

MEASUREMENT VARIABILITY FOLLOWING MRI SYSTEM UPGRADE

Olivier Potvin, PhD¹, April Khademi, PhD², Isabelle Chouinard¹, Farnaz Farokhian, PhD¹, Louis Dieumegarde, BA¹, Ilana Leppert, PhD^{3,4}, Rick Hoge, PhD^{3,4}, Natasha Rajah, PhD^{3,7}, Pierre Bellec, PhD^{5,6}, Simon Duchesne, PhD^{1,8}, and the CIMA-Q* and the CCNA groups

Affiliations

¹ Centre de recherche CERVO Research Center, Institut universitaire de santé mentale de Québec, Québec, Canada

² Image Analysis in Medicine Lab, Ryerson University, Toronto, Canada

³ McGill University, Montréal, Québec

⁴ Montreal Neurological Institute, Montréal, Québec

⁵ Institut universitaire en gériatrie de Montréal, Montréal, Québec

⁶ Université de Montréal, Montréal, Québec

⁷ Douglas Mental Health University Institute, Montréal, Québec

⁸ Université Laval, Québec, Canada

* Part of the data used in this article were obtained from the Consortium pour l'identification précoce de la maladie Alzheimer - Québec (CIMA-Q; cima-q.ca). As such, the investigators within the CIMA-Q contributed to the design, the implementation, the acquisition of clinical, cognitive, and neuroimaging data and biological samples. A list of the CIMA-Q investigators is available on cima-q.ca.

Corresponding Author

Simon Duchesne, PhD
Centre de recherche CERVO Research Center
Institut universitaire en santé mentale de Québec
2601 rue de la Canardière, F-3582
Québec, QC G1J 2G3

Telephone: 418 663-5741, ext. 4777

E-mail : simon.duchesne@fmed.ulaval.ca

Acknowledgments

Financial support for I.C. and O.P. was obtained from the Alzheimer's Society of Canada (#13-32), the Canadian Institute for Health Research (#117121), and the Fonds de recherche du Québec – Santé / Pfizer Canada - Pfizer-FRQS Innovation Fund (#25262). S.D. is a Research Scholar from the Fonds de recherche du Québec – Santé (#30801).

The Consortium pour l'identification précoce de la maladie d'Alzheimer – Québec (CIMA-Q) is financed through the Fonds de recherche du Québec – Santé / Pfizer Canada - Pfizer-FRQS Innovation Fund (#27239). The CCNA is financed through the Canadian Institutes for Health Research (2014-2019) with funding from several partners.

Abstract

Major hardware/software changes to MRI platforms, either planned or unplanned, will almost invariably occur in longitudinal studies. Our objective was to assess the resulting variability on relevant imaging measurements in such context, specifically for three Siemens Healthcare Magnetom Trio upgrades to the Prisma^{fit} platform.

We report data acquired on three healthy volunteers scanned before and after three different platform upgrades. We assessed differences in image signal (contrast-to-noise ratio (CNR)) on T1-weighted images (T1w) and fluid-attenuated inversion recovery images (FLAIR); brain morphometry on T1w image; and small vessel disease (white matter hyperintensities; WMH) on FLAIR image.

Prisma^{fit} upgrade resulted in higher (30%) and more variable neocortical CNR and higher brain volume and thickness mainly in frontal areas. A significant relationship was observed between neocortical CNR and cortical volume. For FLAIR images, no significant CNR difference was observed, but WMH volumes were significantly smaller (-68%) after Prisma^{fit} upgrade, when compared to results on the Magnetom Trio.

Together, these results indicate that Prisma^{fit} upgrade significantly influenced image signal, brain morphometry measures and small vessel diseases measures and that these effects need to be taken into account when analyzing results from any longitudinal study undergoing similar changes.

Keywords

Neuroimaging, magnetic resonance imaging, MRI upgrade, variability, longitudinal studies, morphometry, Siemens Healthcare

INTRODUCTION

Magnetic resonance imaging (MRI) is by now of routine use in neuroscience studies of the living human brain. A multiplicity of contrast mechanisms has been devised to provide anatomical, cerebrovascular, functional, pathological, and metabolite information, in neurological and psychiatric diseases alike. Frequently, in order to increase participation and achieve sample sizes of statistical significance, investigators rely on acquisitions performed at multiple centers – and hence, using multiple imaging systems. Unfortunately, given that MRI signals are not recorded in absolute values, different platforms will produce different intensities for a given contrast, based on the physics of acquisition. These intensity differences will lead to between-system contrast differences, which in turn will impact measurements, for example morphometric estimates [1; 2; 3; 4; 5; 6; 7]. Harmonized protocols, geometric phantom correction and human volunteer calibration are quality control techniques which can be used to reduce these differences, and hence their impact when comparing participant populations between centers.

The situation is otherwise complicated when changes occur within a given center, for example when technical and/or managerial pressure requires preventive (scheduled) or corrective software and especially hardware updates. On the one hand, cross-sectional designs would be mildly affected; one could consider participants scanned pre- and post-upgrades as though they had been seen at two different sites. On the other hand, longitudinal studies risk being affected; disentangling the effect due to the upgrade from that of the phenomena under study becomes intractable.

Very few studies have assessed MRI scanner upgrade effects using short scan intervals. On T1-weighted images (T1w), increased signal-to noise (SNR) and contrast-to-noise (CNR) ratio have

been reported on Siemens Symphony software upgrades [8] and altered brain morphometry have been observed following GE Signa software upgrades using a one-year scan interval [9]. Moreover, using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, other authors [10] reported a slight increase in total brain volume following upgraded 1.5T scanners (0.33% GE Signa Excite to Signa HDx and 0.39% Siemens Symphony to Symphony TIM). However, other studies did not observe any significant difference in neocortical thickness and subcortical regions' volume following Siemens Sonata-Avanto and Trio-Tim Trio software/hardware upgrades, within a 6-week scan interval [4; 7].

This situation is common for studies in the context of neurodevelopment or neurodegeneration, where follow-ups tend to last through the expected mid-life update (~3-4 years) if not expected lifetime of a high-caliber research MRI platform (~6-8 years). It was the case for our involvement in both the longitudinal *Quebec Consortium for Early identification of Alzheimer's disease* (the Consortium d'identification précoce de la maladie d'Alzheimer – Québec; CIMA-Q; www.cima-q.ca) and the *Canadian consortium for neurodegeneration and aging* (CCNA; www.ccna-ccnv.ca). Common to both studies, three Siemens Trio systems have undergone major hardware upgrades (Prisma^{fit}) within the span of one year, while recruitment and follow-up were undergoing.

Faced with this inevitability, our strategy was to measure the variability induced by the upgrade by measuring changes in pre/post scans of human volunteers for measurements of interest to both studies. We argue that this variability can serve as a threshold against which to compare any future change being detected in the course of longitudinal studies that have included these sites. We therefore report in the following chapters our quantifying of differences in image signal, anatomical information (brain morphometry), and small vessel disease as relevant

measurements that exemplify the degree of variability following a major upgrade. To our knowledge, this study is the first to assess the impact of Prisma^{fit} upgrade on image signal and brain morphometry. Since the Siemens Trio scanner is a model widely used in neuroimaging research (e.g. see [1] for the scanner characteristics of 23 openly accessible datasets), this upgrade will be likely prevalent in future neuroimaging studies, hence the importance to measure its impact on image signal and brain morphometry.

MATERIALS AND METHODS

Participants and image acquisition

Three healthy volunteers (all males; age range 43-47 years old) participated in the study, with images acquired at three different sites, each undergoing a complete system overhaul. Specifically, these were: the McConnell Brain Imaging Center (McGill University, Montreal, Canada), the Douglas Mental Health University Institute's Brain Imaging Center (McGill University, Montreal, Canada) and the Unité de Neuroimagerie Fonctionnelle (Université de Montréal, Montreal, Canada). Each site planned and executed a scheduled upgrade of their Magnetom Trio to Prisma^{fit} platforms (Siemens Medical Systems, Erlangen, Germany). These major upgrades involved a complete retrofit of the signal transmission and reception chains, reconstruction hardware, and coils.

