

ORIGINAL ARTICLE

Protocol for magnetic resonance imaging acquisition, quality assurance, and quality check for the Accelerator program for Discovery in Brain disorders using Stem cells

Pravesh Parekh^{1,2,3}  | Gaurav V. Bhalerao^{1,3,4}  | Rashmi Rao¹ |
Vanteemar S. Sreeraj^{1,3,4} | Bharath Holla^{1,3} | Jitender Saini⁵ |
Ganesan Venkatasubramanian^{1,3,4}  | John P. John^{1,2,3}  | Sanjeev Jain³ |
ADBS Consortium

¹ADBS Neuroimaging Centre, National Institute of Mental Health and Neurosciences, Bangalore, India

²Multimodal Brain Image Analysis Laboratory, National Institute of Mental Health and Neurosciences, Bangalore, India

³Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore, India

⁴Translational Psychiatry Lab, National Institute of Mental Health and Neurosciences, Bangalore, India

⁵Department of Neuroimaging and Interventional Radiology, National Institute of Mental Health and Neurosciences, Bangalore, India

Correspondence

Dr. John P. John, Department of Psychiatry,
Multimodal Brain Image Analysis Laboratory,
National Institute of Mental Health and
Neurosciences (NIMHANS), Bangalore
560029, India.

Email: jpj@nimhans.ac.in and
jpjnimhans@gmail.com

Dr. Ganesan Venkatasubramanian,
Department of Psychiatry, Translational
Psychiatry Lab, National Institute of Mental
Health and Neurosciences (NIMHANS),
Bangalore 560029, India.

Email: gvs@nimhans.ac.in and venkat.nimhans@gmail.com

Funding information

Department of Biotechnology, Ministry of
Science and Technology, Grant/Award
Number: BT/PR17316/MED/31/326/2015;
National Institute of Mental Health, Grant/
Award Number: RO1 MH113250

Abstract

Objective: The Accelerator program for Discovery in Brain disorders using Stem cells (ADBS) is a longitudinal study on five cohorts of patients with major psychiatric disorders from genetically high-risk families, their unaffected first-degree relatives, and healthy subjects. We describe the ADBS protocols for acquisition, quality assurance (QA), and quality check (QC) for multimodal magnetic resonance brain imaging studies.

Methods: We describe the acquisition and QC protocols for structural, functional, and diffusion images. For QA, we acquire proton density and functional images on phantoms, along with repeated scans on human volunteer. We describe the analysis of phantom data and test–retest reliability of volumetric and diffusion measures.

Results: Analysis of acquired phantom data shows linearity of proton density signal with increasing proton fraction, and an overall stability of various spatial and temporal QA measures. Examination of dice coefficient and statistical analyses of coefficient of variation in test–retest data on the human volunteer showed consistency of volumetric and diffusivity measures at whole-brain, regional, and voxel-level.

Conclusion: The described acquisition and QA–QC procedures can yield consistent and reliable quantitative measures. It is expected that this longitudinal

Pravesh Parekh and Gaurav V. Bhalerao should be considered joint first author.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. International Journal of Methods in Psychiatric Research published by John Wiley & Sons Ltd.

neuroimaging dataset will, upon its release, serve the scientific community well and pave the way for interesting discoveries.

KEYWORDS

ADBS, magnetic resonance imaging, longitudinal study, quality assurance, quality check

1 | INTRODUCTION

Discovery biology of neuropsychiatric syndromes is an approach that proposes to evaluate overlapping and unique genetic, environmental, and developmental factors across five major psychiatric disorders: schizophrenia, bipolar disorder, obsessive-compulsive disorder, Alzheimer's dementia, and alcohol dependence syndrome (Viswanath et al., 2018). Using this approach, the Accelerator program for Discovery in Brain disorders using Stem cells (ADBS) project aims to create a dataset of patients and their first degree relatives (FDRs) from families with high genetic loading (i.e., with at least another affected FDR having any of the above five disorders), and socio-demographically matched healthy subjects. This dataset will consist of detailed clinical and neuropsychological assessments, electroencephalogram (EEG)/event-related potential (ERP), functional near-infrared spectroscopy (fNIRS), eye tracking, and multimodal magnetic resonance imaging (MRI) data, along with genomics and stem cells, thereby allowing us to characterize the structure and function of the brain at macroscopic and microscopic levels across these five disorders. A longitudinal study that integrates the above clinical, neuropsychological, neurophysiological, psychophysical, neuroimaging, genomic, and stem cell data (Bilderbeck et al., 2019; Clementz et al., 2015; McPhie et al., 2018; Tamminga et al., 2014; Vasistha et al., 2019; de Vrij et al., 2019) using a systems-based approach (Saxe et al., 2016) has the potential to provide deeper insights into the shared and unique neurodevelopmental trajectories of these neuropsychiatric disorders (Etkin, 2019; McGorry et al., 2018) and generate a large database for use by the scientific community.

MRI allows us to noninvasively study the structure and functioning of the brain. However, quantitative variables derived from MRI are susceptible to factors like acquisition protocol and MR hardware (Han et al., 2006; Kruggel et al., 2010); variations in scanner-related factors like drift (Takao et al., 2011), ghosting (Reeder et al., 1997), and signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) (Jovicich et al., 2009); artefacts related to study participants such as motion (Power et al., 2012; Reuter et al., 2015; Yendiki et al., 2014); ambient temperature (Vogelbacher et al., 2018); time of acquisition (Orban et al., 2020); and data processing workflow (Glatard et al., 2015; Gronenschild et al., 2012; Kharabian Masouleh et al., 2020). Therefore, to ensure the validity, precision, and reliability of quantitative MRI outcome measures, it is important to set up optimal scanning parameters (like Alzheimer's Disease Neuroimaging Initiative [ADNI; Jack et al., 2008; <http://adni.loni.usc.edu>] and Human Connectome Project [HCP; Van Essen et al., 2013; <https://neuroscienceblueprint.nih.gov/human-connectome/>

connectome-programs]), adhere to this acquisition protocol (Wunderlick et al., 2009), ensure regular monitoring of scanner performance (Friedman & Glover, 2006; Stöcker et al., 2005), and minimize subject-independent variances in MRI measurements (Bennett & Miller, 2010; Gunter et al., 2009; Maclaren et al., 2014).

Quality assurance (QA) and quality check (QC) principles can be used to ensure reliability of outcome measurements. Typically, QA is done at the MRI site by acquiring phantom scans that can help in identifying scanner related artefacts and signal fluctuation, while QC is done by manual and/or automated inspection of the already acquired human data. Having robust QA and QC protocols is critical, especially for longitudinal studies. This is underscored by observations such as that of (Stöcker et al., 2005), where the authors show that differences in quality measurements across groups of participants, can lead to misinterpretations of the results. The goal of this paper is to provide an overview of the MRI acquisition and QA-QC protocol of the ADBS project. We comprehensively describe each step of acquisition and QA-QC protocols along with example results which will illustrate that the data being acquired under the ADBS project is capable of yielding consistent and reliable outcome measures.

2 | METHODS

2.1 | Participants

The study design, short-, and long-term objectives, and details of sample size calculation of the ADBS project are fully described elsewhere (Viswanath et al., 2018). We aim to examine overlapping and unique genetic, environmental, and developmental factors in five disorders: schizophrenia, bipolar disorder, obsessive-compulsive disorder, Alzheimer's dementia, and alcohol dependence syndrome. A sample of 4500 individuals comprising of index probands belonging to the above five cohorts, their unaffected FDRs and healthy comparison subjects undergo phenotypic assessments and blood sampling. The blood sample is used for molecular genetic studies and cellular modeling of diseases. A subset of 1500 participants (1200 from affected families and 300 healthy controls from unaffected families) from the above sample that constitutes the neurodevelopmental endophenotype cohort, undergo biennial assessments using structural, functional, and diffusion MRI, EEG/ERP, fNIRS, eye movement tracking, and neuropsychological evaluation. Participants are recruited after obtaining written informed consent duly approved by the institutional ethical committee of the National Institute of

Mental Health and Neurosciences (NIMHANS), Bangalore, India, where this study is being carried out (Ethics approval number: Item No. VII, Sl. No. 7.01, Behavioural Sciences & Item No. XI, Sl. No. 11.05, Behavioural Sciences).

2.2 | MRI acquisition

The MRI acquisition is performed temporally as close as possible to other clinical and endophenotype assessments (typically within 1–2 days, with a maximum of up to 1 week). All subjects are comprehensively screened for MR safety by the clinical team, followed by an on-site screening prior to MRI acquisition (see Figure S20 for MRI screening form). The MRI session consists of two task-based functional MRI (fMRI) acquisitions using the Tool for Recognition of Emotions in Neuropsychiatric Disorders (TRENDS; Behere et al., 2008), and verbal fluency task (VFT; John et al., 2011). Training of subjects for task-based fMRI is conducted prior to the MRI acquisition session by trained neuropsychologists. First, the primary language (or the language most comfortable among the available languages) of the participant is identified using a screening questionnaire (see Figure S21 for the questionnaire on determination of primary language). Then, participants are familiarized with fMRI tasks using a video recording. At the end of the training session, the neuropsychologists ensure that the participants have understood the task instructions.

The MRI data are acquired on a 3T Philips Ingenia CX (Philips Healthcare) machine using a 32-channel phased-array coil. The initial software version was R5.3.0.3 and the current software version being used is R5.3.1.3. A fiducial marker (vitamin E capsule) is used for enabling unambiguous delineation of laterality. A plastic mirror is mounted on the head coil such that the participants can view the MR-compatible monitor. Stimulus presentation is done via E-Prime 3.0.3.60 (Psychology Software Tools; <https://pstnet.com/products/e-prime>). Synchronization of fMRI stimulus presentation with the scanner is achieved via a sync box (Nordic Neuro Lab Inc.; SyncBox S, Type Number SB-3.0; <https://www.nordicneurolab.com/en/products>) and the stimulus is displayed on an LCD monitor (Nordic Neuro Lab Inc). An MR-compatible microphone with active noise cancellation (Optoacoustics—Advanced Noise canceling Fiber Optic Microphone for fMRI, FOMRI™ – III, Optoacoustics Ltd., Moshav Mazor; <http://www.optoacoustics.com>) is used to collect vocal responses from participants.

The MRI acquisition protocols were adapted from the ADNI MRI protocols for Philips Ingenia (software version R5). The scan session consists of structural, functional (resting state and task-based fMRI), and diffusion scans. Table 1 lists the sequences and duration of each sequence being acquired during the scan session. A brief description of our acquisition protocol is mentioned below while additional details are presented in Tables S1–S4. A figure showing the contrast of each type of image is shown in Figure S1. Details of post acquisition data integrity check and data storage are mentioned in the Supporting Information Materials.

2.2.1 | Survey

A localizer sequence (“survey”) is initially run which captures the position of the head by acquiring three slices in sagittal, coronal, and axial views. Since the total duration of acquisition is long (1 h 07 min), we acquire three different survey scans during the session: one at the beginning, second before the first task-based (tb) fMRI (tb-fMRI) (~13 min 47 s into the scan), and a third before the second tb-fMRI (~35 min 02 s into the scan). An additional survey is done before the diffusion-weighted imaging (DWI) scan (~53 min 50 s after the beginning of the scan), if there is an indication that the study participant has moved.

2.2.2 | Structural scans

We acquire a high resolution T1-weighted single-shot three-dimensional (3D) turbo field echo (TFE) image at $1 \times 1 \times 1$ mm voxel size, in the sagittal orientation (see Table S1 for details). The field of view (FOV) is set to ensure full brain coverage while ensuring that no fold-over happens. An acceleration factor (SENSE) of 2 is applied in the right to left (RL) direction to reduce the overall time taken for this acquisition.

In addition to the T1-weighted scan, we acquire a 3D multishot TFE phase-sensitive inversion recovery (PSIR) image at $1 \times 1 \times 1$ mm voxel size, in the sagittal orientation (see Table S1 for details). A SENSE factor of 2 in the anterior–posterior (AP) and 1.75 in the RL direction is applied for reducing the amount of time taken. The PSIR images are reconstructed so that cerebrospinal fluid (CSF) voxels have negative values. This can potentially lead to better contrast between tissue types (Hou et al., 2005; Moran et al., 1986; Park et al., 1986).

We also acquire a T2-weighted structural scan at $1 \times 1 \times 1$ mm voxel size in the transverse orientation (see Table S1 for details). This is a multishot turbo spin-echo (SE) sequence with fat-suppression using SPIR technique. Apart from using this image for various structural analyses, this image can be also used for registration and correction of geometrical distortion in DWI.

Additionally, for each study participant, we acquire a fluid-attenuated inversion recovery image to visualize white matter (WM) abnormalities.

