Early Stopping in Experimentation with Real-time Functional Magnetic Resonance Imaging Using the Sequential Probability Ratio Test

Sarah J. A. Carr[1,2], Weicong Chen[3], Jeremy Fondran[4], Harry Friel[5], Curtis Tatsuoka[4,2*]

1. Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
2. Department of Neurology, Case Western Reserve University, Cleveland, OH, USA
3. Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, USA
4. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA
5. Philips Healthcare, Highland Heights, OH, USA

*Corresponding Author:
Curtis Tatsuoka
10900 Euclid Avenue
Case Western Reserve University
Cleveland, Ohio, 44106
USA
Email: cmt66@case.edu

KEY WORDS: real-time fMRI, adaptive fMRI, dynamic experimentation, SPRT, early stopping

Abstract

Introduction: Functional magnetic resonance imaging (fMRI) often involves long scanning durations to ensure the associated brain activity can be detected. However, excessive experimentation can lead to many undesirable effects, such as from learning and/or fatigue effects, as well as undue discomfort for the subject, which can lead to motion artifact and loss of sustained attention on task. Overly long experimentation ironically can thus have a detrimental effect on signal quality and accurate voxel activation detection. Here, we propose a method of dynamic experimentation with real-time fMRI using a novel statistically-driven approach to fMRI analytics. This new approach to experimental design invokes early stopping when sufficient statistical evidence for assessing the task-related activation is observed.

Methods: Voxel-level sequential probability ratio test (SPRT) statistics based on general linear models (GLM) were implemented on fMRI scans of a mathematical 1-back task from 25 subjects, 14 healthy controls and 11 subjects born extremely preterm. This approach is based on likelihood ratios and allows for systematic early stopping based on statistical error thresholds being satisfied. We explored voxel-level serial covariance estimation in real-time using the "sandwich" estimator. We adopted a two-stage estimation approach that allows for the hypothesis tests to be formulated in terms of t-statistic scale, which enhances interpretability. Scan data was collected using a dynamic feedback system that allowed for adaptive experimentation. Numerical parallelization was employed to facilitate completion of computations involving a new scan within every repetition time (TR).

Results: SPRT analytics demonstrate the feasibility and efficiency gains of automated early stopping, while performing comparably in activation detection with full protocols analyzed through standard fMRI software. Dynamic stopping of stimulus administration was achieved in all subjects with typical time savings between 33 - 66% (4 – 8 minutes on a 12 minute scan).

Conclusion: A systematic statistical approach for early stopping with real-time fMRI experimentation has been implemented. This allows for great savings in scan times, while still eliciting comparable activation patterns as full protocols. This dynamic approach has promise for reducing subject burden and fatigue effects.

1.0 Introduction

Analysis of task-based functional MRI (fMRI) scans is typically performed with fixed, predetermined experimental designs. As a result, subjects must often endure stimulus protocols that are overly long in order to ensure the neural activity can be statistically discerned in the noisy data. However, this can lead to fatigue, learning effects and excessive motion, such as from agitation, as well as being costlier to administer due to longer scan times and potentially less reliable measurement. Also, the experimenter does not know if the neural activity is detectable until long after the scanning session is over. Real-time functional MRI (RT-fMRI) provides an opportunity to ameliorate these issues. RT-fMRI has been successfully applied in the field of neurofeedback and biofeedback from neural responses, where subjects may be trained to alter their brain activity based on real-time information provided from the fMRI scans. This has been reported in ADHD (1), healthy subjects with no psychiatric or neurological disorders (2, 3), Alzheimer's disease (4) and Parkinson's disease (5, 6). Its uses have also been described in psychoradiology to aid diagnosis and treatment planning in psychiatric disorders (7). A largely unexplored application of RT-fMRI is to dynamically and statistically determine when a stimulus has been sufficiently presented in terms of replication of blocks to terminate early. The magnitude of effort and variability in neural activity while completing a task will vary from person to person. Trial administration within a block design can be stopped early if sequentially updated statistical inference on activation can be determined with sufficient accuracy based on the observed BOLD (blood oxygen level dependent) signal response up to that point. This application will be explored in detail.

The benefits of adaptive RT-fMRI include: 1) <u>Shorter scan times for fMRI testing</u>: Shorter scan times cannot only save in technology and personnel costs, but fatigue and learning effects can be avoided, improving signal quality. Scanning becomes less burdensome on the subject as well, which is an especially important consideration for children or elderly subjects. 2) <u>Real-time quality control</u>: greater consistency in activation classification error can be obtained, through statistical error-based benchmarks for stopping rules and real-time feedback on classification performance and adjustment of stimulus durations. 3) <u>Richer information</u>: Paradigms can become more complex and sophisticated. With greater time efficiency and flexibility, more variations of a stimulus, such as reflected by a broader range of difficulty levels, can be administered in the same amount of time. 4) <u>Wide applicability</u>: Dynamic adjustment of stimuli based on BOLD response in real time can be generally applied across a range of focus areas that investigate localization of brain activity, including cognition and motor functioning.

Since the advent of RT-fMRI in the mid 1990's (8), a handful of mainstream software packages have been developed for use by the fMRI community. These include Turbo BrainVoyager (9), AFNI's real-time plugin (8) and FSL-based FRIEND (10). There have been few previous studies that have used adaptive RT-fMRI. In one example, it has been used to determine 'good' and 'bad' brain states to optimize learning (11). The presentation of novel scenes was prompted by the detection of 'good' brain states; the 'good' template was determined based on a prior standard acquisition test scan. They used real-time general linear model (GLM) methods described in (12) to estimate the BOLD signal magnitude at each time point (each scan) and compared it to a value within a region of interest from the earlier test scan. Another adaptive RT-fMRI study has used a person's brain state to judge their attention to a task (13). When their attention appeared to be wandering, the difficulty of the task was increased bringing their attention back. The authors used multivariate pattern analysis to determine task-relevant and task-irrelevant activity. Lorenz et al (2016) used FSL to pre-process the scans in real-time before applying a GLM-based analysis. Their study involved eliciting activity in particular brain regions by presenting stimuli chosen based on the response to the previous stimulus. The aim was not to investigate brain activity related to a particular task but simply activate a brain region (14). Another example of adaptive RT-fMRI used a Bayesian optimization algorithm to estimate when brain activity was mapped to a particular network (15). The Bayesian optimization was trained on 4 difficulty levels of a task prior to switching to choosing the optimal difficulty levels to elicit the desired activity, where there were 12 other levels to choose from.

Here, we extend the use of a statistically-based dynamic approach to RT-fMRI experimentation described in (16). This approach involves the sequential updating of voxel-level likelihood ratio tests, known as sequential probability ratio tests (SPRTs) and assessing after each scan whether there is sufficient statistical evidence to determine whether or not an associated parameter value indicates task activation. Such results, considered in aggregate across a collection of voxels, can be used as a basis for early stopping of experimentation. More generally, this approach can apply to competing hypotheses related to linear contrasts of task parameters. Our extension involves the actual implementation of this system on a Philips MRI scanner, in a real-time workflow, and sequential analysis of subjects across a study. This allows us to assess the variability in experimentation lengths across individuals to reach a rigorous threshold for decisive classification of activation. This work supports the premise that adaptive, individualized experimentation is feasible and can lead to savings in experimental durations.

Most off-line, post-hoc analyses of fMRI data use the general linear model to test statistical associations of voxel activation magnitude to task administration (17-19). This approach involves the voxel-level estimation of task-related regression parameters that indicate magnitude of association between an expected hemodynamic response pattern from a task and the observed BOLD signal. In our previous work, we adapted this general method for real-time fMRI by sequentially updating GLM regression parameter estimates as soon as the brain volumes were collected. At the individual voxel level, we can then assess hypothesis tests related to activation that are based on these estimates. In aggregate, the voxel level analyses inform decisions on early stopping and the tailoring of fMRI experimentation. See (16) for more details.

In comparison to (16), we previously assumed independence when estimating the covariance model. Here, we extend these methods by considering the "sandwich" estimator to recognize serial covariance (20, 21). We also adopt a two-stage estimation approach that allows for the hypothesis test parameters to be formulated in terms of z-score scale at the voxel level. Moreover, we now present a novel workflow to apply and implement these methods on a Philips scanner, with a dynamic feedback system that allows for real-time dynamic adjustment of the experimentation with subjects. This was facilitated with adoption of numerical parallelization techniques.

