a regressor in a GLM. This was repeated for each delay of the CO\textsubscript{2} trace (-15s to +15s in 0.1s steps from the globally optimized delay value). On a voxel-wise basis, the delay at which the GLM explained the most variance in that voxel was chosen as the optimal delay. Again, the beta weight in that voxel for that delay gave the breath-hold response in units of %BOLD signal change per mmHg change in end-tidal CO\textsubscript{2}.

The final analysis, RHsig, provides a way of obtaining breath-hold response maps if the CO\textsubscript{2} trace is unavailable or of too poor a quality to use. If an area of healthy tissue can be defined, the average BOLD response across that region provides a good model of the expected breath-hold response that can be included as a regressor in a GLM. In the RHsig analysis, it was assumed that the average response over the entire right hemisphere is a good representation of the BOLD signal response to the breath-hold challenge. A similar approach was used by van Oers and colleagues (van Oers et al., 2010). This average was fitted to each voxel in a GLM, and the derived beta weight for that voxel was the breath-hold response in units of %BOLD (as a ratio of the %BOLD signal change of the RHsig, i.e., 100\% indicates a signal with the same amplitude as the RHsig time series).

Lesion, Peri-infarct and healthy tissue masks

Individual 3D lesion masks were manually delineated by a neurologist (FG) and verified by a second neurologist (RJSW). These were drawn one slice at a time, on T\textsubscript{1}-weighted images using FSLView (http://www.fmrib.ox.ac.uk/fsl/), with additional guidance from the diagnostic acute post-stroke DWI and ADC sequences that were available in 41 subjects. DWI sequences reveal the early infarct with high contrast, and were used to guide the boundaries of the lesion on the high resolution T\textsubscript{1}-weighted structural images (Saur et al., 2006; Mah et al., 2012). Similar lesion identification based on T\textsubscript{1}-weighted images is commonly used in functional neuroimaging studies that investigate recovery of cognitive functions after stroke, which can arguably be confounded by changes in CVR (Bates et al., 2003; Brownsett et al., 2014; Dronkers et al., 2004; Crinion et al., 2006, Teki et al., 2013, Saur et al., 2006). Identification of the lesion manually, is considered gold standard when evaluating the efficacy of automated lesion segmentation algorithms (Mah et al., 2014; Seghier et al., 2008). Session specific lesion masks were defined independently for each of the SPPS and CPPS scans.

Two patients had acute infarcts that included lesions (maximum volume 1.4 cm\textsuperscript{3}) in the right hemisphere in addition to the acute infarct in the left hemisphere (Fig. 1a). These were included in the acute lesion mask. Five patients had an additional remote and small chronic lesion (range of lesion volume: 0.09 – 30.6 cm\textsuperscript{3}) that was excluded from the acute lesion mask.

A schematic of the masks used in this study is shown in Figure 1b. The acute lesion mask was denoted L-Lesion since the acute stroke lesions were confined to the left hemisphere in the majority of cases, and predominantly left-lateralized in the few patients with bilateral acute strokes. A peri-infarct tissue mask, L-PeriInfarct, was defined for each patient by dilating the individual’s L-Lesion by 10mm and then removing voxels that were in L-Lesion. The remaining left hemisphere voxels outside L-Lesion and L-PeriInfarct constituted the L-
Healthy mask. Homologous masks were defined in the right hemisphere and were named \textit{R-Lesion}, \textit{R-PeriInfarct} and \textit{R-Healthy}. This was done by transforming the left hemisphere masks into Montreal Neurological Institute (MNI) standard space (see below), then flipping about the left/right plane, and finally transforming it back into individual subject space. \textit{R-Lesion}, \textit{R-PeriInfarct} and \textit{R-Healthy} contained unaffected tissue. Full left and right hemisphere masks comprising of all voxels (including those in the lesion and peri-infarct regions) in the hemisphere were also defined and called \textit{L-Hemi} and \textit{R-Hemi}, respectively. These were the only masks used for the \textit{HV} group.

