



TITLE:

Identifying and reverting the adverse effects of white matter hyperintensities on cortical surface analyses

AUTHOR(S):

Oi, Yuki; Hirose, Masakazu; Togo, Hiroki; Yoshinaga, Kenji; Akasaka, Thai; Okada, Tomohisa; Aso, Toshihiko; ... Glasser, Matthew F.; Hayashi, Takuya; Hanakawa, Takashi

CITATION:

Oi, Yuki ...[et al]. Identifying and reverting the adverse effects of white matter hyperintensities on cortical surface analyses. *NeuroImage* 2023, 281: 120377.

ISSUE DATE:

2023-11-01

URL:

<http://hdl.handle.net/2433/285323>

RIGHT:

© 2023 The Author(s). Published by Elsevier Inc.; This is an open access article under the CC BY license.



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

NeuroImage

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



Identifying and reverting the adverse effects of white matter hyperintensities on cortical surface analyses

Yuki Oi^{a,b,c}, Masakazu Hirose^{a,b}, Hiroki Togo^{a,c,d}, Kenji Yoshinaga^a, Thai Akasaka^e, Tomohisa Okada^e, Toshihiko Aso^c, Ryosuke Takahashi^b, Matthew F. Glasser^f, Takuya Hayashi^{c,g}, Takashi Hanakawa^{a,c,d,e,*}

^a Department of Integrated Neuroanatomy and Neuroimaging, Kyoto University Graduate School of Medicine, Kyoto, Japan

^b Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto, Japan

^c Laboratory for Brain Connectomics Imaging, Center for Biosystems Dynamics Research, RIKEN, Kobe, Japan

^d Integrative Brain Imaging Center, National Center of Neurology and Psychiatry, Kodaira, Japan

^e Human Brain Research Center, Kyoto University Graduate School of Medicine, Kyoto, Japan

^f Departments of Radiology and Neuroscience, Washington University School of Medicine, St. Louis, MO, United States

^g Department of Brain Connectomics, Kyoto University Graduate School of Medicine, Kyoto, Japan

ARTICLE INFO

Keywords:

White matter hyperintensities
Human connectome project
Neuroanatomy
Brain
Magnetic resonance imaging
Cortical surface analysis

ABSTRACT

The Human Connectome Project (HCP)-style surface-based brain MRI analysis is a powerful technique that allows precise mapping of the cerebral cortex. However, the strength of its surface-based analysis has not yet been tested in the older population that often presents with white matter hyperintensities (WMHs) on T2-weighted (T2w) MRI (hypointensities on T1w MRI). We investigated T1-weighted (T1w) and T2w structural MRI in 43 healthy middle-aged to old participants. Juxtacortical WMHs were often misclassified by the default HCP pipeline as parts of the gray matter in T1w MRI, leading to incorrect estimation of the cortical surfaces and cortical metrics. To revert the adverse effects of juxtacortical WMHs, we incorporated the Brain Intensity Abnormality Classification Algorithm into the HCP pipeline (proposed pipeline). Blinded radiologists performed stereological quality control (QC) and found a decrease in the estimation errors in the proposed pipeline. The superior performance of the proposed pipeline was confirmed using an originally-developed automated surface QC based on a large database. Here we showed the detrimental effects of juxtacortical WMHs for estimating cortical surfaces and related metrics and proposed a possible solution for this problem. The present knowledge and methodology should help researchers identify adequate cortical surface biomarkers for aging and age-related neuropsychiatric disorders.

1. Introduction

Surface-based analysis of brain MRI can more accurately delineate complicated cortical ribbons and more precisely map the functional neuroanatomy of the brain than volume-based analysis (Anticevic et al., 2008; Fischl et al., 2008; Frost and Goebel, 2012; Tucholka et al., 2012; Van Essen et al., 2012; Glasser et al., 2016; Coalson et al., 2018). Among surface-based analysis methods, the Human Connectome Project (HCP) pipeline is a widely used workflow of advanced surface-based analyses of multimodal brain MRIs (Glasser et al., 2013). The usefulness of the HCP-style approach has been shown with the young adult HCP dataset

(YA-HCP) and the subsequent developing and aging connectome datasets (Elam et al., 2021). A human brain atlas was proposed based on YA-HCP multimodal MRI data, including myelin and thickness maps, and functional connectivity (Glasser et al., 2016). Taking advantage of its registration accuracy, the HCP-style approach has now been adopted in cross-scanner harmonization projects worldwide, including the UK biobank (Williams et al., 2023) and the Brain/MINDS-beyond (Koike et al., 2021). Next, the application of HCP-style analysis to the older population is warranted to gain new insights into the pathophysiological mechanisms of age-related neurodegenerative diseases, including dementia and Parkinson's disease (Bookheimer et al., 2019; Li et al., 2021;

* Corresponding author at: Integrated Neuroanatomy and Neuroimaging, Director, Human Brain Research Center, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

E-mail address: hanakawa.takashi.2s@kyoto-u.ac.jp (T. Hanakawa).

<https://doi.org/10.1016/j.neuroimage.2023.120377>

Received 12 January 2023; Received in revised form 22 August 2023; Accepted 12 September 2023

Available online 14 September 2023

1053-8119/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Wakasugi and Hanakawa, 2021).

However, caution must be exercised when applying surface-based analysis to MRI data from older populations, since they often present with age-related structural changes in the brain, including periventricular and deep white matter (WM) hyperintensities (WMHs) on T2 weighted (T2w) MRIs. WMHs are caused mainly by small vessel vasculopathy and resultant ischemic changes (Barkhof and Scheltens, 2002). WMHs are observed not only in neurological conditions such as vascular dementia and parkinsonism (Gootjes et al., 2004; Bohnen and Albin, 2011) but also at a variable degree in healthy seniors (Scott et al., 2015; Phuah et al., 2022). The prevalence of WMHs increases with age; at least 80–90 % of the population over the age of 60 have WMHs (Launer, 2003; Kruit et al., 2004; Liao et al., 1997; de Leeuw et al., 2001; Caunca et al., 2019; Atwood et al., 2004; Silbert et al., 2008). Furthermore, age-related WM abnormalities cannot only be visualized as WMH on T2w MRI but also as low-intensity areas on T1 weighted (T1w) MRI, making the signal intensity of the gray matter (GM) similar to that of the WM. This can be problematic in segmenting between the GM and WM based on the contrast of signal intensity of T1w MRIs. A recent study showed that an automated segmentation algorithm may misclassify parts of WMHs as GM, resulting in an incorrect estimation of GM volumes (Dadar et al., 2021).

Thus, the existence of WMHs may also degrade surface estimation in HCP-style analysis. In the HCP pipeline, the FreeSurfer software suite 6.0 (Fischl, 2012; <https://surfer.nmr.mgh.harvard.edu/>) segregates the cortical GM and WM based on the contrast of the signal intensity of T1w MRI and estimates the boundary between the WM and GM (the WM surface hereafter) along the cortical ribbon. Notably, because the intensity of WMHs in T1w MRI is close to that of the GM, voxels in juxtacortical WMHs may be misclassified as parts of the cerebral cortex, resulting in an incorrect estimation of the WM surface. As the WM surface is also used as an input to the pipeline when the cortical outer surface (pial surface hereafter) is subsequently estimated, the error in the WM surface estimation can easily propagate to the erroneous pial surface estimation. In contrast to commonly used surface analysis methods that use only T1w MRI, the HCP pipeline uses information from T2w MRI as well as T1w MRI for tuned estimation of the pial surfaces to reduce the adverse effects of the dura mater signals (Glasser et al., 2013). The possibility of surface estimation errors due to WMHs has not yet been thoroughly examined. Thus, it is worth investigating the likelihood of surface estimation errors due to WMHs and the propagation of tissue misclassification into the inaccurate computation of thickness and myelin contrast, particularly when analyzing MRIs with WMHs.

A solution for solving the surface estimation errors caused by WMHs is to correct the WM segmentation by manually editing the misclassified voxel label in the WMHs. (https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/WhiteMatterEdits_freeview). Rerunning FreeSurfer after manual editing should revert the erroneous WM surface estimation, followed by better estimation of the pial surfaces, thickness, and myelin contrast in the HCP pipeline. However, manual correction relies on time-consuming hand-editing operations by human experts. These laborious processes may make manual correction impractical for application to large population studies. Another possible solution is to take advantage of recently developed machine learning (ML) algorithms for the automated detection of WMHs (Griffanti et al., 2016) and integrate the algorithms as a module into the HCP pipeline.

Herein, we aimed to elucidate the potentially detrimental effects of juxtacortical WMHs for estimating cortical surfaces and related metrics and to implement a module for correcting surface estimation errors in the HCP pipeline. To automatically detect WMH labels, we used an ML-based Brain Intensity AbNormality Classification Algorithm (BIANCA) (Griffanti et al., 2016). Furthermore, we incorporated BIANCA-derived WMH masks into FreeSurfer to update the WM segments of the FreeSurfer pipeline, followed by re-estimation of the pial surface, thickness, and myelin in the HCP pipeline. The applied pipeline was validated by comparing the results with those obtained using manually edited WMHs

in the same dataset. Using the outputs of the cortical surface analysis (WM and pial surfaces, cortical thickness, and myelin contrast), we compared the surface estimation errors of the automated ML WMH-adapted HCP pipeline (ML pipeline) with those of the manually delineated WMH-adapted HCP pipeline (manual pipeline) and the default HCP pipeline without considering WMHs (default pipeline). To evaluate the effect of the proposed pipeline, we developed two quality control (QC) methods identifying surface reconstruction errors because past QC methods have limitations in terms of inter-rater variability and difficulty of quantification (Backhausen et al., 2016; Monereo-Sánchez et al., 2021). A stereological QC method was developed for semi-quantitative visual estimation of the pial and white matter surface errors. We also developed a fully automated QC algorithm for detecting extreme outliers of surface metrics. We find these QC accurate and reliable for verifying cortical surface reconstruction by the proposed preprocessing pipelines.

2. Method

2.1. Participants

We used MRI data registered in the MRI database of the Integrative Brain Imaging Center of the National Center of Neurology and Psychiatry (NCNP). This MRI database was originally created to serve as control data for neurodegenerative disorders including Parkinson's disease (Togo et al., 2023) and spinocerebellar degeneration (Bando et al., 2019). The research protocol was approved by the Institutional Review Board of NCNP (A2018-086) and was performed in accordance with the Declaration of Helsinki. For the present study, we retrieved MRI data from participants with both 3D T1 and FLAIR images. Resultantly, the present data were derived from 43 individuals who did not report any previous neuropsychiatric disorders (64.4 years old [SD 11.1], age range 41–83 years, 29 males). The exclusion criteria were as follows: a Mini-Mental State Examination (MMSE) score < 24 or local brain lesions (e.g., brain tumor or cerebral infarction) incidentally identified on MRIs. In addition, all study participants provided written informed consent. The mean MMSE score was 29.1 (SD 1.4, range =24–30). The Fazekas score (Fazekas et al., 1987) was 0, 1, 2, and 3 in 10, 21, 8, and 5 participants, respectively, in the periventricular area, and 0, 1, 2, and 3 in 12, 19, 9, and 3 participants, respectively, in the deep WM.

