Physica Medica 101 (2022) 8–17

Contents lists available at ScienceDirect

Physica Medica

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

Reproducibility and repeatability of magnetic resonance imaging in dementia

Catherine A. Morgan [a,b,c,*], Reece P. Roberts [a,b], Tessa Chaffey [a], Lenore Tahara-Eckl [a], Meghan van der Meer [a], Matthias Günther [d], Timothy J. Anderson [b,e,f], Nicholas J. Cutfield [b,g], John C. Dalrymple-Alford [b,e,f,h], Ian J. Kirk [a,b], Donna Rose Addis [a,b,i,j], Lynette J. Tippett [a,b], Tracy R. Melzer [b,e,f,h]

[a] School of Psychology and Centre for Brain Research, The University of Auckland, Auckland, New Zealand
[b] Brain Research New Zealand - Rangahau Roro Aotearoa, Centre of Research Excellence, New Zealand
[c] Centre for Advanced MRI, Auckland UniServices Limited, Auckland, New Zealand
[d] Fraunhofer Institute for Digital Medicine and University of Bremen, Bremen, Germany
[e] Department of Medicine, University of Otago, Christchurch, New Zealand
[f] NZ Brain Research Institute, Christchurch, New Zealand
[g] Department of Medicine, University of Otago, Dunedin, New Zealand
[h] School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand
[i] Rotman Research Institute, Baycrest Health Sciences, Toronto, Canada
[j] Department of Psychology, University of Toronto, Toronto, Canada

ARTICLE INFO

Keywords:
Reproducibility
Repeatability
Dementia
Quantitative MRI
Resting state fMRI

ABSTRACT

Purpose: Individualised predictive models of cognitive decline require disease-monitoring markers that are repeatable. For wide-spread adoption, such markers also need to be reproducible at different locations. This study assessed the repeatability and reproducibility of MRI markers derived from a dementia protocol.
Methods: Six participants were scanned at three different sites with a 3T MRI scanner. The protocol employed: T1-weighted (T1w) imaging, resting state functional MRI (rsfMRI), arterial spin labelling (ASL), diffusion-weighted imaging (DWI), T2-weighted fluid attenuation inversion recovery (FLAIR), T2-weighted (T2w) imaging, and susceptibility weighted imaging (SWI). Participants were scanned repeatedly, up to six times over a maximum period of five years. One participant was also scanned a further three times on sequential days on one scanner. Fifteen derived metrics were computed from the seven different modalities.
Results: Reproducibility (coefficient of variation; CoV, across sites) was best for T1w derived grey matter, white matter and hippocampal volume (CoV $< 1.5\%$), compared to rsfMRI and SWI derived metrics (CoV, 19% and 21%). For a given metric, long-term repeatability (CoV across time) was comparable to reproducibility, with short-term repeatability considerably better.
Conclusions: Reproducibility and repeatability were assessed for a suite of markers calculated from a dementia MRI protocol. In general, structural markers were less variable than functional MRI markers. Variability over time on the same scanner was comparable to variability measured across different scanners. Overall, the results support the viability of multi-site longitudinal studies for monitoring cognitive decline.

1. Introduction

People with mild cognitive impairment (MCI) [1] show age-related decline that is greater than that of their age-matched peers. A diagnosis of dementia is given when a significant loss of everyday cognitive function, unrelated to frailty, is identified. Approximately 50 million people are living with dementia worldwide, a number set to increase three-fold by 2050 [2]. Although MCI can be a precursor, not all people with MCI go on to develop dementia [3]. The reasons why some individuals progress to dementia and others do not is unresolved and

* Corresponding author at: School of Psychology, The University of Auckland, Building 302, Level 2, 23 Symonds Street, Private Bag 92019, Auckland 1142, New Zealand.
E-mail address: c.morgan@auckland.ac.nz (C.A. Morgan).

https://doi.org/10.1016/j.ejmp.2022.06.012
Received 15 March 2022; Received in revised form 9 June 2022; Accepted 27 June 2022
Available online 16 July 2022

remains a topic of focussed research. Biomarkers are needed that accurately track cognitive decline and hold potential to predict which individuals will progress to dementia.

Quantitative imaging biomarkers (QIBs) assist with diagnosis, disease-monitoring, and assessment of treatment and interventions. QIBs derived from magnetic resonance imaging (MRI) that are of interest in dementia studies [4] include grey matter brain volumes derived from T1-weighted (T1w) imaging for tissue atrophy, arterial spin labelling (ASL) metrics for cerebral blood flow (hypo-perfusion/hypo-metabolism), diffusion weighted imaging (DWI) measures for white matter structure, resting state functional MRI (rsfMRI) for functional connectivity and T2w fluid attenuation inversion recovery (FLAIR) derived estimates of white matter hyperintensity (WMH) volume for small vessel disease. Additional, more clinically focused scans in a dementia protocol often include T2-weighted (T2w) and susceptibility weighted imaging (SWI) to assess vascular health and other pathologies.

Multicentre studies allow for larger, more representative cohorts to be recruited for research trials. However, inter-site measurement variability needs to be quantifiable to interpret pooled data. For QIBs to be adopted widely and incorporated into clinical practice, inter-site measurement reliability needs to be established to determine confidence in diagnostic metrics. Knowledge of QIB variability over time due to measurement uncertainty is essential when monitoring longitudinal cognitive changes as a function of normal ageing, disease, or an intervention. Without this crucial information, it cannot be determined whether subsequent changes in the QIB are due to underlying physiological changes, or simply measurement variability.

The Quantitative Imaging Biomarkers Alliance (QIBA) [5,6] defines *reproducibility* as variability due to measurements being collected under different conditions, e.g., at different sites with different hardware or processed with different software. Conversely *repeatability* refers to variability in a measurement collected under the same conditions multiple times. Several studies have examined the reproducibility and repeatability of individual, or single modality QIBs within the context of dementia [7–9]. However multiple QIBs in combination, akin to a "biomarker signature", are likely to have more predictive power of cognitive decline than a single maker alone [10]. Accordingly, we wished to assess measurement variability in a suite of parameters that could be used for this purpose. We recruited "travelling heads" (THs), the same participants who travelled to imaging centres to be scanned at repeated time-points, enabling assessment of the reproducibility and repeatability of quantitative MRI (qMRI) markers for dementia. Each site was part of a multicentre, longitudinal study known as the *Dementia Prevention Research Clinics*.

2. Materials and method

2.1. Participants

Six participants (3 female, 3 male) were recruited as THs. Their average age at commencement of the study was 38.5 years (range 31.9–52.7). The study received ethical approval from the Health and Disability Ethics Committee and all participants provided informed written consent before taking part, per New Zealand National Ethical Standards.

2.2. Data acquisition

Imaging was performed at three different cities in New Zealand: Auckland, Christchurch, and Dunedin. Each site has a similar 3T MRI system (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany). Different coils were used due to equipment availability at each site. In Auckland, data were collected using a manufacturer-supplied 32-channel radio frequency (RF) head coil. In Dunedin and Christchurch, a manufacturer-supplied 64-channel head and neck RF coil were used. All three systems have the same gradient sets of 45 mT/m peak amplitude with a slew rate of 200 mT/m/msec, have identical 1st and 2nd order shim coils, and all are actively shielded. All sites had the same software version installed at the time of the inter-site reproducibility data collection, but this changed over the period of repeatability measurements.

Several steps were taken to harmonise data acquisition across all centres. At initiation of the study, the scan parameters were exported in a format that could be imported at each site directly (.exar files), to minimise error due to user input. The same stimuli presentations for rsfMRI and ASL imaging were sent to each site and consisted of a fixation cross and instructions to keep eyes closed respectively. The same set of briefing instructions for participants were used at all sites. A standard operating procedure for scanner users on how to acquire the data (e.g., landmarking, setting angle of acquisition, shim adjustments) was also adopted by each site.

TH data collection commenced when the Dementia Prevention Research Clinics opened and when participant scanning began at each site: Auckland in March 2016, Dunedin in July 2017, and Christchurch in December 2017. Data to measure reproducibility were acquired in the six THs at the three different sites as close-in-time as possible; the time between scans was on average 10 days (range 5–26). To measure long-term repeatability, data from the Auckland site were chosen as it had the longest data collection period and the greatest number of scans. In Auckland, repeat data were collected in 5 participants; one of the initial 6 participants recruited was only scanned once due to family reasons. Three participants were scanned over a period of 5 years, with 5 to 6 repeat scans per participant. Due to travel restrictions, two participants were scanned over shorter periods (1 and 3 years). To assess repeatability over a shorter time frame, one participant was scanned in Auckland on 3 consecutive afternoons.