The high-resolution structural images were registered to the MNI space using FMRIB's (http://www.fmrib.ox.ac.uk/fsl/) Linear Image Registration Tool (FLIRT) with 12 degrees of freedom and cost-function masking in order to avoid the known problem of stretching normal tissue to fill the infarct during standard registration (Brett et al., 2001). The final transformation matrix was used to transform each mask to standard space as stated above.

\begin{center}
\underline{Insert Figure 1}
\end{center}

\textit{Breath-hold response comparisons}

Breath-hold responses for each of the three analyses were averaged over each mask for comparison. However, to remove voxels in which no breath-hold response was observed (e.g. a white matter voxel in which blood volume is low), each mask was thresholded on the basis of both the individual participant and the analysis method. A liberal threshold of $R^2 = 0.03$ ($P = 0.05$, for the number of time points with no correction for multiple comparisons) was used. If the GLM did not explain this amount of variance in a voxel's data, the voxel was removed from the mask. This ensured that averages across the masks were not skewed by voxels in areas of post-stroke atrophy that may show no response. In each mask, both the percentage of voxels surviving this threshold and their average variance were calculated for all participants. Comparisons between analysis methods and across masks were made using paired and unpaired $t$-tests. Specifically, differences in the values in each mask across the analysis methods were tested using paired $t$-tests. Similarly, differences within analysis method but across the masks were tested using paired $t$-tests. Unpaired $t$-tests were used to test differences between the \textit{HV} and \textit{PT} groups in the \textit{L-Hemi} masks. These results are presented in Figure 2 and discussed below.

The average voxel-wise delays across the masks, calculated using the \textit{VoxOpt} analyses, were also compared for both the \textit{SPPS} and \textit{CPPS} time points. The average delay difference was calculated in the \textit{HV L-Hemi}, \textit{PT L-Healthy}, \textit{PT L-PeriInfarct} and \textit{PT L-Lesion} masks with the corresponding delay in the homologous mask as the baseline. In this way, no difference in delay between the right and left hemisphere yields a value of 0. T-tests against 0 determined whether the delays were significantly different between the hemispheres. The results are shown in Figure 3 and discussed below.

CVR and \%BOLD change responses were compared across left and right masks and SPPS and CPPS time points for both the \textit{VoxOpt} and \textit{RHsig} analyses, respectively. Paired $t$-tests were used to determine significant differences between masks within analysis methods and
between time points. Unpaired t-tests were used to compare the PT and HV groups. The results are shown in Figure 4 and discussed below.

Finally, to determine whether the breath-hold responses in the PT group change over time from SPPS to CPPS, the Sess1 and Sess2 results from the 20 patients that provided data for both phases were compared. Paired t-tests were used to compare CVR results in each of the masks across the time points. These results are shown in Figure 5 and are discussed below.

Results

Lesion distributions

Stroke patients with recent left lateralized infarcts were chosen to take part in the study. Figure 1a shows the distribution of the lesions across the participants at each of the two phases, SPPS and CPPS. For the 24 patients scanned at SPPS, the mean percentage of left hemisphere voxels deemed to be in $L$-Lesion was 1.6%, maximum 9%. $L$-PeriInfarct comprised a mean of 7% of left hemisphere voxels, maximum 22%. In the 42 patients scanned at CPPS, $L$-Lesion comprised a mean of 3.2% of left hemisphere voxels, maximum 21.2%, and $L$-PeriInfarct 10.8%, maximum 28.8%. In the 20 patients scanned at both SPPS and CPPS, the mean lesion size reduced from 1.78% to 1.28% of left hemisphere voxels, a small but significant difference ($t = 3.2, P = 5 \times 10^{-3}$).

Motion related to the breath-hold task

The average scan motion was calculated as the mean frame-wise displacement (FWD) using the method described by (Power et al., 2012). Across both sessions, the mean FWD was $0.21 \pm 0.13$ in the PT group and $0.17 \pm 0.08$ in the HV group. An insignificant trend ($P = 0.06$) towards greater motion in the PT group versus the HV group was observed.

End-tidal CO$_2$ and the global BOLD signal

For the GlobOpt analysis, the delay at which the CO$_2$ trace explained the most variance (as measure by $R^2$) in the BOLD signal was chosen (see Fig. 1c top panel). The average $R^2$ did not differ in either the PT group or the HV group between the first and second time points ($PT$ Sess$1 = 0.38 \pm 0.28$; $PT$ Sess$2 = 0.38 \pm 0.24$; $HV$ Sess$1 = 0.58 \pm 0.21$; $HV$ Sess$2 = 0.47 \pm 0.22$). However, more global signal variance was explained by the CO$_2$ trace in the HV group versus the PT group ($t = 3.2, P = 2 \times 10^{-3}$, combining the data across both sessions).

