Published in final edited form as:
Neuroimage. 2007 March ; 35(1): 175–184.

Reproducibility of BOLD, Perfusion, and $CMRO_2$ Measurements with Calibrated-BOLD fMRI

Oleg Leontiev[1,2] and **Richard B. Buxton**[1]

1 *Department of Radiology and Center for Functional MRI, University of California, San Diego*

2 *School of Medicine, University of California, San Diego*

Abstract

The coupling of changes in cerebral blood flow (CBF) and cerebral metabolic rate of oxygen ($CMRO_2$) during brain activation can be characterized by an empirical index, *n*, defined as the ratio between fractional CBF change and fractional $CMRO_2$ change. The combination of blood oxygenation level dependent (BOLD) imaging with CBF measurements from arterial spin labeling (ASL) provides a potentially powerful experimental approach for measuring *n*, but the reproducibility of the technique previously has not been assessed. In this study, inter-subject variance and intra-subject reproducibility of the method were determined. Block design %BOLD and %CBF responses to visual stimulation and mild hypercapnia (5% CO_2) were measured, and these data were used to compute the BOLD scaling factor *M*, %$CMRO_2$ change with activation, and the coupling index *n*. Reproducibility was determined for three approaches to defining regions-of-interest (ROIs): 1) Visual area V1 determined from prior retinotopic maps, 2) BOLD-activated voxels from a separate functional localizer, and 3) CBF–activated voxels from a separate functional localizer. For estimates of %BOLD, %$CMRO_2$ and *n*, intra-subject reproducibility was found to be best for regions selected according to CBF activation. Among all fMRI measurements, estimates of *n* were the most robust and were substantially more stable within individual subjects (coefficient of variation, CV=7.4%) than across the subject pool (CV=36.9%). The stability of *n* across days, despite wider variability of CBF and $CMRO_2$ responses, suggests that the reproducibility of blood flow changes is limited by variation in the oxidative metabolic demand. We conclude that the calibrated BOLD approach provides a highly reproducible measurement of *n* that can serve as a useful quantitative probe of the coupling of blood flow and energy metabolism in the brain.

Keywords

Cerebral Blood Flow (CBF); Cerebral Metabolic Rate of Oxygen ($CMRO_2$); Blood Oxygen Level Dependent (BOLD); fMRI; reproducibility

Introduction

The relationship between cerebral blood flow (CBF) and the cerebral metabolic rate of oxygen ($CMRO_2$) can be described by a single index *n*, defined as the ratio between the fractional change in CBF and the fractional change in $CMRO_2$ in response to functional activation. The blood oxygenation level dependent (BOLD) effect that is the basis for most functional magnetic

Address correspondence to: Richard B. Buxton, UCSD Center for functional MRI, 9500 Gilman Drive, Mail Code 0677, La Jolla, CA 92093-0677, Phone: (858) 822-0503, Fax: (858) 822-0605, E-mail: rbuxton@ucsd.edu

resonance imaging (fMRI) studies arises because of the larger change in CBF than $CMRO_2$ during activation, so the existence of the BOLD effect implies that $n>1$. Beyond this simple fact, however, our understanding of n is rather poor. Because of this, any quantitative interpretation of the magnitude of the BOLD effect is problematic: differences in n between brain regions or in association with disease could dramatically change the magnitude of the resulting BOLD signal despite similar underlying changes in neural activity and $CMRO_2$. Misestimates of n on the order of 30–50% can substantially confound any conclusions drawn about the underlying neural activity on the basis of BOLD data alone (Buxton 2002) . This effect is especially pronounced for $n<4$.

The variability of n in the reported literature is quite large. Positron emitted tomography (PET) studies, which measure CBF and $CMRO_2$ directly, provide the largest pool of data. Here, the results range from an n value of 0.9 (Roland et al. 1987) to $n=6$ (Fox and Raichle 1986). Calibrated-BOLD functional magnetic resonance imaging (fMRI) permits a non-invasive approach to measuring CBF and $CMRO_2$ (Davis et al. 1998;Hoge et al. 1999b). Through combined measurements of BOLD and CBF using arterial spin labeling (ASL) during periods of activation and mild hypercapnia, $CMRO_2$ can be calculated with a mathematical model of the BOLD signal. In this approach the hypercapnia data is used to define a local BOLD scaling parameter M, with the assumption that mild hypercapnia does not induce $CMRO_2$ changes (Jones et al. 2005;Sicard and Duong 2005). The scaling parameter M is proportional to the product of baseline deoxyhemoglobin content, echo time and a field strength proportionality constant and physiologically represents the maximal BOLD signal available upon washout of all deoxyhemoglobin (Davis, et al. 1998;Hoge et al. 1999a; (Buxton et al. 2004). The few studies using the calibrated-BOLD fMRI approach found n values ranging from 2–4 (Davis et al. 1998;Hoge et al. 1999a;Kastrup et al. 2002;Kim et al. 1999;St Lawrence et al. 2003;Stefanovic et al. 2004;Uludag and Buxton 2004;Uludag et al. 2004).

The question of whether these measurements are reproducible within single subjects needs to be addressed before making comparisons across subjects or between health and disease. A recent study found that in terms of relative signal change, CBF has lower inter-subject variation compared to BOLD, although intra-subject variation between sessions for CBF was not significantly different from BOLD (Tjandra et al. 2005). The aim of the current study was to quantify between-subject variability across the population and compare this with single-subject inter-session reproducibility for several fMRI measurements: %BOLD, %CBF, $\%CMRO_2$, M, n, and the cerebrovascular response to CO_2 (CRC). By measuring these two sources of variance in collected data, we can begin to evaluate the application of calibrated-BOLD to a population by quantifying the degree to which the variance in the population data can be attributed to measurement error. These experiments employed a simple flashing checkerboard visual stimulus with a block-design paradigm. Because these experiments necessarily involve analyzing data collected on different days, the method for choosing a region of interest (ROI) for averaging can potentially have a strong effect on the reproducibility of the measurement. We assessed the reproducibility for three approaches to defining an ROI: 1) area V1 determined by a previous retinotopy experiment; 2) BOLD-activated voxels determined in a previous functional localizer; and 3) CBF-activated voxels determined in a previous functional localizer.

Methods

10 healthy subjects (5 male, 5 female, 32 ± 8 years) were recruited and scanned in a 3 Tesla MR imaging system according to the guidelines set by the University of California San Diego (UCSD) Institutional Review Board (IRB). All subjects underwent a preliminary scan session on a different day where retinotopic mapping was performed. Each calibrated-BOLD imaging session lasted approximately 50 minutes and was performed twice for each subject at the same time of day to minimize effects of diurnal variations in baseline CBF. Approximately half of

the data acquired in this study was also used to evaluate biases in estimates of CBF-$CMRO_2$ coupling associated with voxel selection (Leontiev et al. submitted).

Experimental Task

The first portion of each calibrated-BOLD scan session was dedicated to functional activation and consisted of four imaging runs. The first of these runs was used exclusively to construct a functional localizer (a map of activated voxels) and the remaining runs were averaged to measure BOLD and CBF responses in the defined region of interest (ROI). For all functional runs, a block design paradigm was used with subjects viewing a radial black and white checkerboard flickering at 8Hz, for 4 repeated blocks (20 s on, 60 s off) in each experimental run. Additional baseline data were collected for 60 s before the first block and an additional 30 s after the last off period. During the off period, subjects viewed a gray screen with a white fixation square in the middle.

For the second portion of the experiment, subjects breathed a hypercapnic air mixture (5% CO_2, 21%O_2, 74%N_2) through a non-rebreathing face mask (Hans Rudolph 2700 Series, St. Louis, MI) for two runs (with 4 minutes between runs) without performing any task. Each CO_2 run consisted of 1 minute of breathing air, 2 minutes of breathing the CO_2 mixture, and another 3 minutes of breathing air. Physiological measures were recorded throughout all experiments using a PC located in the MR console room. Respiratory motions were measured with bellows around the lower portion of the thorax, pulse waveforms were measured with a fingertip pulse oximeter, and end tidal CO_2 levels were measured via a sampling port located inside a mixing chamber attached to the face mask worn by subjects.