Another novel aspect of this work is the application of adaptive RT-fMRI in a sample group of 14 healthy adolescent subjects and 11 adolescents born extremely preterm (EPT). Analysis was implemented using sequential probability ratio test (SPRT) statistics and our server was a custom-built Linux computer located in a nearby building. The fMRI stimulus was a mathematical version of the well-known 1-back task. We show a novel workflow for the real-time processing of fMRI scans; a workflow that allows adaption of the stimulus difficulty level; processing of RT-fMRI was completed within 3 seconds before the next scan arrived; average time savings of 33 - 60% based on 80% of voxels being classified were achieved. This equated to 4 – 8 minutes savings with a 12 minute scan.

2.0 Methodology

2.1 fMRI Analysis Methods

2.1.1 General linear model

Briefly, the general linear model involves convoluting a double gamma hemodynamic response function (HRF) with task indicator variables that denote timing of administration to reflect expected task-related BOLD responses. Voxel-level task-related regression parameters are estimated and represent the association of the observed response to expected task-activated BOLD signal. Thus, activation is assessed through statistical inference on regression parameters. For a given voxel up to time t (i.e. for scans 1 through t), the GLM takes the form:

$$Y_t = XB + E_t \tag{1.1}$$

Where Y_t is a $t \times 1$ vector of observed BOLD signal intensities for the voxel up to time t, and E_t is a $t \times 1$ vector that represents the error components. X is a $t \times$ p design matrix and includes the expected BOLD signal values per task. We also include cosine functions of increasing periodicity (scan duration*2, scan duration, scan duration/1.5, scan duration/2 and scan duration/2.5) to model physiological and other low frequency noise (22). For large periodicities, cosine functions are approximately linear for the time frame of scans we consider here, and hence are essentially collinear from a GLM modeling perspective. Five regressors were thus added to the design matrix. $B = \left[b_1 \dots b_j \dots b_p\right]^t$, a p $\times$ 1 regression coefficients vector. In this formulation, a regression parameter b_j can represent magnitude of association with task j. E_t is assumed to be distributed as multivariate normal with mean zero and covariance W_t, where W_t is a $t \times t$ matrix that represents the temporal autocorrelation structure. For spatial correlation, we conduct spatial smoothing, so do not explicitly model the spatial correlation structure. Y_t is assumed to have a multivariate normal probability distribution as follows:

$$f(Y_t, B, \sigma^2 W_t) = \frac{1}{(2\pi)^{t/2}|\sigma^2 W_t|} \exp\left(-\frac{1}{2}(Y_y - XB)'(\sigma^2 W_t)^{-1}(Y_t - XB)\right) \tag{1.2}$$

where $|\sigma^2 W_t|$ is the determinant of $\sigma^2 W_t$. Major sources of noise in fMRI data include brain metabolism, physiology, and spontaneous fluctuations (23).

We fit regression models in parallel for all voxels under consideration in a target region of interest (ROI), which could include the whole brain. Real-time analysis requires signal and image processing steps, as well as the continual updating of statistical estimates as new scan data are received from the scanner. Hence, given the large number of voxels to be analyzed, real-time fMRI presents "big data" computational challenges.

2.1.2 Sandwich Estimator

In our previous work (16), we assumed serial independence for computational simplicity. Here we recognize serial covariance using the sandwich estimator. The sandwich estimator is a robust, model-free covariance estimator that does not require distributional assumptions. Importantly, it still provides asymptotically consistent covariance matrix estimates, although convergence rates can be slow (20, 21). The approach is evaluated for computational feasibility for real time analysis, as well as plausibility in effectively modeling the error structure of the BOLD data over time within a voxel.

2.1.3 Wald's Sequential Probability Ratio Test

At the voxel level, we can use the sequential analytic framework of (16, 24-28), to adaptively assess activation status using real-time fMRI. As we will demonstrate, Wald's SPRT is an efficient, sequential testing approach that can greatly reduce the need for experimental block administrations compared with fixed designs while attaining similar classification performance in simulation, and activation patterns with subject data. This approach relies on a SPRT statistic to conduct hypothesis testing, with the null hypothesis representing no activation with respect to a task, and the alternative hypothesis representing some threshold of activation, as represented by a GLM parameter value (16). This statistic is updated with each new observation, and its value is compared with thresholds for stopping. Exceeding thresholds indicate a decision on activation status can be made within predetermined statistical error levels. Additionally, we conduct two-stage estimation to obtain a preliminary estimate of the voxel-level error variance, so that an associated magnitude of the linear contrast of task parameters $c'\beta$ can be determined as the threshold in the alternative hypothesis that will correspond to a desired t-statistic value. As an illustration, suppose a t-statistic value of 3.1 is selected, as will be done below in our studies. Note 3.1 is the one-sided p-value = 0.001 - critical value for the standard normal distribution. Given an estimated value $\hat{\sigma}^2$ from a first stage, we solve for the value of $c'\beta$ that satisfies $c'\beta$ /sqrt($\widehat{var}[c'\hat{\beta}]$) = 3.1. See, for instance, (16) for the variance formula.

Specifically, the general procedure of Wald's SPRT is described as follows. Consider a one-sided hypothesis of the form $H_0: c'\beta = c'\beta_0$ versus $H_a: c'\beta \geq c'\beta_1$, where $c'(\beta_1 - \beta_0) \geq 0$. Two-sided formulations are described in (24) and (16). Implementation of Wald's SPRT involves updating Wald's likelihood ratio statistic as new data are observed (24):

$$\Lambda_t = log\left(\frac{f\left(Y_t|c'\beta_1, \widehat{var}\left[c\hat{\beta}\right]\right)}{f\left(Y_t|c'\beta_0, \widehat{var}\left[c\hat{\beta}\right]\right)}\right) \tag{1.3}$$

where $f(Y|c'\beta_0, \widehat{var}\left[c\hat{\beta}\right])$ and $f(Y|c'\beta_1, \widehat{var}\left[c\hat{\beta}\right])$ are the respective probability densities functions of Y_t given $c'\beta_0$ or $c'\beta_1$ is the true value of parameter of interest and conditioning on the estimated covariance. After Y_t is observed at a time point, t, one of three possible decisions is made according to the following rules:

1. Continue sampling if $B < \Lambda_t < A$
2. Stop sampling and accept H_0 if $\Lambda_t < B$
3. Stop sampling and accept H_a if $A < \Lambda_t$

where stopping boundaries $(A, B) = (\log((1-\beta_E)/\alpha_E), \log(\beta_E/(1-\alpha_E)))$, and the target Type I and Type II error levels are respectively denoted as α_E and β_E. These error levels are specified before testing. Note that both the Type I and Type II error levels are controlled for with SPRT, as opposed to standard hypothesis test formulations that only control for Type I error level. Multiple SPRTs are conducted concurrently across voxels and boundary error levels can be adjusted for instance by Bonferroni correction to account for this simultaneous testing. A practical modification of the original SPRT formulation for stopping is to consider the truncated SPRT (29), which will additionally call for stopping if an upper bound for the number of observations is reached. In our case, this is reached when a fixed number of blocks have been administered.

Ultimately, we aggregate the findings of the voxel-level SPRTs to determine whether or not experimentation within a block design should be terminated early. A "global" stopping rule that considers all voxels in a region of interest (can be whole brain or smaller ROIs) that we have adopted is to terminate task administration when a predetermined percentage of voxels have been classified by their respective SPRTs. For instance, we have used 80% as a global stopping criterion. Note that 80% classified means either as active or non-active. We choose this cut-off as it is fairly strict, which will facilitate correspondence in full scan data results. We also choose Type I and Type II error levels that are relatively more stringent for Type I error. For the future, we are considering adaptive and more flexible stopping rules beyond predetermined percentages. However, these will rely on vastly more powerful computational resources than available for our current experiments. For now, we stop early only in a conservative manner. Note that for $c'\beta$ parameter values that are "in-between" the null and alternative hypothesis values, the SPRT is indifferent to preferring one hypothesis over the other. This leads to larger numbers of scans needed before a stopping boundary is crossed. So, we have to accept a lack of decisive stopping decisions for these cases in order for overall experimentation to stop early. This is an acceptable trade-off for shorter experimental scan times and the ability to tailor experimentation.

Note that it also is important to balance the stopping times for active and non-active decisions, so that there is not a systematic bias towards one activation status being classified faster and driving attainment of the global stopping rule, while the other status has less decisive classification. For instance, a potential concern is if the non-active voxels do not require as many scans to satisfy the SPRT criterion. If there also is a high percentage of non-actives in the ROI for the experiment, then the global stopping rule would result in decisively classifying only non-active voxels. On the other hand, if the SPRT stopping times for active and non-active voxels are generally similar, then

there will be a proportional balance of decisively classified active and non-active voxels. This mix can be adjusted by selection of the respective error thresholds that determine stopping boundaries.