To test pre-post upgrade changes, the three volunteers were scanned three times before and twice after at McConnell Brain Imaging Center (BIC). One of these participants was scanned five times before and five times after at the Unité de Neuroimagerie Fonctionnelle (UNF); and twice before and once after at the Institut en santé mentale de l'hôpital Douglas (ISMD). Altogether,

there were 16 scans acquired before and 12 scans after the upgrade. The scan interval ranged between 86 to 150 days (mean: 133.8, std: 26.8).

Part of the data used in this article were obtained from the Consortium pour l'identification précoce de la maladie Alzheimer - Québec (CIMA-Q), founded in 2013 with a \$2,500,000 grant from the *Fonds d'Innovation Pfizer - Fond de Recherche Québec – Santé sur la maladie d'Alzheimer et les maladies apparentées*. The main objective was to build a cohort of participants characterized in terms of cognition, neuroimaging and clinical outcomes in order to acquire biological samples allowing 1) to establish early diagnoses of Alzheimer's disease, 2) to provide a well characterized cohort and 3) to identify new therapeutic targets. The principal investigator and director of CIMA-Q is Dr Sylvie Belleville from the Centre de recherche de l'Institut universitaire de gériatrie de Montréal, CIUSSS Centre-sud-de-l'Île-de-Montréal. CIMA-Q represent a common effort of several researchers from Québec affiliated to Université Laval, Université McGill, Université de Montréal, et Université de Sherbrooke. CIMA-Q recruited 290 cognitively healthy participants, with subjective cognitive impairment, mild cognitive impairment, or Alzheimer's disease, between 2013-2016.

Image acquisition protocol

All acquisitions were performed following the *Canadian Dementia Imaging Protocol* [11] (www.cdip-pcid.ca), and consisted in (a) a sagittal 3D isotropic T1-weighted (T1w) scan with 1.0 X 1.0 X 1.0mm³ resolution, 256 X 256 matrix, 192 slices, field of view (FOV) of 256 X 256 mm, repetition time (TR) of 2300 msec, echo time (TE) of 2.98 msec, no inversion time (TI), flip angle of 125, and acceleration factor of 2 (Siemens: MP- RAGE-PAT); and (b) an axial fluid attenuated inversion recovery (FLAIR) with resolution 0.9 X 0.9 X 3 mm³, fat saturation, 256 X 256 matrix, 48

slices, FOV of 240 X 240 mm, TE of 123 msec, TI of 2500 msec, flip angle of 165, and an acceleration factor of 2. Two sites (BIC and UNF) used the *32-Channel Head Coil* for the pre- and post-upgrade acquisitions, while the ISMD used the *Head Matrix Coil* (12 channels) for the pre-upgrade acquisition and the *Head/Neck 20* coil (20 channels) for the post-upgrade acquisition. This report focuses on the impact of Prismafit upgrade on measurements of T1w and FLAIR image, but the CDIP also included other sequences (diffusion-tensor (DWI), and resting state functional T2*-weighted blood-oxygen-level-dependent sensitive sequence (rsfMRI) images). A separated study is planned for DWI and rsfMRI.

Image processing

T1w images were processed using *FreeSurfer* 5.3 (<http://freesurfer.net>) with default pipeline (recon-all -all) without any flag option. The technical details of these procedures are described in prior publications [12; 13].

Signal analysis measurements

Contrast-to-noise ratio (CNR) was assessed using intensities values with the following formula, using voxel intensities from T1w images with the mean and variance (squared standard deviation) of gray matter (GM) and cerebral white matter (WM):

$$CNR = \frac{(GM\ mean - WM\ mean)^2}{(GM\ variance + WM\ variance)}$$

where subcortical and neocortical GM and WM volumes were obtained from the *aparc* and *aseg* labels generated by *FreeSurfer*. We computed the CNR before (*orig.mgz*) and after (*nu.mgz*) inhomogeneity correction. In order to compute CNR for FLAIR images, the same equation as T1w was used. However, since there is little contrast between WM and GM classes in FLAIR, the first tissue class is called the brain class (which is GM + WM as a whole), and the second tissue class

is the CSF. In order to compute the mean and standard deviations for the brain and CSF tissues, segmentations of these tissues were required, which was generated using a standardization and segmentation framework for FLAIR MRI [14; 15; 16]. To ensure that the tissue classes contained pure tissues only, 50% of the middle slices were retained for CNR calculation, ensuring that if the brain extraction algorithm missed any skull at the top or bottom, it would not bias the approach.

Brain morphometry measurements

From *FreeSurfer*, the default subcortical (10 regions per hemisphere) [17] and Desikan-Killiany-Tourville neocortical atlases (31 neocortical regions per hemisphere) volumes and thicknesses were used [18], resulting in 144 regional morphometric measures. Total white matter was defined as the difference between the total brain volume without ventricles and the total gray matter.

White matter hyperintensities measurement

WMH on FLAIR were assessed using Schmidt et al.'s automated LST toolbox [19], including both T1w and FLAIR images as input.

Statistical analyses

For analyses, variables were transformed into comparable scales between participants. For each participant, all values were converted into percentage of its mean. To verify the effect of Prisma^{fit} upgrade on CNR and morphometric measures, we fitted a linear-mixed model for each measure with Prisma^{fit} upgrade as between-factor and subject as a repeated factor with random intercepts for each subject. This model allows to partition out variability due to individual differences and test whether the variance due to the scanner upgrade was higher than error variance which encompasses unknown factors' influence such as noise. Furthermore, we used the Levene's test,

which assesses the homogeneity of variances[20], to test differences in variance before and after upgrade. For the regional morphometric measurements (144 measures), we used the corrected with false discovery rate (FDR) p value correction to adjust for multiple comparisons [21].

Moreover, in order to verify whether signal changes influenced morphometric results, we conducted Pearson's correlation were conducted between CNR and cortical/subcortical volumes.

Furthermore, to assess the pre/post upgrade reliability, we computed intraclass correlations (ICC) two-way random effects with multiple measurements[22] with pre and post measurements each treated as a rater. To assess intra-scanner morphometric reliability, the same ICC was computed, but for pre and post images separately with each measurement treated as a rater.

Using 28 images (16 scans acquired before and 12 scans acquired after the upgrade), we calculated that with a two-tailed alpha of 5% and a power of 80%, medium effect sizes (0.55) could be detected [23].

All statistical analyses were conducted in Python using SciPy [24] and StatsModels [25] modules, except ICC which were computed through the ICC function of the psych R package [26] and interpretation were made based on Cicchetti's guidelines [27].

RESULTS

Signal changes following upgrade

Figures 1 shows the CNR before and after upgrade for neocortical and subcortical areas. The Prisma^{fit} upgraded platforms significantly increased neocortical (model estimate change: 30.0%, $p < .0001$; mean \pm sd: pre=86.8% \pm 4.3, post=117.6% \pm 15.8), but not subcortical CNR (1.4%, $p = .8962$; pre=97.8% \pm 33.8, post=103.0% \pm 40.8), compared to their Trio counterparts. In addition,

Prisma upgrade resulted in higher variability of neocortical, ($p=.0459$), but not subcortical ($p=.2557$) CNR compared to the Magnetom Trio.

After N3 inhomogeneity correction[28], a significant difference in CNR was observed after Prisma^{fit} upgrade for subcortical (-13.8%, $p=.0017$; pre=105.9±10.4, post=92.1 ±12.1), but not neocortical (0.3%, $p=.8300$; pre=99.7±3.1, post=100.4 ±3.9) areas. The upgrade did not significantly impact the variability of neocortical ($p=.4106$) or subcortical ($p=.9021$) CNR compared to the Magnetom Trio.

Figure 1 displays the CNR for FLAIR images. No significant differences were observed before and after Prisma^{fit} upgrade in terms of mean (-6.4%, $p=.4613$; pre=102.8% ±17.7, post=96.3 ±26.5) or variance ($p=.5409$) of FLAIR CNR.

