Machine Learning-Based Reconstruction of 2D MRI for Quantitative Morphometry in Epilepsy

Corey Ratcliffe[1,2,*] Christophe de Bézenac[1] Kumar Das[3]
Shubhabrata Biswas[3] Anthony Marson[1] Simon S. Keller[1,3]

21.06.2024

[1] Department of Pharmacology and Therapeutics, Institute of Systems, Molecular, and Integrative Biology, University of Liverpool, Liverpool, UK
[2] Department of Neuro Imaging and Interventional Radiology, National Institute of Mental Health and Neuro Sciences, Bangalore, India
[3] The Walton Centre NHS Foundation Trust, Liverpool, UK

[*] Correspondence: Corey Ratcliffe <corey.ratcliffe@liverpool.ac.uk>

* Correspondence Email: <corey.ratcliffe@liverpool.ac.uk>

* Correspondence Address: Department of Pharmacology and Therapeutics, Institute of Systems, Molecular, and Integrative Biology, 3rd Floor, William Henry Duncan Building, University of Liverpool, L7 8TX

Christophe de Bézenac <cedb@liverpool.ac.uk>

Kumar Das <Kumar.Das@liverpool.ac.uk>

Shubhabrata Biswas <Shubhabrata.Biswas@nhs.net>

Anthony Marson <A.G.Marson@liverpool.ac.uk>

Simon S. Keller <Simon.Keller@liverpool.ac.uk>

Ethics and Integrity Nothing to declare.

Introduction: Structural neuroimaging analyses require 'research quality' images, acquired with costly MRI acquisitions. Isotropic (3D) T1 images are desirable for quantitative analyses, however a routine compromise in the clinical setting is to acquire anisotropic (2D) analogues for qualitative visual inspection. Machine learning-based software have shown promise in addressing some of the limitations of 2D scans in research applications, yet their efficacy in quantitative research is not well understood. To evaluate the applicability of image preprocessing methods, morphometry in idiopathic generalised epilepsy (IGE)—in which, pathology-related abnormalities of the subcortical structures have been reproducibly demonstrated—was investigated first in 3D scans, then in 2D scans, resampled images, and synthesised images.

Methods: 2D and 3D T1 MRI were acquired during the same scanning session from 31 individuals (males = 14, mean age = 32.16) undergoing evaluation for IGE at the Walton Centre NHS Foundation Trust, Liverpool, as well as 39 healthy age- and sex-matched controls (males = 16, mean age = 32.13). The DL+DiReCT pipeline was used to provide segmentations of the 2D images, and estimates of regional volume and thickness. The 2D scans were also resampled into isotropic images using NiBabel, and preprocessed into synthetic isotropic images using SynthSR. For the 3D scans, untransformed 2D scans, resampled images, and synthesised images, FreeSurfer 7.2.0 was used to create parcellations of 178 anatomical regions (equivalent to the 178 parcellations provided as part of the DL+DiReCT pipeline), defined by the aseg and Destrieux atlases, and FSL FIRST was used to segment subcortical surface shapes. Spatial correspondence and intraclass correlations between the morphometrics of the five parcellations were first determined, then subcortical surface shape abnormalities associated with IGE were identified by comparing the FSL FIRST outputs of patients with controls.

Results: When standardised to the metrics derived from the 3D scans, cortical volume and thickness estimates trended lower for the untransformed 2D, DL+DiReCT, resampled, and SynthSR images, whereas subcortical volume estimates did not differ. Dice coefficients revealed a low spatial similarity between the cortices of the 3D scans and the other images overall, which was higher in the subcortical structures. Intraclass correlation coefficients reiterated this disparity, with estimates of thickness being less similar than those of volume, and DL+DiReCT estimates trending less similar than the other images types. For the people with epilepsy, the 3D scans showed significant surface deflations across various subcortical structures when compared to healthy controls. Analysis of the untransformed 2D scans enabled

the detection of a subset of subcortical abnormalities, whereas analyses of the resampled and synthetic images attenuated almost all significance.

Conclusions: Generalised image synthesis methods do not currently attenuate partial volume effects resulting from low through plane resolution in anisotropic MRI scans, instead quantitative analyses using 2D images should be interpreted with caution, and researchers should consider the potential implications of preprocessing.

Keywords: epilepsy, quantitative MRI, deep-learning, image synthesis, morphometry, shape analysis

Key Points

- Estimates of cortical and subcortical morphometry/volumetry from synthetic images were not closer to the ground truth than estimates from 2D scans.
- Spatial correspondence of regional parcellations to their ground truth analogue was not higher in synthetic images than 2D scans.
- Priors in synthetic image creation models may render them insensitive to subtle biomarkers.
- Resampling and generalised synthesis are not currently replacements for research quality acquisitions in quantitative MRI research.
- The results of studies using synthetic iamges should be interpreted in a separate context to those using untransformed data.

Abbreviations

- ML - Machine Learning
- qMRI - Quantitative Magnetic Resonance Imaging
- PVEs - Partial-volume Effects
- DL - Deep Learning
- CNN - Convolutional Neural Network
- GAN - Generative Adversarial Network
- PWE - People With Epilepsy
- IGE - Idiopathic Generalised Epilepsies
- HC - Healthy Controls
- rT1 - Resampled isometric T1 images
- sT1 - SynthSR isometric T1 images
- dT1 - DL+DiReCT Outputs
- ANTs - Advanced Normalization Tools
- FSL - FMRIB Software Library
- ICCs - Intraclass Correlation Coefficients

Introduction

Clinical data

In the clinical evaluation pathway of non-specialist centres (i.e. general hospitals in the UK), research quality imaging data is not always feasible or accessible. Because of this, there is a growing amount of legacy data on PACS systems that remains relatively underexplored.[1] There is a growing interest in sustainability within science, which extends to minimising research waste, and the utility of repurposed clinical data for quantitative neuroimaging analyses are axiomatic.[2] Many of the benefits associated with open-source datasets also apply to methods that increase the usability of retrospective clinical data, such as reducing research costs and increasing accessibility - repurposed clinical MRI data can be anonymised, distributed, and used by researchers who may not otherwise have access to it.[3,4] Access to large amounts of neuroimaging data is especially valuable in the advent of big data approaches.[5] Alongside (and partly due to) the popularisation of big data research, machine-learning (ML) technologies have also seen adoption within and beyond neuroimaging.[6]

With the proliferation of capable hardware and software, quantitative MRI (qMRI) evaluation has become a common method for exploring structural correlates of various pathologies, providing insights into both the prognostication and characterisation of neurological disease. There is, however, a trade-off between spatial resolution and ease of acquisition, which can influence quantitative analysis.[7–9] It is a recognised limitation of MRI research that analyses are dependent on input data quality, and deviations from the research ideal (i.e. 3D, or near-isotropic, 1mm MP-RAGE, a sequence that analysis software is often designed to favour) are detrimental to the reliability of the results.[10,11]

Partial-volume effects (PVEs), error introduced by the difficulty in deconvolving multiple tissue types from a single voxel intensity, which can lead to misclassification of tissue, are more pronounced when voxels are larger.[12] Whilst qualitative inspection by experts is less susceptible to systematic bias resulting from acquisition parameters than quantitative analyses are, anisotropic voxels are differentially susceptible to PVEs in different orientations, leading to an overestimation of the volume of the cortical surface.[13,14] Low resolution scans are therefore not ideal for brain morphometry and deviations of basic imaging parameters (i.e. voxel size) can lead to systematic bias.[15–17] Whilst the in plane 2D (anisotropic) resolution of the images can be increased without much consequence to the patient/clinician, acquiring data in 3D can significantly increase scan times, which increases cost and patient discomfort. Long scans are undesirable, so there is an argument for increasing the step size (and thus improving

patient experience/reducing artifact susceptibility) through plane, and creating anisotropic voxels in an attempt to facilitate qualitative visual analysis/increase diagnostic yield of the scan without overextending resources.[18]

Processing clinical data

Several methods for coercing raw/clinical imaging data into a more homogenous format (i.e. isotropic) have been proposed, tested, and in some cases even widely adopted. There has been, however, no comparative systematic evaluation of these methods with reference to same session intrasubject 3D 'ideal' data. This paper will therefore apply, and evaluate the performance of, three methods for facilitating the qMRI analysis of 2D MRI scans: resampling, ML-based segmentation, and ML-based synthesis. The first and simplest, resampling, is a context-naïve mathematical operation used to split or combine *k* pre-existing voxels into *j* voxels of arbitrary dimensions, using interpolation algorithms to estimate values which satisfy the (unknown) midpoints between two known voxel values.[19] The `conform` command in the Nibabel software suite is a popular method of accomplishing this.[20] Resampling can be used to create the impression that an image is isotropic by splitting voxels, however, resampling cannot resolve the systematically biased PVEs introduced by an anisotropic acquisition. Nonetheless, it is a common neuroimaging analysis technique that has a myriad of uses, including preparing anisotropic scans for morphometry with the DL+DiReCT tool, the second method we will evaluate in this paper.

