NeuroImage 62 (2012) 1017–1023

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Review

The continuing challenge of understanding and modeling hemodynamic variation in fMRI

Daniel A. Handwerker [a,*], Javier Gonzalez-Castillo [a], Mark D'Esposito [c,d], Peter A. Bandettini [a,b]

[a] Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, 10 Center Dr., Room 1D80, MSC1148, Bethesda, MD 20814, USA
[b] Functional MRI Facility, 10 Center Dr., Room 1D80, MSC1148, National Institute of Mental Health, Bethesda MD, USA
[c] Helen Wills Neuroscience Institute, University of California, Berkeley 94720, USA
[d] Department of Psychology, University of California, Berkeley 94720, USA

ARTICLE INFO

Article history:
Accepted 7 February 2012
Available online 14 February 2012

Keywords:
BOLD
fMRI
Hemodynamic response
Population studies
Regional variation
Blood vessels

ABSTRACT

Interpretation of fMRI data depends on our ability to understand or model the shape of the hemodynamic response (HR) to a neural event. Although the HR has been studied almost since the beginning of fMRI, we are still far from having robust methods to account for the full range of known HR variation in typical fMRI analyses. This paper reviews how the authors and others contributed to our understanding of HR variation. We present an overview of studies that describe HR variation across voxels, healthy volunteers, populations, and dietary or pharmaceutical modulations. We also describe efforts to minimize the effects of HR variation in intrasubject, group, population, and connectivity analyses and the limits of these methods.

Published by Elsevier Inc.

Contents

Introduction

The hemodynamic response (HR) to neural activity is the basis of functional magnetic resonance imaging (fMRI). Neural activity changes can occur on the order of milliseconds. When activity across a population of neurons in a voxel changes, a hemodynamic response can be recorded using T2* weighted acquisitions in an MRI scanner. In contrast to most direct measures of neural activity, HRs are slow (on the order of seconds) and vary in shape and timing between voxels, cortical regions, subjects, subject populations, and experimental tasks. Because hemodynamics depend on the blood vessel structure, changes in this underlying structure will alter the HR even with consistent changes in neural activity.

Fig. 1 is reprinted, with permission, from Handwerker et al. (2004). It shows hemodynamic responses from 20 subjects in primary sensorimotor cortex in response to a single button press. The signal

* Corresponding author at: Bldg 10, Rm 1D80, 10 Center Dr MSC 1148, Bethesda, MD 20892-1148, USA. Fax: +1 301 402 1370.
E-mail address: handwerkerd@mail.nih.gov (D.A. Handwerker).

1053-8119/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.neuroimage.2012.02.015

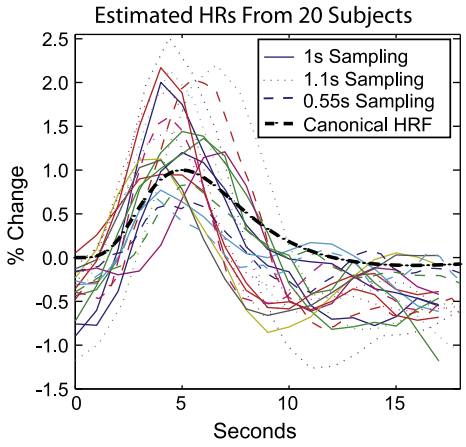

Fig. 1. Hemodynamic responses from 20 subjects averaged across a region of interest in primary sensorimotor cortex in response to a single button press.
Reprinted, with permission, from (Handwerker et al., 2004).

increases over several seconds and then drops below and then returns to baseline. Although we do not fully understand in detail all the nuances of how blood flow, volume, and oxygenation change in response to neural activity, the hemodynamic response is highly consistent. Yet, if one looks closer, it has significant variability. The 20 subjects in Fig. 1 show a wide range of differences in response latency, width, shape, and magnitude.

While we understand many of the hemodynamic mechanisms behind these differences in HR, predicting the precise response shape for a given stimulus and voxel region is still impossible. HR shape predictions are required for many fMRI statistical analyses. In addition, it is difficult to determine if HR shape variability is neural rather than vascular in origin. This article covers the history of our understanding of HR variation and the development of methods to try to account for hemodynamic variation in fMRI research.

Understanding HR variation

The HR becomes central to fMRI analysis

The earliest fMRI studies used block designs to maximize signal-to-noise ratio and increase the likelihood of response detection. The first few papers, for example, (Kwong et al., 1992), didn't model the HR shape. Similar to Positron Emission Tomography analyses of the time, they created statistical measures by subtracting magnitudes from an active condition from magnitudes during a rest condition. Direct observation of raw time series revealed features in the hemodynamic response that required further modeling to more effectively extract task-related signal fluctuations. Soon after the work by Kwong et al., Bandettini and colleagues modeled the full time series response as a box-car function — i.e., zeros for the rest periods, and ones for the active periods — and convolved it with a smoothing function. By calculating the correlation of this time series to the data, they were able to generate more precise statistical maps (Bandettini et al., 1993). This approach of comparing an idealized HR to BOLD time series soon became the standard way of detecting sites of neuronal activity in fMRI experiments. Simple correlation analysis was soon substituted by general linear model (GLM) analyses (Friston et al., 1994), which allow greater flexibility for the experimental design and response modeling. The GLM remains the primary analysis technique for most task-based fMRI studies.