ICCs revealed that CNR reliability between pre and post Prisma^{fit} upgrade ranged between poor and fair (T1w cortex: 0.04 ±95CI: -0.13-0.33, with corrected intensities: 0.52 ±0.34-0.83; T1w subcortical: 0.38 ±0.74-0.78; with corrected intensities: 0.00 ±0-0.66-0.53; FLAIR: 0.46 ±0.46-0.81)

Brain morphometry changes following upgrade

Figures 2 and 3 show the percentage of change in volume and thickness, respectively, of neocortical regions after the Prisma^{fit} upgrade. Thirty-two neocortical measures (16 volumes and 15 thicknesses) were significantly larger after the upgrade with ranges between 1.9% to 6.4% for volumes and 1.9% to 5.3% for thickness. In addition, two regions were significantly thinner (left entorhinal: -2% and left superior parietal: -2%). Figure 4 displays the percentage of change in volume of subcortical regions after the Prisma^{fit} upgrade. Nine regions had significantly larger

volumes (range: 1.9%-8.5%) while the right cerebellum cortex had significantly smaller volume after the upgrade.

In terms of variances, all morphometric measures (neocortical volumes, neocortical thicknesses, and subcortical volumes) did not significantly differ between pre and post Prisma^{fit} upgrade after FDR correction (Supplementary Table 1).

Figures 5, 6, and 7 show ICCs for neocortical volume, neocortical thickness and subcortical volume, respectively. Neocortical regions' volume ICCs were generally excellent (mean: $0.85 \pm \text{sd}: 0.14$) while those for thickness were more variable ranging from poor to excellent (0.60 ± 0.30). As could be expected, brain regions showing the largest thickness differences after the upgrade (i.e., frontal areas) displayed the lowest ICCs. Subcortical regions' volume ICCs were mostly good (0.66 ± 0.31), except for the accumbens area, ventral diencephalon, left caudate and left cerebellum WM, which had poor values.

Correlation between CNR and morphometry

A significant positive correlation was observed between cortical CNR and cortical volume ($r: 0.623, p=.0004$). After inhomogeneity correction, this relationship was lower, but still significant ($r: 0.491, p=.0080$). A significant negative correlation was also observed between subcortical CNR and subcortical volume before ($r: -0.534, p=.0034$), but not after inhomogeneity correction (before $r: -0.332, p=.0840$).

WMH changes following upgrade

Figure 8 illustrates that WMH were significantly smaller after Prisma^{fit} upgrade compared to Magnetom Trio acquired FLAIR images (-68%, $p=.0011$; pre= 123.8 ± 48.6 , post= 68.3 ± 32.5), but

no significant difference in terms of variability were observed ($p=.2260$). ICC revealed that WMH reliability was fair between pre and post Prisma^{fit} upgrade (0.41).

Intra-scanner morphometric reliability

Supplementary Figures 1 and 2 show the pre and post upgrade ICCs for neocortical volume and thickness, and subcortical volumes, respectively. Reliability was excellent for nearly all regions both pre- (0.92 ± 0.19) and post- (0.90 ± 0.16) upgrade. Only four pre-upgrade measures (right pars triangularis thickness, right superior parietal thickness, left ventral diencephalon, and left isthmus cingulate thickness) and four post-upgrade measures (left isthmus cingulate volume, left paracentral thickness, right lingual thickness, right accumbens area, left medial orbitofrontal thickness) had poor ICCs. Furthermore, WMH reliability was excellent pre- (0.78) and post- (0.88) upgrade.

DISCUSSION

We aimed to assess the impact of major MRI system upgrades on relevant signal, brain morphometry, and small vessel disease measurements. Despite good reliability within scanner for the Magnetom Trio and Prisma^{fit} when assess separately, we observed that the upgrade generated notable changes on the order of 30% for neocortical CNR; larger morphometric measures up to 6.4% for neocortical regions' volume, 5.3% for neocortical regions' thickness, 8.5% for subcortical regions' volume; and smaller WMH volume of 68%. Such changes are not to be expected of healthy volunteers (all cognitively intact, and aged < 50 years old) within such short scan intervals (between three to five months). These changes appear consistent with previous results showing increase CNR[8] and larger brain volumes after scanner upgrade [10].

These changes are not negligible and must be taken into consideration when analyzing data; they should be thought of as “floor” values for effect sizes whenever interpreting results coming from systems that have been upgraded, in a longitudinal setting. To wit, these values must be compared to atrophy rates reported for aging (0.83%/year)[29] and Alzheimer’s disease (1.9%/year)[30], that are lower than some of the observed upgrade effects. However, in opposition to these multiple significant mean effects, we did not observe any significant heterogeneity of the variance between pre and post upgrade and therefore cross-sectional estimates using either platform are perfectly sound.

This report focuses on the impact of Prisma^{fit} upgrade on measurements of T1w and FLAIR image, but it is likely that similar variability will affect other sequences e.g. diffusion-tensor functional T2*-weighted blood-oxygen-level-dependent sensitive sequences. A separate study is planned for these acquisitions, as obtained using the CDIP protocol.

Investigators analyzing data from studies using these platforms will of course want to take particular note of the above results. While we acknowledge that Prisma^{fit} upgrade has notable effects, there is no easy solution to counter these effects in multicentric studies. A correction factor would be hazardous to build since the effects that we observed are highly non-linear, driven by MR physics and systems engineering, and not driven by the participants.

Our results were acquired at three different sites undergoing the same type of upgrade. It is likely that they are therefore representative of the variability for other sites undergoing a similar change, should the investigators have not captured pre/post-upgrade data. However, this is only one of the many systems in operation and hence, these results cannot be assumed to apply for any other configuration. What should be assumed however is that any upgrade will generate

variability, even if its magnitude is less than that reported here. It is therefore highly recommended that investigators collect pre/post upgrade data, preferably on a number of volunteers, and test measurements of interest. If this proves impossible, then comparison to other sites undergoing similar changes may pose as a substitute.

AUTHOR CONTRIBUTIONS STATEMENT

OP Guarantor of integrity of entire study. OP, SD Literature research; OP Statistical analysis. All authors study concepts/study design or data acquisition or data analysis/interpretation, manuscript drafting or manuscript revision for important intellectual content, approval of final version of submitted manuscript, and manuscript editing.

CONFLICT OF INTEREST STATEMENT

None.

REFERENCES

- [1] O. Potvin, L. Dieumegarde, and S. Duchesne, Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. *Neuroimage* 156 (2017) 315-339.
- [2] O. Potvin, L. Dieumegarde, and S. Duchesne, Freesurfer cortical normative data for adults using Desikan-Killiany-Tourville and ex vivo protocols. *Neuroimage* 156 (2017) 43-64.
- [3] K.A. Govindarajan, L. Freeman, C. Cai, M.H. Rahbar, and P.A. Narayana, Effect of intrinsic and extrinsic factors on global and regional cortical thickness. *PLoS One* 9 (2014) e96429.
- [4] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, and B. Fischl, Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32 (2006) 180-94.
- [5] F. Kruggel, J. Turner, and L.T. Muftuler, Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *Neuroimage* 49 (2010) 2123-33.
- [6] O. Potvin, A. Mouiha, L. Dieumegarde, and S. Duchesne, Normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage* 137 (2016) 9-20.
- [7] J. Jovicich, S. Czanner, X. Han, D. Salat, A. van der Kouwe, B. Quinn, J. Pacheco, M. Albert, R. Killiany, D. Blacker, P. Maguire, D. Rosas, N. Makris, R. Gollub, A. Dale, B.C. Dickerson, and B. Fischl, MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage* 46 (2009) 177-92.
- [8] B. Shuter, I.B. Yeh, S. Graham, C. Au, and S.C. Wang, Reproducibility of brain tissue volumes in longitudinal studies: effects of changes in signal-to-noise ratio and scanner software. *Neuroimage* 41 (2008) 371-9.
- [9] H. Takao, N. Hayashi, and K. Ohtomo, Effects of the use of multiple scanners and of scanner upgrade in longitudinal voxel-based morphometry studies. *J Magn Reson Imaging* 38 (2013) 1283-91.
- [10] H. Lee, K. Nakamura, S. Narayanan, R.A. Brown, D.L. Arnold, and I. Alzheimer's Disease Neuroimaging, Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. *Neuroimage* 184 (2019) 555-565.
- [11] S. Duchesne, I. Chouinard, O. Potvin, V.S. Fonov, A. Khademi, R. Bartha, P. Bellec, D.L. Collins, M. Descoteaux, R. Hoge, C.R. McCreary, J. Ramirez, C.J.M. Scott, E.E. Smith, S.C. Strother, S.E. Black, C.-Q. group, and C.g. the, The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts. *J Magn Reson Imaging* 49 (2019) 456-465.
- [12] A.M. Dale, B. Fischl, and M.I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9 (1999) 179-94.
- [13] B. Fischl, M.I. Sereno, and A.M. Dale, Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9 (1999) 195-207.
- [14] A. Khademi, A. Venetsanopoulos, and A.R. Moody, Robust white matter lesion segmentation in FLAIR MRI. *IEEE Trans Biomed Eng* 59 (2012) 860-71.
- [15] A. Khademi, A. Venetsanopoulos, and A.R. Moody, Generalized method for partial volume estimation and tissue segmentation in cerebral magnetic resonance images. *J Med Imaging (Bellingham)* 1 (2014) 014002.
- [16] B. Reiche, A.R. Moody, and A. Khademi, Effect of image standardization on FLAIR MRI for brain extraction. *Signal, Image and Video Processing (SIVP)* 9 (2015) 11-16.
- [17] B. Fischl, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A.M. Dale, Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (2002) 341-55.

