



METHOD ARTICLE

REVISED **Active Acquisition for multimodal neuroimaging [version 2; peer review: 2 approved, 1 approved with reservations]**

James H. Cole ¹, Romy Lorenz ^{2,3}, Fatemeh Geranmayeh ⁴, Tobias Wood ¹, Peter Hellyer¹, Steven Williams ¹, Federico Turkheimer ¹, Robert Leech ¹

¹Department of Neuroimaging, Institute of Psychiatry, Psychology and Neurosciences, King's College London, London, UK

²MRC Centre for Cognition and Brain Sciences, University of Cambridge, Cambridge, UK

³Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

⁴Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK

v2 **First published:** 14 Nov 2018, 3:145 (<https://doi.org/10.12688/wellcomeopenres.14918.1>)

Latest published: 23 Sep 2019, 3:145 (<https://doi.org/10.12688/wellcomeopenres.14918.2>)

Abstract

In many clinical and scientific situations the optimal neuroimaging sequence may not be known prior to scanning and may differ for each individual being scanned, depending on the exact nature and location of abnormalities. Despite this, the standard approach to data acquisition, in such situations, is to specify the sequence of neuroimaging scans prior to data acquisition and to apply the same scans to all individuals. In this paper, we propose and illustrate an alternative approach, in which data would be analysed as it is acquired and used to choose the future scanning sequence: Active Acquisition. We propose three Active Acquisition scenarios based around multiple MRI modalities. In Scenario 1, we propose a simple use of near-real time analysis to decide whether to acquire more or higher resolution data, or acquire data with a different field-of-view. In Scenario 2, we simulate how multimodal MR data could be actively acquired and combined with a decision tree to classify a known outcome variable (in the simple example here, age). In Scenario 3, we simulate using Bayesian optimisation to actively search across multiple MRI modalities to find those which are most abnormal. These simulations suggest that by actively acquiring data, the scanning sequence can be adapted to each individual. We also consider the many outstanding practical and technical challenges involving normative data acquisition, MR physics, statistical modelling and clinical relevance. Despite these, we argue that Active Acquisition allows for potentially far more powerful, sensitive or rapid data acquisition, and may open up different perspectives on individual differences, clinical conditions, and biomarker discovery.

Keywords

Neuroimaging, active acquisition, MRI, brain

Open Peer Review

Reviewer Status ? ✓ ✓

	Invited Reviewers		
	1	2	3
REVISION			✓
version 2 published 23 Sep 2019			report
			↑
version 1 published 14 Nov 2018	? report	✓ report	✗ report

- Hugo G Schnack** , Utrecht University, Utrecht, The Netherlands
- Diego Vidaurre**, University of Oxford, Oxford, UK
- Ron Kikinis** , Harvard Medical School, Boston, USA
University of Bremen, Bremen, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert Leech (robert.leech@kcl.ac.uk)

Author roles: **Cole JH:** Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Lorenz R:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Geranmayeh F:** Formal Analysis, Investigation, Methodology, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Wood T:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Hellyer P:** Formal Analysis, Funding Acquisition, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Williams S:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Turkheimer F:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Leech R:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust through a Wellcome Trust Seed Award in Science to JC [213996], and Sir Henry Wellcome Postdoctoral Fellowships to RLo [209139] and PH [106092]. JC is funded by a UKRI Innovation Fellowship. RLe is supported by an MRC Research grant [MR/R005370/1]. RLe and SW are supported by the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z] and the Data to Early Diagnosis and Precision Medicine Industrial Strategy Challenge Fund, UK Research and Innovation (UKRI). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2019 Cole JH *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cole JH, Lorenz R, Geranmayeh F *et al.* **Active Acquisition for multimodal neuroimaging [version 2; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2019, 3:145 (<https://doi.org/10.12688/wellcomeopenres.14918.2>)

First published: 14 Nov 2018, 3:145 (<https://doi.org/10.12688/wellcomeopenres.14918.1>)

REVISED Amendments from Version 1

The revised version includes substantial clarifications and revisions to the text, predominantly in the methods and discussion in response to the reviewers' comments.

Any further responses from the reviewers can be found at the end of the article

Introduction

Neuroimaging involves trade-offs; whether for clinical diagnosis, patient stratification, or biomarker discovery. For example, with a typical MRI scan, there are substantial practical constraints (money, patient comfort and compliance, radiological reporting) which means decisions have to be taken as to what kind of scan to perform, where in the brain scan, and the scan resolution. The standard approach is to make these decisions before scanning commences, acquiring the data then analysing it. However, the optimal resolution/type of scan will depend on what is being investigated, and the type and location of the pathology or abnormalities, and may not be known *a priori*.

Here, we propose an alternative approach using active learning for real-time optimisation of neuroimaging data acquisition; providing illustrative examples. Broadly, in our approach data acquisition and analysis are not separated; instead data is analysed as it is acquired and used to guide subsequent data acquisition, in a closed-loop. The word game *hangman* is a simple illustration of a form of active learning (as is predictive text messages and search engine auto-completion): a letter is guessed, and whether it is present or not is then evaluated; this information is then used to narrow the search for the next letter. Active learning approaches are potentially far more efficient (in terms of scanner time) than treating acquisition and analysis as separate phases. A non-active learning version of hangman would involve guessing all the letters at the start of the game and then evaluating them all at once without any feedback; in most situations, this would be a highly inefficient strategy.

We have previously demonstrated that active learning can be used to guide the choice of experimental paradigm in functional MRI (Lorenz *et al.*, 2016); with substantial increases in terms of speed, searching over many experimental parameters far quicker than an exhaustive search. This allows for far broader research questions to be asked (Lorenz *et al.*, 2018). Active learning also has another important feature; it involves a prediction and testing cycle, with the learner having to make predictions that are then tested with out-of-sample data. This potentially increases the replicability of analyses and reduces the ability for post-hoc bias (Lancaster *et al.*, 2018; Lorenz *et al.*, 2017).

The work presented here investigates the use of sequential decision-making to select the type of scan, using information gained from previous scans to actively seek out brain abnormalities or make diagnostic predictions. This requires data to be collected and analysed in near real-time; however, to illustrate the potential power of this approach our demonstrations use previously collected data, by simulating the real-time analysis aspect.

Video 1 presents a video overview of active acquisition: (i) scan parameters are chosen (e.g., modality or acquisition parameters such as resolution, repetition time or echo time); (ii) the scan is acquired; (iii) pre-processed; and, (iv) acquired data is compared to an existing normative dataset. The loop then continues with the information in (iv) used to optimise the next scan (or decide whether sufficient data have been collected to stop scanning). We explore using Active Acquisition in three different scenarios with T1-weighted MR images:

- 1) Finding a localised structural anomaly (e.g., locating a focal lesion).
- 2) Choosing the optimal scanning modality to actively detect abnormalities.
- 3) Actively choosing the type of scan to characterize an aspect of the individual being scanned (e.g., age).

Active acquisition for multimodal neuroimaging

1 Data File

Video 1: General illustrative video of one active acquisition approach for structural neuroimaging. (If video fails to play, it is also available in supplementary material or from the Github repository).

<https://dx.doi.org/10.6084/m9.figshare.7296920.v1>

Methods

Scenario 1: Changing structural scan resolution to detect stroke pathology

Our rationale is to start at a low image resolution for a (very rapid) whole brain scan, before acquiring higher resolution scans if the brain appears to be abnormal. This way, it is possible to efficiently image a focal pathology, such as a lesion or tumour, and to rapidly estimate its spatial location and establish whether more data needs to be acquired, potentially with a restricted field of view focused on the site of the abnormality.

Choice of scan parameters. For our illustrative simulation, we used structural scans collected offline (Dataset 1 (Leech & Cole, 2018)). Further details on the participants are included in the Data Acquisition section. In practice they would be acquired and analysed online. Practical challenges and limitations to acquiring these data, as well as consider possible methods to mitigate these challenges, as outlined in the Discussion section.

At each iteration (in terms of increasing resolution), the scan is divided into three equally sized volumes, along the z-dimension. The 'outlier distance' (defined below) is then quantified for each third by reference to the distribution in an independent normative sample (in this example, the n=7 healthy controls). The volume with the highest outlier distance is then selected and the next scan "acquired"; covering the same section of the volume, but with the resolution doubled. The process was repeated three times until the maximum resolution of 1mm³ voxel was achieved. The choice of resolution and number of sub-divisions (and other scanner parameters) presented in this scenario is relatively arbitrary. Future work will need to establish the optimal approach for a given clinical or scientific question. There will always be a

trade-off between multiple comparisons and precision when assessing; here, because the aim was to illustrate the logic and operation of the approach, we chose a very coarse approach which should be sufficient, given the focal and macroscopic nature of the brain injury (i.e., lesion). In clinical or scientific applications, a more sophisticated approach would probably be required, that chooses the brain region for the outlier detection (and potentially subsequent more targeted acquisition), related to the size and location of the pathology or abnormality, possibly changing orientation, and the image field-of-view in the process. In this vein, different potential decompositions of the multivariable imaging signal could be applied in parallel and evaluated in terms of outlier distance to normative data; subsequently, a decision could then be made (based on e.g., which is most likely an outlier) as to what decomposition to use to guide further data acquisition.

Outlier distance from normative sample. The extent to which a participant's image was different from the normative sample was quantified, restricted to the resolution and coverage of the specific scan. The median distance between an individual's scan and each participant in the normative dataset was calculated using the median absolute deviation (in Euclidean distance) of signal intensity averaged across all voxels. This results in a single value of outlier distance. The choice of outlier quantification depends on the type of data being acquired and the question being asked. We opted for the median absolute deviation because it is a simple measure that is relatively robust to violations of normality assumptions. However, we note that many other more sensitive outlier measures could also be used (e.g., measures taking into account covariance across voxels, (Fritsch *et al.*, 2012)).

Scenario 2: Active multimodal stratification of individual differences

In this scenario, Active Acquisition is used to choose the modality of the scan to achieve a given goal. The rationale is that the optimal scanning modality for assessing an individual, for example to quantify their relationship to a normative sample, will vary for different individuals; when performing a battery of scans, each individual may have a different set of scans and a different acquisition order.

In Scenario 2, we use multimodal imaging to quantify individual variability. This type of analysis could be relevant when classifying or stratifying individuals into scientifically or clinically relevant groups. To illustrate this, we use the [Cam-CAN dataset](#) (Shafto *et al.*, 2014) and with the task of predicting chronological age from neuroimaging data. Predicting age is a useful example case for active multimodal imaging because there are large datasets available, there is little ambiguity about label validity (unlike many clinical descriptions), age is associated with large-scale neural changes (e.g., (Good *et al.*, 2001) and "brain-predicted age" has been shown to relate to many other health related biomarkers (e.g., Cole *et al.*, 2018a). Cam-CAN is a particularly useful dataset to assess this source of individual variability since the age distribution of the participants is approximately equally balanced across seven decades from 20s to 80s. For details on the modalities please see the Data Acquisition section.