While some early fMRI papers tested different HR shapes or models, the robust response obtained with block design tasks made HR variability a less urgent concern. This started to change with the use of event-related task designs. Contrasting hemodynamic responses to brief stimuli let researchers answer more complex neuroscience questions. The inherently lower signal to noise ratio of event related experiments combined with the increased importance of subtle BOLD signal changes in response of brief events, meant that HR variability had the potential to confound results. Although the risk that HR variation could affect results was acknowledged, it was assumed to be a relatively minor concern (Buckner et al., 1998). As described below, later studies have shown that the problem can be serious in some contexts.

HR variation in healthy volunteers

Even before event-related fMRI started to push the temporal limits of the hardware, early fMRI studies were determining the relative T2* vs. T2 vs. inflow contribution in activated tissue. In addition to showing that fMRI signal changes result from a spatially variable contribution of many factors, these early fMRI studies showed that the HR shape – particularly the response magnitude – varied greatly depending on the vessel diameters and density within each voxel. Another review in this issue covers this topic in more depth (Menon, 2012-this issue). Vessel structure differences across voxels can also cause delay differences, as shown by the early HR delay maps produced by Lee et al. (1995). To produce these delay maps, the authors collected data from a periodic flashing checkerboard task and spatially mapped the phase differences of the responses. They saw latency delays of 4–8 s in gray matter and 8–14 s in visible vessels and sulci (Lee et al., 1995). Later work, at higher temporal and spatial resolutions, showed evidence of hemodynamic magnitude and delay differences across cortical layers (Silva and Koretsky, 2002; Yu et al., 2012). While the goal of much of this work was to identify MRI pulse sequences and field strengths that could best localize neural activity, it also clearly showed that one of the biggest sources of non-neural HR magnitude and latency variation in healthy volunteers was within voxel blood vessel structure.

While the 14 s phase lags from Lee et al. (1995) are more than is typically observed in the literature, other studies have shown HR lag variation in the order of several seconds. For example, Buckner et al. found that when signals were averaged across regions of interest, task responses across healthy subjects often showed a 4–6 s range of time-to-peaks (with 13 subjects and a 2 s TR) (Buckner et al., 1998). In another study, Birn et al. showed 6–11 s visual cortex response latency in response to a brief stimulus (Birn et al., 2001). Interestingly, in this last study, these latencies were not correlated with response magnitude. This is potentially explained by the understanding that while vessels are mostly “downstream” they can also exist more upstream as well — having shorter latencies.

Buckner et al. not only looked at delay differences, but also evaluated HR differences across subjects in terms of how much the HR from one subject could predict the HR on another subject for the same region. They found that HR from one subject was able, on average, to explain 72% of the variance in other subjects. Around the same time, Aguirre was asking how much the hemodynamic response to a simple visuo-motor task varies within the same subject across hours or days and across subjects. While they didn't perform the same variance analysis as Buckner, they did see a time-to-peak range of 2.7–6.2 s across 32 volunteers. This variation didn't correlate with reaction time. The HR shapes across subjects in that paper have very different peak-times, widths, and post-undershoot shapes. This variability is visibly greater than the HR shapes Buckner et al. used to estimate HR similarity. They also showed that the HRs estimated from a region of interest in the same subject in multiple scans varied much less than HRs estimated from the same region of interest across subjects — particularly for multiple scans in a single day (Aguirre et al., 1998).

Known causes of HR variation

As discussed above, one of the biggest causes of non-neural HR variation is vessel size and density within a voxel. Large vessels are typically "downstream" and therefore experience an activation-induced oxygenation change up to 3 s later than the capillaries within parenchyma. Additional factors that can contribute to differences in HR shape between two subsequent measurements of the response to the same task, in the same region of the same subject include: thermal noise, breathing and heart rate changes, hardware instabilities, changes in cognitive state and strategy used to approach the task. If the measurements are performed on different days or we attempt to compare across different subjects or regions then vascular state, region of interest selection, and tissue differences also contribute.

Past research has helped us start to understand the relative contributions of each of these factors to HR variation so that variation of neuronal origin can be segregated from variation due to the other non-neuronal sources. Much of the research is based on dynamic models of BOLD contrast, which try to predict how variation in regional blood flow, volume, and vessel reactivity combines to cause observed BOLD signal changes. This special issue has another article that covers this topic in more depth (Buxton, 2012-this issue).