2.2. Data acquisition

All MRI scans were acquired using a 3-T MRI scanner with a 32-channel phased-array head coil (MAGNETOM Verio Dot, Siemens Medical Systems, Erlangen, Germany). In addition, T1w MRIs were acquired using the three-dimensional (3D) magnetization-prepared rapid gradient-echo (MPRAGE) sequence: repetition time (TR) = 1900 ms; echo time (TE) = 2.52 ms; inversion time (TI) = 900 ms; base resolution = 256; field of view (FoV) read = 250 mm; flip angle (FA) = 9°; sagittal slices = 192; voxel size = 0.98 × 0.98 × 1 mm³. T2w fluid-attenuated inversion recovery (FLAIR) MRIs were acquired using 3D sampling perfection with application-optimized contrast using different flip angle evolution (3D SPACE) sequences: TR = 5000 ms; TE = 383.0 ms; TI = 1800 ms; base resolution = 256; FoV read = 250 mm; FA = 150°; sagittal slices = 176; interpolation = on; voxel size = 0.5 × 0.5 × 1 mm³.

2.3. Data analysis

2.3.1. Preprocessing of MRI data

All structural MRI data were converted from Digital Imaging and Communications in Medicine files to Neuroimaging Informatics Technology Initiative (NIfTI) files and then preprocessed using the HCP pipeline implemented with Connectome Workbench ver. 1.5.0, FMRIB Software Library (FSL) 6.0.4 (Smith et al., 2004) and FreeSurfer 5.3-HCP (Fischl, 2012). The HCP pipeline consists of three steps:

PreFreeSurferPipeline, FreeSurferPipeline, and PostFreeSurferPipeline (Fig. 1a). The initial PreFreeSurfer Pipeline step corrected the T1w and T2w structural MRIs for image distortions related to the gradient nonlinearity inherent to the scanner type, registered MRIs into the anterior-posterior commissural coordinate space of the Montreal Neurological Institute (MNI) templates using functional magnetic resonance imaging of the brain's (FMRIB) Linear Image Registration Tool (FLIRT) algorithm, and resampled the data at a 0.7-mm isovoxel. The pipeline then performs brain extraction, fine-tuned registration of T1w and T2w with boundary-based registration (Greve and Fischl, 2009), biasfield correction of T1w and T2w (Glasser and Van Essen, 2011), and nonlinear registration to the MNI template. The nonlinear registration was conducted using FMRIB's nonlinear Image Registration Tool (FNIRT), and both T1w and T2w volumes were resampled using spline at 0.7-mm and 2-mm isovoxels in the MNI space.

2.3.2. Cortical surface estimation and calculation of T1w/T2w myelin contrast

Subsequently, the FreeSurferPipeline performed cortical surface reconstruction using both T1w and T2w volumes. We used T2-weighted images to minimize the impact of the skull or dura signal on the surface reconstruction. In brief, the process included skull stripping of T1w volume using the standard brain mask template and the deformation field calculated in the PreFreeSurferPipeline (Glasser et al., 2013), classification of brain voxels into subcortical and cortical GM and WM segments (Fischl et al., 2002), intensity normalization (Sled et al., 1998), and extraction of the WM segment of the cerebrum. Next, the WM segment was used for the initial estimation of the tentative GM and WM boundary, forming a 'WM surface' in each hemisphere, followed by fine-tuning of the WM surface placement using the 0.7-mm isovoxel T1w volume (Glasser et al., 2013), the estimation of GM and cerebrospinal fluid (CSF) boundary forming a 'pial surface', fine-tuned registration of T2w to T1w with a boundary-based registration (Greve and Fischl, 2009), and fine-tuning of the pial surface using high-resolution (0.7 mm isovoxel) T2w volume (Glasser et al., 2013). We found that juxtacortical WMHs were often mislabeled as parts of cortical GM because of relatively low- and high-intensity WMHs in T1w and T2w MRIs, respectively, resulting in errors in the white and pial surface estimation. These phenomena are detailed in Section 3.2 and the correction methods in Section 2.3.5. Cortical thickness was calculated as the distance between the white and pial surfaces at each vertex of the cortex (Fischl and Dale, 2000). In the last step, the PostFreeSurferPipeline performed surface registration using a multimodal surface matching (MSM) program

(Robinson et al., 2018) based on a metric of the folding pattern, 'sulc', generated by FreeSurfer (MSMsulc). The white and pial surfaces were then resampled into standardized mesh surfaces using 164k and 32k vertices after symmetrization between the left and right hemispheres (Van Essen et al., 2012). Next, the pipeline created a volume with a ratio of the T1w and T2w signals (myelin mapping) and mapped the values of the voxels between the inner and outer cortical surfaces onto the mid-thickness surface in the participants' native space. The surface metrics including the T1w/T2w myelin contrast and cortical thickness were resampled onto the 164k and 32k mesh surfaces (Glasser and Van Essen, 2011). The result of this analysis was based on the default setting of the HCP pipeline, which we call the "default pipeline".

2.3.3. Manual delineation of WMH masks

In this study, we created manually defined WMH masks on the 2-mm isovoxel image in the MNI space in each participant using FSL eyes in FSL (McCarthy, 2022). In addition, manual delineation of WMHs was achieved through a consensus between two board-certified neurologists (Y.O. and M.H.) who had access to both 0.7-mm and 2-mm versions of the T1w and T2w-FLAIR MRIs. The manually defined WMH masks were used as the training data for BIANCA and were directly fed into the HCP pipeline for surface re-estimation (see Section 2.3.5). The total volume of the WMHs was calculated for each participant after the WMH masks were registered back into the participant's native space using the inversion warp field.

2.3.4. Machine learning prediction of WMHs

Automated supervised ML-based segmentation of WMHs was performed using BIANCA (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BIANCA>). BIANCA is based on a *k*-nearest neighbor algorithm and computes the probability of each voxel being a WMH according to the voxel intensity and its spatial features. Furthermore, BIANCA has the flexibility to achieve accurate WMH segmentation in different MRI acquisition protocols by adapting to data.

To run BIANCA, bias-corrected T1w and T2w-FLAIR MRIs, which are the interim products of the default HCP pipeline, were used as features, and a manually delineated binary WMH mask as a label feature. The leave-one-out cross-validation scheme was used for training the model and predicting WMHs for each participant. The bias-corrected T1w and T2w-FLAIR volumes in the MNI nonlinear space were resampled to a 2 mm isovoxel. The WMH masks, brain mask, and T2w FLAIR volume were fed into the BIANCA. To run BIANCA, the following options were applied: spatial weighting = 1; no patch; selection of the non-lesion

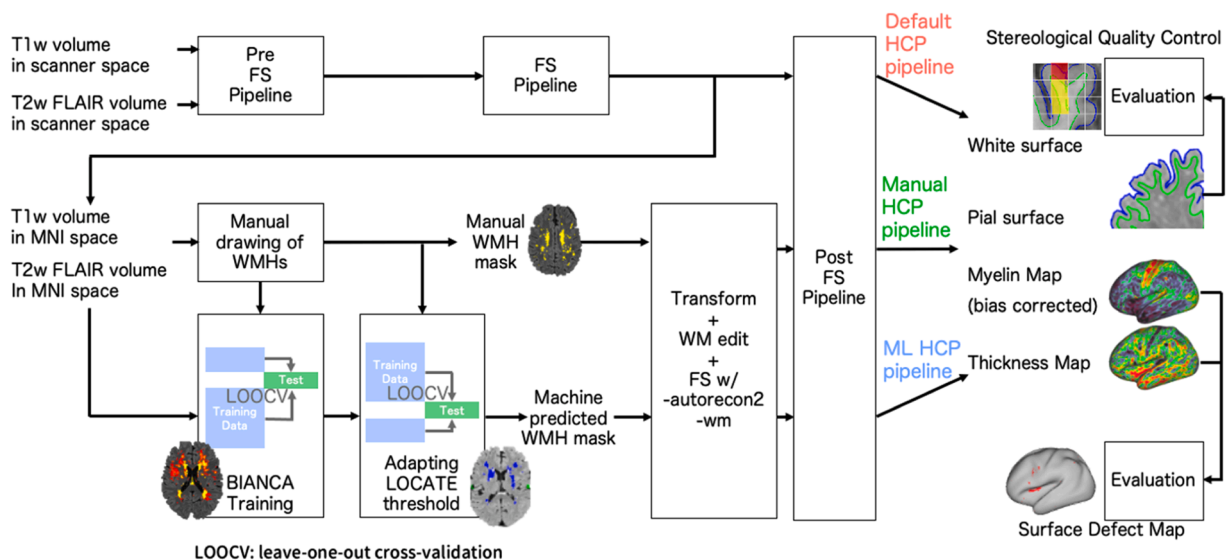


Fig. 1. Overview of processing pipelines for cortical surface analysis in the current study.

points = no border (excluding three voxels close to the lesion's edge); the number of lesion points to use = 2000; the number of non-lesion points to use = 10,000. BIANCA outputs a volume file in which the intensity of each voxel represents the probability of being a WMH, ranging from zero to one (WMH probability map). To make a binary WMH mask from the WMH probability map, an appropriate threshold was chosen by using a region-by-region threshold optimization technique, LOcally Adaptive Threshold Estimation (LOCATE) (Sundaresan et al., 2019). LOCATE refines the estimation accuracy in three steps. First, the lesion probability map was divided into subregions based on Voronoi tessellation. Second, local features within these subregions were extracted. Finally, based on the extracted features, the optimal local threshold was estimated through a supervised learning method using the manually delineated WMH masks as the training data and the leave-one-out cross-validation scheme. Using the optimized threshold by LOCATE, BIANCA yields a binary mask for each individual (ML-predicted WMH mask). The ML-predicted WMH mask was validated using the dice similarity index (SI) with reference to manual WMH masks. The SI was calculated as $2 * (|\text{manually defined WMHs} \cap \text{ML-predicted WMH mask}|) / (|\text{manually defined WMHs}| + |\text{ML-predicted WMH mask}|)$.

2.3.5. Cortical surface re-estimation

The manually defined or ML-predicted WMH masks were both resampled into 0.7-mm isovoxel resolution using FLIRT with spline interpolation and binarized again by thresholding at 0.5, for each participant. The cortical surfaces were re-calculated by running a customized FreeSurferPipeline to rerun the surface reconstruction program ('recon-all' of FreeSurfer), according to 'Manual-Intervention Workflow for WM edit'. In brief, the user-edited, binarized WMH masks (derived from either the manual or the ML pipeline) were labeled with a value of "255", by following the WM editing procedure of FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/WhiteMatterEdits_freeview). The labels were transferred to the corresponding voxels of the WM segmentation file (wm.mgz) as the WM representing the WM mask to be used for subsequent white surface reestimation. Next, the 'recon-all' command was rerun by specifying the workflow to re-estimate the cortical surface (with flags of -autorecon2-wm and -autorecon3). The products of the cortical surface re-estimation, such as the pial and WM surfaces, thickness, and 'sulc', were used to rerun the PostFreeSurfer pipeline, as described above. With the new surface information, the pipeline re-calculated the symmetrization, resampled the surfaces to 164k and 32k meshes, re-registered the surfaces with MSMSulc, and re-calculated the myelin mapping and thickness.