- [18] A. Klein, and J. Tourville, 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci* 6 (2012) 171.
- [19] P. Schmidt, C. Gaser, M. Arsic, D. Buck, A. Forschler, A. Berthele, M. Hoshi, R. Ilg, V.J. Schmid, C. Zimmer, B. Hemmer, and M. Muhlau, An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage* 59 (2012) 3774-83.
- [20] H. Levene, Robust tests for equality of variances, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford University, 1960, pp. 278-292.
- [21] Y. Benjamini, and Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* 57 (1995) 289-300.
- [22] K.O. McGraw, and S.P. Wong, Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods* 1 (1996) 30-46.
- [23] J. Cohen, A power primer. *Psychol Bull* 112 (1992) 155-9.
- [24] E. Jones, T. Oliphant, and P. Peterson, *SciPy: Open Source Scientific Tools for Python*. (2001).
- [25] S. Seabold, and J. Perktold, *Statsmodels: Econometric and statistical modeling with python*. *Proceedings of the 9th Python in Science Conference* (2010).
- [26] W. Revelle, *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA, 2018.
- [27] D.V. Cicchetti, Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 6 (1994) 284-290.
- [28] J.G. Sled, A.P. Zijdenbos, and A.C. Evans, A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Transactions on Medical Imaging* 17 (1998) 87-97.
- [29] F. Crivello, N. Tzourio-Mazoyer, C. Tzourio, and B. Mazoyer, Longitudinal assessment of global and regional rate of grey matter atrophy in 1,172 healthy older adults: modulation by sex and age. *PLoS One* 9 (2014) e114478.
- [30] J.D. Sluimer, H. Vrenken, M.A. Blankenstein, N.C. Fox, P. Scheltens, F. Barkhof, and W.M. van der Flier, Whole-brain atrophy rate in Alzheimer disease: identifying fast progressors. *Neurology* 70 (2008) 1836-41.

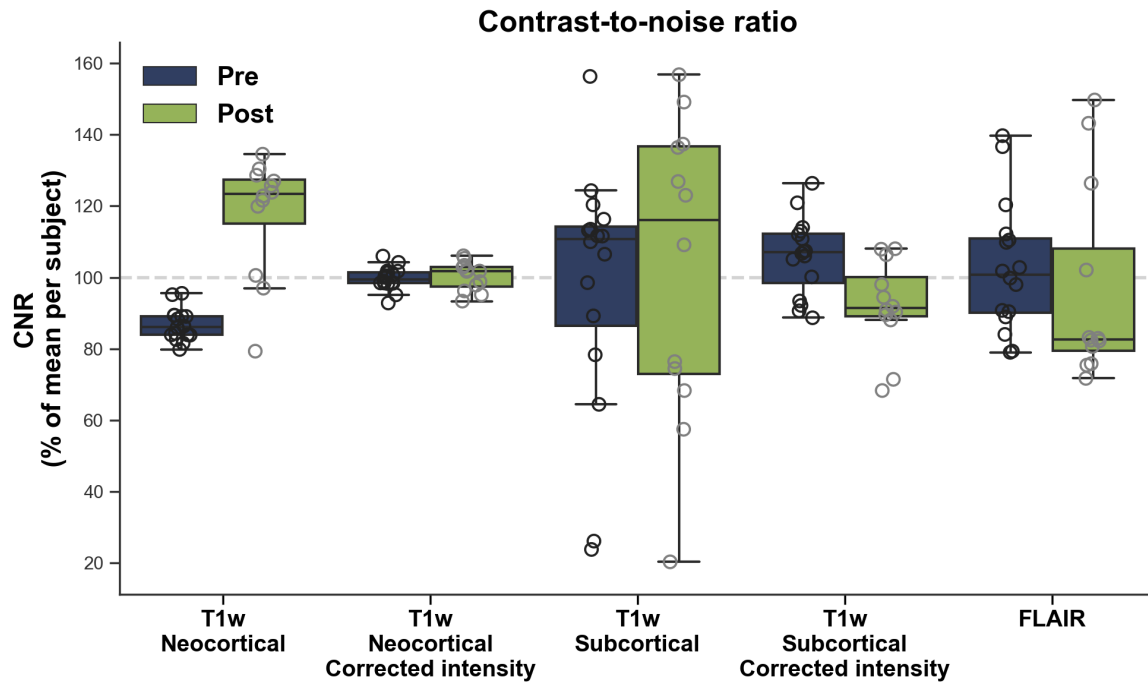


Figure 1. Contrast-to-noise ratio (CNR) before (Pre) and after (Post) Prisma^{fit} upgrade from T1w images and FLAIR images.

Prisma^{fit} upgrade effect on volume

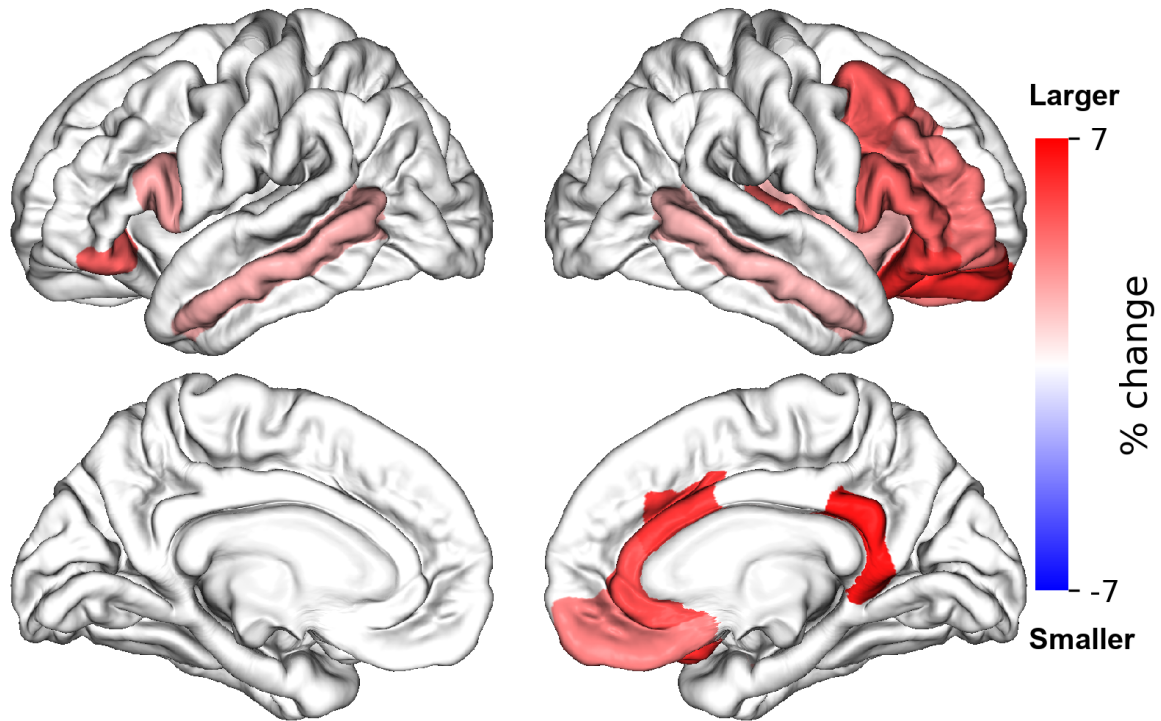


Figure 2. Percentage of change in neocortical regions' volume after Prisma^{fit} upgrade. Only regions that were significant after false-discovery rate correction ($p < .05$) are showed.