The THs were scanned with the same protocol used for clinic participants. This consisted of: T1w, rsfMRI, a B0 field map (for distortion correction of the rsfMRI), ASL, multi-shell DWI, FLAIR, SWI and T2w imaging. Imaging parameters are outlined below.

2.3. Image acquisition

A ***T1-weighted*** magnetisation-prepared rapid gradient-echo (MPRAGE) sequence, repetition time (TR) / echo time (TE) / inversion time (TI) = 2000/2.85/880 ms, flip angle = 8 degrees, receiver bandwidth (rBw) = 240 Hz/pixel, voxel size = 1.0 × 1.0 × 1.0 mm, was collected in a sagittal orientation, with whole-brain coverage, field of view (FoV) = 256 × 256 × 208 mm and GRAPPA acceleration factor = 2, yielding a total scan duration of 4 min 56 s.

Functional blood oxygen level dependant (BOLD) ***rsfMRI*** images were collected using a gradient-echo, echo-planar imaging (EPI) simultaneous multi-slice (SMS) sequence [11] in a transverse orientation, approximately aligned with a line joining the anterior and posterior commissures (ACPC). The acquired voxel size was 2.4 × 2.4 × 2.4 mm, FoV = 210 × 210 × 154 mm, TR/TE = 735/39.0 ms, flip angle = 51 degrees, rBw = 2030 Hz/pixel, multi-band (MB) acceleration factor = 8, and 490 measurements, collected for a period of 6 min 10 s. A field map was collected in the same orientation and FoV as the BOLD scan, with a voxel size of 3.3 × 3.3 × 2.4 mm, TR = 626 ms and two TEs of 4.92 ms and 7.38 ms. Magnitude and phase data were reconstructed.

Whole brain ***ASL*** images were acquired using a 3D gradient and spin echo (GRASE) readout and pseudo-continuous labelling (pCASL) prototype sequence, with background suppression, labelling duration = 1800 ms, and single post-labelling delay = 1800 ms [12]. The acquired voxel size was 3 × 3 × 4 mm, FoV = 192 × 192 × 168 mm, TR/TE = 5000/14.4 ms, GRAPPA = 2, segments = 6, EPI factor = 17, Turbo factor = 14, rBw = 2694 Hz/pixel, with each control-label pair repeated eight times and an M0 scan collected in-line (with the sequence default TR of 4 s) for a total scan duration of 8 min 31 s. In 2020, an alternative sequence [13] was adopted at all sites. Parameters were matched as closely as possible to the previous implementation, but with the following necessary deviations, FoV = 194 × 194 × 168 mm, TE = 14.8

ms, no GRAPPA acceleration, partial Fourier in the phase encode direction = 6/8, EPI factor = 18 and Turbo factor = 21. An M0 image was collected separately.

DWI data were acquired with an echo-planar spin-echo SMS sequence [11] with a MB factor = 3. The voxel size was 2.0 × 2.0 × 2.0 mm, FoV = 210 × 210 × 144 mm, TR/TE = 3600/92 ms, excitation flip angle = 78 degrees and refocusing flip angle = 160 degrees. Diffusion was encoded using a monopolar scheme in a total of 100 non-collinear directions: 50 volumes b = 1000 s/mm^2 volumes, 50 b = 2000 s/mm^2 volumes, and six interleaved volumes without diffusion weighting (b = 0 s/mm^2). Three additional b = 0 s/mm^2 volumes were collected with the reverse (posterior-anterior) phase encoding direction for distortion correction. The DWI acquisition time was 7 min 44 s in total. Both the DWI and rsfMRI and acquisition were harmonised with the UK Biobank protocol [14].

FLAIR images were collected with a 3D T2-w SPACE sequence (variable flip angle), acquired sagittally with TR/TE/TI = 5000/393/1800 ms, FoV = 256 × 256 × 202 mm; voxel size = 1.0 × 1.0 × 1.1 mm, rBw = 781 Hz/pixel, GRAPPA = 2, and partial Fourier in the slice direction, making a scan duration of 4 min 7 s.

A **T2w** scan was collected with a BLADE (radial k-space trajectory) sequence acquired in the same ACPC orientation as the rsfMRI, with TR/TE = 5500/117 ms, FoV = 230 × 230 × 140 mm; voxel size = 0.7 × 0.7 × 3.0 mm, 36 slices, rBw = 120 Hz/pixel, GRAPPA = 2, BLADE coverage of 91%, and two concatenations, making a scan duration of 2 min 3 s.

Lastly, **SWI** data were collected with a 3D gradient echo sequence, in the same ACPC orientation and with the same voxel size as the T2w scan and FoV = 230 × 201 × 144 mm. A TR/TE = 29/20 ms, flip angle of 15, rBw = 120 Hz/pixel and GRAPPA = 2 were used. Total scan time was 2 min 46 s. Magnitude, phase, and SWI data were reconstructed.

Scan parameters were identical at all sites except for the following minor deviations: in Auckland the TE of the T1-weighted MPRAGE scan was 2.83 ms (vs. 2.85 ms in Dunedin and Christchurch); diffusion directions had to be modified slightly in Dunedin for reasons described here: https://www.fmrib.ox.ac.uk/ukbiobank/protocol/UKBB_Porting_Diffusion_Protocol_v2.pdf.

2.4. Image processing

Images were processed centrally, with the same software used for all data sets. DICOM images were converted to NifTi format, following Brain Imaging Data Structure (BIDS) conventions (http://bids.neuroimaging.io/) specified at the time of analysis.

The **T1w** images were processed using the default settings in the Computational Anatomy Toolbox (CAT12 v12.7 [15]) processing pipeline, run within the Statistical Parametric Mapping (SPM12 v7771) software package [16] using MATLAB (R2019b). A spatial-adaptive non-local means denoising filter was applied to the images [17], followed by internal resampling, affine pre-processing, initial bias correction, and affine registration. The initial standard SPM unified segmentation [18] was then applied using tissue probability maps from the International Consortium for Brain space template [19]. Skull stripping, regional parcellation and spatial normalisation then took place, followed by local intensity correction and a final adaptive maximum a posteriori segmentation [20] into total grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF), utilising a Markov Random Field approach. The proportion of each tissue type in every voxel was estimated by performing a partial volume estimation [21], generating a tissue probability map in native space. Total GM (cortical and subcortical) and WM volumes were derived for each subject and regional volumes, specifically the left and right hippocampus given evidence of atrophy in MCI [22,23] were estimated using the Automated Anatomical Labelling version 3 atlas [24,25]. An average of the left and right hippocampal volumes is reported.

The **rsfMRI** images were pre-processed with fmriprep v20.2.1 stable [26], a Nipype [27] based tool. Briefly, a single-band reference volume (SBRef) was co-registered to the T1w image, and motion parameters were estimated for each subsequent volume relative to the reference volume. Slice-timing correction, normalization to MNI space, and spatial smoothing (6 mm FWHM) were then performed. fMRI images were denoised using ICA-AROMA [28]. The following confounds (estimated during fmriprep's pipeline) were then regressed out of each participant's data using Denoiser (https://github.com/arielletambini/denoiser): six motion parameters, the top five aCompCor regressors, outlier detection regressors (framewise displacement, rmsd, dvars, standardised dvars), and discrete cosine-basis regressors for scanner signal drifts. Finally, time-series were band-pass filtered (0.01-0.1 Hz) to isolate low-frequency fluctuations in the BOLD signal.

A set of 48 regions of interest (ROIs) representing the four largest functional networks specified in [29] (default mode, DMN; dorsal attention, DAN; ventral attention, VAN; and frontoparietal control, FPCN, networks) were used for rsfMRI analyses. For each ROI, activity within a sphere (radius = 2 mm) around a central coordinate was averaged to generate a time-series. The time-series of all ROIs were then correlated with each other to generate correlation matrices for each scan. Correlation coefficients between all nodes within a network (excluding self-connections) were averaged to provide an estimate of within-network connectivity; for conciseness, VAN and FPCN results are provided as supplementary material. We also computed modularity, a graph-theoretical measure that captures the extent to which networks can be segregated into smaller communities [30] that has been found to be increased in MCI and dementia [31]. The modularity metric (Q) was estimated across all entire matrices (i.e., including all networks) using the community_louvain function in the Brain Connectivity Toolbox [32].