To instigate Active Acquisition in this case, we simulate the active learning process by fitting a decision tree regression model to the six modalities of Cam-CAN; predictions of chronological age were the outcome measure. This is because: a low-depth decision tree would not include all modalities, just those important for predicting age; making the decision sequentially (i.e., modality by modality) rather than simultaneously, thus it is well-suited for Active Acquisition, and finally; allows different individuals to have different scans and different orders of scans.

A holdout dataset was created with 20% of the individuals, selected randomly (the data partition was performed once rather than pooling across multiple, randomly generated partitions). A decision tree was fit to the remaining 80% of individuals' six imaging modalities as the predictor variables and their ages in years as the outcome variable. The model hyper-parameters (tree depth, number of leaves, etc.) were estimated with Bayesian optimisation (Leech & Cole, 2018). Subsequently, the decision tree was evaluated with the holdout participants. The application of the decision tree (the sequential decision process) to each individual in the holdout group, could be performed in real-time to new participants in exactly the same way. For comparison, we also fit a standard support vector regression, with hyper-parameters also optimised with Bayesian optimisation, to the same data (see Matlab code, (Leech & Cole, 2018)) which used all data modalities simultaneously.

Scenario 3: Active discovery of individual differences with multi-modal imaging

Whereas Scenario 2 focuses on quantifying how an individual varies along some dimension (e.g., age), in Scenario 3, we attempt to actively learn in which modality an individual is most likely to be an outlier (i.e., is a given individual more likely to be an outlier from the norm when using phenotypes derived from: T1-weighted MRI, diffusion-MRI or functional MRI?). This could be useful for efficiently finding pathology in an individual or for discovering biomarkers; particularly, when there are a large number of possible modalities to choose from and a limited amount of scanning time/participant tolerance of scanning (Leech & Cole, 2018).

To illustrate Scenario 3, we again used the Cam-CAN dataset, as per Scenario 2. In addition, we included a Bayesian optimisation algorithm (Shahriari *et al.*, 2016) to actively learn which modality is most abnormal (as quantified by the magnitude of outlier measurement). Bayesian optimisation is particularly well suited for this type of problem when the the general objective underlying function is unknown a priori and is costly to evaluate at any given point; Bayesian optimization is also relatively robust to the presence of noise in the data.

For optimisation to work efficiently, the acquisition function needs to take advantage of existing information; in this case the covariance across individuals for different modalities. Therefore, we split the data into two: 80% of the Cam-CAN participants were used to estimate the space (across modalities) for the algorithm to search across. To do this, we calculated a single summary statistic for each individual for each modality.

Then, we normalised the data within each modality so that each individual had a z-score for each modality. Subsequently, we performed a factor analysis (using Matlab) on the z-scored summary statistics resulting in a single factor. We then reorganized the modalities for the search space for the Bayesian optimisation in terms of weighting on the principle factor; this process estimates how different modalities will co-vary across individuals (approximately) with each other. For this example, with only six modalities to choose from, we opted for a simple experiment space with modalities given an integer between 1–6, based on the output of the factor analysis and the optimisation algorithm output integers. For more realistic situations with more complicated spaces (e.g., with many modalities organised along multiple dimensions and with more continuous modalities) one could use alternative (e.g., ratio) scales.

Subsequently, we performed Bayesian optimisation using the remaining 20% of participants, allowing the algorithm to pick the modality for a given individual, with the target objective of finding the minimum z-score. Given the relatively small number of available modalities, we allowed the algorithm to randomly choose three modalities (the burn-in phase) to sample first, to fit a Gaussian process regression and then to use the expected-improvement acquisition function to choose the next point to sample. The expectation was that after some initial random exploration, the model should be able to take advantage of the covariance across individuals to estimate the modality with the minimum z-score more frequently than expected by chance. Our choice of using the minimum z-score should be considered as illustrative and relatively arbitrary; we could equally have maximised the z-score or the absolute z-score. The actual target value would be based on the clinical/scientific question.

To assess whether this optimisation approach was performing above chance levels, we compared results for each individual with the correct factor ordering of modality (based on the covariance structure across individuals) with a random ordering of modalities. For each individual the order of the modalities was randomised (i.e., the ordering of the modalities was no longer based on prior information about how individuals co-vary across modalities). For both random and true covariance models, we calculated the proportion of participants where the optimisation algorithm correctly found the minimum z-score modality. This assessment process was repeated 100 times with different random seeds, allowing different burn-in sampling trajectories for each individual for each iteration.

Data acquisition

In Scenario 1, data were acquired from 13 participants: seven healthy controls with no history of neurological problems (average age= 56, range = 46 to 67, female=4); and, six patients with chronic left-hemisphere middle-cerebral artery focal strokes (average age= 60, range = 47 to 78, female=2, average lesion volume= 10.6 cm³). For each participant, three T1-weighted scans were acquired at different voxel resolutions: 1mm³, 2mm³, 4mm³. As with other data presented here, the patient and control data were not collected in real-time, but is intended to illustrate the general utility of the approach. In this example, we use seven healthy controls as a “normative” sample; this is,

obviously, far too small for actual practical uses, but was limited by the data set available (multiple resolutions per individual) but does illustrate the potential of the approach if scaled-up.

For Scenarios 2 and 3, multimodal MRI data from 611 people (age range from 18–88, 312, female) were taken from the Cam-CAN dataset. These data consisted of T1-weighted, T2-weighted, diffusion-weighted MRI, three functional scans (resting-state fMRI, movie-watching fMRI and a blocked sensorimotor task-based fMRI). Imaging acquisition has been presented in detail elsewhere (Shafto *et al.*, 2014). Only Cam-CAN participants with complete data from these six sequences were included (n=611, out of n=653).

Implementation

All analyses were performed with Matlab Version 2017b. Given the relatively low computational requirements of the analyses reported here, there are no additional minimum system requirements over and above those required to run this version of Matlab. Matlab code and associated data is available at Github (Leech & Cole, 2018).

Operation

Image pre-processing

Scenario 1:

To explore the feasibility of processing brain images in near real-time and to make minimal assumptions about the location or nature of pathology when calculating outlier distance, we used very simple and rapid pre-processing. T1-weighted images were converted from DICOM to NifTI format before being linearly-registered into MNI152 1mm³ space using the very efficient registration tool NiftyReg Version 1.3.9 (Modat, 2012). The same process was performed for each of the three different image resolutions acquired.

Scenario 2 & 3:

T1-weighted MRI. All T1-weighted structural images from all three datasets were processed in the same way as follows. Grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) volumes were calculated using SPM12 ‘Segment’ (University College London, UK). Voxelwise assessment of changes to brain volume was calculated using the SPM symmetric diffeomorphic registration process (Ashburner, 2007) to a predefined template used in our previous studies (Cole *et al.*, 2018b).

For the Cam-CAN dataset, the other modalities were (briefly) processed as detailed below. These analyses are merely illustrative of the type of data that could be extracted; they have been simplified from multivariate raw data for each individual into a single summary statistic, chosen for its simplicity rather than because it is optimal for measuring individual variability.

Diffusion-weighted MRI. White-matter microstructure was analysed under the diffusion tensor imaging (DTI) paradigm, using FSL Version 5.010 tract-based spatial statistics (Smith *et al.*, 2006) with DTI-TK (Zhang *et al.*, 2007) software for affine then non-linear tensor-based image registration. Normalised tensor images were used to derive voxelwise measures of fractional anisotropy (FA) and mean diffusivity. Mean values across major

WM fibre tracts, taken from the JHU-ICBM tract atlas, were calculated, resulting in an FA average value per individual.

T2-weighted MRI. The same diffeomorphic transformation that was calculated for the grey matter was applied to the T2-weighted scan data to warp each individual's data into the same T1-weighted template space. Subsequently, the average T2-weighted intensity values from the normalised image was calculated.

Resting state functional connectivity. Measures of 'within-network' connectivity were calculated from resting-state fMRI data using FSL 'Dual Regression' (Filippini *et al.*, 2009). Prior to the dual regression, the standard FSL 'MELODIC' analysis pipeline was applied (Smith *et al.*, 2004): high-pass temporal filtering at 100s, spatial smoothing at 5mm FWHM, global intensity normalisation, motion-correction followed by realigning the data into MNI152 space using linear registration before the data were resampled into 4x4x4mm voxel space. Then the data were cleaned by linearly regressing six motion parameters from each voxel's time-course, before nuisance WM and CSF time-courses were linearly regressed from each voxel (using average CSF and WM masks from the segmentation). Subsequently, using canonical spatial maps of twenty networks (including both intrinsic connectivity networks and likely noise networks) (Smith *et al.*, 2009), cleaned data underwent a multiple regression to derive voxelwise measures of connectivity for each network for each individual. Finally, to keep this aspect of the approach as simple as possible, we averaged all voxels within the default mode network (DMN) mask; this process resulted in an individualized 'within-network' connectivity measure for the DMN. Future work (with any fMRI data) could explore only using short segments of the functional time-series (rather than the whole scan), to allow for faster, repeated measurements.

Movie-watching functional connectivity. This was identical to the analysis of the resting state connectivity, calculating individualised within-DMN functional connectivity while watching the movie.

Task fMRI. The sensorimotor task data were analysed following a standard FSL pipeline: global intensity normalisation, high-pass temporal filtering at 100s, spatial smoothing at 5mm FWHM, motion-correction, registration of the data into MNI152 space using linear registration. Subsequently, a general linear model was applied voxelwise (using the standard FSL approach for dealing with the auto-correlation of residuals (Smith *et al.*, 2004)), with separate explanatory variables modelling auditory and visual blocks convolved with a canonical hemodynamic response function. Subsequently, a contrast of all task conditions versus the implicit baseline was calculated and a higher-level group mixed-effects model was used to calculate increased and decreased BOLD activity with task. This resulted in group task positive and task negative networks which were converted into binary mask defined by voxels that survived cluster correction for multiple comparisons. An individualised task fMRI measure was calculated by taking the average activity within the positive network mask and subtracting the average value from the negative network mask.

Results

Scenario 1: Changing structural scan resolution to detect stroke pathology

The simplest Active Acquisition model involves starting with a rapid, low resolution structural scan, analysing it and then deciding to whether to acquire further higher-resolution scan(s). Here, we collected lower-resolution (4mm³) structural scans from six patients with focal brain lesions and seven age-matched controls, followed by intermediate-resolution (2mm³) and higher-resolution (1mm³) scans. An illustrative patient at three resolutions is presented in Figure 1 (left). Even with the lowest resolution scan, patients and control participants (Figure 1 - right) show a large difference in terms of outlier distance. This example, in patients with large focal strokes, illustrates how data simple measures calculated in near real time, and then a decision made as to whether a slower, higher resolution scan is needed or not. As can be seen from the outlier measurements, only a subset of the control participants, close to the boundary with the patients would require slower, additional scans.