Analysis methods comparison

Three types of breath-hold analyses were compared: GlobOpt, VoxOpt and RHsig. Figure 2 shows the relative performance of each of the methods by plotting the percentage of voxels that passed threshold in each mask (top panel), and the average variance explained in each mask across the participants for Sess1 (similar results were observed in Sess2). Results from the HV group are shown with grey bars. The percentage of voxels that passed
threshold in left hemisphere masks, L-Hemi, in this group was 53.1 ± 8.9% for the GlobOpt method, increasing to 68.6 ± 7.4% for the VoxOpt analysis and 64.1 ± 5.7% for the RHsig method. Both the VoxOpt and RHsig values were significantly greater than the GlobOpt values ($P < 10^{-7}$ and $P = 9 \times 10^{-5}$, respectively). In the HV group, the variance explained in L-Hemi voxels was 13.7 ± 3.7%, 14.3 ± 3.6% and 17.9 ± 3.4% for the GlobOpt, VoxOpt and RHsig analysis methods, respectively. Again, both the VoxOpt and RHsig values were significantly greater than the GlobOpt values ($P < 10^{-7}$ and $P = 9 \times 10^{-5}$, respectively). In the HV group, the variance explained in L-Hemi voxels was 13.7 ± 3.7%, 14.3 ± 3.6% and 17.9 ± 3.4% for the GlobOpt, VoxOpt and RHsig analysis methods, respectively. Again, both the VoxOpt and RHsig values were significantly greater than the GlobOpt values ($P < 10^{-7}$ and $P = 9 \times 10^{-5}$, respectively). In the PT group, L-Hemi included not only tissue not directly affected by ischemic damage but also the lesion and the peri-infarct masks. Since this may have caused the lower values in the PT group compared to the HV group, L-Healthy, L-PeriInfarct and L-Lesion were also compared in Fig. 2. For all three analyses, the number of voxels that passed the threshold in each mask, and the variance explained, reduced from L-Healthy to L-PeriInfarct and then to L-Lesion. Only in the GlobOpt analysis was the reduction in the number of voxels significant (L-Healthy versus L-PeriInfarct, $P = 4.7 \times 10^{-4}$; L-Healthy versus L-Lesion, $P = 1.7 \times 10^{-4}$). For the measure of explained variance, all analyses showed significant reduction from L-Healthy to the other two masks. For example, in the VoxOpt analysis, the variance explained in L-Healthy was 11.3 ± 4.1%, which reduced significantly to 9.5 ± 4.1% in L-PeriInfarct ($P = 1.7 \times 10^{-3}$) and to 8.6 ± 6.1% in L-Lesion ($P = 3.3 \times 10^{-5}$).

Similar trends were observed in left hemisphere mask, L-Hemi, of the PT group, with significantly less voxels passing threshold and significantly less variance explained with the GlobOpt analysis than both the VoxOpt and the RHsig analyses methods (% voxels: $P < 10^{-7}$ and $P = 2.8 \times 10^{-5}$, respectively; variance explained: $P = 1.4 \times 10^{-3}$ and $P = 1.2 \times 10^{-5}$, respectively). However, in all three analyses methods, the results were lower in L-Hemi in the PT group compared to the HV group. Less voxels passed threshold in L-Hemi of the PT group compared to the HV group (GlobOpt, $P = 2 \times 10^{-3}$; VoxOpt, $P = 4 \times 10^{-4}$; and RHsig, $P = 9 \times 10^{-3}$). A trend was observed for more variance explained in the HV compared to the PT group for the GlobOpt analysis ($P = 0.014$), with significantly more variance explained in both the VoxOpt and RHsig analyses ($P = 8.6 \times 10^{-3}$ and $P = 7.2 \times 10^{-3}$, respectively). A similar comparison between PT and HV in the R-Hemi mask, in which areas should be unaffected, also shows decreases. Significantly less voxels pass threshold in the PT group for GlobOpt and VoxOpt analyses ($P = 9.4 \times 10^{-3}$ and $P = 2.0 \times 10^{-3}$). A trend for a reduction in the amount of variance explained in the PT is observed in the R-Hemi masks for all analyses methods, with VoxOpt reaching significance ($P = 3.5 \times 10^{-2}$). This suggests that voxel time series are noisier in the PT group than the HV group.