Image Acquisition

Imaging data were acquired on a 3 Tesla whole body system (GE Excite, Milwaukee, WI) with an eight-channel receive-only head coil. For the first functional localizer activation run, data were collected with a PICORE (Wong et al. 1997) arterial spin labeling (ASL) sequence (TR 2s, TI 1400 ms, 20-cm tag width, and a 1-cm tag-slice gap) with a dual-echo gradient echo (GRE) readout and spiral acquisition of *k*-space (TE_1 9.4 ms, TE_2 30ms, flip angle 90º, FOV 24 cm, 64 x 64 matrix). This type of image acquisition allowed for simultaneous acquisition of perfusion and BOLD data. Four oblique 7mm-thick slices centered around the calcarine sulcus were acquired in a linear fashion from bottom to top. Data from this imaging run were used to construct a region of activation and were not otherwise used in determining functional response curves. For the remainder of the experiment quantitative ASL images were acquired with a PICORE QUIPSS II (Wong et al. 1998) sequence (TR 2.0 s, 180 repetitions, TI_1 700ms, TI_2 1400ms, tag thickness 20cm, 1-cm tag to proximal slice gap) with the same in-plane GRE parameters as the PICORE localizer run. Small bipolar crusher gradients were applied to both PICORE and QUIPSS II runs to remove signal from large vessels (b = 2 s/mm^2). The PICORE method is more sensitive than PICORE QUIPSS II for detecting CBF activation, but is less accurate for quantifying the CBF change. For this reason the PICORE method was used as a functional localizer only, while all of the reported changes are based on the QUIPSS II data. At the end of the experiment, a high-resolution structural scan was acquired with a magnetization-prepared 3D fast spoiled grass sequence (FSPGR)

Retinotopic Mapping

In a preliminary scan session on a different day, subjects were presented with standard stimuli for retinotopic mapping (Engel et al. 1997;Press et al. 2001). The flickering rate and the visual field eccentricity of these stimuli were identical to the stimulus presented during calibrated-BOLD experiments. During presentation of visual stimuli, images were acquired with an EPI sequence with the following parameters: TR 2 s, TE 30 ms, flip angle 90º, FOV 19cm, matrix 64 x 64, 3-mm isotropic resolution, 20 interleaved slices. The entire set of stimuli (meridian,

ring and wedge) yielded a single representation of V1. A high-resolution whole brain structural scan (3D FSPGR with 1-mm isotropic resolution) was acquired for each subject. The segmented cortical gray matter of occipital cortices was flattened using surface rendering methods described in (Wandell et al. 2000). Linear trends were removed from the datasets. Activation was assessed by correlating detrended data with the first harmonic of the stimulus variation frequency (Press et al. 2001). Representations of the primary visual area V1 of each subject was delineated on the computationally flattened visual cortex to define the retinotopic ROI for each subject. A combination of in-house Matlab (www.mathworks.com) and *mrLoadRet-1.0* code (http://white.stanford.edu/software/) was used for retinotopic mapping.

Calculation of M and $CMRO_2$

For $CMRO_2$ calculations, the model and methods described by Davis *et al*.,1998 were used. In this model, the fractional BOLD signal change ($\Delta S/S_o$) is related to the underlying changes in CBF and $CMRO_2$ by

$$\frac{\Delta S}{S_0} = M\left[1 - \left(\frac{CBF}{CBF_0}\right)^{\alpha-\beta}\left(\frac{CMRO_2}{CMRO_{20}}\right)^{\beta}\right] \quad [1]$$

The parameter *M* is the proportionality constant that reflects baseline deoxyhemoglobin content and defines the maximum possible BOLD signal change for that region. In the context of the model this parameter is proportional to the baseline blood volume fraction and O_2 extraction fraction. The parameter α is the exponent in an assumed power law relationship between cerebral blood flow and cerebral blood volume, and is taken to be α = 0.38 (Grubb et al. 1974;Mandeville et al. 1998). The parameter β was introduced as an empirical description of the signal reduction found in Monte Carlo simulation studies of spins diffusing near cylinders with altered magnetic susceptibility, and is taken to be β = 1.5 (Boxerman et al. 1995b;Davis et al. 1998). The parameters α and β are assumed to be global properties and all calculations are based on assumed values. Although we assume specific values for α and β, previous error analysis studies have found that the estimated physiological changes are not strongly dependent on the specific values used (Davis et al. 1998;Uludag et al. 2004). The parameter *M* is a local parameter estimated from the hypercapnia experiment using Equation [1] with the assumption that mild hypercapnia does not alter $CMRO_2$. Specifically, the ratio $CMRO_2/CMRO_{2o}$ was assumed to equal one during hypercapnia, and the measured changes in CBF and BOLD with hypercapnia within a specified region of interest (ROI) are combined with Equation [1] to calculate *M*. The derived value of *M* for each ROI was then applied to the activation data to calculate the stimulus-evoked change in O_2 metabolic rate $CMRO_2/CMRO_{2o}$. In addition to this primary use of the hypercapnia experiment to calibrate the BOLD effect, this data also provides a direct measure of the local vascular responsiveness, measured as the percent change in CBF divided by the absolute change in end-tidal CO_2.

Data Analysis

The flattened representation of the boundaries of V1 was rendered on a high-resolution anatomical volume. This volume was registered to the anatomical volume acquired in the calibrated-BOLD scan session using AFNI software with displacement and rotation parameters applied to the V1 representation to create a high resolution registered region of interest (ROI). This ROI was then undersampled to the resolution of the perfusion and BOLD images, including voxels that were at least 50% occupied by the high resolution V1 ROI, to define a V1 ROI on the spiral images of the activation run.

All images were coregistered using the AFNI software package (Cox 1996). The remainder of data processing was accomplished by using in-house Matlab code. The data from the first 8

seconds of each functional run was discarded to ensure steady-state conditions. Perfusion-weighted images were constructed by performing a running subtraction of consecutive control and tag images obtained from the first echo (Liu and Wong 2005). Similarly, BOLD-weighted images were constructed by performing a running average of consecutive control and tag images obtained from the second echo. Zero-order and linear drift were removed from the voxel time series. For physiological noise reduction, Fourier series components of phases corresponding to chest excursions and pulse waveforms were constructed and used as regressors in the BOLD and ASL time series using the modified RETROICOR approach for ASL (Restom et al. 2006). Voxel-wise perfusion (first echo) and BOLD (second echo) time series of the functional localizer run were individually correlated with a function representing our block-design paradigm convolved with a gamma-variate impulse response to create two 3D correlation coefficient (*r*) maps (Friston et al. 1994). Two ROIs were constructed from these maps to include voxels that exceeded a correlation coefficient of 0.5 from the perfusion data (CBF localizer) and BOLD data (BOLD localizer). This particular value for the correlation coefficient was arbitrary, but reasonable for reproducibility studies based on earlier studies (Rombouts et al. 1998).

Voxel-wise perfusion and BOLD data were averaged across functional activation PICORE QUIPSS II runs (runs 2,3,4) to improve signal-to-noise. For functional activation runs, average time courses of voxels for each of the four ROIs were calculated and normalized to a baseline value determined from the initial baseline period of each run and the final 10 seconds of each "off" period within each run and then averaged over 4 cycles of activation. In order to allow dynamic transients to pass, the average magnitude of the %BOLD and %CBF plateau were computed as the average of the final 12 seconds of the stimulus. Similarly, for the hypercapnia runs, raw CBF and BOLD data were averaged for each of the four ROIs and normalized to the 60s baseline period preceeding hypercapnia. The magnitudes of the %BOLD and %CBF responses were calculated as the average of the final 60 seconds of the hypercapnic period. The data from the CO_2 experiment was used to compute the maximal BOLD response, *M*, for each ROI.