Prisma^{fit} upgrade effect on thickness

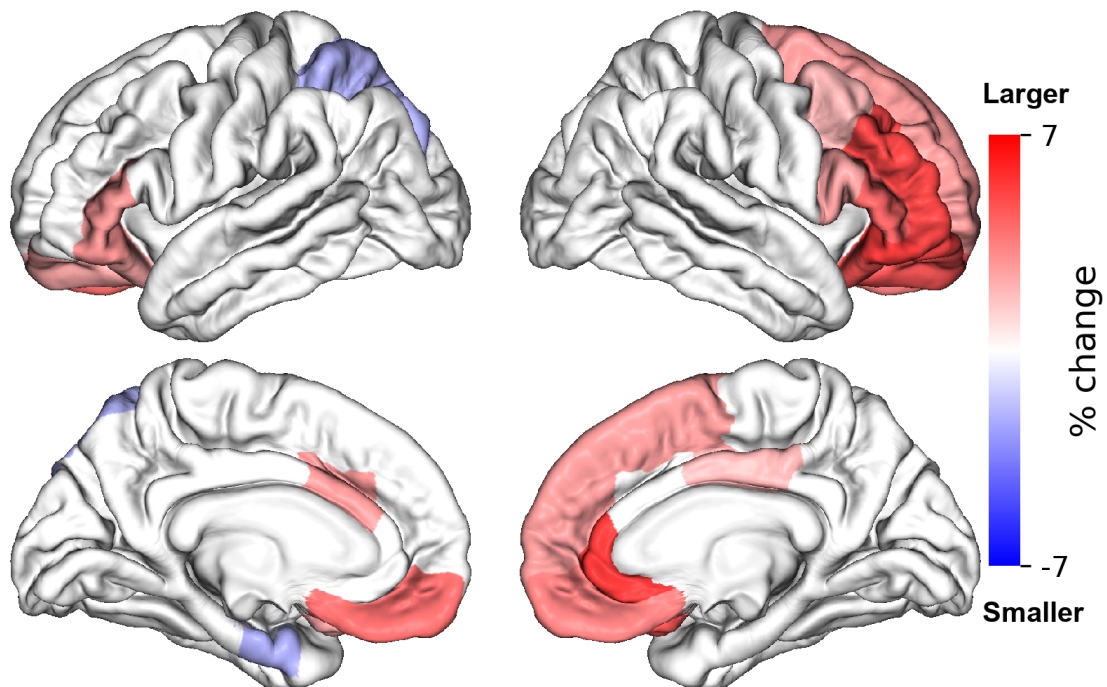


Figure 3. Percentage of change in neocortical regions' thickness after Prisma^{fit} upgrade. Only regions that were significant after false-discovery rate correction ($p < .05$) are showed.

Prisma^{fit} upgrade effect on subcortical volumes

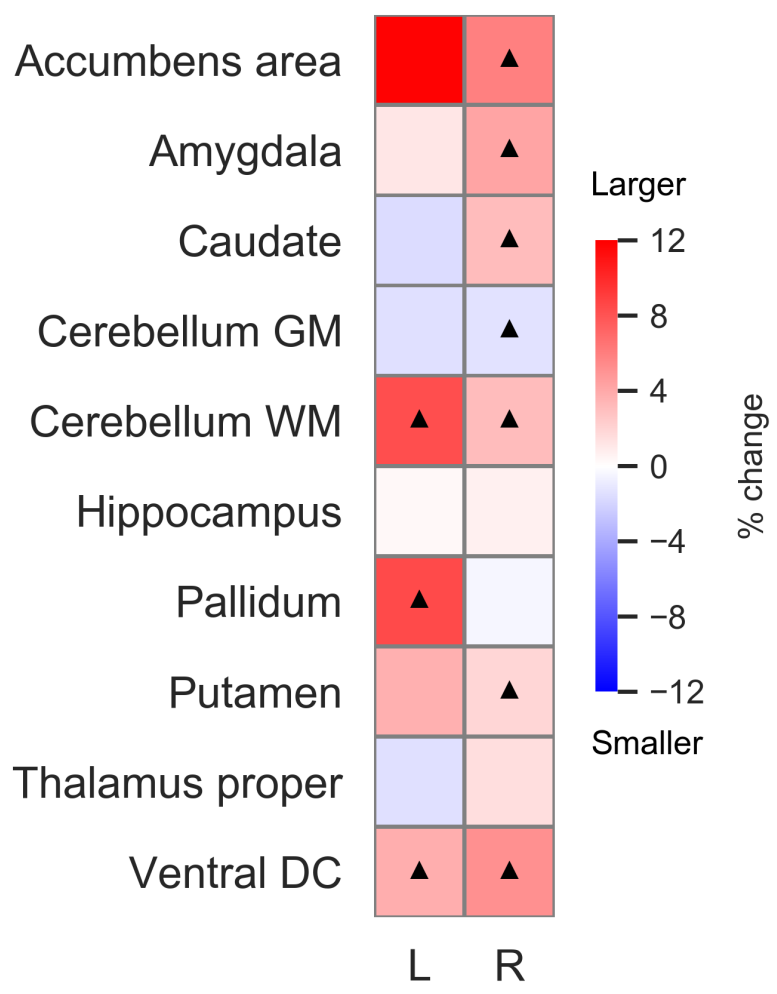


Figure 4. Percentage of change in subcortical regions' volume after Prisma^{fit} upgrade. Black triangles denote regions that were significant after false-discovery rate correction ($p < .05$).

Prisma^{fit} upgrade ICC on volume

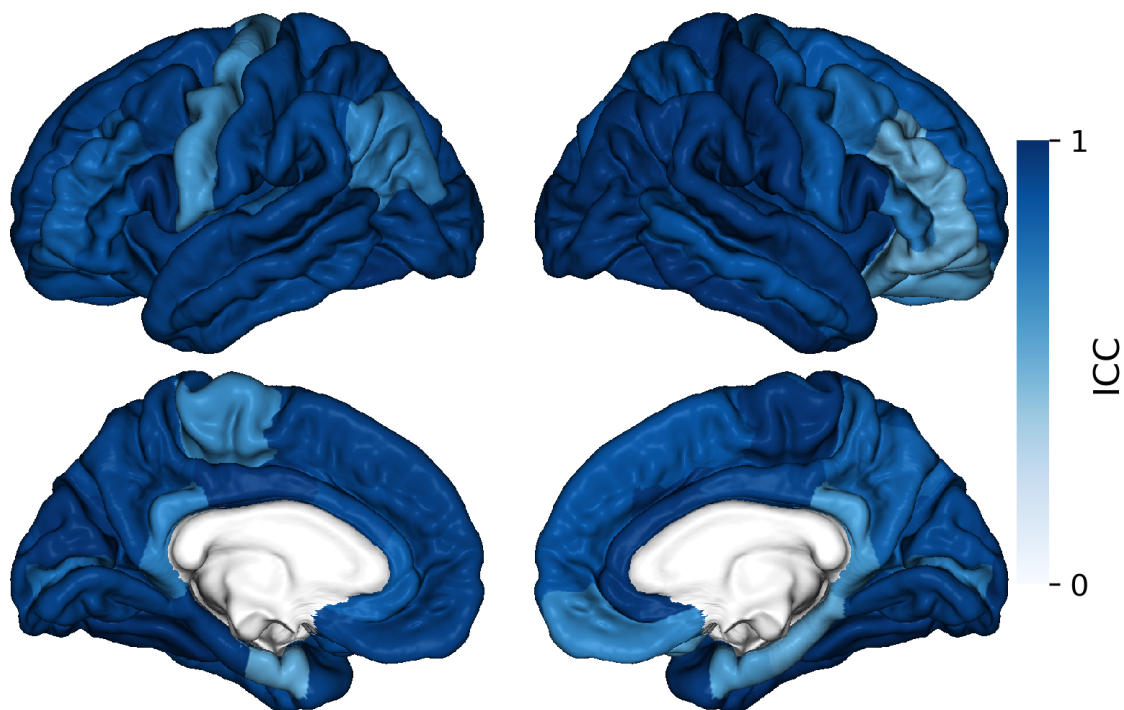


Figure 5. Intraclass correlations (ICC) of neocortical regions' volume before and after Prisma^{fit} upgrade.