In the PT group, L-Hemi included not only tissue not directly affected by ischemic damage but also the lesion and the peri-infarct masks. Since this may have caused the lower values in the PT group compared to the HV group, L-Healthy, L-PeriInfarct and L-Lesion were also compared in Fig. 2. For all three analyses, the number of voxels that passed the threshold in each mask, and the variance explained, reduced from L-Healthy to L-PeriInfarct and then to L-Lesion. Only in the GlobOpt analysis was the reduction in the number of voxels significant (L-Healthy versus L-PeriInfarct, $P = 4.7 \times 10^{-4}$; L-Healthy versus L-Lesion, $P = 1.7 \times 10^{-4}$). For the measure of explained variance, all analyses showed significant reduction from L-Healthy to the other two masks. For example, in the VoxOpt analysis, the variance explained in L-Healthy was 11.3 ± 4.1%, which reduced significantly to 9.5 ± 4.1% in L-PeriInfarct ($P = 1.7 \times 10^{-3}$) and to 8.6 ± 6.1% in L-Lesion ($P = 3.3 \times 10^{-5}$).

Delays in breath-hold responses relative to the contralesional hemisphere

The VoxOpt analysis allowed for a variable delay between the CO₂ trace and each voxel’s time series, independent of other voxels. Fig. 3 shows delays in the left hemisphere masks compared to their homologous mask in the right hemisphere. No difference in delays is
defined as 0 on this graph. In the HV group, who were also scanned at two time points, the relative delay in L-Hemi was -0.03 ± 0.67s and -0.15 ± 0.71s for the Sess1 and Sess2, respectively. Neither was significantly different from zero. Similarly, the response in L-Healthy in the PT group was not significantly delayed relative to the homologous right hemisphere region, with mean delays of 0.03 ± 0.89s and -0.02 ± 0.87s for the Sess1-SPPS and Sess2-CPPS data, respectively. The responses in L-PeriInfarct and L-Lesion showed a significant delay compared to their homologous right hemisphere regions at Sess1-SPPS, with delays of 1.14 ± 2.4s (P = 0.03) and 1.94 ± 4.34s (P = 0.04), respectively. Likewise at Sess2-CPPS, the responses in L-PeriInfarct and L-Lesion were also significantly delayed compared to their right homologues with delays of 1.24 ± 2.14s (P = 5x10^-4) and 1.83 ± 3.27s (P = 7x10^-4), respectively. This finding demonstrates the need to account for regional variability in delays to the CO₂ trace in stroke-affected areas during breath-hold analyses. These delays do not appear to recover over the time periods tested in this study.

Breath-hold responses across all participants

The average breath-hold responses across all participants are displayed in Fig. 4 for the VoxOpt (top panel) and RHsig (bottom panel) analyses. For the HV group, values are shown in L-Hemi and R-Hemi for both Sess1 and Sess2. For the PT group, both Sess1-SPPS and Sess2-CPPS, the results of the ipsilesional mask are displayed in grey bars (L-Healthy, L-PeriInfarct and L-Lesion) and results from the right hemisphere are depicted in white bars (R-Healthy, R-PeriInfarct and R-Lesion).

In the HV group, similar breath-hold response levels were observed in both hemispheres and across sessions. For the VoxOpt analysis, Sess1 showed mean responses of 0.32 ± 0.09 and 0.31 ± 0.09, and the corresponding values for Sess2 were 0.30 ± 0.07 and 0.29 ± 0.06, for R-Hemi and L-Hemi, respectively (units: %BOLD per mmHg change in end-tidal CO₂). For the RHsig analysis, the average responses were 147 ± 11 and 141 ± 12 for Sess1 and 149 ± 12 and 142 ± 15 for Sess2 (units: %BOLD). Although the differences were small between hemispheres, they were significant for the RHsig analysis in both Sess1 (P = 2.1 x 10^-4) and Sess2 (P = 9.6 x 10^-3). This might be expected, because the regressor for the RHsig analysis was drawn from R-Hemi, and therefore was a slightly better fit to the voxels in that hemisphere. However, a similar significant difference was also observed for the VoxOpt analysis in both Sess1 (P = 5.9 x 10^-3) and Sess2 (P = 0.023).