ASL images were processed using toolboxes distributed with FSL (version 5.0.9). First, fsl_anat was applied to the T1w image for tissue segmentation, followed by BASIL (Bayesian Inference for Arterial Spin Labelling MRI) [33] to compute motion corrected, partial volume corrected (PVC) [34] CBF maps in native space. Magnetisation of arterial blood was computed voxel-wise using the acquired M0 image and corrected for T1 relaxation [12]. Calibrated perfusion values were calculated assuming T1 blood at 3T = 1.65 s, fixed bolus duration = 1.8 s, and single compartment fitting [12]. For the first ASL sequence, an experimentally-determined labelling efficiency of 60% was used [35]. For the second ASL sequence variant, the labelling efficiency was assumed to be 85% [12]. The average PVC CBF in GM and WM masks were output.

DWI data were pre-processed using MRtrix3 [36], FMRIB Software Library (FSL) [37], and Advanced Normalisation Tools (ANTs) [38] to perform the following steps: 1) denoising [39–41], 2) Gibbs ringing correction [42], 3) eddy current distortion and motion correction [43,44], 4) brain mask estimation [45], and 5) bias field correction [46]. Next, response functions were estimated for GM, WM, and CSF [47,48], and the diffusion data and brain masks were upsampled to 1.25 mm isotropic voxel size [49]. Fibre orientation distributions (FOD) were estimated using a multi-tissue constrained spherical deconvolution (CSD) [50], a higher-order diffusion model which aims to account for multiple fibre populations (or crossing fibres) in each voxel or 'fixel' [51]. Outputs were joint bias field corrected and intensity normalised, while allowing for multi-tissue components [52].

A white matter FOD study template was generated [53] from 18 scans (6 participants at 3 sites). Each FOD image (for both reproducibility and repeatability analysis) was registered to the template as previously described [53,54]. Further processing was performed as recommended [36]: a WM fixel mask was generated with peak threshold value of 0.06; whole-brain fibre tractography was performed by generating 20 million streamlines, then subsequently filtered to two million streamlines via the spherical-deconvolution informed filtering of tractograms (SIFT) [55] algorithm. Data was smoothed to achieve connectivity-based fixel enhancement [56]. Lastly, fibre density (FD) and fibre-bundle cross-section (FC) were estimated. Using this fixel-based approach [51] (implemented in MRtrix3 [36]), FD represents

the fraction of the intra-axonal compartment within a voxel, while FC aims to account for more macroscopic properties, like the number of voxels the WM fibre bundle occupies. Since WM pathology can result from both microscopic changes in FD and macroscopic changes in FC (e.g. due to atrophy in dementia), a combined measure of fibre density and cross-section (FDC) may give a more comprehensive measure of the ability of the tract to relay information [51]. Therefore, FDC was extracted for each participant, while results for FD and FC separately are provided as supplementary material.

Lesion probability maps of WMH were generated automatically from the **T2-FLAIR** image by the lesion prediction algorithm (LPA) [57] as implemented in the LST toolbox version 3.0 (https://www.statistical-modelling.de/lst.html) for SPM. The lesion probability maps were thresholded using default settings to extract WMH lesion volumes for each scan. Since the middle-aged TH participants were likely to have a minimal/undetectable lesion load, overall FLAIR image contrast was also assessed. Both GM-WM and GM-CSF contrasts were assessed, with the GM-CSF contrast as a measure of CSF signal suppression. The participant's T1w image was input as a reference image to LPA, resulting in a FLAIR image, bias corrected in the participant's T1w space. GM, WM, and CSF whole brain signal intensity (SI) were extracted using the CAT12 tissue probability map previously described. For all analyses, image contrast between two tissue types was computed using the Michelson definition of image contrast = $(SI_1 - SI_2)/(SI_1 + SI_2)$ [58]. WMH lesion volume results are presented in the supplement.

The same contrast measures of GM-WM and GM-CSF were used for assessing the **T2w** data. Signal intensities were extracted from ROIs drawn manually on DICOM format images in the caudate, frontal WM, and anterior horns of the lateral ventricles for GM, WM, and CSF ROIs, respectively (see supplementary Fig. S1 for example ROI placements). Similar to the analysis of Voelker et al., [58], GM-WM and blood vessel contrasts were used for **SWI** derived metrics. SIs were extracted from ROIs manually drawn on the scanner generated SWI DICOM images in the globus pallidum, frontal WM, basal vein and adjacent tissue for GM, WM, vessel, and adjacent to vessel ROIs. For all the manually drawn ROIs, left and right SIs were averaged before computing contrast.

2.5. Data analysis

A total of 15 metrics were computed from the 7 different modalities: GM volume, WM volume, hippocampal (HC) volume, Q (modularity), DMN connectivity, DAN connectivity, GM perfusion, WM perfusion, fibre density CS, FLAIR GM-WM contrast, FLAIR GM-CSF contrast, T2w GM-WM contrast, T2w GM-CSF contrast, SWI GM-WM contrast, and SWI vessel contrast. Values computed for each participant were visualised using R Studio (1.3.1093), with already available plotting packages and code adapted from [59]. *Reproducibility* was defined as the mean of the within-participant coefficient of variation (CoV) of a metric across sites (at a single time-point). The metric average for all participants at one site was calculated and displayed to compare group mean values between scanners. Lines were used to connect individual participant results to examine within-participant variability (since site group means could be similar, even if within-participant results are highly variable).

Repeatability was defined as the mean of within-participant CoV of a metric across time (at a single site). To visualise repeatability, results for participants scanned at a single site were plotted, normalised to their first scan result, to show relative change from baseline. A linear model regression line was fit to the data to observe any obvious trends in the data over time. The 95% confidence interval was also plotted to visualise variability in the longitudinal data. Highly repeatable data would have a zero slope and narrow confidence interval. Further, while all repeated data was graphed, short-term repeatability was calculated separately, with data from the three scans collected on consecutive days.

3. Results

From a total of 266 scans (38 scan sessions × 7 modalities), 4 scans were missing/unusable. One baseline ASL scan was excluded due to poor labelling, evident by very low perfusion-weighted signal. For this participant, repeatability ASL measures were normalised to their second ASL scan. For another participant, the FLAIR, T2w, and SWI scans were not collected due to limited scan time availability. Representative images from a single participant scanned at the three different sites within seven days are shown in Fig. 1. Overall, images show similar signal distribution, contrast, and lack of obvious artefacts. As expected, WMH volumes in the THs were minimal (<0.35 cm^3), and therefore whole brain FLAIR GM-WM and GM-CSF contrast results were analysed.

Fig. 2 shows the data used for the assessment of reproducibility. Overall, group means for most metrics appear to be in good agreement between sites. As an example, group mean HC volumes were 4.40, 4.38 and 4.44 cm^3 for Auckland, Christchurch, and Dunedin respectively. Considering individual participant data, inter-site variability would be low compared to inter-subject variation if lines connecting the same participants data at each site do not cross, which appears to be the case for GM, WM, and HC volume. Conversely, the rsfMRI metrics, (Q, DMN, and DAN connectivity) and ASL metrics (grey and white matter perfusion) have crossing lines, suggesting inter-site variability is greater than inter-subject variability. Group means for GM perfusion were consistent across the sites (60 ml/min/100 g to 0 d.p.), although individual GM CBF measurement is more variable in Auckland than Dunedin. For the DWI metric, while the mean appears to be highly reproducible (0.29, 0.29 and 0.28 for Auckland, Christchurch, and Dunedin respectively), there is a greater range of fibre density CS values measured in Dunedin compared to Auckland. The FLAIR GM-WM contrast also appears reproducible, with one participant having greater GM-WM contrast at all sites. Regarding the T2w metrics, the mean of data from the Christchurch and Dunedin sites is more similar than that from the Auckland site, with higher GM-WM contrast (0.25 and 0.26 in Christchurch and Dunedin vs. 0.22 in Auckland) and lower and GM-CSF contrast (0.50 and 0.48 in Christchurch and Dunedin vs. 0.56 in Auckland). This trend is also evident when looking at individual participants' data points. SWI venous contrast is higher in one participant but is also different across sites for that participant.