In the particular case of studies across populations, the major concern should be that an HR difference is mistakenly considered a neural difference or that a true neuronal difference is obliterated by the HR variability. Virtually every examined disease state, age difference, or ingested substance causes changes in HR shape. Alcohol (Levin et al., 1998), caffeine (Mulderink et al., 2002), fat (oral canola oil) (Noseworthy et al., 2003), inhaled CO_2 and O_2 concentration changes (Sicard and Duong, 2005), and intravenous saline (Levin et al., 2001) all alter HR response magnitude. Nicotine is one notable exception with one study showing that HR magnitude remains constant across dose even with a heart rate change (Jacobsen et al., 2002). Population and case studies have found HR differences that were likely vascular in origin in aging (D'Esposito et al., 1999), dementia (Buckner et al., 2000), and stroke (Hamzei et al., 2003; Pineiro et al., 2002). Even if variation across voxels due to vessel effects is much larger than these group differences, a systematic change under large amounts of within-group variance can still produce a population effect. The larger the population size, the more likely a population effect due to HR differences will be statistically significant.

In many of these examples, the underlying assumption is that anything that can systematically alter baseline cerebral blood flow (CBF) or hematocrit levels will alter the HR response. While others have systematically examined the effects of baseline CBF on the HR (Bandettini and Wong, 1997), an elegantly designed study by Cohen et al. provides a good illustration of the effects of baseline CBF on HR responses. The authors had volunteers breathe normocapnic, hypercapnic (raises baseline CBF), and hypocapnic (lowers baseline CBF) air. When the baseline CBF stabilized in each condition, they ran a visual stimulation task. Thus, in the same volunteers, they had HRs at three different baseline CBF levels. As CBF rose, the magnitude of the response to visual stimulation decreased. Interestingly, the onset time and time-to-peak also increased with baseline CBF. When there was a physiologic modulation in BOLD magnitude in an individual, it appeared that time-to-peak increased by 0.28 s for every 1% change in the baseline BOLD signal (Cohen et al., 2002). This lag difference could appear in any population with a baseline CBF change.

Hemodynamic response changes can be due to more than altered CBF. Arterial spin labeling measures depend primarily on blood flow. Multiple arterial spin labeling studies have shown lower HR response magnitude variability across scans within subjects and across subjects than BOLD (Aguirre et al., 2002; Raoult et al., 2011; Tjandra et al., 2005). This means that variation in volume and vessel elasticity can increase variability of the BOLD hemodynamic response in ways that are distinct from blood flow variation.

Drug studies also show how HR variability across scans is due to more than just blood flow variation. Caffeine, a known vasoconstrictor, is one of very few drugs whose effects on the HR have been examined in multiple experiments. While an early study noted the CBF baseline shift and response magnitude increase and proposed intentionally giving caffeine as a BOLD contrast enhancer (Mulderink et al., 2002), later work creates a more complex story (Behzadi and Liu, 2006; Liu et al., 2004). If all caffeine did was shift the baseline CBF then the BOLD response would shift to result in a similar estimate of oxygen metabolism in response to a task. Two groups have shown that caffeine alters the relationship between CBF and oxygen metabolism (Chen and Parrish, 2009; Perthen et al., 2008). In addition, BOLD response is measurably better modeled as a linear system after caffeine ingestion (Liu and Liau, 2010) and caffeine alters post-stimulus undershoot dynamics (Liu et al., 2004). Perhaps some of these effects are neural, but it appears that the effects of caffeine on the HR shape include more than a CBF baseline shift. Even with this information, there is no field-wide consensus or even significant public discussions regarding how we record or control for caffeine intake (or drug intake in general) in fMRI studies.

Another mechanism to alter neurovascular coupling involves cyclooxygenase (COX). COX has several key roles in how astrocytes and neurons interact to alter vessel dilation (Koehler et al., 2006). Indomethacin, a COX inhibitor, has been shown to decrease the BOLD response to a task despite being a vasoconstrictive drug that should lower baseline CBF (Bruhn et al., 2001). HR magnitude also changes in humans depending on volunteer's specific COX-1 genotype (Hahn et al., 2011). Given that Ibuprofen, an over-the-counter pain reliever, is a COX inhibitor, population differences in the intake of a drug of such common use as this have the potential to become a confound in fMRI studies. Caffeine and COX inhibitors are just two examples, but the broader question of what other common substances can cause similarly complex changes in the HR remains unanswered.

In population studies, there also seem to be changes that can't simply be attributed to baseline CBF changes. In one case, disturbed cerebrovascular reserve capacity from an intracranial arterial occlusion, resulted in a negative HR (Hamzei et al., 2003). Aging studies have shown conflicting reports of differences in HR time-to-peak, time-to-baseline, and fit residuals (D'Esposito et al., 1999; Huettel et al., 2001; Richter and Richter, 2003). Other potential aging related variations in BOLD signal that might be hemodynamic in origin were summarized in a review article (D'Esposito et al., 2003). For example, changes in vessel elasticity were implicated as causing a difference in the time course of cerebrovascular reactivity (Hajdu et al., 1990). Some of the challenges mentioned in that review can be generalized to other populations.