Briefly, ML is a branch of artificial intelligence wherein algorithms are developed from a subset of data, and extrapolated to represent a broader range of input/outcome operations.[21,22] When a machine learning model contains multiple hidden computational layers, it is considered a Deep Learning (DL) model. Developed for use with contrast-enhanced clinical MRI scans, DL+DiReCT uses a proprietary DL-based segmentation tool, DeepSCAN, to parcellate tissue classes in images with systematic bias, before estimating cortical thickness using the ANTs DiReCT tool.[20,23–25] Through the implementation of a U-net CNN (Convolutional Neural Network) trained on parcellations from a large training data set, DeepSCAN uses generalised information on the cortical structure (i.e. context priors) to partially attenuate the bias introduced by anisotropy when classifying tissue.[26,27]

The third method of image manipulation we will apply in this paper is image synthesis. SynthSR, a ML-based joint super-resolution and contrast synthesis method, uses a 3D U-net CNN to create 1mm isotropic MP-RAGE versions of input images. More specifically,

a model trained on paired images generates an isotropic analogue of the input using context priors to supplement the information in the input, and then a model trained on source and target modality pairs (augmented with a further GAN (Generative Adversarial Network) model that supplements 'real scan' pairs with synthetic ones to enhance the training data) re-estimates the tissue response intensities that would be expected in an MP-RAGE scan.[28] This approach is novel for its ability to compute super-resolution and contrast synthesis of an image without paired data (i.e. explicitly coupled source and target images), its ability to combine data from multiple heterogenous scans, and its ability to synthesise a reference contrast (MP-RAGE).[29–32]

Structural biomarkers in idiopathic generalised epilepsy

The current understanding of epilepsy—one of the most common neurological diseases globally—as a network disorder has been heavily influenced by the results of qMRI studies based on prospectively acquired, research quality data.[33] Patterns of volumetric, morphometric, and network abnormalities have been associated with common, syndrome-specific, and prognostic phenotypes of epilepsy.[34–39] People with epilepsy (PWE) routinely present with marked (yet subtle) patterns of subcortical (thalamic, hippocampal, striatal) atrophy that may not appear visually abnormal on clinical MRI, even at the point of diagnosis.[40] Atrophy of the subcortical structures has been repeatedly observed in cohorts of people with Idiopathic Generalised Epilepsy (IGE), potentially related to syndrome-specific pathomechanisms; abnormalities of the hippocampi, putamen, and thalami are of particular relevance to epileptogenesis, constituting biomarkers of IGE.[41–43] The reliability of these structural correlates makes them an appropriate model for testing the veracity of image manipulation.

Objectives

The purpose of this study was to systematically evaluate which of three openly-available methods, when provided with 2D MRI scans, returned the output (i.e. the iso-analogue) with the closest similarity to 'fit for purpose' 3D MRI data: resampling with NiBabel, ML-based segmentation with DL+DiReCT, or ML-based super-resolution and contrast synthesis with SynthSR. We quantified this through comparison of the iso-analogue derived volumetry and morphometry with that from the silver-standard isotropic MRI. With this data, we first determined within subject variation, and then explored IGE-related subcortical surface shape abnormalities. This follows numerous studies testing the efficacy of DL+DiReCT and SynthSR, alongside other methods with more specific

utility—such as super-resolution of isotropic images, or in neonates.[48] This is, however, the first investigation into and comparison of the relative efficacy of several common and generalisable image synthesis methods. To help neuroimagers make a more informed choice, we demonstrate the relative caveats and benefits of four potential structural preprocessing pipelines.

Materials and Methods

Participants and scanning protocols

This study was conducted using participants and data that has been previously described.[9,49–51] All participants provided informed written consent, and data collection was approved by the local ethics committee (UK Research Ethical Committee ref. 14/NW/0332). 33 individuals diagnosed with IGE (People With Epilepsy, PWE) at the Walton Centre NHS Foundation Trust Liverpool were recruited, as well as 39 age- and sex-matched healthy controls (HC). Diagnoses of IGE were made by an epileptologist based on semiology, EEG characteristics, and clinical history, and are consistent with current International League Against Epilepsy guidelines.[52] Pertinent demographic information was recorded at the time of recruitment.

Scanning took place in a GE Discovery MR750 3.0T MRI scanner at the Walton Centre NHS Foundation Trust between 2014 and 2016. For each participant, two scans were acquired over a single session: a 3D T1-FSPGR PURE (Fast Spoiled Gradient Echo with Phased Array Uniformity Enhancement signal inhomogeneity correction) with TR = 8.2ms, TI = 450ms, TE = 3.22ms, flip angle = 12°, voxel size = 0.9 × 0.9 × 1.0mm, acquisition matrix = 256 × 256 × 135, and FOV = 240 × 240 × 135mm; and a 2D T1-FLAIR (Fluid Attenuated Inversion Recovery) with TI = 920ms, TE = 9.94ms, flip angle = 111°, voxel size = 0.4 × 3.0 × 0.4mm, acquisition matrix = 512 × 52 × 396, and FOV = 220 × 156 × 170mm. See Figure 1. for a comparison of anisotropic and isotropic image acquisitions.

Image synthesis

The code used for image synthesis, processing, and analysis is provided on GitHub. As shown in Figure 2, the 2D T1 scans were, in parallel, preprocessed using DL+DiReCT and SynthSR. The intermediate 3D T1 image (voxel size 1 × 1 × 1mm, matrix = 256 × 256 × 256), which had been skull-stripped with `HD-BET` and resampled with

`nibabel.processing.conform` as part of the DL+DiReCT pipeline, was extracted, as well as used as the input for the DeepSCAN (model 7) segmentation step.[20,23,53]

After ML-based segmentation, ANTs DiReCT was automatically applied to estimate cortical thickness and volume measurements for regions based on a modified version of the Destrieux atlas provided in FreeSurfer 5.3.[24–27,54] For each 2D scan, the standalone implementation of SynthSR was used to predict a best-guess estimate of a 3D T1 image (voxel size 1 × 1 × 1mm, matrix = 221 × 157 × 171), using a model based on the included training dataset.[32,48] After this stage, we had four structural images per participant, as well as DL+DiReCT metrics:

- 3D Scan: T1-FSPGR-PURE, acquired with a near-isotropic resolution
- (Untransformed) 2D Scan: T1-FLAIR, acquired with an anisotropic resolution
- rT1: Resampled (mathematically interpolated) and skull-stripped 2D T1
- sT1: SynthSR derived pseudo-isotropic T1
- dT1: The quantitative outputs of the DL+DiReCT pipeline

Image preprocessing

To concisely reflect structural neuroimaging clinical research, as clinically interpretable metrics, our comparisons were limited to volumes and thicknesses of the cortical parcellations, and volumes of the subcortical regions. Assessment of volumetric and morphometric measurements began with the reconstruction of tissue surfaces across all images and participants using the `recon-all` command in FreeSurfer 7.2.0.[55] From the DL+DiReCT pipeline, regional volume and thickness measurements for ML-based segmentations based on the rT1 images were also available, which we have included as a ML-based volumetric alternative to the surface-based morphometry in FreeSurfer.[23,24] To assure consistent comparisons, parcellations redefined by the more recent releases of FreeSurfer were combined to match those provided by DL+DiReCT. The final parcellation maps contained 148 cortical regions and 30 subcortical segmentations.

To evaluate the ability of resampling and synthesis to compensate for resolution-based information loss when distinguishing between groups, subcortical surface shapes were compared between PWE and HC, first in the 3D scans, and then in the non-3D images (i.e. the untransformed 2D scans, rT1, and sT1). For subcortical volumetry and shape analysis, all four structural images across all participants were first bias-corrected with the `N4BiasFieldCorrection` algorithm, brain extracted with the `BET` tool, linearly registered to the brain-extracted MNI template with `FLIRT`, and then segmented and

modelled with the `run_first_all` pipeline (from the Advanced Normalization Tools [ANTs] library and FMRIB Software Library [FSL], respectively).[56,57]

Statistical analyses

To quantify the similarity of the non-3D images with the 'silver-standard' ground truth represented by the 3D scans, Dice coefficients across all parcellations/segmentations were computed using FSL maths to first create 'overlap' images; a demonstration of how Dice coefficients can be used to assess regional overlap is provided in figure 3.[56,58] Further comparisons of cortical metrics were carried out in R, the details of which (including packages, and visualisation tools) are presented in Appendix 1.[59,60] Descriptive assessment of deviation between the quantitative metrics derived from the non-3D images were carried out after standardisation (i.e. z-scoring) of the cortical volumes, cortical thicknesses, and subcortical volumes to distributions defined by the respective 3D image metric. Similarity between the metric datasets were measured using intraclass correlation coefficients (ICCs). Following the recommended pipeline, subcortical surface shape comparisons between PWE and HC were modelled via general linear model using the FSL `randomise` permutation testing tool, with age and sex as covariates, and threshold free cluster enhancement for multiple comparisons.[56]

Results

Descriptive volumetric and morphometric comparisons

Standardised grey matter volume and thickness measurements derived from the non-3D images were consistently lower than those recorded in the 3D scans, as shown in Table 1, whereas deviations in the estimates of the subcortical volumetry were more subtle. Metrics computed using the untransformed 2D scans deviated the least from our ground truth, whereas the dT1 metrics tended to deviate the most. Figure 4 presents the distributions of the measurements from the non-3D images when standardised to the 3D metrics, and highlights the abnormally large variability of the cortical dT1 estimates.

Regionwise spatial correspondence

Dice coefficients were computed to quantify the spatial overlap between the 178 regions/segmentations of the 3D scans and the non-3D images (Figure 5). Overlap for the 148 grey matter regions was small on average (2D scan = 0.244, dT1 = 0.110, rT1 = 0.211, sT1 = 0.227), but slightly larger for the 30 subcortical segmentations (2D scan =

0.376, dT1 = 0.203, rT1 = 0.346, sT1 = 0.359). Again, the parcellations defined in the untransformed 2D scans were the most comparable to those of the 3D scans, and the dT1 parcellations were the least. Interestingly, parcellations created using the sT1 were closer to our ground truth than the rT1, although by a small margin.

Metric correspondence across image types

To quantify the similarity of the metrics extracted from each image type, ICCs were computed between the FreeSurfer 3D scan metrics and each of the non-3D image metrics (FreeSurfer and DiReCT). For cortical volume (2D scan = 0.680, dT1 = 0.476, rT1 = 0.624, sT1 = 0.567), cortical thickness (2D scan = 0.399, dT1 = 0.162, rT1 = 0.344, sT1 = 0.201), and subcortical volume (2D scan = 0.681, dT1 = 0.608, rT1 = 0.671, sT1 = 0.629), the untransformed 2D scan metrics had the highest correlation with the 3D scan metrics, followed by those from the rT1 images, then the sT1 images. The dT1 values were the least similar in all three included measures, and volume measurements differed less than thickness (Figure 6).

Demographics

Two PWE were excluded due to subsequent rediagnosis, resulting in a sample of 31 PWE (males = 14, mean age = 32.16) and 39 HC (n = 39, males = 16, mean age = 32.38). The two groups did not significantly differ in age or sex. Demographic and clinical characteristic summaries are presented in table 2, alongside the results of difference tests. Full sample characteristics are presented in Appendix 2.