2.2.3 | Functional scans

For functional blood oxygenation level-dependent (BOLD) contrast, echo-planar imaging (EPI) method is used. We acquire and discard (at source) five dummy scans to allow magnetization level to reach a steady-state. During EPI scans, brain areas such as OFC are susceptible to signal loss; to mitigate this, we employ a second-order pencil beam shimming (PBS) covering the entire brain which attempts to homogenize the MR signal in the shim region. Further, EPI

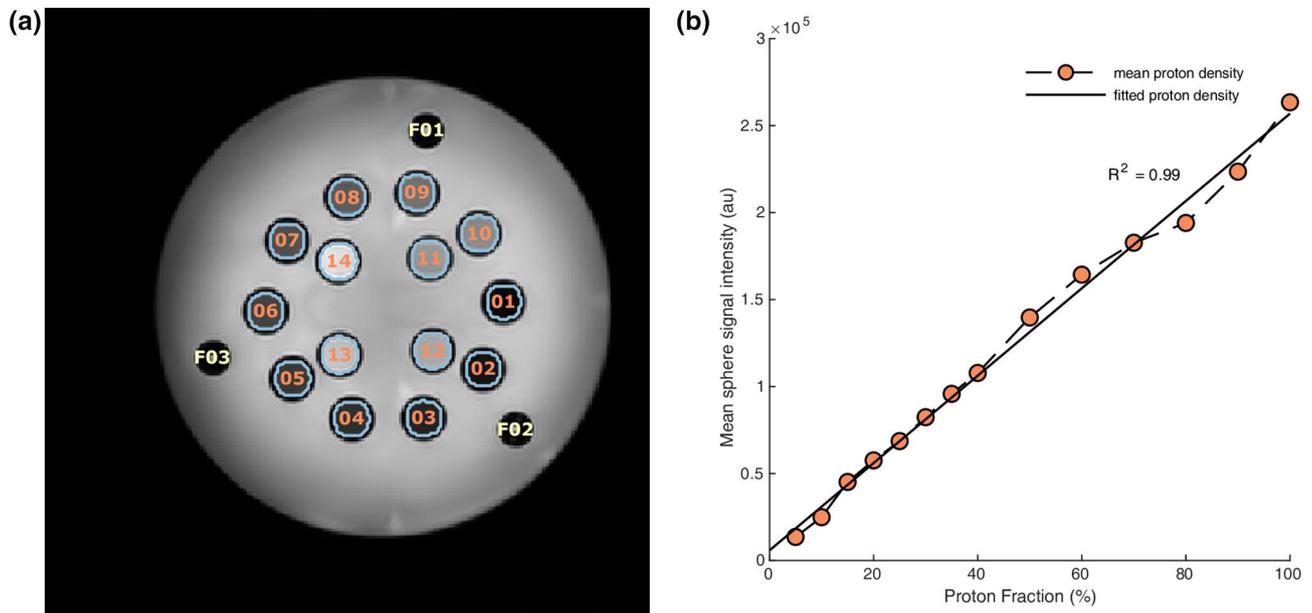


FIGURE 1 Analysis of proton density images acquired on the HPD phantom; (a) fiducials F01, F02, and F03 are first detected. These locations are then used for labeling the other circles using heuristics (see Supporting Information Material). The boundaries of the circles are eroded twice to ensure that only the signal within the circle is used for any calculation; (b) mean signal from each circle is plotted against the proton fraction in these spheres. Since the signal is supposed to linearly increase with increasing proton fraction, we fit a linear regression model and calculate the R^2 value. A poorer fit of this linear model would indicate a problem in the acquisition

Sequence	Duration (MM:SS)
Survey	00:31.5
Structural scans	
T1w	05:39.2
T1w-phase sensitive inversion recovery	08:59.2
T2w	04:48.0
Fluid-attenuated inversion recovery	04:48.0
Functional scans	
Resting state	10:20.3
Resting state (reference scans)	00:32.0 each
Verbal fluency task	07:35.9
Verbal fluency task (reference scans)	00:32.0 each
Tool for Recognition of Emotions in Neuropsychiatric Disorders (TRENDS)	05:33.9
TRENDS (reference scans)	00:32.0 each
Fieldmap	00:56.2
Diffusion scans	
Multishell diffusion scan	10:35.4
Diffusion reference scans	00:32.0

TABLE 1 List of magnetic resonance imaging sequences and their duration (in minutes and seconds)

scans often lead to geometrical distortion. To correct for potential geometrical distortion, we acquire fieldmap and opposite polarity phase encoding EPI scans (Jezzard & Balaban, 1995; Jezzard & Clare, 1999; Smith et al., 2004).

Resting state fMRI

During resting-state fMRI (rsfMRI) scan (see Table S2 for details), study participants are asked to keep their eyes open. Acquisition is carried out without any stimuli, leaving the monitor screen blank.

Since previous research has indicated that at least 10 min of rsfMRI data are required for reliable characterization of various network properties (though the timing required varied from 10 to 80 min) (Birn et al., 2013; Gordon et al., 2017), we acquire 10 min and 20.3 s of rsfMRI data (275 volumes, TR 2.2 s).

Task-based fMRI

Abnormalities in emotion recognition (Castellano et al., 2015; Daros et al., 2014; Elferink et al., 2015; Homorogan et al., 2017; Kohler et al., 2010) and verbal fluency (Cardenas et al., 2016; Clark et al., 2009, 2016; Liang et al., 2016; Snyder et al., 2015; Weiner et al., 2015) have been reported in all the disorders being studied in this project. We use tasks that were previously developed at NIMHANS for tb-fMRI acquisitions: TRENDS (Behere et al., 2008), which is a standardized validated tool for studying emotion recognition deficits in the Indian context, and VFT (John et al., 2011), which is a semantic category overt word generation task that is implemented as a blocked design. A brief description of these tasks including the acquisition parameters (Table S2), as well as the acquisition parameters for reverse polarity phase encoded reference scans (Table S3) are provided in the Supporting Information Material. The task design for VFT was exactly the same as described in (John et al., 2011; with a different set of acquisition parameters, as summarized in Table S2); for TRENDS, the original paradigm had six emotions: happy, sad, fear, anger, surprise, and disgust, along with neutral expression. Patients with severe psychiatric disorders like schizophrenia have been shown to have intact recognition of positive expressions (i.e., happiness; Bediou et al., 2005; Kohler et al., 2010). In order to shorten the length of the paradigm given the time constraints, we did not include happy facial expression trials in the paradigm since recognition of happy faces may not differentiate patient groups from healthy subjects.

To prevent any carry over effect of one task to another, task scans are separated by the PSIR scan (described above) which lasts for about 9 min. The tasks are further spaced by the preceding two reverse polarity phase-encoded reference scans (for EPI distortion correction) before the beginning of the actual task (32 s each). The tasks are administered in a counterbalanced manner (i.e., the task order is alternated across subjects) to avoid order effects across participants.

2.2.4 | Diffusion-weighted imaging

The DWI scan (see Table S4 for details) is preceded by reference images which can be used for correction of EPI-induced geometrical distortions. These are SE images acquired in opposite phase encoding directions (posterior–anterior and AP), two volumes in each direction. The main scan is acquired with a multi-shell sampling protocol, optimized to provide uniform coverage on each shell, and a global uniform angular coverage (Caruyer et al., 2013). Three high b -value shells corresponding to 1000, 2000, and 3000 (smm^{-2}) with 25, 24, and 24 gradient directions respectively and seven interspersed volumes without diffusion weighting (b -value = 0) are acquired using a

second-order PBS covering whole brain to homogenize the signal in regions susceptible to signal loss.

2.3 | Quality Assurance

2.3.1 | MRI phantoms

Acquisition of data on MRI phantoms is a way to ensure that the scanner stability is maintained over a period of time. Since there are no sources of variation in a phantom, if acquisition protocol and related factors are kept constant, the scan-to-scan change seen in phantom data should be minimal. Depending on the type of phantom and the data being acquired, a variety of parameters targeting specific scanner or image properties can be studied. In this study, we use two kinds of phantoms: an agar gel phantom for ensuring temporal stability of EPI data, and a geometric phantom for ensuring the stability of acquired MR signal.

Agar gel phantom

We use an agar gel phantom (http://pro-project.pl/pro-mri_agar) to quantify the temporal stability of EPI BOLD signal. The phantom is placed as centrally as possible in the 32-channel head coil as shown in Figure S2a with appropriate paddings. The phantom is placed in the horizontal direction and is corrected for any misalignment based on the localizer scan as shown in Figure S3.

Since the goal of the QA procedure is to examine scanner stability under similar conditions as an fMRI experiment (Friedman & Glover, 2006), we acquire EPI BOLD images on the phantom with the same protocol as human subjects (summarized in Table S2), except that only 140 volumes are acquired and dummy scans are not acquired. The shim volume is adjusted so that it covers the entire phantom as shown in Figure S3.

To examine the properties of EPI images acquired on the phantom, we followed the methods detailed earlier (Vogelbacher et al., 2018). The implementation is adapted from the MATLAB scripts for the gel phantom implemented in LAB-QA2GO (Vogelbacher et al., 2019; see Supporting Information Material for details on the modifications to the code for adapting to our data). We quantified the SNR, the percent integral uniformity (PIU), percent signal ghosting (PSG), signal to fluctuation noise ratio (SFNR), drift, percent fluctuation, and percent signal change (PSC) for the acquired data (see Table 2).

High precision devices phantom

The second phantom which we use is the quantitative MRI system phantom from high precision devices (HPD; <http://hpd-online.com/system-phantom.php>). The phantom is placed as centrally as possible within the head coil and a notch in the mounting plate is used for further alignment (see Figure S2b). A single coronal slice covering the proton density (PD) array, consisting of 14 spheres with differing proton fractions (5–40 incrementing by 5% and 40–100 incrementing by 10%), is acquired using a SE sequence (slice thickness 6 mm,

TABLE 2 Summary of quality measures from (Vogelbacher et al., 2018)

Measure	Significance
Signal-to-noise ratio (SNR)	Quantification of dominance of meaningful signal over noise in the image
Percent integral uniformity	Describes uniformity of an image
Percent signal ghosting	Assessment of the amount of ghosting in the image
Signal-to-fluctuation-noise-ratio	Measure of relative temporal noise
Percent fluctuation Drift	Assessment of signal variance
Percent signal change	Assessment of homogeneity of SNR

TR = 5000 ms, TE = 10 ms, flip angle = 90°). The acquisition protocol is based on the manufacturer guidelines. Since the PD signal is expected to linearly increase with proton fraction; this property can be used as a proxy for scanner stability.

For each PD image, we detect PD circles (since the image is a 2D slice) using the MATLAB (<https://www.mathworks.com/>) image processing toolbox (currently used version R2019b). First, we detect the three fiducial markers on the 2D image and label them. Then, the detected PD circles are labeled based on their proximity to the fiducials. The detected boundary is eroded twice to ensure that only intensities within the circles are considered. An example of the detected circles is shown in Figure 1a and a stepwise depiction of the procedure is shown in Figure S5. Finally, we calculate the mean intensity within the circles and fit a linear regression model, explaining the mean PD signal as a function of proton fraction percentages. The percentage variance explained by the model, percentage R^2 , is then a summary measure of the linear increase of signal with increasing proton fraction (see Figure 1b).

2.3.2 | Schedule of QA runs

Till June 2019, HPD and agar gel phantom scans were being acquired approximately once per week. From July 2019, we started to acquire EPI data on agar gel phantom daily (8–9 a.m.; Table S6), and the PD acquisition on HPD phantom on a weekly basis (9–10 a.m.; Table S8). Therefore, an increased number of time points was available from July 2019 onwards for phantom data analyses (Figures 2 and 3). This decision to increase the number of phantom scans was made in accordance with the observations by Vogelbacher et al. (2018), in order to more closely monitor scanner performance and quality of data acquisition.

2.3.3 | Test–retest reliability analysis

Unlike phantom data, outcome measures on human data can vary based on factors like (but not limited to) scanner variations, physiological conditions, circadian variations, and so forth. Establishing the validity of outcome measures is challenging

given the lack of ground truth; however, reliability of these measures can be established by repeated measurements on the same subject.

Since, we are interested in ensuring reliability of outcome measurements over time, we perform repeated scans on two volunteers who are scanned on three consecutive days approximately in the middle of each month (henceforth referred as acquisition series). These series of scans can be used to establish the variation in quantitative outcome measures (such as brain volumes) in a short span of time (such as within an acquisition series) and over relatively longer span of time (such as across acquisition series). We would expect that even though outcome measures may change over time, the variation seen in them would remain comparable. As proof of concept, we present reliability analysis on T1-weighted and DWI scans on data from one volunteer (male, right-handed, 29 years old) for the time between July 2019 and December 2019 (three acquisitions per month, 18 total scans; see Table S9 for scan schedule). Between July and September, the volunteer was on oral medication for a medical condition; medication details are mentioned in Table S10. Specifically, we have examined whole brain, regional, and voxel-level reliability of various quantitative measurements across two acquisition series: July–September (acquisition series 1, nine scans) and October–December (acquisition series 2, nine scans).