We also simulated optimising the scan field-of-view in near-real-time. In this case, at each resolution the brain is divided into thirds, and the negative outlier distance calculated for each third. The third that is most strongly classed as an outlier is then retained and subsequent, higher resolution scans, acquired just within that third. The process then repeats (Figure 2, top). This illustrates how a composite brain image can be built up out of increasing resolution scans. This could trade-off sensitivity for tissue contrast with increasing quantification of brain structure, while limiting scanning time.

Scenario 2: Active multimodal stratification of individual differences

When fitting the decision tree regression (Figure 3) to predict chronological age from neuroimaging data, the regression model contained multiple modalities (indicating its utility in a sequential acquisition and analysis procedure). It started with GM volume, consistent with previous data suggesting a strong relationship between GM and age (Good *et al.*, 2001), with lower z-scores indicating older age. Subsequently, average WM FA was chosen, again with lower values relating to older age. Next, the model's branches become very different, both in terms of modality chosen and number of scans required, depending on the route through the tree. By way of individual examples, if a participant had a GM z-score = -0.8 and FA z-score = -0.7, then they would have a predicted age of 77.6 years (following the left most branches of the decision tree in Figure 3). However, if a participant also has GM z-score = -0.8, but their FA z-score = 0.5, then task BOLD data would be necessary to make a prediction. If that individual's task BOLD z-score \geq -0.18 then their predicted age would be 55.7 years, whereas a task BOLD z-score $<$ -0.18 would give a predicted age of 63.1 years.

We observed that the mean absolute error (MAE) of age prediction is 10.47 years and the median error 8 years. For comparison, the MAE calculated on the same data using a support vector regression approach with all of the data is very similar, was 10.42 years, with a median error of 9.4 years. The predicted

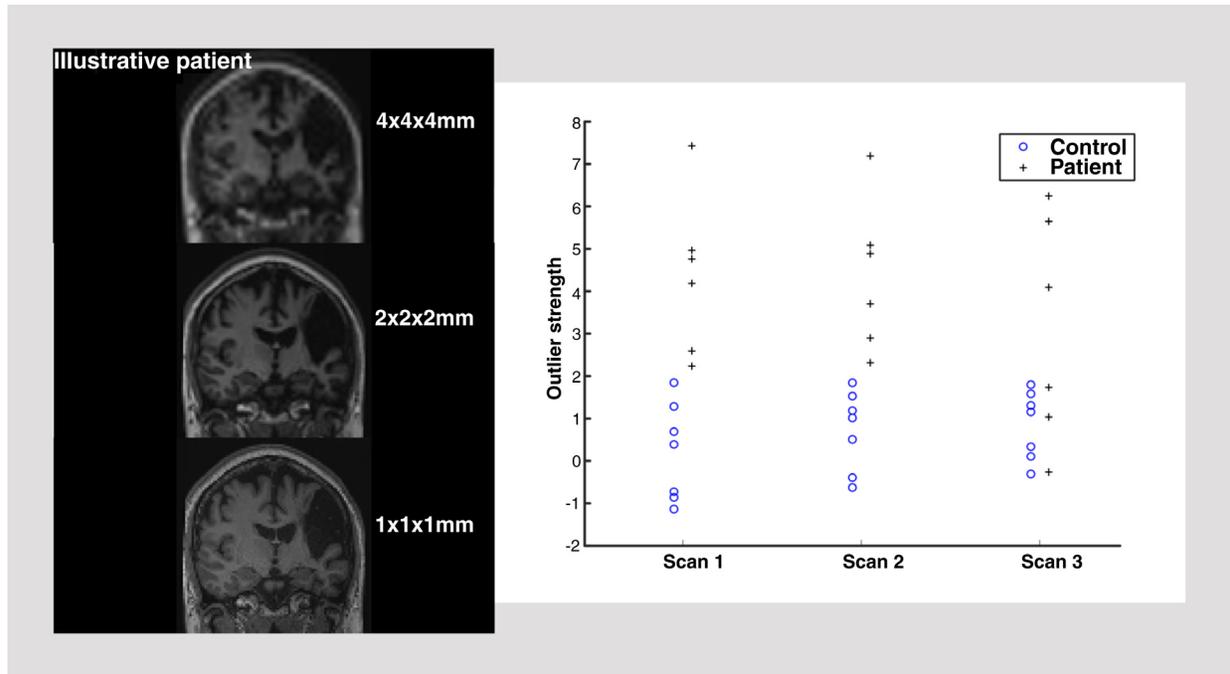


Figure 1. Left, a stroke patient with three different resolution T1-weighted scans. Right, outlier distance from control participants, for each participant for the three different scans, and combining all three scans. For each scan, the scan is subdivided into three, and the maximum outlier distance (out of the three subdivisions assessed) from the control data is plotted. This shows a relatively clear difference in outlier distance between patients and controls. For most patients and controls (either far from 0 or close to 0 respectively), there is no need to collect additional higher resolution (slower) scans to differentiate the two groups.

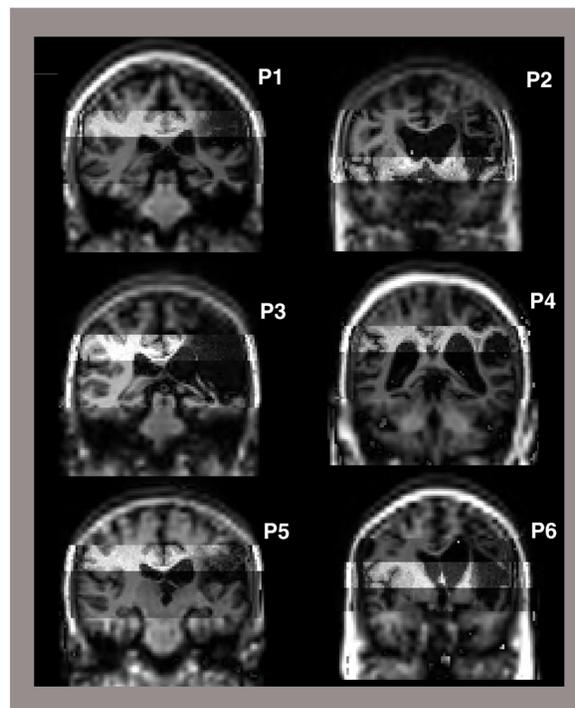


Figure 2. top, composite T1-weighted coronal slices (one for each of the six patients), comprise out three scans with increasing resolution, each with different coverage of the field-of-view (less field-of-view, higher resolution). The reduced field-of-view is centred on the slice from the previous image that is quantified as most abnormal relative to the norm (i.e., the control group). This demonstrates that the very simple approach to subdividing the brain and quantifying outliers can be used to ‘zoom’ in on areas of pathology that are specific for individual patients.

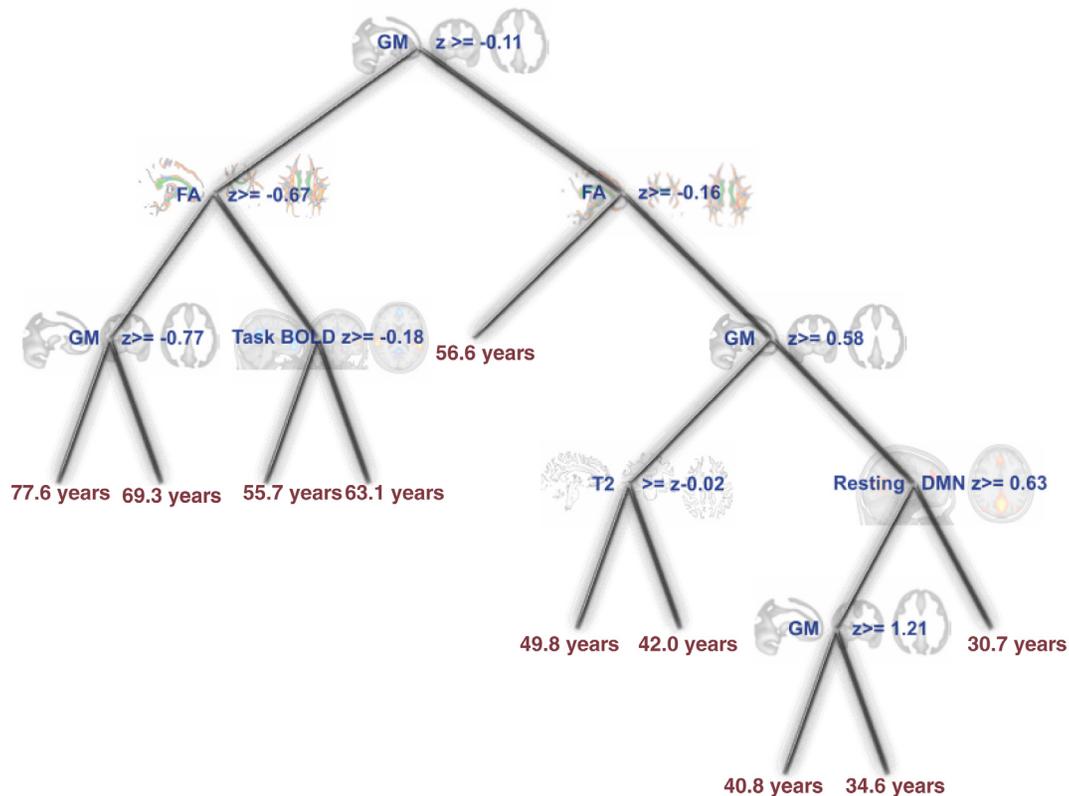


Figure 3. The decision tree regression model calculated on summary statistics for each of six modalities to predict individual age. At each node in the tree the z-scored data for a given individual are used to decide modality to use next or whether to stop at this point. This can happen in near-real time, with different individuals taking different routes through the tree, and with different numbers of scans. The estimated age is then approximated by the age at the leaf nodes.

age performance is considerably worse than has been reported elsewhere for single modalities from the same dataset e.g., (Lancaster *et al.*, 2018); this is to be expected given that, for illustrative simplicity, we have collapsed large, multivariable datasets into single summary statistics (i.e., a single value for grey matter probability per individual etcetera). In practice, sequential decision methods incorporating multivariate datasets to utilise the full richness of the underlying data are needed to realise the potential of the approach.

Scenario 3: Active discovery of individual differences with multi-modal imaging

Scenario 3 used a normal distribution (from a normative dataset) to derive individual z-scores for each MRI modality, with the goal of identifying for each individual in which modality they are most outlying (i.e., which modality has the highest absolute z-score for that person). To achieve this we simulated closed-loop Bayesian optimisation to identify the highest z-score from across all modalities for a given individual (from the holdout dataset), as shown in Figure 4. For individual depicted in Figure 4A, the highest z-score was in T2-weighted MRI ($z = -1.2$), followed by resting-state fMRI, sensorimotor task BOLD, FA, movie BOLD and then GM. This suggests that for this individual, the T2-weighted is likely

to be most valuable in determining whether or not they are experiencing some underlying pathology. While this approach can be used to rank z-scores across modalities, the magnitude of the z-score can also be informative, to provide insight into how much of an outlier this individual is in that modality (which may or may not indicate the presence of a pathology).