%BOLD and %CBF responses from the functional activation portion of the experiment were used along with the calculated value of *M* to compute percent $CMRO_2$ changes. CBF-$CMRO_2$ coupling was characterized as the ratio *n* of the percent blood flow change (%CBF) to the percent change in oxygen metabolism (%$CMRO_2$). To compute vascular responsiveness to CO_2, an estimate of Cerebrovascular Response to CO_2 (CRC) was made by dividing the fractional blood flow change in percent by the absolute change in end-tidal CO_2 (%/ΔmmHg).

Reproducibility of each measurement for each ROI was computed for both inter and intra-subject data. The inter-subject data gives information about the variability of measurements across our subject pool. For this purpose, the coefficient of variation (CV_{inter}) was computed (Tjandra et al. 2005). The CV is defined as the standard deviation normalized to the mean of the task-related signal change across all subjects for each ROI, expressed as a percent. This allows adjustment for the difference in magnitude between measurements. For example, the average BOLD signal change is on the order of 2% while CBF is on the order of 70%.

Within-subject reproducibility of measurements was computed in a way that would allow direct comparisons to be made between intra-subject and inter-subject reproducibility. That is, the goal was to assess whether the observed variance of these measurements in the population could be explained just as the intrinsic variance of the methodology itself. To this end we calculated for each subject the fractional difference between the two measurements on different days, defined as a percentage difference relative to the mean:

$$\Delta = 100\frac{x_1 - x_2}{x_{avg}} \quad [2]$$

where x_1 and x_2 are measurements acquired on day 1 and day 2, respectively, and x_{avg} is the mean value of these measurements. The mean of this difference ($\langle\Delta\rangle$) provides information about any order effects that may exist across scan sessions for a single measurement (e.g., a greater BOLD response on day 1 than day 2), and the standard deviation (σ) of these Δ values provides information about the reproducibility of the measurement. However, for proper comparisons with the inter-subject CV defined above, the relevant value of CV for these intra-subject comparisons is not simply the standard deviation divided by the mean. Our goal is to define a representative reproducibility measure that is equivalent to asking how variable a single measurement should be compared to the true underlying mean of the population. By taking the individual differences of two measurements, each drawn from the same underlying probability distribution, we are effectively constructing samples from a distribution that is broader by a factor of the square root of 2 compared to the true reproducibility distribution. Because of this, we also calculate an appropriate CV as the ratio of the standard deviation to the mean of the set of Δ's divided by the square root of 2 (CV_{intra}). In order to test for significant differences of population variance between ROIs, an *F*-test (test of variance) was performed for the pooled data comparing V1 and BOLD localizer to the CBF localizer ROI. To determine whether Δ significantly differs from zero (the case where no bias in measurement exists between the two scan sessions), a two-tailed *t*-test was performed for each ROI comparing the pooled data acquired during the first scan session with the data acquired during the second scan session. To determine whether the variance of Δ differs between the ROIs, an *F*-test was performed comparing V1 and BOLD localizers to the CBF localizer distribution of Δ values. Lastly, for the ROI that gives the most stable measurements, an *F*-test was performed comparing the distribution of Δ values for each response metric to the Δ values for %BOLD.

Results

Figure 1 shows the ROIs generated for a typical subject. The BOLD localizer clearly has the largest volume of activation, reflecting superior SNR in the BOLD signal compared to the perfusion signal. Figure 2 demonstrates typical average response curves for a visual stimulus (A and B) and hypercapnia (C and D) in a single subject for the CBF localizer. Evidence of a $CMRO_2$ increase with activation is given by comparing the BOLD signal change to activation and hypercapnia for similar flow changes. Figure 3 illustrates average measurements across sessions for each subject in the BOLD and CBF localizer. The red line with a slope of 1 represents the ideal case of perfect reproducibility. It is evident that the reproducibility of several response indices is superior for the CBF localizer (black data points) compared to the BOLD localizer (red data points) with significant differences found for %BOLD, $\%CMRO_2$ and *n* measurements (Table 3). As such, all comparisons of reproducibility between different response measurements will be made for data generated with the CBF localizer ROI.

Reproducibility of hypercapnia data

Table 1 summarizes the results of intra-subject reproducibility of hypercapnia measurements in terms of the mean and standard deviation of the fractional difference between measurements acquired across different days ($\langle\Delta\rangle\pm$SD), and the coefficient of variation (CV_{intra}) between measurements acquired across different days. An asterisk next to $\langle\Delta\rangle$ denotes a significant order effect ($p<0.05$) between measurements across different days with positive values reflecting a larger measurement on day 1 and a negative value reflecting a larger measurement on day 2. The CBF response to hypercapnia has the poorest reproducibility of any response metric (CV=40.6%). However, when this measurement is normalized to the change in end-tidal CO_2 (CRC), reproducibility is improved (CV=23.7%), reflecting inconsistencies in the

administration of CO_2 (see bottom caption of Table 1). The BOLD response is more reproducible in individual subjects than the CBF response (CV=26.1%), perhaps as a result of noisy %CBF measurements for small flow changes. A significant order effect exists for %BOLD and %CBF measurements in area V1 with larger measurements on day 1 compared to day 2 (%BOLD, $\langle\Delta\rangle$ = 30.3±31.8%; %CBF $\langle\%\Delta\rangle$ = 43.6±50.4%). The choice of ROI has a non-significant effect on intra-subject reproducibility of BOLD and CBF responses to hypercapnia. Despite relatively poor reproducibility of BOLD and CBF responses to CO_2, calculated *M* values have a significantly improved reproducibility with a CV value roughly one-half the %BOLD and one-quarter the %CBF responses (*M*, CV_{intra} =12.3%).

Table 2 reports the mean, standard deviation, and coefficient of variation (CV_{inter}) of subject-pooled hypercapnia measurements. An asterisk next to CV_{inter} denotes a significant difference ($p<0.05$) in the variance of the data observed for the specified ROI compared to the CBF localizer as determined by an *F*-test. The mean values of the parameters differed depending on the choice of ROI, possibly reflecting effects such as the inclusion of draining veins. A systematic study of potential biases due to ROI selection has been reported as an abstract (Leontiev and Buxton 2006) and will be presented in a separate paper (Leontiev et al. submitted). For inter-subject data, the variability of CBF measurements is relatively insensitive across functional localizers. However, the pooled variance of the BOLD-response in area V1 and the BOLD localizer were significantly greater than the CBF localizer. Comparing corresponding CV values across Table 1 and Table 2, a greater variability is observed across the subject pool than the reproducibility of individual measurements across scan sessions. The reproducibility of *M* in individual subjects (CV_{intra} =12.3%) was nearly three times better than its pooled variance (CV_{inter} =36.1%) for the CBF localizer.

Reproducibility of activation data

In the same manner for reporting data as Table 1 (see above), Table 3 summarizes the results of intra-subject reproducibility of functional activation measurements as the mean and standard deviation of the fractional difference between measurements acquired across different days ($\langle\Delta\rangle$±SD), and the coefficient of variation (CV_{intra}) between measurements acquired across different days. An asterisk next to $\langle\Delta\rangle$ denotes a significant order effect ($p<0.05$) between measurements across different days with positive values reflecting a larger measurement on day 1 and a negative value reflecting a larger measurement on day 2. An asterisk next to CV_{intra} denotes a significant difference ($p<0.05$) in the reproducibility of individual measurements for the specified ROI compared to the CBF localizer. %BOLD and $\%CMRO_2$ reproducibility in the BOLD localizer was found to be significantly worse (%BOLD, CV_{intra} =20.2%; $\%CMRO_2$, CV_{intra} =22.6%) in comparison to the CBF localizer (%BOLD, CV_{intra}=11.7%; $\%CMRO_2$, CV_{intra} =10.7%). For the CBF localizer, the reproducibility of %CBF and %BOLD activation was found to be similar. However, a significantly lower CV value was found for *n* (CV_{intra}=7.4%) compared to the %BOLD-response (CV_{intra}=11.7%). The BOLD localizer gave the poorest reproducibility of *n* (CV_{intra}=20.8%), possibly reflecting its poor $\%CMRO_2$ reproducibility. Table 4 reports the variance of these measurements across the subject pool as the mean, standard deviation and coefficient of variation (CV_{inter}) of pooled measurements for all subjects in all sessions. For the pooled subject data, the variance of BOLD responses was significantly greater in area V1 (CV_{inter} =33.9%) and the BOLD localizer (CV_{inter} =24.8%) compared to the CBF localizer (CV_{inter}=16.1%). For calculated *n* values, area V1 was found to have significantly smaller inter-subject variability than the CBF localizer (CV_{inter}= 22.4% vs CV_{inter}= 36.9%).