T1-weighted scans

For quantifying the reliability of T1-weighted scans, we performed segmentation of T1-weighted images using SPM12 (v7487; <https://www.fil.ion.ucl.ac.uk/spm/>) running on MATLAB R2019b. Using the normalized, modulated gray matter (GM), WM, and CSF tissue class segmentation images, we performed test–retest reliability at whole brain, regional, and voxel-level.

Whole brain reliability. Using the normalized, modulated, GM, WM, and CSF tissue class segmentation images, we calculated whole brain GM, WM, CSF, and total intracranial volumes (TIVs); the TIV was defined as the sum of GM, WM, and CSF volumes. For each of the volumes, we calculated the percentage coefficient of variation (CV), defined as the ratio of standard deviation of tissue volume to the mean tissue volume, for each acquisition series. We then tested

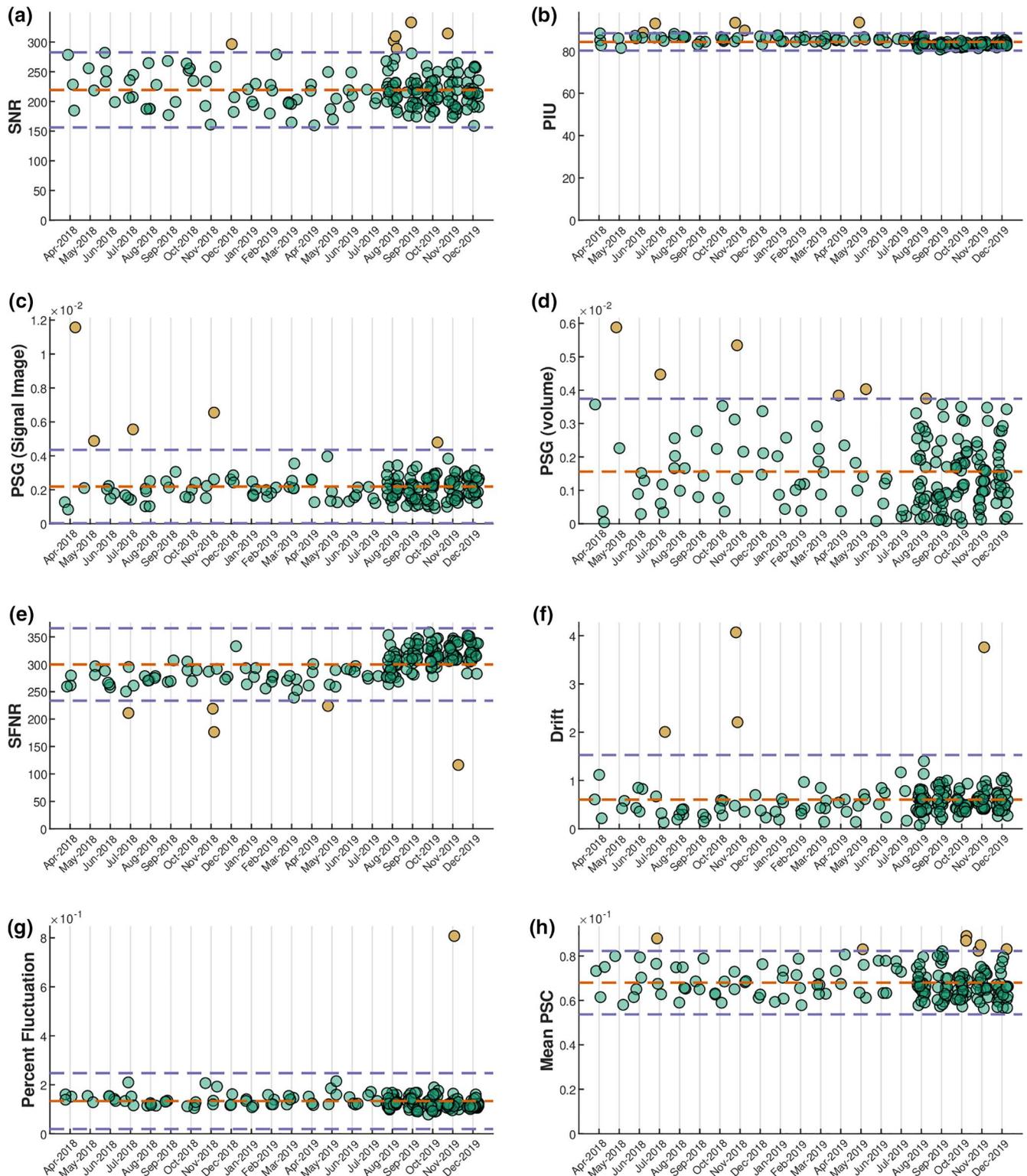


FIGURE 2 Variation of EPI quality measures on agar gel phantom (see Table S5 for a breakup of number of scans per month and Table S6 for scan schedule); the average value within each measurement is shown by a dashed orange line while mean \pm twice the standard deviation of the measurement is shown by a dashed purple line; outlier values are measurements beyond this range and are marked with orange colored circles; note that the lower mean \pm twice the standard deviation limit is not shown for PSG (volume) and drift as there were no values beyond this range. PIU, percent integral uniformity; PSC, percent signal change; PSG, percent signal ghosting; SFNR, signal-to-fluctuation-noise-ratio; SNR, signal-to-noise ratio

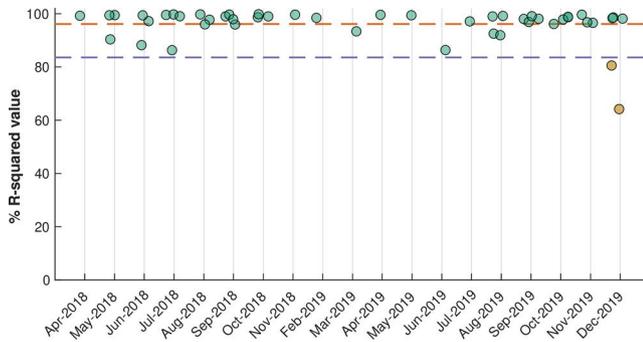


FIGURE 3 Percentage R^2 value from proton density images acquired on HPD phantom between April 2018 and December 2019 (see Table S6 for a breakup of number of scans per month and Table S7 for scan schedule); the average percentage R^2 is shown by a dashed orange line while mean \pm twice the standard deviation of the percentage R^2 is shown by a dashed purple line; outlier values beyond this range and are marked with orange-colored circles; note that the upper mean \pm twice the standard deviation limit is not shown. HPD, high precision devices

whether the CVs were significantly different between the two acquisition series (see below).

Region of interest-level reliability. Using the GM segmentation image (from above), we calculated the GM volumes of various regions of interest, for each acquisition series. These region of interests (ROIs) were defined using the Hammers Atlas (Faillenot et al., 2017; Gousias et al., 2008; Hammers et al., 2003); we excluded the following non-GM ROIs from this calculation: brainstem, corpus callosum, lateral ventricles (and its temporal horn), and the third ventricles. We then calculated the regional GM CVs and tested whether they were significantly different between the two acquisition series (see below).

Voxel-level reliability. Similar to whole brain and regional reliability, we calculated voxel-wise CV for GM segmentation images. We used the GM segmentation images and restricted the CV calculation to only GM voxels by applying a threshold of 0.1 (i.e., any voxel with a value less than or equal to 0.1 was not considered). This threshold is commonly used in voxel-based morphometry analyses; additionally, we wanted to be as inclusive of the GM voxels as possible. We then used the resulting CV images to test whether the voxel-wise CVs were significantly different between the acquisition series (see below).

Hypothesis testing of CVs. As mentioned before, the goal of test-retest reliability is to test that the variation seen in quantitative variables across the two acquisition series remains the same. Therefore, we tested the hypothesis that there was no difference in CVs of quantitative variables across the two acquisition series. First, we quantified the absolute difference in CVs between the two series (CV_{diff}). Then, we randomized the series labels. The maximum number of ways that this can be done is 48,620 (including the original series membership). For each of these arrangements of permuted series labels, we

computed the new CV. The two-sided p -value was calculated as the fraction of times the absolute difference in permuted CV became equal to or exceeded the originally observed difference CV_{diff} . We then rejected the null hypothesis at $\alpha < 0.05$ at whole brain level, $\alpha < 0.05/76$ at regional level (Bonferroni adjusted for number of ROIs, $n = 76$), and $\alpha < 0.001$ at voxel-level.

Segmentation consistency. In addition to examining the variability in quantitative estimates, we were interested in examining the consistency of voxels being classified as a particular tissue type. Since the data is from the same volunteer, all tissue specific segmentation images should have a high degree of overlap with each other. To quantify this overlap, we calculated the pairwise dice coefficient (DC) between the segmentation images (for each tissue class), which would reveal the consistency of segmentation within and across acquisition series. Since the goal is to measure the whole brain overlap in segmentation (for each tissue class; e.g., overlap of GM segmentation of one scan with GM segmentation of another), we binarized (i.e., converted the intensities above the threshold to one and set the rest of them to zero) the GM, WM, and CSF segmentation using a liberal threshold of 0.1; then for each pair of images (18 total images, 153 pairwise combinations), we calculated DC for each tissue class.

Diffusion scans

For quantifying the reliability of diffusion scans, we used FSL 6.0.1 (fsl.fmrib.ox.ac.uk/) to perform distortion, eddy current, and motion correction as described in QC section (see below). We then fit a weighted least squares (WLS) tensor model and obtained fractional anisotropy (FA), radial diffusivity (RD), axial diffusivity (AD), and mean diffusivity (MD) maps for each image, followed by tract-based spatial statistics (TBSS) pipeline (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS>) to obtain the mean and skeleton image which was used for further calculations. Specifically, we used TBSS steps 1 and 2 to nonlinearly register the FA images into standard space (using FSL's nonlinear registration tool FNIRT (Andersson et al., 2007a, 2007b)), followed by TBSS step 3 to create mean and skeletonized FA images, and then used TBSS step 4 to project all FA images onto the mean skeleton image. We then used the non-FA pipeline of TBSS to apply the nonlinear transforms and projection to RD, AD, and MD images.

Whole brain reliability. From the skeletonized images (from above), we calculated the average FA, AD, MD, and RD values across the entire skeleton. For each acquisition series, we then calculated the CV of FA, AD, MD, and RD and tested the hypothesis whether the CVs of any of these whole brain average values were significantly different between the two acquisition series (see below).

Region of interest-level reliability. We defined 48 WM ROIs using the Johns Hopkins University (JHU) WM labels Atlas (Mori et al., 2008; Oishi et al., 2008) shipped with FSL. From each of these ROIs, we summarized the average FA, AD, MD, and RD values by masking with the skeletonized image from the TBSS pipeline. The averages were

calculated only from voxels which belonged to the skeleton. For both the acquisition series, we then calculated the CV of each of the diffusion measures for each ROI. We tested the hypothesis whether these ROI measures were significantly different between the two acquisition series (see below).

Voxel-level reliability. To quantify voxel-level reliability, we calculated the voxel-level CVs for FA, AD, MD, and RD skeletonized images for both acquisition series. Then, we tested the hypothesis whether these CVs were significantly different between the two series for any of the voxels (see below).

Hypothesis testing of CVs. To hypothesis testing procedure for diffusion images was similar to the one described for T1-weighted scans. We rejected the null hypothesis at $\alpha < 0.05$ at whole brain level, $\alpha < 0.05/48$ at regional level (Bonferroni adjusted for number of ROIs, $n = 48$), and $\alpha < 0.001$ at voxel-level.

2.4 | Quality check

The following section describes the QC pipeline for structural, functional, and diffusion images; the software versions used to implement these pipelines are noted at the end of this section.

2.4.1 | Structural pipeline

Each structural image is reviewed by a neuroradiologist and a report is generated. If any significant structural abnormality is found, the study participant is flagged for further clinical evaluation. Following this, we set the origin of the images to the anterior commissure (AC). This step is done using `acpcdetect v2` (Ardekani, 2018; Ardekani & Bachman, 2009; Ardekani et al., 1997) (part of Automatic Registration Toolbox) (<https://www.nitrc.org/projects/art>). In case this step fails, we manually set the origin using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) running on MATLAB.

Computational anatomy toolbox and MRIQC

All T1-weighted images (with origin set to the AC) are segmented using the Computational Anatomy Toolbox (CAT) (<http://www.neuro.uni-jena.de/cat/>) (running on MATLAB). The QC module of CAT returns an image quality grade (A+, A, A-, to F). Any image receiving a rating of "C" or below is flagged. We also use automated quality prediction and visual reporting of MRI scans by running the docker version of MRIQC (Esteban et al., 2017) on structural scans which generates a detailed visual and quantitative report for each image. Any image which is an outlier (examined by box-plot) in any of the quality measurement is visually examined for suitability for further analysis. Additionally, we use the pre-trained random forest classifier of MRIQC to predict if a T1-weighted image is of good quality or not.