Cortical surfaces are reconstructed using three pipelines: (1) the original HCP pipeline (default pipeline), (2) the white matter (WM) hyperintensity (WMH)-adapted pipeline with manually-drawn WMHs (manual pipeline), and (3) the WMH-adapted pipeline with the machine learning (ML) algorithm (ML pipeline). T1w and T2w-fluid-attenuated inversion recovery (FLAIR) images are used as inputs of the HCP pipeline consisting of PreFreeSurferPipeline, FreeSurferPipeline, and PostFreeSurferPipelines in the default pipeline (the upper row). In the manual pipeline, manually-drawn WMHs are used for editing WM, and FreeSurfer is run with -autorecon2 and -wm options to re-estimate cortical surface (middle row). The ML pipeline (lower row) used an ML algorithm for WMH estimation (FSL BIANCA), trained using manually-made WMH masks, followed by optimized thresholding (LOCATE) and the same edits of WM for re-estimation. The WM surface, pial surface, and bias-corrected T1w/T2w myelin and thickness maps are used to evaluate the validity of the customized pipelines.

2.3.6. Quality control (QC) of cortical surface

2.3.6.1. Stereological QC by Experts. Two expert radiologists (T.O. and T.A.) were asked to independently evaluate the quality of the pial and WM surfaces of the three pipelines and to count the number of surface

errors in a blinded manner with respect to the difference in pipelines. For the semi-quantification of errors, we applied a stereological method, which is commonly used in quantitative neuroanatomy, for example, to count the number of neurons with a microscope (Saper, 1996; Zhao and van Praag, 2020). The coronal sections of T1w images and pial and WM surfaces in the MNI nonlinear standard space are displayed in a 4×7 matrix (a total of 28 sections at an interval of 7 mm in the y-direction) with grid lines overlaid at 7-mm intervals (Fig. 2). The images were imported into the background image and grids using Microsoft Excel (Microsoft Corporation 2019). The aforementioned radiologists were asked to identify the surface errors in each grid, depending on the type of surface errors: pial surface error only, WM surface error only, and both errors. The radiologists were asked to exclude medial temporal regions (including the hippocampus and amygdala) from their evaluation, based on anatomical location on structural MRIs. This policy was set because the cortical surfaces were not necessarily well estimated in these regions in some cases, regardless of the WMHs in the default pipeline. The number of each category was computed cell by cell across the 43 datasets separately for the manual, ML, and default pipelines, yielding two sets of each category data for each pipeline.

2.3.6.2. QC using surface defect score. Automation or quantification of surface QC has not yet been achieved. To date, QC of cortical surface estimation has commonly been performed by visual inspection of the cortical surface boundaries overlain over the orthogonal sections of T1w images (e.g., https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/OutputData_freeview).

In this study, we developed an original QC method for the cortical surfaces. We assume that if there are cortical surface defects, then the values of both T1w/T2w myelin and thickness at a surface vertex deviate from the standard values. One may expect that cortical surface defects may cause an error exclusively in thickness, but this is unlikely because T1w/T2w values should take much lower (CSF) or higher signals (WM) than those in the cortical ribbons. The deviation from the standard values can be estimated using robust Z-scores (RZ) based on the standard values. The HCP pipeline can internally define a "medial wall" (representing the inner/medial surface of 3rd ventricles, hippocampus, amygdala and thalamus, and the midsection of the corpus callosum) so as to automatically exclude the medial temporal regions. With the use of the medial wall, therefore, the defect scores were not calculated in the medial temporal regions to be consistent with the policy for the stereological QC.

We defined a vertex-wise surface defect when absolute RZ values for both T1w/T2w myelin (RZm) and thickness (RZt) took outlier values from the standard. Therefore, the surface defect at surface vertex i can be expressed as follows:

$$\text{Surfacedefect}(i, Z_0) = \text{abs}(\text{RZm}(i)) > Z_0 \cap \text{abs}(\text{RZt}(i)) > Z_0$$

where Z_0 is the Z-threshold for outliers. RZ can be estimated using the following equation:

$$\text{RZ}(i) = (X(i) - Q2(i)) / \text{NIQR}(i)$$

where $X(i)$ is the measured T1w/T2w myelin (i.e., MyelinMap_BC) or thickness at a vertex, i , in 164k Connectivity Informatics Technology Initiative (CIFTI) dscalar format in a participant, $Q2$ is the median of the standard values at vertex i , and NIQR is the normalized interquartile range at vertex i :

$$\text{NIQR}(i) = (Q3(i) - Q1(i)) / (F(0.75) - F(0.25)),$$

where $Q3$ and $Q1$ are 75 percentile and 25 percentile of the standard values, respectively, and $F(0.75)$ and $F(0.25)$ are the probability density functions of the normal distribution at the probability of 0.75 and 0.25, respectively.

For the $Q2$, $Q3$, and $Q4$ of standard values, we used 164k

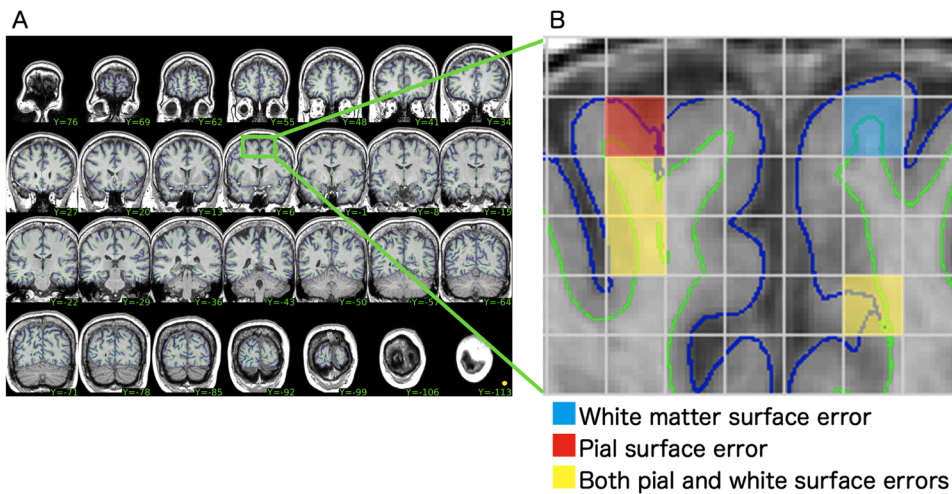


Fig. 2. A stereological method used in visual quality control (QC) of cortical surfaces.

A) The coronal sections of T1w images (in gray color) as well as the pial (blue line) and WM surfaces (lime-colored line) in the MNI nonlinear standard space are displayed in a 4×7 matrix (a total of 28 sections at an interval of 7 mm in the y direction) with grid lines overlain at 7-mm intervals (in the x and z directions). B) A zoomed panel showing grids over the medial frontal area. Two qualified radiologists identified and color-coded the surface errors in each grid: the pial surface error only (red), WM surface error only (light blue), and both errors (yellow).

MyelinMap_BC and thickness data in a CIFTI dscalar format, which was pre-generated from the YA-HCP data S1200 ($n=1096$) and released at BALSa database (<https://balsa.wustl.edu/reference/pkXDZ>). Next, the vertices with the surface defect were mapped onto the surface in each individual, which we called the ‘surface defect map’ (SDM). For the group-wise evaluation, we created a surface defect frequency map. The total number of vertices with surface defects at $Z_0=5$ was termed the ‘Surface Defect Score’ (SDS5) and was used for the subsequent analysis. That was, to define outliers, we used the robust Z score of $= 5$, which corresponds to the extreme outliers in the literature (Tukey, 1977). We furthermore checked the validity of this threshold based on preliminary analyses in which Z score thresholds between 2 and 5 were applied to small data subsets sampled randomly from the entire datasets. Note, however, that optimization of thresholding based on the robust z-scoring is recommended for future studies since an optimal threshold is dependent on the distribution and size of the data. To calculate SDM and SDS5, we used a quality check and reporting tool, *hcpipe_qc*, (https://github.com/RIKEN-BCIL/bcil/blob/master/bin/hcpipe_qc), which can be applied to the data analyzed using the HCP or non-human primate-HCP pipeline. Note that the current SDM and SDS5 are based on the standard values of the surface metric distribution of young adults (YA-HCP), as well as on a threshold corresponding to extreme outliers. These operational parameters may be optimized in the future depending on the research questions, protocols of data acquisition, and data quality (see Discussion).

We tested whether the SDS5 values were correlated with the WMH volume in each pipeline. In addition, we computed the degree to which SDS5 was improved by the proposed pipeline, thereby defining the cases in which the proposed pipelines improved the surface analyses above chance. We defined the improved cases as participants whose differences in SDS5 (between the default pipeline and ML pipeline) were outliers ($Q3+1.5IQR$) in the distribution of the SDS5 changes.

2.3.7. Statistical analysis

To estimate how the number of surface errors was affected by the raters and pipelines, a three-way analysis of variance (ANOVA) was applied to the number of errors, with factors of the rater (raters 1 and 2), the pipelines (default, manual, and ML pipelines), and surfaces (pial and WM). For the comparison between the pipelines, a post-hoc comparison was performed using the Wilcoxon matched-pairs signed-rank test, with p-values adjusted by Bonferroni correction for multiple comparisons. To assess inter-rater agreement, error detection in each grid was analyzed between the two raters using Cohen’s kappa. To confirm the effect of age on WMH volume in previous studies (Sachdev et al., 2007; Ylikoski et al., 1995; Atwood et al., 2004; Silbert et al., 2008), the correlation between the logarithm of WMH volume and age was tested using

Spearman’s rank correlation. We also analyzed the correlation between WMH volume and SDS5 in the default, manual, and ML pipelines using Spearman’s rank correlation.

Furthermore, we adopted a data-driven approach to define the improvement in the surface analysis of WMH-adapted pipelines. A substantial portion of the SDS5 changes after the application of WMH-adapted pipelines was distributed around zero, indicating that many MRI data with low WMH load did not show changes by the pipeline. There were, however, asymmetric changes in SDS5, that is, some data showed considerable decreases, while no data showed such increases. These interpretations were supported by the correlation analysis between the WMH volume and SDS5, as well as by the visual inspection of representative cases.

This information was used to define cases in which cortical analyses were improved by the WMH-adapted pipelines; the upper inner fence ($Q3+1.5 \times IQR$ of the data) was adopted as the criterion for the meaningful SDS5 decrease (thus, the improvement of cortical analysis). We then performed a receiver-operating characteristic (ROC) curve analysis to determine the optimal cutoff value of WMH volume, above which cortical estimation procedure likely benefits from the implementation of the proposed procedure. With the determined cutoff WMH volume, the sensitivity, specificity, AUC, and data with improvement were computed.