Repeated scan metrics (including short-term repeatability) are plotted relative to baseline values in Fig. 3. Volumetric measures derived from T1w images are highly repeatable, as indicated by the flat regression lines and very narrow confidence intervals. Much larger confidence intervals are seen for the rsfMRI metrics. For example, for DMN connectivity for two participants in particular (blue and pink data points); their repeated scan results are similar to each other but over 50% different to their baseline scan. GM perfusion changes by up to 25% in the same participant, but there is a trend for consistent CBF measurement (indicated by a model slope close to zero). WM perfusion on the other hand is more variable and there appears to be a trend for increasing WM CBF over time. Fibre density CS also appears highly repeatable, evidenced by consistent values over time and very narrow confidence intervals. For the clinical scan contrast measures, SWI image contrast is the most variable, with large relative differences over time of approximately 50% for some participants.

The averages of individual participants' CoV across sites, over several years, and over three days, representing a measure of reproducibility and long and short-term repeatability respectively, are summarised in Table 1, along with the mean and standard deviation for each metric. For CoV, lower values demonstrate a more consistent measurement and higher values are more variable. Generally, reproducibility was comparable to long-term repeatability, and short-term repeatability was better than long-term repeatability (as indicated by lower CoV values; e.g., GM volume average CoVs of 1.3, 1.2, and 0.1% for reproducibility, long-term and short-term repeatability, respectively). As suggested by data presented in Figs. 2 and 3, rsfMRI, ASL, and SWI

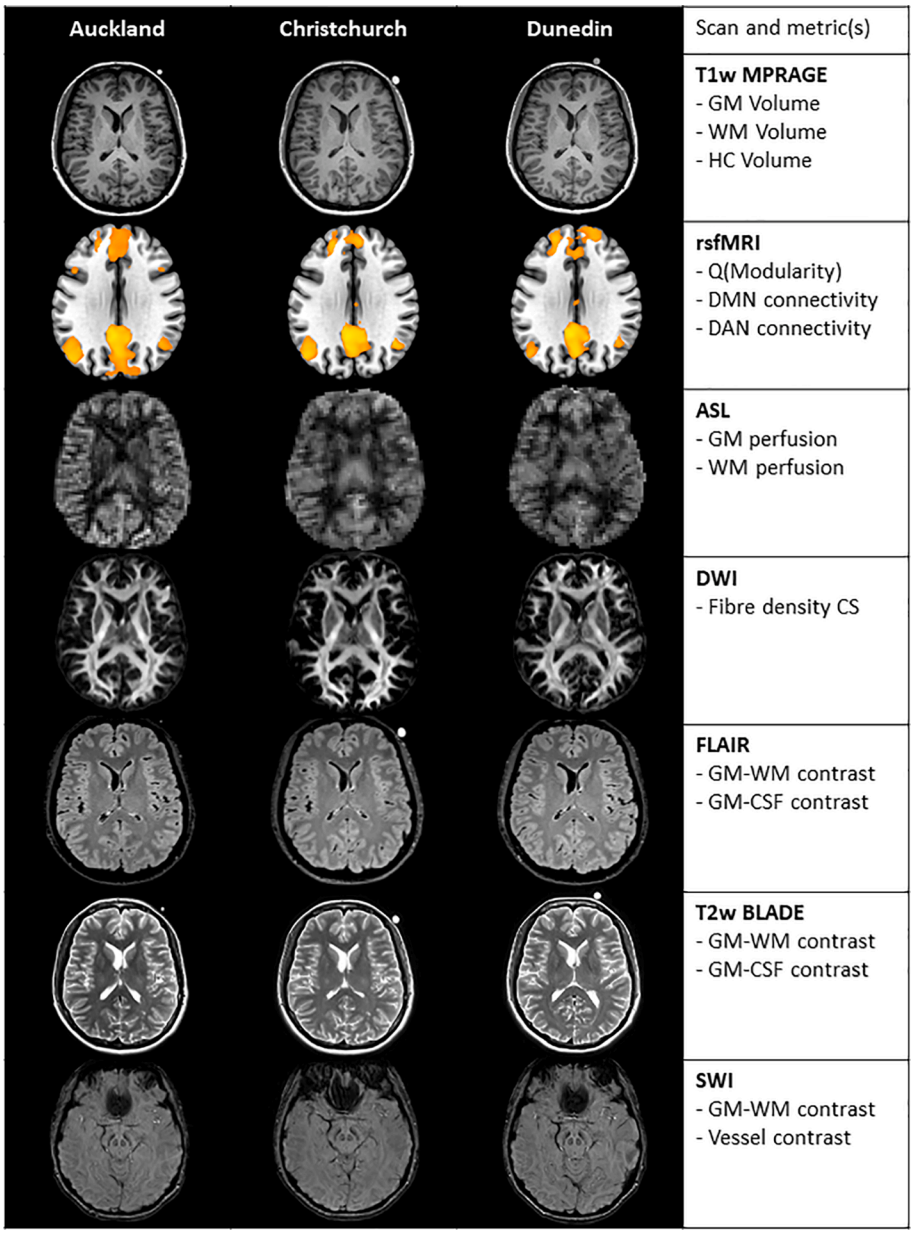

Fig. 1. Representative data for a single participant collected at three different sites within seven days (columns 1 to 3) and a list of the 15 derived metrics (column 4). From the top, row 1 shows the acquired T1w MPRAGE scan in participant's native space, row 2 shows DMN connectivity maps (precuneus seed) processed in MNI template space, row 3 shows calculated CBF maps in native space, row 4 shows white matter FOD images in study template space, row 5 shows acquired FLAIR images, row 6 shows the acquired T2w images, and row 7 the SWI images. An oil capsule affixed to the left side of the head is seen in some images as an hyperintense circle on the right side of the head (images are presented in radiological orientation). Abbreviations: ASL – arterial spin labelling, CS – cross section, CBF – cerebral blood flow, CSF – cerebrospinal fluid, DAN – dorsal attention network, DMN – default mode network, DWI – diffusion weighted imaging, FLAIR – fluid attenuation inversion recovery, FOD – fibre orientation density, GM – grey matter, HC – hippocampal, MNI - Montreal Neurological Institute, rsfMRI – resting state functional magnetic resonance imaging, SWI – susceptibility weighted imaging, T1w – T1-weighted, T2w – T2-weighted, WM – white matter.

metrics were more variable across sites and over time than the other metrics (as evidenced by higher average CoVs), with SWI and rsfMRI repeatability CoVs as high as 25%.

4. Discussion

Overall, we found that 15 QIBs derived from MRI modalities typically found in a dementia imaging protocol [4], were comparable across all sites participating in our Dementia Prevention Research Clinics. Generally, reproducibility of metrics derived from brain structure (e.g., tissue volume) was better than those measuring physiological brain processes (e.g., resting state connectivity and perfusion). For a given metric, reproducibility CoV was of similar magnitude to long-term repeatability CoV. This result suggests that between-site variability was comparable to variability in repeat scanning over several years; the length of time that may be examined in longitudinal studies on MCI. Short-term repeatability CoV (measured over 3 days) was considerably better than long-term repeatability (measured over 5 years).

Different levels of reproducibility and repeatability among MRI modalities have important implications for the interpretation of results. Our structural MRI metrics showed excellent reproducibility and repeatability, consistent with previous work [60], with CoVs of $< 1.5\%$ for GM, WM and HC volume. The CoV for GM perfusion (11% between sites and 13% over five years) are similar to those reported in the literature of 5–13% for GM regions [61,62]. This is particularly encouraging, especially given a sequence variant was required near the end of the long-term repeatability study. Also encouraging is the short-term repeatability of GM perfusion (CoV = 2.7%), with results overall suggesting consistent acquisition with the current pCASL implementation.

While DTI reliability has previously been shown to exhibit good test–retest reliability and repeatability [9,63,64], variability associated with newer alternative metrics e.g. fixel based analysis [51] requires further investigation. Recent work demonstrated that a three-tissue CSD technique (as used in this work) provided reliable and stable estimates of tissue microstructure composition, up to 3 months longitudinally in a control population (ICCs > 0.8) [65]. Our work builds on these results, measuring downstream metrics from the pipeline, fibre density cross section. We found FDCS showed only slightly higher reproducibility and repeatability CoVs (2–4%) than T1w-volumetric based metrics; lending confidence to their application in the investigation of MCI and dementia.