Subcortical surface shape

The subcortical shape differences commonly reported in IGE were observable in our sample (see table 3). Following permutation testing (`HC - PWE + age + sex`, [Appendix 6]), clusters of significant inwards surface deflation (as a proxy of regional subcortical atrophy) were reported in the caudate (left *p corr* = .012, right *p corr* = .017), pallidum (left *p corr* = .019, right *p corr* = .003), putamen (left *p corr* = .039, right *p corr* = .003), and thalamus (left *p corr* = .001, right *p corr* = .002) bilaterally, as well as in the left accumbens (*p corr* = .034) and right amygdala (*p corr* = .013) of our PWE sample when compared to controls. The same comparisons in the untransformed 2D scan data returned significant clusters in the left accumbens (*p corr* = .021), right caudate (*p corr* = .005), right putamen (*p corr* = .030), and thalamus bilaterally (left *p corr* = .002, right *p corr* < .000). No significant clusters could be identified in the rT1, or, aside from a

significant inwards deflation of the right thalamus (*p corr* = .028) in the sT1 images. All clusters which reached *p corr* < .050 are presented in figure 7. Regions of significant outward surface inflation in the PWE group relative to controls were also reported in the caudate and hippocampus (reported fully in table 4).

Discussion

Study summary

This study aimed to evaluate the efficacy of resampling, DL+DiReCT, and SynthSR for synthesising images suitable for morphometric analysis when using anisotropic MRI scans representative of that available retrospectively. Rigorous control during the acquisition of our dataset (2D and 3D scans acquired in the same session and scanner) makes this sample ideal for resolution comparisons, despite its unremarkable size, and the inclusion of both HC and PWE provides a means of assessing clinical viability.

Estimates of cortical volume and cortical thickness were markedly reduced in all of the non-3D images, however variations in subcortical volume estimates were more evenly dispersed around the ground truth. Dice analyses showed the greatest concordance between the 3D scans and the untransformed 2D images, although overlaps were small for the cortical parcellations in general—this is likely influenced by the granularity of our parcellation atlas and relatively greater structural complexity of grey matter regions. Overlap improved slightly for the subcortical structures, and again, the untransformed 2D images were the most similar to our ground truth data. In the context of the previous analyses, it is unsurprising that ICCs were highest between the metrics extracted from the 3D scans and untransformed 2D scans, followed by the rT1, sT1, and lastly the dT1 metrics.

In the 3D data comparisons, PWE presented with evidence of distributed subcortical shape differences indicative of both atrophy and growth, relative to HC. There is an abundance of evidence for subcortical abnormalities related to the development and maintenance of an epileptogenic environment, which supports the hypothesis that these shape differences are representative of 'real' properties.[29–32] Whilst a subset of these abnormalities were also identifiable in the untransformed 2D scans (left accumbens, right putamen, caudate and thalamus bilaterally), false positives (right hippocampus) and false negatives (right amygdala, left putamen, pallidum bilaterally) draw into question the reliability of analyses based on data with a low spatial resolution. Greater statistical rigour could indeed be employed to control for false positives, but the possibility of false-negative results might encourage researchers to make use of

experimental image synthesis methodologies that are insufficiently validated. In this study, both resampling and image synthesis with SynthSR led to a near-total attenuation of group-level true-positive differences, suggesting that—in the context of data with low spatial resolution—even with increased statistical rigour, evaluation of untransformed data might prove more illuminating.

Morphometry with synthetic data

The consistent underestimation of grey matter volume and thickness in the non-3D images suggests that lower spatial resolution biases PVEs towards white matter and extraparenchymal tissue (i.e. the meninges). This is seemingly exacerbated by preprocessing methods, hence the greater disparity between our ground truth and the estimates from the rT1, sT1, and dT1 images. This behaviour is difficult to avoid in interpolation, however there is potential for a model-based solution, provided that appropriate data is used in training.

The choice of training data used when creating a DL model is crucial to the function of the resultant network. For example, in the presence of uncertainty, a super-resolution model trained on healthy data could be expected to attenuate pathology present in input data, making it more difficult to identify despite the improved clarity of the output image. Generally, the data a model is trained on will provide a measure of bias (reflecting priors) with downstream impacts that make it difficult to quantify, and the variety of the data should be proportional to the breadth of the datasets it is applicable to.[61,62] Caution is therefore advisable when using models claiming input agnosticism, due to the massive variability of neurological pathology.

DL+DiReCT

Disparity as a result of inappropriate generalisation was demonstrated in this study by the overall poor performance of DL+DiReCT. DL+DiReCT was developed (and trained) for the purpose of estimating cortical morphometry in contrast enhanced (Multiple Sclerosis) imaging data. Furthermore, the DiReCT cortical thickness measurements employed in the DL+DiReCT have previously been shown to systematically differ from those provided by FreeSurfer.[63] Whilst DL+DiReCT's potential application in generalised contrast enhanced data is promising, it also warrants further independent investigation beyond the scope of this paper. We therefore present our results not as a commentary on the ability of DL+DiReCT to perform its stated purpose, but rather to demonstrate

the variance that can be introduced through methodological decisions, compounded by inappropriate use of experimental preprocessing steps.

SynthSR

Unlike DL+DiReCT, SynthSR claims applicability to a broad range of use cases under the blanket of generalised contrast and resolution synthesis.[28] The authors posit that due to its training data, SynthSR is robust to variations in resolution, contrast, and pathology, providing a general algorithm for transforming any clinical MRI into a synthetic 1mm MP-RAGE image, which is supported by its ability to facilitate distinction between the brains of people with Alzheimer's Disease and healthy controls, or improve the application of preprocessing steps in lesioned images. In this study, we have shown that there are contexts in which SynthSR is not an appropriate solution for image anisotropy—namely in the evaluation of subcortical surface shape. This is possibly a consequence of the specificity of the abnormalities (IGE is typically MR-negative after visual inspection), which causes them to be insufficiently represented in SynthSR's training data. Furthermore, we have demonstrated that synthetic images do not increase agreement (compared to untransformed 2D images) with the ground truth for the estimation of morphometric properties. Synthetic images represent data that has been fundamentally changed, which is reflected in morphometric analysis, and should be clearly stated throughout dissemination.

Resampling

Perhaps surprisingly, metrics estimated from the resampled images were also further from our ground truth than those from the untransformed 2D scans. Even more strikingly, no subcortical surface shape abnormalities reached significance in the resampled data, which suggests that researchers hoping to upsample 2D data into 3D (to facilitate examination with software requiring isotropic inputs) may inadvertently attenuate the same morphometric biomarkers they aim to quantify

Conclusion

Any one of the methods presented herein may be readily employed in an attempt to coerce morphometric information from low-resolution (retrospective or clinical) imaging data. However, we have shown that contrary to popular opinion, the evaluation of anisotropic data may constitute the most appropriate strategy when feasible. In a dataset typical of a clinical context, iso-analogue images provided by resampling and

synthesis techniques showed a reduction in the relative preservation of morphometric properties otherwise discernible from anisotropic T1 scans. Neuroimaging researchers should carefully evaluate their choice of preprocessing before proposing morphometric relationships, and when applicable, clearly state that results are based on manipulated data.

Postscript

Data and code availability Data available on request from the authors. Code freely available on GitHub.

Author contributions CR: Conceptualisation, investigation, visualisation, writing - original draft preparation, writing - review and editing. ORCID ID: https://orcid.org/0000-0002-3824-1681

CdB: Conceptualisation. ORCID ID: https://orcid.org/0000-0002-2433-9776

KD: Resources.

SB: Resources. ORCID ID: https://orcid.org/0000-0002-6817-6497

AM: Writing - review and editing. ORCID ID: https://orcid.org/0000-0002-6861-8806

SSK: Writing - review and editing, supervision. ORCID ID: https://orcid.org/0000-0001-5247-9795

Funding and acknowledgements SK acknowledges support from the UK Medical Research Council (Grant Number MR/S00355X/1).

Declaration of competing interests The authors declare no conflicts of interest. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Supplementary material Scripts: GitHub

References

1. Papanicolas I, Woskie LR, Jha AK. Health Care Spending in the United States and Other High-Income Countries. *JAMA*. 2018;319(10):1024-1039. doi:10.1001/jama.2018.1150
2. Samuel G, Lucivero F, Lucassen AM. Sustainable biobanks: A case study for a green global bioethics. *Global Bioethics*. 2022;33(1):50-64. doi:10.1080/11287462.2021.1997428
3. Laird AR. Large, open datasets for human connectomics research: Considerations for reproducible and responsible data use. *NeuroImage*. 2021;244:118579. doi:10.1016/j.neuroimage.2021.118579
4. Madan CR. Scan Once, Analyse Many: Using Large Open-Access Neuroimaging Datasets to Understand the Brain. *Neuroinformatics*. 2022;20(1):109-137. doi:10.1007/s12021-021-09519-6
5. Van Horn JD, Toga AW. Human Neuroimaging as a "Big Data" Science. *Brain Imaging Behav*. 2014;8(2):323-331. doi:10.1007/s11682-013-9255-y
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553, 7553):436-444. doi:10.1038/nature14539
7. Despotović I, Goossens B, Philips W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Comput Math Methods Med*. 2015;2015:450341. doi:10.1155/2015/450341
8. Vidal-Jordana A, Pareto D, Sastre-Garriga J, et al. Measurement of Cortical Thickness and Volume of Subcortical Structures in Multiple Sclerosis: Agreement between 2D Spin-Echo and 3D MPRAGE T1-Weighted Images. *AJNR Am J Neuroradiol*. 2017;38(2):250-256. doi:10.3174/ajnr.A4999
9. Brownhill D, Chen Y, Kreilkamp BAK, et al. Automated subcortical volume estimation from 2D MRI in epilepsy and implications for clinical trials. *Neuroradiology*. Published online October 18, 2021. doi:10.1007/s00234-021-02811-x
10. Mugler III JP, Brookeman JR. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*. 1990;15(1):152-157. doi:10.1002/mrm.1910150117
11. Edelman RR, Dunkle E, Koktzoglou I, et al. Rapid whole-brain magnetic resonance imaging with isotropic resolution at 3 Tesla. *Invest Radiol*. 2009;44(1):54-59. doi:10.1097/RLI.0b013e31818eee3c