Prisma^{fit} upgrade ICC on thickness

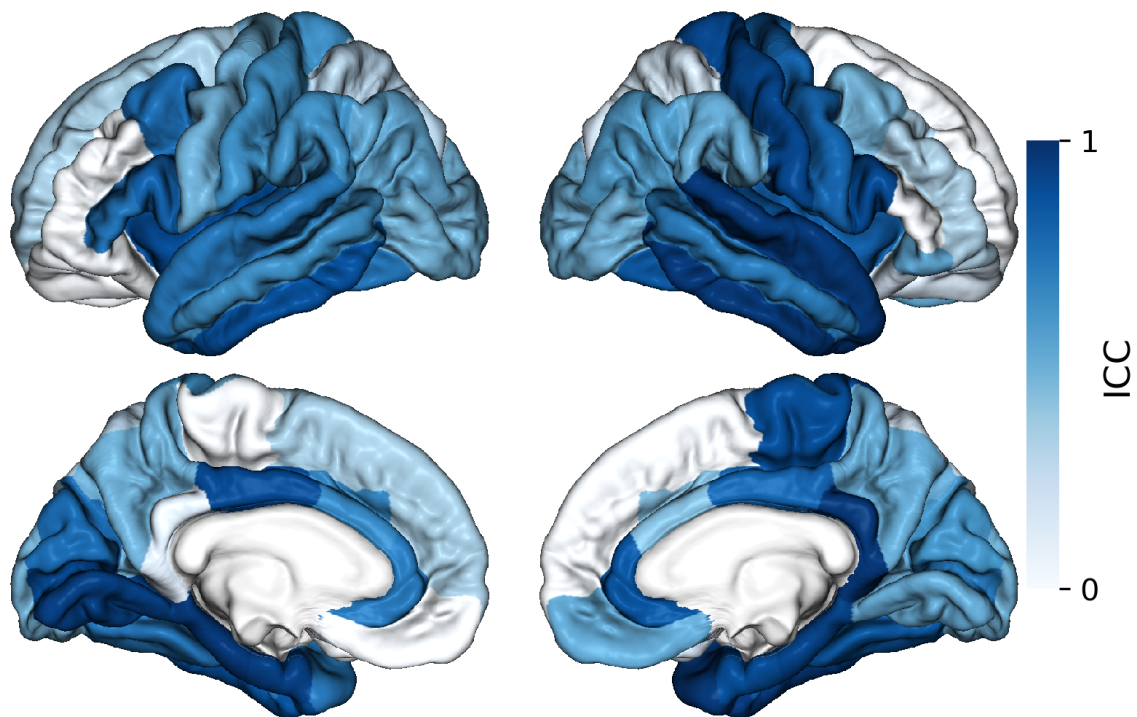


Figure 6. Intraclass correlations (ICC) of neocortical regions' thickness before and after Prisma^{fit} upgrade.

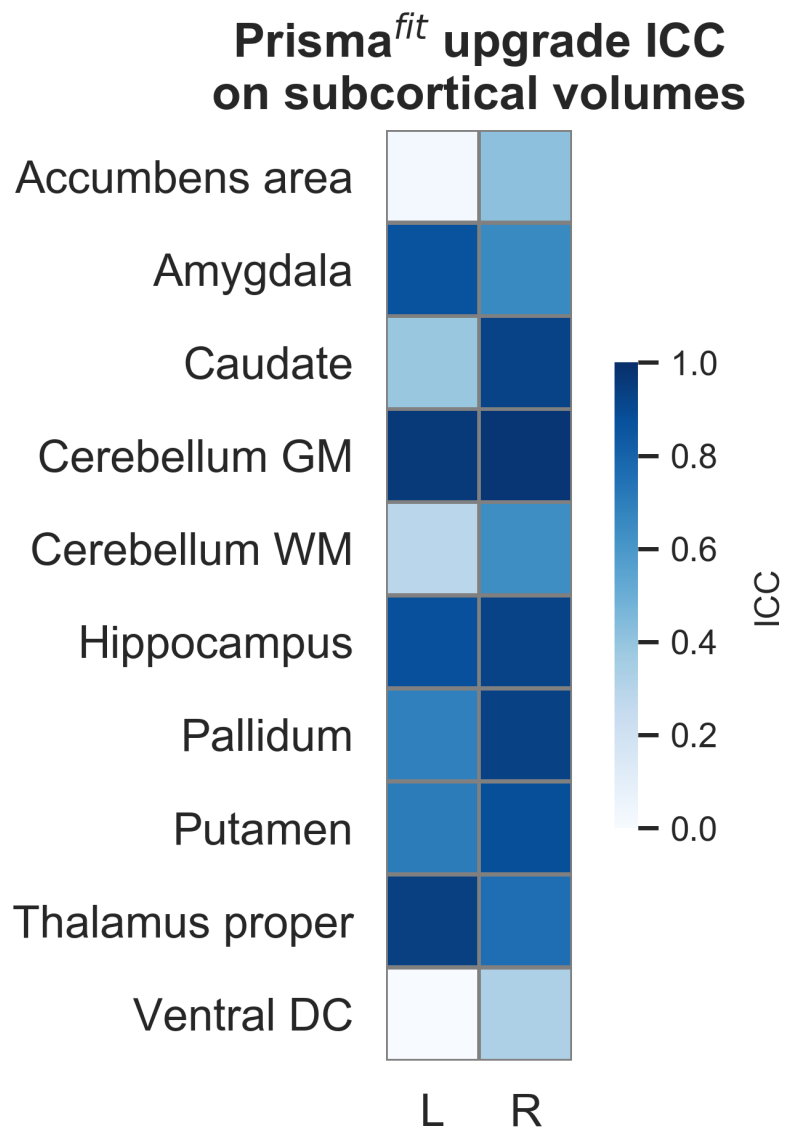


Figure 7. Intra-class correlations (ICC) of subcortical regions' volume before and after Prisma^{fit} upgrade.