To gain insight into the potential effects of WMHs on the group comparison, the myelin and thickness maps were parcellated into 180 parcels using an HCP multimodal parcellation atlas (HCP-MMP, Version 1.0) (Glasser et al., 2016), upon completion of the cortical surface re-estimation process. We compared the myelin maps and cortical thickness maps of the default and ML pipelines. A paired t-test was applied as implemented in the Permutation Analysis of Linear Models (PALM) tool, which provided a non-parametric family-wise error (FWE) correction over multiple parcellations (Winkler et al., 2014, 2016).

3. Results

3.1. WMHs in the dataset

During the manual delineation procedure, WMHs were observed in all 43 participants. In the group-level frequency map of the WMHs, they were distributed most densely in the periventricular zones but also in the deep WM just beneath the frontal, temporal and parietal cortices, corresponding to the juxtacortical WMHs (Supplementary Fig. S1-a). The number of WMHs (combining all the periventricular WMH and deep WMH) ranged from 7–268 (median = 80.2), with a mean volume of $6.1 \pm 8.4 \text{ cm}^3$. Hereafter, unless otherwise noted, the term “WMH volume” refers to the volume of WMH including both periventricular WMH and

deep WMH. The age of the participants linearly correlated with the volume of manually defined WMHs after logarithmic transformation ($r = 0.63, p < 0.001$; Supplementary Fig. S1-b). The ML-predicted WMH volume showed a strong correlation with the manually defined WMH volume (Spearman's $\rho = 0.95, p < 0.001$). The ML-predicted WMHs had a sensitivity of 0.63, specificity of 0.999, and SI of 0.51 ± 0.16 , using the manually defined WMHs as the standard. Overall, from visual inspection, the ML-predicted WMH agreed well with the manually defined WMH. Moreover, a mismatch between the manually defined WMH and the ML-predicted WMH was present at the periphery of WMHs in the deep or periventricular WM; however, the mismatch was not evident in the juxtacortical area.

3.2. Characteristics of surface estimation errors

We observed an extreme case in which a small WMH caused extensive cortical surface errors when analyzed using the default pipeline (Fig. 3). In such cases, WMHs were mislabeled as parts of the GM because of the relatively low and high intensities of WMHs in T1w and T2w MRIs, and the WM surface was incorrectly estimated. The WM surfaces often invaded the WMHs, which was accompanied by pial surface errors (Fig. 3A, B). Across the participants, the WM surface errors often led to overestimation and sometimes underestimation (Fig. 3C) of the cortical thickness. Consequently, the bias-corrected myelin maps, which were calculated based on an inadequate definition of the cortical ribbon, often yielded a mosaic of exceptionally high/low myelin signals (Fig. 3D) in the vertices near the WMHs (Fig. 3B). SDM clearly revealed these cortical surface abnormalities (Fig. 3E).

The ML-predicted WMH masks reasonably overlapped with the WMHs (yellow outline in Fig. 3B, lower panel). After re-estimating the surfaces using either the manual or ML pipeline, the pial and WM surfaces were corrected (Fig. 3A and B, lower panels). Subsequently, the abnormalities in the myelin and thickness maps, as well as the SDM surface defect, disappeared (Fig. 3C-E, lower panels). Overall, at the individual level, the existence of WMHs increased errors in the estimation of both WM and pial surfaces in some cases, resulting in abnormalities in the myelin and thickness maps and SDM. Both the manual and ML pipelines appeared to reduce estimation errors.

3.3. Effects of WMH-adapted pipelines in surface QC

Two raters (T.O. and T.A.) blindly performed visual QC for the cortical surfaces across the three pipelines using the stereological method (Figs. 2 and 4). We assessed the inter-rater reliability of stereological QC by independently treating each grid-wise rating. The result showed a fair agreement between the two raters; Cohen's kappa was 0.27 and 0.34 in the WM and pial surface, respectively. Moreover, when the error count was pooled in a grid across the error types, the agreement reached Cohen's kappa of 0.57.

In the default pipeline, at least one surface error was found in either the WM or the pial surface in 100 % of the cases by rater 1 (median= 25, IQR= 20) and 88 % by rater 2 (median= 7, IQR= 14.5). ANOVA with rater, pipeline, and surface (WM or pial) as factors revealed significant effects of these factors on the error count: rater ($F(1,42)=212.7, p=2.2 \times 10^{-16}$), pipeline ($F(2,41)=17.3, p=3.7 \times 10^{-6}$) and surface ($F(1,42)=4.1, p=0.048$). There were also significant interactions between all three factors: the rater-by-pipeline interaction ($F(2,41)=13.0, p=4.1 \times 10^{-5}$), pipeline-by-surface interaction ($F(2,41)=7.7, p=1.5 \times 10^{-3}$), and surface-by-rater interaction ($F(1,42)=28.9, p=3.1 \times 10^{-6}$). Overall, the quantitative results of stereological QC indicated that the proposed pipeline significantly reduced the estimation errors in both the WM and pial surfaces. A post hoc pairwise comparison of the error count between the pipelines was performed for each rater (Fig. 4). For the WM surface error, significant differences were found between the default and ML pipelines in both raters (corrected $p = 1.6 \times 10^{-5}$ and 9.7×10^{-3} for rater 1 and 2, respectively). Between the manual and ML pipelines, rater 1 (corrected $p = 2.5 \times 10^{-6}$) but not rater 2 (corrected $p = 0.33$), found differences. No differences were noted between the default and manual pipelines for either rater. For the count of pial error, rater 1 found a significant difference between the default and ML pipelines (corrected $p = 4.2 \times 10^{-4}$) and between the manual and ML pipelines (corrected $p = 8.6 \times 10^{-4}$), but not between the default and manual pipelines (corrected $p = 0.31$). In addition, we did not find any difference in the pial error count between the pipelines for rater 2.

The SDS5 analysis indicated that both the manual and ML pipelines significantly reduced outliers compared with the default pipeline (Fig. 5). The median SDS5 values were 11 (44.5 IQR), 4 (14), and 4 (15.5) in the default, manual, and ML pipelines, respectively (Fig. 5A). For reference, the median (IQR) of SDS5 computed from the 1113

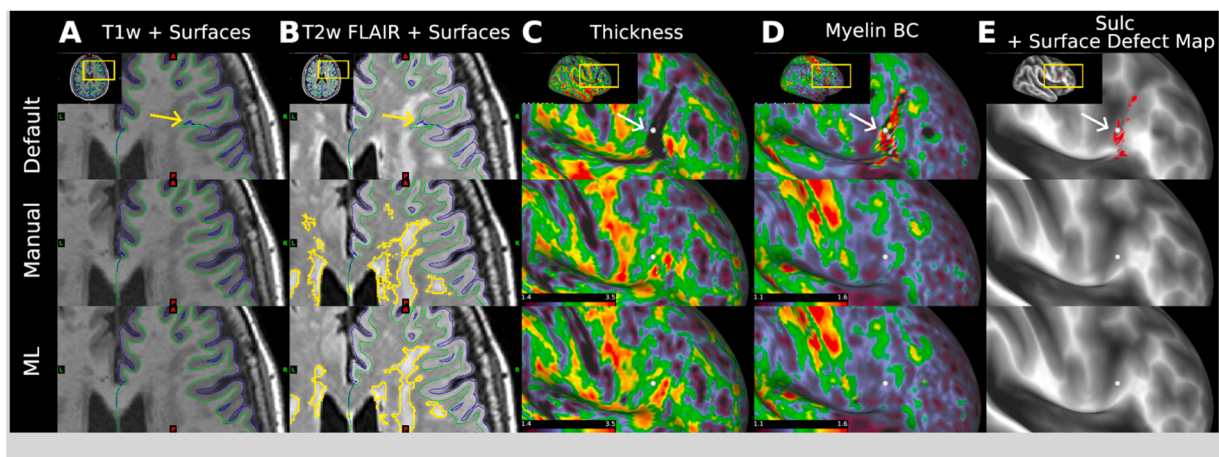


Fig. 3. A representative of white matter (WM) hyperintensity (WMH) in T1w and T2w FLAIR MRI and cortical surface errors. The top row shows the default findings in a representative case in whom the left frontal subcortical WMH (yellow arrows) caused errors of WM surfaces (lime line) and pial surfaces (blue line) overlain on T1w MRI (A) and T2w MRI (B). The surface analysis yields errors at the corresponding surface vertex (white arrows in C, D, E). Cortical thickness of the corresponding vertex is likely underestimated as compared with those in the surrounding vertices (C), whereas bias-corrected myelin (myelin BC) is likely over- and underestimated (D). The surface defect map (SDM) detects outliers (red areas in E). The middle and bottom rows show the corrected WM and pial surfaces after the reanalysis with the manual pipeline and the machine learning (ML) pipeline, respectively. The yellow line in (B) shows the contour of WMH defined by manually or estimated by ML. The corrected thickness (C) and myelin BC (D) maps, and SDM (E). The white spheres in C, D, E show the identical vertex corresponding to the tip of the yellow arrow in panel A and B.

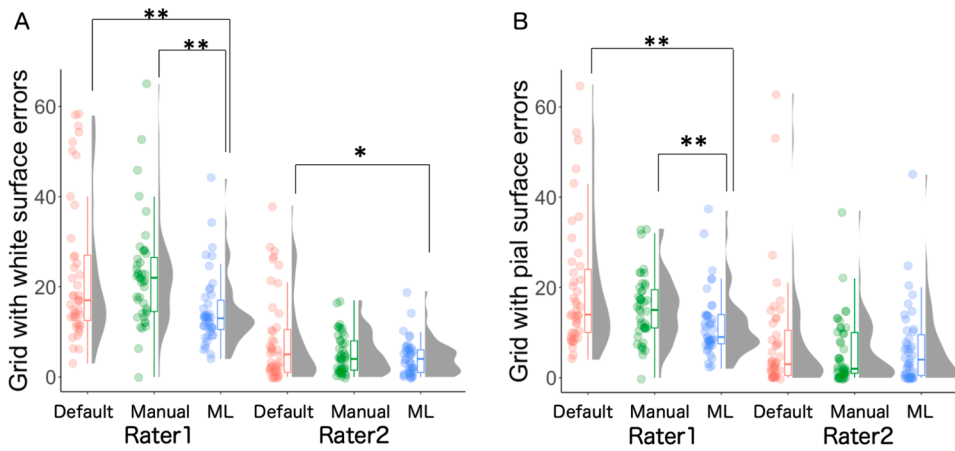


Fig. 4. Effects of WMH-adapted pipelines on the visual QC by the two raters in the rain cloud plots. The number of errors at the WM surface (A) and the pial surface (B) per participant is shown for each rater across the default (red), manual (green) and ML (blue) pipelines (see also Fig. 2). Shown are data points (each participant), box plot (w/ median (Q2), Q1 and Q3 with whiskers showing the largest and smallest data point excluding outliers), and smoothed histogram (gray). **Corrected $P < 0.01$ and *corrected $P < 0.001$ in Wilcoxon Signed rank test.