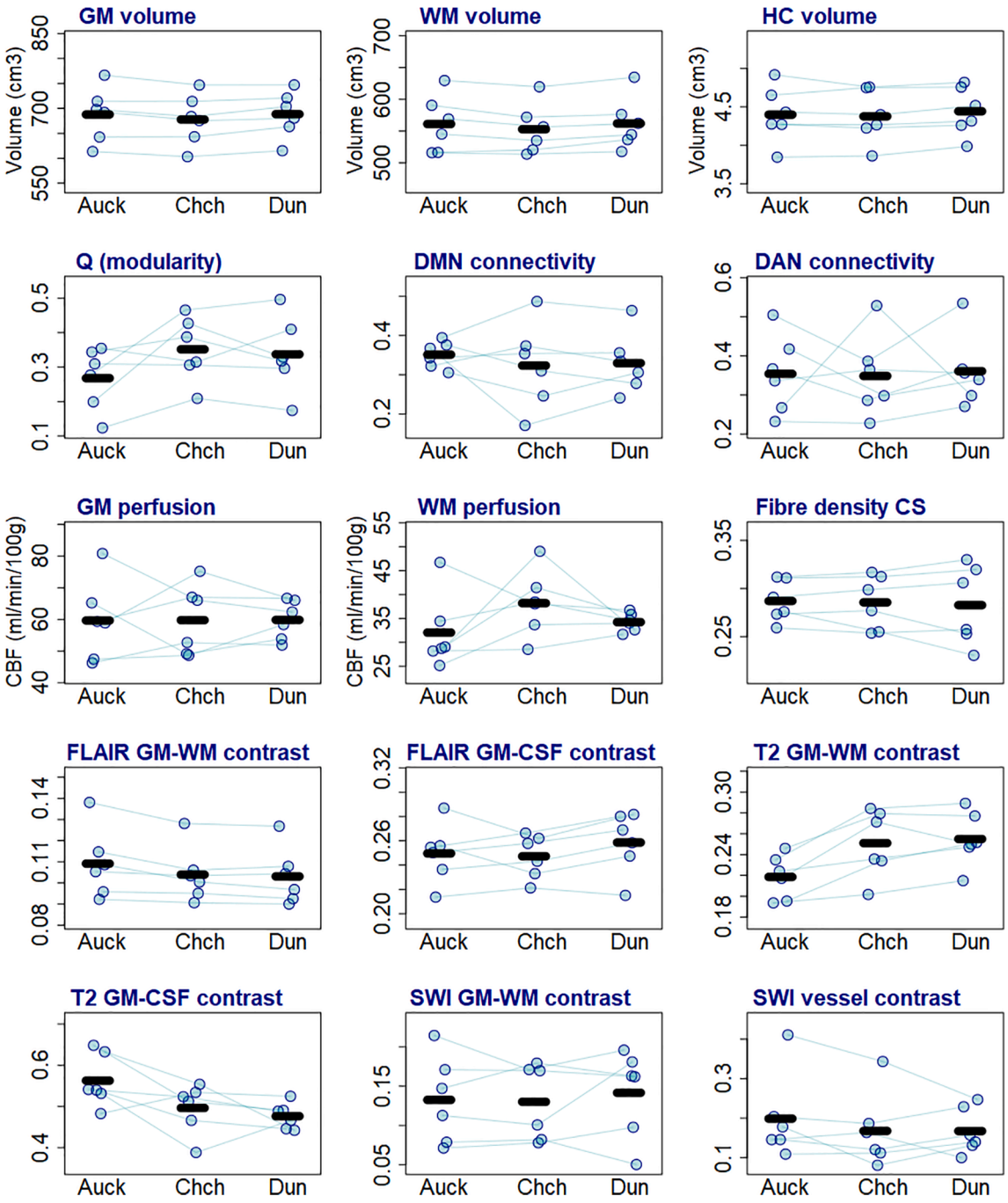

Fig. 2. Reproducibility of metrics of interest measured at three different sites. Blue circles represent values measured in individual Travelling Heads at each site and lines connect the same participant. Black thick line is the mean value for all subjects at that site. Abbreviations: Auck – Auckland, CS – cross section, CSF – cerebrospinal fluid, Chch – Christchurch, Dun – Dunedin, DAN – dorsal attention network, DMN – default mode network, FLAIR – fluid attenuation inversion recovery, GM – grey matter, HC – hippocampal, SWI – susceptibility weighted imaging, WM – white matter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The result that reproducibility (inter-site variation) and long-term repeatability (within-site variation) were similar in magnitude for a given metric was somewhat surprising. One would expect data collected under near identical conditions (i.e., on the same scanner) to be less variable than data collected on different scanners. There are, however, several factors that might explain this result. First, even when closely matching sequence parameters between vendors [66], some differences remain. The data we present, although collected at three different sites, were acquired on MRI scanners of the same model, from a single manufacturer, with the same operating system (OS) software version at the time of reproducibility scanning. This meant that the same versions of pulse sequences (including those supplied as standard, prototype sequences from the vendor, and "research" sequences developed by other centres) were available at each site. Therefore, sequence parameters could be better matched across sites in our study than in multi-vendor studies, thus improving reproducibility. Second, although there were some hardware differences between the sites, these were unlikely to impact our metrics of interest. Specifically, the 64-channel coil in Christchurch and Dunedin compared to the 32-channel head coil in Auckland is likely to have better SNR of the infratentorial brain due to greater coil coverage in this region and may be the reason for