2.2 MRI Data Collection Methods

Fourteen healthy subjects were recruited, 7 males. They were aged 15-16 years old and 13 were right-handed. They had no known neurological conditions and a normal developmental history. A group of 11 adolescents born EPT were also recruited, 1 male. EPT is defined as being born at < 26-week gestation and weighing < 1000g. All were aged 15-17 years old and 8 were right-handed, 2 left-handed and 1 ambidextrous. All subjects were recruited as part of a larger study to evaluate functional and structural differences associated with mathematical abilities and working memory between those born EPT and those born at normal term. Adolescents were recruited as they can handle the stress of fMRI experimentation, are mathematically advanced enough and have had time to master the subject area. EPT subjects were included to show that differences with patient populations are detectable with our methods. A subsection of the full study is reported here to demonstrate the real-time analysis.

The subjects made one two-hour visit to the MRI department at University Hospitals Cleveland Medical Center (UHCMC). Ethics approval was obtained from the UHCMC Institutional Review Board office prior to the study and complied with the Declaration of Helsinki for human subject research. Subjects and their parents gave informed consent prior to taking part.

2.2.1 MRI protocols

The subjects were positioned head-first supine on the scanner bed with their head fixed in position using inflatable pads. An 8-channel head coil was used for data acquisition. Echo planar imaging scans were acquired on a Philips Ingenuity 3T PET/MR imager at UHCMC. The following fMRI scan parameters were used: TR = 3.0 s, TE = 35 ms, in-plane resolution was 1.797 mm^2 (matrix 128 x 128), slice thickness was 4 mm, number of slices = 36 slices and flip angle = 90°. A SENSE P reduction factor of 2 was implemented and scans were acquired in an ascending interleaved fashion.

In addition to the fMRI scans, a high-resolution T1-weighted anatomical image of the brain was also acquired. This was taken using a standard gradient-echo sequence with TurboFLASH. Imaging parameters were: TR = 7.5 ms, TE = 3.7 ms, in-plane resolution was 1 mm^2 (matrix 256 x 256), slice thickness was 1 mm, number of slices = 200 slices and flip angle = 8°.

2.2.2 Stimulus protocols

During data acquisition subjects were presented with a mathematical version of the well-known 1-back memory task. It involved performing basic addition and subtraction calculations and required the answer to be remembered and compared to the next answer. Two difficulty levels were included. The protocol was developed by our lab as part of a battery to assess mathematical and working memory abilities in 14 – 17-year olds to evaluate the functional differences between those born EPT and those born at normal term. The stimulus was presented on an MRI compatible LCD monitor (manufactured by Cambridge Research Systems, Rochester, UK) positioned at the end of

the bore and viewed via a mirror attached to the head coil. Equations were presented, for example, the subject may see "2 + 3 = ?". The subject was required to work out the answer and then remember it while working out the next equation, for example "1 + 4 = ?". If they thought the answers matched, then the subject pressed a button on a response box held in their right hand. If they thought the answers did not match, then they did nothing but remember the new answer to compare to the answer of the next equation. An example sequence is shown in Figure 1.

The stimulus was presented in a block design. 8 equations were presented per block. Each block lasted 36 seconds followed by 21 seconds of rest condition (fixation dot). Two difficulty levels were presented. The easier level consisted of single digit numbers to add or subtract and the answers were always a single digit. The harder level involved addition or subtraction of single or two-digit numbers and the answers were always two digits. Blocks of difficulty levels were alternated during the scan and a total of 6 blocks per level were presented. Note: although only 2 difficulty levels are used here, the setup is able to accommodate any number of difficulty levels. The full duration of the task was 238 scans or 11 minutes and 54 seconds. This was based on a moderate length of experimentation for a 1-back block design (e.g. see (30-33)), allowing 6 minutes for each task.

The visual stimulus was presented using an in-house custom written program that was developed using the Python programming language (Python Software Foundation, https://www.python.org/) and libraries from Psychopy - an open source visual presentation program (34-36). The program connected to a Cedrus Lumina controller to receive stimulus responses from the subject and trigger pulses from the MRI scanner (outputted every dynamic). The timing of the presentation of the visual stimulus was synchronized to the trigger pulses to ensure that stimulus images were displayed at the expected time. A Supervisor Window displayed on the experimenter's computer screen allowed the visual stimulus to be tracked throughout. It displayed the current block number being presented, how many remaining blocks there were and when the subject responded. The program was also able to terminate one or both of the difficulty levels if it received a signal indicating the relevant areas in the fMRI data were sufficiently classified across voxels. The software is freely available from the Bitbucket repository: https://bitbucket.org/tatsuoka-lab/fmri-presentation.

2.2.3 Real-time fMRI acquisition

Real-time image transfer was achieved by XTC (eXTernal Control). This is a program integrated into the Philips scanner software and enabled by a research clinical science key. XTC communicates with the reconstruction and scanner processes on the scanner computer and interfaces to a network Client application using a minimalistic CORBA (Common Object Request Broker Architecture) (37) interface which uses TCP/IP as the transport layer. CORBA is platform independent, reliable, and has the ability to process large amounts of data with minimum overhead. Each CORBA message consisted of a hierarchical attribute collection identified with UUIDs (universally unique identifiers) (38). Messages carried reconstructed image data and meta-data containing details of scan protocols. Due to hospital network security protocols the reconstructed images were placed in a folder on the scanner computer and then pushed across to a Linux computer. To achieve necessary image transfer speeds to the scanner computer folder a modification to XTC was installed on the scanner to disable two-way communications as only

one-way image transfer functionality was required. However, XTC does support two-way communication between the scanner and the Client.

The Linux computer was a custom-built server equipped with a solid state hard drive and two 8-core Intel Xeon E5-2687W processors running at 3.1 GHz and providing 40 MB L3 cache. It was installed with Centos 7.4 operating system. As the scans were received, custom written Python and Bash scripts implemented the analysis using core-based parallelization to preprocess the data and perform the SPRT statistical analysis. Preprocessing was performed using standard modules from AFNI (Analysis of Functional NeuroImages, https://afni.nimh.nih.gov) and FSL (FMRIB's Software Library, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). The analysis sequence is detailed in the following section. The setup is shown in Figure 2.

2.2.4 MRI preprocessing and ROI identification

At the beginning of the scanning session a single fMRI scan (3 sec) was acquired and used for coregistration (motion correction) purposes. In preparation, the skull was removed using FSL's Brain Extraction Tool (BET) (39) and a mask of the full brain was created. During the real-time adaptive fMRI scan session, new scans arrived every 3 seconds and was dumped in a folder on the Linux workstation where the following actions were applied to each one. AFNI's 'dcm2niix_afni' command was used to convert the .par/.rec files to nifti. Motion correction was performed using coregistration techniques. Every fMRI scan was realigned to the initial scan that was acquired before the task began, and AFNI's '3dvolreg' command was used. Spatial smoothing was also applied using an 8 mm kernel with AFNI's '3dmerge' command. The full brain mask created at the beginning of the session was applied using FSL's 'fslmaths' command to remove noisy voxels outside the brain (voxels of no interest). The resulting images were then converted to ascii format for statistical analysis with SPRT.

2.2.5 fMRI SPRT analysis

The SPRT analysis was applied using highly-optimized C++ program that used Intel Cilk Plus library for multicore and vector processing of data. BLAS routines from Intel MKL were used to enable instruction-based acceleration for matrix computation. They are available from the Bitbucket repository at https://bitbucket.org/tatsuoka-lab. The design matrix was created prior to the scan session using AFNI's '3dDeconvolve' command to model the stimulus and HRF. It is possible to include the temporal derivatives of the HRF or other regressors in the design matrix where applicable in studies. Temporal derivatives were not included here due to the long durations of the block design used to present the task. Statistical analysis included the modeling of low frequency physiological noise and the associated removal of serial correlation using discrete cosine transforms. Motion parameters are also frequently used as regressors to remove correlated activations produced by movement. Here motion parameter regressors were not included with the estimation of the discrete cosine transforms due to the time limitation of conducting the analysis on each scan within the 3 second TR. The automatic determination of when to terminate the scanning is based on the Type I and Type II errors, α_E and β_E, as described above. A percentage of voxels that must be classified before termination was also specified during the setup, such as 80%.