For the Bayesian optimisation to work efficiently (i.e., faster than exhaustive search across modalities), it needs to take advantage of covariance across modalities in individual differences. In Figure 4A, the order that the optimisation procedure tests the modalities (Movie, resting-state fMRI, task BOLD, T2, FA, GM) reflects this covariance structure. This provides prior information that the optimisation algorithm can combine with some random initial samples (numbers 1–3 in Figure 4A, left) to build a Gaussian process regression model to predict the modality with minimum z-score (in this case number 4, T2).

By chance, the proportion of participants for whom the algorithm finds the modality with the minimum z-score is 0.67 (given that it sampled four modalities in total). When the Bayesian optimisation algorithm utilises the estimated covariance structure from the training dataset, the proportion increases to >0.72 on average (results in Figure 4B are presented from

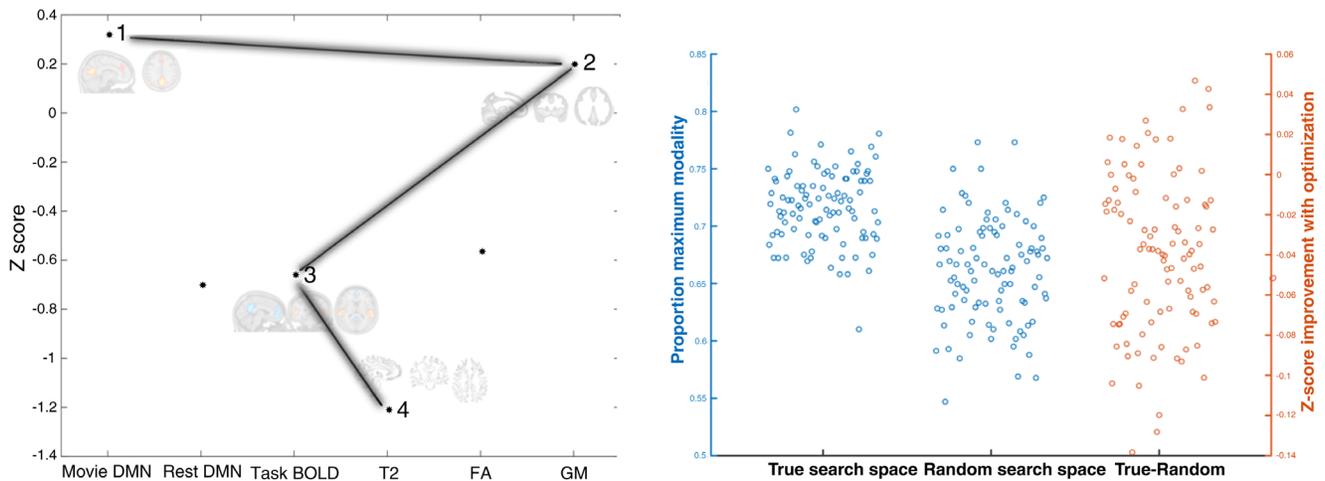


Figure 4. Active discovery of individual differences across modalities, controlled by a closed-loop optimisation algorithm. Left, the trajectory of the algorithm as it traverses the modality space, estimating a model of which modalities a specific individual appears most abnormal in, without exhaustively sampling every point, before guessing which is the most abnormal. Right, proportion of participants in the holdout set where the optimisation algorithm correctly chose the modality most sensitive to abnormalities for both true and random modalities, and the decrease in estimated minimum z-statistic for true versus random organisation of modalities (repeated 100 times with different burn-in random initialisation of the models).

100 replications). We see that if the modality ordering is chosen randomly (rather than based on covariance across individuals from the training set) the average proportion of participants where the minimum modality is selected approximates that expected by chance (i.e., 0.67). We also see that this translates into an increase in the estimated minimum z-score found when using the optimisation algorithm compared to the random modality ordering. This difference between true and random ordering of modality search space is relatively modest (approximately 5%). However, the dataset used in this example has very few modalities and thus a restricted search space and has a relatively limited sample size. Also, we used somewhat coarse pre-processing and summary statistics. Applying this approach to on-going data collection in much larger projects or at clinical neuroimaging centres that scan large numbers of people, alongside the myriad of different MRI scan modalities available, means that this approach could be substantially improved and used much more powerfully for biomarker discovery.

Discussion

Here, we have outlined the Active Acquisition approach for optimising multimodal neuroimaging scan protocols. The analytical examples are intended to illustrate the potential utility of Active Acquisition; by using this approach important decisions about the scan do not need to be in advance; how long to scan for, what modalities to acquire, which regions of the brain to focus on. Rather, the precise nature of the scanning protocol is determined online, adapting to the individual in the scanner, optimising acquisition for a given set of circumstances. Our current goal has been to outline several broad scenarios that suggest how Active Acquisition could progress and its general potential, rather than provide evidence of a specific biomarker or indeed specific pipelines or analysis approaches. Here, we discuss

future potential directions for Active Acquisition, in particular for diagnosis and stratification as well as for biomarker discovery. We envisage these two directions developing along independent but complementary lines. We also consider some practical issues that need to be overcome to take the approach forward and maximise its potential for clinical and scientific neuroimaging.

Clinical diagnosis

Perhaps the more obvious use case for Active Acquisition is in clinical diagnostics, and the stratification of individuals into subgroups. Incorporating Active Acquisition could lead to either shorter scanning sessions, or more accurate and more reliable data collection. Multiple imaging modalities are typically collected in a diagnostic clinical scanning session, many of which end up being unnecessary for accurate diagnosis. If the scanning session can be terminated early, when sufficient diagnostic certainty has been reached (as in Scenario 1), there would be a significant reduction in scanning time, reducing patient discomfort and scanning costs. Equally, by optimising the order of the scans (as in Scenario 2), tailored to the targeted disorder, this would potentially remove the need to collect all modalities, leading to the same benefits in terms of time, cost and patient comfort.

Alternatively, Active Acquisition could be used to produce more accurate diagnoses and to optimise certain modalities for clinical use that are currently not used in clinical settings. Active acquisition could make use of scanning time and resources more efficiently; collecting repetitions of important scans (until a sufficient signal-to-noise ratio has been reached), or changing the scanning resolution or field-of-view to focus on potential abnormalities. This may be of particular use in relatively low

signal-to-noise imaging modalities. For example, the pattern of brain damage presented Scenario 1 (focal ischaemic stroke) is evident even on very low resolution and low signal-to-noise structural scans; however, other neurological conditions may have far more subtle abnormalities and other modalities (e.g., arterial spin labelling, diffusion tensor imaging, resting state or task BOLD scans) have lower signal-to-noise, and may benefit from more spatially focused, repeated data acquisition.

A pertinent issue facing neuroimaging research in clinical samples is how to deal with heterogeneity within patient groups; particularly common in chronic neuropsychiatric diseases. The “average” best scanning protocol sequence may well not be optimal at identifying clinically relevant abnormalities in a specific individual. Potentially, different scans may be optimal for a given diagnosis in different individuals and at different points in the natural history of a disease. One major strength of active acquisition approaches is that they can more easily locate an individual patient’s “sweet-spot” from a large menu of possible scan types/parameters in a time-efficient manner, without having to exhaustively search through all possibilities.

Biomarker discovery

Finding biomarkers that sensitively detect individual variability linked to clinical and scientific questions is an important precursor to improving diagnosis and stratification. The application of active acquisition illustrated in Scenario 3 presents a radically different way to achieve this: actively searching for modalities or scanning parameters give abnormal readouts for a single individual. This approach contrasts with the current typical approach to biomarker discovery which can be characterised as choosing a set of modalities prior to scanning that are thought to be related to the clinical question, and then assessing them on a large group of patients and controls or subgroups of patients, to provide sufficient statistical power to detect average group differences. Active acquisition also has the benefit of attempting to focus on modalities only when they are likely to be abnormal for an individual relative to a normative dataset, which is potentially much more powerful than the comparison of group averages, as well as leading intuitively to clinical applications of personalised medicine. Active acquisition also has the advantage of relying less on relatively arbitrary decisions that lead to a limited number of modalities being acquired, which means that the clinically-relevant sweet-spot for data acquisition is more likely to be found.

Active Acquisition could also avoid the potential problem of scanning protocols being determined based on biased or inaccurate previous studies. Given the replication ‘crisis’ in biomedical research, such issues are becoming increasingly recognised as a serious problem in medical imaging. Active optimisation approaches (such as in Scenario 3) involve repeatedly cycling between prediction and hypothesis testing on out-of-sample data, and as such are less susceptible to data overfitting. Equally, active optimisation approaches like these also involve a form of implicit “pre-registration” (Lorenz *et al.*, 2017). This makes it harder to engage in certain questionable research practices (e.g., p-hacking, post-hoc hypothesising (Poldrack *et al.*, 2017)) that are currently thought to hamper the development of neuroimaging biomarkers.

One additional advantage of active optimisation is that it is able to estimate how an individual varies from normality across the whole of the search space, despite only sampling a subset of the modalities tested included in the space. While the gains observed were relatively minor in the current example, where only six modalities organised along one searchable dimension were considered, the potential benefit would grow as the space becomes larger and multidimensional. Using the optimisation algorithm to map out the entire possible space offers the potential for a very rich, but efficiently collected, description of how an individual differs from normality. The search space mapped out could involve observing multiple optima in a given individual and estimating modalities with higher and lower than typical signal. Subsequent offline higher-level modelling (e.g., clustering or other data reduction approaches) could then be applied across individuals to find frequent patterns of abnormality from across all modalities.

Need for different types of normative datasets

One major limiting step to the development of active acquisition is the need to have well-characterised variability across individuals in both healthy or ‘normal’ participants as well as clinical samples and relevant subgroups. Achieving this will require developing large and representative datasets from which to derive estimates of between-individual covariance. Currently for Scenario 1, the normative dataset is simply the n=7 healthy controls; in reality the size of normative cohorts will have to be much larger.

Some simpler applications of Active Acquisition could be built with existing normative datasets. For example, when the problem involves deciding when to stop collecting more data because a sufficient signal-to-noise ratio has been reached, increasing confidence in the inferences made from these data. Other approaches could take advantages of new acquisition methods such as the very rapid multi-contrast images at the start of a scan (Skare *et al.*, 2018) or synthetic imaging which are then used to decide whether to collect slower, higher resolution scans. To utilise these types of scans, existing datasets could be utilised to create sufficiently large normative models.