The effect of HR variation in data analysis

Significance testing across varying HR shapes

Any statistical method that includes a model of the HR risks missing the aspects of the response that are not included in the model. Some of the earliest analysis methods examined how significance varied depending on the modeled HR shape (Bandettini et al., 1993; Friston et al., 1994). In most of these cases, they were trying to identify a single HR shape that was best for most studies. Aguirre et al. suggested including a separate scan to estimate a separate HR for each subject (Aguirre et al., 1998). They made it easy to include these individualized estimates in GLM analyses using the VoxBo software. This proposal was not widely adapted. Hesitance to use 6 additional minutes of scan time to independently estimate a subject-specific HR played a role, but more prosaic issues also mattered. There was no "push-button" way to insert a custom HR into SPM, the very widely used fMRI data processing software. There was also skepticism that a motor HR from one region could better estimate

an HR in other brain regions of the same person despite solid research showing this was true. Perhaps the primary reason why it was not adapted was, that no one showed precisely, at the time, how much an HR misestimate could alter results.

When Dan Handwerker joined Mark D'Eposito's group as a graduate student, they decided to examine this issue. They repeated the basic experimental design of Aguirre, but designed a task to generate robust HR's in 4 brain regions. These responses were used to set a range of HR shapes to simulate the effect of HR variation on GLM statistics. They discovered that a 1 s misestimate of the HR time-to-peak could decrease the fit magnitude by 10% and a 2 s misestimate decreases the magnitude by 38%. An accurate subject-specific HR estimate would clearly improve statistical results compared to using the same shape in all subjects. Still, variation across voxels within subjects was still large enough to cause non-trivial shifts in statistics to limit the benefits of using one HR shape per subject (Handwerker et al., 2004).

While a robust enough task would still show significance, the drop is large enough to create false negative results. Including a wider range of response HR shapes in an analysis, such as modeled transients, increases the extent of significant activation (Harms and Melcher, 2003; Uludag, 2008). While these studies hypothesized neural origins of HR shape differences, they clearly show that unmodeled variation affects significance. In addition, if there were a systematic HR shape difference between populations it would incorrectly appear as a response magnitude difference in many group analyses.

The basic approach to account for HR variability in GLM-based analyses is to include multiple basis functions for each condition. For example, a Taylor expansion using an HR and its derivatives allows for some lag and width variation around the initial HR shape (Henson et al., 2002). While beneficial, it rapidly loses effectiveness if the true HR in a voxel has more than a 1 s lag from the modeled HR. In addition, if all the basis functions aren't included in the magnitude estimates in a group analysis (Calhoun et al., 2004), the benefits of flexible HR modeling are lost (Handwerker et al., 2004). The opposite extreme would be to deconvolve any task-locked HR with a series of sinusoids in a GLM (Josephs et al., 1997), without a GLM (Glover, 1999), or using a series of lagged impulse responses in a GLM (Ollinger et al., 2001). These are powerful methods for significance testing within subject, but they use many more degrees of freedom and it is more challenging to design and interpret contrasts between task conditions and groups using these methods. Intermediate methods can model an unlimited range of lags for a specific HR shape (Saad et al., 2001) or use a specially designed set of basis functions in a GLM and prior information to estimate the HR (Woolrich et al., 2004). While it is possible to compare results across several methods, no one method is designed to handle the full range of known HR lag and shape variation observed across voxels in individual subjects.

Correcting for non-neural magnitude differences

HR magnitude variation based on the underlying vessel structure, adds noise to group results and can bias the foci of activation towards larger vessels. At best, this is just another source of noise in fMRI data. For studies across populations, a vascular difference could be interpreted as a neural difference. One robust way to address this issue is to make sure one condition of a study shows no population difference in the same voxels that the condition of interest does show a population difference. This is not always feasible.

A series of methods have sought to calibrate the fMRI signal to remove large vessel effects (or more precisely, to remove blood volume variation across voxels — the primary source of magnitude differences). Peter Bandettini and Eric Wong first proposed collecting a run with volunteers breathing CO_2 enriched air. This would cause a global BOLD signal change that would scale with venous blood volume in each voxel. Simply dividing by vessel magnitudes would remove spatial vascular differences (Bandettini and Wong, 1997). This concept was extended to collecting BOLD/CBF data at a single (Davis et al., 1998) or multiple (Hoge et al., 1999) air CO_2 concentrations to calibrate responses to a more direct measure of cerebral metabolism. A large methodological literature is being built around this concept, but these types of methods are used in only a small fraction of population studies. One reason is that CO_2 inhalation systems might be unpleasant for volunteers, and the best methods require a lot of scan time. Also, these calibrations involve subtraction or division of data, decreasing the signal to noise ratio. These inconvenience factors tend to outweigh the benefit of this type of calibration for most studies. Handwerker and D'Esposito tested whether a more convenient, but slightly less controllable method, global signal changes from breath hold, could be sufficient to calibrate some signal changes with aging (Handwerker et al., 2007). That manuscript includes a more thorough review of earlier work in this area. We showed that breath holding revealed age-related BOLD differences. Accounting for those global signal changes altered the HR magnitude in regions that showed significant task-related changes with aging. While simpler to implement than CO_2 breathing, breath holding tasks still require people who are able to hold their breath for around 20 s and follow instructions. This challenge, in addition to the requirements of extra scan time and the decreased SNR from dividing data for calibration, explains why this method also hasn't been widely adapted.