3.0 Results

3.1 Individual Subject Results of SPRT

All subjects were able to complete the 1-back task. The median control subject response time across both difficulty levels was 1.44 sec (SD 0.51 sec), and median task accuracy was 90.8 % (SD 20.2 %). When these are broken down by difficulty level, the easy level median task accuracy was 86.1 % (SD 22.6 %) with median response time of 1.28 sec (SD 0.54 sec); and the hard level median task accuracy was 90.0 % (SD 18.4 %) with median response time of 1.56 sec (SD 0.51 sec). EPT subjects had a slightly longer overall median response time of 1.91 sec (SD 0.48 sec) and overall median task accuracy was lower at 65.8 % (SD 21.2 %). For the easy level, the median accuracy was 72.2 % (SD 24.2 %) and median response time was 1.63 sec (SD 0.49 sec). For the hard level the median accuracy was 70.0 % (SD 19.8 %) with a median response time of 2.10 sec (SD 0.54 sec).

Functional MRI data were analyzed using typical Type I and II error levels, to test stopping time performance, with $\alpha_E = 0.001$, $\beta_E = 0.1$ (40, 41). The results of the SPRT versions are presented in Tables 1 A-B. A full-length task protocol comprised 238 scans and lasted 12 minutes. The first 78 scans were used for the first stage of experimentation and included two blocks per difficulty level. Note: data collection of subjects 1 and 3 were terminated early to test the early stopping functionality in real time and those datasets contain only 78 scans.

Real-time transfer speeds between the scanner and the Linux computer were consistently fast, with individual scan files taking less than 150 milliseconds to transfer. All subject scans were processed within the 3 second TR period. Offline testing showed that the subject with the largest number of voxels (subject 21 with 135,379 voxels) was processed in just 5 minutes and 45 seconds. The subject with the fewest number of voxels (subject 5 with 77,359 voxels) was processed in 5 minutes and 2 seconds. Thus, it is feasible for a TR of 2 seconds to be used with the software, depending on transfer speeds.

Stopping was reached at 80% and 90% classified in most instances. At 80% classification for controls, 86% of subjects stopped early for the easy level and 93% for the hard level. The median stopping time at 80% for controls was 84 scans (SD 18 scans) for the easy level and 85 scans (SD 10 scans) for the hard level. At 90% classification, 71% and 79% of control subjects stopped early for the easy and hard levels, respectively. For EPT subjects, at 80% classification, 91% of subjects stopped early for the easy level and all subjects stopped early for the hard level. At 90% classified, 91% stopped early for both levels. At 80% classified, the median stopping scan was 104 scans (SD 46 scans) for the easy level and 80 scans (SD 13 scans) for the hard level. At 90% classification, the median stopping scan times between the groups differed more significantly for the easy level: controls – 115 scans (SD 53 scans) and EPT – 141 scans (SD 44 scans). However, they are very similar for the hard level: controls 103 scans (SD 43 scans) and EPT – 106 scans (SD 46 scans).

In summary, both groups had a good rate of stopping early for 80 % of voxels classified and a slightly lower rate for 90 %. Note: it is not possible to know if subjects 1 and 3 would have stopping early if the full 238 scans had been collected. It should also be noted that those born EPT can have structural abnormalities of the brain which can affect fMRI results. Subject 23 had an enlarged right ventricle that was obvious on the T1 and fMRI images. This may have affected the results

preventing early stopping occurring during the easy level and at 90% classification for the hard level.

In Table 1, far right-hand column, are listed the active voxel counts if a full duration analysis was performed using all 238 scans (- 78 scans for subjects 1 and 3). For the easy level, two-thirds of subjects show an increase in the number of active voxels at scan 238 compared to the scan where 80% of voxels are classified. At the hard level, the majority of subjects (96% of subjects) increase the active voxel count compared to early stopping. Visual inspection of the t-score maps at the stopping scan and at the full duration revealed similar patterns of clusters with centers overlapping, see example in Figure 3. In most instances, the additional active voxels at full duration were around the edges of existing clusters at the early stopping scan. This is explored further in Table 2 where we show the number of active voxels in common spatially between the two durations.

In Table 2, the values given are for the scans where 80% of voxels were classified. The median spatial overlap for controls was 69.6% (SD 31.2%) for the easy level and 79.2% (SD 25.0%) for the hard level. For EPT subjects, the median was 64.3% (SD 32.6%) and 63.3% (31.6%) for the easy and hard levels, respectively. There are a few instances where the active voxel counts differ substantially between early stopping and full duration analyses (See Tables 1 A-B) - for example Subject 2, hard level: 80% classified with early stopping at scan 80, active voxel count = 2,026, full duration active voxel count = 23,351. The low early stopping active voxel count leads to a low percentage of spatial overlap with full duration, here 29.1 %. However, for most subjects, there is a good amount of spatial overlap between the two time points. Generally, across subjects the largest activations were centered bilaterally around the inferior and superior parietal areas, taking in the intraparietal sulcus, a region highly associated with mathematical functioning. Further activations were seen in the cuneus. These are most likely correlated with the visual processing associated with the task. Additional activations were seen in the precuneus, bilateral areas in the medial frontal gyrus, anterior cingulate, insula and inferior frontal gyrus. These areas are often associated with attention and memory systems (42, 43).

3.2 Group Analysis Results

The preprocessed files from the SPRT analysis can be used directly to perform a group analysis using AFNI's 3dMEMA command (Mixed Effects Meta Analysis tool) (44). However, A group analysis was carried out using FSL to demonstrate that the data collected in real-time can still be used in a typical post-hoc analysis. Raw data was preprocessed with FSL FEAT (45). Motion correction was performed using a rigid body transform, spatial smoothing with a full-width-at-half-maximum Gaussian kernel of 6mm was applied, high pass temporal filtering of 90 s was carried out and coregistration to (MNI) standard space was done before performing a first level individual GLM analysis. The statistical output from these were used to perform the higher level group statistics using FLAME 1 (FMRIB's Local Analysis of Mixed Effects, (46)). The results for the 1-back easy and hard contrasts are given for EPT and controls in Table 3 and Figure 4. The group results of full scan durations are compared to the group results using only the scans up to the early stopping point for each subject for each difficulty level.

For controls, there is consistency in location and overlap of clusters between the group activations for each level with strong activations in the anterior cingulate and parietal regions, See Tables 3

and 4, and Figure 4. This suggests that the data is being stopped at an optimal point and that scanning for longer might not be adding useful information to the dataset.

The EPT subjects demonstrate some activity in the parietal regions and, at full duration, the anterior cingulate too. However, there is much less activity as a group overall compared to control subjects. In order to understand this result it is necessary to consider neuropsychological skills and structural and functional brain changes within the group. Behavioral data collected as part of our wider study shows that nearly two thirds (63.6 %) of the EPT cohort have lower working memory function, compared to just over one third (35.7 %) of controls subjects. Working memory is a key skill required for both mathematics and this numerical 1-back task. Recall the lower accuracy and longer response times in the EPT group. fMRI studies on dyscalculia (difficulty in learning and performing mathematics) suggest that there is greater heterogeneity in activations with a more diffuse pattern being apparent (47, 48). Additionally, there is overlap in structural differences in white matter integrity, as measured from diffusion weighted imaging studies, between those born EPT and those with dyscalculia including inferior fronto-occipital fasciculus and the inferior and superior longitudinal fasciculi (49-52). These connect crucial areas associated with mathematics and working memory. A more diffuse and variable pattern of functional activity, perhaps partly due to structural differences, may confound a group analysis in this instance. More data points from individuals do seem to improve the results, perhaps allowing the variability to converge somewhat. This is supported by the change in variance for the group between early stopping and full duration analyses, see right-hand column of Table 4. The control group variances are much lower throughout. A group analysis of the EPT subjects based on early stopping at 90% classification does produce a map of activations similar to the full duration due to the longer scan times involved (not shown).

4.0 Discussion

We present a workflow for the implementation of an adaptive real-time fMRI system that allows for statistically-driven dynamic adjustment of experimentation based on voxel-level SPRT. We show that this dynamic and adaptive statistical approach is comparable to corresponding fixed experimental designs in terms of detected activation. At the same time, time savings in experiment durations ranged between approximately one third and two thirds of the full protocol. The SPRT results in Tables 1 and 2 indicate that results are fairly comparable with a full protocol analysis. The statistical stopping criteria of the approach help ensure accurate characterizations of activation patterns while reducing time needed for scanning and providing a framework for flexible yet systematic dynamic paradigms.

Our SPRT approach was effective at detecting brain activity at the individual level with early stopping in both the controls and EPT groups. We note the individual variability among subjects in early stopping performance. Factors that can affect stopping times include the magnitudes of activation, variability in task performance and the noise levels in the BOLD signal. In the future, we will consider development of adaptive approaches to individualize global stop rule threshold selection that can encourage early stopping while balancing the potential for improved activation classification with continued experimentation.