However, for other applications, such as when searching across modalities (Scenarios 2 and 3), the benefits of Active Acquisition may be most evident when the space of possible modalities/parameters to be considered is large but structured in some way. Indeed, while at present only a small number of imaging modalities are employed clinically, more modalities could be useful but only for stratifying specific subgroups. An accurate understanding of the covariance between modalities/scan parameters relevant to the clinical or scientific question will be necessary for maximising the benefit from these approaches

In Scenario 3, where the optimisation algorithm maps out where an individual is maximally abnormal, understanding the covariance across imaging modalities in a healthy control group (possibly controlling for factors such as age) may suffice. Existing large-scale projects to produce large normative databases have focused on small numbers of modalities collected in large numbers of people (e.g., UK Biobank (Sudlow *et al.*, 2015),

Human Connectome Project (Van Essen *et al.*, 2013) and the Cam-CAN dataset presented here). One possible approach is to use meta-analyses of different imaging modalities to try to estimate covariance structure across modalities (capitalising on the fact that different large-scale projects have some shared modalities but also differ from each other). An even better approach would be to have large-scale data collection projects that explicitly seek to quantify covariance across many different imaging modalities/scan parameters. Ideally, this would involve many different representative individuals being scanned, but each with different subsets of modalities/scan parameters; subsequently, a large, comprehensive covariance matrix across individuals can be assembled out of the incomplete datasets from each individual. These normative datasets will allow active searching for how individual patients vary from normality across many modalities, useful for biomarker discovery, without requiring dedicated large multimodal datasets for each clinical condition. Approaches such as Bayesian optimisation with Gaussian processes will allow us to start with relatively few assumptions (i.e., only approximate similarity across modalities near each other in the experimental search space which can be based on health control data); importantly, the approach should work for individuals even when there are areas of the experimental space that deviate from the normative data.

There are also likely to be some situations, however, where acquiring targeted multi-modal normative datasets for specific clinical conditions will also be important. For example, when performing diagnostics rather than discovery of biomarkers (more like in Scenario 2). In these situations, bespoke multi-modal datasets may be necessary to arrive at a very specific quantification of the covariance between different modalities, in order to accurately guide the sequential decision making. In such situations, particularly, with rare disease groups, acquisition of such datasets would be far more challenging and may not be practical.

At present, it is unknown how much benefit we can derive by collecting additional normative databases covering a wider array of scan types. The potential benefit depends on whether or not types of scans not already collected in existing large normative samples capture clinically or scientifically useful variability and the potential benefit of that clinical and scientific value. Equally, it depends on the scientific or clinical question, and whether we already know the optimal scans for assessing individual variability. To assess the potential benefit, one approach is to start with relatively small-scale normative data collection with many types of scans in the same individuals; analysis of covariance between scans will allow us to estimate individual variability not captured by the small number of modalities typically collected. Consequently, we will be able to estimate how much benefit can be gained by individually-tailoring scan sequences compared to acquiring the same small subset of scans on everybody.

Methodological considerations

All methodological approaches come with costs and benefits; with Active Acquisition approaches one concern is that early mismeasurement can lead to serious failures later on. For example, in Scenario 1, this could result in terminating scans

prematurely without collecting sufficient data; or, e.g., in Scenario 2, this could involve travelling down the wrong branch of the decision tree. In such situations, important information for diagnosis or biomarker discovery may not be collected. This cost of using active approaches will be most acute when the underlying covariation between scan modalities is well understood and the optimal scan type is known. In contrast, the way that we currently collect data in many exploratory studies (e.g., UK Biobank), it is likely that optimal scans for assessing variability in an individual are being omitted. This reflects the classic exploration versus exploitation trade off well-known in computer science. There is a potential risk when designing adaptive experiments that the acquisition function is too exploitative and acquired data will fail to be broad enough to allow for future serendipitous discovery about unrelated scientific or clinical questions. The choice of an appropriate exploratory acquisition function guiding data collection has the potential to balance between efficient imaging while also estimating individual variability across a larger space of different types of scans, that may be relevant to other questions or future studies.

For the benefits of active exploration to be maximised, many choices have to be made regarding the acquisition function to guide exploration, how to decide when to stop searching, how to quantify abnormality or predict an individual's classification. We have suggested several simple illustrative scenarios, but each comes with its own specific challenges and future directions. There is a long history of methodological developments for adaptive studies in clinical situations (Cornfield *et al.*, 1969) (Hauskrecht & Fraser, 2000) (Alagoz *et al.*, 2010). Future work is needed to incorporate some of the more sophisticated approaches developed in these other domains to neuroimaging and ideally combine them multivariate classification and clustering approaches increasingly commonly used with MRI. In Scenarios 2 and 3, work is needed to understand what happens when there is not a single optimum modality to maximally quantify abnormality (Scenario 3) or multiple equally good paths through the decision tree (Scenario 2).

Future work is also needed to evaluate how to robustly quantify the abnormality of an individual's scan, considering the large number of voxels and possibly heterogeneous or diffuse pathologies. The examples presented were designed to clearly illustrate different adaptive approaches rather than show state-of-the-art outlier detection or age classification. As such, performance at age classification in Scenario 2 was based on whole-scan based on summary statistics and a transparent decision tree technique; therefore, performance is expected to be considerably inferior to more sophisticated voxelwise approaches. Equally, in Scenario 3, we chose to only update the optimization with a single summary statistic, but multiple, complementary measures from a single scan could be calculated in parallel and used to update multiple points in the search space simultaneously. More generally, there are a whole range of future potential avenues for developing and applying more sophisticated, but less transparent approaches to be explored in future work.

Another major consideration is the inherent trade-offs between how long the analysis takes set against the potential benefits of the adaptive approach and potential time savings. Current

image analysis pipelines are often very slow and potentially costly in terms of computational processing power, making near real-time analysis infeasible. For example, the image analysis in Scenario 1 would be possible in near real time, (full pre-processing and analysis pipeline using a standard, quad-core personal computer in < 1 minute); however, for Scenarios 2 and 3, data was processed by a high-performance compute cluster with some of the pipelines (e.g., fitting tensor models to calculate fractional isotropy or non-linear image registration) being far slower than would be feasible for active acquisition. These timing challenges have to be offset against any potential gains from adaptive approaches. To address these challenges will require the use of parallelization and dedicated computational hardware that could be used to substantially improve speed as well as optimized pipelines, maximising speed; this will bring processing time down to a fraction of the current time (as has been achieved with e.g., fMRI for brain computer interfaces) with minimal loss of image quality. Finally, recent developments in deep learning offer considerable promise; deep learning approaches, that are slow and costly to train, requiring large datasets, but are very fast to apply. Existing work suggests that many of the more time-consuming steps of pre-processing could be accomplished in near real time using these approaches. For example, structural MR image can undergo an analogue of a complete pre-processing pipeline in a matter of seconds (Cole *et al.*, 2017), making near real-time applications applied to multimodal imaging practical.

Finally, from an MR physics perspective, there are also a number of limitations and challenges. Actively altering the field-of-view and resolution (as suggested in Scenario 1 where the scan zooms in on the site of injury) for 3D structural imaging may not have any benefits (in terms of time saved, increased resolution) given inherent trade-offs between tissue contrast, signal to noise and number of measurements acquired. However, a similar approach could be taken with other imaging modalities (e.g., arterial spin labelling, diffusion imaging) where increased signal to noise from restricting the number of slices or increasing the resolution may be beneficial. Equally, there may be different sources of information that different resolutions and fields-of-view could acquire (e.g., rapidly assessing geometry at higher resolution and tissue contrasts at a lower resolution).

In summary, here we have presented Active Acquisition, a novel conceptual approach to how neuroimaging data could be collected. We have utilised advances in optimisation algorithms and harnessed large publicly-available neuroimaging databases

to develop Active Acquisition. This approach embeds data analysis into the acquisition process, allowing information to be obtained and employed for making online decisions about the optimal scans or parameters for a given clinical or scientific goal. While Active Acquisition is still at the embryonic stage, our intention with this manuscript and the illustrative examples contained herein, is to provide the groundwork for future conceptual and experimental work aimed at optimising the acquisition of neuroimaging data for clinical and scientific purposes.

Software availability

Underlying code used to perform this method is available from GitHub: <https://github.com/ActiveNeuroImaging/MultimodalActiveAcquisition>.

Archived source code at time of publication: <http://doi.org/10.5281/zenodo.1478784> (Leech & Cole, 2018)

Licence: MIT

Data availability

The anonymised and pre-processed data used in all scenarios is available from the same Github repository as the code. The MRI data has been provided in an anonymised format and registered to an average template space.

GitHub: <https://github.com/ActiveNeuroImaging/MultimodalActiveAcquisition>

Zenodo. Dataset 1: ActiveNeuroImaging/MultimodalActiveAcquisition: v1.0 <http://doi.org/10.5281/zenodo.1478784> (Leech & Cole, 2018)

This dataset is available under a MIT license.

The original data used in Scenarios 2 and 3 was from the Cambridge Centre for Ageing and Neuroscience (CamCAN) project dataset: <https://camcan-archive.mrc-cbu.cam.ac.uk/dataaccess/>

Media

Video 1: General illustrative video of one active acquisition approach for structural neuroimaging. (If video fails to play, it is also available from the Github repository).

<https://doi.org/10.6084/m9.figshare.7296920.v1> (Cole *et al.*, 2018)

Acknowledgements

Thanks to Dr František Váša for comments on the paper.

References

Alagoz O, Hsu H, Schaefer AJ, *et al.*: **Markov decision processes: a tool for sequential decision making under uncertainty.** *Med Decis Making.* 2010; **30**(4): 474–83.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ashburner J: **A fast diffeomorphic image registration algorithm.** *NeuroImage.* 2007; **38**(1): 95–113.
[PubMed Abstract](#) | [Publisher Full Text](#)

Cole J, Leech R, Lorenz R, *et al.*: **Video 1. General illustrative video of one**

