Evaluating models of the ageing BOLD response

Henson, R.N.[1,2], Olszowy, W.[3,4], Tsvetanov, K.A.[5,6] Cam-CAN[*] & Zeidman, P. [7]

[1] Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, UK
[2] Department of Psychiatry, University of Cambridge, UK
[3] Wolfson Brain Imaging Centre, Department of Clinical Neurosciences, University of Cambridge, UK
[4] Data Science Unit, Science and Research, dsm-firmenich AG, Kaiseraugst, Switzerland
[5] Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0SP, UK
[6] Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK
[7] Wellcome Centre for Human Neuroimaging, University College London, UK
[*] list of authors at end of paper.

*Corresponding author:
Prof Richard Henson
Medical Research Council Cognition and Brain Sciences Unit
University of Cambridge
15 Chaucer Road
Cambridge CB2 7EF
United Kingdom
Telephone: +44 (0)1223 355 294

Email: rik.henson@mrc-cbu.cam.ac.uk

Running title: Evaluating models of the ageing BOLD response

Abstract

Neural activity cannot be directly observed using fMRI; rather it must be inferred from the haemodynamic responses that neural activity causes. Solving this inverse problem is made possible through the use of forward models, which generate predicted haemodynamic responses given hypothesised underlying neural activity. Commonly-used haemodynamic models were developed to explain data from healthy young participants; however studies of ageing and dementia are increasingly shifting the focus towards elderly populations. We evaluated the validity of a range of haemodynamic models across the healthy adult lifespan: from basis sets for linear convolution models (implemented in the general linear models used to analyse most fMRI studies) to nonlinear fitting of a parameterised haemodynamic response function (HRF) and nonlinear fitting of a biophysical generative model (haemodynamic modelling, HDM). Using an exceptionally large sample of participants, and a sensorimotor task optimized for detecting the shape of the HRF, we first characterised the effects of age on descriptive features of the HRF (e.g., peak amplitude and latency). We then compared these to features from more complex nonlinear models, fit to four regions of interest engaged by the task, namely left auditory cortex, bilateral visual cortex, left (contralateral) motor cortex and right (ipsilateral) motor cortex. Finally, we validated the extent to which parameter estimates from these models have predictive validity, in terms of how well they predict age in cross-validated multiple regression. We conclude that age-related differences in the HRF can be captured effectively by models with three free parameters. Furthermore, we show that biophysical models like the HDM have predictive validity comparable to more common models, while additionally providing insights into underlying mechanisms, which go beyond descriptive features like peak amplitude or latency. Here, the HDM revealed that most of the effects of age on the HRF could be explained by an increased rate of vasoactive signal decay and decreased transit rate of blood, rather than changes in neural activity per se. However, in the absence of other types of neural/haemodynamic data, unique interpretation of HDM parameters is difficult from fMRI data alone, and some brain regions in some tasks (e.g, ipsilateral motor cortex) can show responses that are more difficult to capture using current models.

Introduction

Functional magnetic resonance imaging (fMRI) is often used to infer changes in neural activity across the brain as a function of a task or the individual being scanned. However, the Blood Oxygenation Level Dependent (BOLD) signal measured in most fMRI experiments is a complex function of neural activity, neurovascular coupling, blood flow, blood oxygenation and blood volume (Buxton et al., 1998). Therefore, differences in the BOLD signal across individuals can be caused by differences in haemodynamics, rather than in neural activity per se. Here we focus on the effects of age, since it is likely that age affects vascular factors as well as neural activity (Handwerker et al., 2012; Wright and Wise, 2018). Indeed, neurovascular ageing is not only a confound in fMRI, but an important process to assess alongside neural ageing (Abdelkarim et al., 2019; Tsvetanov et al., 2021).

The most common way to model the BOLD impulse response to an experimental input (e.g., stimulus or trial, assumed to cause a short burst of neural activity) is with a linear convolution model, in which a delta function is convolved with a basis set of temporal functions, typically lasting 0-30s post-stimulus. After convolution with multiple, sequential inputs in a typical fMRI experiment, each convolved basis function forms a separate regressor within a general linear model (GLM). When fitting the GLM to the fMRI timeseries from a voxel or region of interest (ROI), the weights (parameter estimates) for each such function can be estimated using ordinary least squares minimisation. There are multiple options for such basis sets (Henson, 2004), but a common one is

the Finite Impulse Response (FIR) basis set, which consists of multiple "top-hat" functions, e.g., 16 contiguous time bins of 2 s duration, spanning from 0 to 32 s post-stimulus (Figure 1).

Using such a flexible basis set (whether the FIR set in the time domain, or sinusoidal functions of periodic frequencies in the Fourier domain), the first study to address age effects on the BOLD response failed to find differences in HRF shape between young versus older participants (D'Esposito et al., 1999). However, like many studies, this study used relatively small groups (n=32 young and n=20 older), and the older volunteers may have been healthier than average for their age (i.e, not representative of population). Furthermore, tests using such flexible basis sets may be less sensitive to subtle statistical differences in HRF shape.

Even though the BOLD response can vary across individuals and brain regions, it has a typical shape: peaking around 5s and with an undershoot lasting 10-30s: often called the canonical haemodynamic response function (HRF; see Figure 1 for an example). Thus a common approach is to assume this canonical form, and allow for some variability around this by adding, e.g, its partial derivatives in SPM's "informed basis set", or the first few singular vectors in FSL's data-driven FLOBS approach (Woolrich et al., 2004; see Lindquist et al., 2009, for yet further models of HRF shape). Greater sensitivity to ageing may be achieved by focusing on specific features of the HRF shape, such as latency of its peak. For example, Taoka et al. (1998) and Handwerker et al. (2007) found an age-related delay in the time to peak, while Richter and Richter (2003) and Aizenstein et al. (2004) found age-related reductions in the magnitude of the undershoot. By contrast, other studies have reported no effect of age. For example, Ward et al. (2008) found no age effects on the HRF in contralateral motor cortex during hand grips, while Grinband et al. (2017) found no effects of age on various HRF features, once controlling for possible age-related differences in neural activity. Moreover, even if differences are found in the shape of the HRF (e.g, peak latency), relating these to differences in underlying neural activity is far from simple, given that the mapping from neural activity to BOLD response is nonlinear (see Discussion).

Again, the above studies involved relatively small samples, whose recruitment may have been biased towards younger and/or older volunteers who were healthier than average. In the largest study of HRF shape of which we are aware, West et al. (2019) compared a younger (n=74) and an older (n=173) group. They found that HRFs differed between groups in auditory, visual and motor cortices, with increased time-to-peak and decreased peak amplitude in older compared to younger adults. These groups are actually a sub-set of the full set of n=645 participants used here. These participants come from the CamCAN cohort (www.cam-can.org), approximately uniformly distributed across 18-88 years of age. These participants were recruited using an opt-out procedure after contact from local surgeries, making them more representative than typical studies with opt-in recruiting via adverts. More specifically, these data come from the CamCAN "sensorimotor" task, in which participants saw brief trials consisting of bilateral visual checkerboards and/or auditory tones (see Shafto et al., 2014, for more details). Importantly, the time between trials (in terms of the stimulus onset asynchrony, SOA) was deliberately optimised for detecting the shape of the HRF, by using a 255-length "m-sequence" (Buračas and Boynton, 2002). Here we extend the work of West et al. (2019) by examining the parametric effect of age across the whole sample, and more importantly, by directly comparing different linear and nonlinear methods to characterise age-related changes in the HRF.

This study addresses four objectives:

1. To characterise the effects of age on the BOLD impulse response, using a model that makes minimal assumptions about the form of the response, at the cost of a large number of unconstrained parameters (a 32-parameter Finite Impulse Response model, FIR).

2. To compare these effects of age against more parsimonious models that are intended for routine use in neuroimaging analysis. These models that have fewer parameters (and are therefore statistically more efficient) but make stronger assumptions: specifically, the 3-parameter, "canonical" temporal basis set used for GLMs in the SPM software package (Can3) and a potentially more flexible, 4-parameter Non-Linear Fitting approach (NLF4).

3. To go beyond the descriptive statistics provided by the models above, and fit a mechanistic, biophysical Haemodynamic Model (HDM) to the same data. This model explains the effects of ageing in terms of 3 biologically-meaningful parameters: neural efficacy, neurovascular signal decay rate and haemodynamic transit rate. A version of this HDM is widely used as part of the Dynamic Causal Modelling (DCM) framework used for connectivity analysis (Friston et al., 2003).

4. To test whether the models' parameters have real-world (predictive) validity. We did this by evaluating how well all the above models can predict the age of previously unseen participants, using cross-validation accuracy.

Methods

Participants

The N=645 participants were aged 18-88 (approximately half of each sex), fluent English speakers and in good physical and mental health based on the CamCAN cohort exclusion criteria, which excluded volunteers with a low Mini Mental State Examination score (<24), serious current medical or psychiatric problems or poor hearing or vision, and standard MRI contraindications. None showed evidence of neurodegeneration in their structural MRIs (though we cannot rule out sub-clinical effects of neural and/or vascular disease that increase with age). For more details, see Shafto et al. (2014).

Paradigm

Each trial consisted of two checkerboards presented either side of a central fixation cross (34 ms duration) and/or a simultaneously-onsetting binaural auditory tone (300 ms duration), followed a fixation-only screen. The auditory tones were one of three equiprobable frequencies (300 Hz, 600 Hz, or 1200 Hz). Frequency was not relevant to the task or current hypotheses, so is collapsed here. The majority of the trials (120 of the 128) involved both visual and auditory stimuli, which required a keypress with the right index finger; the remaining 8 randomly-intermixed "catch" trials had only visual (4 trials) or only auditory (4 trials) stimuli, and participants were told to withhold their responses to such "unimodal" trials. The latter ensured that participants needed to pay attention to both modalities in all trials.

The SOA between trials was determined by a second-order m-sequence (m = 2; Buračas and Boynton, 2002), in which half of the sequence elements were null events. The minimal stimulus onset asynchrony (SOA) was 2 s, with the null events resulting in SOAs that ranged from 2 to 26 s. There was an additional random jitter drawn from a uniform distribution of 0.1-0.3s before stimulus

onset, which meant that the effective sampling rate of the HRF across trials was 0.12s (Josephs & Henson 1999).

MRI data acquisition

The MRI data were collected using a Siemens Trio 3T MRI Scanner system with a 32-channel head coil. A T2*-weighted echo planar imaging (EPI) sequence was used to collect 261 volumes, each containing 32 axial slices (acquired in descending order) with slice thickness of 3.7 mm and an interslice gap of 20% (for whole-brain coverage including cerebellum; repetition time = 1970 ms; echo time = 30 ms; flip angle = 78°; field of view = 192 mm x 192 mm; voxel size 3 x 3 x 4.44 mm). Higher resolution (1 x 1 x 1 mm) T1- and T2-weighted structural images were also acquired to aid registration across participants. For more details, see https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/pdfs/CAMCAN700_MR_params.pdf.

MRI data preprocessing

MRI preprocessing used the SPM12 software (Wellcome Centre for Human Neuroimaging; https://www.fil.ion.ucl.ac.uk/spm), release 4537, implemented in the Automatic Analysis pipeline, release 4.2 (Cusack et al., 2015). Preprocessing is described in Taylor et al. (2017), but in brief, structural images were rigid-body registered to a Montreal Neurological Institute (MNI) template brain, bias corrected, segmented, and warped to match a grey matter template created from the whole CamCAN Stage 2 sample using SPM's DARTEL toolbox. This template was subsequently affine transformed to standard MNI space. The functional images were spatially realigned, interpolated in time to the middle slice to correct for the different slice acquisition times, rigid-body coregistered to the structural image, transformed to MNI space using the warps and affine transforms from the structural image, and resliced to 3 mm x 3 mm x 3 mm voxels. Finally, residual noise was minimised by performing wavelet-based de-spiking (Patel et al., 2014) and the despiked images were smoothed by a 10mm isotropic FWHM Gaussian kernel. An XML summary of preprocessing can be found here: https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/ImagingScripts/mri_aa_release004_roistreams_v1_tasklist.xml.

Single-participant (1st-level) models

Two models were run, where neural activity was assumed to be: 1) locked to the onset of each audiovisual stimulus or 2) locked to the subsequent key-press. We included unimodal trials for the stimulus-locked model (since they still involved stimulation), but only included them for the response-locked model if the participant incorrectly pressed the key.