There are many possible applications in the research setting where individual level results may be the focus. A possible clinical application may be in clinical assessments for presurgical evaluation for brain surgery in patients with brain tumors or epilepsy. A group analysis was also feasible using the early stopping data in controls. A possible limitation was discovered in performing a group analysis of the EPT group who demonstrated greater variability in location at the individual level. While it is feasible to apply our approach for patient group studies, consideration should be given to the particular patient group of interest and the likely differences in brain activity when making the decision to terminate early. If early stopping is attractive, such as to reduce burden for an impaired population, we conjecture that larger sample sizes or stricter early stopping criteria may help overcome this larger variability.

Use of SPRT allows for voxel-level control of both Type I (α_E) and Type II (β_E) error relative to a hypothesized GLM parameter or contrast. Other α_E, β_{E-} pairs were considered as well, to test how different combinations impact activity detection and early termination. For instance, given selection of $\alpha_E = 0.001$ and $\beta_E = 0.01$, overall stopping did not occur. In this case, the more stringent choice of β_E makes it more difficult to cross either of the SPRT thresholds, although here the active voxels are still classified relatively quickly. We also saw that for $\alpha_E = 0.001$ and $\beta_E = 0.1$, we gain dramatic gains in shorter scan durations for experimentation, across both of the conditions that are considered. Note, that for SPRT, stopping is less likely for values in-between the null and alternative hypothesis values. As we selected a stringent alternative value (with t-statistic approximately equal to 3.1), this makes stopping less likely if we focus on reducing Type II error. Also, since we are stopping "globally", SPRT values exceed stopping thresholds in many cases.

In terms of the global stop rule threshold, we observed that for the cases under consideration, stopping when 80% of voxels in the full brain (or smaller ROI) respectively satisfy their SPRT-based stopping criterion generally leads to early termination of stimulus administration, while also leading to comparable activation classification as with the full protocol. The stricter 90% criterion is not always satisfied, and hence does not always lead to early stopping of experimentation. Recall that when GLM parameter values are "in-between" the null and alternative hypothesis values, SPRT-based stopping is less likely at the voxel level. A 100% stopping rule is thus not feasible. Importantly, note that group-level analyses can still be conducted with varying experimentation durations across subjects, for example using AFNI's '3dMEMA' module (44), and data collected in real-time may also be analysed using normal post-hoc methods implemented, for example, in FSL.

Here we demonstrated full brain analytics with parallelization using MKL Intel libraries for matrix computation with two Xeon E5-2687W 8-core processors. It is also feasible to consider only partial brain volumes where experiments demand more consideration of a particular area. Future directions for the study are to implement the SPRT and Bayesian sequential estimation methods using distributed computing approaches to increase processing speed allowing full brain real-time analyses and advance stopping rule methods in shorter scan times. This could include dynamic selection of global stopping rule thresholds, and look-ahead stopping rules that rely on probabilistic prediction.

5.0 Conclusion

We introduce a systematic, statistically-based approach to dynamic experimentation with real-time fMRI. Saving in scan time and accurate voxel activation detection can be achieved, while redundant experimentation in block designs is reduced. These methods lay a foundation for future dynamic experimentation approaches with fMRI, and for real-time quality control analyses to assess statistical error performance across voxels. Use of high performance computing enables the advent of these more sophisticated real-time experimental designs.

Declaration of conflicts of interest: All authors declare no conflicts of interest.

Author contributions

SC – Study design, analysis and interpretation of data, drafting of manuscript
WC – Study design, analysis and interpretation of data, software development, drafting of manuscript
JF – Study design and software development
HF – Study design and technical support, drafting of manuscript
CT – Study design, analysis and interpretation of data, drafting of manuscript

Funding: This study was supported by Phillips Healthcare and the National Science Foundation (Award number: 1561716).

Acknowledgements
None

Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

References

1. A. A. Alegria, M. Wulff, H. Brinson, G. J. Barker, L. J. Norman, D. Brandeis, D. Stahl, A. S. David, E. Taylor, V. Giampietro and K. Rubia: Real-Time fMRI Neurofeedback in Adolescents with Attention Deficit Hyperactivity Disorder. *Human Brain Mapping*, 38(6), 3190-3209 (2017) doi:10.1002/hbm.23584
2. E. J. Lawrence, L. Su, G. J. Barker, N. Medford, J. Dalton, S. C. R. Williams, N. Birbaumer, R. Veit, S. Ranganatha, J. Bodurka, M. Brammer, V. Giampietro and A. S. David: Self-regulation of the anterior insula: Reinforcement learning using real-time fMRI neurofeedback. *Neuroimage*, 88, 113-124 (2014) doi:10.1016/j.neuroimage.2013.10.069
3. M. S. Sherwood, J. H. Kane, M. P. Weisend and J. G. Parker: Enhanced control of dorsolateral prefrontal cortex neurophysiology with real-time functional magnetic resonance imaging (rt-fMRI) neurofeedback training and working memory practice. *Neuroimage*, 124, 214-223 (2016) doi:10.1016/j.neuroimage.2015.08.074
4. C. Hohenfeld, N. Nellessen, I. Dogan, H. Kuhn, C. Muller, F. Papa, S. Ketteler, R. Goebel, A. Heinecke, N. J. Shah, J. B. Schulz, M. Reske and K. Reetz: Cognitive Improvement and Brain Changes after Real-Time Functional MRI Neurofeedback Training in Healthy Elderly and Prodromal Alzheimer's Disease. *Frontiers in Neurology*, 8 (2017) doi:10.3389/fneur.2017.00384

5. L. Subramanian, J. V. Hindle, S. Johnston, M. V. Roberts, M. Husain, R. Goebel and D. Linden: Real-Time Functional Magnetic Resonance Imaging Neurofeedback for Treatment of Parkinson's Disease. *Journal of Neuroscience*, 31(45), 16309-16317 (2011) doi:10.1523/jneurosci.3498-11.2011
6. L. Subramanian, M. B. Morris, M. Brosnan, D. L. Turner, H. R. Morris and D. E. J. Linden: Functional Magnetic Resonance Imaging Neurofeedback-guided Motor Imagery Training and Motor Training for Parkinson's Disease: Randomized Trial. *Frontiers in Behavioral Neuroscience*, 10 (2016) doi:10.3389/fnbeh.2016.00111
7. S. Lui, X. H. J. Zhou, J. A. Sweeney and Q. Y. Gong: Psychoradiology: The Frontier of Neuroimaging in Psychiatry. *Radiology*, 281(2), 357-372 (2016) doi:10.1148/radiol.2016152149
8. R. W. Cox, A. Jesmanowicz and J. S. Hyde: REAL-TIME FUNCTIONAL MAGNETIC-RESONANCE-IMAGING. *Magnetic Resonance in Medicine*, 33(2), 230-236 (1995) doi:10.1002/mrm.1910330213
9. R. Goebel: BrainVoyager - Past, present, future. *Neuroimage*, 62(2), 748-756 (2012) doi:10.1016/j.neuroimage.2012.01.083
10. J. R. Sato, R. Basilio, F. F. Paiva, G. J. Garrido, I. E. Bramati, P. Bado, F. Tovar-Moll, R. Zahn and J. Moll: Real-Time fMRI Pattern Decoding and Neurofeedback Using FRIEND: An FSL-Integrated BCI Toolbox. *Plos One*, 8(12) (2013) doi:10.1371/journal.pone.0081658
11. J. J. Yoo, O. Hinds, N. Ofen, T. W. Thompson, S. Whitfield-Gabrieli, C. Triantafyllou and J. D. E. Gabrieli: When the brain is prepared to learn: Enhancing human learning using real-time fMRI. *Neuroimage*, 59(1), 846-852 (2012) doi:10.1016/j.neuroimage.2011.07.063
12. O. Hinds, S. Ghosh, T. W. Thompson, J. J. Yoo, S. Whitfield-Gabrieli, C. Triantafyllou and J. D. E. Gabrieli: Computing moment-to-moment BOLD activation for real-time neurofeedback. *Neuroimage*, 54(1), 361-368 (2011) doi:10.1016/j.neuroimage.2010.07.060
13. M. T. Debettencourt, J. D. Cohen, R. F. Lee, K. A. Norman and N. B. Turk-Browne: Closed-loop training of attention with real-time brain imaging. *Nature Neuroscience*, 18(3), 470-165 (2015) doi:10.1038/nn.3940
14. R. Lorenz, R. P. Monti, I. R. Violante, C. Anagnostopoulos, A. A. Faisal, G. Montana and R. Leech: The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *Neuroimage*, 129, 320-334 (2016) doi:10.1016/j.neuroimage.2016.01.032
15. R. Lorenz, I. R. Violante, R. P. Monti, G. Montana, A. Hampshire and R. Leech: Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. *Nature Communications*, 9 (2018) doi:10.1038/s41467-018-03657-3
16. I. J. Feng, A. I. Jack and C. Tatsuoka: Dynamic Adjustment of Stimuli in Real Time Functional Magnetic Resonance Imaging. *Plos One*, 10(3) (2015) doi:10.1371/journal.pone.0117942
17. M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich and S. M. Smith: FSL. *Neuroimage*, 62(2), 782-790 (2012) doi:10.1016/j.neuroimage.2011.09.015
18. N. Lazar: The Statistical Analysis of Functional MRI Data. Springer-Verlag New York, (2008) doi:10.1007/978-0-387-78191-4
19. R. W. Cox: AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 162-173 (1996) doi:10.1006/cbmr.1996.0014
20. R. J. Carroll, S. Wang, D. G. Simpson, A. J. Stromberg and D. Ruppert: The Sandwich (Robust Covariance Matrix) Estimator. In: Texas A&M University, (1998)
21. G. Kauermann and R. J. Carroll: A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96(456), 1387-1396 (2001) doi:10.1198/016214501753382309
22. K. J. Friston, O. Josephs, E. Zarahn, A. P. Holmes, S. Rouquette and J. B. Poline: To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *Neuroimage*, 12(2), 196-208 (2000) doi:10.1006/nimg.2000.0609