In the PT group, similar levels of response compared to the HV group were observed in the healthy tissue masks. For the VoxOpt analysis, the response in L-Healthy was 0.29 ± 0.12
and $0.30 \pm 0.10$ for Sess1 and Sess2, respectively (%BOLD per mmHg). The corresponding results for the \textit{RHsig} analysis were $143 \pm 30$ and $143 \pm 17$ (%BOLD). For the two methods, there was a drop in response when going from \textit{Healthy} to \textit{PeriInfarct} to \textit{Lesion} in both the stroke-affected left hemisphere masks as well as the contralesional right hemisphere masks. For example, at SPPS with the VoxOpt method, the response dropped from $0.29 \pm 0.12$ to $0.21 \pm 0.09$ ($P = 4 \times 10^{-5}$), and to $0.19 \pm 0.17$ (although not significantly, $P = 0.1$) in \textit{L-Healthy}, \textit{L-PeriInfarct} and \textit{L-Lesion}. Similar declines from $0.30 \pm 0.12$ to $0.23 \pm 0.10$ ($P = 5.3 \times 10^{-4}$) and then to $0.19 \pm 0.11$ ($P = 2 \times 10^{-3}$) were also observed in the right hemisphere. This demonstrated that the definition of the masks changed the response levels on the supposedly unaffected right hemisphere side.

In the comparison between left and right hemisphere masks, the \textit{RHsig} method appears to have distinguished the affected ipsilesional from the unaffected contralesional masks better, because the observed decline in response was greater on the left side (assuming that we would expect a decline in CVR on the lesioned side). For example, at CPPS the response difference was $156 \pm 16$ in \textit{R-Healthy} compared to $113 \pm 61$ in \textit{R-Lesion}, but on the affected side the difference was much greater, $142 \pm 17$ compared to $75 \pm 56$. Although, the left/right differences in the VoxOpt analysis appear small, paired t-tests showed differences in \textit{Healthy} (Sess1, $P = 1.1 \times 10^{-3}$; Sess2, $P = 8.3 \times 10^{-4}$) and \textit{PeriInfarct} (Sess1, $P = 7.8 \times 10^{-4}$; Sess2: $P = 9.5 \times 10^{-4}$) but not \textit{Lesion} in both sessions. The same comparisons in the three masks for the \textit{RHsig} analysis showed more significant differences: $P = 6.4 \times 10^{-4}$, $P = 4 \times 10^{-4}$ and $P = 0.015$ for Sess1-SPPS; $P < 10^{-7}$, $P < 10^{-7}$ and $P < 3.6 \times 10^{-3}$ for Sess2-CPPS, respectively.

\hspace{1cm} Insert Figure 5

\hspace{1cm} \textit{Longitudinal changes in breath-hold responses in 20 patients with two scans}

To determine whether reduced CVR recovers over time, Sess1-SPPS and Sess2-CPPS were compared in the 20 patients who were scanned twice (Figure 5). For the VoxOpt analysis method (top panel) at the SPPS time point (grey bars), there was no significant difference between the mean responses in the \textit{R-Hemi} and \textit{L-Healthy} masks ($0.29 \pm 0.13$ and $0.28 \pm 0.13$, respectively). The breath-hold responses were significantly reduced to $0.20 \pm 0.10$ in the \textit{L-PeriInfarct} mask ($p = 6.7 \times 10^{-3}$). Although the \textit{L-PeriInfarct} had a similar response to the \textit{L-Lesion} mask ($0.19 \pm 0.18$), due to a large standard deviation of the \textit{L-Lesion} responses, the difference between the \textit{L-Lesion} and \textit{L-Healthy} was not significant ($P = 0.093$). Almost identical values were observed for all four masks at CPPS (white bars). Importantly there was no significant difference between SPPS and CPPS in all the masks ($P > 0.5$).

Using the \textit{RHsig} analysis (bottom panel of Figure 5) at the SPPS time point (grey bars), the breath-hold response in the \textit{L-Healthy} mask was significantly less than that in \textit{R-Hemi} mask ($134 \pm 29$ and $153 \pm 23$, respectively: $P = 2.8 \times 10^{-3}$). Again, since the \textit{RHsig} draws its regressor from the right hemisphere, this may not be surprising. There was a significant reduction from the \textit{L-Healthy} response to $96 \pm 56$ in the \textit{L-PeriInfarct} mask ($P = 8.1 \times 10^{-4}$).
and to $75 \pm 99$ in the L-Lesion mask ($P = 6.5 \times 10^{-3}$). As with the VoxOpt analysis, no differences existed between responses in the SPPS and CPPS time points ($P > 0.2$), demonstrating that reductions in the breath-hold response in the Lesion and PeriInfarct masks did not recover over the time course of this study.

**Results summary**

The VoxOpt and RHsig analyses fit the breath-hold data better than the GlobOpt analysis (Fig. 2). Delays in the breath-hold response in the L-PeriInfarct and L-Lesion masks are observed in the PT group at both the SPPS and CPPS time points (Fig. 3). At both time points, the CVR measures decrease from the Healthy to the PeriInfarct to the Lesion regions (Fig. 4). The decreased CVR in the L-PeriInfarct and L-Lesion masks does not recover over time (Fig. 5).