GLMs for each participant were created and estimated in SPM release 7771[1] in Matlab R2020b (https://www.mathworks.com). A single vector of delta functions at the onset of each trial was convolved with each of the temporal basis functions (FIR or SPM's informed basis set; see below) in a high-resolution space with 32 time bins every TR, and subsequently down-sampled at the middle time bin (to match the reference slice for the slice-timing correction). Six additional regressors, representing the three rigid body translations and the three rotations estimated in the realignment stage, were included to capture residual movement-related artifacts, plus the timeseries extracted for WM and CSF compartments, to further remove non-neural noise. Finally, the data were scaled to a grand mean of 100 over all voxels and scans, and the model fit to the data in each voxel. The autocorrelation of the error was modelled using a "first-order autoregressive plus

[1] except that orthogonalisation of regressors within each condition was switched off within spm_fMRI_design.m (see https://github.com/RikHenson/AgeingHRF)

white-noise" model, together with a set of cosines that were used to high-pass filter both the model and the data at a frequency cutoff of 1/128 Hz. The GLM and noise model were estimated using restricted maximum likelihood (ReML), and the resulting estimate of error autocorrelation used to prewhiten both the model and the data. Finally, OLS was used to estimate the model parameters for the whitened data.

Group model for FIR basis set

For the FIR basis set, the parameter estimate for each of the 32 time bins for each participant was entered into a second-level GLM (equivalent to a repeated-measures ANOVA), as well as a (linear) modulation by each FIR parameter by age (see Figure 2).

ROI definition

To maximise sensitivity to age, while not making assumptions about the shape of the HRF, masks for each ROI were defined from an F-contrast in the second-level group FIR analysis that spanned the (linear) effects of age. For the left auditory cortex (lAC) and bilateral visual cortex (bVC), these came from the stimulus-locked model, while for the left and right motor cortex (lMC and rMC), these came from the response-locked model. The F-values were thresholded at F>12 to define clusters of contiguous voxels[2]. This resulted in a mask of 559 and 371 voxels for lAC and bVC, respectively, and 219 and 481 voxels for lMC and rMC, respectively.

To allow for variability across participants in their most responsive voxels, we sub-selected those voxels within the above masks that showed a significant F-contrast across time bins within their first-level FIR models. We started with a threshold of p<.005 uncorrected, but iteratively relaxed this (to p<.05 and p<.5) until at least 3 voxels remained for each participant. The number of remaining voxels did not covary appreciably with age and had median values of 351 for lAC, 139 for bVC, 158 for lMC and 41 for rMC.

The fMRI timeseries for each ROI were then summarised in terms of the first temporal component from a singular-value decomposition (SVD) across the voxels remaining from the above procedure, using standard tools for ROI extraction in SPM. This entailed the high-pass filtering, pre-whitening and correction for confounds described above during GLM estimation.

HRF models

For each ROI, we fit 4 types of model: two linear models (estimated using OLS) and two nonlinear models (fit by gradient descent). The two linear models corresponded to 1) the 32 x 1s bin FIR basis set described above (and used to define the ROIs) – henceforth the "FIR32" model – and 2) SPM's informed basis set, consisting of a canonical HRF and two partial derivatives – henceforth the "Can3" model. The two nonlinear models were based on shape-matching of the HRF using 4 parameters – the "NLF4" model – and a biophysical generative model with 3 parameters – the "HDM3" model. These are detailed below.

FIR32 model

The "FIR32" model consisted of 32 top-hat temporal basis functions of 1s duration that captured the first 32 seconds of post-stimulus time (Figure 1A). This FIR basis set is flexible, but can also capture

[2] The degrees of freedom in this model are so high (>20,000) that nearly every active voxel survives correction for multiple comparisons, hence the arbitrary threshold to define reasonably-sized ROIs.

non-haemodynamic trial-locked effects (e.g, motion of the eyeballs during first time bin, in response to a visual stimulus).[3] This approach is also called selective averaging (Dale and Buckner, 1997).

Can3 model

The canonical HRF used in SPM is a mixture of two gamma functions, one capturing the positive peak around 5s and one capturing the negative undershoot peaking around 15s. The parameters of these gamma functions were estimated from approximating the first singular vector of stimulus-locked responses in an fMRI experiment (Friston et al., 1998). The second basis function in this "informed" basis set is the temporal derivative of the canonical HRF, obtained by the finite difference after shifting the onset of the canonical HRF by 1 second. The third basis function is the dispersion derivative, obtained by the finite difference after increasing the dispersion of the gamma function for the peak response by 1%. See Figure 1B and Henson (2004).

NLF4 model

The NLF4 model refers to nonlinear fitting of a template HRF to each individual's FIR fit, using a first order expansion with respect to amplitude and with respect to time (i.e., four parameters in total). More precisely, the following model is fit:

$$\hat{Y}(t) = a_1 Y\left((t - t_0) + \frac{(t - t_0)}{t_1}\right) + a_0$$

where vector Y is the template HRF, derived from the first singular vector from an SVD of the FIR values across all participants, t_0 is the latency offset (constant delay), t_1 is the latency scaling (cumulative delay), a_0 is the amplitude offset and a_1 is the amplitude scaling – see Figure 2C. Thus for latency, t_0 moves the whole HRF forwards or backwards in time, while t_1 extends the HRF in proportion to the time relative to trial onset. This model has greater flexibility, e.g, in separating amplitude from different types of latency, than the above linear basis sets. It is fit by gradient ascent on the correlation between (normalised) template and (normalised) individual FIR. We have previously applied this approach to modelling the effect of age on the latency of evoked MEG responses in the same task and participants (Price et al., 2017), and use the same fitting procedure described there. For similar approaches to nonlinear HRF fitting, see Kruggel & Yves Von Cramon (1999).[4]

[3] Trial-locked FIR effects that do not resemble a typical HRF are sometimes rejected as "non-haemodynamic", but in this case, such an implicit model of the HRF would be better put explicitly into the GLM in the first place, e.g, by using a more constrained basis set. Nonetheless, if the neural activity associated with a trial is sustained (more than a second or so) and unknown in form, then a FIR basis set can be a useful catch-all.

[4] Note that, unlike the other 3 models, the NLF4 approach is fit to the FIR parameters (i.e, peristimulus time), rather than the original fMRI timeseries. One could take the best-fitting HRF shape from such nonlinear fitting of the FIR parameters, and insert it back into the first-level GLM for the fMRI timeseries, in order to estimate a single scaling (amplitude) parameter. This would enable inference to be performed on a single parameter, while allowing for variability in the HRF shape across ROIs and/or across participants. This has the potential to be more robust than re-inserting the FIR estimate of HRF shape for a given participant and ROI, if the latter is noisy, owing to the effective "regularisation" of the HRF by the average shape across participants and/or ROIs (similar to the formal hierarchical model used for HDM below). Nonetheless, care would be needed for regions like rMC that do not show a consistent HRF shape across participants (see Results).

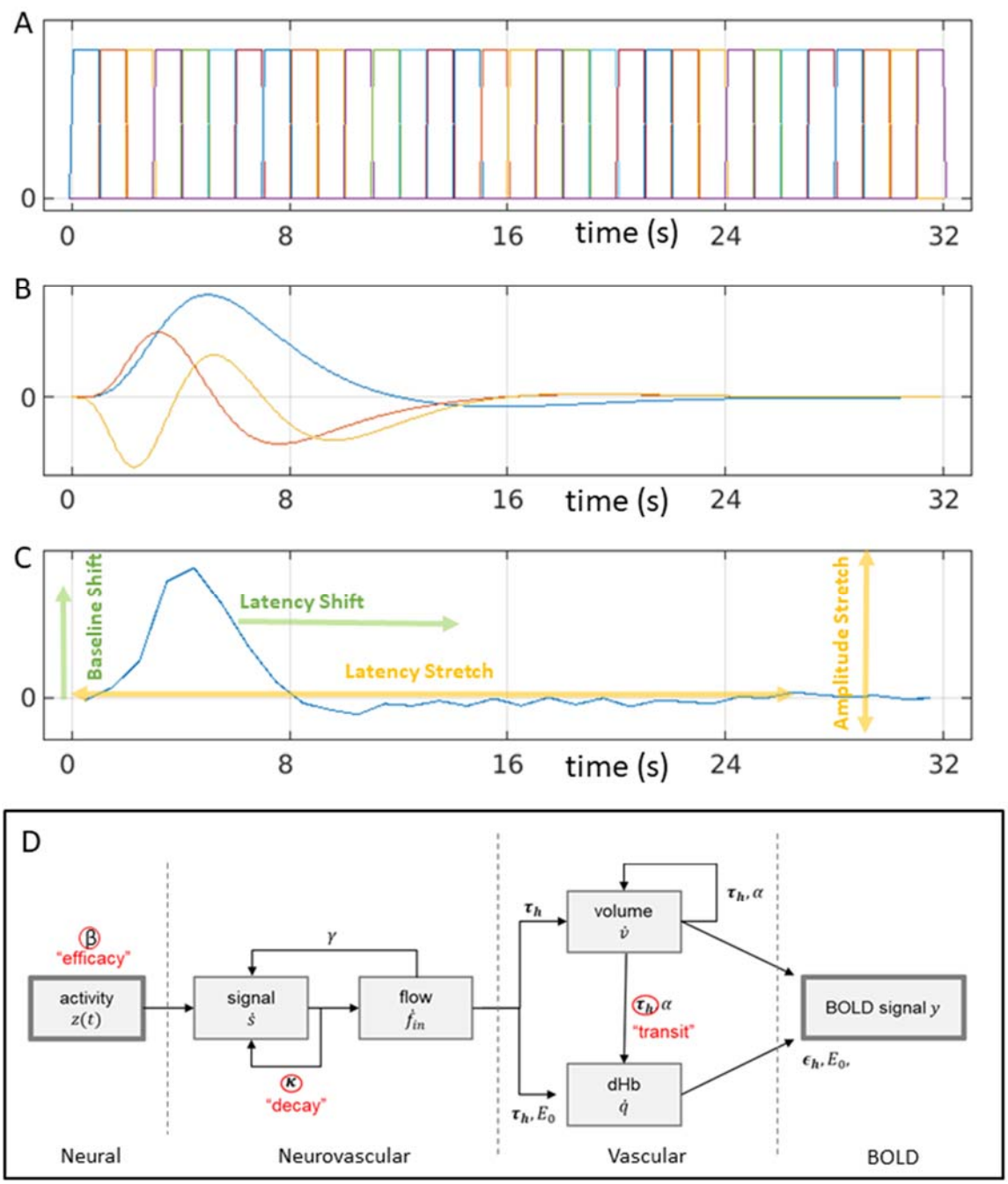

Figure 1. HRF models: (A) linear Finite Impulse Response basis set with 32 x 1s bins (FIR32); (B) canonical HRF linear basis set (blue), its temporal derivative (red) and dispersion derivative (yellow) (Can3); (C) nonlinear fitting of average empirical HRF for a given ROI (here lAC) using amplitude and latency offsets and scalings (NLF4); (D) nonlinear haemodynamic modelling using differential equations with three free parameters (in red) (HDM3).

HDM3 model

This nonlinear model is a biophysical, generative model based on differential equations that can be integrated to produce HRFs, as a function of biologically relevant parameters (e.g, vasodilatory time constants), rather than parameters that simply describe the shape of the HRF (as in NLF4 above). We used a well-established haemodynamic model (Friston et al., 2000; Stephan et al., 2007) with some updates to the parameters. The form of the model is explained in detail in Appendix 5 of Zeidman et al. (2019), and it is shown schematically in Figure 1D. To briefly reprise, there are four hidden states: vasoactive signal s, regional cerebral blood flow f_{in}, venous volume v and deoxyhaemoglobin q. Their dynamics are governed by the following equations:

$$\dot{f}_{in} = s \tag{1}$$

$$\dot{s} = \beta z(t) - \kappa s - \gamma(f_{in} - 1) \tag{2}$$

$$\tau_h \dot{v} = f_{in}(t) - f_{out}(v,t) \tag{3}$$

$$\tau_h \dot{q} = f_{in}(t)\frac{1-(1-E_0)^{\frac{1}{f_{in}(t)}}}{E_0} - \frac{f_{out}(v,t)q(t)}{v(t)} \tag{4}$$

$$y = 100 \cdot V_0\left(k_1(1-q) + k_2\left(1-\frac{q}{v}\right) + k_3(1-v)\right) \tag{5}$$

$$k_1 = 4.3 \cdot \vartheta_0 \cdot E_0 \cdot T_E \tag{6}$$

$$k_2 = \epsilon_h \cdot r_0 \cdot E_0 \cdot T_E \tag{7}$$

$$k_3 = 1 - \epsilon_h \tag{8}$$

Equations 1-2 model the vasoactive signal s, which induces blood flow f_{in} in response to neural activity $z(t) \in \{0,1\}$ with neural efficacy parameter β. This part of the model subsumes a variety of neurovascular coupling mechanisms, such as nitric oxide signalling. The decay of the vasoactive signal is governed by rate parameter κ, while rate parameter γ provides autoregulatory feedback[5]. Together these equations form a coupled oscillator.