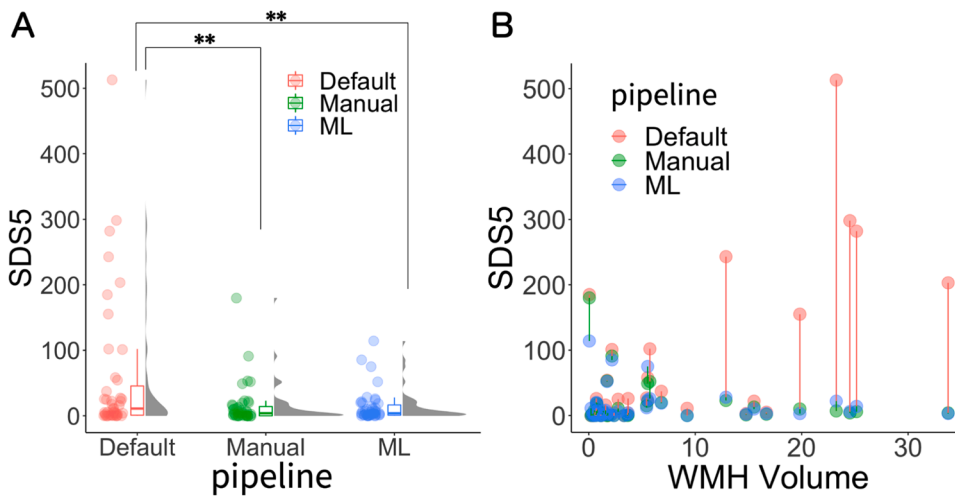


Fig. 5. Effects of WMH-adapted pipelines in the automatic QC of cortical surfaces.

A Comparison of surface defect scores (SDS5) between three groups of default, manual and ML pipeline. SDS5 in 3 pipelines are shown by the datapoint, boxplot and histogram in the rain cloud plots. Wilcoxon signed rank test, **corrected $p < 0.001$.

B A scatter plot of the SDS5 against the manually defined WMH volume at the individual level (filled circles). In many cases, the SDS5 decreased in the manual (green) and ML (blue) pipeline compared with the default pipeline (red). The three data points of the same individual are connected with colored lines so that the line color corresponds to the color of the data symbol with a higher SDS5 value. The WMH volume had a correlation with SDS5 only in the default pipeline (corrected $p = 3.6 \times 10^{-5}$, $\rho = 0.61$), but this correlation disappeared in the manual (corrected $p = 0.39$, $\rho = 0.23$) and ML pipeline (corrected $p = 0.45$, $\rho = 0.23$). The surface estimation errors due to

WMH and their improvement by the proposed pipeline were found mainly in the cases with WMH volume $> 5.6 \text{ cm}^3$ when the improvement of SDS5 of > 42.5 ($Q3 + 1.5IQR$) was adopted as a criterion (sensitivity of 0.80, specificity of 0.86 and accuracy of 0.84). Among twelve individuals with WMH volume greater than this cutoff, seven showed the reduction of SDS with the ML pipeline. Among eight individuals showing the reduction of SDS5 with the ML pipeline, seven individuals showed WMH volume greater than this cutoff.

participants of the YA-HCP dataset was 0 (3). The ANOVA results showed significant differences in SDS5 across the pipelines ($p = 0.001$). Wilcoxon signed-rank tests with Bonferroni correction revealed a significant reduction in SDS5 ($p < 0.001$) in both the manual and ML pipelines compared to the default pipeline.

The manually defined WMH volume was positively correlated with SDS5 in the default pipeline (corrected $p = 3.6 \times 10^{-5}$, $\rho = 0.61$) (Fig. 5B). Notably, the correlation of the WMH volume with SDS5 was no longer significant in either the manual pipeline (corrected $p = 0.39$, $\rho = 0.23$) or ML pipeline (corrected $p = 0.45$, $\rho = 0.23$).

The difference in SDS5 between the default and ML pipelines was not normally distributed (see method 2.3.7); hence, we used a value of 42.5, corresponding to the upper inner fence ($Q3 + 1.5 \times IQR$ of the data) as the criterion for improvement with the proposed pipeline. With this criterion, eight individuals showed improvement. In the ROC curve analysis, the detection of the improved cases was maximum with a cutoff WMH volume of 5.6 cm^3 with an AUC of 0.85 (Supplementary Fig. S2). We identified twelve individuals with WMH volumes greater than this cutoff, among which seven showed a reduction in SDS5 with the ML pipeline.

When we compared the group-wise surface defect frequency map

between the default and ML pipelines, it was obvious that the application of the proposed pipeline substantially reduced the outliers likely caused by WMHs (Supplementary Fig. S3a). However, there might also be other types of surface errors, which were not reduced even after the application of our proposed pipeline, in the medial surface of the brain (cingulate areas) and pre- and post-central gyri. Although such non-WMH causes of the surface error were outside the scope of the present study, we examined individual cases caused by apparently non-WMH causes to explore the future application of the SDS method. The surface errors in the pre- and post-central gyri were potentially because of the residual B1 bias (Supplementary Fig. S3b) as suggested by the hemispheric asymmetry (Glasser et al., NeuroImage 2022). To summarize, the SDS seemed useful to detect non-WMH-related surface errors as well, but its potential should be addressed in future studies.

The impact of WMHs on surface estimation can be extended even to the HCP-style group-level parcellation analysis that is widely performed in the field. We compared cortical thickness and myelin maps before and after consideration of WMHs. In the paired t-test of the parcellation-wise thickness map the differences between the default and ML pipelines reached statistical significance after the correction for multiple comparisons (PALM, $p < 0.05$, family wise error corrected) in several surface

areas (R_6a, R_46, R_9-46d, R_9a, R_8BL, R_9m, R_p32, R_p32pr, R_v2, R_PIT, R_IFSa, R_a47r, R_a10p, R_11l, R_10pp, R_OFc, R_FOP4, L_10r, L_VMV1, Supplementary Fig. S4-a). As for the myelin BC map, there was no statistically significant difference between the default and ML pipelines.

4. Discussion

Herein, we report that cortical surface estimation errors are algorithmically caused by age-related WMHs and propose automated correction and QC methods for cortical surface errors due to WMHs. Surface estimation errors occurred because WMHs were mislabeled as parts of the GM because of the relatively low and high intensities of WMHs in T1w and T2w MRIs, respectively. We used previously proposed ML algorithms (BIANCA and LOCATE) to automatically generate WMH masks to correct mislabeled WM segments and re-estimate cortical surfaces. The correlation between age and the logarithm of WMH volume is consistent with previous reports in which the WMH volume increased monotonically as a function of age (Atwood et al., 2004; Silbert et al., 2008). Two blinded raters (T.O. and T.A.) warranted a decrease in the surface estimation errors in the proposed pipeline. Furthermore, SDS5, a novel surface QC metric, sensitively detected a reduction in outliers in the proposed pipeline compared with the default pipeline. Finally, the adverse effects of WMHs on the surface estimation accuracy can be observed even in HCP-style group-level parcellation analysis, which is widely performed in the field. The proposed pipeline improves the reliability of surface-based MRI analyses in middle-aged to older people with WMHs and may contribute to disentangling the effects of WMHs and cortical abnormalities following aging and pathological processes. The present methodology may help researchers find reliable imaging markers for aging and age-related neuropsychiatric disorders.

Hunting imaging markers have become increasingly important in the clinical application of neuroimaging for neurodegenerative and psychiatric disorders. Recent studies have used cortical-surface analysis to identify neurobiological changes in the cortex during health and disease (Bethlehem et al., 2022; Cho et al., 2013; Lemaitre et al., 2012; Salat et al., 2004, 2009). WM often affects various brain disorders related to aging; however, it has been difficult to uncover the contribution of WMHs to their pathophysiology. Alzheimer's disease and Parkinson's disease accompanying WMHs likely form a continuum with vascular dementia and vascular parkinsonism, respectively, making it difficult to disentangle the effects of WMHs on pathophysiology. Future neuroimaging studies in neurodegenerative disorders should consider the effects of vasculopathy, including the disconnecting effects of WMHs, overlain by those of proteinopathy and neurocircuitopathy (Wakasugi and Hanakawa, 2021). WMHs are frequently found on brain MRIs in older populations (Sachdev et al., 2007; Ylikoski et al., 1995; Atwood et al., 2004; Silbert et al., 2008). Therefore, a detailed investigation of WM and cortical integrity and their relationship with WMHs is important for a mechanistic understanding of the pathophysiology of cognitive and motor disturbances, which are cardinal symptoms of Alzheimer's disease and Parkinson's disease, respectively.

In this study, the current ML algorithm for predicting WMH worked reasonably well, and the validation was comparable to those in the literature. Furthermore, we created a WMH probability map using BIANCA and adopted LOCATE for threshold optimization, yielding ML-predicted WMHs. The ML-predicted WMH volume correlated with the manually defined WMH volume, supporting the hypothesis that BIANCA plus LOCATE reasonably predicted WMHs. In a previous study, the performance (SD) of BIANCA plus LOCATE varied across the datasets (from 0.64 ± 0.23 to 0.73 ± 0.13) (Sundaresan et al., 2019). The present BIANCA plus LOCATE showed poorer sensitivity (0.63) and better specificity (0.99) than the sensitivity (0.81) and specificity (~ 0.98) of the Vrije Universiteit Amsterdam dataset (Sundaresan et al., 2019) which showed WMH volumes similar to ours. A reason behind the relatively low performance of BIANCA plus LOCATE in our setting may

be that we were not able to fully optimize the BIANCA plus LOCATE mainly because the WMHs in our MRI dataset were limited. Still ML-generated WMHs were effective in reducing surface errors, thereby fulfilling our study purpose. Therefore, although ML performance can further be improved, we were able to demonstrate the utility of the proposed method to reduce errors in cortical surface estimation.

Overall, the stereological visual assessment successfully detected a reduction in cortical surface errors using the ML-assisted HCP pipeline relative to the default pipeline. To date, many attempts have been made to establish a QC workflow of cortical surfaces estimated by surface reconstruction algorithms. Most have proposed QC workflows based on visual information, for which many documents are available online (Supplementary Table S1). Visual QCs rely on the visual inspection of preprocessed outputs through graphical user interfaces and are considered to be the gold standard for surface-based analysis of neuroimaging data. These visual QC approaches allow assessment by qualitative grading (Backhausen et al., 2016) subject to raters' experience and time limitations (Monereo-Sánchez et al., 2021). Visual QCs have obvious limitations in terms of inter-rater variability and quantification difficulty. Several studies have proposed semi-quantified visual QCs. A QC rating framework, the 'VisualQC' (Raamana et al., 2021; Raamana et al., 2018), by creating a visualization system that allows easy assessment of cortical parcellations by raters. However, the 'VisualQC' uses relatively small numbers and orientations of image slices (two rows by six slices by default), which can be customizable by the users; thus, visual input information is not necessarily standardized. In addition, the rater's assessment in VisualQC is not based on stereology; thus, rater responses are recorded for each individual's subject of interest. Therefore, previous visual QC methods may not be fully quantitative, and no previous QC methods provide specific types of surface errors (e.g., errors in the WM, pial matter, or both). With this background, we applied a stereological approach for quantitative estimation of the surface errors and validation of the WMH-adapted pipeline.