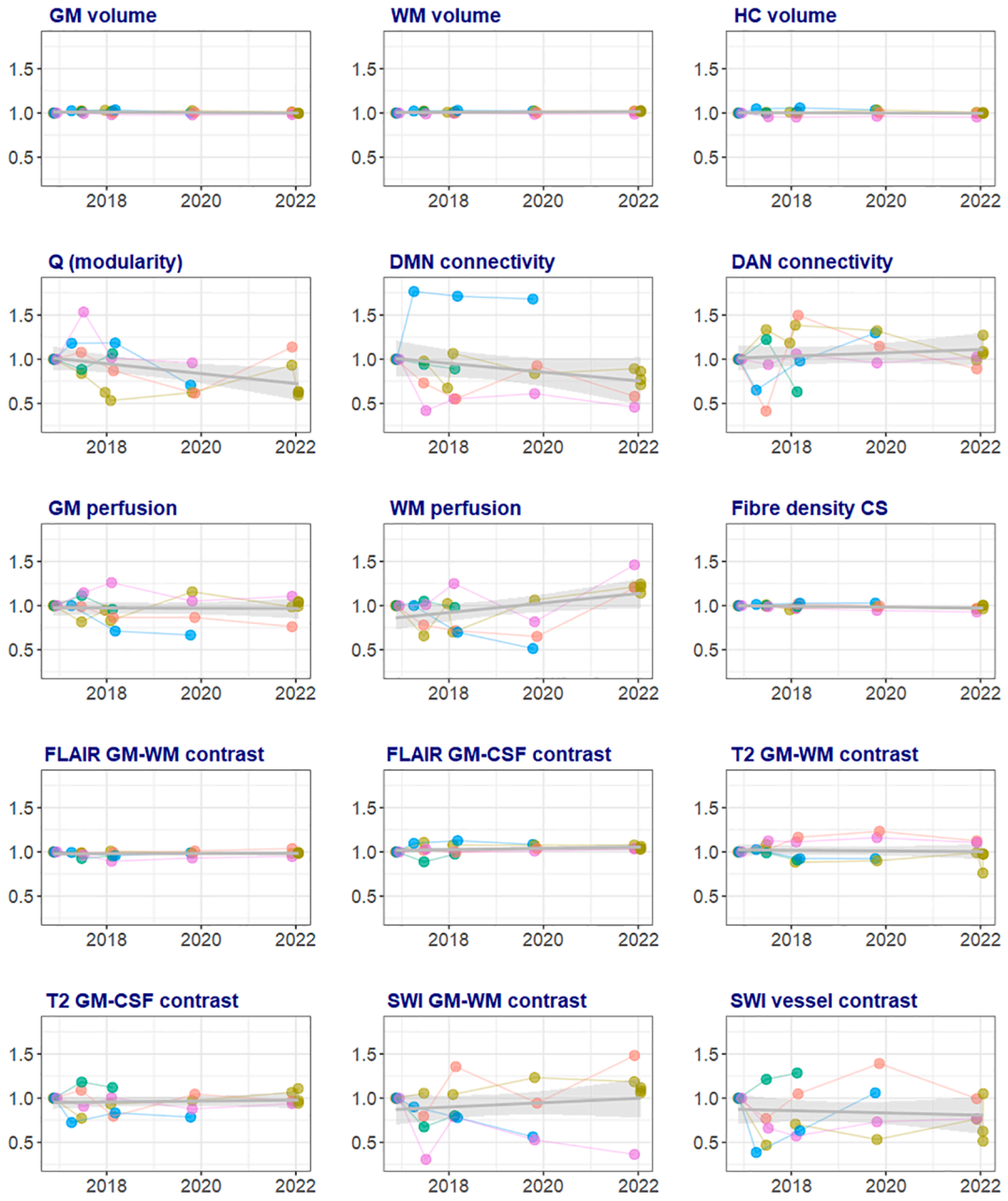

Fig. 3. Repeatability of metrics of interest measured at one site (Auckland). Circles represent relative values compared to baseline, measured in individual THs. Colours represent each participant. Grey lines indicate regression lines and shaded areas are 95% confidence intervals. Abbreviations: CS – cross section, CSF – cerebrospinal fluid, DAN – dorsal attention network, DMN – default mode network, FLAIR – fluid attenuation inversion recovery, GM – grey matter, HC – hippocampal SWI – susceptibility weighted imaging, WM – white matter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more consistent T2w contrast at the two sites compared to Auckland. However, all our metrics of interest are in the supratentorial brain, and other QIBs investigated do not appear to have a step change at the two 64-channel coil sites. One reason for the absence of a step change could be that, although it has a greater total number of RF receive coils, the 64-channel coil has these elements distributed around the head and neck, and is likely to have a comparable number of coils around the head to the 32-channel head only coil used in Auckland. Given the aforementioned factors, our reproducibility results are likely to be better than those in other studies that measured reproducibility across scanners from multiple vendors. Minimising measurement variability due to differing hardware configurations is an important step in the process of validating QIBs [67], and our results indicate that our efforts in doing so have been successful.

The overarching aim of the Dementia Prevention Research Clinics is to find biomarkers, or combinations thereof, that predict the development of dementia. Given that the process of cognitive decline is likely to occur over several years, the long-term stability of QIBs needs to be assessed. We scanned the THs repeatedly over a period of up to five years at the Auckland site to assess long-term repeatability. Strictly speaking, repeatability is measured under identical conditions [4], but over five years there are factors beyond our control that may cause measurements to deviate from this ideal scenario. For instance, required updates to operating system OS software and sequences is one such factor. Although an OS update was implemented at the end of 2017 in Auckland, our data do not show any obvious step changes in any of the QIBs collected after this time points (see Fig. 3). There were several deviations

Table 1
Summary of reproducibility and long and short-term repeatability results for metrics of interest. Horizontal line groups metrics from the same modality. Mean is reported to 3 significant figures and CoV reported to 2 decimal places. **Reproducibility results**: Mean results are the average of the within-participant mean of a metric across sites (at a single time-point). S.D. results are the average of the within-participant standard deviation of a metric across sites (at a single time-point). CoV results are the average of the within-participant coefficient of variation of a metric across sites (at a single time-point). **Repeatability (long term) results**: Mean results are the average of within-participant mean of a metric across time (over 5 years, at a single site). S.D. results are the average of within-participant standard deviation of a metric across time (over 5 years, at a single site). CoV results are the average of within-participant coefficient of variation of a metric across time (over 5 years, at a single site). **Repeatability (short term) results:** Mean results are the average of within-participant mean of a metric across time (over 3 days, at a single site). S.D. results are the average of within-participant standard deviation of a metric across time (over 3 days, at a single site). CoV results are the average of within-participant coefficient of variation of a metric across time (over 3 days, at a single site). **Abbreviations**: CS – cross section, CSF – cerebrospinal fluid, DAN – dorsal attention network, DMN – default mode network, FLAIR – fluid attenuation inversion recovery, GM – grey matter, HC – hippocampal, SWI – susceptibility weighted imaging, WM – white matter.

	Reproducibility			Repeatability (Long term)			Repeatability (Short term)		
Total data points	*18*			*23*			*3*		
Total participants	*6*			*5*			*1*		
Total data points per participant	*3 sites*			*3–6 time points*			*3 timepoints (over 3 days)*		
Across participant average	*Mean*	*S.D.*	***CoV (%)***	*Mean*	*S.D.*	***CoV (%)***	*Mean*	*S.D.*	***CoV (%)***
GM volume (cm^3)	685	9	**1.3**	698	8	**1.2**	744	1	**0.1**
WM volume (cm^3)	559	7	**1.2**	570	5	**0.9**	590	4	**0.7**
HC volume (cm^3)	4.41	0.06	**1.3**	4.36	0.06	**1.3**	4.76	0.03	**0.7**
Q (modularity) (a.u.)	0.319	0.061	**19.3**	0.314	0.069	**23.5**	0.306	0.011	**3.7**
DMN connectivity (a.u.)	0.335	0.044	**14.4**	0.335	0.072	**21.9**	0.285	0.028	**9.7**
DAN connectivity (a.u.)	0.355	0.061	**16.7**	0.373	0.082	**23.5**	0.290	0.030	**10.2**
GM perfusion (ml/min/100 g)	59.8	6.7	**11.1**	59.7	7.9	**12.7**	71.6	1.9	**2.7**
WM perfusion (ml/min/100 g)	34.8	5.3	**14.9**	36.9	8.4	**21.8**	52.7	2.3	**4.4**
Fibre density CS (a.u.)	0.285	0.010	**3.6**	0.293	0.005	**1.9**	0.313	0.007	**2.3**
FLAIR GM-WM contrast (a.u.)	0.105	0.003	**3.1**	0.108	0.003	**2.6**	0.092	0.001	**0.7**
FLAIR GM-CSF contrast (a.u.)	0.252	0.010	**4.1**	0.240	0.009	**3.7**	0.299	0.004	**1.7**
T2 GM-WM contrast (a.u.)	0.242	0.021	**8.7**	0.214	0.014	**6.7**	0.195	0.027	**13.7**
T2 GM-CSF contrast (a.u.)	0.512	0.059	**11.6**	0.570	0.058	**10.3**	0.560	0.050	**8.8**
SWI GM-WM contrast (a.u.)	0.135	0.020	**16.4**	0.134	0.030	**25.4**	0.189	0.005	**2.7**
SWI vessel contrast (a.u.)	0.178	0.037	**20.7**	0.194	0.046	**25.5**	0.200	0.077	**38.7**

in the ASL pulse sequence over the period of data collection. Although updates to a prototype 3D pCASL GRASE sequence were provided in late 2017 and 2018, generally settings used at baseline could still be reimplemented. In early 2020, however, a switch to an alternative to 3D pCASL GRASE [13] sequence was required. The GM-CBF results collected with the new variant look consistent with those collected beforehand, but WM-CBF does appear to be higher. That WM-CBF was affected but GM-CBF was not is likely because WM perfusion is considerably lower than GM, yielding lower perfusion weighted signal that may, in turn, lead to noisier estimates of CBF. CBF is scaled by labelling efficiency which will also vary between pulse sequences [35]. Ideally, stable versions of sequences should be used in longitudinal clinical studies, however, there is an ongoing tension between stability of measurement and rapid pace of progress in the field. Thus, in order to leverage the current best-practice in terms of ASL acquisition protocol (i.e. 3D pCASL) [12] we opted to use development versions.

Of course, another deviation from identical conditions is that the brains of the THs also inevitably age and thus may change over the period of scanning. Our travelling heads had a mean age of approximately 40 years at the start of the study. It is therefore unlikely that they underwent potentially pathological declines (e.g. from undiagnosed MCIs) over a five year period and that any age-related changes in the QIBs should be minimal [68,69]. The younger age of our THs compared to the target cohort of the Clinics (55+ years) means that we are better able to isolate the true repeatability in metrics from underlying age-related brain changes. It should be noted however, that a disadvantage in studying this younger age group is that volume of WMHs, a QIB widely studied in the context of normal and pathological aging, was very low (0 to 0.3 cm^3) compared to approximately 5 cm^3 in cognitively-normally 60–64 year olds [70], and over 10 cm^3 in our own cohort of probable AD participants (preliminary unpublished data).

Taken together, although minimal, there were nevertheless some minor OS software and sequence variations, and the potential for brain changes related to aging over the study period. Therefore, it is unsurprising that long-term repeatability (up to five years) was considerably poorer than short-term repeatability (over three days). These findings emphasise the need to minimise changes to software and sequences during longitudinal studies, particularly when collecting 'noisier' QIBs (e.g., DMN-connectivity), and to quantify and account for these variations in subsequent analyses where possible.