Equations 3-4 model the cerebral blood volume v relative to the blood volume at steady state, as well as the proportion of deoxyhaemoglobin relative to rest, q. These equations correspond to the balloon model of Buxton et al. (1998), where parameter $1/\tau_h$ is the rate of blood flow through the vessel, and parameter E_0 is the resting oxygen extraction fraction. The blood outflow from the vascular compartment is $f_{out}(v,t) = v(t)^{1/\alpha}$, in which Grubb's exponent α relates cerebral blood flow (CBF) and cerebral blood volume (CBV).

Finally, Equations 5-8 model the generation of the fMRI BOLD signal (Obata et al., 2004, Stephan et al., 2007). Parameter V_0 is the resting blood volume fraction, ϑ_0 is the frequency offset at the outer surface of magnetised vessel, r_0 is the slope of intravascular relaxation rate against O_2, T_E is the echo time and ϵ_h is the ratio of intra-vascular to extra-vascular components of the gradient echo signal.

[5] The autoregulatory feedback parameter may not have a direct physiological correlate at the vascular level. Indeed, it has been proposed that neurovascular coupling can be modelled as a purely feedforward system, in conjunction with inhibitory neural activity (Havlicek et al., 2015). In the absence of a neural model, the purpose of the feedback parameter here is simply to enable damped oscillatory dynamics on a physiologically plausible timescale. See Appendix 5 of Zeidman et al. (2019) for details.

In the present HDM3 model, we allowed three parameters to be estimated from the data: 1) neural efficacy parameter, β (as a proxy for neural activity), 2) rate of decay of vasoactive signal, κ and 3) transit rate of blood flow, $1/\tau_h$. The remaining parameters were fixed based on empirical priors (see Supplementary Material).

To fit the HDM, we used https://github.com/pzeidman/HDM-toolbox, which requires the SPM software package (https://www.fil.ion.ucl.ac.uk/spm/software). The model was fit using the standard Bayesian model fitting scheme in the SPM software, called Variational Laplace (Friston et al., 2007). For participant i with timeseries Y_i, this returns a multivariate normal probability density over the parameters, $P(\theta_i|Y_i) \sim N(\mu_i, \Sigma_i)$. The estimated values of the parameters μ_i were those that maximised the log-evidence for the model, as approximated by the free energy $F_i \approx \ln P(Y_i)$.

While this scheme provides parameter estimates for individual participants, μ_i, it also yields the estimated covariance of the parameters Σ_i. To convey both μ_i and Σ_i to the group level, we used the Parametric Empirical Bayes (PEB) framework in SPM (Friston et al., 2016, Zeidman et al., 2019). This is a Bayesian hierarchical linear regression model applied to the estimated parameters of all participants' models (see Supplementary Material for more details). We then tested the evidence for the presence versus absence of each covariate (mean and effect of age) on each haemodynamic parameter using Bayesian Model Reduction (BMR, Friston et al., 2018), which iteratively prunes mixtures of group-level parameters from the PEB model, where doing so does not reduce the free energy.

Access to data and code

The raw functional and structural images and their preprocessed versions are available from https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/. The Matlab code for extracting data from ROIs and fitting the models above is available in https://github.com/RikHenson/AgeingHRF.

Results

Whole-brain analyses

Separate analyses were performed to identify BOLD responses time-locked to stimulus presentation and BOLD responses time-locked to button presses. For each analysis, the parameter estimates for each basis function (time bin) from the FIR32 GLM were entered into a second-level group model across participants to estimate the mean and the linear effect of age for each basis function.

An F-test across the mean values from the stimulus-locked analysis showed a large number of voxels that survived correction for multiple comparisons, the most significant of which were in bilateral auditory cortex, bilateral visual cortex and left (contralateral) motor cortex (Figure 2A). Similar regions showed an effect of age, as did similar regions in the response-locked analyses, but with the addition of right (ipsilateral) motor cortex (Figure 2B).[6]

The stimulus-locked, mean FIR estimates from the peak of the left auditory cortex cluster showed a typical HRF, peaking around 4-5s and with a more sustained undershoot (Figure 2C, top). Interestingly, the estimates for the effect of age on each time bin (Figure 2C, bottom) showed a profile that resembled a temporal derivative of the mean HRF, such that the HRF became more delayed with age (which can be visualised by adding the bottom plot in Figure 2C to the top plot). Similar results were found for the visual and left motor clusters (not shown). The response-locked mean FIR for the rMC (Figure 2D, top) however was not typical, and it was the effect of age (Figure 2D, bottom) that appeared more like an HRF. This pattern can be explained when plotting the HRF by age in the next section, where it becomes apparent that the young showed a negative BOLD response that disappeared with age (see also Mayhew et al., 2022; Tak et al., 2021).

[6] The SPMs for the stimulus-locked and response-locked group FIR models were very similar, as would be expected since the median across participants of their median reaction time (RT) across trials was 282ms (range = 182ms - 679ms). Nonetheless, this difference in RTs (and their variance over trials) is likely to be important for latency analyses below.

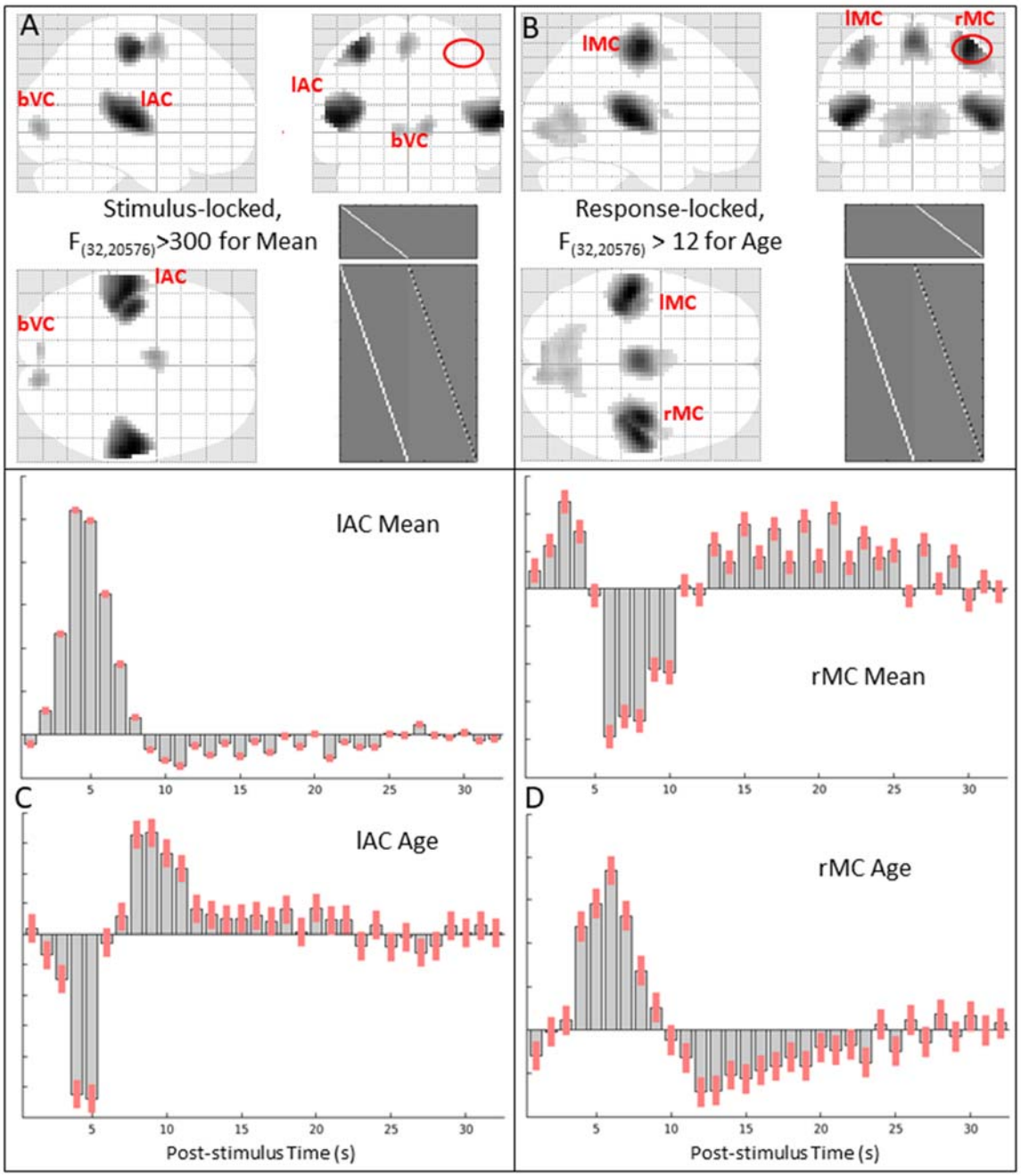

Figure 2. Panels A and B show maximal intensity projections (MIPs) of statistical parametric maps (SPMs) of the thresholded F-contrast for (A) the mean effect across participants, locked to audio-visual stimulation, and (B) the (linear) effect of age, locked to the right finger press. The clusters corresponding to the four functional ROIs analysed below (lAC, bVC, lMC and rMC) are labelled on the sagittal, coronal and transverse sections, along with the F-contrast and the design matrix (bottom right). The red ellipse highlights that the rMC cluster does not show a strong (compared to the significance in other ROIs) main effect across participants, but does show a strong effect of age. Panels C and D show the mean of each FIR parameter (upper plot) and effect of age on each parameter (lower plot) for the peak voxel from the (C) lAC ROI [-48 -24 +6] and (D) rMC ROI [+39 -18 +57] (note that the scale of the y-axis is in arbitrary units and differs across plots; for % signal change, see Figure 4).

314

Preliminary FIR models were used to define four functional ROIs: lAC, bVC, lMC and rMC (see Methods). FIR models were then fit to the first temporal component across voxels for each ROI. These are shown for each participant as heatmaps in the top row of Figure 3, while corresponding HRFs averaged within age tertiles are plotted in the top row of Figure 4. The effect of age for lAC, bVC and lMC can be seen as a reduction in peak amplitude, increase in peak dispersion (duration) and attenuation of the undershoot. The latter effect is particularly strong, with the undershoot almost vanishing for the oldest study participants. Interestingly, rMC shows a more triphasic response, with a large negative peak in young people that becomes attenuated as age increases (explaining the mean and age effects in Figure 2D). Note also that there is some variability in HRF shape across ROIs (see Supplementary Figure S3).