**Discussion**

Erroneous interpretation of BOLD fMRI group results can be caused by stroke-induced changes in cerebrovascular reactivity (CVR). Bright and Murphy (2013) have previously established that CVR can be reliably quantified with breath-hold tasks, even when poorly performed, demonstrating that this approach is useful in patients who may not comply fully with the task. Since stroke has a vascular etiology, changes to CVR, especially within the infarct and the peri-infarct tissue, might be expected, making this a pathology for which this simple measure of CVR may be well utilized.

In the present study, patients with left hemisphere infarcts were examined. In the sub group of patients that were scanned twice, SPPS and CPPS time points were approximately 4 months apart. A steady state of recovery is assumed at the CPPS time point since it has been shown that clinically, maximum recovery has been reached after 3 months, beyond which the rate of recovery plateaus and very little further recovery occurs ((Laska et al., 2001; Pedersen et al., 1995; Ward et al., 2003).

In all patients, masks of four separate regions were defined: the lesion itself (L-Lesion); the peri-infarct tissue (L-PeriInfarct); the regions in the left hemisphere unaffected by stroke (L-Healthy); and the unaffected contralesional right hemisphere (R-Hemi). The peri-infarct tissue was arbitrarily defined as a 10mm wide region around the lesion. This mask was included in the analysis as some studies have attributed activity or lack of activity, in the peri-infarct tissue to stroke recovery (Heiss et al., 1999; Heiss et al., 2006; Rosen et al., 2000; Saur, 2006; Ward et al., 2003). Although this mask was arbitrarily defined and may not fully overlap with affected areas, it was included as an example of a region that might have reduced CVR. In a BOLD study investigating recovery after stroke, any other region could be analysed in a similar way to determine if CVR changes are likely to cause interpretability problems for that region.

One aspect of the lesion and peri-infarct BOLD data that might change compared to healthy tissue is the response delay to the breath-hold. Previously, studies have used the end-tidal CO$_2$ trace as a regressor in a GLM (Bright and Murphy, 2013; Murphy et al., 2011) to produce a CVR measure that is comparable across subjects. If a region has a delayed
response that is unaccounted for, the CVR measure might be artificially reduced. For this reason, we performed two methods of analysis with the CO₂ trace: GlobOpt and VoxOpt. The GlobOpt method represents the standard analysis for breath-hold data in which the same regressor is used for all voxels. The VoxOpt analysis allows for a voxel-wise delay. Fig. 3 demonstrates that the delay between the responses within the peri-infarct and lesion tissue response compared to the responses in their homologous regions on the contralateral side can be appreciable (1.24 ± 2.14s (P = 5x10⁻⁴) and 1.83 ± 3.27s (P = 7x10⁻⁴), respectively). This can explain why less voxels passed threshold in GlobOpt analysis versus the VoxOpt analysis (see Fig. 2). For this reason, the VoxOpt analysis was considered superior to the GlobOpt analysis, which was dropped from further study.

The VoxOpt analysis requires an end-tidal CO₂ trace, which may not be available in all situations. In addition, patients may have difficulties using the nasal cannula, especially if they prefer to breathe through the mouth, or the equipment may not be available in all settings. The finding that more global signal variance is explained by the end-tidal CO₂ trace in the HV group versus the PT group suggests that the end-tidal CO₂ recordings in the patient group may be less reliable. For this reason, the RHsig analysis, in which the model of breath-hold signal change is derived from the contralesional hemisphere, was also compared. A similar number of voxels passed threshold with this method compared to the VoxOpt method. However, the RHsig method also explained significantly more variance in those voxels that passed threshold (Fig. 2). Perhaps this is not so surprising since the regressor derived from the RHsig is more likely to resemble a BOLD response than an end-tidal CO₂ trace convolved with a standard HRF. The advantages of the RHsig method are that it is a better fit to the individual voxel BOLD responses, and it does not need end-tidal CO₂ traces that require equipment and setup time that may not result in accurate recordings. On the other hand, there are a number of disadvantages to the RHsig compared to the VoxOpt technique: it may not be truly comparable across subjects, since it is not a quantitative measure and, therefore, not suitable for correcting task-related BOLD responses; it requires an assumption of where unaffected regions lie (indeed, in this study, 7 of the R-Hemi masks were slightly contaminated by small lesions in the right hemisphere); it does not account for delays between the model and a voxel’s BOLD response; the model is an average over voxels that may have differing delays, and will therefore represent a more dispersed HRF; and finally, the CVR measures are biased towards the mask from which the regressor is derived, in this case R-Hemi, and so may not be reliable across the entire brain.