Figure 4 compares individual-subject reproducibility (CV_{intra}) with the population variance (CV_{inter}) of measurements available with calibrated-BOLD fMRI for the CBF localizer. It is evident that the population variance of calibrated bold estimates of *M*, $CMRO_2$ and *n* in the

visual cortex cannot be explained by poor reproducibility of this technique. For this ROI, reproducibility of n was found to be significantly better ($p<0.05$) than the reprodubility of the BOLD-response. The inter-subject variation of the BOLD-response was found to be lower than all other measurements. This is likely due to suppression of the extra-vascular BOLD signal by diffusion gradients.

Discussion

The calibrated BOLD approach provides a potentially powerful tool for investigating the coupling of blood flow and oxygen metabolism in different brain structures and in different disease states. However, in order to draw conclusions about differences between response measurements of individual subjects or between patient populations, knowledge about the natural variance of these measurements in the healthy population and the reproducibility within individual subjects is essential. In addition to the intrinsic day-to-day variability of the underlying physiology in an individual and intrinsic noise in the measurements, the method used for choosing an appropriate ROI for averaging can also affect the reproducibility of the measurement. In these studies we did not attempt to produce a precise co-registration of the data sets collected on different days. Instead, we tested several approaches for defining an ROI for averaging based just on data collected separately on each day so that we could directly compare the measured reproducibility with the observed variation across the population. We constructed ROIs in the visual cortex based on different anatomical and functional contrast (retinotopy vs CBF or BOLD functional localizers) in order to test whether this has an effect on reproducibility. This study demonstrates that 1) calculated measurements (M and n) are more stable across days than direct measurements (%BOLD and %CBF) and 2) ROI selection plays a significant role in the reproducibility of individual measurements.

Ideally, we would like to quantify within-subject reproducibility as the standard deviation, σ', of a response measurement, x, around some true value in a single subject after an infinite number of repeated trials. Practical consideration places a limit on this type of experimental design, so instead we measured responses in 10 subjects on two different days and computed the fractional difference of individual measurements across different days (Δ) and determined the mean ($\langle\Delta\rangle$) and standard deviation (σ) of this value across our subject pool (Tables 1 and 3). However, because this difference measurement involves two samples from the desired population, the distribution of Δ is broadened by the square root of 2, and the values of CV reported are corrected for this. With this definition, the intra-subject CV's are directly comparable to the inter-subject CV's, in the sense that they should be equal if there is no variation in the population and all of the variability in the measurements is due to the intrinsic reproducibility of the measurement.

In order to evaluate any effects of habituation that may occur between two scan sessions, we tested whether the mean Δ deviates significantly from zero. The only measures that demonstrated a significant order effect were the BOLD and CBF responses in area V1 to hypercapnia with larger responses on day 1 for both measurements (Δ = 22–31 ± 37–57%). This result was likely due to a leak in the the non-rebreathing face mask during delivery of CO_2 for two subjects (subjects 6 and 7 in figure 3) on day 2 of scanning, resulting in end-tidal CO_2 increases of less than 5 mmHg. Although this will have the effect of degrading the reproducibility of the blood flow change or the BOLD change to the hypercapnic stimulus, it serves to demonstrate the stability of M at different levels of administered CO_2. This can be explained by the fact that the calculation of M is essentially a reflection of the BOLD-CBF ratio and therefore any inconsistencies in the delivery of CO_2-enriched air will offset both BOLD and CBF measurements.

BOLD reproducibility is influenced by changes in the baseline state caused by differences in physiologic or pharmacologic states (Brown et al. 2003;Cohen et al. 2002), or simply by differences in attention during different runs. The majority of BOLD reproducibility studies either examined: 1) the spatial reproducibility of the volume of activation (Rombouts et al. 1998;Tegeler et al. 1999) or 2) the stability of the dynamics of the hemodynamic response function (HRF) in the context of an event-related study (Aguirre et al. 1998;Handwerker et al. 2004). Our study differs from these previous studies in that it evaluates the utility of fMRI in generating robust functional response data that have direct physiological significance (%CBF, $\%CMRO_2$, M, n) rather than evaluating the robustness of fMRI in the context of brain mapping. A recent study by Tjandra and co-workers (Tjandra et al. 2005) looked at both the reproducibility of the spatial localization and percent signal change of BOLD and ASL data to a block-design stimulus. They found that for relative signal change, CBF had a lower inter-subject variation than BOLD. However, they found no significant differences in reproducibility between relative BOLD and CBF signal change within individual subjects across scan sessions. The CV value we report for intra-subject BOLD reproducibility in the CBF localizer (CV=11.7%) is smaller than the reported numbers in the Tjandra study (CV=20–30%). However, for the BOLD localizer, our CV value is more similar (CV=20.2%). Similarly for inter-subject data, we found a decreased variance of the BOLD-effect across the subject pool with the CBF localizer compared to either the BOLD or V1 localizers. We hypothesize that through the use of mild diffusion gradients and selecting voxels according to perfusion activation, both the inter-subject variability and the reproducibility of the relative BOLD signal change can be improved by reducing the signal contribution from structures that have the largest shifts in deoxyhemoglobin (draining veins and large venules). We found that $CMRO_2$ increases show similar reproducibility as CBF changes, suggesting that the reproducibility of blood flow changes is limited by variation in the oxidative metabolic demand across days.

The coupling of CBF and $CMRO_2$, described by the single variable n, was found to be the most reproducible measure with a coefficient of variation of roughly 7% for the flow localizer (Table 3). This finding supports the basic idea that day to day fluctuations in CBF and $CMRO_2$ responses to the same stimulus are coupled. Moreover, calculated responses ($\%CMRO_2$ and n) are relatively insensitive to any changes in baseline CBF across scan sessions (e.g., due to differences in caffeine intake or anxiety) because any differences in the baseline state will similarly scale BOLD-responses to both activation and hypercapnia. The intra-subject reproducibility of n, in part a reflection of the robustness of the measurement technique, was found to be much better than the variance of the measurement across the subject pool (CV_{intra} =7.4% vs. CV_{inter} =36.9%). We conclude, therefore, that the calibrated BOLD technique can confidently discriminate n measurements between individuals and that the large population variance of n in the visual cortex cannot be attributed to poor reproducibility. However, the large variance of this measurement in the visual cortex across subjects suggests that at least for population studies, a large subject pool must be recruited. Further studies are needed to understand the physiological source of the differences in n across a healthy population.

We found that the reproducibility of %BOLD, $\%CMRO_2$ and n measurements was significantly better in regions selected according to perfusion activation (CBF localizer) compared to BOLD activation (BOLD localizer) (Figure 3). The volume of activated tissue was found to be at least twice as large for the BOLD localizer for the same level of significance (r=0.5). In order to test whether 1) these two activated volumes of cortex correspond to spatially distinct areas and 2) that the benefits in reproducibility of selecting ROIs according to perfusion-activation is not merely an effect of thresholding, the statistical threshold for BOLD-activation was adjusted for each subject until the number of voxels was made equal to the number of voxels in the CBF localizer. We found that 1) the average percent overlap between these two regions was 50% (range 29–62%) and 2) reproducibility was degraded even further

for the threshold-adjusted BOLD localizer, especially for measurements of M, $\%CMRO_2$ and n. Figure 5 compares reproducibility of calibrated-BOLD measurements for the CBF localizer (r=0.5), BOLD localizer (r=0.5) and the threshold-adjusted BOLD localizer. The average r-value applied to BOLD-activation that produced the same number of voxels as the CBF localizer was found to be 0.7. Upon examination of Figure 5, it appears that the source of this phenomenon is poor reproducibility of the BOLD-response to hypercapnia. It is possible that this ROI is dominated by draining veins, leading a high variability of the response. Further work needs to be done to address the cause of the observed poor reproducibility of M for voxels chosen according to a very high level of functional BOLD activation.