23. C. Yan, D. Liu, Y. He, Q. Zou, C. Zhu, X. Zuo, X. Long and Y. Zang: Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS One*, 4(5), e5743 (2009) doi:10.1371/journal.pone.0005743
24. A. Wald: Sequential Analysis. John Wiley and Sons, (1947)
25. A. Wald and J. Wolfowitz: OPTIMUM CHARACTER OF THE SEQUENTIAL PROBABILITY RATIO TEST. *Annals of Mathematical Statistics*, 19(3), 326-339 (1948) doi:10.1214/aoms/1177730197
26. D. R. Cox: Large Sample Sequential Tests for Composite Hypotheses. *The Indian Journal of Statistics, Series A (1961-2002)*, 25(1), 5-12 (1963)
27. S. Tantaratana and J. B. Thomas: TRUNCATED SEQUENTIAL PROBABILITY RATIO TEST. *Information Sciences*, 13(3), 283-300 (1977) doi:10.1016/0020-0255(77)90050-0
28. J. X. Li: Sequential probability ratio tests for generalized linear mixed models. In: University of California, Riverside, (2010)
29. J. B. T. Sawasd Tantaratana: Truncated sequential probability ratio test. *Information Sciences*, 13(3), 283-300 (1977) doi:http://dx.doi.org/10.1016/0020-0255(77)90050-0
30. X. Di, H. M. Zhang and B. B. Biswal: Anterior cingulate cortex differently modulates frontoparietal functional connectivity between resting-state and working memory tasks. *Human Brain Mapping*, 41(7), 1797-1805 (2020) doi:10.1002/hbm.24912
31. J. Z. Huang, L. Xie, R. W. Guo, J. H. Wang, J. Q. Lin, Z. B. Sun, S. X. Duan, Z. R. Lin, H. Li and S. H. Ma: Abnormal brain activity patterns during spatial working memory task in patients with end-stage renal disease on maintenance hemodialysis: a fMRI study. *Brain Imaging and Behavior* (2020) doi:10.1007/s11682-020-00383-7
32. M. Daamen, J. G. Bauml, L. Scheef, C. Sorg, B. Busch, N. Baumann, P. Bartmann, D. Wolke, A. Wohlschlager and H. Boecker: Working Memory in Preterm-Born Adults: Load-Dependent Compensatory Activity of the Posterior Default Mode Network. *Human Brain Mapping*, 36(3), 1121-1137 (2015) doi:10.1002/hbm.22691
33. G. R. Poudel, J. C. Stout, J. F. Dominguez, M. A. Gray, L. Salmon, A. Churchyard, P. Chua, B. Borowsky, G. F. Egan and N. Georgiou-Karistianis: Functional changes during working memory in Huntington's disease: 30-month longitudinal data from the IMAGE-HD study. *Brain Structure & Function*, 220(1), 501-512 (2015) doi:10.1007/s00429-013-0670-z
34. J. W. Peirce: PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13 (2007) doi:10.1016/j.jneumeth.2006.11.017
35. J. W. Peirce: Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2 (2009) doi:10.3389/neuro.11.010.2008
36. J. Peirce and M. MacAskill: Building Experiments in Pscychopy. Sage Publications Ltd, London, England (2018)
37. D. C. Schmidt, D. L. Levine and S. Mungee: The design of the TAO real-time object request broker. *Computer Communications*, 21(4), 294-324 (1998) doi:10.1016/s0140-3664(97)00165-5
38. ITU-T: Series X: Data Networks and Open System Communications. In: Ed I. T. S. S. o. ITU. ISO/IEC, (2005)
39. S. M. Smith: Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143-155 (2002) doi:10.1002/hbm.10062
40. H. R. Cremers, T. D. Wager and T. Yarkoni: The relation between statistical power and inference in fMRI. *Plos One*, 12(11) (2017) doi:10.1371/journal.pone.0184923
41. A. Banerjee, U. B. Chitnis, S. L. Jadhav, J. S. Bhawalkar and S. Chaudhury Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, 18, 127-131 (2009) doi:0.4103/0972-6748.62274
42. V. Menon: Memory and cognitive control circuits in mathematical cognition and learning. *Mathematical Brain across the Lifespan*, 227, 159-186 (2016) doi:10.1016/bs.pbr.2016.04.026

43. A. E. Cavanna and M. R. Trimble: The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129, 564-583 (2006) doi:10.1093/brain/awl004
44. G. Chen, Z. S. Saad, A. R. Nath, M. S. Beauchamp and R. W. Cox: FMRI group analysis combining effect estimates and their variances. *Neuroimage*, 60(1), 747-765 (2012) doi:10.1016/j.neuroimage.2011.12.060
45. M. W. Woolrich, B. D. Ripley, M. Brady and S. M. Smith: Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage*, 14(6), 1370-1386 (2001) doi:10.1006/nimg.2001.0931
46. M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, M. Jenkinson and S. M. Smith: Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage*, 21(4), 1732-1747 (2004) doi:10.1016/j.neuroimage.2003.12.023
47. K. Kucian, T. Loenneker, T. Dietrich, M. Dosch, E. Martin and M. von Aster: Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behav Brain Funct* . 2, 31 (2006) doi:10.1186/1744-9081-2-31
48. M. M. Berl, C. J. Vaidya and W. D. Gaillard: Functional imaging of developmental and adaptive changes in neurocognition. *Neuroimage*, 30(3), 679-691 (2006) doi:10.1016/j.neuroimage.2005.10.007
49. I. M. Loe, J. N. Adams and H. M. Feldman: Executive Function in Relation to White Matter in Preterm and Full Term Children. *Frontiers in Pediatrics*, 6 (2019) doi:10.3389/fped.2018.00418
50. J. M. Young, B. R. Morgan, H. E. A. Whyte, W. Lee, M. L. Smith, C. Raybaud, M. M. Shroff, J. G. Sled and M. J. Taylor: Longitudinal Study of White Matter Development and Outcomes in Children Born Very Preterm. *Cerebral Cortex*, 27(8), 4094-4105 (2017) doi:10.1093/cercor/bhw221
51. E. Rykhlevskaia, L. Q. Uddin, L. Kondos and V. Menon: Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience*, 3 (2009) doi:10.3389/neuro.09.051.2009
52. K. Kucian, S. S. Ashkenazi, J. Hanggi, S. Rotzer, L. Jancke, E. Martin and M. von Aster: Developmental dyscalculia: a dysconnection syndrome? *Brain Structure & Function*, 219(5), 1721-1733 (2014) doi:10.1007/s00429-013-0597-4

Figure 1: Sample 1-back protocols demonstrating the two difficulty levels.

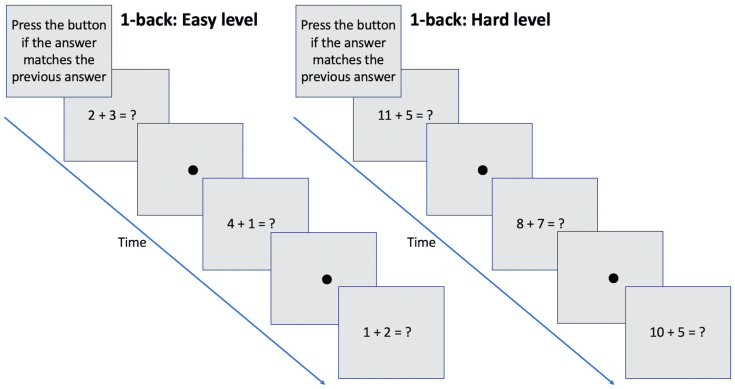

Figure 2: Schematic of the experimental setup of the dynamic real-time fMRI process. The equations were presented to the subject while the scans were acquired and exported to the Linux workstation for processing with SPRT statistics. The results were relayed back to the presentation program with instruction to either continue or terminate the stimulus.