12. Firbank MJ, Coulthard A, Harrison RM, Williams ED. Partial volume effects in MRI studies of multiple sclerosis. *Magnetic Resonance Imaging*. 1999;17(4):593-601. doi:10.1016/S0730-725X(98)00210-0
13. Rebsamen M, Capiglioni M, Hoepner R, et al. Growing importance of brain morphometry analysis in the clinical routine: The hidden impact of MR sequence parameters. *J Neuroradiol*. Published online April 26, 2023:S0150-9861(23)00198-0. doi:10.1016/j.neurad.2023.04.003
14. Lie IA, Kerklingh E, Wesnes K, et al. The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. *Eur Radiol*. 2022;32(5):3576-3587. doi:10.1007/s00330-021-08405-8
15. Tor-Díez C, Pham C-H, Meunier H, et al. Evaluation of cortical segmentation pipelines on clinical neonatal MRI data. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.; 2019:6553-6556. doi:10.1109/EMBC.2019.8856795
16. Haller S, Falkovskiy P, Meuli R, et al. Basic MR sequence parameters systematically bias automated brain volume estimation. *Neuroradiology*. 2016;58(11):1153-1160. doi:10.1007/s00234-016-1737-3
17. May A, Gaser C. Magnetic resonance-based morphometry: A window into structural plasticity of the brain. *Current Opinion in Neurology*. 2006;19(4):407. doi:10.1097/01.wco.0000236622.91495.21
18. Mayberg M, Green M, Vasserman M, et al. Anisotropic neural deblurring for MRI acceleration. *Int J Comput Assist Radiol Surg*. 2022;17(2):315-327. doi:10.1007/s11548-021-02535-6
19. Li A, Mueller K, Ernst T. Methods for efficient, high quality volume resampling in the frequency domain. In: *IEEE Visualization 2004*.; 2004:3-10. doi:10.1109/VISUAL.2004.70
20. Brett M, Markiewicz CJ, Hanke M, et al. *Nipy/Nibabel: 5.1.0*. Zenodo; 2023. doi:10.5281/zenodo.7795644
21. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol*. 2020;9(2):14. doi:10.1167/tvst.9.2.14
22. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *Journal of Medical Imaging and Radiation Sciences*. 2019;50(4):477-487. doi:10.1016/j.jmir.2019.09.005

23. McKinley R, Wepfer R, Aschwanden F, et al. Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks. *Sci Rep*. 2021;11(1, 1):1087. doi:10.1038/s41598-020-79925-4

24. Das SR, Avants BB, Grossman M, Gee JC. Registration based cortical thickness measurement. *Neuroimage*. 2009;45(3):867-879. doi:10.1016/j.neuroimage.2008.12.016

25. Tustison NJ, Avants BB, Cook PA, et al. The ANTs cortical thickness processing pipeline. In: *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol 8672. SPIE; 2013:126-129. doi:10.1117/12.2007128

26. Rebsamen M, Rummel C, Reyes M, Wiest R, McKinley R. Direct cortical thickness estimation using deep learning-based anatomy segmentation and cortex parcellation. *Human Brain Mapping*. 2020;41(17):4804-4814. doi:10.1002/hbm.25159

27. Rebsamen M, McKinley R, Radojewski P, et al. Reliable brain morphometry from contrast-enhanced T1w-MRI in patients with multiple sclerosis. *Hum Brain Mapp*. 2022;44(3):970-979. doi:10.1002/hbm.26117

28. Iglesias JE, Billot B, Balbastre Y, et al. SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci Adv*. 2023;9(5):eadd3607. doi:10.1126/sciadv.add3607

29. Du J, He Z, Wang L, et al. Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing*. 2020;392:209-220. doi:10.1016/j.neucom.2018.10.102

30. Zhao C, Dewey BE, Pham DL, Calabresi PA, Reich DS, Prince JL. SMORE: A Self-Supervised Anti-Aliasing and Super-Resolution Algorithm for MRI Using Deep Learning. *IEEE Transactions on Medical Imaging*. 2021;40(3):805-817. doi:10.1109/TMI.2020.3037187

31. Huang Y, Shao L, Frangi AF. Simultaneous Super-Resolution and Cross-Modality Synthesis of 3D Medical Images Using Weakly-Supervised Joint Convolutional Sparse Coding. In:; 2017:6070-6079. Accessed September 26, 2023. https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Simultaneous_Super-Resolution_and_CVPR_2017_paper.html

32. Billot B, Greve DN, Puonti O, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*. 2023;86:102789. doi:10.1016/j.media.2023.102789

33. Beghi E. The Epidemiology of Epilepsy. *Neuroepidemiology*. 2020;54(2):185-191. doi:10.1159/000503831

34. Whelan CD, Altmann A, Botía JA, et al. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. *Brain*. 2018;141(2):391-408. doi:10.1093/brain/awx341

35. Hatton SN, Huynh KH, Bonilha L, et al. White matter abnormalities across different epilepsy syndromes in adults: An ENIGMA-Epilepsy study. *Brain*. 2020;143(8):2454-2473. doi:10.1093/brain/awaa200

36. Bonilha L, Keller SS. Quantitative MRI in refractory temporal lobe epilepsy: Relationship with surgical outcomes. *Quant Imaging Med Surg*. 2015;5(2):204-224. doi:10.3978/j.issn.2223-4292.2015.01.01

37. Larivière S, Rodríguez-Cruces R, Royer J, et al. Network-based atrophy modeling in the common epilepsies: A worldwide ENIGMA study. *Science Advances*. Published online November 2020. doi:10.1126/sciadv.abc6457

38. Keller SS, Schoene-Bake JC, Gerdes JS, Weber B, Deppe M. Concomitant Fractional Anisotropy and Volumetric Abnormalities in Temporal Lobe Epilepsy: Cross-Sectional Evidence for Progressive Neurologic Injury. *PLoS One*. 2012;7(10):e46791. doi:10.1371/journal.pone.0046791

39. Keller SS, Richardson MP, O'Muircheartaigh J, Schoene-Bake JC, Elger C, Weber B. Morphometric MRI alterations and postoperative seizure control in refractory temporal lobe epilepsy: Morphometry and Outcome in Epilepsy. *Hum Brain Mapp*. 2015;36(5):1637-1647. doi:10.1002/hbm.22722

40. Leek NJ, Neason M, Kreilkamp BaK, et al. Thalamohippocampal atrophy in focal epilepsy of unknown cause at the time of diagnosis. *Eur J Neurol*. 2021;28(2):367-376. doi:10.1111/ene.14565

41. Du H, Zhang Y, Xie B, et al. Regional atrophy of the basal ganglia and thalamus in idiopathic generalized epilepsy. *Journal of Magnetic Resonance Imaging*. 2011;33(4):817-821. doi:10.1002/jmri.22416

42. Keller SS, Ahrens T, Mohammadi S, et al. Microstructural and volumetric abnormalities of the putamen in juvenile myoclonic epilepsy. *Epilepsia*. 2011;52(9):1715-1724. doi:10.1111/j.1528-1167.2011.03117.x

43. Caciagli L, Wandschneider B, Xiao F, et al. Abnormal hippocampal structure and function in juvenile myoclonic epilepsy and unaffected siblings. *Brain*. 2019;142(9):2670-2687. doi:10.1093/brain/awz215

44. Fiscone C, Curti N, Ceccarelli M, et al. Generalizing the Enhanced-Deep-Super-Resolution neural network to brain MR images: A retrospective study on the Cam-CAN dataset. *eNeuro*. Published online May 10, 2024:ENEURO.0458-22.2023. doi:10.1523/ENEURO.0458-22.2023

45. Rusak F, Santa Cruz R, Lebrat L, et al. Quantifiable brain atrophy synthesis for benchmarking of cortical thickness estimation methods. *Medical Image Analysis*. 2022;82:102576. doi:10.1016/j.media.2022.102576

46. Tian Q, Bilgic B, Fan Q, et al. Improving in vivo human cerebral cortical surface reconstruction using data-driven super-resolution. *Cerebral Cortex*. 2021;31(1):463-482. doi:10.1093/cercor/bhaa237

47. Delannoy Q, Pham CH, Cazorla C, et al. SegSRGAN: Super-resolution and segmentation using generative adversarial networks — Application to neonatal brain MRI. *Computers in Biology and Medicine*. 2020;120:103755. doi:10.1016/j.compbiomed.2020.103755

48. Iglesias JE, Billot B, Balbastre Y, et al. Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage*. 2021;237:118206. doi:10.1016/j.neuroimage.2021.118206

49. Pegg EJ, McKavanagh A, Bracewell RM, et al. Functional network topology in drug resistant and well-controlled idiopathic generalized epilepsy: A resting state functional MRI study. *Brain Commun*. 2021;3(3):fcab196. doi:10.1093/braincomms/fcab196

50. McKavanagh A, Kreilkamp BAK, Chen Y, et al. Altered Structural Brain Networks in Refractory and Nonrefractory Idiopathic Generalized Epilepsy. *Brain Connectivity*. Published online August 5, 2021. doi:10.1089/brain.2021.0035

51. McKavanagh A, Ridzuan-Allen A, Kreilkamp BAK, et al. Midbrain structure volume, estimated myelin and functional connectivity in idiopathic generalised epilepsy. *Epilepsy & Behavior*. 2023;140:109084. doi:10.1016/j.yebeh.2023.109084

52. Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*. 2017;58(4):512-521. doi:10.1111/epi.13709

53. Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*. 2019;40(17):4952-4964. doi:10.1002/hbm.24750

54. Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*. 2010;53(1):1-15. doi:10.1016/j.neuroimage.2010.06.010

55. Fischl B. FreeSurfer. *NeuroImage*. 2012;62(2):774-781. doi:10.1016/j.neuroimage.2012.01.021

56. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012;62(2):782-790. doi:10.1016/j.neuroimage.2011.09.015

57. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908

58. Yao AD, Cheng DL, Pan I, Kitamura F. Deep Learning in Neuroradiology: A Systematic Review of Current Algorithms and Approaches for the New Wave of Imaging Technology. *Radiology: Artificial Intelligence*. 2020;2(2):e190026. doi:10.1148/ryai.2020190026

59. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2023. <URL: https://www.R-project.org/>.

60. RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, PBC.; 2020. http://www.rstudio.com/

61. Gambhir R, Nachman B, Thaler J. Bias and priors in machine learning calibrations for high energy physics. *Phys Rev D*. 2022;106(3):036011. doi:10.1103/PhysRevD.106.036011

62. Konell HG, Junior LOM, Dos Santos AC, Salmon CEG. Assessment of U-Net in the segmentation of short tracts: Transferring to clinical MRI routine. *Magn Reson Imaging*. Published online May 14, 2024:S0730-725X(24)00158-9. doi:10.1016/j.mri.2024.05.009

63. Tustison NJ, Cook PA, Klein A, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage*. 2014;99:166-179. doi:10.1016/j.neuroimage.2014.05.044

Figures

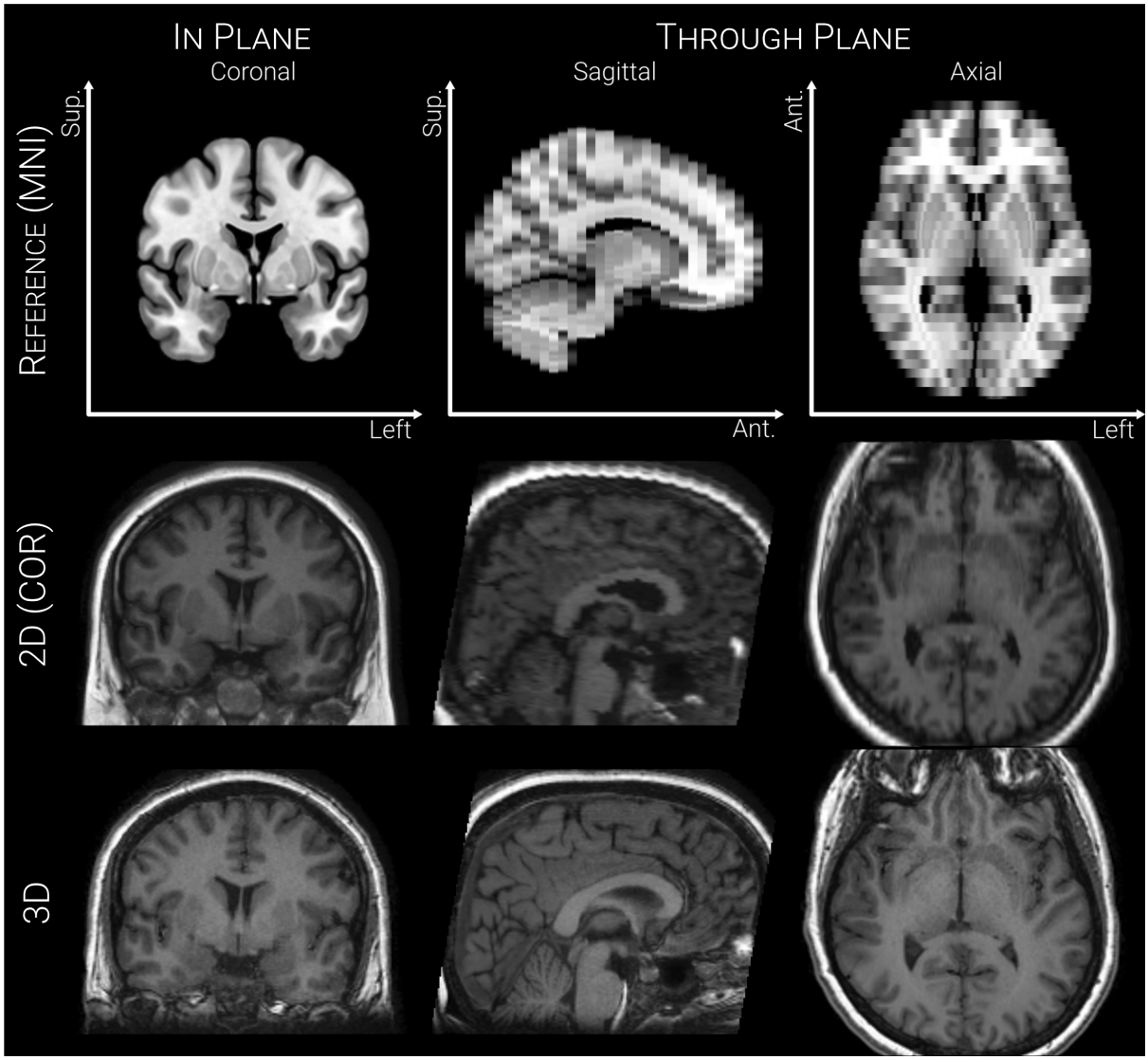


Figure 1. A 2D and 3D T1-weighted MRI scan of the same participant, annotated with the planes defined by acquisition.

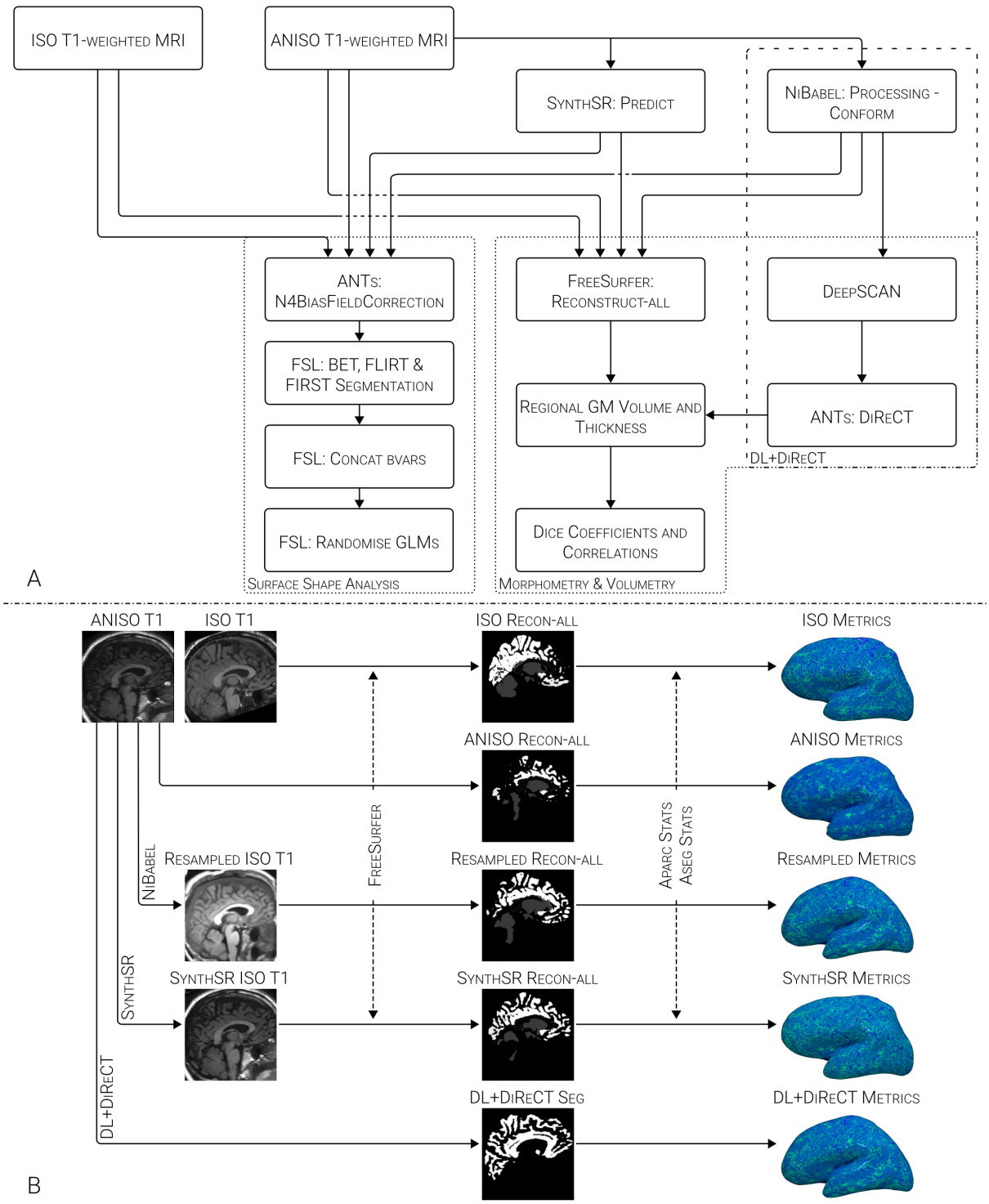
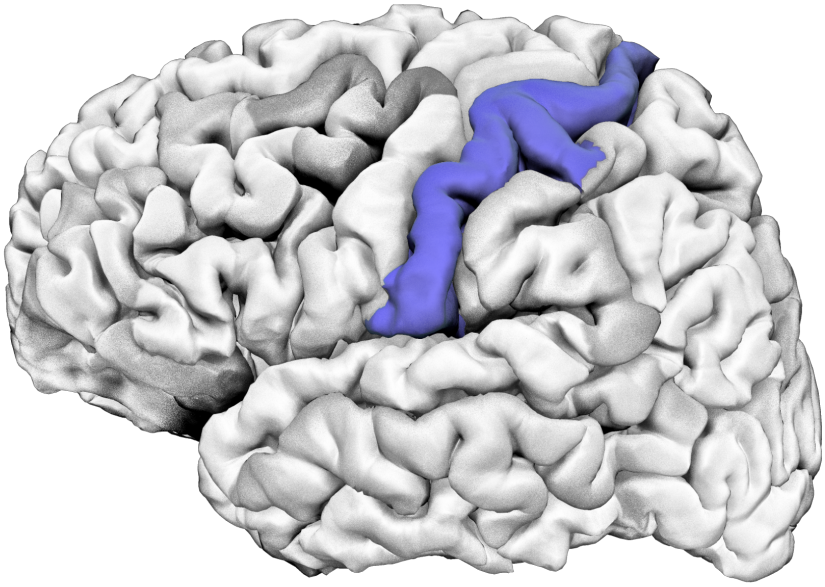
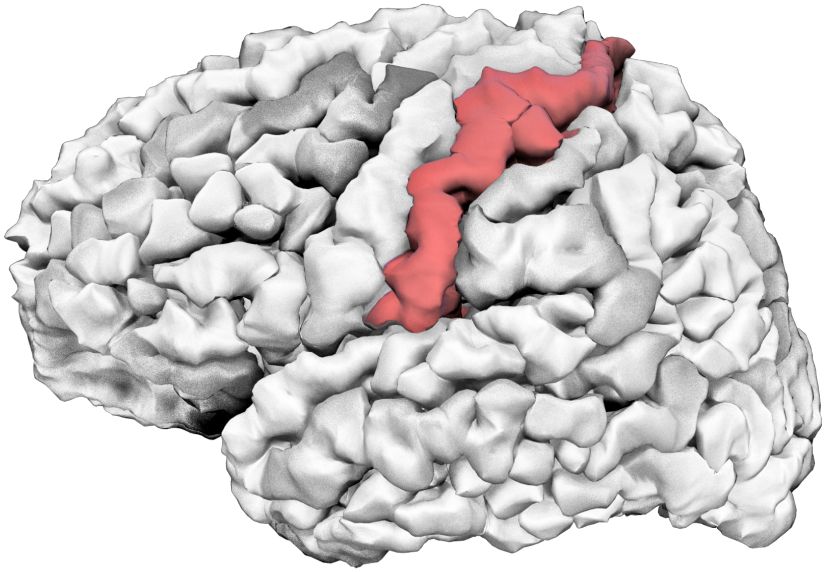