It is clear that selecting a region-of-interest according to activation will likely ensure confident estimates of activation response characteristics. However, the experimenter is still left with the task of arbitrarily defining a minimum threshold of activation for potentially multiple types of functional contrast (flow, BOLD, intersection of flow and BOLD). With this in mind, we chose to perform retinotopic mapping on all our subjects in order to delineate area V1. Although determined functionally, area V1 represents a relatively constant, well-defined anatomical ROI. However, we found that in our hands the V1 ROI suffered from relatively poor reproducibility of measurements within individual subjects (Tables 1 and 3). A possible explanation for this finding is that we did not sub-mask this region for voxels activated by our full-field flickering checkerboard stimulus and therefore it is incorrect to assume that all voxels in this region demonstrate appreciable stimulus-modulated signal change. Additionally, acquiring the retinotopic maps during a different scan session introduces the potential for misalignment of ROIs due to imperfect registration. Lastly, since area V1 is projected onto a high-resolution anatomical volume, undersampling is necessary in order to apply these masks to functional data sets. In order to define V1 voxels on a 64x64 grid with 7mm slice thickness, a percentage cut-off for voxels contained in the high-resolution V1 ROI has to be defined.

Conclusions

The variability of calibrated-BOLD measurements across subjects, and the reproducibility of measurements in individual subjects across different days, were determined within the visual cortex for regions-of-interest based on prior retinotopic mapping, CBF activation and BOLD activation. Reproducibility of measurements within subjects across different days was best for regions defined according to CBF activation. The poorest reproducibility of any response metric was observed for BOLD and CBF changes to hypercapnia, likely due to inconsistencies of CO_2 administration. However, the BOLD scaling factor, M, was found to be significantly more reproducible than individual CBF and BOLD responses to hypercapnia. For the CBF localizer, similar reproducibility was observed for BOLD, CBF and $CMRO_2$ responses. However, the ratio n of CBF to $CMRO_2$ responses was observed to have the best reproducibility of any fMRI measurement. Two major conclusions can be drawn from this study: 1) the index n describing the coupling of blood flow and oxygen metabolism is a remarkably stable measurement in individual subjects and 2) selecting ROIs according to flow activation significantly improves the robustness of measurements available with fMRI. These conclusions support the use of calibrated fMRI as a quantitative probe of the coupling of blood flow and energy metabolism. In particular, this approach may greatly expand the role of fMRI methods in studies of disease and development where the complexity of the BOLD effect limits the quantitative interpretation of BOLD responses alone

Acknowledgements

We thank Giedrius Buracas for providing support with protocols for retinotopic mapping and Beau Ances for his ongoing support during preparation of this manuscript. We would also like to thank Khaled Restom, Yashar Behzadi and Thomas Liu for providing in-house Matlab code for general data processing, and David Dubowitz for help with the experimental setup and design. This work was supported by National Institutes of Health grants (RO1 NS36722-08 and RO1 NS42069-04) and the UCSD General Clinical Research Center grant MO1 RR000827.

References

Aguirre GK, Zarahn E, D'Esposito M. The variability of human, BOLD hemodynamic responses. NeuroImage 1998;8:360–376. [PubMed: 9811554]

Boxerman JL, Hamberg LM, Rosen BR, Weisskoff RM. MR contrast due to intravascular magnetic susceptibility perturbations. Magn Reson Med 1995b;34:555–566. [PubMed: 8524024]

Brown GG, Eyler Zorrilla LT, Georgy B, Kindermann SS, Wong EC, Buxton RB. BOLD and Perfusion Response to Finger-Thumb Apposition After Acetazolamide Administration: Differential Relationship to Global Perfusion. J Cereb Blood Flow Metab 2003;23:829–837. [PubMed: 12843786]

Buxton, RB. Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques. Cambridge: Cambridge University Press; 2002.

Buxton RB, Uludag K, Dubowitz DJ, Liu TT. Modeling the hemodynamic response to brain activation. Neuroimage 2004;23(Suppl 1):S220–233. [PubMed: 15501093]

Cohen ER, Ugurbil K, Kim SG. Effect of basal conditions on the magnitude and dynamics of the blood oxygenation level-dependent fMRI response. J Cereb Blood Flow Metab 2002;22:1042–1053. [PubMed: 12218410]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29:162–173. [PubMed: 8812068]

Davis TL, Kwong KK, Weisskoff RM, Rosen BR. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc Natl Acad Sci USA 1998;95:1834–1839. [PubMed: 9465103]

Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex 1997;7:181–192. [PubMed: 9087826]

Fox PT, Raichle ME. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci USA 1986;83:1140–1144. [PubMed: 3485282]

Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Human Brain Mapping 1994;1:153–171.

Grubb RL, Raichle ME, Eichling JO, Ter-Pogossian MM. The effects of changes in $PaCO_2$ on cerebral blood volume, blood flow, and vascular mean transit time. Stroke 1974;5:630–639. [PubMed: 4472361]

Handwerker DA, Ollinger JM, D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 2004;21:1639–1651. [PubMed: 15050587]

Hoge RD, Atkinson J, Gill B, Crelier GR, Marrett S, Pike GB. Investigation of BOLD signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model. Magn Reson Med 1999a;42:849–863. [PubMed: 10542343]

Hoge RD, Atkinson J, Gill B, Crelier GR, Marrett S, Pike GB. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. Proc Natl Acad Sci, USA 1999b;96:9403–9408. [PubMed: 10430955]

Jones M, Berwick J, Hewson-Stoate N, Gias C, Mayhew J. The effect of hypercapnia on the neural and hemodynamic responses to somatosensory stimulation. Neuroimage 2005;27:609–623. [PubMed: 15978844]

Kastrup A, Kruger G, Neumann-Haefelin T, Glover GH, Moseley ME. Changes of cerebral blood flow, oxygenation, and oxidative metabolism during graded motor activation. Neuroimage 2002;15:74–82. [PubMed: 11771975]

Kim SG, Rostrup E, Larsson HBW, Ogawa S, Paulson OB. Determination of relative CMRO2 from CBF and BOLD changes: significant increase of oxygen consumption rate during visual stimulation. Magn Reson Med 1999;41:1152–1161. [PubMed: 10371447]

Leontiev, O.; Buxton, RB. CBF/CMRO2 coupling measurements with calibrated BOLD-fMRI: Bias due to voxel selection. 14th meeting, International Society for Magnetic Resonance in Medicine; Seattle. 2006. p. 454

Leontiev O, Dubowitz DJ, Buxton RB. CBF/CMRO2 coupling measured with calibrated-BOLD fMRI: sources of bias. Neuroimage. submitted

Liu TT, Wong EC. A signal processing model for arterial spin labeling functional MRI. Neuroimage 2005;24:207–215. [PubMed: 15588612]

Mandeville JB, Marota JJA, Kosofsky BE, Keltner JR, Weissleder R, Rosen BR, Weisskoff RM. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. Magn Reson Med 1998;39:615–624. [PubMed: 9543424]

Press WA, Brewer AA, Dougherty RF, Wade AR, Wandell BA. Visual areas and spatial summation in human visual cortex. Vision Res 2001;41:1321–1332. [PubMed: 11322977]

Restom K, Behzadi Y, Liu TT. Physiological noise reduction for arterial spin labeling functional MRI. Neuroimage 2006;31:1104–1115. [PubMed: 16533609]

Roland PE, Eriksson L, Stone-Elander S, Widen L. Does mental activity change the oxidative metabolism of the brain? J Neuroc 1987;7:2373–2389.