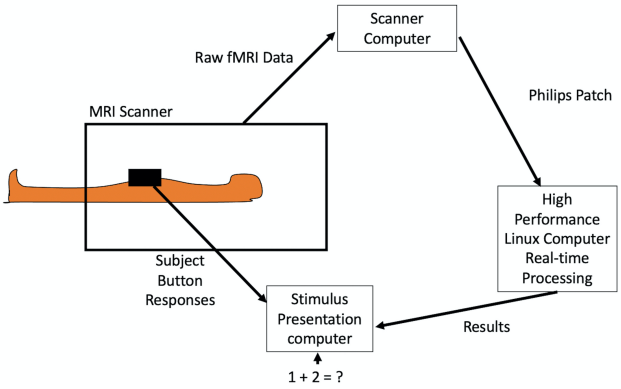

Figure 3: Full brain activation t-score maps that are associated with the easy and hard level 1-back task for 1 EPT subject (number 20). Left in red shows SPRT results with early stopping. Right in blue shows full duration results (activations at scan 238). $P \leq 0.001$ uncorrected. Shown in subject own space. R = right, L = left, A = anterior, P = posterior.

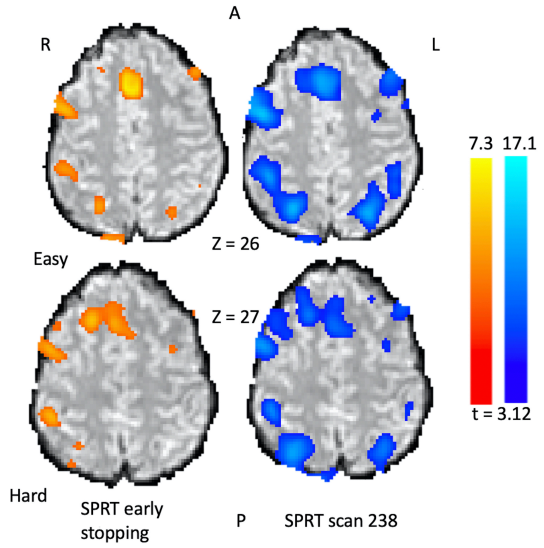

Figure 4: Group results for the 1-back task. Analysis performed for controls and EPT subjects using FSL. Early stopping is compared to full duration. Activations are overlaid on the MNI template brain. Red (Left) = early stopping group results, Blue (Right) full duration (238 scans) group results. $P < 0.001$ uncorrected. Slices z = 58 is shown. R = right, L = left, A = anterior, P = posterior.

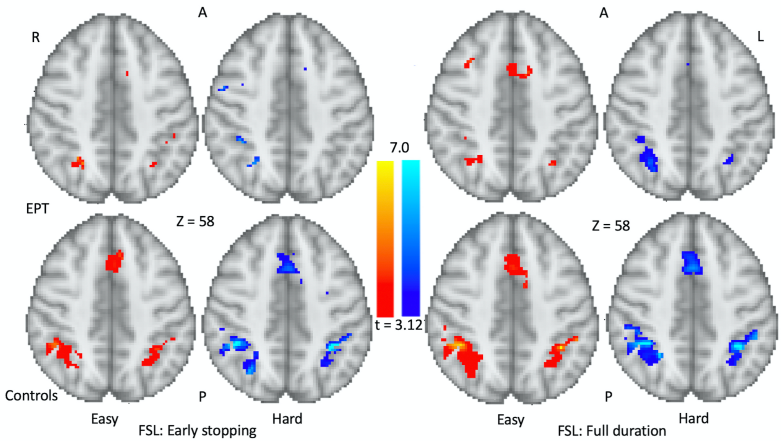

Table 1: Subject results of the 1-back task using SPRT to analyse the data. Analysis reported here uses $\alpha_E = 0.001$, $\beta_E = 0.1$ and thresholded at $p < 0.001$. A) Easy level, B) Hard level.

A) Easy level

Subject	No of Voxels in ROI (full brain)	No of Voxels Classified Active at 80%	No of Voxels Classified Active at 90%	Scan when 80% Reached	Scan when 90% Reached	No of Voxels Classified Active at Scan 238
Controls						
1*	91,936	7,935^	7,935^	Not reached	Not reached	7,935
2	115,062	6,182	2,717	121	140	3,556
3*	103,128	1,010^	1,010^	Not reached	Not reached	1,010
4	113,564	5,236	9,049	83	155	12,803
5	77,359	4,531	8,233	79	109	13,472
6	103,591	15,065	13,680^	132	Not reached	13,680
7	114,260	2,583	6,505	100	108	10,540
8	121,353	979	979	79	79	4,674
9	107,406	115	1,892	79	108	4,115
10	106,267	7,242	9,341	107	156	7,124
11	121,195	349	349	79	79	9,329
12	96,565	549	3,596	103	139	7,340
13	107,016	2,146	2,398	84	120	6,357
14	96,936	10,150	12,429^	104	Not reached	12,429
Median values				84 (SD 18)	115 (SD 53)	
EPT						
15	94,623	64	56	79	100	49
16	94,905	4,870	10,779	104	141	15,866
17	98,799	795	580	105	108	750
18	118,098	16,139	14,598	111	177	13,484
19	124,749	4,782	5,192	79	165	3,487
20	97,437	734	1,263	79	105	3,925

21	135,379	246	6,854	85	140	7,402
22	89,609	1,397	3,664	79	108	8,093
23	104,584	10,039^	10,039^	Not reached	Not reached	10,039
24	114,201	5,881	11,637	107	203	13,817
25	86,177	11,354	9,437	106	148	6,715
Median values				104 (SD 46)	141 (SD 44)	

B) Hard level

Subject	**No of Voxels in ROI (full brain)**	**No of Voxels Classified Active at 80%**	**No of Voxels Classified Active at 90%**	**Scan when 80% Reached**	**Scan when 90% Reached**	**No of Voxels Classified Active at Scan 238**
Controls						
1*	91,936	2,708^	2,708^	Not reached	Not reached	2,708
2	115,062	2,026	2,737	80	102	23,351
3*	103,128	2,237	2,237^	78	Not reached	2,237
4	113,564	12,051	10,462	102	121	20,749
5	77,359	4,127	8,409	89	104	14,976
6	103,591	5,322	5,722	84	98	9,874
7	114,260	3,743	11,787	79	114	14,939
8	121,353	3,784	5,233	87	113	7,829
9	107,406	1,194	330	85	95	2,494
10	106,267	7,229	7,977	97	117	14,634
11	121,195	5,527	3,380	85	101	8,898
12	96,565	6,048	3,353	80	101	10,921
13	10,7016	5,205	7,609	110	183	11,355
14	96,936	6,352	29,279^	98	Not reached	29,279
Median values				85 (SD 10)	103 (SD 43)	
EPT						
15	94,623	139	139	79	79	860
16	94,905	7,779	19,974	79	158	21,074
17	98,799	1,325	922	84	87	1,696
18	118,098	11,491	10,401	99	126	14,984
19	124,749	1,848	2,212	79	96	4,335
20	97,437	16,677	9,877	119	156	10,967
21	135,379	1,310	1,223	79	86	7,609
22	89,609	1,563	4,979	79	106	8,754
23	104,584	1,822	10,038^	80	Not reached	10,038
24	114,201	2,651	8,197	84	114	15,374
25	86,177	6,110	8,715	97	105	8,088
Median values				80 (SD 13)	106 (SD 46)	

* The datasets of these subjects contain only 78 scans in total.

^ Values taken from last scan – scan 78 in subjects 1 and 3, scan 238 in subjects 6, 14 and 23.

Table 2: At each subject's stopping scan, the number of active voxels that spatially overlap between early stopping and full duration are listed. The percentage of voxels-in-common is also given relative to the total number of active voxels detected at the stopping scan, given in Tables 1 A-B. Stopping based on 80% classification. Maximum number of possible scans is 238, minimum is 78 scans, equal to two blocks of easy and hard stimulus administration.