All these approaches are trying to strike a balance between an accurate and precise calibration method and a practical protocol to add to a large population study. Until that happens, best research practice will continue to require study designs with a dissociation between population responses between conditions. Studies that only show a population difference to a single condition should continue to be treated with healthy skepticism.

HR variation and brain networks

With the rapid growth of network modeling of brain connections – particularly using fMRI data with spontaneous fluctuations – there is increased interest and attempts to derive functional brain networks; and to assign a temporal direction to information flow between brain regions. A recent paper, Smith et al. (2011), provides a good summary of existing network modeling methods and uses simulations to compare method quality. The simulations included HR delay variation with a standard deviation of 0.5 s. Since this was well below the observed variation across voxels, they considered this a “best-case scenario while remaining reasonably realistic.” They show that some methods are robust to this level of HR lag variation, but lag-based methods, like Granger causality perform particularly poorly. Even on stimulated data with best-case HR variation, none of the existing methods performs exceptionally at assigning directionality to the connections between regions. The best method tested, Patel's τ, showed a 65% accuracy with 50% being chance (Smith et al., 2011). While these results show some existing methods are robust to a best-case scenario of random HR noise, none was tested on the full range of known HR lag differences, other HR shape differences, or systematic HR differences between network nodes.

One well-known network directionality method that was not tested by Smith et al. is Dynamic Causal Modeling (DCM). An appealing aspect of this method is that it attempts to use Bayesian methods to fit an optimal HR for each node and then use that model to estimate the temporal patterns of the underlying neural activity (Friston et al., 2003). Although the creators of this method have tested it using some HR variation (Stephan et al., 2007), we decided to test the accuracy of DCM at making the simplest possible prediction. In a two node network, do the estimated neural events of node 1 predict the estimated neural events in node 2 or vice versa? These two models are depicted in Fig. 2E. For each of the two nodes, we took a