2.4.2 | Functional pipeline

Minimal preprocessing

The minimal preprocessing pipeline for functional images is implemented in SPM running on MATLAB and consists of motion correction (realign and unwarp), followed by slice timing correction. Then, the functional images are AC-posterior commissure (PC) aligned (using the transforms from the T1-weighted AC-PC detection step). The mean functional image is then coregistered to the AC-PC aligned T1-weighted image and the estimated transforms are applied to all functional volumes. Then, the T1-weighted scan is segmented, and a deformation field is estimated which nonlinearly warps the brain from native space to the Montreal Neurological Institute (MNI) space. This forward deformation field is applied to the coregistered functional scans to warp functional images to the MNI space. The bounding box during normalization step is calculated to allow full brain coverage (rather than using SPM's default bounding box). Additionally, a smoothing of 6 mm full-width at half-maximum (FWHM) is applied to the normalized functional images.

Motion profile

At each stage of preprocessing of functional data, we estimate the number of time points that can be considered as motion outliers. This is done using a modification of the DVARS (D referring to temporal derivative of time courses, VARS referring to root-mean-squared variance over voxels) approach (Afyouni & Nichols, 2018). A DVARS p -value is computed which can be used for declaring a pair of time points as outliers. Additionally, a liberal user-defined threshold is also applied to detect outliers. An example plot is shown in Figure S18.

Brainmask profile

Brain regions such as the orbitofrontal cortex (OFC) are susceptible to signal loss. To quantify signal loss, we eliminate all voxels which have a mean signal less than 80% of global signal (GS; similar to an approach used by SPM). We calculate this using `spm_global` function which first eliminates voxels which have signal less than 80% of the whole brain mean signal; then returns the mean signal from the remaining voxels. This operation is performed on smoothed functional images (from the minimal preprocessing step). We create a binary mask of the voxels which survive this threshold and examine them visually to identify subjects in which the mask is of poor quality. We create this mask using `spm_mask` function by passing the smoothed images and a whole brain mask as inputs, along with the previously estimated GS. An example of these brain mask profiles are shown in Figure S19b. Additionally, we calculate the ratio of voxels remaining in the mask and the template mask (shipped with SPM). Subjects with drastically reduced proportion of voxels can then be flagged using a box-plot method.

Signal profile and left-right correlation

To flag images with severe signal loss, we derive mean time series from regions that are susceptible to signal loss, regions not susceptible to signal loss, GS from left hemisphere, right hemisphere, and average GS. These regions are defined using the Hammers Atlas

(Faillenot et al., 2017; Gousias et al., 2008; Hammers et al., 2003) and comprise of bilateral orbitofrontal gyri, gyrus rectus, superior temporal gyrus, and the lateral occipital lobe. The mean percentage difference can be calculated from each region with respect to the GS. Any subject with a significant reduction in signal with this reference signal will be flagged and eliminated. Additionally, looking at GS independently from left and right hemispheres can be useful in diagnosing any hemispheric signal abnormalities which might happen during acquisition (such as a bad channel in the head coil). We further quantify the correlation between left and right hemispheres and outliers are detected using a box-plot method. An example of this signal profile is shown in Figure S19a.

MRIQC

Apart from the above-mentioned methods of assessing the quality of functional images, we also run the docker version of MRIQC for all functional images. For each functional scan, we look at the distribution of various quality measurements and flag any images which are potential outliers (using a box-plot method), especially in terms of number of functional time points identified as outliers for each image.

2.4.3 | Diffusion pipeline

The diffusion pipeline is implemented in FSL (<https://fsl.fmrib.ox.ac.uk>). Briefly, the pipeline consists of distortion correction, eddy current and motion correction, subject and group-level quality assessment, and tensor fitting operation.

Distortion correction

We correct EPI-induced distortions by using opposite phase encoding directions reference scans. These non-diffusion scans are passed as an input to FSL's *topup* (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/topup>) to estimate susceptibility-induced off-resonance field using a method similar to that described in Andersson et al. (2003) as implemented in FSL (Smith et al., 2004) and the two images are combined into a single corrected one. Once the images are distortion corrected, we derive a binary brain mask using the FSL brain extraction toolbox (BET). We have noticed that a fractional intensity threshold of 0.35 with robust BET usually results in brain extraction but this threshold can vary depending on the input image.

Eddy current and motion correction

We employ FSL's *eddy* (Andersson & Sotiropoulos, 2016) to simultaneously correct eddy current and motion. The previously calculated brain mask is passed to *eddy* along with information on which direction distortion is expected, the *b*-vector and *b*-value files, and the estimated field from *topup*. Additionally, we use the "repol" option (Andersson et al., 2016) to replace volumes or slices with partial or complete signal loss (due to movement) with their Gaussian process predictions. Finally, we also use the "mporder" option (Andersson et al., 2017) to find and replace those slices (within a volume) which are corrupted by motion. We are using the following *eddy* settings:

niter = 8 (number of iterations to run), fwhm = 10,8,4,2,0,0,0, (full width at half maximum of the Gaussian filter used during estimation of distortion), ol_type = both (slice and group wise outlier detection), mporder = 6 (slice-to-volume motion correction parameter); the remaining parameters are default settings. These parameters are based on the recommendations on FSL's eddy wiki (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy>).

Subject and group level quality assessment

Once the data is distortion-, eddy current-, and motion-corrected, we use FSL's eddy's QUAD (Bastiani et al., 2019) module for quality assessment. This subject level visual and quantitative report that is generated, derives QC measures based on volume-to-volume motion, within-volume motion, eddy current-induced distortion, susceptibility induced distortion, and the number of outliers replaced. SNR and CNR ratios are additionally calculated and incorporated in the subject level report. Once QC reports at subject level are ready, group-level reports are generated/updated by using FSL's eddy's SQUAD tool.

Tensor fitting

Once the data are distortion-, eddy current-, and motion-corrected (and the *b*-vectors rotated), we fit a WLS tensor model to the data using FSL dtfit. Voxel-wise sum of squared errors is also saved, and the principle diffusion direction is visualized by overlaying on the FA map.

2.4.4 | Software versions for QC

The analyses pipelines requiring MATLAB have been setup using MATLAB R2019b (<https://www.mathworks.com/>) with SPM v7487; the CAT toolbox (<http://www.neuro.uni-jena.de/cat/>) currently being used is v1450 running on MATLAB R2019b with SPM v7487. The container version of MRIQC (<https://mriqc.readthedocs.io/en/latest/>) used is 0.14.2. The diffusion pipeline was implemented in FSL (<https://fsl.fmrib.ox.ac.uk>) 6.0.1; BET version 2.1, CUDA version 9.1, QUAD and SQUAD versions 1.0.2. We note that the QC methods described above are not bound to a particular software version and are periodically updated as new versions get released.

3 | RESULTS

3.1 | Phantom QA

3.1.1 | Agar gel phantom

For agar gel phantom, we calculated various measurements like SNR, PIU, PSG, SFNR, drift, percent fluctuation, and PSC, which quantify the spatial and temporal characteristics of the EPI images. Descriptive statistics of these measures for 172 time points collected between April 2018 and December 2019 presented in

Table 3 (see Tables S5 and S6 for scan schedule) and Figure 2 shows the variation in these measurements over time. Outliers in the phantom measurements were identified as two standard deviations away from the mean measurement over time. We noticed a few outliers in each measurement category. The outliers in the ghosting category can be explained by an error in the automatic ROI placement during the ghosting calculation and skewed placement of the phantom within the head coil (see Figure S6 for a few examples). We found that four phantom measurements (three in 2018 and one in 2019) showed a different intensity profile than the rest of the measurements, characterized by lower intensities in the center of the phantom (see Figure S7). This has led to outliers in various measurements such as SFNR, PIU, and drift; we could not find any obvious reason for these fluctuations, and therefore need to closely examine the human data acquired on these days. We discovered a fault in the head coil during one of our scans in November 2019 which resulted in altered SFNR, drift, percent fluctuation, and mean PSC values (a representative time series of this scan is shown in Figure S8 which shows a sharp signal change indicative of the malfunctioning head coil). For six measurements, the overall SNR is higher than average; since the SNR is measured as the ratio of mean signal to the standard deviation of the signal in the centrally placed ROI (Friedman & Glover, 2006; Friedman et al., 2006), this could indicate more uniformity in the central region of the phantom for those time points, thereby leading to increased SNR. For the remaining six outliers, we have not found any obvious visual or processing problems. These fluctuations could have been caused by several other factors like phantom orientation, slice selection, and temperature and humidity during the time of acquisition, although it is not possible to conclusively identify the causative factor for these variations.

3.1.2 | HPD phantom

For HPD phantom (PD images), we calculated the percentage R^2 value for each time point and examined this value over time. We have shown the fitted percentage R^2 value for 48 time points between April 2018 and December 2019 in Figure 3 (see Tables S7 and S8 for scan schedule). We noticed two outliers in December 2019. On visual examination of these scans, we observed that the central region of the images was darker and therefore the mean intensities in the central spheres was lower. This resulted in a poorer R^2 value, as compared to other time points (see Figure S9). This could be caused due to errors in phantom handling or an incorrect slice being selected during the acquisition.

3.2 | Test-retest reliability analysis

A summary of MRIQC measures on T1-weighted images and SQUAD summary on DWI are shown in Figures S10 and S11 respectively, while the test-retest reliability results are presented below.

3.2.1 | T1-weighted scans

Whole brain reliability

At the whole brain level, the CVs of GM, WM, CSF, and TIV were not significantly different between the two acquisition series. The TIV showed a trend level significance with p -value 0.06; since the TIV is the sum of GM, WM, and CSF, the CV of TIV would be influenced by the cumulative variances in each of these measures, thereby leading to a trend level p -value. Additionally, we note that these p -values are unadjusted for multiple comparisons. A summary of the whole brain reliability is shown in Table 4.

Region of interest-level reliability

At an ROI level, the CV of GM volumes of left fusiform gyrus (p -value 0.03), left posterior temporal lobe (p -value 0.03), right inferiolateral remainder of the parietal lobe (p -value 0.05), left precentral gyrus (p -value 0.02), and left presubgenual frontal cortex (p -value 0.03) were significantly different; however, none of the CVs were significant after correcting for multiple comparisons for the number of ROIs (Bonferroni adjustment). A summary plot showing the absolute difference in the CV between acquisition series 1 and 2 (for left and right hemisphere ROIs) is shown in Figure 4 and the absolute difference in CV values and associated p -values are listed in Table S11. We additionally note that the difference in the CV of GM in substantia nigra and pallidum (left and right hemispheres) are relatively on the higher side. This is explained by the relatively poor overlap of the substantia nigra and pallidum and the GM segmentation (see Figure S14 and S15).

Voxel-level reliability

At the voxel-level, we created a p -value map after eliminating voxels with $p < 0.001$. On visual inspection of this map, we did not find any clusters of significantly different CVs (see Figure S12 for voxel-wise absolute difference in CVs). We found a total of 394 voxels spread over multiple ROIs (labeled using Hammers Atlas). To summarize these results, we counted the number of voxels in each ROI with $p < 0.001$. These summary counts are presented in Figure 5 while an overlay of these significant voxels are shown on a template brain in Figure S13.

Consistency of segmentation

Examining the pairwise DC for each tissue class, (153 pairs of images, 9 acquisitions per series), we observed excellent agreement between the segmentation images. The mean and standard deviations of DC were 0.97 ± 0.001 (GM), 0.96 ± 0.001 (WM), and 0.95 ± 0.003 (CSF). A summary of each pairwise DC for each tissue class is shown in Figure 6. The DC of all tissue segmentations were above 0.95 for all pairs of images (except one pair with CSF DC = 0.94).

3.2.2 | Diffusion scans

Whole brain reliability

At the whole brain level, the CVs of FA, MD, RD, and AD were not significantly different between the two acquisition series. The CV of

Measure	Mean	Standard deviation	Min	Max
SNR	219.4163	31.6091	158.8507	333.1398
Mean PIU2	84.2961	2.0638	80.7046	93.5360
PSG _{signalimage} ^a	0.0022	0.0011	0.0008	0.0116
PSG _{volume} ^b	0.0016	0.0011	0.0000	0.0059
SFNR	299.6195	33.0544	116.2972	358.1444
Drift	0.6043	0.4615	0.0753	4.0682
Percent fluctuation	0.1335	0.0570	0.0786	0.8075
Mean PSC	0.0680	0.0071	0.0564	0.0890

Abbreviations: EPI, echo-planar imaging; PIU, percent integral uniformity; PSC, percent signal change; PSG, percent signal ghosting; SFNR, signal-to-fluctuation noise ratio; SNR, signal-to-noise ratio.

^aSlice (mean_PSG).

^bVolume (PSG2).