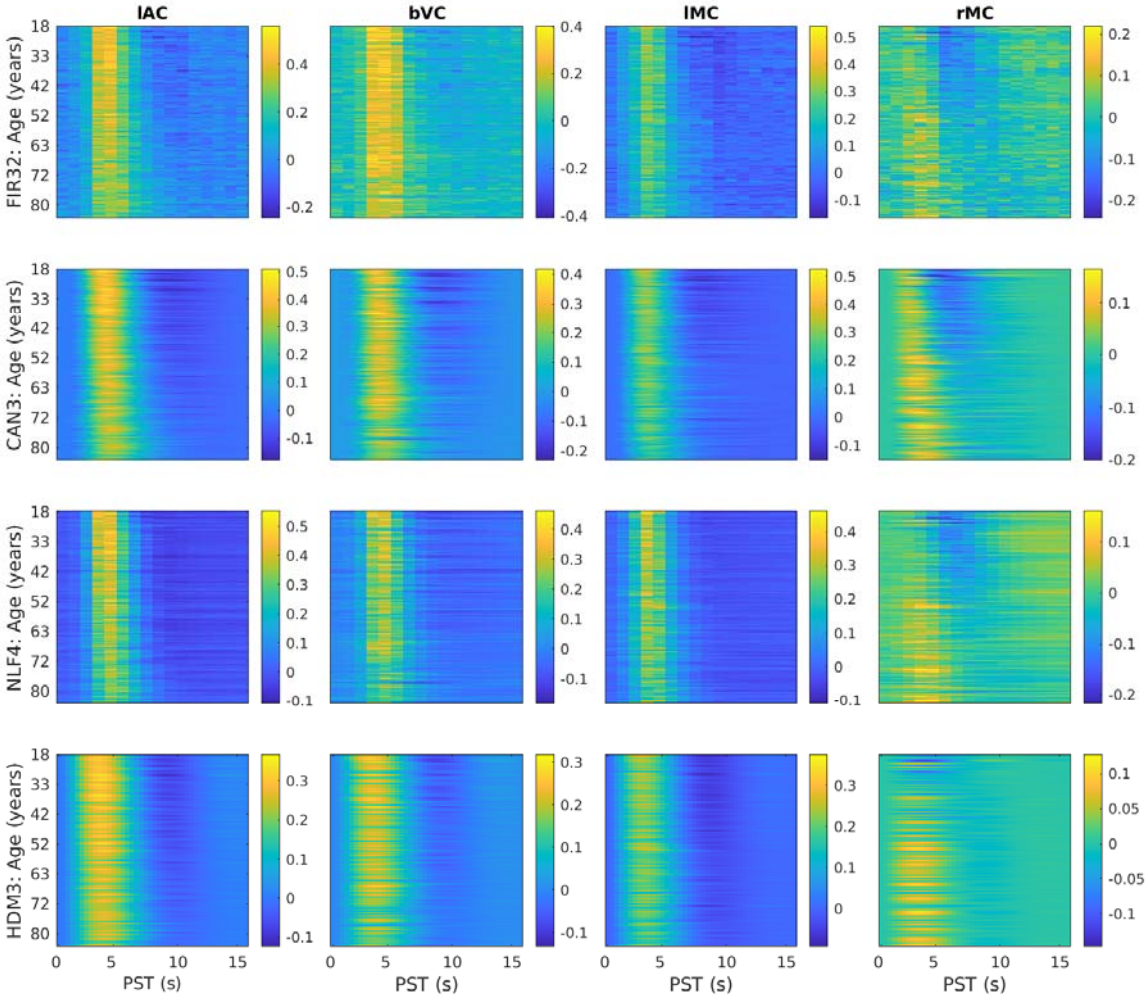

Figure 3. HRFs shown as heatmaps for each ROI (columns) and model (rows). The y-axis represents each participant, sorted by age, while the x-axis represents post-stimulus time (PST). The fits were smoothed across participants with a 5-participant running average. lAC = left auditory cortex, bVC = bilateral visual cortex, lMC = left (contralateral) motor cortex, rMC = right (ipsilateral) motor cortex. The lAC and bVC data are from the stimulus-locked model; lMC and rMC are from the response-locked model.

331

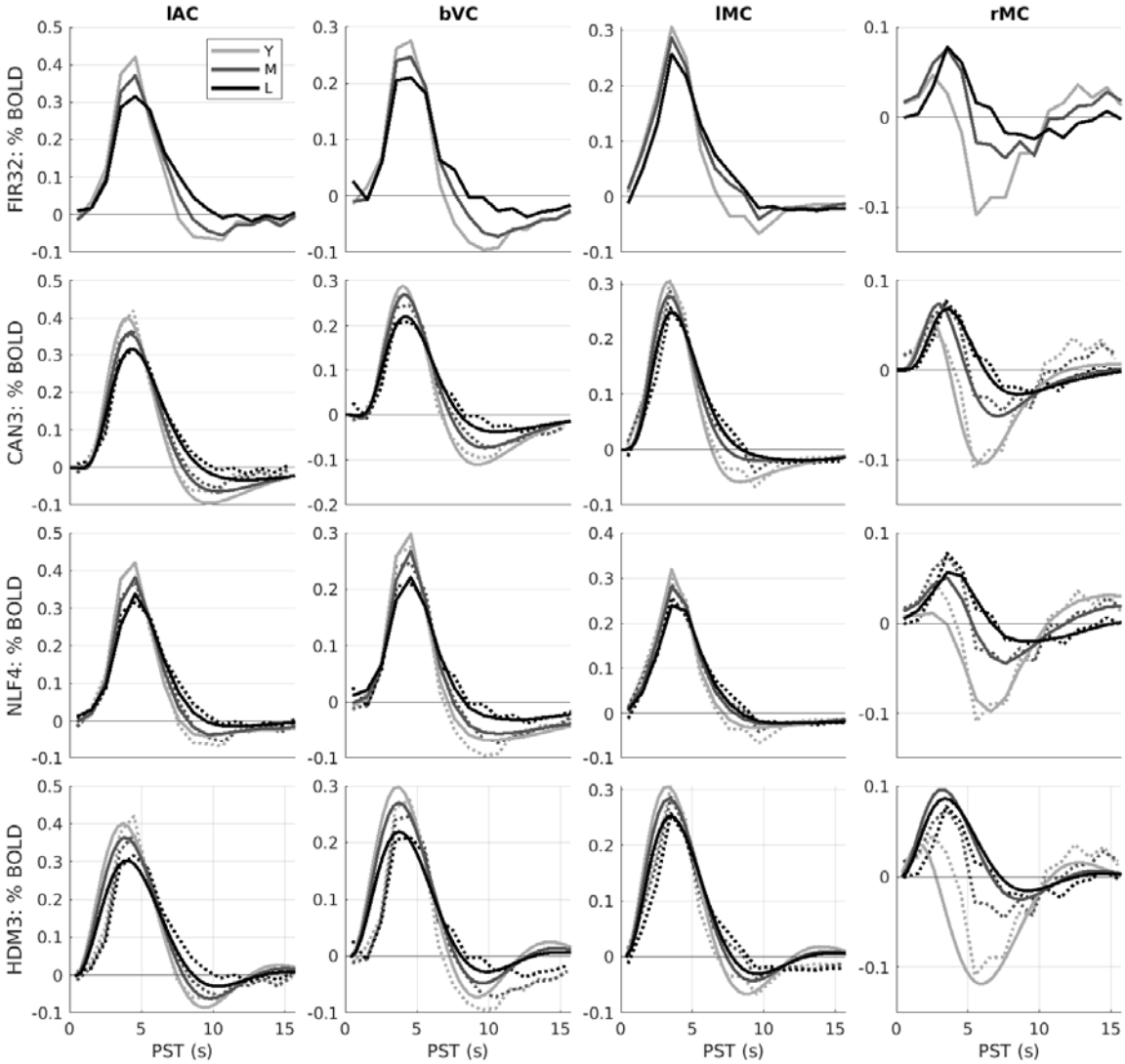

Figure 4. Average HRFs within age tertiles from each for each ROI (columns) and model (rows). Y = young; M = mid-life; L = late-life. The dotted lines in rows 2-4 are the (replotted) FIR estimates (i.e. same as top row) for reference. See Figure 3 legend for more details.

Features of FIR32 fits

One could extract a number of features (i.e., descriptive statistics) from the FIR fits to each participant. As an example, we defined the peak amplitude and peak latency from the FIR time bin with the maximal absolute value within the first 16s. There are clearly some outliers (and limited resolution for the peak latency), but for lAC, bVC and lMC, Spearman rank correlations showed significant effects of age in decreasing peak amplitude for lAC, bVC and lMC (top row of Figure 5), and also increasing peak latency in lAC (top left panel of Figure 6).

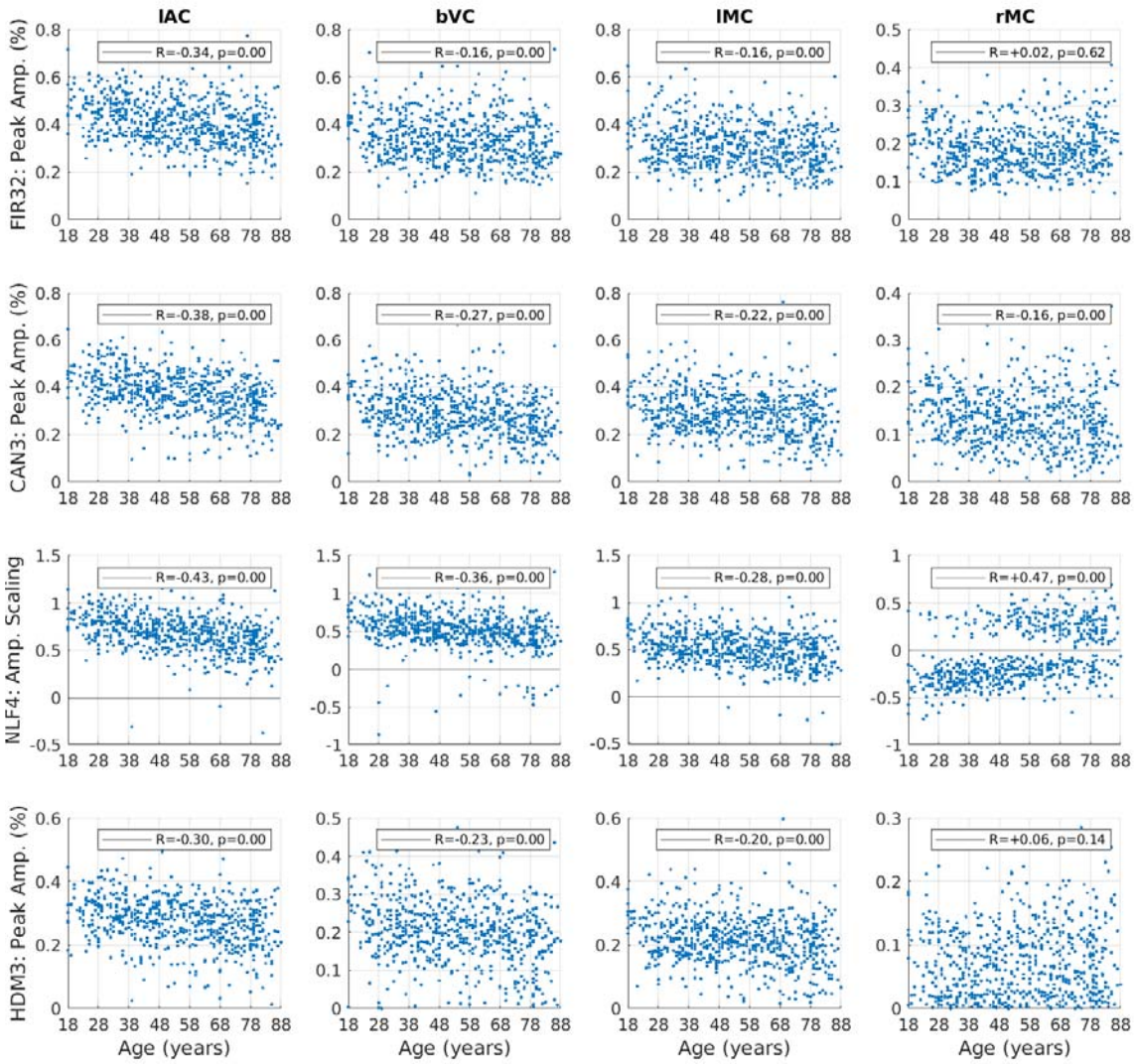

Figure 5. Amplitude estimates for each ROI (columns) and model (rows), plotted for each participant as a function of age. For all except NLF4, peak is determined by maximum absolute value of the fit within the first 16s. The R and p-value for Spearman rank correlations with age are shown in legend. See Figure 3 legend for more details.