Figure 2. Schematic of (a) the image processing pipeline and (b) the morphometry/volumetry pipeline.

3D
Postcentral
Gyrus



2D
Postcentral
Gyrus



Overlap

Dice Score = 0.23

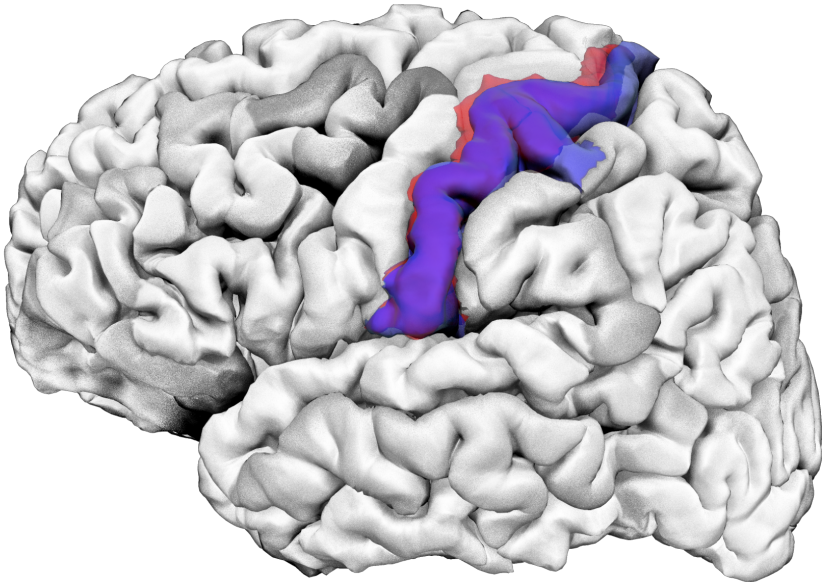


Figure 3. Spatial overlap is sensitive to slight differences in location, as well as in shape.

Figure 4. Violin plots representing the deviations in volume (across 148 cortical and 30 subcortical parcellations) and thickness (148 parcellations only) derived from the different image types. The metrics for each parcellation are standardised to the distribution created from the 3D scan metrics.

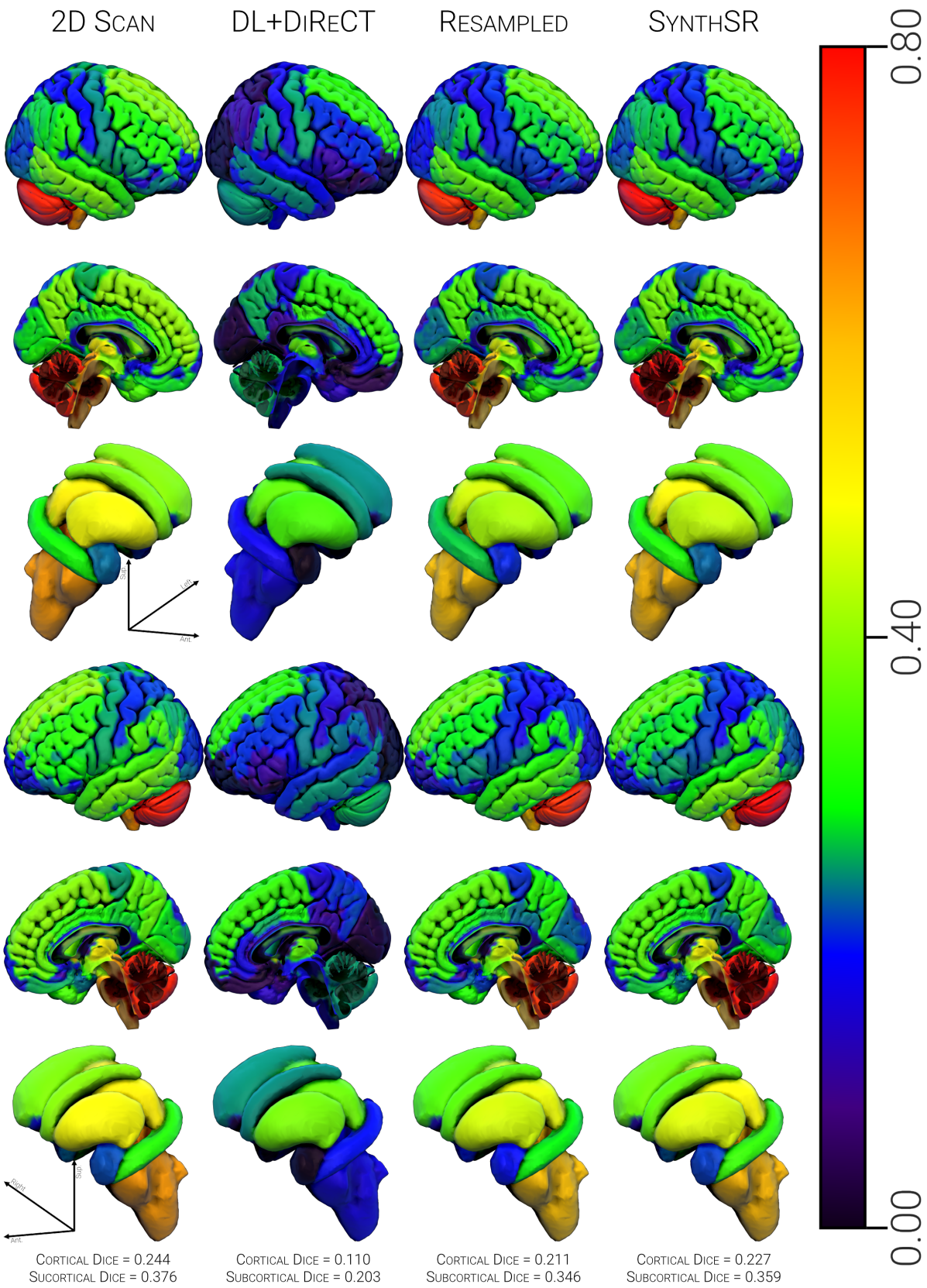


Figure 5. Heatmap representations of the Dice overlap of parcellations created from 3D scans, and from 2D, DL+DiReCT, resampled, and SynthSR-derived analogues. The average Dice coefficients are presented underneath for both the 148 cortical parcellations and 30 subcortical parcellations - note that some segmentations may not be visible.