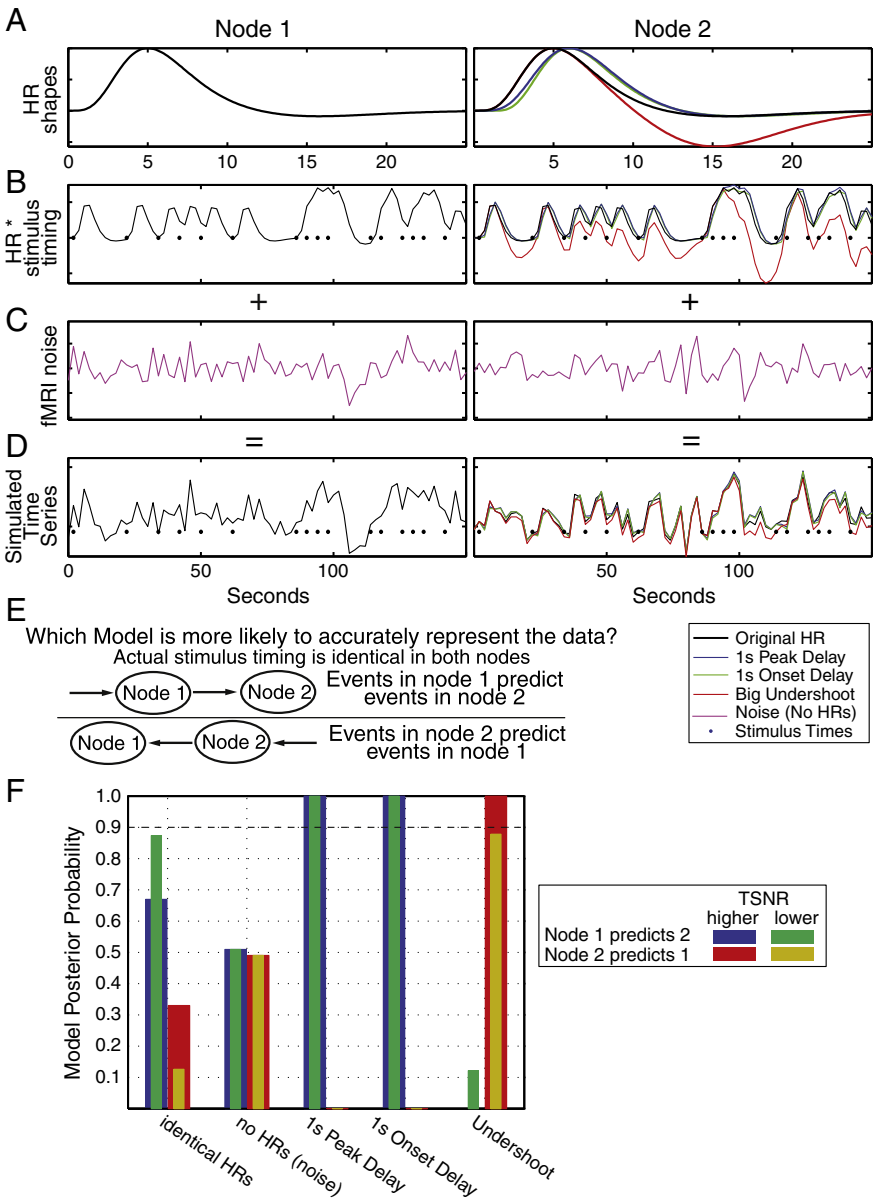

Fig. 2. The left column shows the inputs to node 1 of the dynamic casual modeling (DCM) simulation and the right column shows the inputs to node 2. (A) Node 1 always includes a single HR shape. Node 2 includes the same HR shape or HR shapes that have delayed onsets, peak times, or undershoot magnitudes. (B) The HRs are convolved with an event-related design of neural event times (black dots). This shows a 150 s window of the 300 s time series (C) Each node has a noise time series from different subjects' scans of spontaneous fluctuations. (D). The HR time series in B are scaled and added to the noise in C. This figure shows how the time series look for the lower TSNR condition. (E) Schematics of the two models that were compared using DCM. (F) Comparisons of the two models in E. For each HR shape tested, if the blue or green lines are higher, that means node 1 is more likely to predict node 2. If the red or yellow bars are higher, node 2 is more likely to predict node 1. The dashed line at 0.9 is a typical significance threshold.

hemodynamic response function shape (Fig. 2A), convolved it with an event-related time series of neural events (Fig. 2B), and scaled and added it to noise (Figs. 2C and D). The HR was constant in node 1. The HR in node 2 was identical to node 1, had a 1 s delayed peak or onset, or a larger relative post-stimulus undershoot. For the event-related design, there was a 50% chance of an event appearing at each time point (sampling rate = 2 s). The identical event pattern with no time lag was used in both nodes. While a 1 s HR onset delay is equivalent to a 1 s shift in event timing, all other cases should be null results (i.e. neither node significantly predicts the other). To get realistic fMRI noise with minimal causality and correlation, each of the two nodes contained spontaneous fluctuations from a single voxel taken from a different subject and brain region (Fig. 2C). These data had 300 time points and were collected as part of Murphy et al. (2009). The noise time series had a correlation magnitude of $r = -0.08$. Simulated data were created by adding these noise time series to the scaled synthetic responses (Fig. 2B). The data were scaled so that the correlations between the nodes when the HRs were identical would represent a higher ($r = 0.78$) and lower ($r = 0.46$) temporal signal-to-noise ratio (TSNR). A DCM analysis in the SPM8 software package was run on both of the networks in Fig. 2E at both TNSR levels and all 4 pairings of HR shapes (Fig. 2D). The DCM analysis was also run on the noise time series (Fig. 2C). The underlying noise time series in nodes 1 and 2 were also switched to confirm that the presented results were a function of HR variation and not noise characteristics.

Fig. 2F shows the results comparing posterior probabilities from each model. If one model is higher than the other, it is more likely to represent the data. A typical significance cutoff is a posterior probability greater than 0.9. As expected, neither model is significantly better

when they contain identical HR shapes or with the noise time series. The identical HR shapes show a causality bias towards node 1 predicting node 2, but this doesn't cross significance. This non-significant bias might exist because there is very small bias towards the same causal model in the noise time series ("no HRs" in Fig. 2F). With just the noise time series, neither model fits the data well so there's little evidence either is correct. With the identical stimuli included, there is the same bias for the node 1 predicts node 2 causal model as in the noise time series, but each node's time series fits the stimulus timing better and shows more evidence for each model. Thus, even with identical HRs, one model seems more accurate.

When the HR in node 2 has a 1 s peak or onset delay compared to node 1, the DCM analysis shows node 1 probably predicts node 2. Since an HR onset delay would look identical to a neural event delay, one would expect DCM to show this model difference. A 1 s slower rise-time for an HR is well within the range of normal HR variation and shouldn't be considered to represent a delay in neural timing. When the HR in node 2 had a larger post-stimulus undershoot, the DCM analysis shows that node 2 probably predicts node 1. While the tested undershoot is very similar to HRs observed in real data (Fig. 1), it doesn't perfectly fit into the standard HR estimation method used in DCM. We suspect that, if the HR is misestimated, any systematic variation in responses could alter the predicted timing for neural events. For all models with different HR shapes in each node, the higher the TSNR of the data, the more likely a significant model difference will be found — since the variability of the HR's was more fully expressed.

While our test is on an overly simplistic model it does highlight that the hemodynamics estimation portion of the DCM analysis shouldn't be assumed to be accurate and cannot be finessed away no matter how sophisticated the models. Even though DCM may model some HR variation it is important to both examine the Bayesian priors for the HR estimation and compare them to the actual HR shape in each voxel or region of interest.