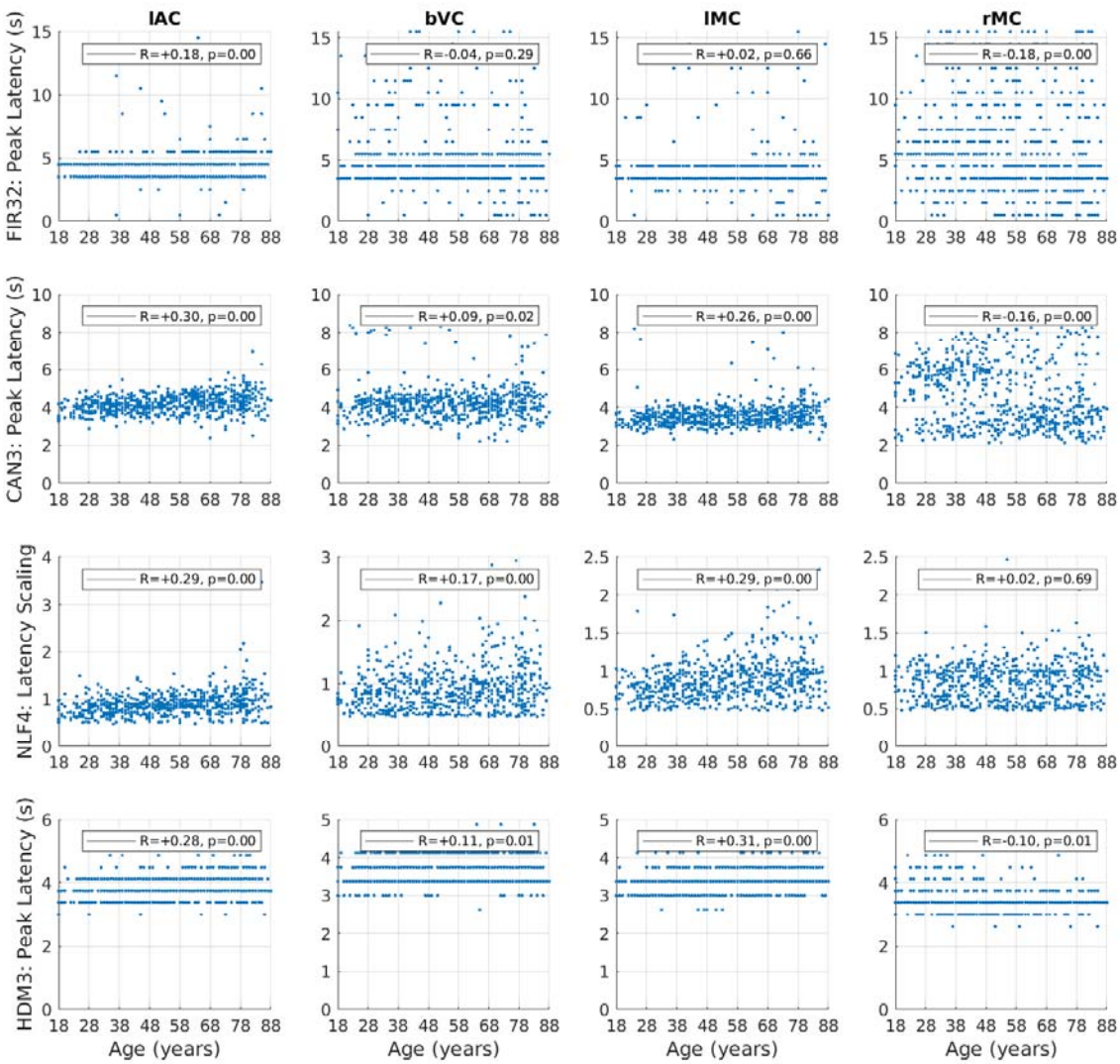

Figure 6. Latency estimates for each ROI (columns) and model (rows), plotted for each participant as a function of age. For all except NLF4, peak is determined by maximum absolute value of the fit within the first 16s. The R and p-value for Spearman rank correlations with age are shown in legend. See Figure 3 legend for more details.

We also examined the effect of age on the fMRI timeseries residuals for each ROI (Supplementary Figure S4). For lAC and bVC, there was negligible effect of age on the residual error in any model. This suggests, contrary to our expectations, that older participants did not have more (non-stimulus-locked) noise remaining in their fMRI data (e.g, residual effects of head motion), nor demonstrated more trial-to-trial variability in their HRF. However for lMC and rMC, the residual error did increase slightly with age. This could reflect greater age-related variability intrinsic to motor cortex, and/or greater such variability associated with analyses locked to button presses.

348 Can3 fits and HRF features

349 The second row of Figures 3-4 shows the fits of the Can3 basis set (see Methods), while the
350 second row of Figures 5-6 shows its estimates of peak amplitude and peak latency derived from the
351 reconstructed HRF for each participant. Figure 4 shows that the Can3 basis set is accurate in
352 capturing variation in the HRF shape across age tertiles and across ROIs (closely following the FIR
353 estimates in dotted lines). One possible limitation of the Can3 set is the inability to decouple the
354 amplitude of the peak and that of the subsequent undershoot (see, for example, lAC and bVC where
355 the magnitude of the undershoot is slightly over-estimated), though this could be fixed by adjusting
356 the scaling ratio of the two gamma functions that capture the peak and undershoot assumed by the
357 canonical HRF.[7]

358 The effects of age on peak amplitude and peak latency were stronger (at least in terms of
359 higher Spearman correlations) for the Can3 fits than the FIR fits (cf. first and second rows of Figures
360 5-6). For example, the effect of age on peak latency was now significant for all ROIs. This suggests
361 that the Can3 basis set is more sensitive to effects of age on features of the HRF shape (as would be
362 expected when the peak amplitude and latency are effectively derived from a weighted sum of many
363 time bins, rather than a single time bin in case of the FIR). Note however that for the rMC, Can3
364 shows that age decreases the (absolute) amplitude (contrary to the NLF4 model considered next).

365 NLF4 fits and HRF features

366 The third row of Figures 3-4 shows the fits of the NLF4 model (see Methods), while the third
367 row of Figures 5-6 shows its estimates of amplitude and latency. Note that the purpose of the NLF4
368 model is to estimate amplitude and latency directly, rather than derive post hoc from the peak of
369 the fitted responses, so the third row in Figure 5 reflects the NLF4 amplitude scaling parameter and
370 the third row in Figure 6 reflects the NLF4 latency scaling parameter. Indeed, this model
371 distinguishes two types of amplitude and latency – an offset (shift) and a scaling (stretch) – see
372 Methods (and see Supplementary Figure S5 for all four parameter estimates). Note that also that
373 this model is fit to the FIR (i.e. post-stimulus time) rather than the original timeseries.

374 The advantage of the NLF4 model can be seen in the generally higher correlations with age.
375 In Figure 5 for example, the amplitude scaling parameter shows a stronger effect of age than does
376 the peak amplitude estimated from fits of the FIR or Can3 basis sets. This is because the NLF4 model
377 explicitly separates the amplitude from the latency. Note that age increases the amplitude in rMC
378 because this is now a signed value (whereas the age decrease in rMC with the Can3 model reflects
379 the absolute value), while the bimodal distribution of NLF4 amplitude estimates in rMC reflects
380 uncertainty about that sign when amplitude is close to zero.

381 The NLF4 latency scaling parameter in Figure 6 also tends to show stronger effects of age, at
382 least in bVC and lMC. Interestingly, there is no significant effect of age on the latency parameter in
383 rMC, unlike the negative effect of age for the Can3 peak latency in rMC. This again reflects the better
384 ability of the NLF4 model to separate amplitude and latency. This can also be seen in the heatmaps
385 for rMC in Figure 3: the NLF4 model is able to capture a transition from a negative peak response in
386 younger people to a positive peak response in older people, with less change in latency of the peak,

[7] We did try a fifth model in which the gamma function for the peak and the gamma function for the undershoot were separate basis functions (along with their temporal derivatives, i.e. 4 basis functions in total), but this model did not perform as well (e.g., in the cross-validated prediction of age reported later). Also note that, according to the HDM model, the peak and undershoot are highly coupled, since they are part of the same damped oscillation in response to a brief burst of neural activity.

whereas the Can3 model must capture this transition by using the derivatives to shift earlier the positive peak of the canonical HRF and augment its undershoot, resulting in an age-related shift in peak latency when that latency is defined as the time of maximal absolute deviation from zero (i.e., corresponding to the peak of the "undershoot" for younger participants). This does not mean the NLF4 model is closer to the truth (the Can3 model may capture what is really happening); only that it is better suited to estimating features like amplitude and latency, if those features are of interest (and can be meaningfully interpreted; see Discussion).

HDM3 fits and its parameters

Whereas the NLF4 model is designed to fit features of the HRF (specifically amplitude and latency changes), the final model considered here is a biologically-plausible generative model of how the HRF is caused. While the full HDM potentially has multiple free parameters (see Supplementary Material), we focus on an HDM3 version in which we fixed several parameters to their prior expected value, leaving three free parameters: neural efficacy, neurovascular decay rate and haemodynamic transit rate.

The bottom row of Figures 3-4 shows the fits of the HDM3 model (see Methods), while the bottom row of Figures 5-6 shows estimates of amplitude and latency from the HDM3 fit. The HDM3 model does reasonably well for all ROIs, though appears to peak earlier than the FIR model, and struggles to capture age effects in rMC.

The three free parameters of the HDM3 model are shown as a function of age in Figure 7. Neural efficacy (top row) showed little effect of age in lAC, bVC and lMC, but a positive effect of age in rMC. Neurovascular decay rate (second row) showed strong positive effects of age in all ROIs except rMC, though there is evidence, particularly in rMC, of participants for whom the default value of 0.64 was sufficient. The haemodynamic transit rate (bottom row) showed strong negative effects of age in lAC, bVC and lMC, but not rMC. Again, there were many participants, particularly in rMC, for whom the default value of 1.02 was sufficient.

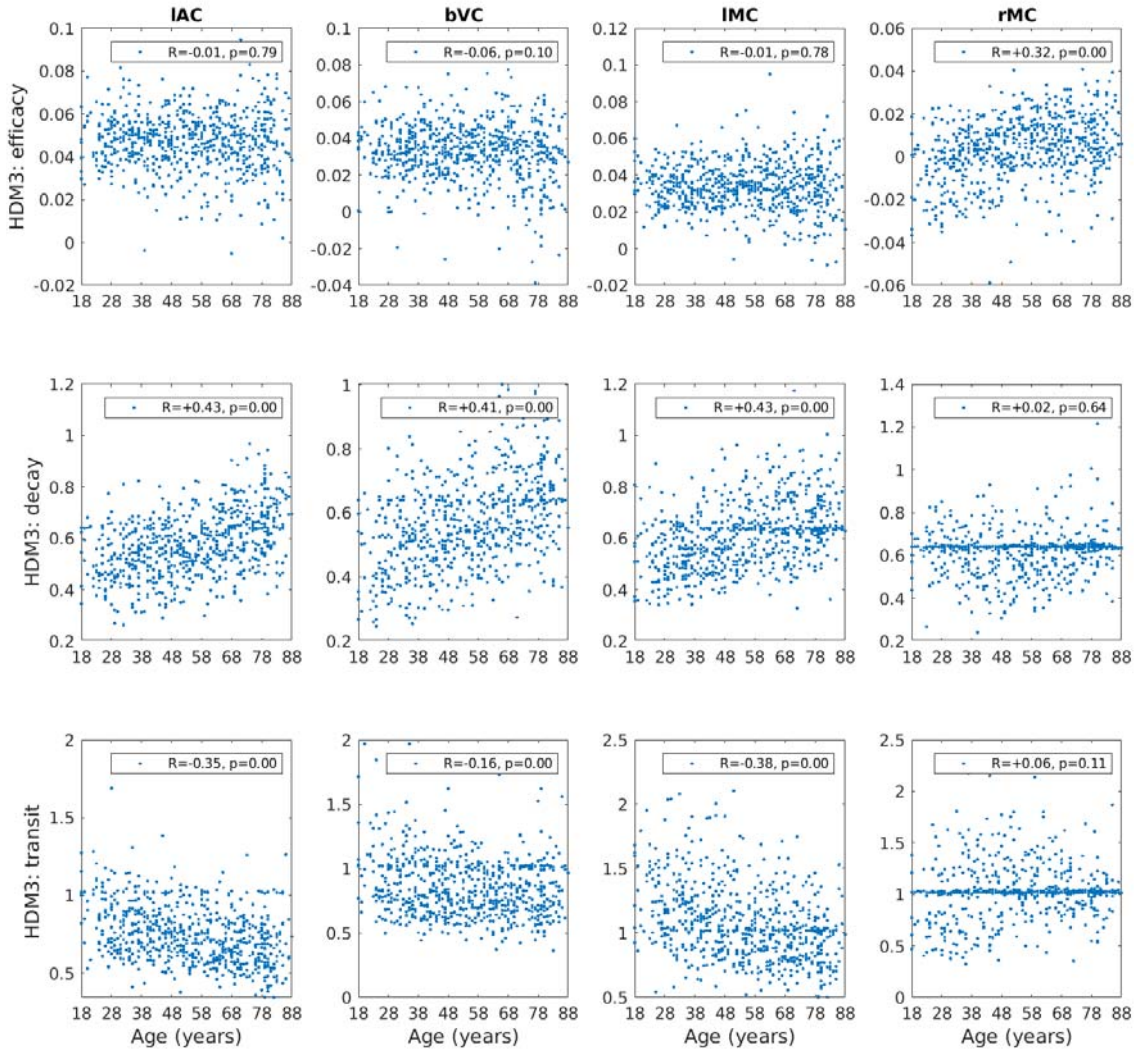

Figure 7. HDM3 parameters (rows) for each ROI (column) plotted for each participant as a function of age. The R and p-value for Spearman rank correlations with age are shown. Parameter estimates for decay and transit rates are after transforming the posterior expected values from their log-values back into original units (see Supplementary Material). See Figure 3 legend for more details.

Furthermore, we can incorporate the HDM results into a hierarchical linear model, in order to regularise the parameter estimates across participants, using SPM's PEB framework (see Methods). More specifically, we fit a group-level model that included a constant term (average parameter value across participants) and a linear age effect (slope of parameter change with age). We then used BMR to prune parameters that are not needed in order to maximise the model evidence for this group-level model. This analysis takes into account the posterior covariance between parameters,

effectively dropping parameters whose effects on the HRF fitting can be accommodated by another parameter.