It should be noted that in order to be 'closer' to the clinic and understand the variability in the metrics of interest to researchers or clinicians, we compared endpoint parameters of interest (e.g. DMN correlation coefficients) rather than raw data (e.g. temporal signal to noise ratio of the BOLD-weighted time series). Arguably this is a strength of this study, and to this end, we selected analytic approaches and software packages that are commonly used in the field. However, it remains possible that other software packages may produce output metrics that are more (or less) sensitive to variation of the raw data. This question sits outside of the scope of the present study, but we note that there are multiple ongoing investigations on this topic [60,71].

The usefulness of any given QIB in monitoring disease processes depends both on its measurement error and the magnitude of the disease-related change. Here, we provide a quantitative estimate of reproducibility and repeatability, encompassing systematic error and random error. This approach allows informative comparison to potential disease effect sizes. Measures with excellent between- and within-site stability (e.g., GM and HC volumes) facilitate the detection of even subtle disease-related changes. Our reported long-term repeatability of hippocampal volume CoV of 1.3% suggests that annualised rates of

hippocampal atrophy in AD of 4.7% (3.3% greater than control subjects) [72] should be reliably detectable. Using the same ASL acquisition and processing as in the present study, we report reduced perfusion (PVC CBF in GM) of approximately 50 ± 15 ml/min/100 g (group mean ± S. D) in a probable Alzheimer's disease group compared to 75 ± 20 ml/min/100 g in a cognitively normal group [73]. For the THs, we found the average of the within-participant standard deviation of PVC CBF in GM, across either site or across timepoint, to be < 8 ml/min/100 g (see supplementary Table 2). This observation provides confidence in measuring changes due to perfusion in Alzheimer's disease, since the magnitude of the TH variability—both between sites and over time—is small compared to the observed group difference. Recent work suggests that rsfMRI metrics exhibit low test–retest reliability (*meta*-analysis ICC = 0.29) [74]. Our findings of relatively high within-subject CoV values for modularity, DMN connectivity, and DAN connectivity are consistent with this finding. While still informative, this finding suggests that effect sizes must be large and/or more subjects are needed to detect differences in rsfMRI metrics; new advanced multimodal methods are also proposed to address this issue [75].

5. Conclusion

In this work, we investigated the reproducibility (inter-site) and repeatability (intra-site, over both short (days) and long (years) time periods) of 15 quantitative MRI metrics in the context of an ongoing longitudinal investigation of MCI and dementia. Structural metrics exhibited excellent reproducibility across three sites and repeatability over both days and up to five years. Resting state fMRI showed poorer reproducibility and repeatability, while perfusion MRI showed intermediate levels. Variability over time on the same scanner was comparable to variability measured on different scanners, and generally short term repeatably was much better than long term repeatability. This work provides both confidence in the robustness of many MRI-based metrics and highlights areas for improvement.

Funding source

This study was funded in whole by Brain Research New Zealand - Rangahau Roro Aotearoa, a government funded Centre of Research Excellence. DRA receives salary support from the Canada 150 Research Program.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the travelling head participants for their time and dedication to the study. We thank the imaging staff of the Centre for Advanced Magnetic Resonance Imaging (Auckland) and Pacific Radiology Group (Dunedin and Christchurch) for MRI scanning, and the Centre for eResearch, The University of Auckland, for computing support. The authors acknowledge Siemens Healthcare for the provision of a 3D pCASL prototype sequence and Dr. Marta Vidorreta De Cerio and Josef Pfeuffer for advice on its use. We thank Simon Konstandin and Klaus Eickel, along with author Matthias Günther at Fraunhofer Institute for Digital Medicine, Germany, for provision of a 3D pCASL sequence. Finally, we thank Essa Yacoub and Edward Auerbach at the Center for Magnetic Resonance Research, The University of Minnesota, for providing the multi-band accelerated EPI sequences used for the rsfMRI and DWI acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2022.06.012.

References

[1] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004; 256:183–94. https://doi.org/10.1111/j.1365-2796.2004.01388.x.
[2] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020;396(10248):413–46.
[3] Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 2018;90(3):126–35.
[4] Vernooij MW, Pizzini FB, Schmidt R, Smits M, Yousry TA, Bargallo N, et al. Dementia imaging in clinical practice: a European-wide survey of 193 centres and conclusions by the ESNR working group. Neuroradiology 2019;61(6):633–42.
[5] Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27–67.
[6] Wang Y, Tadimalla S, Rai R, Goodwin J, Foster S, Liney G, et al. Quantitative MRI: Defining repeatability, reproducibility and accuracy for prostate cancer imaging biomarker development. Magn Reson Imaging 2021;77:169–79.
[7] Leutritz T, Seif M, Helms G, Samson RS, Curt A, Freund P, et al. Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3 T: A multicenter dual-vendor reproducibility and repeatability study. Hum Brain Mapp 2020;41:4232–47. https://doi.org/10.1002/hbm.25122.
[8] Mutsaerts HJMM, Petr J, Thomas DL, De Vita E, Cash DM, van Osch MJP, et al. Comparison of arterial spin labeling registration strategies in the multi-center GENetic frontotemporal dementia initiative (GENFI). J Magn Reson Imaging 2018; 47(1):131–40.
[9] Palacios EM, Martin AJ, Boss MA, Ezekiel F, Chang YS, Yuh EL, et al. Toward Precision and Reproducibility of Diffusion Tensor Imaging: A Multicenter Diffusion Phantom and Traveling Volunteer Study. AJNR Am J Neuroradiol 2017;38(3): 537–45.
[10] Kim J, Lee B. Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine. Hum Brain Mapp 2018;39(9):3728–41.
[11] Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, et al. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn Reson Med 2010;63(5):1144–53.
[12] Alsop DC, Detre JA, Golay X, Günther M, Hendrikse J, Hernandez-Garcia L, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn Reson Med 2015;73(1):102–16.
[13] Günther M, Bock M, Schad LR. Arterial spin labeling in combination with a look-locker sampling strategy: Inflow turbo-sampling EPI-FAIR (ITS-FAIR). Magn Reson Med 2001;46:974–84. https://doi.org/10.1002/mrm.1284.
[14] Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage 2018;166:400–24.
[15] Dahnke R, Gaser C. Voxel-based Preprocessing in CAT. 2017. https://doi.org/10.13140/RG.2.2.11653.70887.
[16] Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical parametric mapping: the analysis of functional brain images. Elsevier; 2011.
[17] Manjón JV, Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging 2010;31(1):192–203.
[18] Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005;26(3):839–51.
[19] Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). NeuroImage 1995;2:89–101. https://doi.org/10.1006/nimg.1995.1012.
[20] Rajapakse JC, Giedd JN, Rapoport JL. Statistical approach to segmentation of single-channel cerebral MR images. IEEE Trans Med Imaging 1997;16:176–86.
[21] Tohka J, Zijdenbos A, Evans A. Fast and robust parameter estimation for statistical partial volume models in brain MRI. Neuroimage 2004;23(1):84–97.
[22] Tabatabaei-Jafari H, Shaw ME, Cherbuin N. Cerebral atrophy in mild cognitive impairment: A systematic review with meta-analysis. Alzheimer's Dementia: Diagnosis Assessment Disease Monitor 2015;1:487–504. https://doi.org/10.1016/j.dadm.2015.11.002.
[23] Mueller SG, Schuff N, Yaffe K, Madison C, Miller B, Weiner MW. Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. Hum Brain Mapp 2010;31(9):1339–47.
[24] Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M. Automated anatomical labelling atlas 3. NeuroImage 2020;206:116189. https://doi.org/10.1016/j.neuroimage.2019.116189.
[25] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 2002; 15(1):273–89.

[26] Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods 2019;16(1):111–6.

[27] Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. Front Neuroinform 2011;5. https://doi.org/10.3389/fninf.2011.00013.

[28] Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 2015;112:267–77. https://doi.org/10.1016/j.neuroimage.2015.02.064.

[29] Thomas Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 2011;106(3):1125–65.

[30] Newman MEJ. Modularity and community structure in networks. PNAS 2006;103:8577–82. https://doi.org/10.1073/pnas.0601602103.

[31] Pereira JB, Mijalkov M, Kakaei E, Mecocci P, Vellas B, Tsolaki M, et al. Disrupted Network Topology in Patients with Stable and Progressive Mild Cognitive Impairment and Alzheimer's Disease. Cereb Cortex 2016;26(8):3476–93.

[32] Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage 2010;52:1059–69. https://doi.org/10.1016/j.neuroimage.2009.10.003.

[33] Chappell MA, Groves AR, Whitcher B, Woolrich MW. Variational Bayesian Inference for a Nonlinear Forward Model. Trans Sig Proc 2009;57:223–36. https://doi.org/10.1109/TSP.2008.2005752.