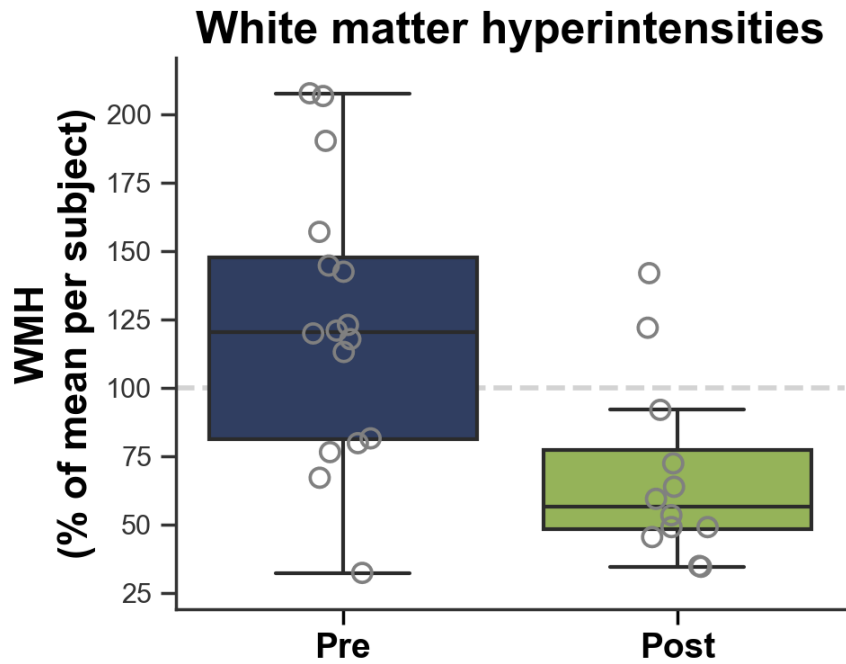
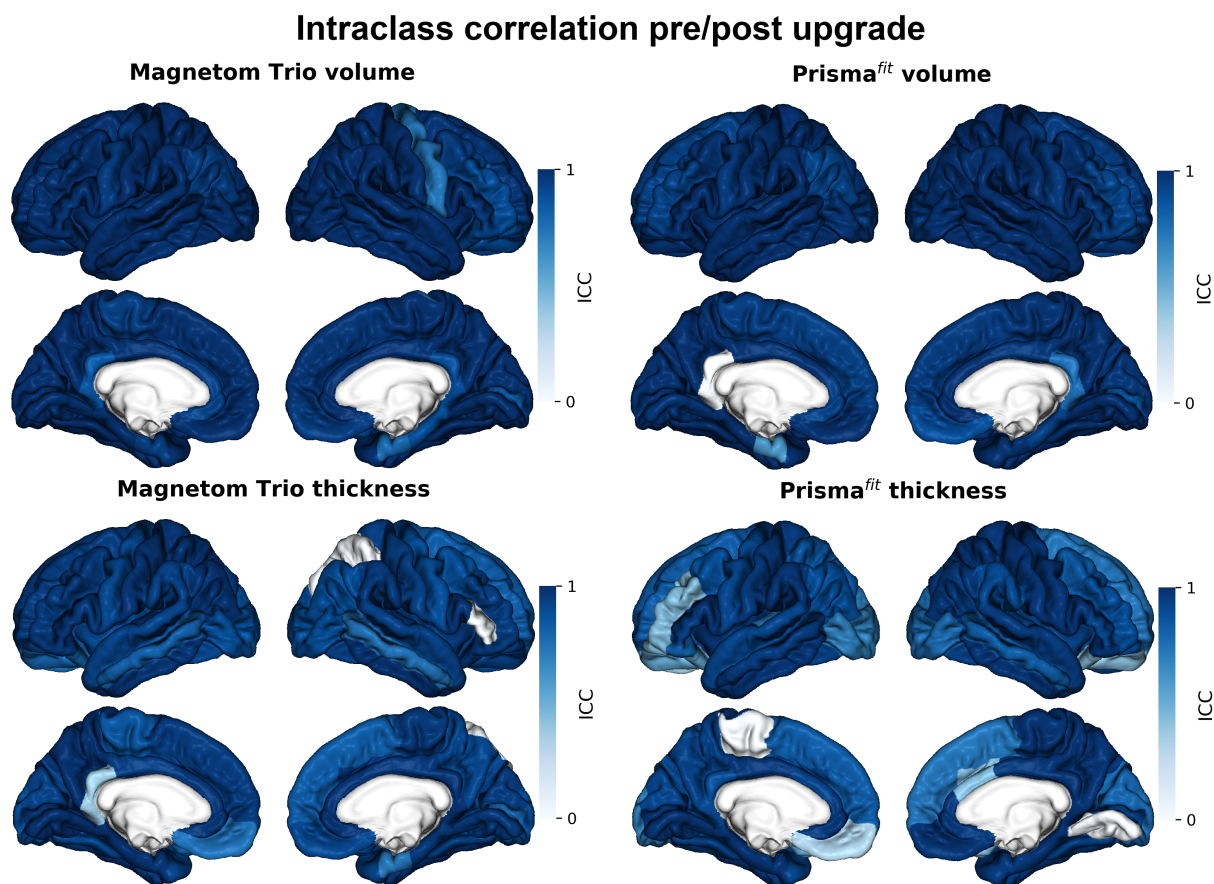
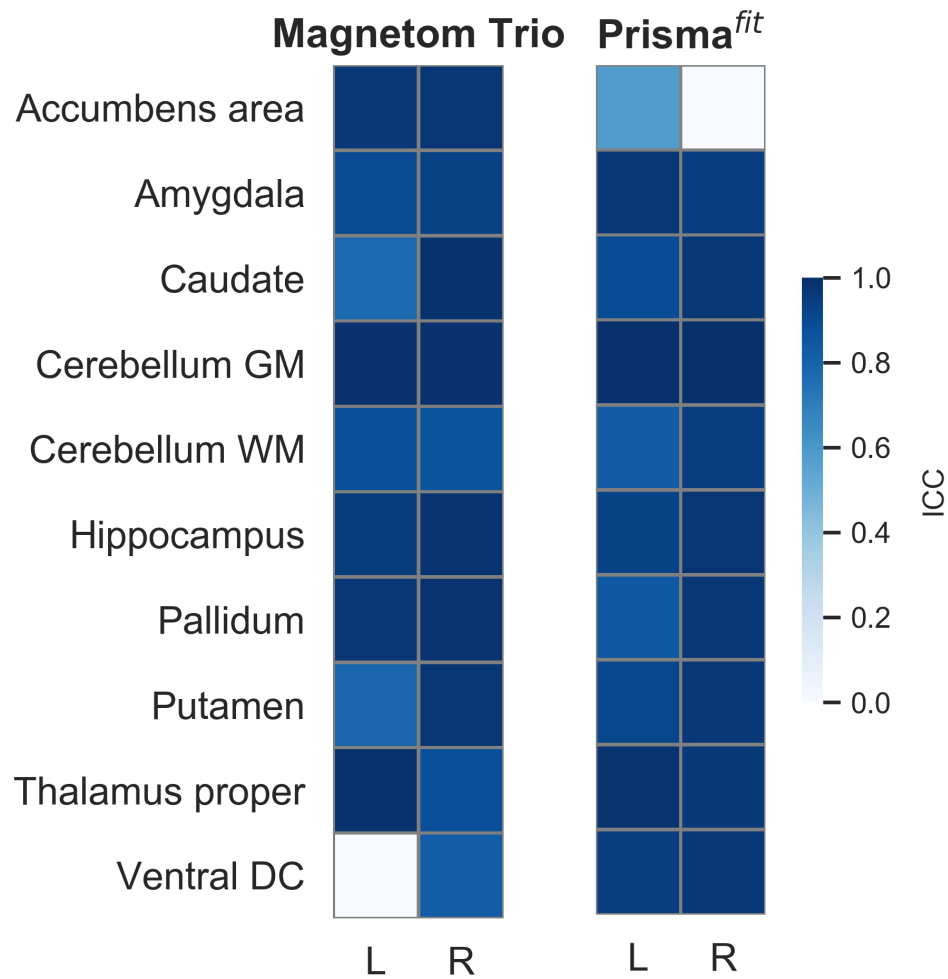


Figure 8. White matter hyperintensities (WMH) measured on FLAIR images according to before (Pre) and after (Post) Prisma^{fit} upgrade.



Supplementary Table 1. Pre and post Prisma^{fit} upgrade intraclass correlation for neocortical volume and thickness.



Supplementary Table 2. Pre and post Prisma^{fit} upgrade intraclass correlation for subcortical volumes.

Supplementary Table 1. Levene's test of homogeneity of variance on morphometric measures between pre and post Prisma^{fit} upgrade.