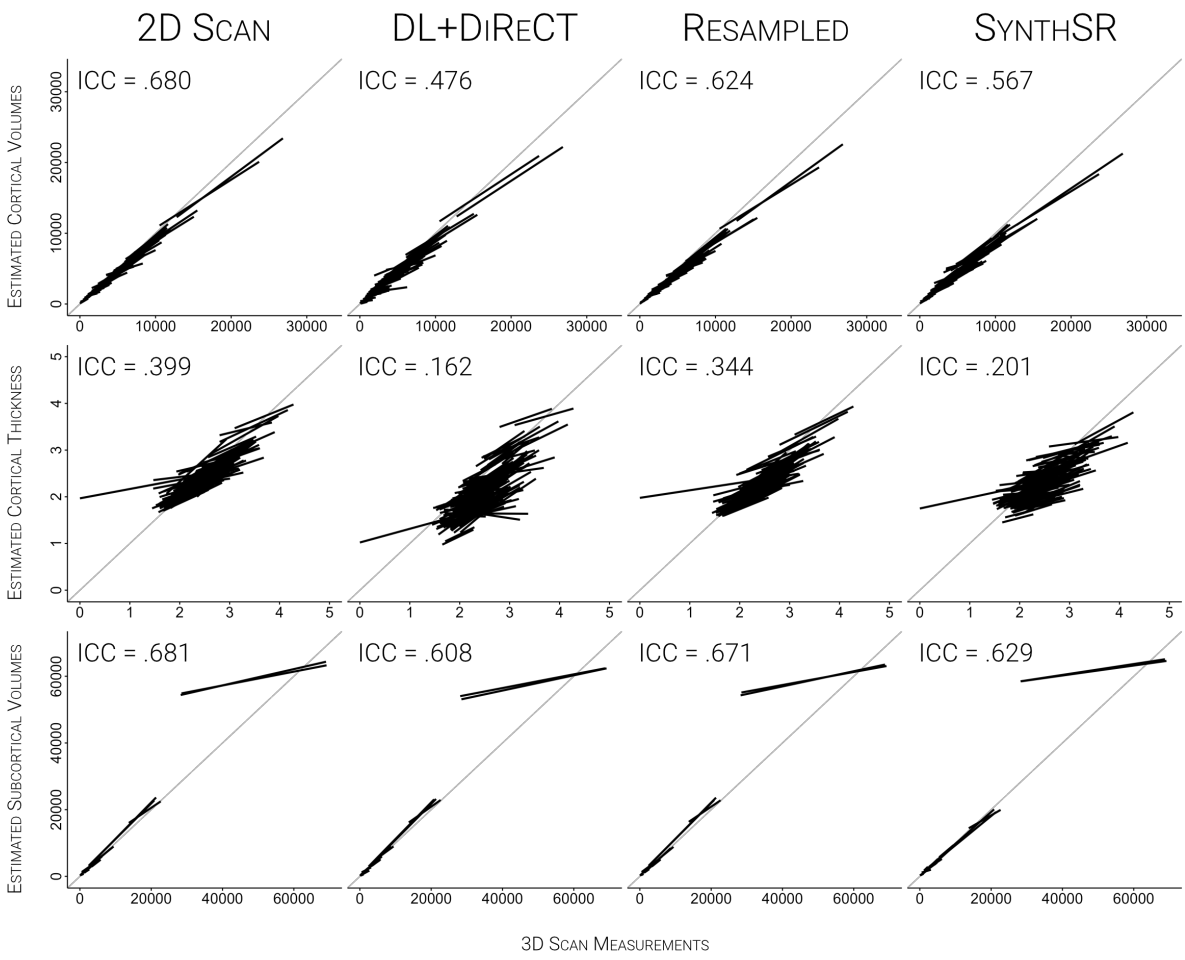


Figure 6. Trend lines of the relationships between FreeSurfer-extracted volume and thickness measurements of parcellated regions in the 3D scans (y-axes) and the 2D scan-, DL+DiReCT-, resampled-, and SynthSR-derived analogues (x-axis). Intraclass correlation coefficients (ICCs) of the morphometrics are presented alongside each group of trendlines, as a quantifiable measure of the similarity.

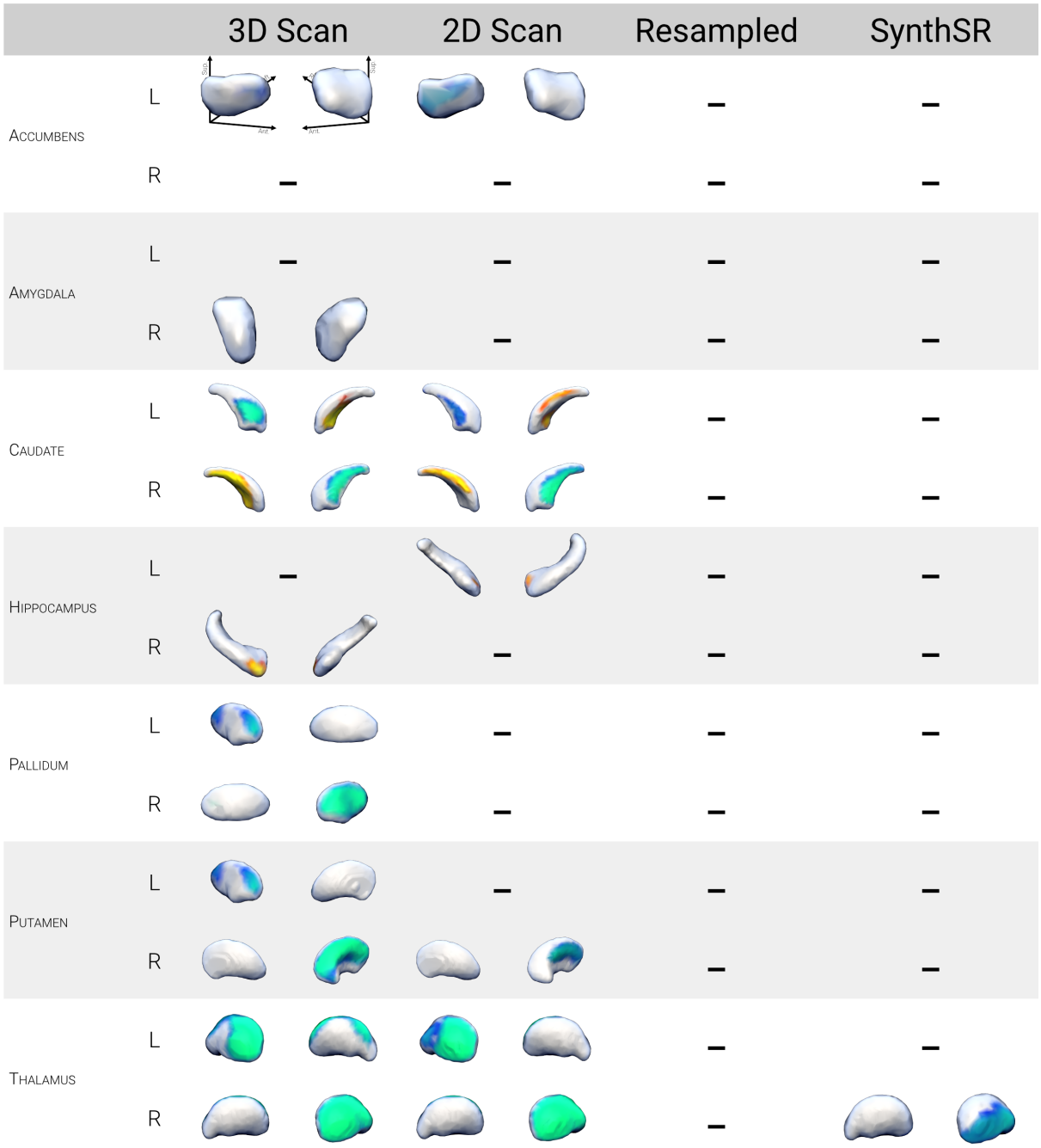


Figure 7. Clusters of subcortical surface shape deflations. Design: Shape ~ Group + Age + Sex. Blue-green represents significant surface deflation for PWE compared to HC, and red-yellow represents significant inflation. Only volumes including clusters with *p corr* < .050 are shown.

Tables

		3D Scan	2D Scan	DL+DiReCT	Resampled	SynthSR
Cortical Regions	Dice score (mean)	-	0.244	0.110	0.211	0.227
	Volume z-score (mean)	-	-0.443	-0.687	-0.652	-0.604
	Volume ICC	-	0.680	0.476	0.624	0.567
	Thickness z-score (mean)	-	-0.499	-2.227	-1.117	-1.133
	Thickness ICC	-	0.399	0.162	0.344	0.201
Subcortical Regions	Dice score (mean)	-	0.376	0.203	0.346	0.359
	Volume z-score (mean)	-	0.015	-0.096	0.020	0.121
	Volume ICC	-	0.681	0.608	0.671	0.629

Table 1. Regional volumetric measurements, extracted from FreeSurfer and DeepSCAN parcellations of 148 cortical and 30 subcortical regions. Regions were parcellated using FreeSurfer in the 3D scans, 2D scans, Resampled images, and SynthSR-derived images. DeepSCAN was applied to the resampled images as part of the DL+DiReCT pipeline.

		HC (n = 39)	PWE (n = 31)	*p* =
Age at Scan	Mean (SD)	32.38 (8.65)	32.16 (14.20)	0.936
Sex	F	23 (58.97)	17 (54.84)	0.917
	M	16 (41.03)	14 (45.16)	

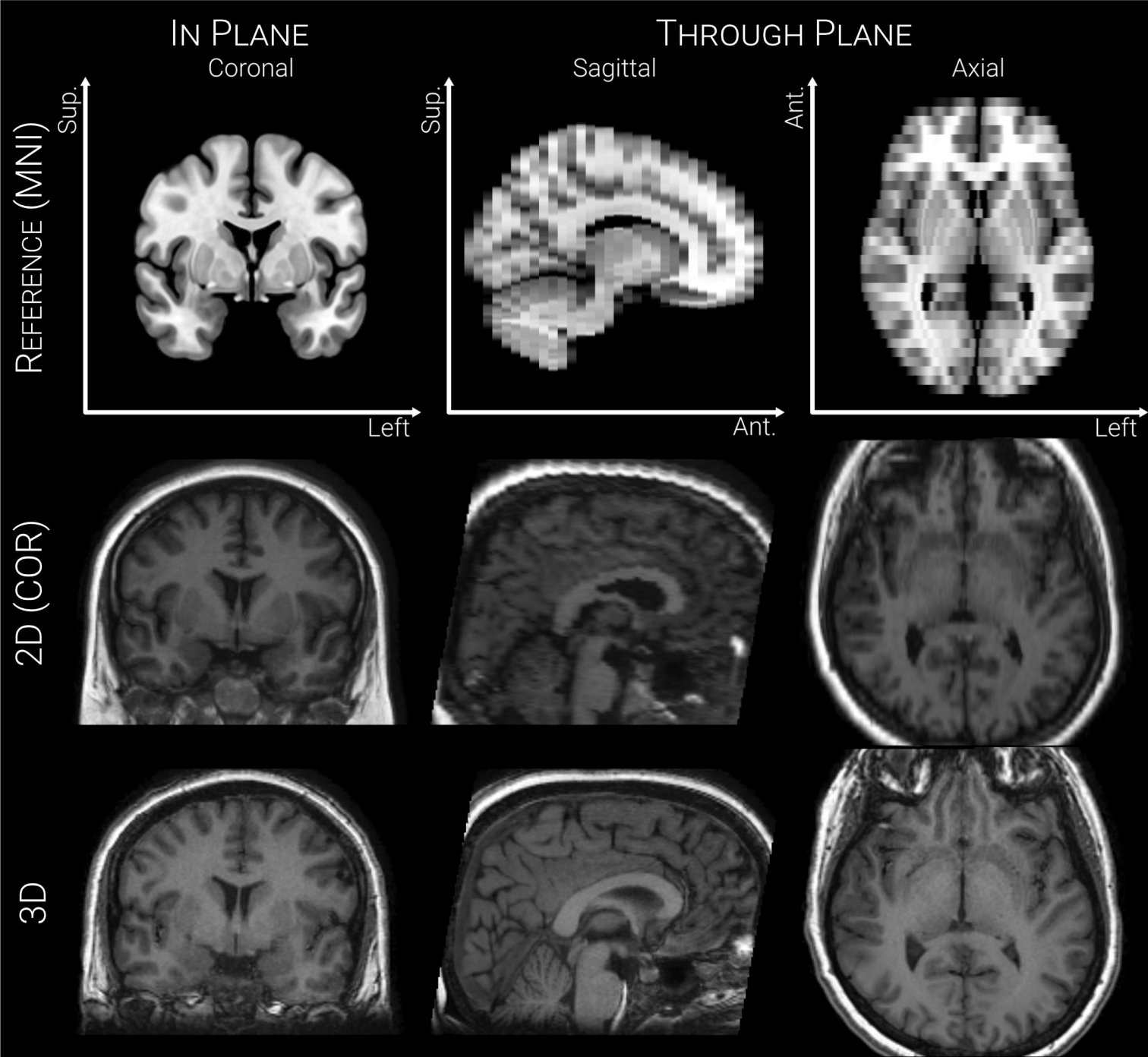
Table 2. Sample demographics.