- active acquisition approach for structural neuroimaging. *Wellcome Open Research. Media*. 2018.
- Cole JH, Poudel RPK, Tsagkrasoulis D, *et al.*: Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*. 2017; **163**: 115–124.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cole JH, Ritchie SJ, Bastin ME, *et al.*: Brain age predicts mortality. *Mol Psychiatry*. 2018a; **23**(5): 1385–92.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cole JH, Caan MWA, Underwood J, *et al.*: No Evidence for Accelerated Aging-Related Brain Pathology in Treated Human Immunodeficiency Virus: Longitudinal Neuroimaging Results From the Comorbidity in Relation to AIDS (COBRA) Project. *Clin Infect Dis*. 2018b; **66**(12): 1899–1909.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cornfield J, Halperin M, Greenhouse SW: An Adaptive Procedure for Sequential Clinical Trials. *J Am Stat Assoc*. 1969; **64**(327): 759–70.
[Publisher Full Text](#)
- Filippini N, MacIntosh BJ, Hough MG, *et al.*: Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. *Proc Natl Acad Sci U S A*. 2009; **106**(17): 7209–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fritsch V, Varoquaux G, Thyreau B, *et al.*: Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators. *Med Image Anal. Special Issue on the 2011 Conference on Medical Image Computing and Computer Assisted Intervention*. 2012; **16**(7): 1359–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Good CD, Johnsrude IS, Ashburner J, *et al.*: A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*. 2001; **14**(1 Pt 1): 21–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hauskrecht M, Fraser H: Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artif Intell Med*. 2000; **18**(3): 221–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lancaster J, Lorenz R, Leech R, *et al.*: Bayesian Optimization for Neuroimaging Pre-processing in Brain Age Classification and Prediction. *Front Aging Neurosci*. 2018; **10**: 28.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leech R, Cole JH: Multimodal Active Acquisition (version 1.0). 2018.
<http://www.doi.org/10.5281/zenodo.1478783>
- Lorenz R, Hampshire A, Leech R: Neuroadaptive Bayesian Optimization and Hypothesis Testing. *Trends Cogn Sci*. 2017; **21**(3): 155–67.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lorenz R, Monti RP, Violante IR, *et al.*: The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *NeuroImage*. 2016; **129**: 320–34.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lorenz R, Violante IR, Monti RP, *et al.*: Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. *Nat Commun*. 2018; **9**(1): 1227.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Modat M: Efficient Dense Non-Rigid Registration Using the Free-Form Deformation Framework. UCL (University College London). 2012.
[Reference Source](#)
- Poldrack RA, Baker CI, Durnez J, *et al.*: Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat Rev Neurosci*. 2017; **18**(2): 115–126.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shafto MA, Tyler LK, Dixon M, *et al.*: The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurol*. 2014; **14**: 204.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shahriari B, Swersky K, Wang Z, *et al.*: Taking the Human out of the Loop: A Review of Bayesian Optimization. *Proc IEEE*. 2016; **104**(1): 148–75.
[Publisher Full Text](#)
- Skare S, Sprenger T, Norbeck O, *et al.*: A 1-minute full brain MR exam using a multicontrast EPI sequence. *Magn Reson Med*. 2018; **79**(6): 3045–54.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith SM, Fox PT, Miller KL, *et al.*: Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A*. 2009; **106**(31): 13040–45.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith SM, Jenkinson M, Johansen-Berg H, *et al.*: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*. 2006; **31**(4): 1487–1505.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith SM, Jenkinson M, Woolrich MW, *et al.*: Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004; **23** Suppl 1: S208–19.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sudlow C, Gallacher J, Allen N, *et al.*: UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015; **12**(3): e1001779.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Essen DC, Smith SM, Barch DM, *et al.*: The WU-Minn Human Connectome Project: an overview. *NeuroImage*. 2013; **80**: 62–79.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhang H, Avants BB, Yushkevich PA, *et al.*: High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *IEEE Trans Med Imaging*. 2007; **26**(11): 1585–97.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:



Version 2

Reviewer Report 21 October 2019

<https://doi.org/10.21956/wellcomeopenres.16937.r36546>

© 2019 Kikinis R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ron Kikinis 

¹ Surgical Planning Laboratory, Harvard Medical School, Boston, MA, USA

² University of Bremen, Bremen, Germany

I am satisfied with the changes.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medical image computing.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 14 March 2019

<https://doi.org/10.21956/wellcomeopenres.16265.r34521>

© 2019 Kikinis R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ron Kikinis 

¹ Surgical Planning Laboratory, Harvard Medical School, Boston, MA, USA

² University of Bremen, Bremen, Germany

This is a paper about a method called active acquisition. The authors propose a framework where neuroimaging sequence selection is iteratively optimized while the patient is on the table using an active

learning approach. The method is currently a simulation. At this point the paper is an outline of a research strategy not a presentation of a method that is actually working.

In practical terms, the speed of the computation will be critical. If the computation increases the time that the patient spends in the scanner, this will result in resistance to using this approach.

My main concern is that the three scenarios are spread widely and mostly talk about the concept. In my opinion the authors should limit themselves to one scenario and flesh it out with enough detail that the work could be duplicated.

Introduction:

- “Active learning also has another important feature; they involve a prediction and testing cycle..” they or it?
- “using information gained from previous scans actively seek out brain abnormalities or make diagnostic predictions...” should read “to actively seek..”

Methods:

Scenario 1: detection of stroke pathology

Choice of parameters:

- How about a short summary of Dataset 1? The paper should be self-contained.
- “At each iteration the scan...” which scan? Is it at full resolution or, as implied in the introduction, at reduced resolution? How much reduction?
- What is a normative sample and where is it coming from?

Outlier distance:

- When you say median distance, what do you mean? Distance in voxels, signal intensities, something else?

Scenario 2: choose a modality

- Again, provide a short description of the data set, so the reader does not have to visit a different webpage just to know what data you are using. Also, which modalities are you planning to use in the simulation and why? I went to the Cam-CAN webpage and modalities include task based fMRI and MEG. It is not clear to me how such acquisitions could be performed dynamically at need.

Scenario 3: discovery of individual differences.

- “we attempt to actively learn which modality an individual is most likely to be an outlier in” - Please reword.
- The section data acquisition provides the details that would have been helpful to have when reading the scenarios. There were no cross-references in the previous sections. The scenario descriptions 2 and 3 do not mention that only a subset of the Cam-CAN data is being used.

Results:

Scenario 1: Figure 2 is not clear to me. What do you show when you mean composite?

Scenario 2: It would be helpful to have two examples: a young and an old person and how their trajectory in the regression is different.

Scenario 3: I do not understand what the results are. “Here we simulated closed-loop Bayesian optimisation used to discover the modality for a given individual (from the holdout dataset) where the negative outlier distance is most (i.e., relative to normative data from the training dataset), shown in Figure 4.”

Discussion:

- The authors should begin the discussion of the results in the three scenarios and how they relate to existing literature. Just stating that the current goal was to outline several broad scenarios before going on to a broad discussion of what the simulated technique could be used for is not to my liking and rather unconventional.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Partly

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medical image computing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 23 January 2019

<https://doi.org/10.21956/wellcomeopenres.16265.r34533>

© 2019 Vidaurre D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Diego Vidaurre

Department of Psychiatry, Oxford Centre for Human Brain Activity (OHBA), University of Oxford, Oxford, UK

In this paper, Cole *et al.* advocate for using adaptive decision procedures during the data acquisition stage such that data collection is optimised with regard to the question at hand.

I think this is conceptually a quite valuable contribution. Although the concepts are clear, my only

negative comment is that it is sometimes slightly hard to follow in the details – although this does not get in the middle of the general message too much.

For example: "To do this, we converted each modality to a z-score, then performed a factor analysis (using Matlab) and calculated a single factor". One eventually gets an idea of what the authors mean, but this could be expressed more clearly.

- What do they mean by convert each modality to a z-score, exactly? How do they do it? First I understood that they normalised each voxel across subjects within each modality, but that's not what it's meant.

- Perform factor analysis; is it done across subjects? i.e. the input factor is (no. subjects by modalities)? To walk the reader through these details and being more explicit would make the reading more amenable.

Another example: why the objective in Scenario 3 is to find the minimum z-score? Could the authors elaborate to make it more accessible?

Another slightly negative comment is that perhaps Example 1 is somewhat trivial. Isn't it standard practice to do this? In any case, I guess it works to illustrate the point.

I particularly liked the Discussion, which I found informative and honest. I am a bit worried about the normative data sets, given that differences in acquisition and preprocessing can make a huge difference (for example, HCP and UKBiobank data are hardly comparable between them).

Relatedly, would it be possible to extend this paradigm of "choosing modality" to "tuning the preprocessing pipeline"? Admittedly the acquisition is the most costly thing, but the same idea could guide perhaps questions like whether to work in volumetric or surface space, which we know makes a big difference in the HCP for example.

In this line, I was wondering whether having different versions of each modality (for example sensor and source space in MEG) would enrich and improve the optimisation in the end? Intuitively I would think that should be the case.

Minor:

"the objective underlying function is unknown and costly to evaluate: -> I would say that either it's unknown or is costly. One can't evaluate (no matter at what cost) something that is unknown?"

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroscience, Machine Learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 21 January 2019

<https://doi.org/10.21956/wellcomeopenres.16265.r34522>

© 2019 Schnack H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Hugo G Schnack 

Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

This manuscript discusses the possible need to change the way neuroimaging is used for diagnostic purposes. The authors argue that, currently, an (MRI) examination uses a fixed set of scan acquisitions (a multimodal imaging protocol), i.e., the same protocol for every patient. This possibly means that either some scans are acquired that are unnecessary for diagnosing this patient, and/or some other scans that are not part of the protocol should have provided better diagnostic information regarding this patient. The authors suggest that changing to an adaptive way of acquiring neuroimaging data -which they call 'active acquisition'- could address this issue of obtaining more efficiently better neuroimaging data, both for clinical and scientific use.

Their method basically starts with acquiring some base scan, which is then analyzed. The information obtained from this first scan is then used to decide what will be the next scan, and so on. The method thus uses a decision tree to determine the multimodal acquisition set for each individual patient. Each patient's acquisition set can thus differ both in length (number of scans) and in kind (modality, resolution, ...) from other patients' sets.

The authors illustrate their ideas with three examples of possible implementations. The authors also state that there are quite a number of (challenging) issues that need to be addressed before such an approach could be of use. In the following I will discuss some of these points and raise some other issues that were not addressed (enough) in the current manuscript.

Time

The factor time (or cost) is one of the main ingredients in the proposed method. Although the authors state that faster processing tools need to be developed for their approach, the reader may at least expect an educated estimate of the time that can be saved, by not acquiring unnecessary scans. This gain in time is compensated by a loss of time due to image processing and analysis. Image reconstruction can take quite some time, images probably have to be transferred to a dedicated computer system for further processing, images need to be processed and analyzed and decisions need to be converted to setting up the next acquisition. The current manuscript does not mention at all times, neither actual times for the steps taken in their examples nor estimates for the steps not currently executed. This information, together with estimates of savings due to new developments, should provide the reader a good impression of whether or not the proposed approach is viable.

Economics

While time saving definitely benefits the patients, the effect on cost is less clear. If scanning takes less time this might reduce the cost of scanning, but it is not clear how this is compensated by increased costs for compute servers, data servers, network equipment and computer programmers on the one hand, and the (for some scenarios) required building of normative databases on the other hand.

Also, although current examination protocols may (to some extent) lead to acquisition of scans that are unnecessary for the diagnosis of this patient, these scans may later turn out to be useful, either for additional diagnostic purposes, for contributing to (new) normative databases, or for scientific research. It happens quite often that neuroimaging data acquired for some (scientific) purpose can be used to answer new and different research questions. Being able to use these data lowers the cost involved in these new studies. It would be worthwhile to have some discussion about this, including estimates regarding the economical feasibility of the proposed method.

Validity

While walking along different investigation paths for different individual cases (scenario 1) clearly is a sensible approach for making clinical diagnosis, for quantitative research this is different. Scenario 2 is used to show how the proposed method can be used 'to quantify [an individual's] relationship to a normative sample'. In this example, each patient's age is predicted based on 1 out of 6 different selections of scans (Figure 3). It is unclear what the effect of individual-based feature selections is on the resulting output (age, in this case). One may expect bias here and differences in accuracy and reliability, all because of using scans that differ in number and/or modality. Apart from showing that the overall prediction accuracy is comparable to that of a state-of-the-art traditional modeling approach (as has been done), the authors should provide a thorough analysis, comparing the different performance measures in a detailed manner. More generally, the potential effects of the proposed method on the statistical validity should be discussed.