Rombouts SA, Barkhof F, Hoogenraad FG, Sprenger M, Scheltens P. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. Magn Reson Imaging 1998;16:105–113. [PubMed: 9508267]

Sicard KM, Duong TQ. Effects of hypoxia, hyperoxia, and hypercapnia on baseline and stimulus-evoked BOLD, CBF, and CMRO2 in spontaneously breathing animals. Neuroimage 2005;25:850–858. [PubMed: 15808985]

St Lawrence KS, Ye FQ, Lewis BK, Frank JA, McLaughlin AC. Measuring the effects of indomethacin on changes in cerebral oxidative metabolism and cerebral blood flow during sensorimotor activation. Magn Reson Med 2003;50:99–106. [PubMed: 12815684]

Stefanovic B, Warnking JM, Pike GB. Hemodynamic and metabolic responses to neuronal inhibition. Neuroimage 2004;22:771–778. [PubMed: 15193606]

Tegeler C, Strother SC, Anderson JR, Kim SG. Reproducibility of BOLD-based functional MRI obtained at 4 T. Hum Brain Mapp 1999;7:267–283. [PubMed: 10408770]

Tjandra T, Brooks JC, Figueiredo P, Wise R, Matthews PM, Tracey I. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. Neuroimage 2005;27:393–401. [PubMed: 15921936]

Uludag K, Buxton RB. Measuring the effects of indomethacin on changes in cerebral oxidative metabolism and cerebral blood flow during sensorimotor activation. Magn Reson Med 2004;51:1088–1089. [PubMed: 15122696]author reply 1090

Uludag K, Dubowitz DJ, Yoder EJ, Restom K, Liu TT, Buxton RB. Coupling of cerebral blood flow and oxygen consumption during physiological activation and deactivation measured with fMRI. Neuroimage 2004;23:148–155. [PubMed: 15325361]

Wandell BA, Chial S, Backus BT. Visualization and measurement of the cortical surface. J Cogn Neurosci 2000;12:739–752. [PubMed: 11054917]

Wong EC, Buxton RB, Frank LR. Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. NMR in Biomed 1997;10:237–249.

Wong EC, Buxton RB, Frank LR. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). Magn Reson Med 1998;39:702–708. [PubMed: 9581600]

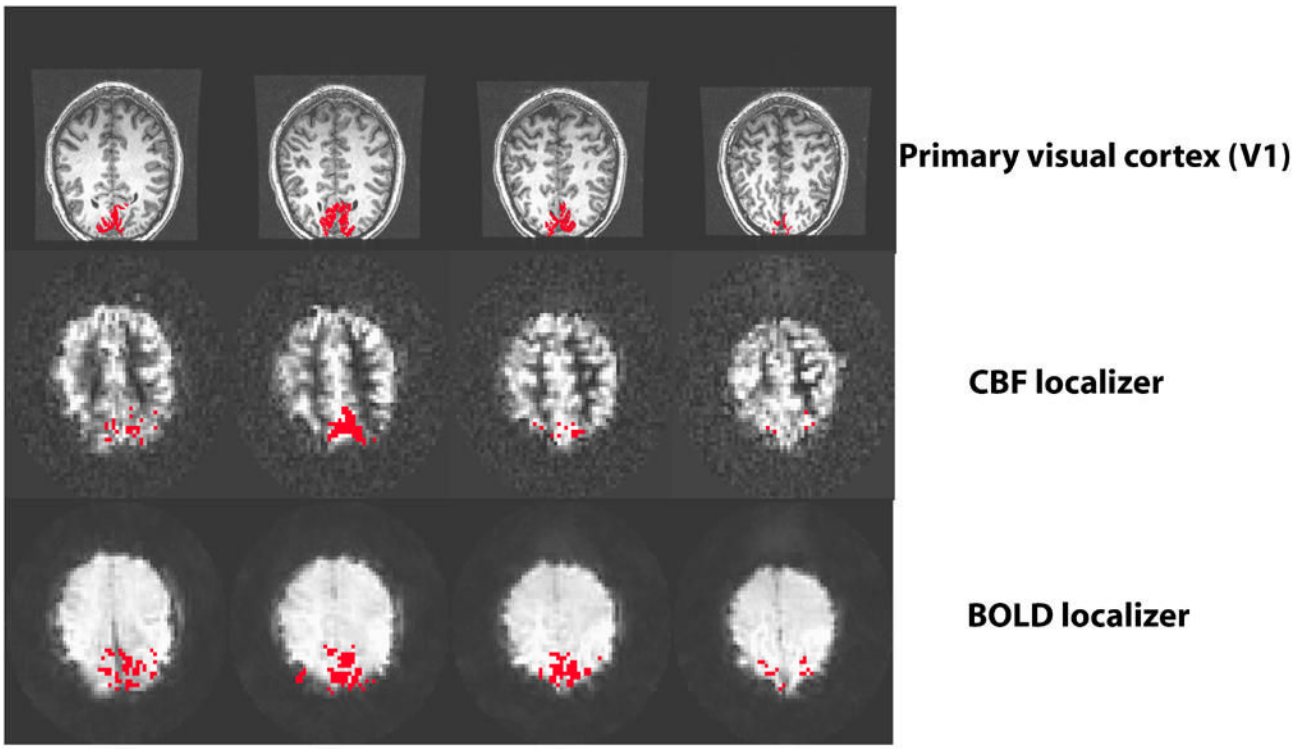

Figure 1.
Visual cortex regions-of-interest generated for a typical subject. The top row of images depicts 4 slices of visual area V1 overlaid onto an oblique representation of a high resolution anatomical scan (FSPGR). Retinotopic mapping was performed in a preliminary scan session on a different day. The second and third rows of images depict voxels with significant CBF and BOLD activation (r=0.5) overlaid onto their respective average perfusion and BOLD images.