[34] Chappell MA, Groves AR, MacIntosh BJ, Donahue MJ, Jezzard P, Woolrich MW. Partial volume correction of multiple inversion time arterial spin labeling MRI data. Magn Reson Med 2011;65:1173–83. https://doi.org/10.1002/mrm.22641.

[35] Vidorreta M, Wang Z, Rodríguez I, Pastor MA, Detre JA, Fernández-Seara MA. Comparison of 2D and 3D single-shot ASL perfusion fMRI sequences. NeuroImage 2013;66:662–71. https://doi.org/10.1016/j.neuroimage.2012.10.087.

[36] Tournier JD, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. NeuroImage 2019;202:116137. https://doi.org/10.1016/j.neuroimage.2019.116137.

[37] Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. Review FSL. NeuroImage 2012;62:782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.

[38] Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage 2011;54:2033–44. https://doi.org/10.1016/j.neuroimage.2010.09.025.

[39] Cordero-Grande L, Christiaens D, Hutter J, Price AN, Hajnal JV. Complex diffusion-weighted image estimation via matrix recovery under general noise models. NeuroImage 2019;200:391–404. https://doi.org/10.1016/j.neuroimage.2019.06.039.

[40] Veraart J, Fieremans E, Novikov DS. Diffusion MRI noise mapping using random matrix theory. Magn Reson Med 2016;76:1582–93. https://doi.org/10.1002/mrm.26059.

[41] Veraart J, Novikov DS, Christiaens D, Ades-Aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. NeuroImage 2016;142:394–406. https://doi.org/10.1016/j.neuroimage.2016.08.016.

[42] Kellner E, Dhital B, Kiselev VG, Reisert M. Gibbs-ringing artifact removal based on local subvoxel-shifts. Magn Reson Med 2016;76:1574–81. https://doi.org/10.1002/mrm.26054.

[43] Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage 2016;125:1063–78. https://doi.org/10.1016/j.neuroimage.2015.10.019.

[44] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004;23:S208–19.

[45] Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143–55. https://doi.org/10.1002/hbm.10062.

[46] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction. IEEE Trans Med Imag 2010;29:1310–20. https://doi.org/:10.1109/TMI.2010.2046908.

[47] Dhollander T, Raffelt D, Connelly A. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. ISMRM Workshop on Breaking the Barriers of Diffusion MRI 2016:5.

[48] Dhollander T, Mito R, Raffelt D, Connelly A. Improved white matter response function estimation for 3-tissue constrained spherical deconvolution. Proc Intl Soc Mag Reson Med 2019:555.

[49] Dyrby TB, Lundell H, Burke MW, Reislev NL, Paulson OB, Ptito M, et al. Interpolation of diffusion weighted imaging datasets. NeuroImage 2014;103:202–13.

[50] Jeurissen B, Tournier JD, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage 2014;103:411–26. https://doi.org/10.1016/j.neuroimage.2014.07.061.

[51] Raffelt DA, Tournier J-D, Smith RE, Vaughan DN, Jackson G, Ridgway GR, et al. Investigating white matter fibre density and morphology using fixel-based analysis. NeuroImage 2017;144:58–73.

[52] Raffelt D, Dhollander T, Tournier JD, Tabbara R, Smith RE, Pierre E, et al. Bias field correction and intensity normalisation for quantitative analysis of apparent fiber density. Proc Intl Soc Mag Reson Med 2017;25:3541.

[53] Raffelt D, Tournier JD, Fripp J, Crozier S, Connelly A, Salvado O. Symmetric diffeomorphic registration of fibre orientation distributions. NeuroImage 2011;56:1171–80. https://doi.org/10.1016/j.neuroimage.2011.02.014.

[54] Raffelt D, Tournier J-D, Rose S, Ridgway GR, Henderson R, Crozier S, et al. Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. NeuroImage 2012;59(4):3976–94.

[55] Smith RE, Tournier JD, Calamante F, Connelly ASIFT. Spherical-deconvolution informed filtering of tractograms. NeuroImage 2013;67:298–312. https://doi.org/10.1016/j.neuroimage.2012.11.049.

[56] Raffelt DA, Smith RE, Ridgway GR, Tournier J-D, Vaughan DN, Rose S, et al. Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. Neuroimage 2015;117:40–55.

[57] Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. NeuroImage 2012;59(4):3774–83.

[58] Voelker MN, Kraff O, Brenner D, Wollrab A, Weinberger O, Berger MC, et al. The traveling heads: multicenter brain imaging at 7 Tesla. MAGMA 2016;29(3):399–415.

[59] Politzer-Ahles S, Piccinini P. On visualizing phonetic data from repeated measures experiments with multiple random effects. J Phonetics 2018;70:56–69. https://doi.org/10.1016/j.wocn.2018.05.002.

[60] Hedges EP, Dimitrov M, Zahid U, Brito Vega B, Si S, Dickson H, et al. Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. NeuroImage 2022;246:118751. https://doi.org/10.1016/j.neuroimage.2021.118751.

[61] Rischka L, Godbersen GM, Pichler V, Michenthaler P, Klug S, Klöbl M, et al. Reliability of task-specific neuronal activation assessed with functional PET, ASL and BOLD imaging. J Cereb Blood Flow Metab 2021;41(11):2986–99.

[62] Melzer TR, Keenan RJ, Leeper GJ, Kingston-Smith S, Felton SA, Green SK, et al. Test-retest reliability and sample size estimates after MRI scanner relocation. NeuroImage 2020;211:116608. https://doi.org/10.1016/j.neuroimage.2020.116608.

[63] Madhyastha T, Mérillat S, Hirsiger S, Bezzola L, Liem F, Grabowski T, et al. Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. Hum Brain Mapp 2014;35(9):4544–55.

[64] Shahim P, Holleran L, Kim JH, Brody DL. Test-retest reliability of high spatial resolution diffusion tensor and diffusion kurtosis imaging. Sci Rep 2017;7:11141. https://doi.org/10.1038/s41598-017-11747-3.

[65] Newman BT, Dhollander T, Reynier KA, Panzer MB, Druzgal TJ. Test-retest reliability and long-term stability of three-tissue constrained spherical deconvolution methods for analyzing diffusion MRI data. Magn Reson Med 2020;84:2161–73. https://doi.org/10.1002/mrm.28242.

[66] Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry 2010;15(12):1128–39.

[67] Smith EE, Biessels GJ, De Guio F, Leeuw FE, Duchesne S, Düring M, et al. Harmonizing brain magnetic resonance imaging methods for vascular contributions to neurodegeneration. Alzheimers Dement (Amst) 2019;11(1):191–204.

[68] Juttukonda MR, Davis LT, Lants SK, Waddle SL, Lee CA, Patel NJ, et al. A Prospective, Longitudinal Magnetic Resonance Imaging Evaluation of Cerebrovascular Reactivity and Infarct Development in Patients With Intracranial Stenosis. J Magn Reson Imaging 2021;54:912–22. https://doi.org/10.1002/jmri.27605.

[69] Garnier-Crussard A, Bougacha S, Wirth M, André C, Delarue M, Landeau B, et al. White matter hyperintensities across the adult lifespan: relation to age, Aβ load, and cognition. Alzheimer's Res Therapy 2020;12(1). https://doi.org/10.1186/s13195-020-00669-4.

[70] Wen W, Sachdev P. The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals. NeuroImage 2004;22:144–54. https://doi.org/10.1016/j.neuroimage.2003.12.027.

[71] Bergamino M, Keeling EG, Walsh RR, Stokes AM. Systematic Assessment of the Impact of DTI Methodology on Fractional Anisotropy Measures in Alzheimer's Disease. Tomography 2021;7:20–38. https://doi.org/10.3390/tomography7010003.

[72] Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, et al. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. Neurobiol Aging 2009;30(11):1711–23.

[73] Morgan CA, Melzer TR, Roberts RP, Wiebels K, Mutsaerts HJMM, Spriggs MJ, et al. Spatial variation of perfusion MRI reflects cognitive decline in mild cognitive impairment and early dementia. Sci Rep 2021;11(1). https://doi.org/10.1038/s41598-021-02313-z.

[74] Noble S, Scheinost D, Constable RT. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage 2019;203:116157. https://doi.org/10.1016/j.neuroimage.2019.116157.

[75] Elliott ML, Knodt AR, Hariri AR. Striving toward translation: strategies for reliable fMRI measurement. Trends Cogn Sci 2021;25:776–87. https://doi.org/10.1016/j.tics.2021.05.008.