Since it was difficult to determine which analysis approach was superior, results from both VoxOpt and RHsig analyses were investigated further. Comparisons of the responses in healthy unaffected tissue in the PT group with responses in the HV group showed no differences (Fig. 4). Although the number of voxels passing threshold is less, this suggests the CVR measures are as reliable in the patient group as the healthy controls. In both analyses, the CVR measures reduces as the mask size decreases (from Healthy to PeriInfarct and then Lesion) on both the stroke-affected left hemisphere and the contralesional right hemisphere. It is possible that this represents a reduction in CVR due to the anatomical location of the masks. However, since the lesions are in largely heterogeneous locations (Fig. 1), it is more likely that this reduction is simply a function of mask size.
The interesting question is how the masks on the left side compare with their homologous masks on the right. A small but significant difference in CVR between the healthy tissue masks of the right and left hemisphere for the PT group might be indicative of increased cerebrovascular disease on the stroke-affected left side of the brain. However, since a similar small significant difference is observed between the L-Hemi and R-Hemi mask in the HV group, it is difficult to determine whether this is the case. For the peri-infarct mask compared to its homologue, small but significant differences were observed with the VoxOpt analysis with much larger and significant differences observed with the RHsig analysis. Only the RHsig analysis showed differences in CVR between the lesion and its homologue (Fig. 4). If we expect that CVR is reduced in the lesion, this might suggest that the RHsig is better at distinguishing CVR changes in that area. However, from Figure 3 we know that there was a delay in the BOLD response in the lesion that was not accounted for by the RHsig analysis, which might have lead to the reduced CVR measures in the lesion compared to its homologue. From Figure 2, we can observe that the number of voxels that passed threshold was reduced in the lesion but it is possible that these voxels had intact CVR, albeit delayed, which the VoxOpt analysis captured but the RHsig method could not. Although differences exist, both methods provide similar interpretations of the data.

It is important to note that a breath-hold challenge results in mild hypoxia alongside the intended hypercapnia. Since oxygen is vasoactive, this might directly influence the BOLD contrast (Bulte et al., 2012). A previous study has shown that breath-hold induced changes in end-tidal O₂ and CO₂ traces are significantly negatively correlated (Bright and Murphy, 2013), thus their respective effects on BOLD signal changes are difficult to disentangle. However, it has been shown that there is minimal impact on cerebral blood flow when end-tidal O₂ ranges from 60-150 mmHg (Brugniaux et al., 2007); all changes to the breath-hold task fall well within this range. A recent study comparing CVR measures derived from breath-hold paradigms and gas challenges demonstrated that mild hypoxia caused by breath-holds did not significantly alter CVR results compared to an iso-oxic CO₂ challenge (Tancredi and Hoge, 2013). These studies indicate that the mild hypoxia caused by the breath-hold challenge should not influence the CVR values in this study.

The peri-infarct region is important when studying stroke recovery. Many longitudinal BOLD fMRI studies focus on activity in the peri-infact tissue, to determine if neural activity has returned to what would be expected in healthy brains (Saur, 2006; Ward et al., 2003). Others have attributed activity in the peri-infarct region, or lack of it, to stroke recovery (Heiss et al., 1999; Heiss et al., 2006; Rosen et al., 2000). Alterations in the peri-infarct CVR over time, may adversely affect the interpretation of task-related BOLD signal change in fMRI studies of patients in the acute stages of stroke or in longitudinal studies. In the present study, for both the VoxOpt and RHsig analysis methods, the CVR in the peri-infarct tissue was shown to be reduced compared to the healthy tissue in the stroke-affected hemisphere, but did not change between the sub-acute and chronic phases after the ictus (Fig. 5). These findings suggests that in the cohort of patients studied, a relative lack of group-level task-related BOLD signal in the peri-infarct tissue compared to the healthy controls, may not accurately reflect neural deficits. However, since CVR in the peri-infarct tissue remains unchanged over time, a finding of increased group-level peri-infarct activity in the chronic phase in such a longitudinal study is less likely to be due to changes in vascular reactivity, and may be more reliably interpreted as a change in neural activity in the peri-infarct tissue after stroke.
Conclusions