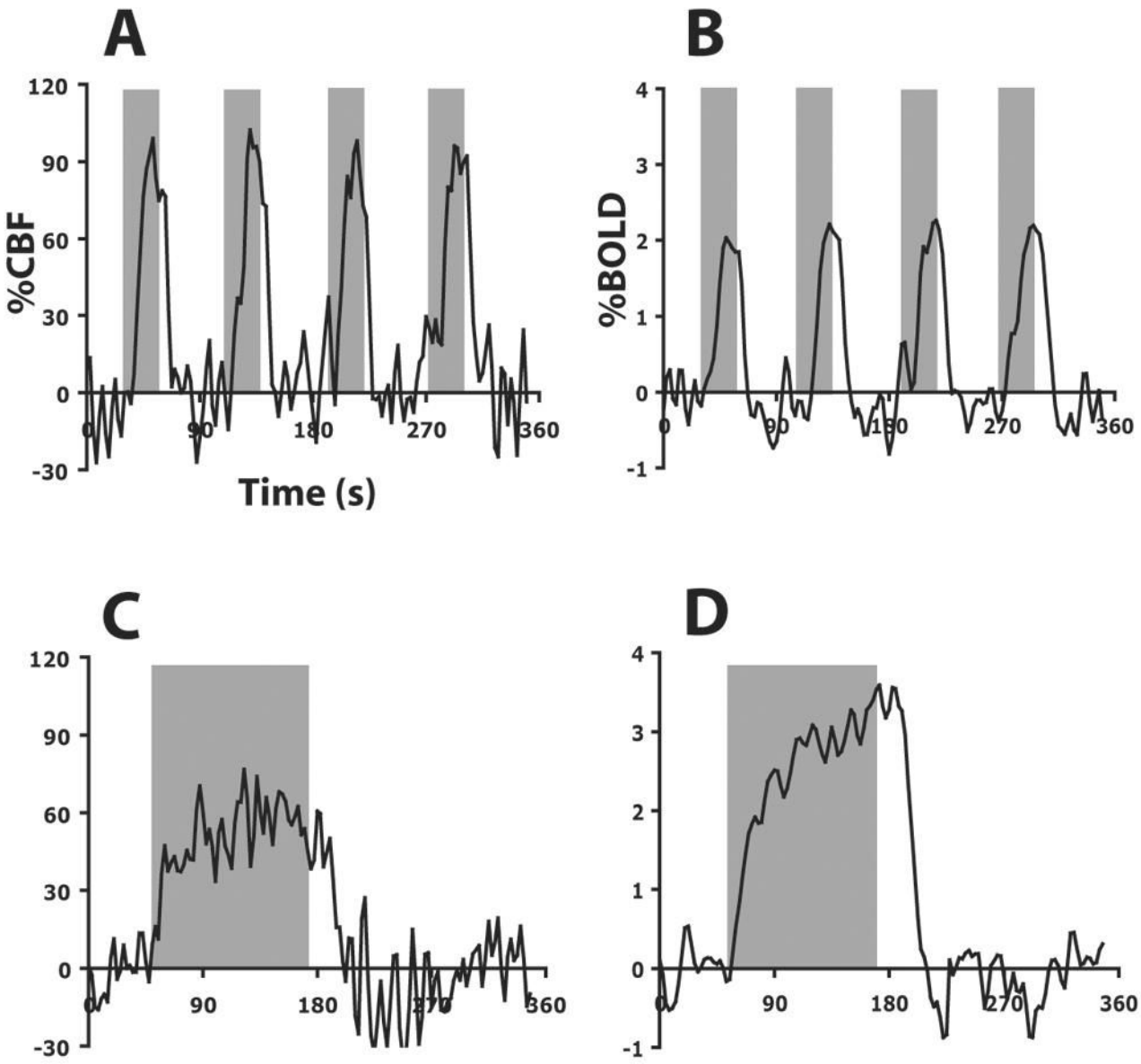

Figure 2.
Average BOLD and CBF responses for a typical subject under activation (Aand B) and hypercapnia (C and D) conditions for the CBF localizer. Evidence of $CMRO_2$ increase with activation is given by the observation that a larger BOLD signal is elicited by hypercapnia than activation for similar flow changes.

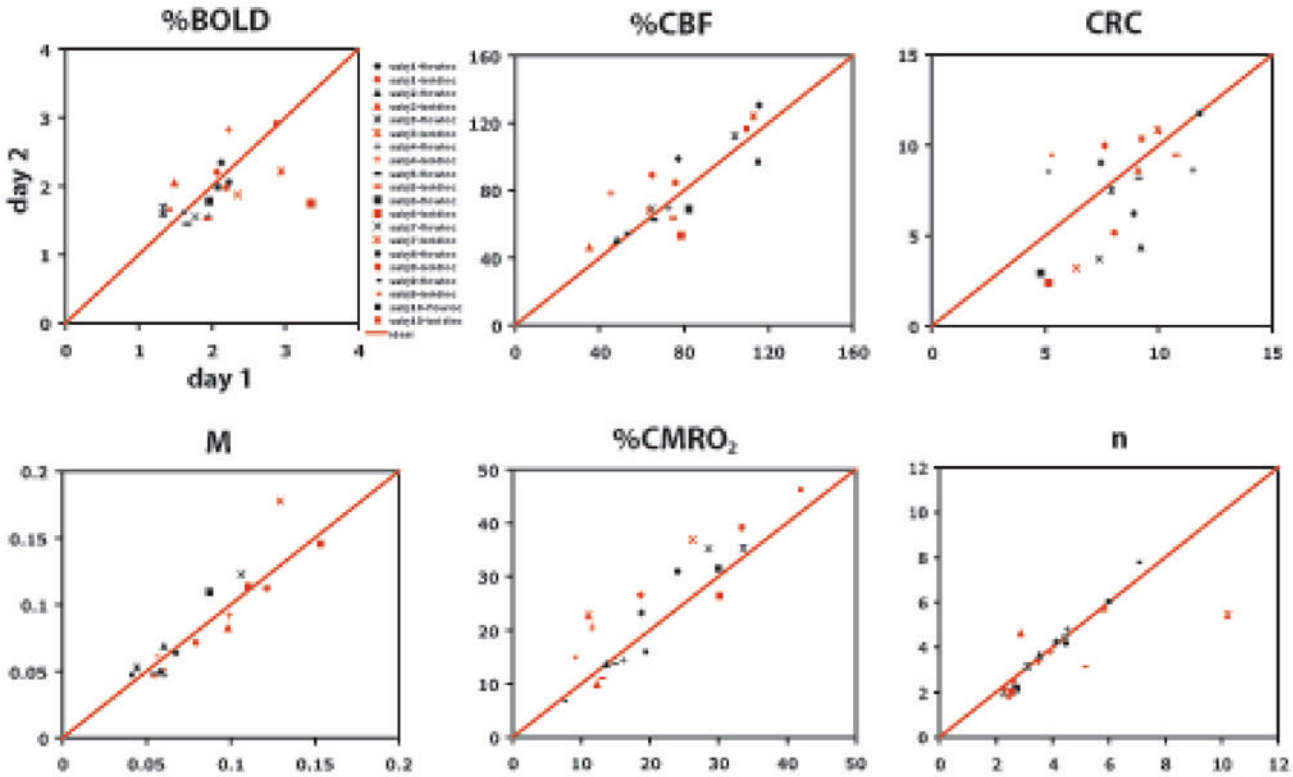

Figure 3.

Scatter plots of day 1 versus day 2 measurements for six physiological quantities available with calibrated BOLD fMRI. Red and black data points correspond to average measurements for individual subjects for the BOLD and CBF localizer, respectively. Different shapes and objects are used to identify individual subject data. The red line of slope one represents the case of perfect reproducibility. Data from subjects 6 and 7 is complicated by a CO_2 leak on day 2 as described in the results section. It is evident that the reproducibility of calculated measurements (*M* and *n*) are more stable across days than direct measurements (%BOLD, %CBF and CRC).

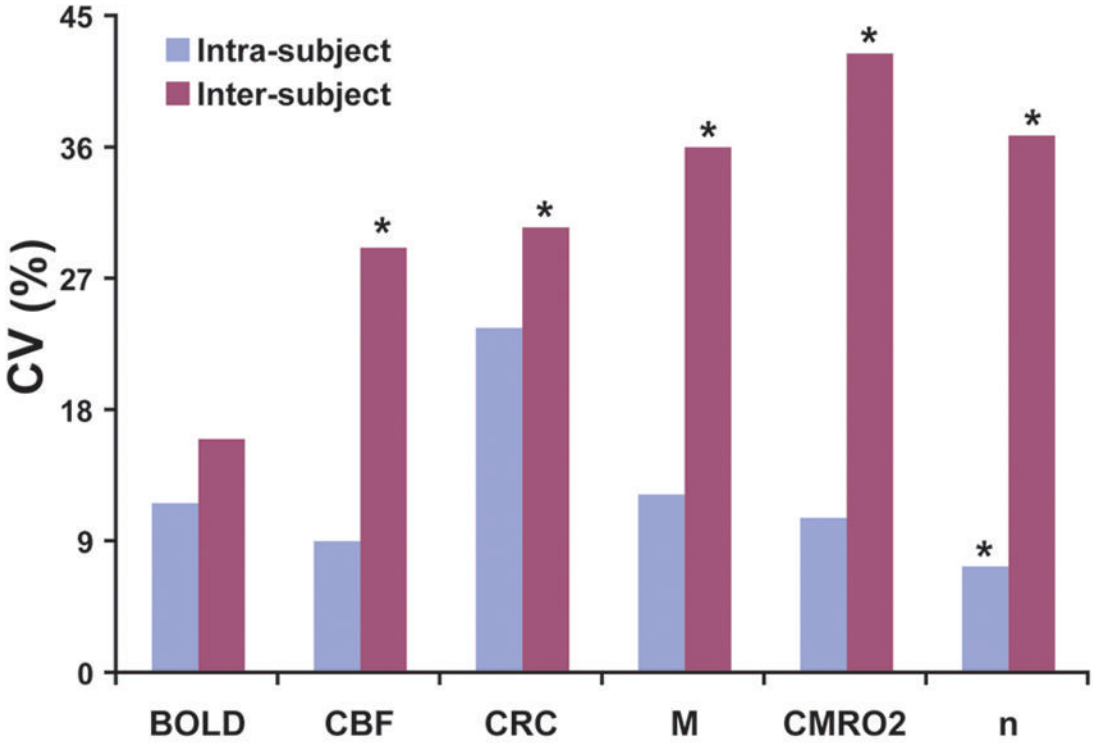

Figure 4.
Bar graph depicting the coefficient of variation (CV) for inter-subject variance (purple) and single subject reproducibility (blue-gray) of calibrated-BOLD fMRI measurements. An asterisk next to a measurement denotes a significantly different ($p<0.05$) CV_{inter} or CV_{intra} value compared to the BOLD response. This figure demonstrates that discrimination of calibrated BOLD measurements between healthy individuals is not limited by poor reproducibility.