The current status of HR variation

The challenges and solutions regarding HR variation seem to have changed little over the history of fMRI. We now have a better understanding of the range of HR variation and some factors that can systematically alter HR shape. We know observed HR variation is sufficient, at best, to increase noise in most fMRI analyses and, at worst, to systematically bias results. An array of methods has been developed to deal with some of this variation, but they are all far from perfect and quite a few studies don't consider HR variation at all.

Since neurovascular coupling can vary based on how the vascular and tissue composition of voxels interact and can be altered in complex ways by disease or medication, we are not capable of robustly separating the voxel-wise, region-wise, subject-wise, and population-wise variability into variability of hemodynamic origin and variability of neuronal origin. While better scanners and pulse sequences will improve our temporal signal to noise ratio, the issues surrounding HR variation will continue to fundamentally limit the conclusions we can reach using fMRI data.

The best research uses known sources of HR variation and consistency and makes careful assumptions where possible and necessary. Even after 20 years, we are still developing a deeper understanding of the hemodynamic response and better methods to robustly characterize it, work with it, calibrate it, or work around it. In addition, we are finding that even with the dominance of the hemodynamic variability, still more subtle neuronal information can and will be extracted in the future.

Acknowledgments

Supported by the NIMH Intramural research program. Carlton Chu helped make sure the DCM analyses were properly run.

References

Aguirre, G.K., Zarahn, E., et al., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8 (4), 360–369.
Aguirre, G.K., Detre, J.A., et al., 2002. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. Neuroimage 15 (3), 488–500.
Bandettini, P.A., Wong, E.C., 1997. A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. NMR Biomed. 10 (4–5), 197–203.
Bandettini, P.A., Jesmanowicz, A., et al., 1993. Processing strategies for time-course data sets in functional MRI of the human brain. Magn. Reson. Med. 30 (2), 161–173.
Behzadi, Y., Liu, T.T., 2006. Caffeine reduces the initial dip in the visual BOLD response at 3 T. Neuroimage 32 (1), 9–15.
Birn, R.M., Saad, Z.S., et al., 2001. Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. Neuroimage 14 (4), 817–826.
Bruhn, H., Fransson, P., et al., 2001. Modulation of cerebral blood oxygenation by indomethacin: MRI at rest and functional brain activation. J. Magn. Reson. Imaging 13 (3), 325–334.
Buckner, R.L., Koutstaal, W., et al., 1998. Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. Neuroimage 7 (3), 163–175.
Buckner, R.L., Snyder, A.Z., et al., 2000. Functional brain imaging of young, nondemented, and demented older adults. J. Cogn. Neurosci. 12 (Suppl. 2), 24–34.
Buxton, R.B., 2012. Dynamic models of BOLD Contrast. Neuroimage 62 (2), 952–960 (this issue).
Calhoun, V.D., Stevens, M.C., et al., 2004. fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. Neuroimage 22 (1), 252–257.
Chen, Y., Parrish, T.B., 2009. Caffeine's effects on cerebrovascular reactivity and coupling between cerebral blood flow and oxygen metabolism. Neuroimage 44 (3), 647–652.
Cohen, E.R., Ugurbil, K., et al., 2002. Effect of basal conditions on the magnitude and dynamics of the blood oxygenation level-dependent fMRI response. J. Cereb. Blood Flow Metab. 22 (9), 1042–1053.
Davis, T.L., Kwong, K.K., et al., 1998. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc. Natl. Acad. Sci. U. S. A. 95 (4), 1834–1839.
D'Esposito, M., Zarahn, E., et al., 1999. The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. Neuroimage 10 (1), 6–14.
D'Esposito, M., Deouell, L.Y., et al., 2003. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nat. Rev. Neurosci. 4 (11), 863–872.
Friston, K.J., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Hum. Brain Mapp. 1, 153–171.
Friston, K.J., Harrison, L., et al., 2003. Dynamic causal modelling. Neuroimage 19 (4), 1273–1302.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9 (4), 416–429.
Hahn, T., Heinzel, S., et al., 2011. Neurovascular coupling in the human visual cortex is modulated by cyclooxygenase-1 (COX-1) gene variant. Cereb. Cortex 21 (7), 1659–1666.
Hajdu, M.A., Heistad, D.D., et al., 1990. Effects of aging on mechanics and composition of cerebral arterioles in rats. Circ. Res. 66 (6), 1747–1754.
Hamzei, F., Knab, R., et al., 2003. The influence of extra- and intracranial artery disease on the BOLD signal in FMRI. Neuroimage 20 (2), 1393–1399.
Handwerker, D.A., Ollinger, J.M., et al., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21 (4), 1639–1651.
Handwerker, D.A., Gazzaley, A., et al., 2007. Reducing vascular variability of fMRI data across aging populations using a breathholding task. Hum. Brain Mapp. 28 (9), 846–859.
Harms, M.P., Melcher, J.R., 2003. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum. Brain Mapp. 20 (3), 168–183.
Henson, R.N., Price, C.J., et al., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15 (1), 83–97.
Hoge, R.D., Atkinson, J., et al., 1999. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. Proc. Natl. Acad. Sci. U. S. A. 96 (16), 9403–9408.
Huettel, S.A., Singerman, J.D., et al., 2001. The effects of aging upon the hemodynamic response measured by functional MRI. Neuroimage 13 (1), 161–175.
Jacobsen, L.K., Gore, J.C., et al., 2002. Impact of intravenous nicotine on BOLD signal response to photic stimulation. Magn. Reson. Imaging 20 (2), 141–145.
Josephs, O., Turner, R., et al., 1997. Event-related fMRI. Hum. Brain Mapp. 5 (4), 243–248.
Koehler, R.C., Gebremedhin, D., et al., 2006. Role of astrocytes in cerebrovascular regulation. J. Appl. Physiol. 100 (1), 307–317.
Kwong, K.K., Belliveau, J.W., et al., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Natl. Acad. Sci. U. S. A. 89 (12), 5675–5679.
Lee, A.T., Glover, G.H., et al., 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. Magn. Reson. Med. 33 (6), 745–754.
Levin, J.M., Ross, M.H., et al., 1998. Reduction in BOLD fMRI response to primary visual stimulation following alcohol ingestion. Psychiatry Res. 82 (3), 135–146.
Levin, J.M., Frederick, B.B., et al., 2001. Influence of baseline hematocrit and hemodilution on BOLD fMRI activation. Magn. Reson. Imaging 19 (8), 1055–1062.