The results of PEB-BMR on the linear effect of age on each parameter are shown in Figure 8 (values for the mean across participants are shown in Supplementary Figure S6). There was no need for age to moderate neural efficacy in lAC, bVC or lMC, but a compelling need to increase efficacy with age in rMC. Neurovascular decay rate showed an increase with age across all ROIs, while haemodynamic transit rate showed a decrease with age in lAC, bVC and lMC, but moderate increase with age in rMC.

Note that while these Bayesian linear regression (i.e., PEB) results were generally consistent with the classical correlations with age reported in Figure 7, this was not the case for the decay and transit parameters in rMC: the PEB analysis showed a positive effect of age on both, whereas there were no significant Spearman correlations in Figure 7. This could be due to the greater power afforded by the (parametric) PEB approach, or the fact that PEB takes into account the posterior covariance among parameters, which was ignored by the correlation analyses above.

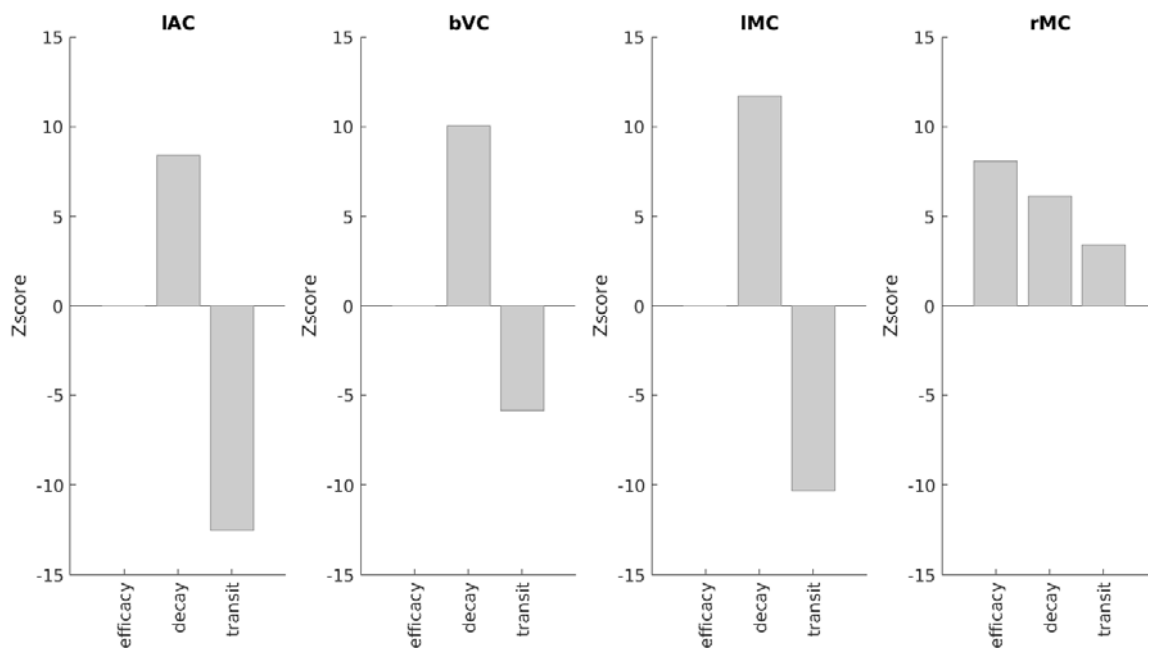

Figure 8. Results of PEB BMR on HDM parameters for each ROI for the linear effect of age across participants. Grey bars show Z-scores. Missing bars are parameters that BMR has removed as unnecessary (in terms of maximising evidence for PEB model). For the decay and transit parameters, units are log deviations.

Cross-validated Prediction of Age

Finally, we tested how well each of the four HRF models above could predict participants' chronological age based on their parameter estimates, using multiple linear regression. We used leave-one-out cross-validation (LOOCV) and calculated the absolute difference between the predicted and actual age. The results are shown in Figure 9. The performance of FIR32 and Can3

models did not differ significantly for any ROI (with a median error of approximately 11 years). It is interesting to note that FIR32 did not overfit, despite having ~10 times more parameters than the Can3 model. On the other hand, performance of the NLF4 model was significantly worse than the FIR32 model in all ROIs, suggesting that it was failing to capture an important aspect of ageing. The HDM3 model was better than the NLF4 model in all ROIs, but was significantly worse than the FIR32 model in lAC and rMC. It was particularly worse in rMC, suggesting that the BOLD impulse response in this region may be atypical (see Discussion).

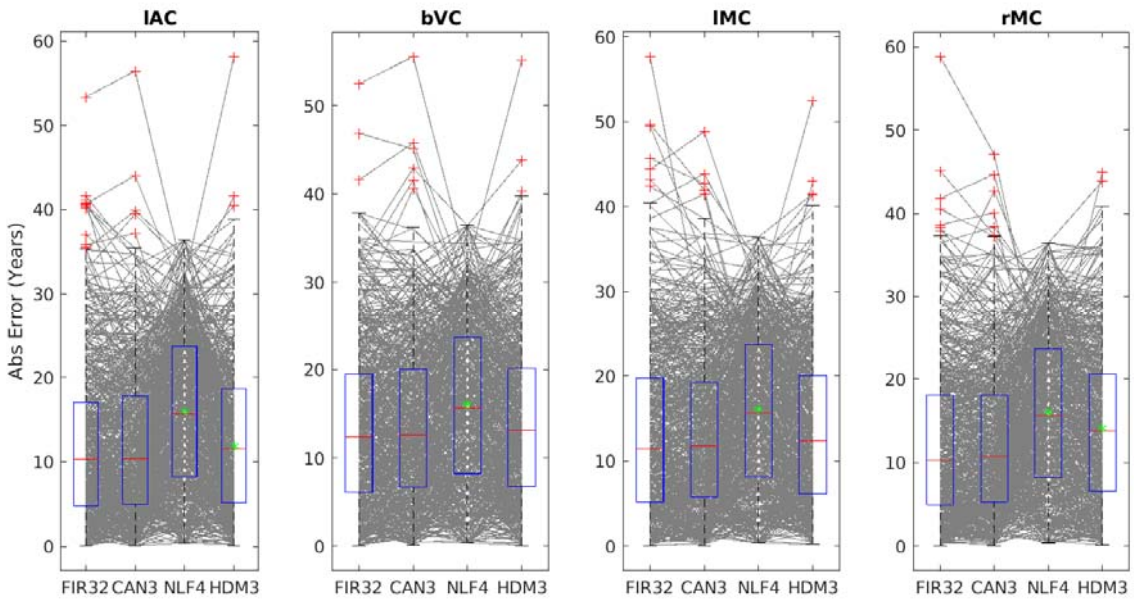

Figure 9. Cross-validated absolute error in predicting age for each model and each ROI. Each grey line corresponds to one participant. Superimposed on these lines are boxplots together with outliers (red crosses). A green asterisk means that a sign-test revealed significantly worse prediction than the FIR32 model. The average across ROIs of the median error was 11.1 years (FIR32), 11.4 years (Can3), 13.0 years (NLF4) and 12.7 years (HDM3), respectively.

For comparison, Supplementary Figure S7 shows the LOOCV error for the Can3 model, when fitting the canonical HRF only, or adding its temporal derivative, or adding both its temporal and its dispersion derivatives. This shows clearly that all three basis functions are needed in all ROIs.

Supplementary Figure S8 shows the LOOCV error for four versions of the HDM model (the HDM3 used here, plus versions with 4, 5 and 6 free parameters). Even though the HDM3 model did not differ significantly from more complex variants in lMC and rMC (indeed, models with more parameters did numerically better on rMC), the HDM3 model did perform best for lAC and bVC, which is why we focused on the HDM3 model here.

Discussion

We compared the ability of four models of the HRF to capture age-related changes in the BOLD impulse response, using a large (N=645) and representative dataset containing individuals uniformly distributed across the adult lifespan, and an fMRI sensorimotor paradigm optimised to estimate the shape of the HRF.

Whole-brain analysis using the first model – the "FIR32" model with 32 temporal basis functions capturing the HRF every 2s – revealed strong effects of age in many brain regions, particularly bilateral visual, auditory and motor cortices. The predominant effects of increased age were to reduce the amplitude of the BOLD response and to shift it later in time.

Focusing on four of these ROIs, we compared the FIR fit to a more parsimonious linear model using a canonical HRF and two of its partial derivatives ("Can3" model), as well as two nonlinear models. In terms of capturing features of the HRF such as peak amplitude and peak latency, the Can3 model showed stronger effects of age than the FIR32 model. However, a nonlinear model in which 4 parameters were fit to the FIR estimates ("NLF4" model, comprising zero- and first-order expansions of amplitude and post-stimulus time) showed even stronger effects of age in most ROIs, suggesting better de-coupling of amplitude and latency.

However, it is unclear whether such HRF features are informative with respect to the underlying neural activity, which is normally what fMRI is used to infer. Indeed, the mapping from neural activity to BOLD response is nonlinear, depending on vascular and haemodynamic parameters that are also likely to change with age. Only the fourth model, based on prior work on biophysical, generative haemodynamic modelling (Buxton et al., 1998; Friston et al., 2000), attempted to estimate such neural and vascular parameters separately. When allowing three of these parameters to vary (neural efficacy, neurovascular decay rate and haemodynamic transit rate), this "HDM3" model suggested that the majority of age-effects on the BOLD impulse response arise from vascular/haemodynamic differences. Indeed, only one ROI – right motor cortex (rMC) – showed an effect of age on HDM3's neural parameter, and this actually reflected increased neural activity in this ipsilateral motor region with age (see below for further discussion about this region).

Finally, we compared the models in terms of the accuracy with which the age of new participants could be predicted from the fitted parameters of each model, i.e., their out-of-sample cross-validation accuracy. Given that the four models are typically fitted using different statistical frameworks (classical and Bayesian), the cross-validation accuracy was useful as a common basis for comparing all the models' sensitivity to ageing. Comparison of the average (absolute) prediction error for of 1, 2 and 3 basis functions in the Can3 model (Supplementary Figure S7), and of the four versions of the HDM model with 3-6 parameters (Supplementary Figure S8), confirmed the value of this approach. When applied to the four main models considered here, performance was similar for FIR32, Can3 and HDM3 models and worst for the NLF4 model, except for rMC where HDM3 was worst.

Studying Ageing with fMRI

If one just wants to detect (e.g, localise in the brain) the effects of age on the BOLD impulse response, rather than interpret those effects, then the present results demonstrate that SPM's canonical basis set, Can3, is sufficient. It did not differ from the FIR32 model in its cross-validated

ability to predict age, but will offer greater statistical power by virtue of fewer degrees of freedom than the FIR.[8]

However, if one wishes to interpret such age effects in terms of neural activity, then such linear basis functions are not sufficient. For this, a biophysically-plausible generative model is needed. The HDM3 model used here is one such model, which showed cross-validated age-prediction performance comparable to the Can3 basis set for three of the four ROIs considered here, with the exception of rMC (see below for a possible explanation for latter). Importantly, if the HDM3 model correctly distinguishes neural and haemodynamic contributions to the BOLD response, then the present results have profound implications for previous fMRI studies that have interpreted age-related differences in the BOLD response in terms of differences in neural activity (at least in sensorimotor tasks like the present one). This is because the present HDM3 results suggest that most age effects on the BOLD response arise from vascular factors instead. This is consistent with recent work combining BOLD fMRI with glucose PET (Stiernman et al., 2023), which showed that effects of age on the BOLD response (during a working memory task) were not accompanied by effects of age on glucose metabolism. Indeed, we found that age need not affect neural efficacy in early visual, auditory and contralateral motor regions, at least when using a Bayesian approach to prune away parameters whose effects are correlated with those of other parameters. This is consistent with a previous study (Tsvetanov et al., 2015) that adjusted activations in the present task by empirical estimates of vascular reactivity from an independent resting-state fMRI dataset, called Resting-State Fluctuation Amplitudes (RSFA). That study also found dramatically attenuated effects of age on visual, auditory and contralateral motor BOLD responses, once they were scaled by RSFA. Interestingly, that study also found that the BOLD increases with age in rMC remained after RSFA scaling, again consistent with the present HDM3 results in suggesting a neural origin for age effects in ipsilateral motor responses.

The positive effect of age on the decay parameter of the HDM3 model in all ROIs suggests that vasodilatory signals decay faster in older people. This seems plausible, and may reflect changes in both the underlying neural dynamics and the efficacy of neurovascular coupling. Increasing this decay rate reduces the peak response and attenuates the post-peak undershoot (Supplementary Figure S1). Meanwhile, the negative effect of age on the haemodynamic transit rate in most ROIs (except rMC) suggests that blood flow is reduced (i.e., longer transit times) in older people. This also seems plausible; cerebral blood flow decreases 0.3-0.5% per year in healthy ageing (Graff et al., 2023). Decreasing this haemodynamic transit rate delays the HRF (Supplementary Figure S1).