Measure	FDR corrected p value	Levene	uncorrected p value
Pericalcarine rv	0.9998	7.2344	0.0123
Superiorparietal rt	0.9998	5.1629	0.0316
Inferiorparietal rt	0.9998	5.1584	0.0316
Pericalcarine rt	0.9998	4.0850	0.0537
Superiorparietal lt	0.9998	3.6909	0.0657
Inferiorparietal lt	0.9998	3.5427	0.0710
Parsorbitalis rv	0.9998	3.4301	0.0754
Left pallidum	0.9998	3.2444	0.0833
Supramarginal rt	0.9998	3.1243	0.0889
Insula lt	0.9998	3.0381	0.0931
Lateraloccipital lt	0.9998	2.5460	0.1227
Parsorbitalis rt	0.9998	2.2331	0.1471
Rostralmiddlefrontal lt	0.9998	2.0440	0.1647
Precentral lv	0.9998	1.8610	0.1842
Left accumbens area	0.9998	1.8275	0.1881
Right accumbens area	0.9998	1.7682	0.1952
Insula rt	0.9998	1.5463	0.2248
Lingual rt	0.9998	1.4980	0.2320
Cuneus rt	0.9998	1.4330	0.2421
Medialorbitofrontal rt	0.9998	1.3911	0.2489
Right amygdala	0.9998	1.2010	0.2832
Rostralmiddlefrontal rt	0.9998	1.1846	0.2864
Lateralorbitofrontal lt	0.9998	1.1522	0.2929
Fusiform lt	0.9998	1.0152	0.3229
Pericalcarine lv	0.9998	0.9509	0.3385
Superiorfrontal rt	0.9998	0.8977	0.3521
Precentral lt	0.9998	0.8444	0.3666
Left cerebellum white matter	0.9998	0.8356	0.3691
Postcentral lt	0.9998	0.8294	0.3708
Superiorfrontal lt	0.9998	0.8192	0.3737
Rostralanteriorcingulate lt	0.9998	0.7656	0.3896
Fusiform rt	0.9998	0.7414	0.3971
Left caudate	0.9998	0.7063	0.4083
Inferiorparietal lv	0.9998	0.6886	0.4142
Right putamen	0.9998	0.6543	0.4259
Rostralmiddlefrontal lv	0.9998	0.6432	0.4298
Rostralmiddlefrontal rv	0.9998	0.6412	0.4305
Transverse temporal lv	0.9998	0.6224	0.4373
Caudalanteriorcingulate lt	0.9998	0.5686	0.4576

Supramarginal lt	0.9998	0.5573	0.4620
Caudalmiddlefrontal rt	0.9998	0.4976	0.4868
Parahippocampal rv	0.9998	0.4663	0.5007
Paracentral rt	0.9998	0.4576	0.5047
Rostralanteriorcingulate rt	0.9998	0.4046	0.5303
Pericalcarine lt	0.9998	0.3802	0.5429
Lateraloccipital rt	0.9998	0.3660	0.5505
Precentral rv	0.9998	0.3627	0.5522
Isthmuscingulate lv	0.9998	0.3625	0.5523
Right hippocampus	0.9998	0.3499	0.5593
Left thalamus proper	0.9998	0.3478	0.5604
Lingual rv	0.9998	0.3387	0.5656
Left ventraldc	0.9998	0.3330	0.5689
Paracentral rv	0.9998	0.3289	0.5712
Precuneus rv	0.9998	0.2967	0.5906
Isthmuscingulate rv	0.9998	0.2908	0.5943
Cuneus lv	0.9998	0.2741	0.6050
Entorhinal rv	0.9998	0.2678	0.6092
Parsopercularis rt	0.9998	0.2646	0.6113
Parsopercularis lt	0.9998	0.2436	0.6258
Parahippocampal rt	0.9998	0.2338	0.6328
Precentral rt	0.9998	0.2314	0.6345
Cuneus lt	0.9998	0.2212	0.6421
Transversetemporal rv	0.9998	0.2149	0.6468
Parstriangularis lv	0.9998	0.2133	0.6480
Superiortemporal rv	0.9998	0.2087	0.6516
Inferiortemporal lt	0.9998	0.2024	0.6565
Lateralorbitofrontal lv	0.9998	0.1972	0.6607
Middletemporal lt	0.9998	0.1960	0.6616
Right cerebellum white matter	0.9998	0.1956	0.6619
Posteriorcingulate lv	0.9998	0.1948	0.6626
Transversetemporal rt	0.9998	0.1843	0.6712
Medialorbitofrontal lt	0.9998	0.1804	0.6745
Cuneus rv	0.9998	0.1665	0.6866
Caudalmiddlefrontal rv	0.9998	0.1574	0.6948
Rostralanteriorcingulate rv	0.9998	0.1533	0.6986
Lateraloccipital lv	0.9998	0.1493	0.7024
Lingual lt	0.9998	0.1461	0.7054
Inferiorparietal rv	0.9998	0.1327	0.7186
Left hippocampus	0.9998	0.1318	0.7195
Superiortemporal lv	0.9998	0.1262	0.7253
Caudalmiddlefrontal lv	0.9998	0.1205	0.7313
Rostralanteriorcingulate lv	0.9998	0.1186	0.7333
Postcentral rt	0.9998	0.1169	0.7352

Parsopercularis rv	0.9998	0.1162	0.7359
Transversetemporal lt	0.9998	0.1101	0.7427
Right caudate	0.9998	0.1032	0.7506
Parsorbitalis lv	0.9998	0.0992	0.7553
Middletemporal lv	0.9998	0.0915	0.7646
Middletemporal rv	0.9998	0.0881	0.7690
Superiortemporal rt	0.9998	0.0875	0.7697
Posteriorcingulate lt	0.9998	0.0857	0.7721
Precuneus lt	0.9998	0.0852	0.7727
Parsorbitalis lt	0.9998	0.0845	0.7736
Postcentral rv	0.9998	0.0814	0.7776
Right ventraldc	0.9998	0.0808	0.7785
Caudalanteriorcingulate rv	0.9998	0.0784	0.7818
Fusiform rv	0.9998	0.0766	0.7842
Superiorparietal lv	0.9998	0.0710	0.7920
Lateralorbitofrontal rt	0.9998	0.0637	0.8028
Parstriangularis lt	0.9998	0.0628	0.8040
Insula rv	0.9998	0.0600	0.8083
Superiorparietal rv	0.9998	0.0567	0.8136
Lateralorbitofrontal rv	0.9998	0.0518	0.8217
Inferiortemporal rv	0.9998	0.0503	0.8244
Superiortemporal lt	0.9998	0.0491	0.8264
Parstriangularis rt	0.9998	0.0474	0.8294
Parahippocampal lt	0.9998	0.0459	0.8321
Supramarginal rv	0.9998	0.0433	0.8367
Paracentral lv	0.9998	0.0416	0.8400
Entorhinal lt	0.9998	0.0412	0.8408
Lateraloccipital rv	0.9998	0.0379	0.8471
Entorhinal rt	0.9998	0.0354	0.8523
Superiorfrontal rv	0.9998	0.0330	0.8573
Caudalmiddlefrontal lt	0.9998	0.0278	0.8689
Insula lv	0.9998	0.0207	0.8868
Left amygdala	0.9998	0.0180	0.8942
Supramarginal lv	0.9998	0.0174	0.8962
Entorhinal lv	0.9998	0.0138	0.9074
Left cerebellum cortex	0.9998	0.0117	0.9145
Superiorfrontal lv	0.9998	0.0112	0.9166
Left putamen	0.9998	0.0111	0.9169
Inferiortemporal rt	0.9998	0.0109	0.9178
Middletemporal rt	0.9998	0.0105	0.9193
Fusiform lv	0.9998	0.0095	0.9233
Right thalamus proper	0.9998	0.0066	0.9359
Postcentral lv	0.9998	0.0061	0.9383
Posteriorcingulate rv	0.9998	0.0047	0.9457

Medialorbitofrontal rv	0.9998	0.0041	0.9492
Isthmuscingulate lt	0.9998	0.0041	0.9493
Parstriangularis rv	0.9998	0.0040	0.9500
Parsopercularis lv	0.9998	0.0026	0.9596
Parahippocampal lv	0.9998	0.0025	0.9605
Isthmuscingulate rt	0.9998	0.0021	0.9637
Caudalanteriorcingulate rt	0.9998	0.0016	0.9683
Medialorbitofrontal lv	0.9998	0.0011	0.9738
Lingual lv	0.9998	0.0010	0.9746
Precuneus rt	0.9998	0.0010	0.9755
Paracentral lt	0.9998	0.0007	0.9794
Right cerebellum cortex	0.9998	0.0007	0.9795
Precuneus lv	0.9998	0.0007	0.9798
Posteriorcingulate rt	0.9998	0.0004	0.9849
Inferiortemporal lv	0.9998	0.0001	0.9924
Caudalanteriorcingulate lv	0.9998	0.0001	0.9929
Right pallidum	0.9998	0.0000	0.9998

FDR: false-discovery rate. l: left. r: right. v: volume. t: thickness.