Liu, T.T., Liau, J., 2010. Caffeine increases the linearity of the visual BOLD response. Neuroimage 49 (3), 2311–2317.
Liu, T.T., Behzadi, Y., et al., 2004. Caffeine alters the temporal dynamics of the visual BOLD response. Neuroimage 23 (4), 1402–1413.
Menon, R.S., 2012. The great brain versus vein debate. Neuroimage 62 (2), 969–973 (this issue).
Mulderink, T.A., Gitelman, D.R., et al., 2002. On the use of caffeine as a contrast booster for BOLD fMRI studies. Neuroimage 15 (1), 37–44.
Murphy, K., Birn, R.M., et al., 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage 44 (3), 893–905.
Noseworthy, M.D., Alfonsi, J., et al., 2003. Attenuation of brain BOLD response following lipid ingestion. Hum. Brain Mapp. 20 (2), 116–121.
Ollinger, J.M., Corbetta, M., et al., 2001. Separating processes within a trial in event-related functional MRI. Neuroimage 13 (1), 218–229.
Perthen, J.E., Lansing, A.E., et al., 2008. Caffeine-induced uncoupling of cerebral blood flow and oxygen metabolism: a calibrated BOLD fMRI study. Neuroimage 40 (1), 237–247.
Pineiro, R., Pendlebury, S., et al., 2002. Altered hemodynamic responses in patients after subcortical stroke measured by functional MRI. Stroke 33 (1), 103–109.
Raoult, H., Petr, J., et al., 2011. Arterial spin labeling for motor activation mapping at 3T with a 32-channel coil: reproducibility and spatial accuracy in comparison with BOLD fMRI. Neuroimage 58 (1), 157–167.
Richter, W., Richter, M., 2003. The shape of the fMRI BOLD response in children and adults changes systematically with age. Neuroimage 20 (2), 1122–1131.
Saad, Z.S., Ropella, K.M., et al., 2001. Analysis and use of FMRI response delays. Hum. Brain Mapp. 13 (2), 74–93.
Sicard, K.M., Duong, T.Q., 2005. Effects of hypoxia, hyperoxia, and hypercapnia on baseline and stimulus-evoked BOLD, CBF, and CMRO2 in spontaneously breathing animals. Neuroimage 25 (3), 850–858.
Silva, A.C., Koretsky, A.P., 2002. Laminar specificity of functional MRI onset times during somatosensory stimulation in rat. Proc. Natl. Acad. Sci. U. S. A. 99 (23), 15182–15187.
Smith, S.M., Miller, K.L., et al., 2011. Network modelling methods for FMRI. Neuroimage 54 (2), 875–891.
Stephan, K.E., Weiskopf, N., et al., 2007. Comparing hemodynamic models with DCM. Neuroimage 38 (3), 387–401.
Tjandra, T., Brooks, J.C., et al., 2005. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. Neuroimage 27 (2), 393–401.
Uludag, K., 2008. Transient and sustained BOLD responses to sustained visual stimulation. Magn. Reson. Imaging 26 (7), 863–869.
Woolrich, M.W., Behrens, T.E., et al., 2004. Constrained linear basis sets for HRF modelling using Variational Bayes. Neuroimage 21 (4), 1748–1761.
Yu, X., Glen, D., et al., 2012. Direct imaging of macrovascular and microvascular contributions to BOLD fMRI in layers IV–V of the rat whisker-barrel cortex. Neuroimage 59 (2), 1451–1460.