	Easy Level			**Hard Level**		
Subject	Stopping Scan	No. of Voxels in Common at Scan 238	% of Common Voxels with Scan 238	Stopping Scan	No. of Voxels in Common at Scan 238	% of Common Voxels with Scan 238
Controls						
1*	78	7,935	100	78	2,708	100
2	121	1,797	29.1	80	1,208	59.6
3*	78	1,010	100	78	2,237	100
4	83	1,888	36.1	102	10,610	88.0
5	79	4,410	97.3	89	4,098	99.3
6	132	11,278	74.9	84	3,724	70.0
7	100	2,414	93.5	79	3,608	96.4
8	79	432	44.1	87	2,597	68.6
9	79	103	89.6	85	334	28.0
10	107	4,653	64.3	97	5,838	80.8
11	79	92	26.4	85	4,289	77.6
12	103	456	83.1	80	2,829	46.8
13	84	194	9.0	110	1,460	28.0
14	104	5,447	53.7	98	5,451	85.8
Median values	84 (SD 18)		69.6 (SD 31.2)	85 (SD 10)		79.2 (SD 25.0)
EPT						
15	79	0	0.0	79	0	0.0
16	104	4,028	82.7	79	7,233	93.0
17	105	137	17.2	84	294	22.2
18	111	10,379	64.3	99	7,278	63.3
19	79	1,251	26.2	79	930	50.3
20	79	361	49.2	119	5,904	35.4
21	85	161	65.4	79	549	41.9
22	79	1,358	97.2	79	1,510	96.6
23	238	10,039	100	80	1,194	65.5
24	107	4,576	77.8	84	2,495	94.1
25	106	5,361	47.2	97	4,801	78.6
Median values	104 (SD 46)		64.3 (SD 32.6)	80 (SD 13)		63.3 (SD 31.6)

Table 3: List of activations from a group analysis using FSL. Easy and hard levels for controls and EPT subjects are reported for early stopping and full scan durations. Thresholds of $p < 0.001$ uncorrected and minimum cluster extent = 10 voxels have been applied.

Group	Cluster No.	Coordinates (MNI)	No. of Voxels	Peak t-score	Side	Location
Early stopping:						
Easy						
EPT	1	-44 4 30	52	4.92	Left	Precentral Gyrus
	2	-42 8 0	40	4.68	Left	Insula
	3	2 14 38	36	5.61	Left	Cingulate Gyrus
	4	-40 -40 36	31	4.57	Left	Supramarginal Gyrus
	5	-52 10 0	29	5.81	Left	Superior Temporal Gyrus
	6	10 4 28	12	3.88	Right	Cingulate Gyrus
	7	48 6 48	11	4.83	Right	Middle Frontal Gyrus
Controls	1	-40 28 26	690	6.66	Left	Middle Frontal Gyrus
	2	38 -46 46	382	7.1	Right	Inferior Parietal Lobule
	3	-6 8 52	352	5.81	Left	Medial Frontal Gyrus
	4	-34 -50 48	134	4.99	Left	Inferior Parietal Lobule
	5	-42 14 2	69	4.53	Left	Insula
	6	22 -32 18	64	4.04	Right	Caudate
	7	14 6 64	57	3.84	Right	Superior Frontal Gyrus
	8	-8 8 18	25	3.97	Left	Caudate
	9	-14 12 34	24	4.75	Left	Cingulate Gyrus
	10	-42 44 16	21	5	Left	Middle Frontal Gyrus
	11	-20 8 66	18	5.05	Left	Superior Frontal Gyrus
	12	38 16 2	18	3.55	Right	Claustrum
	13	6 12 18	14	3.62	Right	Anterior Cingulate
	14	-32 60 16	10	3.56	Left	Superior Frontal Gyrus
	15	-46 8 50	10	3.48	Left	Middle Frontal Gyrus
Hard						
EPT	1	28 -62 48	88	6.45	Right	Superior Parietal Lobule
	2	40 -42 48	84	6.06	Right	Inferior Parietal Lobule
	3	-42 2 26	48	5.35	Left	Precentral Gyrus
	4	32 -54 66	26	4.36	Right	Superior Parietal Lobule
	5	18 12 62	26	4.81	Right	Medial Frontal Gyrus
	6	-32 20 28	17	3.67	Left	Middle Frontal Gyrus
	7	8 28 26	16	4.6	Right	Cingulate Gyrus
	8	-8 8 52	13	4.2	Left	Medial Frontal Gyrus
	9	-6 0 70	12	4.45	Left	Superior Frontal Gyrus
	10	50 0 44	10	4.87	Right	Precentral Gyrus
Controls	1	-10 14 36	509	5.99	Left	Cingulate Gyrus
	2	40 -44 44	481	8.54	Right	Inferior Parietal Lobule
	3	-44 2 30	237	5.09	Left	Precentral Gyrus

	4	-32 -48 46	236	9.48	Left	Inferior Parietal Lobule
	5	-54 22 26	82	5.7	Left	Middle Frontal Gyrus
	6	-32 22 0	30	4.32	Left	Claustrum
	7	-18 6 66	28	4	Left	Middle Frontal Gyrus
	8	-6 0 52	25	4.93	Left	Medial Frontal Gyrus
	9	-38 22 26	21	4.41	Left	Middle Frontal Gyrus
	10	-30 42 30	14	3.75	Left	Middle Frontal Gyrus
	11	-40 -54 62	10	5.62	Left	Superior Parietal Lobule
Full duration:						
Easy						
EPT	1	28 -56 50	150	4.93	Right	Superior Parietal Lobule
	2	-10 8 44	144	4.49	Left	Cingulate Gyrus
	3	-40 12 32	44	4.81	Left	Precentral Gyrus
	4	52 20 38	28	5.48	Right	Middle Frontal Gyrus
	5	32 18 44	28	4.54	Right	Middle Frontal Gyrus
	6	-30 -60 44	17	4.30	Left	Angular Gyrus
	7	8 30 38	14	4.16	Right	Cingulate Gyrus
	8	46 14 28	13	5.75	Right	Precentral Gyrus
Controls	1	-46 0 26	754	6.52	Left	Inferior Frontal Gyrus
	2	40 -44 46	612	9.21	Right	Inferior Parietal Lobule
	3	-12 14 36	573	6.05	Left	Cingulate Gyrus
	4	-36 -46 44	225	9.53	Left	Inferior Parietal Lobule
	5	20 -30 18	65	5.06	Right	Caudate
	6	-40 18 -2	17	3.62	Left	Insula
Hard						
EPT	1	-14 14 34	446	6.78	Left	Cingulate Gyrus
	2	36 -42 48	418	5.93	Right	Precuneus
	3	-44 2 38	66	5.09	Left	Precentral Gyrus
	4	-30 -58 42	40	4.31	Left	Angular Gyrus
	5	-50 22 24	15	3.65	Left	Inferior Frontal Gyrus
	6	-32 16 34	15	3.98	Left	Middle Frontal Gyrus
	7	8 0 68	14	4.2	Right	Medial Frontal Gyrus
	8	12 4 28	11	4.26	Right	Cingulate Gyrus
	9	-10 -4 66	11	3.68	Left	Medial Frontal Gyrus
	10	40 10 26	10	4.25	Right	Precentral Gyrus
Controls	1	-10 14 36	680	7.70	Left	Cingulate Gyrus
	2	38 -44 44	590	9.12	Right	Inferior Parietal Lobule
	3	-34 -46 44	277	9.47	Left	Inferior Parietal Lobule
	4	-44 0 28	259	5.16	Left	Precentral Gyrus
	5	-46 22 22	128	5.00	Left	Middle Frontal Gyrus
	6	-16 -24 22	68	4.02	Left	Caudate
	7	20 -32 18	63	3.94	Right	Caudate
	8	-40 16 -6	45	5.62	Left	Insula
	9	-22 6 66	12	3.43	Left	Superior Frontal Gyrus

10	0 -86 6	11	3.69	Left	Lingual Gyrus

Table 4: The number of active voxels that spatially overlap between early stopping and full duration group analyses are listed. Images thresholded at $p < 0.001$. The percentage of voxels-in-common is given relative to the total number of active voxels detected at early stopping. Stopping based on 80% classification at the individual level.

	No of Active Voxels	No. of Voxels in Common at Scan 238	% of Common Voxels with Scan 238	Variance Magnitudes
EPT - Easy				
Early stopping	409			45,939 (SD 3,664)
Full duration	528	320	21.8	18,351 (SD 1,135)
EPT - Hard				
Early stopping	477			35,842 (SD 4,177)
Full duration	1,186	319	33.1	27,110 (SD 1,054)
Controls - Easy				
Early stopping	2,034			8,819 (SD 631)
Full duration	2,276	699	65.6	11,082 (SD 469)
Controls - Hard				
Early stopping	1,765			11,156 (SD 943)
Full duration	2,520	545	69.1	9,104 (SD 685)