Example scenarios - some specific comments:

Scenario 1.

This is a very straightforward example perfectly illustrating a potential application.

Some more information, however, may be expected regarding the following potential issues.

- I can imagine that the process described here is already part of the current examination process, in the sense that an MRI technician decides, after a first quick scan, how to acquire the next (high-resolution) scan. How does the time-saving (if at all) relate to the potential loss in accuracy?

This scenario uses 'the highest outlier distance':

- What if the lesion (or brain abnormality) is too small to have an effect that is statistically detectable?
- What if the lesion lies half-way two thick slices from the first scan?

- What if there are two (or more) lesions in different slices, or if there is longer lesions oriented orthogonal to the slice orientation?

It is not clear what is depicted in Figure 1's right panel. Please clarify the caption.

Scenario 2.

This example is not very convincing, for two reasons:

1. The age estimation model is based on summary statistics rather than fully exploiting the richness of the multimodal datasets.
2. The resulting age prediction is categorical, while a fully continuous prediction may be expected (especially because of the first author's expertise in this field).

Scenarios 2 and 3.

The authors, at the end, state that more advanced ways to incorporate the methods are needed. I recommend that the authors carry out such an implementation themselves, for at least one of these scenarios (preferably no. 2). Since, as the authors acknowledge, the advantages from the current implementations are modest (scenario 3), such an implementation would be helpful to fully show the advances of their proposed method.

All in all, the current manuscript reports about a very interesting approach to innovate the way neuroimaging data are acquired. However, to convince the reader of the potential of this approach, more supportive data and discussion is needed.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroimaging, data analysis, machine learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Comments on this article

Version 1

Author Response 17 Sep 2019

Robert Leech, King's College London, London, UK

Reviewer 1:

This manuscript discusses the possible need to change the way neuroimaging is used for diagnostic purposes. The authors argue that, currently, an (MRI) examination uses a fixed set of scan acquisitions (a multimodal imaging protocol), i.e., the same protocol for every patient. This possibly means that either some scans are acquired that are unnecessary for diagnosing this patient, and/or some other scans that are not part of the protocol should have provided better diagnostic information regarding this patient. The authors suggest that changing to an adaptive way of acquiring neuroimaging data -which they call 'active acquisition'- could address this issue of obtaining more efficiently better neuroimaging data, both for clinical and scientific use.

Their method basically starts with acquiring some base scan, which is then analyzed. The information obtained from this first scan is then used to decide what will be the next scan, and so on. The method thus uses a decision tree to determine the multimodal acquisition set for each individual patient. Each patient's acquisition set can thus differ both in length (number of scans) and in kind (modality, resolution, ...) from other patients' sets.

The authors illustrate their ideas with three examples of possible implementations. The authors also state that there are quite a number of (challenging) issues that need to be addressed before such an approach could be of use. In the following I will discuss some of these points and raise some other issues that were not addressed (enough) in the current manuscript.

Time

The factor time (or cost) is one of the main ingredients in the proposed method. Although the authors state that faster processing tools need to be developed for their approach, the reader may at least expect an educated estimate of the time that can be saved, by not acquiring unnecessary scans. This gain in time is compensated by a loss of time due to image processing and analysis. Image reconstruction can take quite some time, images probably have to be transferred to a dedicated computer system for further processing, images need to be processed and analyzed and decisions need to be converted to setting up the next acquisition. The current manuscript does not mention at all times, neither actual times for the steps taken in their examples nor estimates for the steps not currently executed. This information, together with estimates of savings due to new developments, should provide the reader a good impression of whether or not the proposed approach is viable.

Response: The reviewer makes important points regarding the trade-offs between analysis time set against the benefits of the adaptive approach and potential time savings. We now discuss

this issue in more detail: 1) we provide some approximate timing information regarding processing T1-weighted images in Scenario 1; using a standard, quad-core computer, the pre-processing and normative modelling was performed in < 1 minute. For the other scenarios, the large CamCAN dataset was processed by a high-performance compute cluster, and we did not record timing information; however, some of the pipelines (e.g., fitting tensor models to calculate fractional isotropy or non-linear image registration) were slower than would be feasible for adaptive acquisition; 2) we discuss the use of parallelization and dedicated computational hardware that could be used to substantially improve speed; 3) we discuss how pipelines have not been optimized for speed and how this is likely to bring timing down to a fraction of the current time (as has been achieved with e.g., fMRI for brain-computer interfaces) with minimal loss of image quality. 4) Finally, we discuss the role of deep learning approaches, that may be slow to train and require large datasets, but are extremely fast to apply to new data, making image analysis possible within seconds.

Economics

While time saving definitely benefits the patients, the effect on cost is less clear. If scanning takes less time this might reduce the cost of scanning, but it is not clear how this is compensated by increased costs for compute servers, data servers, network equipment and computer programmers on the one hand, and the (for some scenarios) required building of normative databases on the other hand.

Response: This is an important issue and will depend on the relative cost of scanning/radiographer time versus computing time as well as the level of increased efficiency or equally improved detection of clinically relevant individual variability. We discuss this in more detail in the revised Discussion section.

It is an empirical question whether additional normative databases in addition to the (e.g., UKBiobank) will be good value for money. This will depend on whether types of scans not already collected in existing large normative samples capture clinically or scientifically useful variability that could be imaged or not and the economic value of that clinical and scientific value. Equally, it depends on whether we already know the best scans for the types of individual variability or not. One approach is to start with relatively small-scale (i.e., cheap) normative data collection with far more types of scans, to quantify how a more exploratory approach, with individually tailored scan sequences can outperform a small subset of scans applied to everybody.

Also, although current examination protocols may (to some extent) lead to acquisition of scans that are unnecessary for the diagnosis of this patient, these scans may later turn out to be useful, either for additional diagnostic purposes, for contributing to (new) normative databases, or for scientific research. It happens quite often that neuroimaging data acquired for some (scientific) purpose can be used to answer new and different research questions. Being able to use these data lowers the cost involved in these new studies. It would be worthwhile to have some discussion about this, including estimates regarding the economical feasibility of the proposed method.

Response: We have added to the discussion substantially regarding this point. Broadly, we strongly agree that discovery is an important consequence of current neuroimaging practice. However, we would like to make two points: (1) the amount of discovery and how this proceeds will depend on the clinical/scientific question; and, (2) if data serendipitously providing insight into new clinical/scientific questions is a goal, then this can be performed in a controlled way,

e.g., through the choice of a more exploratory acquisition function guiding the active acquisition, with the potential to allow for efficient imaging while also efficiently exploring a larger space of different types of scans.

Validity

While walking along different investigation paths for different individual cases (scenario 1) clearly is a sensible approach for making clinical diagnosis, for quantitative research this is different. Scenario 2 is used to show how the proposed method can be used 'to quantify [an individual's] relationship to a normative sample'. In this example, each patient's age is predicted based on 1 out of 6 different selections of scans (Figure 3). It is unclear what the effect of individual-based feature selections is on the resulting output (age, in this case). One may expect bias here and differences in accuracy and reliability, all because of using scans that differ in number and/or modality. Apart from showing that the overall prediction accuracy is comparable to that of a state-of-the-art traditional modeling approach (as has been done), the authors should provide a thorough analysis, comparing the different performance measures in a detailed manner. More generally, the potential effects of the proposed method on the statistical validity should be discussed.

Response: We have amended the discussion to consider this in much more detail, see also our response to the specific concerns about Scenario 2 below.

Example scenarios - some specific comments:

Scenario 1.

This is a very straightforward example perfectly illustrating a potential application. Some more information, however, may be expected regarding the following potential issues.

I can imagine that the process described here is already part of the current examination process, in the sense that an MRI technician decides, after a first quick scan, how to acquire the next (high-resolution) scan. How does the time-saving (if at all) relate to the potential loss in accuracy?

This scenario uses 'the highest outlier distance':

What if the lesion (or brain abnormality) is too small to have an effect that is statistically detectable?
Computer vision

What if the lesion lies half-way two thick slices from the first scan?

What if there are two (or more) lesions in different slices, or if there is longer lesions oriented orthogonal to the slice orientation?

Response: These are important issues that will need to be considered in future work focusing on more specific scientific/clinical questions. However, we now mention these limitations in the revised manuscript and also discuss that active acquisition approaches need not be limited to a single target function such as outlier distance for large areas of the brain. Instead, distinct target functions (e.g., based on different ways of decomposing the multivariable imaging signal) can be applied in parallel and a decision rule based on e.g., which is most likely an outlier, used to guide further data acquisition.

It is not clear what is depicted in Figure 1's right panel. Please clarify the caption.

Scenario 2.

This example is not very convincing, for two reasons:

1. The age estimation model is based on summary statistics rather than fully exploiting the richness of the multimodal datasets.
2. The resulting age prediction is categorical, while a fully continuous prediction may be expected (especially because of the first author's expertise in this field).

Response: We agree that it is not convincing for the two reasons stated by the reviewer and note that performance is considerably worse than would be achieved with a whole brain single modality approach where the mean brain predicted age may be <5 years from the true age. However, the objective, here, was emphatically not to demonstrate the optimal method of calculating brain age, but a different approach to acquiring data, that may in future be used to provide more accurate neuroimaging measures, including potentially more precise brain age data. As such, we chose to use a transparent machine learning technique to illustrate how such an approach could be used; this motivated the use of decision tree regression with single summary measures which shows the acquisition trajectory; this would be far less accessible if, e.g., a succession of support vector regressions. We now acknowledge this very clearly along with the reviewer's issues and future directions in the revised version of the manuscript both in the methods/results and in the discussion.

Scenarios 2 and 3.

The authors, at the end, state that more advanced ways to incorporate the methods are needed. I recommend that the authors carry out such an implementation themselves, for at least one of these scenarios (preferably no. 2). Since, as the authors acknowledge, the advantages from the current implementations are modest (scenario 3), such an implementation would be helpful to fully show the advances of their proposed method.

All in all, the current manuscript reports about a very interesting approach to innovate the way neuroimaging data are acquired. However, to convince the reader of the potential of this approach, more supportive data and discussion is needed.

Response: Broadly, we agree with the reviewer that the examples are limited. Unfortunately, this is inevitable given the constraints of analysing existing data that was not designed to demonstrate the potential of active acquisition. Fundamentally, unless data is analysed as it is collected and a normative dataset of many different types of scans acquired, then it is not possible to properly demonstrate the benefits of the approach. At present we are restricted to only a small subset (i.e., 6 modalities), and are unable to utilise a wider variety of MRI modalities, which is a key element of active acquisition. This manuscript aims to outline a conceptual framework rather than provide definitive evidence. Our current ongoing research is moving beyond the current simulations and we are collecting a much wider normative dataset, with the goal of providing support evidence for the adaptive acquisition approach. Until this work is completed, we hope that the current manuscript serves to outline our motivation for how active acquisition could be implemented.