TABLE 3 Descriptive summary of EPI quality check on agar gel phantom (see text for details)

Measurement	CV Series 1 (%)	CV Series 2 (%)	CV _{diff} (%)	p-Value
Total GM volumes	0.5964	0.8234	0.2270	0.4272
Total WM volumes	0.5815	0.5698	0.0118	0.9210
Total CSF volumes	1.0910	1.5059	0.4149	0.5917
TIV	0.1032	0.2101	0.1069	0.0564

Note: Series 1 refers to nine scans on the same subject between July and September 2019 and Series 2 refers to nine scans on the same subject between October and December 2019 (three scans per month) (see Table S9 for scan schedule).

Abbreviations: CSF, cerebrospinal fluid; GM, gray matter; TIV, total intracranial volume; WM, white matter.

TABLE 4 Coefficient of variation (CV) and *p*-values for whole brain volumetric measurements from T1-weighted scans for test-retest reliability

MD showed a trend level significance ($p = 0.08$); MD $\left(\frac{\lambda_1 + \lambda_2 + \lambda_3}{3}\right)$ is derived from AD (λ_1) and RD $\left(\frac{\lambda_2 + \lambda_3}{2}\right)$, therefore the CV of MD would be influenced by the cumulative variances in both these measures, thereby possibly leading to a trend level *p*-value. Additionally, we note that these *p*-values are unadjusted for multiple comparisons. A summary of the whole brain reliability of diffusion scans is shown in Table 5.

Region of interest-level reliability

At the ROI-level, the FA of the right anterior corona radiata ($p = 0.04$) and the left cingulum (cingulate gyrus) ($p = 0.05$), the AD of the right corticospinal tract ($p = 0.003$) and the left posterior thalamic radiation (including optic radiation) ($p = 0.02$), the RD of the left inferior cerebellar peduncle ($p = 0.02$), the left anterior corona radiata ($p = 0.04$), the right uncinate fasciculus ($p = 0.01$), and the left uncinate fasciculus ($p = 0.01$), and the MD of the right corticospinal tract ($p = 0.01$) and the left anterior corona radiata ($p = 0.03$), were statistically significant; however, none of these values survived a Bonferroni correction for the number of ROIs. A summary plot showing the absolute difference in the CV between acquisition series 1 and 2 (for each of the diffusion measures) is shown in Figure 7 and the absolute difference in CV values and associated *p*-values are listed in Table S13. Additionally, we note that the difference in CV of

left tapetum was relatively on the higher side; this is because the left tapetum has minimal overlap with the mean FA skeleton from which the diffusion measures are summarized (see Figure S17).

Voxel-level reliability

At the voxel-level, we created a *p*-value map after eliminating voxels with $p < 0.001$ for each of the diffusion measures. On visual inspection of these maps, we did not find any clusters of significantly different CVs. To summarize these results, we counted the number of voxels in each JHU WM label ROIs and found that the CVs of 124 voxels for AD, 223 voxels for RD, 161 voxels for MD, and 114 voxels for FA were significantly different between the two acquisition series at uncorrected $\alpha < 0.001$. These summary counts are presented in Figure 8 while an overlay of these significant voxels are shown on the mean FA image in Figure S16.

4 | DISCUSSION

In this paper we have presented the QA and QC pipeline for MRI studies of the ADBS project. In the literature, various QA methods have been proposed to monitor scanner stability and to ensure reliability of quantitative measurements. Measurements like SNR, SFNR,

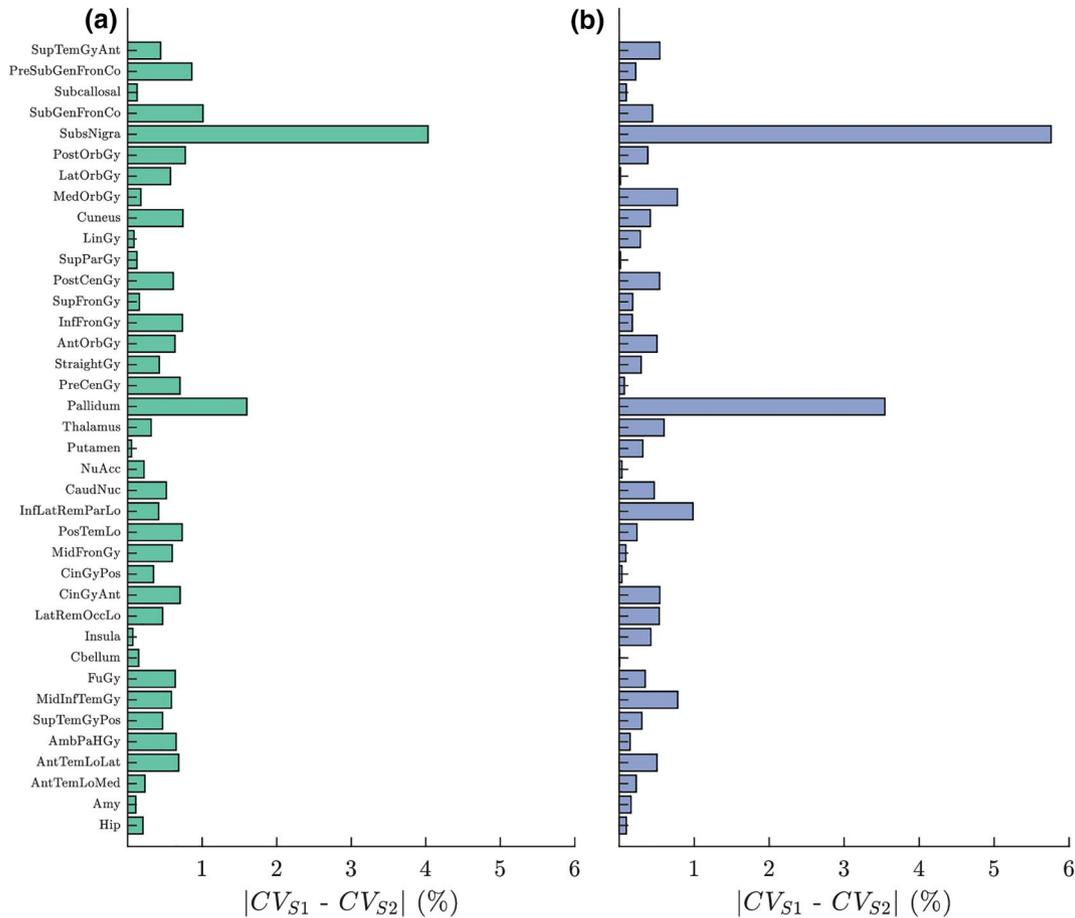


FIGURE 4 Absolute percentage difference in coefficient of variation (CV) of GM volumes between acquisition Series 1 (S1) and Series 2 (S2), for regions of interest from Hammers Atlas for test retest reliability data; (a) regions from the left hemisphere, and (b) regions from the right hemisphere; see Table S11 for full names of the brain regions. GM, gray matter

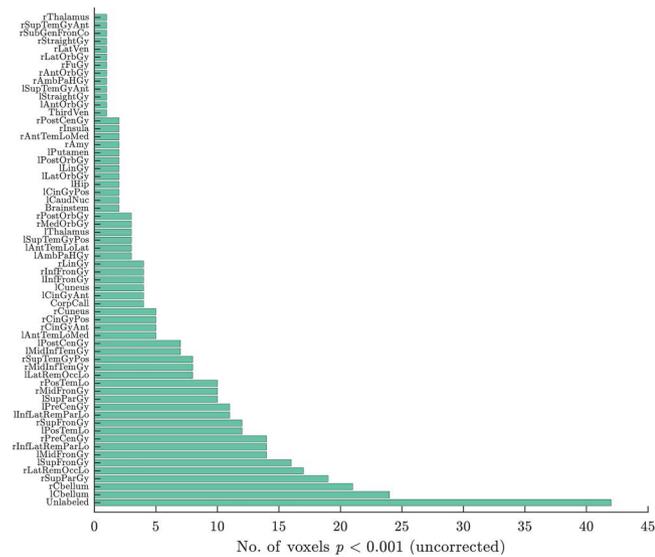


FIGURE 5 Region-wise counts of the number of voxels which had significantly different coefficient of variation of GM between acquisition Series 1 and Series 2 at $\alpha < 0.001$ (uncorrected); see Table S11 for full names of the brain regions. GM, gray matter

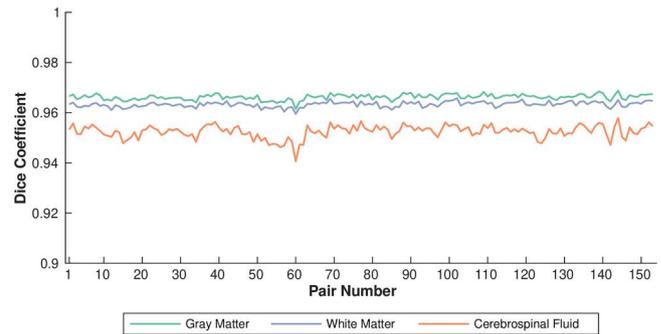


FIGURE 6 Pairwise Dice coefficient for gray matter (green color), white matter (purple line), and cerebrospinal fluid (orange color); between July and December 2019, a volunteer underwent three scans per month (total 18 scans); Dice coefficient between the segmentation of the three tissue classes was calculated for all pairs of images (total 153 pairs) to examine the consistency of segmentation (see Table S9 for details of reliability schedule)

drift, and percent fluctuation (Friedman & Glover, 2006; Friedman et al., 2006), PSC (Stöcker et al., 2005) and PIU and PSG (The American College of Radiology, 2005), and so forth, can be used on

Measurement	CV Series 1 (%)	CV Series 2 (%)	CV _{diff} (%)	p-value
Fractional anisotropy	0.5388	0.4224	0.1164	0.4485
Axial diffusivity	0.4988	0.3220	0.1768	0.1186
Radial diffusivity	0.8355	0.5157	0.3198	0.1132
Mean diffusivity	0.6529	0.3824	0.2705	0.0798

Note: Series 1 refers to nine scans on the same subject between July and September 2019 and Series 2 refers to nine scans on the same subject between October and December 2019 (three scans per month) (see Table S9 for scan schedule).

TABLE 5 Coefficient of variation (CV) and *p*-values for whole brain average diffusivity measurements from diffusion weighted scans for test-retest reliability

phantom or acquired human data for the identification of scanner instabilities or time points with large fluctuations. Further, test-retest analysis on acquired human data can reveal the reliability of quantitative outcome variables (for example [Maclaren et al., 2014], examined CV of regional brain volumes over data from three subjects; [Madan & Kensinger, 2017] examined reliability of volumes and surface features using two open-access datasets; [McGuire et al., 2017] worked on three assessments for 25 subjects using multimodal brain imaging data). Test-retest approaches show that tracking of outcome variables over repeated assessments can not only highlight the natural biological variability but also serve as a measure for consistent scanner performance. Over and above these, various subject-level QC processes can be employed for identifying and flagging images that could potentially influence data analysis. These steps could involve identification of incidental structural abnormalities, identification of MRI artifacts (for example, using MRIQC [Esteban et al., 2017]), data points with high amount of motion (by methods like framewise displacement and DVARS for fMRI [Afyouni & Nichols, 2018; Power et al., 2012], and volume and slice-level motion correction using eddy for DWI [Andersson et al., 2017, 2016]), and so forth.

4.1 | QA approach of the ADBS project

4.1.1 | Phantom measurements

For monitoring scanner performance, we have used data from two phantoms: PD images from HPD phantom and EPI BOLD images from an agar gel phantom. The results from the PD images show a general trend of increasing PD values with increasing proton fraction, as indicated by high percentage R^2 values (see Figure 3), with the exception of two time points where the acquired images exhibited a central darkening. We believe this could be caused due to errors in phantom handling or an incorrect slice being selected during the acquisition. The results from agar gel phantom images show a general trend of lower ghosting, drift, and fluctuations, with the exception of a few outliers. These are explained by incorrect ROI placement during the analysis step, issues related to phantom handling, and a head coil problem. For outliers with no obvious explanation, we need to closely examine the already acquired human data from those dates and ensure their usability. Vogelbacher et al. (2018) have shown a

strong dependence of QA measurements on the positioning of the phantom, and suggest that the only way to alleviate this is by using a phantom holder. We are in the process of implementing this recommendation at our site to further improve the consistency of our phantom measurements. Given differences in scanners, acquisition protocol, types of phantom, and other reasons, it is not straightforward to establish a range for these QA measurements. Therefore, steps to integrate an automated warning system to promptly detect scanner instabilities are currently underway. This is similar to the idea of LAB-QA2GO (Vogelbacher et al., 2019) but incorporating modifications relevant to our site and phantoms (see Supporting Information Material). Since standardizing various QA measurements across sites and scanners is not straightforward, the warning system will use a distribution of QA measurements from our site (reported in this paper) and check a new measurement against this distribution; if the new measurement is an outlier, the system will generate a warning (the cause of which can be investigated by the site technicians). As measurements accumulate, it should be possible to train a machine learning classifier to predict scanner anomalies.