This study demonstrates that CVR can be measured successfully in a stroke patient population using a breath-hold task. The resulting CVR measures can be used to disentangle vascular and neural changes caused by stroke, increasing confidence when interpreting group-level BOLD signal results in terms of neural activity (oxygen consumption). A lack of CVR differences in healthy tissue between patients and controls in this cohort suggests that task-related BOLD signal changes in these regions should be unaffected by the stroke and be more representative of task-related oxygen metabolism. In contrast, since CVR is reduced in the stroke peri-infarct tissue, a lack of BOLD signal in that area compared to controls may not accurately reflect neural deficits. However, the CVR in the peri-infarct tissue remains unchanged over time, suggesting that any group-level BOLD increases in this region over time in a longitudinal fMRI study are less likely to be due to an improvement in vascular reactivity, and may be more confidently attributed to changes oxygen metabolism caused by altered neural function. This study provides a framework for stroke researchers to account for CVR-related confounds in BOLD fMRI stroke studies. By including a similar breath-hold protocol in longitudinal studies of stroke recovery and performing similar analyses, researchers can more confidently interpret BOLD signal changes by disentangling vascular and neural influences using this method.

Acknowledgement

The authors thank all the patients and healthy participants who took part in this study. This work was supported by funding from the Wellcome Trust to FG [WT093957], and KM [WT090199].
Figures

a) Lesion distributions

b) Masks

c) Analyses

Figure 1: a) Lesion Distributions. The spatial distribution of the patient’s acute infarct lesions is shown for both the Subacute Phase Post Stroke (SPPS also referred to as Sess1) and Chronic Phase Post Stroke (CPPS also referred to as Sess2) time points. The colour-code of each voxel indicates how many patients had a lesion that included that voxel. The Venn diagram shows how many patients were scanned at each time point and the overlap between them. b) Masks. The schematic shows the masks used in the analysis. c) Analyses. The delay of each participant’s end-tidal CO2 trace was initially optimised by aligning it to the corresponding global BOLD signal time series. Three subsequent analyses types were performed. GlobOpt: the optimised CO2 trace was fitted to each voxel’s time series using a GLM. VoxOpt: the same CO2 trace was fitted to the voxel’s time series but was allowed to vary temporally to optimise the fit. RHsig: the average BOLD time series signal from the right hemisphere was fitted to the time series of every voxel in the brain.
Figure 2: The three analyses methods (GlobOpt, VoxOpt and RHsig) were compared across different masks for Sess1 (similar results were found for Sess2): the L-Hemi mask for both the HV and PT groups and the L-Healthy (LH), L-PeriInfarct (LP) and L-Lesion (LL) masks for the PT group. Grey bars depict the HV results and white bars the PT results (bars show means, error bars show standard deviations). The top panel shows the percentage of voxels in each mask that passed the liberal fitting threshold ($R^2 = 0.0288$, $P = 0.05$, no correction for multiple comparisons) averaged across subjects in each group. The bottom panel shows the across voxel average variance explained in each thresholded mask, averaged across the subjects in each group.
Figure 3: Using the VoxOpt analysis, the delay between each individual voxel time series and the end-tidal CO₂ time series can be calculated. With the assumption that the right hemisphere masks are largely unaffected by the stroke, the delay in the left hemisphere masks compared to their right homologous regions are shown for both the SPPS and CPPS time points (bars show means, error bars show standard deviations).
Figure 4: Breath-hold responses averaged across participants (bars show means, error bars show standard deviations). The top panel displays results from the VoxOpt analysis and the bottom panel shows results from the RHsig analysis. White bars show results from masks in the right hemisphere and grey bars show results from the left hemisphere masks. For the HV group, whole hemisphere masks were used. For the PT group at the two time points (Sess1-SPPS and Sess2-CPPS), results from three masks in each hemisphere are shown: Healthy, PeriInfarct and Lesion.
Figure 5: Breath-hold responses averaged across patients with a scans at both SPPS and CPPS time points (n = 20; bars show means, error bars show standard deviations). The top panel displays results from the VoxOpt analysis and the bottom panel shows results from the RHsig analysis. The grey bars show results from the first SPPS session and the white bars from the second CPPS session. In both analyses, there were no significant differences between the SPPS and CPPS responses in each of the 4 masks; R-Hemi, L-Healthy, L-PeriInfarct and L-Lesion.
References


Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878. doi:10.1038/nature06976