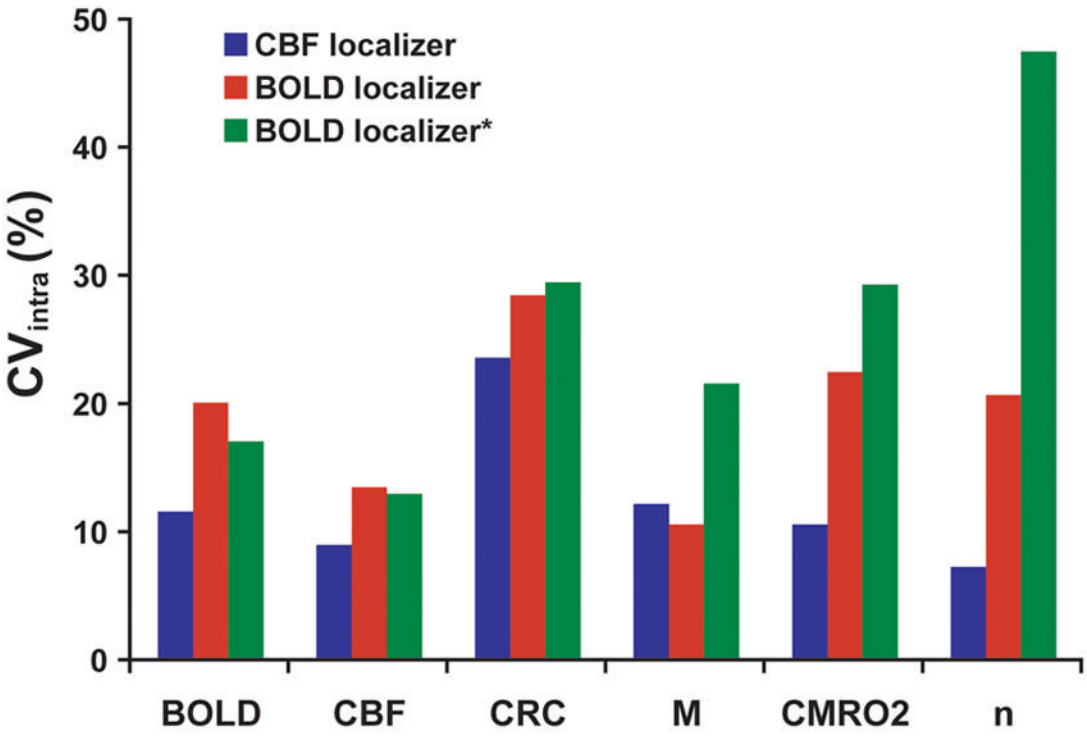

Figure 5.
Intra-subject reproducibility (CV_{intra}) for the CBF localizer (r=0.5), the BOLD localizer (r=0.5) and the threshold-adjusted BOLD localizer (BOLD localizer*). The threshold-adjusted BOLD localizer was constructed to contain the same number of voxels as the CBF localizer. It is evident that BOLD-activation regions, whether constructed at the same level of significance of activation or the same volume of activation as perfusion-activated regions, are associated with poorer reproducibility.

Table 1

Reproducibility of individual subject hypercapnia measurements

	V1		CBF-localizer		BOLD-localizer	
	$\langle\%\Delta\rangle\pm SD$[a]	CV_{intra}[b]	$\langle\%\Delta\rangle\pm SD$	CV_{intra}	$\langle\%\Delta\rangle\pm SD$[c]	CV_{intra}
BOLD[d]	30.3*±31.8	21.4	21.5*±37.0	26.1	24.8±49.5	35.0
CBF[d]	43.6*±50.4	35.7	31.0±57.4	40.6	29.3±68.3	48.3
CRC[c]	19.0±32.9	23.3	12.3±33.4	23.7	9.[1]±40.5	28.6
M	1.4±32.7	23.1	−3.8±[1]7.4	12.3	2.8±[1]5.1	10.7

[a] mean (±SD) of the individual subject fractional difference between measurements acquired on different days, expressed as percent

[b] standard deviation of the individual subject fractional difference between measurements acquired on different days for individual subjects divided by the square root of 2, expressed as a percent

[c] cerebrovascular reactivity to CO_2, defined as the percent change in CBF divided by the absolute change in end-tidal CO_2 (1/ΔmmHg)

* denotes a significant ($p<0.05$) difference between the means of measurements acquired on different days

[d] this category of measurements is complicated by a leak in the CO_2 delivery system for two subjects on day 2 of scanning. This experimental error will therefore falsely increase CV_{intra} as well as bias $\langle\%\Delta\rangle$ towards positive values.

Table 2

Variability of hypercapnia measurements across subjects

	V1		CBF-localizer		BOLD-localizer	
	mean±SD	CV_{inter} [b]	mean±SD	CV_{inter}	mean±SD	CV_{inter}
BOLD[a]	3.99±1.97	50.3*	2.52±0.82	32.7	3.77±1.54	40.9*
CBF[a]	58.1±25.7	44.2	66.0±27.4	41.5	66.0±27.4	41.5
CRC	7.3±2.4	33.0	7.7±2.6	30.6	7.8±2.9	32.2
M	0.111±.059	41.8*	0.065±.023	36.1	0.097±.037	38.5

[a] expressed as the mean (±SD) percent increase from baseline for all subjects all sessions

[b] coefficient of variation (expressed as a percent) representing the pooled data

* denotes a significant ($p<0.05$) difference in the variance compared to the CBF-localizer

Table 3

Reproducibility of individual subject activation measurements

	V1		CBF-localizer		BOLD-localizer	
	⟨%Δ⟩±SD[b]	CV_{intra}[c]	⟨%Δ⟩±SD	CV_{intra}	⟨%Δ⟩±SD	CV_{intra}
BOLD	8.6±25.7	19.2	−0.1±16.5	11.7	6.0±28.6	20.2*
CBF	4.4±30.2	21.2*	−2.7±12.9	9.1	−5.4±19.2	13.6
$CMRO_2$	−2.8±30.1	21.3	−4.9±15.1	10.7	−22.7**±31.9	22.6*
n	7.1±12.4	8.8	2.4±10.4	7.4	12.2±29.5	20.8*

* denotes a significant ($p<0.05$) difference in the variance compared to the CBF-localizer

** denotes a significant ($p<0.05$) difference between the means of measurements acquired on different days

Table 4

Variability of activation measurements across subjects

	V1		CBF - localizer		BOLD - localizer	
	mean±SD[b]	CV_{inter}[c]	mean±SD	CV_{inter}	mean±SD	CV_{inter}
BOLD[a]	1.87±0.67	33.9*	1.81±0.27	16.1	2.17±0.55	24.8*
CBF[a]	72.9±28.8	39.5	81.5±23.8	29.2	74.3±25.7	34.6
[a]$CMRO_2$	27.2±10.4	38.0	21.5±9.1	42.5	23.7±11.4	48.0
n	2.63±.66	22.4*	3.70±1.5	36.9	3.16±2.0	51.8

[a] expressed as the mean (±SD) percent increase from baseline for all subjects all sessions