4.1.2 | Test-retest reliability on human volunteer

In our approach to test-retest reliability, we have analyzed 18 repeated scans of a human volunteer, over a period of 6 months. We compared the global, regional, and voxel-level CVs of quantitative measures from T1w and DWI images between the first and last three months. We did not find any significant differences between the CVs at global and local levels. Additionally, examination of DC shows the consistency of segmentation at short-term (within series) and long-term (between series) basis. These results lead us to believe that the structural and diffusion images that we are acquiring have the potential to yield consistent volumes and diffusivity measures. As the scanner-related variation remains consistent, the acquired images could potentially be sensitive enough to reveal differences in longitudinal scans.

4.2 | QC approach of the ADBS project

The ADBS QC approach is based on automated and semi-automated tools that can help identify and flag problematic images, the analysis

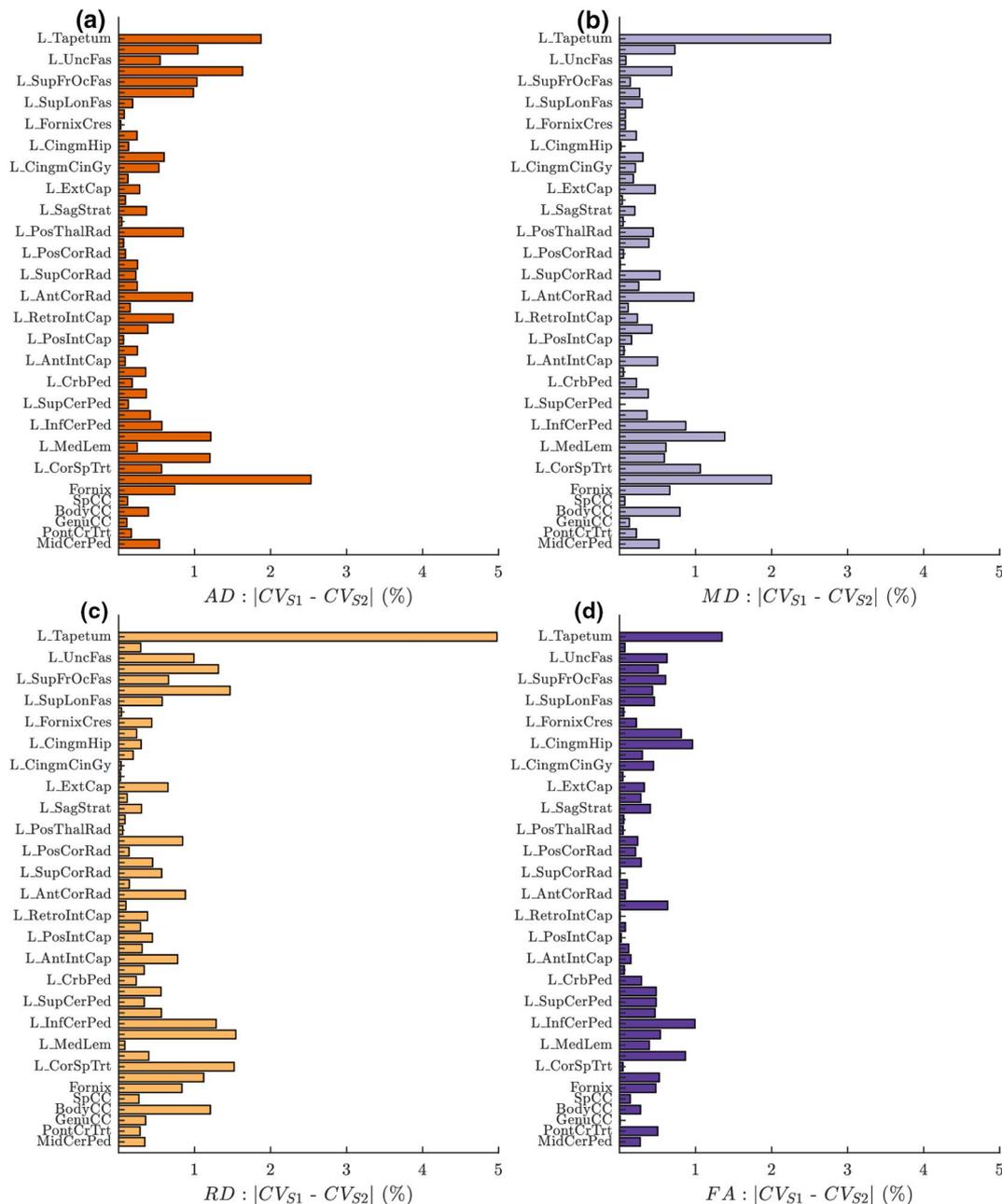


FIGURE 7 Absolute percentage difference in coefficient of variation (CV) of (a) axial diffusivity (AD), (b) mean diffusivity (MD), (c) radial diffusivity (RD), and (d) fractional anisotropy (FA) between acquisition Series 1 (S1) and Series 2 (S2), for regions of interest from JHU white matter labels Atlas for test-retest reliability data; see Table S12 for full names of the brain regions; unlabeled ticks are right hemisphere regions

of which could lead to incorrect interpretation. The QC procedure aims at identifying outliers for each imaging modality (structural, functional, and diffusion) by using state of the art tools such as MRIQC (Esteban et al., 2017) and QUAD/SQUAD (Bastiani et al., 2019), along with in-house workflows which provide additional information about the quality of acquired scans. Specifically, for structural images, each acquired image is reviewed by a trained neuroradiologist to flag any incidental finding(s). In addition, artefacts and images of low quality are identified using MRIQC (Esteban et al., 2017) and CAT (<http://www.neuro.uni-jena.de/cat/>) reports. In

addition, we visually examine any structural image which is flagged as an outlier. For functional images, we comprehensively examine the motion profile of each image using MRIQC (Esteban et al., 2017) and DVARS (Afyouni & Nichols, 2018), along with quantifying the degree of signal loss within susceptibility regions, visually examining the level of distortion by brain mask profile, and calculating left-right correlation for identifying images with signal anomalies using in-house workflows. Finally, for diffusion images, we use eddy for volume and slice-level motion correction and use QUAD/SQUAD (Bastiani et al., 2019) for identification of problematic images.

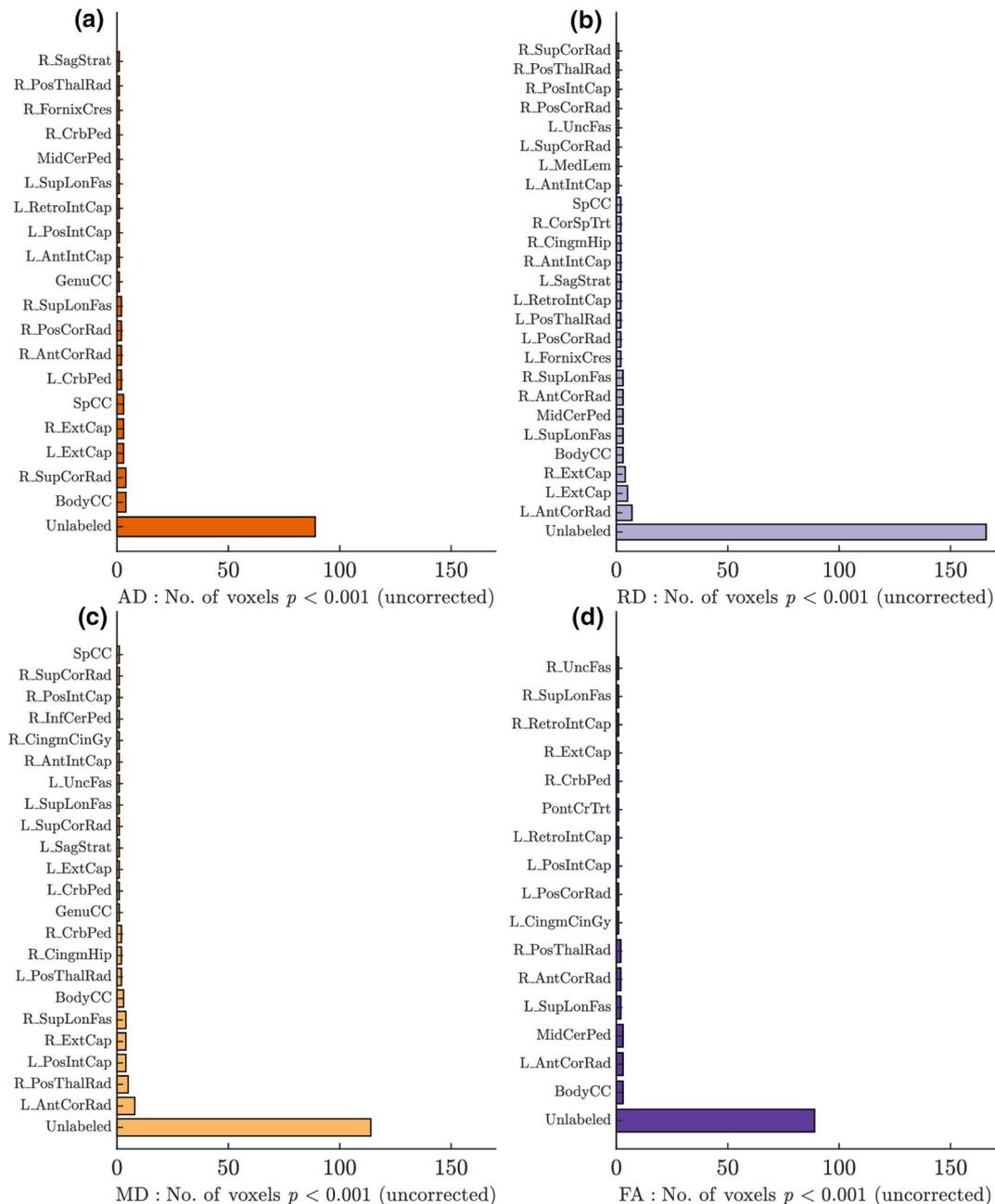


FIGURE 8 Region-wise counts of the number of voxels which had significantly different coefficient of variation of (a) axial diffusivity (AD), (b) mean diffusivity (MD), (c) radial diffusivity (RD), and (d) fractional anisotropy (FA) between acquisition Series 1 and Series 2 at $\alpha < 0.001$ (uncorrected); see Table S12 for full names of the brain regions

4.3 | Conclusions and future directions

The establishment and implementation of strong QA–QC procedures is critical in neuroimaging studies as exemplified by the success of studies like the ADNI (Jack et al., 2008; <http://adni.loni.usc.edu>) and HCP (Van Essen et al., 2013; <https://neuroscienceblueprint.nih.gov/human-connectome/connectome-programs>). It is important to emphasize that monitoring of both QA and QC measures is critical for any study. Having only a good QC protocol does not necessarily ensure that the quantitative outcome variables will be reliable as the QC protocol does not capture the repeatability value of the

acquisition. Additionally, QC protocols may not be able to capture scanner related instabilities. Similarly, having only a good QA protocol does not necessarily ensure that the acquired scans will be acceptable for quantitative analyses because of factors like subject-level motion, incidental structural abnormalities that could interfere with image processing steps, and so forth. In the ADBS project, we have implemented rigorous QA as well as QC steps, which sets a precedent for other studies to adapt and build on (a summary of these QA–QC steps is presented in Figure 9).