The HDM3 finding that age increases the neurovascular decay rate is also consistent with a previous study that used DCM to examine resting-state connectivity in the CamCAN dataset (Tsvetanov et al., 2016), in which there was a significant positive effect in three of the four resting-state networks. That study also found the same negative effect of age on the haemodynamic transit rate, though it was only significant in one resting-state network. Nonetheless, that study used a different neural model (based on a 1/f neural power spectrum in the resting-state, rather than the brief burst of neural activity in response to a stimulus assumed here) and a different formalisation of the HDM (in which the ratio of intra-vascular to extra-vascular components of the signal, ϵ_h, was also estimated; see Supplementary Material). Moreover, it examined different (higher-order) brain regions. One or

[8] The more constrained nature of the Can3 basis set also prevents false detection of trial-locked effects that are implausible as haemodynamic responses, such as “activity” in the eyeballs for the first FIR time bin only, which reflects stimulus-driven eye-motion instead.

more of these differences might explain why it did not find significant effects of age on the haemodynamic parameters in more networks.

Note that we are not claiming that there are absolutely no age-related differences in neural activity in the present sensorimotor task, particularly given that we have previously shown age-related delays in evoked responses measured by MEG on the same participants in the same task (Price et al., 2017). Indeed, we used the same NLF4 model in that study to demonstrate that age increased the latency offset (shift) in visual evoked responses and the latency scaling (stretch) in auditory evoked responses. However, these latency differences were small (approximately 20ms offset and 20% scaling), and so unlikely to have much effect on the HRF, given that the HRF integrates over several seconds of neural activity, and likely includes induced electrophysiological activity across a range of frequencies. Thus we are claiming only that our HDM3 results demonstrate that neurovascular and vascular changes are *sufficient* to explain age effects on the BOLD response in most brain regions, without necessitating effects of age on neural activity per se. Future work could examine the effects of age on the power spectrum of neural responses (evoked and induced) in the present task, for example to see if BOLD amplitude relates to a shift in power towards higher frequencies (Kilner et al., 2005).

Linear Basis Sets and HRF Features

Given that the Can3 basis set showed cross-validated prediction of age as good as that of the FIR32 basis set, with far fewer parameters, it appears to be sufficient to capture the range of BOLD impulse responses across the healthy adult lifespan. Note that the partial derivatives of the canonical HRF are important to capture this variability (Supplementary Figure S7). Thus, studies of ageing that use only a single canonical HRF are likely to miss age-related differences in the HRF. Such studies need to ensure that parameter estimates for the derivatives are also taken forward to any group-level (2nd-level) analyses (or else combined to estimate a shape-independent effect size, Calhoun et al., 2004; Cignetti et al., 2016). The additional flexibility provided by the partial derivatives is also important to capture variability across brain regions (Supplementary Figure S3).

Supplementary Figure S3 also shows that the canonical HRF used in SPM is slightly delayed relative to typical impulse responses, at least in the primary sensorimotor regions examined here. The reason for this is likely that the canonical HRF was derived from a dataset in which the data were not slice-time corrected, so there was a typical delay of TR/2, relative to stimulus onset, in the activated brain regions (Friston et al., 1998). Here we were careful to synchronise the GLM with the reference slice, ensuring no such delay. Indeed, the present data could be used to define a more accurate canonical HRF, or even canonical HRFs for different age ranges.

Beyond localisation, it is unclear whether there are effects of age on the shape of the HRF that can be directly interpreted (at least without a biophysical model). There may be information about the latency of neural activity in the latency of BOLD impulse responses, but this seems most likely for latency differences between two or more conditions (e.g. types of stimuli), found within the same brain region and same individual tested within the same session. In such situations, one can reasonably assume that haemodynamic variables are constant, so any differences are neural in origin. In this case, temporally-extended basis sets like the Can3 can better estimate latencies than individual FIR time bins, which are more prone to noise, while the present results show that non-linear fitting of models that explicitly parametrise different types of latency (e.g, offset and scaling) leads to even higher sensitivity. Even so, reverse inferences about the latency of neural activity from latency of the BOLD response are complicated due to convolution (e.g, a difference in peak latency of the HRF can arise from differences in the duration rather than onset of neural activity, Henson et

al., 2002), not to mention potential nonlinearities in the neural-BOLD mapping. Most importantly however, for the study of ageing, or other individual/group differences, it seems unlikely that differences between individuals in the latency (or indeed amplitude) of their BOLD responses could ever be uniquely attributed to differences in neural activity, at least in the absence of a validated biophysical model or independent data about haemodynamic differences.

Conversely, it is important to remember that any individual differences in the HRF (e.g. as a function of age) cannot be uniquely attributed to haemodynamic factors, unless one knows that neural activity is invariant across individuals (or can estimate that activity separately, e.g. with a biophysical model). This point was made by Grinband et al. (2017) in the context of ageing. We believe that our paradigm minimised factors that could cause age-related differences in neural activity (compared to some previous paradigms) because our trial-based design entailed very brief, simple, meaningless stimuli (<300ms) that required minimal processing. While older people are generally slower to react in such tasks, we accommodated this delay (for analysis of motor cortices) by locking analysis to the keypress rather than the stimulus, which should allow for age-related variability in execution time at least (though differences could remain in motor force and other kinematic variables). However, our task did involve long and variable SOAs, in order to maximise the ability of fMRI to estimate HRF shape (Josephs and Henson, 1999), so it is possible that age-related differences in sustained attention modulated neural activity (Grinband et al., 2017). But this is exactly why more sophisticated models like the HDM are needed (in the absence of independent measures of neural/haemodynamic factors), to try to separate effects of age on neurodynamics and haemodynamics.

Caveats with HDM

Though the present HDM3 model offers the potential to separate neural and vascular contributions to the BOLD impulse response, this is subject to several assumptions, as detailed in the Supplementary Material. We fixed several parameters of the full HDM to their prior expected values, whereas in reality they may also have changed with age. The reason for fixing some of them is that the effects of several parameters on the BOLD response are very similar, so they are difficult to disentangle in the absence of independent data (e.g, about blood flow or blood volume). This is reflected in high posterior covariances between some parameters when fitting a 6-parameter HDM (Supplementary Figure S2), and the fact that freeing up more than 3 HDM parameters did not significantly improve cross-validated age prediction (Supplementary Figure S8). Thus, interpretation of some of the age effects that we found with the HDM3 model could be wrong, mandating follow-up studies that complement the BOLD response with other measurement modalities.

For example, regarding the neural efficacy parameter, its effects on the scaling of HRF are very similar to those of the venous blood volume fraction (V_0), and V_0 is known to change with age. Nonetheless, since V_0 must be positive, it cannot be solely responsible for the negative BOLD peak we found in rMC. The effects of the neural efficacy parameter can also be similar to those of the resting oxygen extraction fraction E_0 (Supplementary Figure S1), resulting in a high negative correlation in their parameter estimates (Supplementary Figure S2). However, we argue in the Supplementary Material that effects of age on oxygen extraction fraction are unlikely to be detectable in the BOLD signal measured here. The effects of the neural efficacy parameter are also related to vessel stiffness parameter α (Grubb et al., 1974), though in this case, they tend to have opposite effects on the HRF amplitude (inducing positive correlation in their parameter estimates). Nonetheless, our application of PEB to the HDM3 model resulted in no effects of age on neural efficacy β in any ROI except rMC, and the latter is unlikely to be explained by parameters such as blood volume fraction, oxygen extraction or vessel stiffness, because the effects of age on these are

unlikely to differ dramatically across cortical regions, particularly across contralateral ROIs like lMC and rMC.

Likewise, the effects of age that we did find in all ROIs on neurovascular decay rate and haemodyamic transit rate could also be attributed to effects of age on other HDM parameters that we fixed, particularly the neurovascular feedback parameter, whose effects on the latency of the HRF are similar to those of the haemodynamic transit rate (Supplementary Figure S1). To properly decouple these HDM parameters, one would need independent measurements of blood flow, oxygenation, vasodilatory signal and neural activity, to simultaneously constrain their values, for example using Arterial Spin Labelling (ASL) fMRI for blood flow, or Vascular Space Occupancy (VASO) fMRI for blood volume. Future work could also add cardiovascular measurements like blood pressure and/or heart-rate variability to the PEB model, to see if they relate to HDM vascular parameters, or more direct measures of neural activity, such as from Electro / Magneto-encephalography (E/MEG). Higher order (polynomial) expansions of age could also be added to the PEB model, in case there are nonlinear effects of age on the HDM parameters.

Thus in summary, we are not claiming that the HDM3 model used here is the best or most accurate model, and conclusions from this model should be treated cautiously. Rather our aim is to illustrate how biophysical models like the HDM could be used in principle to better interpret effects of age on the BOLD impulse response, e.g. in separating neural and vascular components. Thus while the present work supports the face validity and predictive validity of HDM, further support for its construct validity will require independent measurement of several of its variables, beyond the final BOLD response. For example, EEG/MEG can provide independent measurements of neural activity from the same people doing the same task, while other types of fMRI contrast, which depend on blood flow or blood volume, could be acquired concurrently.

Ipsilateral motor cortex

Finally, the BOLD response in rMC is noteworthy in its difference from the other ROIs: it showed a negative and delayed BOLD response versus baseline in young people, which became more positive with age (see also Mayhew et al., 2022; Ward et al., 2008; Tak et al., 2021; Knights et al., 2021; or Mattay & Weinberger, 1999, for review). Indeed, the HRFs in rMC were the hardest to fit, particularly for the HDM3 model. There may be several reasons for this. For example, it has been suggested that the motor cortex ipsilateral to the effector is inhibited by the contralateral motor cortex, but that this inhibition decreases with age, explaining the increase in neural efficacy with age disclosed by the HDM3 model. Such activity in inhibitory interneurons may have metabolic consequences (haemodynamics) that differ from those assumed by the HDM for excitatory activity (Vazquez et al., 2018). Indeed, the neural component of the HDM could be augmented by distinguishing excitatory and inhibitory cell populations, as was proposed for DCM (Marreiros et al., 2008; Havlicek et al., 2015). Furthermore, the neural activity in ipsilateral motor cortex in such tasks could be more delayed/dispersed than the brief burst assumed by all models used here. Ipsilateral motor cortex would therefore be an interesting region to explore more closely in terms of its neurodynamics and haemodynamics.

Summary

Age has strong effects on the form of the HRF, as shown using all four models we tested, and this should be taken into account when analysing fMRI data. The model one would select for a study depends on its aims. Where the priority is regional localisation, our results demonstrate that the three parameter model (Can3) works well. It is parsimonious, has similar sensitivity to ageing effects as the other models we compared, and has predictive validity. However, when interpreting differences between individuals that have been localised this way, such as due to their age, caution should be exercised before interpreting those differences in terms of neural differences (rather than vascular ones). By contrast, where the aim is to ask mechanistic questions about the underlying causes of ageing effects, there is an opportunity to deploy physiological models such as the HDM. As illustrated here, these kinds of models can generate mechanistic hypotheses, which could then be tested with other modalities that can provide direct measures of neural activity, blood perfusion, or blood volume. Thus, with the appropriate model and data, we can go beyond descriptive statistics such as time-to-peak or peak amplitude, and ask mechanistic questions about the genesis of the BOLD response and its dysfunction in ageing.

Acknowledgements

We thank the Cam-CAN respondents and their primary care teams in Cambridge for their participation in this study, and colleagues at the MRC Cognition and Brain Sciences Unit MEG and MRI facilities for their assistance. Further information about the Cam-CAN corporate authorship membership can be found at http://www.cam-can.com/publications/Cam-CAN_Corporate_Author.html (list #14). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. We are grateful for discussions with Pranay Yadav.

Author contributions

Designed research: R.N.H. Performed research: R.N.H., W.O. & P.Z. Analysed data: R.N.H., W.O. & P.Z. Writing—original draft: R.N.H. & P.Z. Writing – review & editing: R.N.H., W.O., K.A.T. & P.Z.

Funding

Cam-CAN was supported by the Biotechnology and Biological Sciences Research Council Grant BB/H008217/1. R.N.H. was supported by the UK Medical Research Council [SUAG/046/G101400]. P.Z. was supported by core funding from Wellcome [203147/Z/16/Z]. W.O. was in receipt of scholarships from the Cambridge Trust and from the Mateusz B. Grabowski Fund. K.A.T. was supported by the Guarantors of Brain [G101149].