Reviewer 2

In this paper, Cole et al. advocate for using adaptive decision procedures during the data acquisition stage

such that data collection is optimised with regard to the question at hand.

I think this is conceptually a quite valuable contribution. Although the concepts are clear, my only negative comment is that it is sometimes slightly hard to follow in the details – although this does not get in the middle of the general message too much.

For example: "To do this, we converted each modality to a z-score, then performed a factor analysis (using Matlab) and calculated a single factor". One eventually gets an idea of what the authors mean, but this could be expressed more clearly.

- What do they mean by convert each modality to a z-score, exactly? How do they do it? First I understood that they normalised each voxel across subjects within each modality, but that's not what it's meant.

Response: We have clarified this in the revised manuscript.

- Perform factor analysis; is it done across subjects? i.e. the input factor is (no. subjects by modalities)? To walk the reader through these details and being more explicit would make the reading more amenable.

Response: We are now clearer about how the factor analysis was performed.

Another example: why the objective in Scenario 3 is to find the minimum z-score? Could the authors elaborate to make it more accessible?

Response: This choice of target function was somewhat arbitrary, and intended as illustrative, given that there is no clear direction to the z-scores for in terms of pathology (e.g., resting state connectivity measures). We could equally have maximised the z-score or the absolute z-score. The actual target value would be based on the clinical/scientific question.

Another slightly negative comment is that perhaps Example 1 is somewhat trivial. Isn't it standard practice to do this? In any case, I guess it works to illustrate the point.

Response: We agree, that the scenario is highly trivial; our objective was to highlight the approach and how it proceeds with an almost trivial example. We state this more clearly in the revised manuscript.

I particularly liked the Discussion, which I found informative and honest. I am a bit worried about the normative data sets, given that differences in acquisition and preprocessing can make a huge difference (for example, HCP and UKBiobank data are hardly comparable between them).

Response: We completely agree that there are likely to be many challenges and, in a related vein, that simulations based on our current datasets are fundamentally inadequate; we need to collect far broader multimodal datasets (i.e., more types of scan), and then build bespoke preprocessing and normative modelling aimed at finding clinically or scientifically useful sources of individual variability. We discuss this in the revised discussion.

Relatedly, would it be possible to extend this paradigm of "choosing modality" to "tuning the preprocessing pipeline"? Admittedly the acquisition is the most costly thing, but the same idea could guide perhaps questions like whether to work in volumetric or surface space, which we know makes a big difference in the HCP for example.

In this line, I was wondering whether having different versions of each modality (for example sensor and source space in MEG) would enrich and improve the optimisation in the end? Intuitively I would think that should be the case.

Response: We agree with the reviewer, and mention that multiple representation of the same imaging data could be used e.g., to simultaneously update multiple points in the search space.

Minor:

"the objective underlying function is unknown and costly to evaluate: -> I would say that either it's unknown or is costly. One can't evaluate (no matter at what cost) something that is unknown?"

Response: We have revised this to state "the general objective underlying function is unknown a priori and is costly to evaluate at any given point"

Reviewer 3

This is a paper about a method called active acquisition. The authors propose a framework where neuroimaging sequence selection is iteratively optimized while the patient is on the table using an active learning approach. The method is currently a simulation. At this point the paper is an outline of a research strategy not a presentation of a method that is actually working.

In practical terms, the speed of the computation will be critical. If the computation increases the time that the patient spends in the scanner, this will result in resistance to using this approach.

My main concern is that the three scenarios are spread widely and mostly talk about the concept. In my opinion the authors should limit themselves to one scenario and flesh it out with enough detail that the work could be duplicated.

Introduction:

- "Active learning also has another important feature; they involve a prediction and testing cycle.. they or it?"

Response: Corrected

- "using information gained from previous scans actively seek out brain abnormalities or make diagnostic predictions..." should read "to actively seek.."

Response: Corrected

Methods:

Scenario 1: detection of stroke pathology

Choice of parameters:

- How about a short summary of Dataset 1? The paper should be self-contained.

Response: *We have added a sentence to the Methods (page 4) directing readers to the short summary of Dataset 1 that is included in Data Acquisition (page 6):*

“Further details on the participants are included in the Data Acquisition section.”

- “At each iteration the scan...” which scan? Is it at full resolution or, as implied in the introduction, at reduced resolution? How much reduction?

Response: *To clarify, here when we say each iteration, we mean at each resolution. In other words, each iteration is getting progressively higher resolution, with voxel size reducing by a third each time. We have updated the sentence to read as follows:*

“At each iteration (in terms of increasing resolution), the scan is divided into three equally sized volumes, along the z-dimension.”

- What is a normative sample and where is it coming from?

Response: *In Scenario 1, the ‘normative’ sample in this case was simply the n=7 healthy controls, as mentioned in the Data Acquisition section. We’ve added a mention of this as a caveat in the Discussion as follows:*

“Currently for Scenario 1, the normative dataset is simply the n=7 healthy controls; in reality the size of normative will have to be much larger.”

Outlier distance:

- When you say median distance, what do you mean? Distance in voxels, signal intensities, something else?

Response: *In this we mean median signal intensities. We have included the following:*

“The median distance between an individual’s scan and each participant in the normative dataset was calculated using the median absolute deviation (in Euclidean distance) of signal intensity averaged across all voxels.”

Scenario 2: choose a modality

- Again, provide a short description of the data set, so the reader does not have to visit a different webpage just to know what data you are using. Also, which modalities are you planning to use in the simulation and why? I went to the Cam-CAN webpage and modalities include task based fMRI and MEG. It is not clear to me how such acquisitions could be performed dynamically at need.

Response: *The data are described in the Data Acquisition section (page 7) as follows:*

“For Scenarios 2 and 3, multimodal MRI data from 611 people (age range from 18-88, 312, female) were taken from the Cam-CAN dataset. These data consisted of T1-weighted, T2-weighted, diffusion-weighted MRI, three functional scans (resting-state fMRI, movie-watching fMRI and a blocked sensorimotor task-based fMRI). Imaging acquisition has been presented in detail elsewhere (Shafto et al. 2014).”

Scenario 3: discovery of individual differences.

- “we attempt to actively learn which modality an individual is most likely to be an outlier in” - Please reword.

Response: *We have reworded as follows:*

“... in Scenario 3, we attempt to actively learn in which modality an individual is most likely to be an outlier (i.e., is a given individual more likely to be an outlier from the norm when using phenotypes derived from: T1-weighted MRI, diffusion-MRI or functional MRI?).”

- The section data acquisition provides the details that would have been helpful to have when reading the scenarios. There were no cross-references in the previous sections. The scenario descriptions 2 and 3 do not mention that only a subset of the Cam-CAN data is being used.

Response: As suggested, we have added cross-references between the Scenarios section and the Data Acquisition section, as mentioned above. We have also added the following to clarify why only $n=611$ Cam-CAN participants were included:

“Only Cam-CAN participants with complete data from these six sequences were included ($n=611$, out of $n=653$).”

Results:

Scenario 1: Figure 2 is not clear to me. What do you show when you mean composite?

Response: The composite images are the combination of three different images: 1) a whole brain low-resolution image, 2) a one-third field-of-view image at a higher resolution, 3) a one-ninth field-of-view image at even higher resolution. We have reworded the figure legend to clarify, as follows:

“Figure 2: top, composite T1-weighted coronal slices (one for each of the six patients), comprise out three scans with increasing resolution, each with different coverage of the field-of-view (less field-of-view, higher resolution). The reduced field-of-view is centred on the slice from the previous image that is quantified as most abnormal relative to the norm (i.e., the control group). This demonstrates that the very simple approach to subdividing the brain and quantifying outliers can be used to ‘zoom’ in on areas of pathology that are specific for individual patients.”

Scenario 2: It would be helpful to have two examples: a young and an old person and how their trajectory in the regression is different.

Response: We agree that examples are useful here, thus have provided a walk-through of different trajectories through the decision tree that result in different age predictions and require different combinations of MRI modalities. Notably, the decision tree does not know (or need to know) the actual age of the individual being scanned, the age prediction is made solely from neuroimaging data. The walk-throughs are as follows:

“By way of individual examples, if a participant had a GM z-score = -0.8 and FA z-score = -0.7 , then they would have a predicted age of 77.6 years (following the left most branches of the decision tree in Figure 3). However, if a participant also has GM z-score = -0.8 , but their FA z-score = 0.5 , then task BOLD data would be necessary to make a prediction. If that individual’s task BOLD z-score ≥ -0.18 then their predicted age would be 55.7 years, whereas a task BOLD z-score < -0.18 would give a predicted age of 63.1 years.”

Scenario 3: I do not understand what the results are. “Here we simulated closed-loop Bayesian optimisation used to discover the modality for a given individual (from the holdout dataset) where the negative outlier distance is most (i.e., relative to normative data from the training dataset), shown in Figure 4.”

Response: We agree that the results in Scenario 3 were incompletely articulated. We have added a passage to outline the results more clearly, as follows:

“Scenario 3 used a normal distribution (from a normative dataset) to derive individual z-scores for each MRI modality, with the goal of identifying for each individual in which modality they are most outlying (i.e., which modality has the highest absolute z-score for that person). To achieve this we simulated closed-loop Bayesian optimisation to identify the highest z-score from across all modalities for a given individual (from the holdout dataset), as shown in Figure 4. For individual depicted in Figure 4A, the highest z-score was in T2-weighted MRI ($z = -1.2$), followed by resting-state fMRI, sensorimotor task BOLD, FA, movie BOLD and then GM. This suggests that for this individual, the T2-weighted is likely to be most valuable in determining whether or not they are experiencing some underlying pathology. While this approach can be used to rank z-scores across modalities, the magnitude of the z-score can also be informative, to provide insight into

how much of an outlier this individual is in that modality (which may or may not indicate the presence of a pathology).”

Discussion:

- The authors should begin the discussion of the results in the three scenarios and how they relate to existing literature. Just stating that the current goal was to outline several broad scenarios before going on to a broad discussion of what the simulated technique could be used for is not to my liking and rather unconventional.

Response: *We appreciate that our Discussion is rather conventional, however, our preference would be to retain the current structure and content. Given that the aim of the article is to outline a potential methodological approach, we believe the Discussion section as is meets that goal. While we do include some empirical data to better illustrate our ideas, there is no existing literature that has looked to optimise how neuroimaging data are acquired using real-time analysis. Admittedly, the specific examples, such as age prediction, do have a body of literature supporting them, however the current goal was not to present an improved method of (offline) age prediction, but to present age prediction of one goal that could be achieved (and potentially optimised) using real-time analysis to decide which data to acquire. We believe that the results of the three scenarios are discussed adequately under the current structure, though this has been done in the context of more general applications (Clinical diagnosis; Biomarker discovery). Since the current data were intended to be illustrative, we feel that this approach broader contextual discussion is more appropriate.*

Competing Interests: No competing interests were disclosed.