The present stereological techniques allow for visual QC with a reasonable level of quantification and standardization. Stereological assessment is an unbiased quantification method for neuroanatomy (Saper 1996; Zhao and van Praag 2020). It allows the standardized evaluation of microscopic neurobiological units (e.g., the number of neurons, synapses, and axons) to be compared across a large tissue space. However, if the stereological approach is not well-standardized, the quantified results may be highly variable across laboratories or raters by as much as 300 % (Herculano-Houzel et al., 2015). The bias is caused by various factors, including differences in sampling strategies across raters. Therefore, we developed a stereological QC system by standardizing the evaluation units across different MRIs, reducing arbitrariness, and increasing the spatial uniformity of the error counts. Unexpectedly, however, the results from the stereological QC showed only a fair level of inter-rater agreement of the error count when analyzed separately for the white and pial surfaces ($\kappa = \sim 0.3$) or a moderate level of agreement when both error types were pooled ($\kappa = 0.57$). In addition, we found a significant interaction effect among the rater, pipeline, and surface type factors. This means that error detectability is influenced by many factors known for radiological diagnosis, such as spectrum bias and misclassification bias (Pavlou et al., 2021). The low agreement between the two raters can be attributed to the limitation of subjective visual assessment, and our stereological QC method should be further validated in terms of intra-rater agreement. Together, our results indicate that the stereological method applies to the visual QC of cortical surface errors, but further elaboration is needed to achieve fully reliable stereological QC.

Furthermore, in this study, we developed an alternative automated QC algorithm that calculates the SDM and derived score (SDS5) for vertex-wise surface error detection. We found that SDM/SDS5 was useful for assessing the presence of surface errors and their reduction using the elaborated pipelines. There has been an attempt to automate quantifiable QC for cortical surface errors using quality metrics, such as

FreeSurfer's topological defect count and curvature smoothness (Ségonne et al., 2007; Tian et al., 2021). However, these topological errors are counted during the initial estimation of WM surfaces; thus, it is not clear whether such errors at an intermediate process are directly associated with the errors in the final results, including the thickness or myelin maps. Moreover, all thickness or myelin errors may not necessarily accompany topological defect count. The SDM in our study allows the joint assessment of thickness and myelin maps quantitatively based on the reference values of cortical metrics in the YA-HCP1200 database. The vertices on the abnormally estimated cortical surfaces often have extreme outlier values in both thickness and myelin values; thus, we defined the statistical threshold of SDS5 in reference to the SDS5 computed from the thickness and myelin maps of YA-HCP1200. Both SDM and SDS5 can be calculated automatically as long as an adequate structural MRI dataset is analyzed using HCP pipelines. The positive correlation of SDS5 with the WMH volume in the default pipeline indicated that SDS5 indeed reflected errors in the presence of WMHs (see Fig. 4b). Moreover, this correlation was abolished after the application of WMH-adapted pipelines (see Fig. 4d). The SDM also proved useful for visualizing the location of the surface errors and their reduction using the proposed WMH-adapted pipelines. Accordingly, SDS and SDM5 provide useful information for automatic and quantitative estimation of surface QC.

Mapping the surface estimation error by SDM is also likely useful for identifying the various error sources. SDM detected errors related to not only WMHs but also B1 transmission bias contamination in spin echo-based T2-weighted images. The errors in the pre- and post-central gyri would primarily be explained by the residual B1 bias. The B1 transmission bias remains even though T1w- and T2w-based B1 biasfield corrections are applied in the PreFreeSurfer pipeline. This is primarily because T1w- and T2w-based B1 biasfield correction removes the bias of the B1 receive field rather than the transmission field (Glasser and Van Essen, 2011; Glasser et al., 2022). The presence of hemispheric asymmetry supports the possibility of B1 transmission bias (Glasser et al., NeuroImage 2022). Future studies may be required to address this issue by estimating the B1 transmission field in volume space from add-on scanning sequences such as Sa2RAGE (Eggenschwiler et al., 2012) and spin-echo and gradient-echo EPI (Glasser et al., 2022) and by feeding them into the FreeSurfer-based re-estimation of the cortical surface. We also note that we could not exclude the possibility that the surface errors in the pre- and post-central gyri reflected, at least in part, age-related changes in the cortical thickness and myelin maps because of the age-gap between the HCP-YA and the current dataset.

The present study had some limitations. First, the generalizability of the proposed WMH-adapted pipeline is limited. The current method was built on a small amount of data; thus, further investigation using a larger dataset may be required. Current MRI data were collected using a legacy MRI protocol (spatial resolution of 1 mm isovoxel), which was not optimized for surface reconstruction; thus, the results of the surface errors might have been overestimated. Future studies with MRIs acquired with a higher spatial resolution, such as the HCP-style protocol (spatial resolution of 0.8 mm), would be less likely to have surface estimation errors.

Second, the differences in scanning protocol and populations may bias the results of SDS5 because SDS calculation is based on the YA-HCP dataset. Recent studies indeed suggest that even using identical scanning protocols, measured cortical thickness and myelin map are affected by the differences in the sites and MRI scanners (Koike et al., 2021). We also found SDS5 was sensitive to the B1 biasfield in the pre- and post-central gyri, which remained unchanged by the proposed pipeline. Further elaboration is needed for qualifying surface metrics and their errors in a more generalized manner, for example, by implementing harmonization methods across the MRI protocols and scanners (Maikusa et al., 2021 and Sun et al., 2022) and by using an age-matched MRI database. The parameters and threshold of SDS5/SDM may need to be adjusted depending on the future optimization for each study. That being said, we

consider that the SDS5/SDM providing a relative yet quantitative vertex-by-vertex QC value is a useful method among the currently available surface QC methods. The proposed pipeline can be applied to HCP-style MRI protocols in the Brain/MINDS-beyond project (Koike et al. 2021) to test the adverse effects of WMHs. That said, our proposed pipeline should be valid because the reconstruction algorithm itself is subject to segmentation errors due to altered contrast in WMH (see Section 3.2) and is not directly related to the spatial resolution. The validity of the training of the WMH discriminator machine (BIANCA) may be evaluated more rigorously using independent datasets.

Third, it has not yet been investigated whether the proposed pipelines can be generally used for other neurological conditions accompanying hyperintensities in the WM, such as multiple sclerosis (Wattjes et al., 2015), progressive multifocal leukoencephalopathy, neuro-myelitis optica spectrum disorder (Pache et al., 2016), cerebral amyloid angiopathy (Subotic et al., 2021), cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) (De Lucia et al., 2022), Binswanger disease (Yin et al., 2014), schizophrenia, and mood disorders (Zanetti, et al 2018; Serafini et al., 2014; Lyoo et al., 2002). Indeed, some of these studies applied cortical surface analysis to understand cortical pathophysiology, but its association with WMHs has not been well characterized. The WMH discriminator machine and surface re-reconstruction algorithm may need to be optimized for each disease population because the pattern of the spatial distribution of WMH is often different between diseases, for example, those related to small vasculopathy vs. leukoaraiosis (Rosenberg et al., 2016).

Fourth, the application of our proposed method to the medial temporal regions was outside the scope of this study. Indeed, the medial temporal regions were excluded from both stereological QC (by means of the instructions) and SDS evaluations (by means of the "medial wall") due to potential inaccuracies in the estimation of cortical surfaces in these regions, regardless of the WMHs. Poor segmentation performance is especially problematic in aged patients with medial temporal atrophy. To address the effects of WMHs on the surface analysis of the medial temporal lobe, one possible approach would be to incorporate a correction for volumetric measures for, both WMHs and atrophy (Griffanti et al., 2022 Neuroimage: Clinical). Therefore, this combined correction method should be tested in the future to address the surface analysis of the medial temporal lobe in aged patients with WMHs and atrophy.

Given the ensured generalizability, the proposed pipeline will allow for reliable assessment of cortical indices in large-scale lifespan cohort studies. Although recent literature suggests that aging is an independent factor affecting cortical thickness (Salat et al., 2004; Bethlehem et al., 2022) and myelin (Grydeland et al., 2013), it remains unclear how such age-related cortical structural changes perturb cortical functions and the resultant behaviors. Structural disconnection in WM might be more specifically linked to the disorganization of functional connectivity responsible for behavioral deficits than the disruption of GM areas (Griffis et al., 2019). Disintegrity in WM may disconnect cortico-cortical and subcortical fiber bundles, thus causing dysfunction of the connected brain structures (Thiebaut de Schotten, Foulon, and Nachev 2020). Therefore, future studies are needed to investigate the effects of age-related WMHs on functional activity or connectivity in cortical and subcortical brain areas. In such studies, it may be worth testing the usefulness of the proposed pipeline for segregating subcortical from cortical views and refining the pathophysiology of neuropsychiatric disorders.

5. Conclusion

In conclusion, the current study revealed that surface-based analysis is prone to surface estimation errors due to WMHs and that these errors can be corrected by automated WMH prediction followed by surface re-estimation based on multimodal MRI datasets. The expected correction was confirmed by both stereological visual QC and automated surface

QC with SDM/SDS5; however, the latter seemed more sensitive to subtle changes in quality. Validation of the proposed WMH-adapted pipeline warrants further investigation into its usefulness and refinement in a larger population study.

Funding and acknowledgments

This study was in part supported by a grant (Brain/MINDS-beyond) from the Japan Agency of Medical Research and Development (AMED) (JP18dm0307006, JP19dm0307004, JP22dm0307002 to Takuya Hayashi and JP18dm0307003 and 18dm0207070 to Takashi Hanakawa) and JSPS KAKENHI (19H05726 and 19H03536) to Takashi Hanakawa.

Data availability

The processing toolbox is available at <https://surfer.nmr.mgh.harvard.edu>.

The analysis pipeline is available at <https://github.com/Washington-University/HCPpipelines>.

The locally adaptive thresholds estimation is available at <https://git.fmrib.ox.ac.uk/vaanathi/LOCATE-BIANCA>.

The Human Connectome Project 1200 Subjects Release is available at <https://balsa.wustl.edu/reference/pkXDD>

The QC and reporting tool is available at <https://github.com/RIKEN-BCIL/bcil/blob/master/bin/hcpipe.qc>.