***Cam-CAN corporate author**

Lorraine K Tyler, Carol Brayne, Edward T Bullmore, Andrew C Calder, Rhodri Cusack, Tim Dalgleish, John Duncan, Richard N Henson, Fiona E Matthews, William D Marslen-Wilson, James B Rowe, Meredith A Shafto; Karen Campbell, Teresa Cheung, Simon Davis, Linda Geerligs, Rogier Kievit, Anna McCarrey, Abdur Mustafa, Darren Price, David Samu, Jason R Taylor, Matthias Treder, Kamen A Tsvetanov, Janna van Belle, Nitin Williams, Daniel Mitchell, Simon Fisher, Else Eising, Ethan Knights; Lauren Bates, Tina Emery, Sharon Erzinçlioglu, Andrew Gadie, Sofia Gerbase, Stanimira Georgieva, Claire Hanley, Beth Parkin, David Troy; Tibor Auer, Marta Correia, Lu Gao, Emma Green, Rafael Henriques; Jodie Allen, Gillian Amery, Liana Amunts, Anne Barcroft, Amanda Castle, Cheryl Dias, Jonathan Dowrick, Melissa Fair, Hayley Fisher, Anna Goulding, Adarsh Grewal, Geoff Hale, Andrew Hilton, Frances Johnson, Patricia Johnston, Thea Kavanagh-Williamson, Magdalena Kwasniewska, Alison McMinn, Kim Norman, Jessica Penrose, Fiona Roby, Diane Rowland, John Sargeant, Maggie Squire, Beth Stevens, Aldabra Stoddart, Cheryl Stone, Tracy Thompson, Ozlem Yazlik, Dan Barnes, Marie Dixon, Jaya Hillman, Joanne Mitchell, Laura Villis.

723 **References**

724 Abdelkarim D, Zhao Y, Turner MP, et al. (2019) A neural-vascular complex of age-related changes in
725 the human brain: Anatomy, physiology, and implications for neurocognitive aging.
726 *Neuroscience and Biobehavioral Reviews* 107(December 2018). Elsevier: 927–944. DOI:
727 10.1016/j.neubiorev.2019.09.005.

728 Aizenstein HJ, Clark KA, Butters MA, et al. (2004) The BOLD hemodynamic response in healthy aging.
729 *Journal of Cognitive Neuroscience* 16(5): 786–793. DOI: 10.1162/089892904970681.

730 Buračas GT and Boynton GM (2002) Efficient design of event-related fMRI experiments using m-
731 sequences. *NeuroImage* 16(3 I). Academic Press Inc.: 801–813. DOI: 10.1006/nimg.2002.1116.

732 Buxton RB, Wong EC and Frank LR (1998) Dynamics of blood flow and oxygenation changes during
733 brain activation: The balloon model. *Magnetic Resonance in Medicine* 39(6): 855–864. DOI:
734 10.1002/mrm.1910390602.

735 Calhoun VD, Stevens MC, Pearlson GD, et al. (2004) fMRI analysis with the general linear model:
736 Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms.
737 *NeuroImage* 22(1): 252–257. DOI: 10.1016/j.neuroimage.2003.12.029.

738 Cignetti F, Salvia E, Anton JL, et al. (2016) Pros and Cons of Using the Informed Basis Set to Account
739 for Hemodynamic Response Variability with Developmental Data. *Frontiers in Neuroscience*
740 10(JUL). Frontiers Media SA: 322. DOI: 10.3389/FNINS.2016.00322.

741 Cusack R, Vicente-Grabovetsky A, Mitchell DJ, et al. (2015) Automatic analysis (aa): Efficient
742 neuroimaging workflows and parallel processing using Matlab and XML. *Frontiers in*
743 *Neuroinformatics* 8(JAN): 1–13. DOI: 10.3389/fninf.2014.00090.

744 D'Esposito M, Zarahn E, Aguirre GK, et al. (1999) The effect of normal aging on the coupling of neural
745 activity to the bold hemodynamic response. *NeuroImage* 10(1): 6–14. DOI:
746 10.1006/nimg.1999.0444.

747 Dale AM and Buckner RL (1997) Selective Averaging of Rapidly Presented Individual Trials Using
748 fMRI. *Hum. Brain Mapping* 5. Wiley-Liss, Inc: 329–340. DOI: 10.1002/(SICI)1097-0193(1997)5:5.

749 Friston KJ, Fletcher P, Josephs O, et al. (1998) Event-Related fMRI : Characterizing Differential
750 Responses. 40(7): 30–40.

751 Friston KJ, Mechelli A, Turner R, et al. (2000) Nonlinear Responses in fMRI: The Balloon Model,
752 Volterra Kernels, and Other Hemodynamics. *NeuroImage* 12(4). Academic Press: 466–477. DOI:
753 10.1006/NIMG.2000.0630.

754 Friston KJ, Harrison L and Penny W (2003) Dynamic causal modelling. *NeuroImage* 19(4): 1273–1302.
755 DOI: 10.1016/S1053-8119(03)00202-7.

756 Graff BJ, Harrison SL, Payne SJ, et al. (2023) Regional Cerebral Blood Flow Changes in Healthy Ageing
757 and Alzheimer's Disease: A Narrative Review. *Cerebrovascular Diseases* 52(1): 11–20. DOI:
758 10.1159/000524797.

759 Grinband J, Steffener J, Razlighi QR, et al. (2017) BOLD neurovascular coupling does not change
760 significantly with normal aging. *Human Brain Mapping* 38(7): 3538–3551. DOI:
761 10.1002/hbm.23608.

762 Grubb RL, Raichle ME, Eichling JO, et al. (1974) The Effects of Changes in PaCO2 Cerebral Blood
763 Volume, Blood Flow, and Vascular Mean Transit Time. *Stroke* 5(5). Lippincott Williams &
764 Wilkins: 630–639. DOI: 10.1161/01.STR.5.5.630.

Handwerker DA, Gazzaley A, Inglis BA, et al. (2007) Reducing vascular variability of fMRI data across aging populations using a breathholding task. *Human Brain Mapping* 28(9): 846–859. DOI: 10.1002/hbm.20307.

Handwerker DA, Gonzalez-Castillo J, D'Esposito M, et al. (2012) The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage* 62(2). Elsevier B.V.: 1017–1023. DOI: 10.1016/j.neuroimage.2012.02.015.

Henson RN (2004) Analysis of fMRI Timeseries: Linear Time-Invariant Models, Event-related fMRI and Optimal Experimental Design. In: *Human Brain Function (Second Edition)*, pp. 793–822. DOI: 10.1006/nimg.1995.1023.

Henson RN, Price CJ, Rugg MD, 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. DOI: 10.1006/nimg.2001.0940.

Josephs O and Henson RN (1999) Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philosophical Transactions of the Royal Society, B, 354, 1215-1228.

Kilner JM, Mattout J, Henson R, et al. (2005) Hemodynamic correlates of EEG: A heuristic. *NeuroImage* 28(1): 280–286. DOI: 10.1016/j.neuroimage.2005.06.008.

Knights E, Morcom AM and Henson RN (2021) Does hemispheric asymmetry reduction in older adults in motor cortex reflect compensation? *Journal of Neuroscience* 41(45). Society for Neuroscience: 9361–9378. DOI: 10.1523/JNEUROSCI.1111-21.2021.

Kruggel F and Yves Von Cramon D (1999) Modeling the Hemodynamic Response in Single-Trial Functional MRI Experiments. *Magnetic Resonance in Medicine* 42: 787–797. DOI: 10.1002/(SICI)1522-2594(199910)42:4.

Lindquist MA, Meng Loh J, Atlas LY, et al. (2009) Modeling the Hemodynamic Response Function in fMRI: Efficiency, Bias and Mis-modeling. *Neuroimage* 45(1 Suppl). NIH Public Access: S187. DOI: 10.1016/J.NEUROIMAGE.2008.10.065.

Marreiros AC, Kiebel SJ and Friston KJ (2008) Dynamic causal modelling for fMRI: A two-state model. *NeuroImage* 39(1): 269–278. DOI: 10.1016/j.neuroimage.2007.08.019.

Mattay VS and Weinberger DR (1999) Organization of the human motor system as studied by functional magnetic resonance imaging. *European journal of radiology* 30(2). Eur J Radiol: 105–114. DOI: 10.1016/S0720-048X(99)00049-2.

Mayhew SD, Coleman SC, Mullinger KJ, et al. (2022) Across the adult lifespan the ipsilateral sensorimotor cortex negative BOLD response exhibits decreases in magnitude and spatial extent suggesting declining inhibitory control. *NeuroImage* 253. Academic Press Inc. DOI: 10.1016/j.neuroimage.2022.119081.

Patel AX, Kundu P, Rubinov M, et al. (2014) A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage*. DOI: 10.1016/j.neuroimage.2014.03.012.

Price D, Tyler LK, Neto Henriques R, et al. (2017) Age-related delay in visual and auditory evoked responses is mediated by white-and grey-matter differences. *Nature Communications* 8(May 2016). DOI: 10.1038/ncomms15671.

Richter W and Richter M (2003) The shape of the fMRI BOLD response in children and adults changes systematically with age. *NeuroImage* 20(2): 1122–1131. DOI: 10.1016/S1053-8119(03)00347-1.

Shafto MA, Tyler LK, Dixon M, et al. (2014) The Cambridge Centre for Ageing and Neuroscience

(Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology* 14(1). BMC Neurol. DOI: 10.1186/S12883-014-0204-1.

Stephan KE, Weiskopf N, Drysdale PM, et al. (2007) Comparing hemodynamic models with DCM. *NeuroImage* 38(3). Neuroimage: 387–401. DOI: 10.1016/J.NEUROIMAGE.2007.07.040.

Stiernman L, Grill F, McNulty C, et al. (2023) Widespread fMRI BOLD signal overactivations during cognitive control in older adults are not matched by corresponding increases in fPET glucose metabolism. *The Journal of Neuroscience* 43(14): JN--RM--1331--22. DOI: 10.1523/jneurosci.1331-22.2023.

Tak YW, Knights E, Henson R, et al. (2021) Ageing and the ipsilateral m1 bold response: A connectivity study. *Brain Sciences* 11(9). MDPI. DOI: 10.3390/BRAINSCI11091130/S1.

Taoka T, Iwasaki S, Uchida H, et al. (1998) Age Correlation of the Time Lag in Signal Change on EPI-fMRI. *Journal of Computer Assisted Tomography* 22(4). Available at: https://journals.lww.com/jcat/Fulltext/1998/07000/Age_Correlation_of_the_Time_Lag_in_Signal_Change.2.aspx.

Taylor JR, Williams N, Cusack R, et al. (2017) The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage* 144(Pt B). Neuroimage: 262–269. DOI: 10.1016/J.NEUROIMAGE.2015.09.018.

Tsvetanov KA., Henson RN, Tyler LK, et al. (2016) Extrinsic and Intrinsic Brain Network Connectivity Maintains Cognition across the Lifespan Despite Accelerated Decay of Regional Brain Activation. *Journal of Neuroscience* 36(11): 3115–3126. DOI: 10.1523/JNEUROSCI.2733-15.2016.

Tsvetanov KA, Henson RN and Rowe JB (2021) Separating vascular and neuronal effects of age on fMRI BOLD signals: Neurovascular ageing. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1815). DOI: 10.1098/rstb.2019.0631.

Vazquez AL, Fukuda M and Kim SG (2018) Inhibitory Neuron Activity Contributions to Hemodynamic Responses and Metabolic Load Examined Using an Inhibitory Optogenetic Mouse Model. *Cerebral Cortex* 28(11). Oxford Academic: 4105–4119. DOI: 10.1093/CERCOR/BHY225.

Ward NS, Swayne OBC and Newton JM (2008) Age-dependent changes in the neural correlates of force modulation: An fMRI study. *Neurobiology of Aging* 29(9). Elsevier Inc.: 1434–1446. DOI: 10.1016/j.neurobiolaging.2007.04.017.

West KL, Zuppichini MD, Turner MP, et al. (2019) BOLD hemodynamic response function changes significantly with healthy aging. *NeuroImage* 188. Neuroimage: 198–207. DOI: 10.1016/J.NEUROIMAGE.2018.12.012.

Wright ME and Wise RG (2018) Can Blood Oxygenation Level Dependent Functional Magnetic Resonance Imaging Be Used Accurately to Compare Older and Younger Populations? A Mini Literature Review. *Frontiers in aging neuroscience* 10. Front Aging Neurosci. DOI: 10.3389/FNAGI.2018.00371.

Zeidman P, Jafarian A, Seghier ML, et al. (2019) A guide to group effective connectivity analysis, part 2: Second level analysis with PEB. *NeuroImage* 200. Academic Press: 12–25. DOI: 10.1016/J.NEUROIMAGE.2019.06.032.