As the ADBS data acquisition relies on a single site/scanner, we do not face intersite/interscanner variability as a factor (unlike

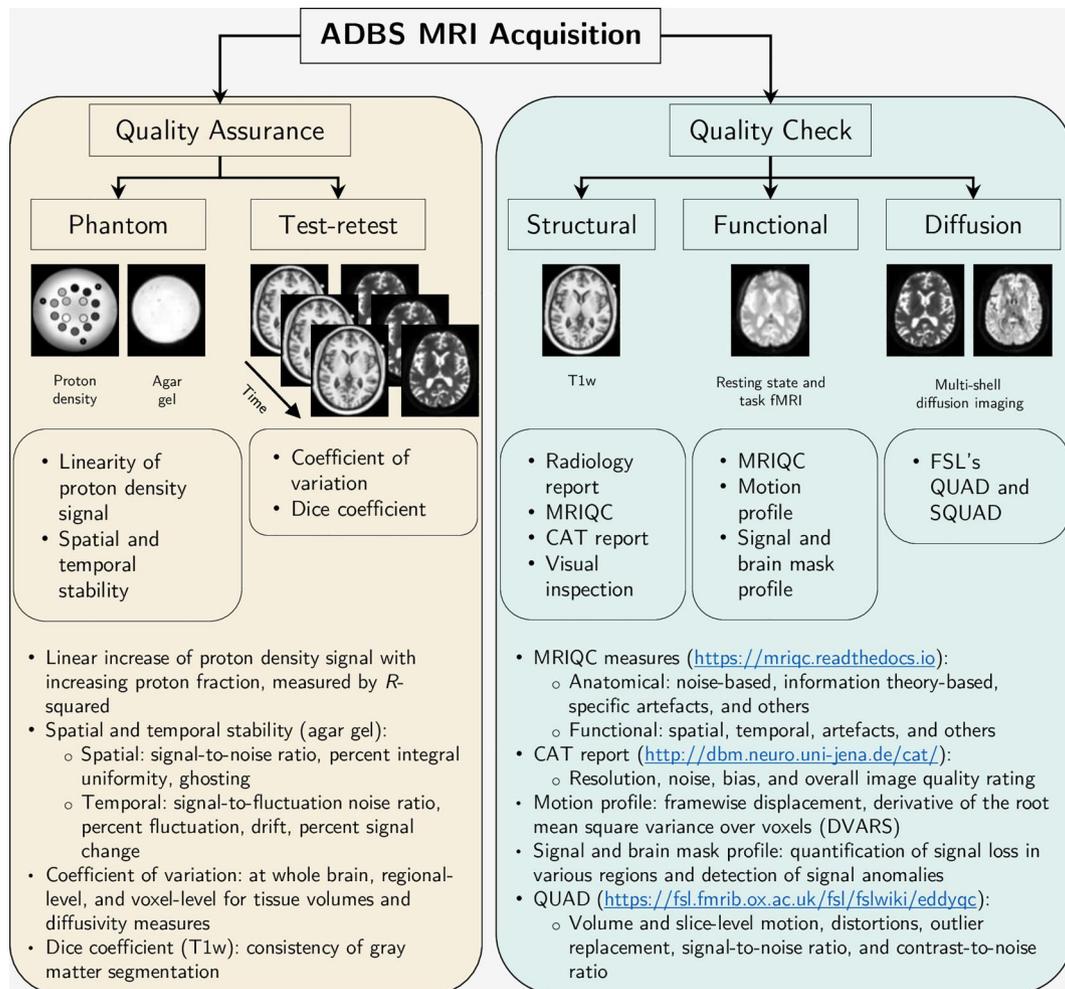


FIGURE 9 Summary of the quality assurance and quality check pipeline being followed in the Accelerator program for Discovery in Brain disorders using Stem cells project

multisite studies). This also allows for a stricter control on variability caused by human factors (e.g., planning of an acquisition), time window during which acquisition happen, and the incorporation of methods to ensure scanner stability. One of the key aspects that we have focused here is ensuring scanner stability by using different MRI phantoms and incorporating test-retest reliability on human volunteers. As the study progresses, it will be necessary to update some of these QA-QC measures to incorporate the longitudinal nature of the dataset. These methods would also need to be periodically revised as newer methods get proposed and adapted in the field. Currently, we have adapted methods in the field which require little to no manual intervention. With the influx of data, it will be possible to develop/incorporate machine learning methods for classifying problematic images (we note that the Alfaro-Almagro et al. (2018) have used a similar idea for identifying problematic T1-weighted images). We hope to curate a high quality neuroimaging database that can serve the scientific community and contribute to understanding the neurobiology of various psychiatric disorders. We also expect that the pipeline for rigorous quality control and reliability measures that are detailed in this manuscript would benefit researchers while

planning similar longitudinal neuroimaging studies in various neuropsychiatric disorders.

ACKNOWLEDGEMENTS

This work is supported by the Department of Biotechnology (DBT), Government of India [BT/PR17316/MED/31/326/2015]; phantom scan data presented in this article were acquired as part of two funded projects: BT/PR17316/MED/31/326/2015 (DBT, Government of India) and RO1 MH113250 (National Institute of Mental Health in the United States). The authors thank the magnetic resonance imaging (MRI) technicians, Ms. Manasa Venkat and Mr. Vinod Kumar for their assistance in setting up the acquisition protocol and for acquisition of MRI data; and Mr. Bopanna Sugandh, and Mr. Anand Yadav for acquisition of MRI data. The authors thank Prof. Thomas Nichols and Dr. Soroosh Afyouni for their inputs on the usage of the DVARS toolbox. The authors are grateful to Dr. Christoph Vogelbacher and his team for their inputs on phantom analysis. Color scheme for the figures are from ColorBrewer 2.0 (<http://colorbrewer2.org/>) by Cynthia A. Brewer, Geography, Pennsylvania State University (Accessed October 25, 2019). Some of the figures use the

tight_subplot function (Pekka Kumpulainen (2020). tight_subplot(Nh, Nw, gap, marg_h, marg_w) (https://www.mathworks.com/matlabcentral/fileexchange/27991-tight_subplot-nh-nw-gap-marg_h-marg_w), MATLAB Central File Exchange. Retrieved January 27, 2020).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Pravesh Parekh  <https://orcid.org/0000-0002-1930-1167>

Gaurav V. Bhalerao  <https://orcid.org/0000-0002-3325-3413>

Ganesan Venkatasubramanian  <https://orcid.org/0000-0002-0949-898X>

John P. John  <https://orcid.org/0000-0001-6756-0297>

REFERENCES

- Afyouni, S., & Nichols, T. E. (2018). Insight and inference for DVARS. *NeuroImage*, 172, 291–312. <https://doi.org/10.1016/j.neuroimage.2017.12.098>
- Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., Miller, K. L. Smith, S. M. (2018). Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, 166, 400–424. <https://doi.org/10.1016/j.neuroimage.2017.10.034>
- Andersson, J. L. R., Graham, M. S., Drobnyak, I., Zhang, H., Filippini, N., & Bastiani, M. (2017). Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage*, 152, 450–466. <https://doi.org/10.1016/j.neuroimage.2017.02.085>
- Andersson, J. L. R., Graham, M. S., Zsoldos, E., & Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, 141, 556–572. <https://doi.org/10.1016/j.neuroimage.2016.06.058>
- Andersson, J. L. R., Jenkinson, M., & Stephen, S. (2007a). *Non-linear optimisation* (No. FMRIB technical report TR07JA1). www.fmrib.ox.ac.uk/analysis/techrep
- Andersson, J. L. R., Jenkinson, M., & Stephen, S. (2007b). *Non-linear registration, aka Spatial normalisation* [FMRIB technical report TR07JA1]. www.fmrib.ox.ac.uk/analysis/techrep
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2), 870–888. [https://doi.org/10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7)
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Ardekani, B. A. (2018). A new approach to symmetric registration of longitudinal structural MRI of the human brain. *BioRxiv*. <https://doi.org/10.1101/306811>
- Ardekani, B. A., & Bachman, A. H. (2009). Model-based automatic detection of the anterior and posterior commissures on MRI scans. *NeuroImage*, 46(3), 677–682. <https://doi.org/10.1016/j.neuroimage.2009.02.030>
- Ardekani, B. A., Kershaw, J., Braun, M., & Kanuo, I. (1997). Automatic detection of the mid-sagittal plane in 3-D brain images. *IEEE Transactions on Medical Imaging*, 16(6), 947–952. <https://doi.org/10.1109/42.650892>
- Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., Jbabdi, S., Andersson, J. L. R. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, 184, 801–812. <https://doi.org/10.1016/j.neuroimage.2018.09.073>
- Bediou, B., Franck, N., Saoud, M., Baudouin, J.-Y., Tiberghien, G., Daléry, J., & d'Amato, T. (2005). Effects of emotion and identity on facial affect processing in schizophrenia. *Psychiatry Research*, 133(2), 149–157. <https://doi.org/10.1016/j.psychres.2004.08.008>
- Behere, R., Raghunandan, V., Venkatasubramanian, G., Subbakrishna, D., Jayakumar, P., & Gangadhar, B. (2008). Trends—A Tool for Recognition of Emotions in Neuropsychiatric Disorders. *Indian Journal of Psychological Medicine*, 30(1), 32. <https://doi.org/10.4103/0253-7176.43132>
- Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, 1191, 133–155. <https://doi.org/10.1111/j.1749-6632.2010.05446.x>
- Bilderbeck, A. C., Penninx, B. W. J. H., Arango, C., van der Wee, N., Kahn, R., Winter-van Rossum, I., Hayen, A., Kas, M. J., Post, A., Dawson, G. R. (2019). Overview of the clinical implementation of a study exploring social withdrawal in patients with schizophrenia and Alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, 97, 87–93. <https://doi.org/10.1016/j.neubiorev.2018.06.019>
- Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E. Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, 83, 550–558. <https://doi.org/10.1016/j.neuroimage.2013.05.099>
- Cardenas, S. A., Kassem, L., Brotman, M. A., Leibenluft, E., & McMahon, F. J. (2016). Neurocognitive functioning in euthymic patients with bipolar disorder and unaffected relatives: A review of the literature. *Neuroscience & Biobehavioral Reviews*, 69, 193–215. <https://doi.org/10.1016/j.neubiorev.2016.08.002>
- Caruyer, E., Lenglet, C., Sapiro, G., & Deriche, R. (2013). Design of multi-shell sampling schemes with uniform coverage in diffusion MRI. *Magnetic Resonance in Medicine*, 69(6), 1534–1540. <https://doi.org/10.1002/mrm.24736>
- Castellano, F., Bartoli, F., Crocamo, C., Gamba, G., Tremolada, M., Santambrogio, J., Clerici, M. Carrà, G. (2015). Facial emotion recognition in alcohol and substance use disorders: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 59, 147–154. <https://doi.org/10.1016/j.neubiorev.2015.11.001>
- Clark, L. J., Gatz, M., Zheng, L., Chen, Y.-L., McCleary, C., & Mack, W. J. (2009). Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. *American Journal of Alzheimer's Disease and Other Dementias*, 24(6), 461–468. <https://doi.org/10.1177/1533317509345154>
- Clarke, T.-K., Smith, A. H., Gelernter, J., Kranzler, H. R., Farrer, L. A., Hall, L. S., Fernandez-Pujals, A. M., MacIntyre, D. J., Smith, B. H., Hocking, L. J., Padmanabhan, S., Hayward, C., Thomson, P. A., Porteous, D. J., Deary, I. J. McIntosh, A. M. (2016). Polygenic risk for alcohol dependence associates with alcohol consumption, cognitive function and social deprivation in a population-based cohort. *Addiction Biology*, 21(2), 469–480. <https://doi.org/10.1111/adb.12245>
- Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., Keshavan, M. S., Tammimga, C. A. (2015). Identification of distinct psychosis biotypes using brain-based biomarkers.