CRedit authorship contribution statement

Yuki Oi: Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Masakazu Hirose:** Investigation. **Hiroki Togo:** Formal analysis, Investigation, Visualization. **Kenji Yoshinaga:** Formal analysis, Investigation, Visualization. **Thai Akasaka:** Methodology, Formal analysis, Investigation. **Tomohisa Okada:** Methodology, Formal analysis, Investigation. **Toshihiko Aso:** Methodology. **Ryosuke Takahashi:** Project administration. **Matthew F. Glasser:** Methodology, Writing – review & editing. **Takuya Hayashi:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Resources, Supervision, Project administration, Funding acquisition. **Takashi Hanakawa:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Resources, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Tomohisa Okada received a grant from Siemens Healthcare K.K., Japan for a 7 Tesla MR research. The remaining authors declare no conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2023.120377](https://doi.org/10.1016/j.neuroimage.2023.120377).

References

Anticevic, A., Dierker, D.L., Gillespie, S.K., Repovs, G., Csernansky, J.G., Van Essen, D.C., Barch, D.M., 2008. Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. *Neuroimage* 41, 835–848.

Atwood, L.D., Wolf, P.A., Heard-Costa, N.L., Massaro, J.M., Beiser, A., D'Agostino, R.B., DeCarli, C., 2004. Genetic variation in white matter hyperintensity volume in the framingham study. *Stroke* 35, 1609–1613.

Backhausen, L.L., Herting, M.M., Buse, J., Roessner, V., Smolka, M.N., Vetter, N.C., 2016. Quality control of structural MRI images applied using freesurfer-a hands-on workflow to rate motion artifacts. *Front. Neurosci.* 10, 558.

Bando, K., Honda, T., Ishikawa, K., Takahashi, Y., Mizusawa, H., Hanakawa, T., 2019. Impaired adaptive motor learning is correlated with cerebellar hemispheric gray matter atrophy in spinocerebellar ataxia patients: a voxel-based morphometry study. *Front. Neurol.* 10, 1183. <https://doi.org/10.3389/fneur.2019.01183>.

Barkhof, F., Scheltens, P., 2002. Imaging of white matter lesions. *Cerebrovasc. Dis.* 13 (Suppl 2), 21–30.

Bethlehem, R.A.I., Seidlitz, J., White, S.R., Vogel, J.W., Anderson, K.M., Adamson, C., Adler, S., Alexopoulos, G.S., Anagnostou, E., Arces-Gonzalez, A., Astle, D.E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S.A., Benegal, V., Beyer, F., Blangero, J., Blesa Cabez, M., Boardman, J.P., Borzage, M., Bosch-Bayard, J.F., Bourke, N., Calhoun, V.D., Chakravarty, M.M., Chen, C., Chertavian, C., Chetelat, G., Chong, Y.S., Cole, J.H., Corvin, A., Costantino, M., Courchesne, E., Crivello, F., Cropley, V.L., Crosbie, J., Crossley, N., Delarue, M., Delorme, R., Desrivieres, S., Devenyi, G.A., Di Biase, M.A., Dolan, R., Donald, K.A., Donohoe, G., Dunlop, K., Edwards, A.D., Ellison, J.T., Ellis, C.T., Elman, J.A., Eyler, L., Fair, D.A., Feczko, E., Fletcher, P.C., Fonagy, P., Franz, C.E., Galan-Garcia, L., Gholipour, A., Giedd, J., Gilmore, J.H., Glahn, D.C., Goodyer, I.M., Grant, P.E., Groenewold, N.A., Gunning, F.M., Gur, R.E., Gur, R.C., Hammill, C.F., Hansson, O., Hedden, T., Heinz, A., Henson, R.N., Heuer, K., Hoare, J., Holla, B., Holmes, A.J., Holt, R., Huang, H., Im, K., Ipser, J., Jack Jr, C.R., Jackowski, A.P., Jia, T., Johnson, K.A., Jones, P.B., Jones, D.T., Kahn, R.S., Karlsson, H., Karlsson, L., Kawashima, R., Kelley, E.A., Kern, S., Kim, K.W., Kitzbichler, M.G., Kremen, W.S., Lalonde, F., Landeau, B., Lee, S., Lerch, J., Lewis, J.D., Li, J., Liao, W., Liston, C., Lombardo, M.V., Lv, J., Lynch, C., Mallard, T.T., Marcelis, M., Markello, R.D., Mathias, S.R., Mazoyer, B., McGuire, P., Meaney, M.J., Mechelli, A., Medic, N., Misić, B., Morgan, S.E., Mothersill, D., Nigg, J., Ong, M.Q.W., Ortinau, C., Ossenkopppe, R., Ouyang, M., Palaniyappan, L., Paly, L., Pan, P.M., Pantelis, C., Park, M.M., Paus, T., Pausova, Z., Paz-Linares, D., Pichet Binette, A., Pierce, K., Qian, X., Qiu, J., Qiu, A., Raznahan, A., Rittman, T., Rodrigue, A., Rollins, C.K., Romero-Garcia, R., Ronan, L., Rosenberg, M.D., Rowitch, D.H., Salum, G.A., Satterthwaite, T.D., Schaare, H.L., Schachar, R.J., Schultz, A.P., Schumann, G., Schöll, M., Sharp, D., Shinohara, R.T., Skoog, I., Smyser, C.D., Sperling, R.A., Stein, D.J., Stolicyn, A., Suckling, J., Sullivan, G., Taki, Y., Thyreau, B., Toro, R., Traut, N., Tsvetanov, K.A., Turk-Browne, N.B., Tuulari, J.J., Tzourio, C., Vachon-Preseuse, E., Valdes-Sosa, M.J., Valdes-Sosa, P.A., Valk, S.L., van Amelsvoort, T., Vandekar, S.N., Vasung, L., Victoria, L.W., Villeneuve, S., Villringer, A., Vertes, P.E., Wagstyl, K., Wang, Y.S., Warfield, S.K., Warrior, V., Westman, E., Westwater, M.L., Whalley, H.C., Witte, A.V., Yang, N., Yeo, B., Yun, H., Zalesky, A., Zar, H.J., Zettergren, A., Zhou, J.H., Ziauddeen, H., Zugman, A., Zuo, X.N., 3R-BRAIN, AIBL, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Repository Without Borders Investigators, CALM Team, Cam-CAN, CCNP, COBRE, cVEDA, ENIGMA Developmental Brain Age Working Group, Developing Human Connectome Project, FinnBrain, Harvard Aging Brain Study, IMAGEN, KNE96, Mayo Clinic Study of Aging, NSPN, POND, PREVENT-AD Research Group, VETSA, Bullmore, E.T., Alexander-Bloch, A.F., 2022. Brain charts for the human lifespan. *Nature* 604, 525–533.

Bohnen, N.L., Albin, R.L., 2011. White matter lesions in Parkinson disease. *Nat. Rev. Neurol.* 7, 229–236.

Bookheimer, S.Y., Salat, D.H., Terpstra, M., Ances, B.M., Barch, D.M., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Diaz-Santos, M., Elam, J.S., Fischl, B., Greve, D.N., Hagy, H.A., Harms, M.P., Hatch, O.M., Hedden, T., Hodge, C., Japardi, K.C., Kuhn, T.P., Ly, T.K., Smith, S.M., Somerville, L.H., Uğurbil, K., van der Kouwe, A., Van Essen, D., Woods, R.P., Yacoub, E., 2019. The lifespan human connectome project in aging: an overview. *Neuroimage* 185, 335–348.

Caunca, M.R., De Leon-Benedetti, A., Latour, L., Leigh, R., Wright, C.B., 2019. Neuroimaging of Cerebral Small Vessel Disease and Age-Related Cognitive Changes. *Front. Aging Neurosci.* 11, 145.

Coalson, T.S., Van Essen, D.C., Glasser, M.F., 2018. The impact of conventional neuroimaging methods on the spatial localization of cortical areas. In: *Proceedings of the National Academy of Sciences of the United States of America*, 115, pp. E6356–E6365.

Cho, H., Jeon, S., Kang, S.J., Lee, J.M., Lee, J.H., Kim, G.H., Shin, J.S., Kim, C.H., Noh, Y., Im, K., Kim, S.T., Chin, J., Seo, S.W., Na, D.L., 2013. Longitudinal changes of cortical thickness in early- versus late-onset Alzheimer's disease. *Neurobiol. Aging* 34 (1921), e9–1921 e15.

Dadar, M., Potvin, O., Camicioli, R., Duchesne, S., Alzheimer's Disease Neuroimaging Initiative, 2021. Beware of white matter hyperintensities causing systematic errors in FreeSurfer gray matter segmentations! *Hum. Brain Mapp.* 42, 2734–2745.

de Leeuw, F.E., de Groot, J.C., Achten, E., Oudkerk, M., Ramos, L.M., Heijboer, R., Hofman, A., Jolles, J., van Gijn, J., Breteler, M.M., 2001. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam scan study. *J. Neurol. Neurosurg. Psychiatry* 70, 9–14.

De Lucia, N., Ragno, M., Paci, C., Cacchiò, G., Caiazzo, A.R., Tiberi, S., De Rosa, A., Navarra, R., Caulo, M., De Michele, G., Trojano, L., 2022. Constructional impairments and their neural correlates in nondemented adults with cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Cogn. Behav. Neurol.* <https://doi.org/10.1097/WNN.0000000000000318>.

Eggenchwil, F., Kober, T., Magill, A.W., Gruetter, R., Marques, J.P., 2012. SA2RAGE: a new sequence for fast B1+ -mapping. *Magn. Reson. Med.* 67, 1609–1619.

Elam, J.S., Glasser, M.F., Harms, M.P., Sotiropoulos, S.N., Andersson, J.L.R., Burgess, G.C., Curtiss, S.W., Oostenveld, R., Larson-Prior, L.J., Schoffelen, J.M., Hodge, M.R.,