- American Journal of Psychiatry*, 173(4), 373–384. <https://doi.org/10.1176/appi.ajp.2015.14091200>
- Daros, A. R., Zakzanis, K. K., & Rector, N. A. (2014). A quantitative analysis of facial emotion recognition in obsessive-compulsive disorder. *Psychiatry Research*, 215(3), 514–521. <https://doi.org/10.1016/j.psychres.2013.11.029>
- Elferink, M. W.-O., van, T. I., & Kessels, R. P. C. (2015). Perception of emotions in mild cognitive impairment and Alzheimer's dementia: Does intensity matter?. *Translational Neuroscience*, 6(1), 139–149. <https://doi.org/10.1515/tnsci-2015-0013>
- Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE*, 12(9), e0184661. <https://doi.org/10.1371/journal.pone.0184661>
- Etkin, A. (2019). A reckoning and research agenda for neuroimaging in psychiatry. *American Journal of Psychiatry*, 176(7), 507–511. <https://doi.org/10.1176/appi.ajp.2019.19050521>
- Faillenot, I., Heckemann, R. A., Frot, M., & Hammers, A. (2017). Macro-anatomy and 3D probabilistic atlas of the human insula. *NeuroImage*, 150, 88–98. <https://doi.org/10.1016/j.neuroimage.2017.01.073>
- Friedman, L., & Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging: JMIRI*, 23(6), 827–839. <https://doi.org/10.1002/jmri.20583>
- Friedman, L., Glover, G. H., & Fbirn Consortium (2006). Reducing inter-scanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, 33(2), 471–481. <https://doi.org/10.1016/j.neuroimage.2006.07.012>
- Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., Rioux, P., Rousseau, M.-E., Sherif, T., Deelman, E., Khalili-Mahani, N., & Evans, A. C. (2015). Reproducibility of neuroimaging analyses across operating systems. *Frontiers in Neuroinformatics*, 9, 12. <https://doi.org/10.3389/fninf.2015.00012>
- Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., & Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, 95(4), 791–807. <https://doi.org/10.1016/j.neuron.2017.07.011>
- Gousias, I. S., Rueckert, D., Heckemann, R. A., Dyet, L. E., Boardman, J. P., Edwards, A. D., & Hammers, A. (2008). Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage*, 40(2), 672–684. <https://doi.org/10.1016/j.neuroimage.2007.11.034>
- Gronenschild, E. H. B. M., Habets, P., Jacobs, H. I. L., Mengelers, R., Rozen-daal, N., van Os, J., & Marcelis, M. (2012). The effects of FreeSurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS One*, 7(6), e38234. <https://doi.org/10.1371/journal.pone.0038234>
- Gunter, J. L., Bernstein, M. A., Borowski, B. J., Ward, C. P., Britson, P. J., Felmler, J. P., & Jack, C. R. (2009). Measurement of MRI scanner performance with the ADNI phantom. *Medical Physics*, 36(6), 2193–2205. <https://doi.org/10.1118/1.3116776>
- Hammers, A., Allom, R., Koeppe, M. J., Free, S. L., Myers, R., Lemieux, L., Mitchell, T. N., Brooks, D. J., Duncan, J. S. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*, 19(4), 224–247. <https://doi.org/10.1002/hbm.10123>
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., & Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32(1), 180–194. <https://doi.org/10.1016/j.neuroimage.2006.02.051>
- Homorogan, C., Adam, R., Barboianu, R., Popovici, Z., Bredicean, C., & Ienciu, M. (2017). Emotional face recognition in bipolar disorder. *European Psychiatry*, 41, S117. <https://doi.org/10.1016/j.eurpsy.2017.01.1904>
- Hou, P., Hasan, K. M., Sitton, C. W., Wolinsky, J. S., & Narayana, P. A. (2005). Phase-sensitive T1 inversion recovery imaging: A time-efficient interleaved technique for improved tissue contrast in neuroimaging. *American Journal of Neuroradiology*, 26(6), 1432–1438.
- Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmler, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., ... Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685–691. <https://doi.org/10.1002/jmri.21049>
- Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, 34(1), 65–73.
- Jezzard, P., & Clare, S. (1999). Sources of distortion in functional MRI data. *Human Brain Mapping*, 8(2–3), 80–85. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:2/3<80::AID-HBM2>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<80::AID-HBM2>3.0.CO;2-C)
- John, J. P., Halahalli, H. N., Vasudev, M. K., Jayakumar, P. N., & Jain, S. (2011). Regional brain activation/deactivation during word generation in schizophrenia: FMRI study. *The British Journal of Psychiatry*, 198(3), 213–222. <https://doi.org/10.1192/bjp.bp.110.083501>
- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B. C., & Fischl, B. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage*, 46(1), 177–192. <https://doi.org/10.1016/j.neuroimage.2009.02.010>
- Kharabian Masouleh, S., Eickhoff, S. B., Zeighami, Y., Lewis, L. B., Dahnke, R., Gaser, C., Chouinard-Decorte, F., Lepage, C., Scholtens, L. H., Hoffstaedter, F., Glahn, D. C., Blangero, J., Evans, A. C., Genon, S., & Valk, S. L. (2020). Influence of processing pipeline on cortical thickness measurement. *Cerebral Cortex*, 30(9), 5014–5027. <https://doi.org/10.1093/cercor/bhaa097>
- Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, 36(5), 1009–1019. <https://doi.org/10.1093/schbul/sbn192>
- Kruggel, F., Turner, J., & Muftuler, L. T. (2010). Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *NeuroImage*, 49(3), 2123–2133. <https://doi.org/10.1016/j.neuroimage.2009.11.006>
- Liang, S., Deng, W., Wang, Q., Ma, X., Li, M., Brown, M. R. G., Hu, X., Li, X., Greenshaw, A. J., & Li, T. (2016). Performance of verbal fluency as an endophenotype in patients with familial versus sporadic schizophrenia and their parents. *Scientific Reports*, 6, 32597. <https://doi.org/10.1038/srep32597>
- Maclaren, J., Han, Z., Vos, S. B., Fischbein, N., & Bammer, R. (2014). Reliability of brain volume measurements: A test-retest dataset. *Scientific Data*, 1, 140037. <https://doi.org/10.1038/sdata.2014.37>
- Madan, C. R., & Kensinger, E. A. (2017). Test-retest reliability of brain morphology estimates. *Brain Informatics*, 4(2), 107–121. <https://doi.org/10.1007/s40708-016-0060-4>
- McGorry, P. D., Hartmann, J. A., Spooner, R., & Nelson, B. (2018). Beyond the “at risk mental state” concept: Transitioning to transdiagnostic psychiatry. *World Psychiatry*, 17(2), 133–142. <https://doi.org/10.1002/wps.20514>

- McGuire, S. A., Wijtenburg, S. A., Sherman, P. M., Rowland, L. M., Ryan, M., Sladky, J. H., & Kochunov, P. V. (2017). Reproducibility of quantitative structural and physiological MRI measurements. *Brain and Behavior*, 7(9), e00759. <https://doi.org/10.1002/brb3.759>
- McPhie, D. L., Nehme, R., Ravichandran, C., Babb, S. M., Ghosh, S. D., Staskus, A., Kalinowski, A., Kaur, R., Douvaras, P., Du, F., Ongur, D., Fossati, V., Eggen, K., & Cohen, B. M. (2018). Oligodendrocyte differentiation of induced pluripotent stem cells derived from subjects with schizophrenia implicates abnormalities in development. *Translational Psychiatry*, 8(1), 230. <https://doi.org/10.1038/s41398-018-0284-6>
- Moran, P. R., Kumar, N. G., Karstaedt, N., & Jackels, S. C. (1986). Tissue contrast enhancement: Image reconstruction algorithm and selection of T1 in inversion recovery MRI. *Magnetic Resonance Imaging*, 4(3), 229–235.
- Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., Toga, A. W., Pike, G. B., Neto, P. R., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P., & Mazziotta, J. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *NeuroImage*, 40(2), 570–582. <https://doi.org/10.1016/j.neuroimage.2007.12.035>
- Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A. W., Pike, G. B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P. C. M., Mazziotta, J., & Mori, S. (2008). Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. *NeuroImage*, 43(3), 447–457. <https://doi.org/10.1016/j.neuroimage.2008.07.009>
- Orban, C., Kong, R., Li, J., Chee, M. W. L., & Yeo, B. T. T. (2020). Time of day is associated with paradoxical reductions in global signal fluctuation and functional connectivity. *PLoS Biology*, 18(2), e3000602. <https://doi.org/10.1371/journal.pbio.3000602>
- Park, H. W., Cho, M. H., & Cho, Z. H. (1986). Real-value representation in inversion-recovery NMR imaging by use of a phase-correction method. *Magnetic Resonance in Medicine*, 3(1), 15–23. <https://doi.org/10.1002/mrm.1910030104>
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142–2154. <https://doi.org/10.1016/j.neuroimage.2011.10.018>
- Reeder, S. B., Atalar, E., Bolster, B. D., & McVeigh, E. R. (1997). Quantification and reduction of ghosting artifacts in interleaved echo-planar imaging. *Magnetic Resonance in Medicine*, 38(3), 429–439. <https://doi.org/10.1002/mrm.1910380312>
- Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage*, 107, 107–115. <https://doi.org/10.1016/j.neuroimage.2014.12.006>
- Saxe, G. N., Statnikov, A., Fenyó, D., Ren, J., Li, Z., Prasad, M., Wall, D., Bergman, N., Briggs, E. C., & Aliferis, C. (2016). A complex systems approach to causal discovery in psychiatry. *PLoS One*, 11(3), e0151174. <https://doi.org/10.1371/journal.pone.0151174>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(Suppl 1), S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Snyder, H. R., Kaiser, R. H., Warren, S. L., & Heller, W. (2015). Obsessive-compulsive disorder is associated with broad impairments in executive function: A meta-analysis. *Clinical Psychological Science*, 3(2), 301–330. <https://doi.org/10.1177/2167702614534210>
- Stöcker, T., Schneider, F., Klein, M., Habel, U., Kellermann, T., Zilles, K., & Shah, N. J. (2005). Automated quality assurance routines for fMRI data applied to a multicenter study. *Human Brain Mapping*, 25(2), 237–246. <https://doi.org/10.1002/hbm.20096>
- Takao, H., Hayashi, N., & Ohtomo, K. (2011). Effect of scanner in longitudinal studies of brain volume changes. *Journal of Magnetic Resonance Imaging*, 34(2), 438–444. <https://doi.org/10.1002/jmri.22636>
- Tammimga, C. A., Pearson, G., Keshavan, M., Sweeney, J., Clementz, B., & Thaker, G. (2014). Bipolar and schizophrenia network for intermediate phenotypes: Outcomes across the psychosis continuum. *Schizophrenia Bulletin*, 40(Suppl_2), S131–S137. <https://doi.org/10.1093/schbul/sbt179>
- The American College of Radiology (2005). *Phantom test guidance for the ACR MRI accreditation program*. Reston, Virginia.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The Wu-minn Human Connectome Project: An overview. *NeuroImage*, 80, 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- Vasistha, N. A., Johnstone, M., Barton, S. K., Mayerl, S. E., Thangaraj Selvaraj, B., Thomson, P. A., Dando, O., Grunewald, E., Alloza, C., Bastin, M. E., Livesey, M. R., Economides, K., Magnani, D., Makedonopoulou, P., Burr, K., Story, D. J., Blackwood, D. H. R., Wyllie, D. J. A., McIntosh, A. M., ... Chandran, S. (2019). Familial t(1;11) translocation is associated with disruption of white matter structural integrity and oligodendrocyte–myelin dysfunction. *Molecular Psychiatry*, 24(11), 1641–1654. <https://doi.org/10.1038/s41380-019-0505-2>
- Viswanath, B., Rao, N. P., Narayanaswamy, J. C., Sivakumar, P. T., Kandasamy, A., Kesavan, M., Mehta, U. M., Venkatasubramanian, G., John, J. P., Mukherjee, O., Purushottam, M., Kannan, R., Mehta, B., Kandavel, T., Binukumar, B., Saini, J., Jayarajan, D., Shyamsundar, A., Moirangthem, S., Vijay Kumar, K. G., ... Jain, S. (2018). Discovery biology of neuropsychiatric syndromes (DBNS): A center for integrating clinical medicine and basic science. *BMC Psychiatry*, 18(1), 106. <https://doi.org/10.1186/s12888-018-1674-2>
- Vogelbacher, C., Bopp, M. H. A., Schuster, V., Herholz, P., Jansen, A., & Sommer, J. (2019). LAB-QA2GO: A free, easy-to-use toolbox for the quality assessment of magnetic resonance imaging data. *Frontiers in Neuroscience*, 13, 688. <https://doi.org/10.3389/fnins.2019.00688>
- Vogelbacher, C., Möbius, T. W. D., Sommer, J., Schuster, V., Dannlowski, U., Kircher, T., Dimpfle, A., Jansen, A., & Bopp, M. H. A. (2018). The marburg-münster affective disorders cohort study (MACS): A quality assurance protocol for MR neuroimaging data. *NeuroImage*, 172, 450–460. <https://doi.org/10.1016/j.neuroimage.2018.01.079>
- de Vrij, F. M., Baghdadi, M., Beverloo, H. B., Bouwkamp, C. G., Bouwkamp, C. G., Breedveld, G. J., Coesmans, M., Cohen, D., Ghazvini, M., Gopalakrishna, S., Gunhanlar, N., Kushner, S. A., Lendemeijer, B., Li, T. M., Mientjes, E., Olgiati, S., Quadri, M., Shpak, G., Verheijen, F. W., ... de Wit, T., (2019). Candidate CSPG4 mutations and induced pluripotent stem cell modeling implicate oligodendrocyte progenitor cell dysfunction in familial schizophrenia. *Molecular Psychiatry*, 24(5), 757–771. <https://doi.org/10.1038/s41380-017-0004-2>
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Luthman, J., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L., Shen, L., Schwarz, A., Toga, A. W., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging Initiative. (2015). 2014 update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's and Dementia*, 11(6), e1–e120. <https://doi.org/10.1016/j.jalz.2014.11.001>
- Wonderlick, J. S., Ziegler, D. A., Hosseini-Varnamkhashi, P., Locascio, J. J., Bakkour, A., van der Kouwe, A., Triantafyllou, C., Corkin, S., & Dickerson, B. C. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence,

voxel geometry, and parallel imaging. *NeuroImage*, 44(4), 1324–1333. <https://doi.org/10.1016/j.neuroimage.2008.10.037>

Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., & Fischl, B. (2014). Spurious group differences due to head motion in a diffusion MRI study. *NeuroImage*, 88, 79–90. <https://doi.org/10.1016/j.neuroimage.2013.11.027>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Parekh, P., Bhalerao, G. V., Rao, R., Sreeraj, V. S., Holla, B., Saini, L., Venkatasubramanian, G., John, J. P., & Jain, S. (2021). Protocol for magnetic resonance imaging acquisition, quality assurance, and quality check for the Accelerator program for Discovery in Brain disorders using Stem cells. *International Journal of Methods in Psychiatric Research*, 30(3), e1871. <https://doi.org/10.1002/mpr.1871>