- Cler, E.A., Marcus, D.M., Barch, D.M., Yacoub, E., Smith, S.M., Ugurbil, K., Van Essen, D.C., 2021. The human connectome project: a retrospective. *Neuroimage* 244, 118543.
- Fazekas, F., Chawluk, J.B., Alavi, A., Hurtig, H.I., Zimmerman, R.A., 1987. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *Am. J. Neuroradiol.* 8, 421–426.
- Fischl, B., 2012. FreeSurfer. *Neuroimage* 62, 774–781.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11050–11055.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klavness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Fischl, S., Busa, A., Dieterich, H., Kouwe, K., Kennedy, K., Montillo, M., Rosen, D., 2008. Whole brain segmentation: neurotechnique automated labeling of neuroanatomical structures in the human brain. *Neuron*.
- Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59, 1369–1381.
- Glasser, M.F., Coalson, T.S., Harms, M.P., Xu, J., Baum, G.L., Autio, J.A., Auerbach, E.J., Greve, D.N., Yacoub, E., Van Essen, D.C., Bock, N.A., Hayashi, T., 2022. Empirical transmit field bias correction of T1w/T2w myelin maps. *Neuroimage* 258, 119360.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124.
- Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616.
- Gootjes, L., Teipel, S.J., Zebuhr, Y., Schwarz, R., Leinsinger, G., Scheltens, P., Möller, H. J., Hampel, H., 2004. Regional distribution of white matter hyperintensities in vascular dementia, Alzheimer's disease and healthy aging. *Dement. Geriatr. Cogn. Disord.* 18, 180–188.
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72.
- Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., Schulz, U.G., Kuker, W., Battaglini, M., Rothwell, P.M., Jenkinson, M., 2016. BIANCA (brain intensity AbNormality classification algorithm): a new tool for automated segmentation of white matter hyperintensities. *Neuroimage* 141, 191–205.
- Griffis, J.C., Metcalf, N.V., Corbetta, M., Shulman, G.L., 2019. Structural disconnections explain brain network dysfunction after stroke. *Cell Rep.* 28, 2527–2540 e9.
- Grydeland, H., Walhovd, K.B., Tamnes, C.K., Westlye, L.T., Fjell, A.M., 2013. Intracortical myelin links with performance variability across the human lifespan: results from T1- and T2-weighted MRI myelin mapping and diffusion tensor imaging. *J. Neurosci.* 33, 18618–18630.
- Koike, S., Tanaka, S.C., Okada, T., Aso, T., Yamashita, A., Yamashita, O., Asano, M., Maikusa, N., Morita, K., Okada, N., Fukunaga, M., Uematsu, A., Togo, H., Miyazaki, A., Murata, K., Urushibata, Y., Autio, J., Ose, T., Yoshimoto, J., Araki, T., Glasser, M.F., Van Essen, D.C., Maruyama, M., Sadato, N., Kawato, M., Kasai, K., Okamoto, Y., Hanakawa, T., Hayashi, T., 2021. Brain/MINDS beyond human brain MRI project: A protocol for multi-level harmonization across brain disorders throughout the lifespan. *NeuroImage Clin.* 30, 102600.
- Kruit, M.C., van Buchem, M.A., Hofman, P.A.M., Bakkers, J.T.N., Terwindt, G.M., Ferrari, M.D., Launer, J.J., 2004. Migraine as a risk factor for subclinical brain lesions. *JAMA* 291, 427–434.
- Launer, J.J., 2003. Epidemiology of white-matter lesions. *Int. Psychogeriatr.* 1, 99–103, 15 Suppl.
- Lemaitre, H., Goldman, A.L., Sambataro, F., Verchinski, B.A., Meyer-Lindenberg, A., Weinberger, D.R., Mattay, V.S., 2012. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiol. Aging* 33 (617), e1–e9.
- Liao, D., Cai, J., Rosamond, W.D., Barnes, R.W., Hutchinson, R.G., Whitsel, E.A., Rautaharju, P., Heiss, G., 1997. Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study: the ARIC study. *Am. J. Epidemiol.* 145, 696–706.
- Li, B., Jang, I., Riphagen, J., Almkott, R., Yochim, K.M., Ances, B.M., Bookheimer, S. Y., Salat, D.H., Alzheimer's Disease Neuroimaging Initiative, 2021. Identifying individuals with Alzheimer's disease-like brains based on structural imaging in the Human Connectome Project Aging cohort. *Hum. Brain Mapp.* 42, 5535–5546.
- Lyoo, I.K., Lee, H.K., Jung, J.H., Noam, G.G., Renshaw, P.F., 2002. White matter hyperintensities on magnetic resonance imaging of the brain in children with psychiatric disorders. *Compr. Psychiatry* 43, 361–368.
- McCarthy, P., 2022. *FSleyes (Version 1.5.0)*. doi:10.5281/zenodo.7038115.
- Monereo-Sánchez, J., de Jong, J.J.A., Drenthen, G.S., Beran, M., Backes, W.H., Stehouwer, C.D.A., Schram, M.T., Linden, D.E.J., Jansen, J.F.A., 2021. Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations - insights from the Maastricht study. *Neuroimage* 237, 118174.
- Pache, F., Zimmermann, H., Finke, C., Lacheta, A., Papazoglou, S., Kuchling, J., Wuerfel, J., Hamm, B., Rupprecht, K., Paul, F., Brandt, A.U., Scheel, M., 2016. Brain parenchymal damage in neuromyelitis optica spectrum disorder - a multimodal MRI study. *Eur. Radiol.* 26, 4413–4422.
- Phuah, C.L., Chen, Y., Strain, J.F., Yechoor, N., Laurido-Soto, O.J., Ances, B.M., Lee, J.M., for the Alzheimer's Disease Neuroimaging Initiative, 2022. Association of data-driven white matter hyperintensity spatial signatures with distinct cerebral small vessel disease etiologies. *Neurology* 99, e2535–e2547. <https://doi.org/10.1212/WNL.0000000000201186>.
- Raamana, P.R., Theyers, A., Selliah, T., 2018. VisualQC: assistive tools for easy and rigorous quality control of neuroimaging data. *April*. doi:10.5281/ZENODO.
- Raamana, P.R., Theyers, A., Selliah, T., Bhati, P., Arnott, S.R., Hassel, S., Nanayakkara, N.D., Scott, C.J.M., Harris, J., Zamyadi, M., Lam, R.W., Milev, R., Müller, D.J., Rotzinger, S., Frey, B.N., Kennedy, S.H., Black, S.E., Lang, A., Masellis, M., Symons, S., Bartha, R., MacQueen, G.M., 2021. The ONDRI Investigators. Visual QC Protocol for FreeSurfer Cortical Parcellations from Anatomical MRI. *bioRxiv*, Strother, S.C.. <https://doi.org/10.1101/2020.09.07.286807>
- Robinson, E.C., Garcia, K., Glasser, M.F., Chen, Z., Coalson, T.S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P.V., Van Essen, D.C., Smith, S.M., Edwards, A.D., Hajnal, J., Jenkinson, M., Glocker, B., Rueckert, D., 2018. Multimodal surface matching with higher-order smoothness constraints. *Neuroimage* 167, 453–465.
- Sachdev, P., Wen, W., Chen, X., Brodaty, H., 2007. Progression of white matter hyperintensities in elderly individuals over 3 years. *Neurology* 68, 214–222.
- Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S.R., Busa, E., Morris, J. C., Dale, A.M., Fischl, B., 2004. Thinning of the cerebral cortex in aging. *Cereb. Cortex* 14, 721–730.
- Salat, D.H., Lee, S.Y., van der Kouwe, A.J., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. *Neuroimage* 48, 21–28.
- Saper, C.B., 1996. Any way you cut it: a new journal policy for the use of unbiased counting methods. *J. Comp. Neurol.* 364, 5.
- Scott, J.A., Braskie, M.N., Tosun, D., Thompson, P.M., Weiner, M., DeCarli, C., Carmichael, O.T., 2015. Alzheimer's disease neuroimaging initiative. *Front. Aging Neurosci.* 7, 221. Cerebral Amyloid and Hypertension are Independently Associated with White Matter Lesions in Elderly.
- Ségonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans. Med. Imaging* 26, 518–529.
- Silbert, L.C., Nelson, C., Howieson, D.B., Moore, M.M., Kaye, J.A., 2008. Impact of white matter hyperintensity volume progression on rate of cognitive and motor decline. *Neurology* 71, 108–113.
- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97.
- 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.
- Subotic, A., McCreary, C.R., Saad, F., Nguyen, A., Alvarez-Veronesi, A., Zwiers, A.M., Charlton, A., Beaudin, A.E., Ismail, Z., Pike, G.B., Smith, E.E., 2021. Cortical thickness and its association with clinical cognitive and neuroimaging markers in cerebral amyloid angiopathy. *J. Alzheimers. Dis.* 81, 1663–1671.
- Sundaresan, V., Zamboni, G., Le Heron, C., Rothwell, P.M., Husain, M., Battaglini, M., De Stefano, N., Jenkinson, M., Griffanti, L., 2019. Automated lesion segmentation with BIANCA: Impact of population-level features, classification algorithm and locally adaptive thresholding. *Neuroimage* 202, 116056.
- Thiebaut de Schotten, M., Foulon, C., Nachev, P., 2020. Brain disconnections link structural connectivity with function and behaviour. *Nat. Commun.* 11, 5094.
- Tian, Q., Zaretskaya, N., Fan, Q., Ngamsombat, C., Bilgic, B., Polimeni, J.R., Huang, S.Y., 2021. Improved cortical surface reconstruction using sub-millimeter resolution MPRAGE by image denoising. *Neuroimage* 233, 117946.
- Togo, H., Nakamura, T., Wakasugi, N., Takahashi, Y., Hanakawa, T., 2023. Interactions across emotional, cognitive and subcortical motor networks underlying freezing of gait. *Neuroimage Clin* 37, 103342. <https://doi.org/10.1016/j.nicl.2023.103342>.
- Tucholka, A., Fritsch, V., Poline, J.B., Thirion, B., 2012. An empirical comparison of surface-based and volume-based group studies in neuroimaging. *Neuroimage* 63, 1443–1453.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. AddisonWesley, Reading, MA.
- Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., WU-Minn HCP Consortium, 2012. The human connectome project: a data acquisition perspective. *Neuroimage* 62, 2222–2231.
- Wakasugi, N., Hanakawa, T., 2021. It is time to study overlapping molecular and circuit pathophysiology in Alzheimer's and Lewy body disease spectra. *Front. Syst. Neurosci.* 15, 777706.
- Wattjes, M.P., Steenwijk, M.D., Stangel, M., 2015. MRI in the diagnosis and monitoring of multiple sclerosis: an update. *Clin. Neuroradiol.* 25 (Suppl 2), 157–165.
- Williams, L.Z.J., Glasser, M.F., Alfaro-Almagro, F., Dahan, S., Fawaz, A., Coalson, T.S., Fitzgibbon, S.P., Suliman, M.A., Van Essen, D.C., Smith, S.M., Edwards, A.D., Robinson, E.C., 2023. Generalising the HCP multimodal cortical parcellation to UK Biobank. doi:10.1101/2023.03.14.532531.
- Winkler, A.M., Ridgway, G.R., Douaud, G., Nichols, T.E., Smith, S.M., 2016. Faster permutation inference in brain imaging. *Neuroimage* 141, 502–516.
- Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. *Neuroimage* 92, 381–397.

Y. Oi *et al.*

NeuroImage 281 (2023) 120377

- Yin, X., Liu, C., Gui, L., Zhao, L., Zhang, J., Wei, L., Xie, B., Zhou, D., Li, C., Wang, J., 2014. Comparison of medial temporal measures between Binswanger's disease and Alzheimer's disease. *PLoS One* 9, e86423.
- Ylikoski, A., Erkinjuntti, T., Raininko, R., Sarna, S., Sulkava, R., Tilvis, R., 1995. White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. *Stroke* 26, 1171–1177.
- Zhao, X., van Praag, H., 2020. Steps towards standardized quantification of adult neurogenesis. *Nat. Commun.* 11, 4275.

Further reading

- Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L., Auerbach, E.J., Behrens, T.E., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., Robinson, E.C., Sotiropoulos, S.N., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D.C., 2016. The neuroimaging methodology of the human connectome project. *Nat. Neurosci.* 19, 1175–1187.