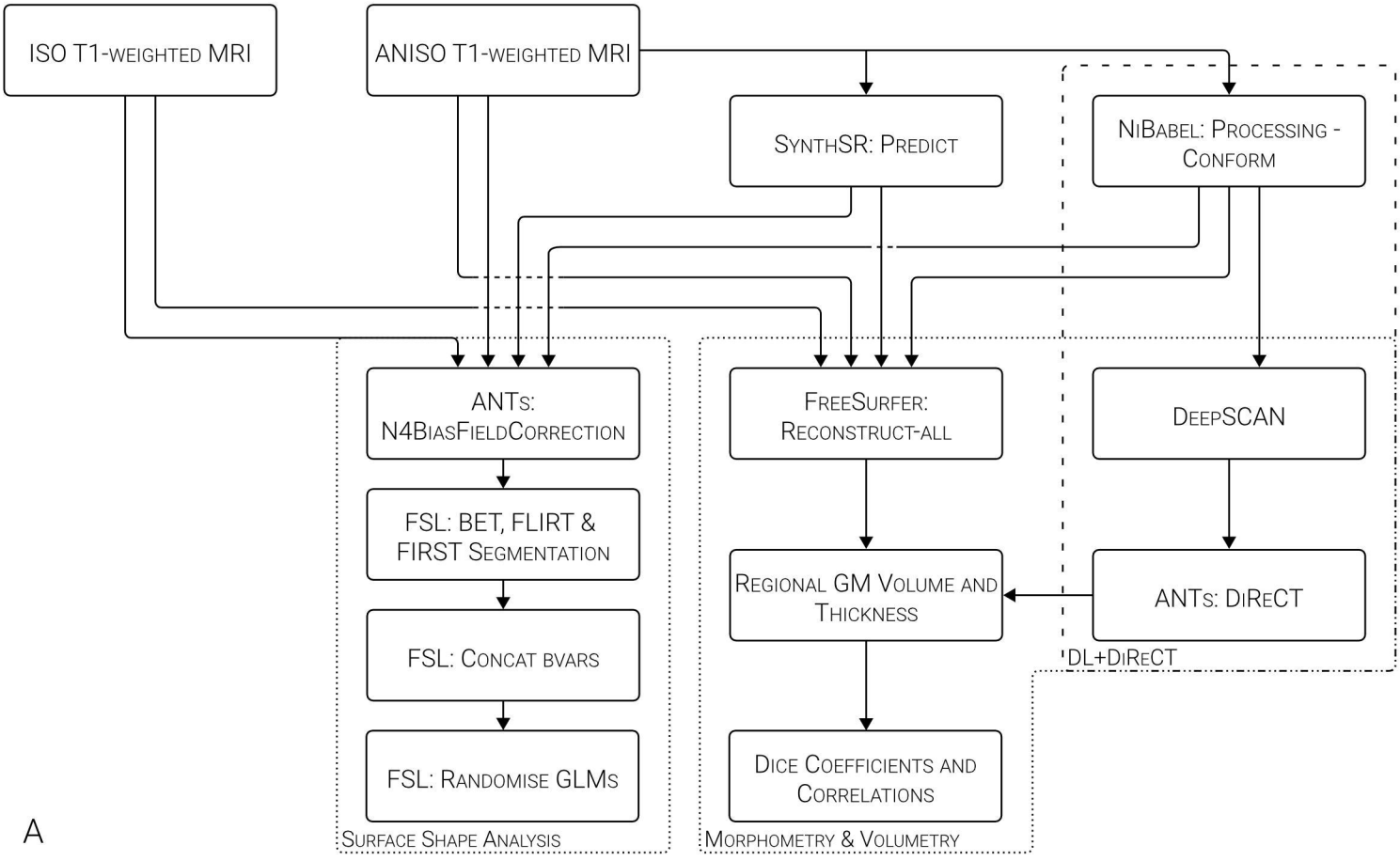
		3D Scan	2D Scan	Resampled	SynthSR
Accumbens	L	.034	.021	.534	.068
	R	.404	.089	.289	.129
Amydala	L	.670	.130	.222	.268
	R	.013	.360	.319	.383
Caudate	L	.012	.065	.879	.110
	R	.017	.005	.746	.422
Hippocampus	L	.195	.117	.583	.088
	R	.238	.158	.090	.447
Pallidum	L	.019	.247	.514	.232
	R	.003	.109	.199	.505
Putamen	L	.039	.140	.832	.266
	R	.003	.030	.718	.351
Thalamus	L	.001	.002	.316	.077
	R	.002	.000	.100	.028

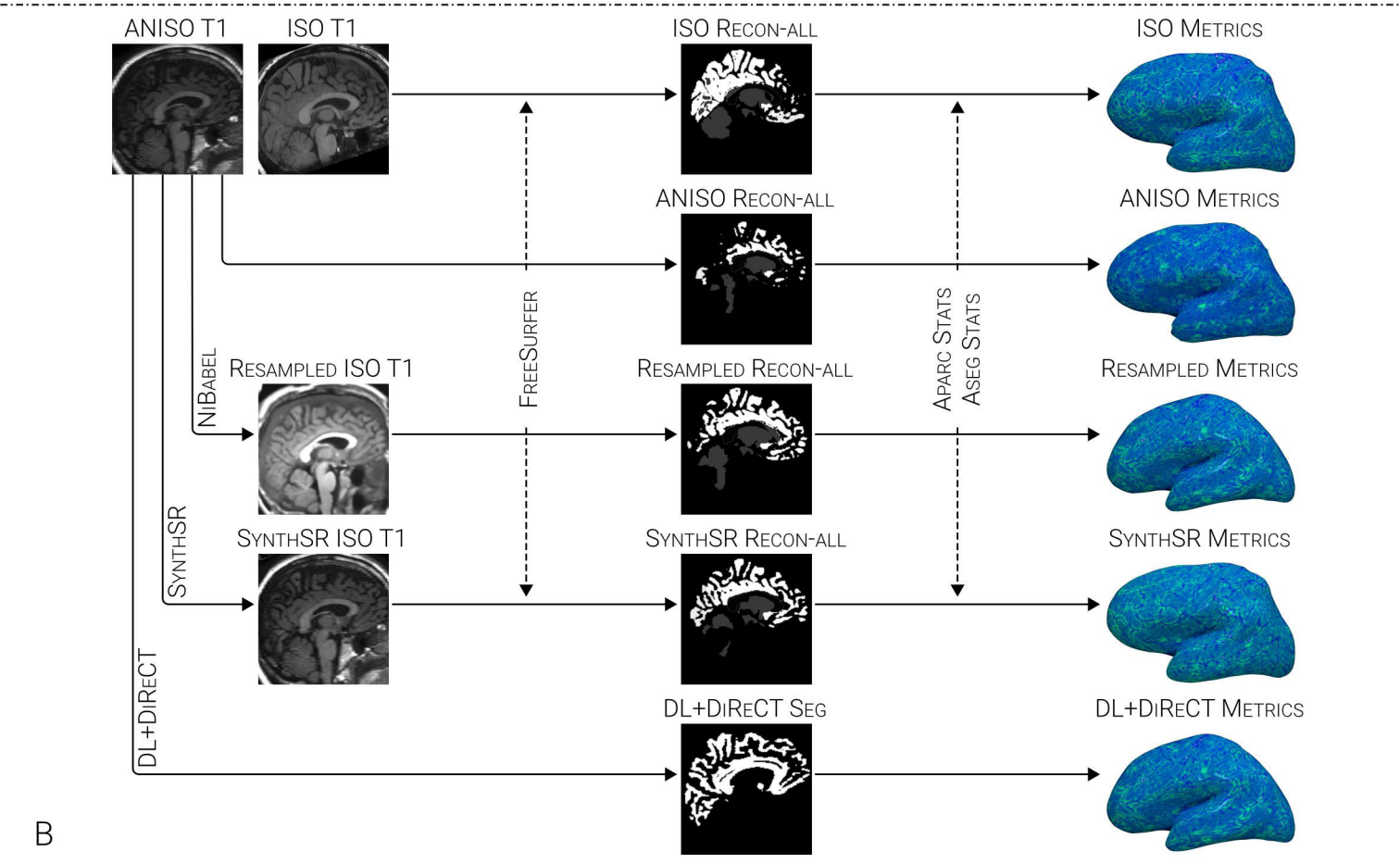
Table 3. Maximal significance of subcortical surface shape deflation clusters in PWE (i.e. contrast: PWE < HC), observed across image types.

		3D Scan	2D Scan	Resampled	SynthSR
Accumbens	L	.127	.210	.771	.471
	R	.563	.413	.605	.703
Amydala	L	.782	.495	.754	.806
	R	.481	.501	.827	.949
Caudate	L	.000	.016	.779	.417
	R	.004	.011	.127	.267
Hippocampus	L	.176	.035	.861	.233
	R	.012	.051	.585	.398
Pallidum	L	.826	.446	.863	.522
	R	.177	.622	.904	.785
Putamen	L	.766	.416	.114	.999
	R	.397	.421	.698	.793
Thalamus	L	.482	.774	.677	.539
	R	.456	.372	.821	.217

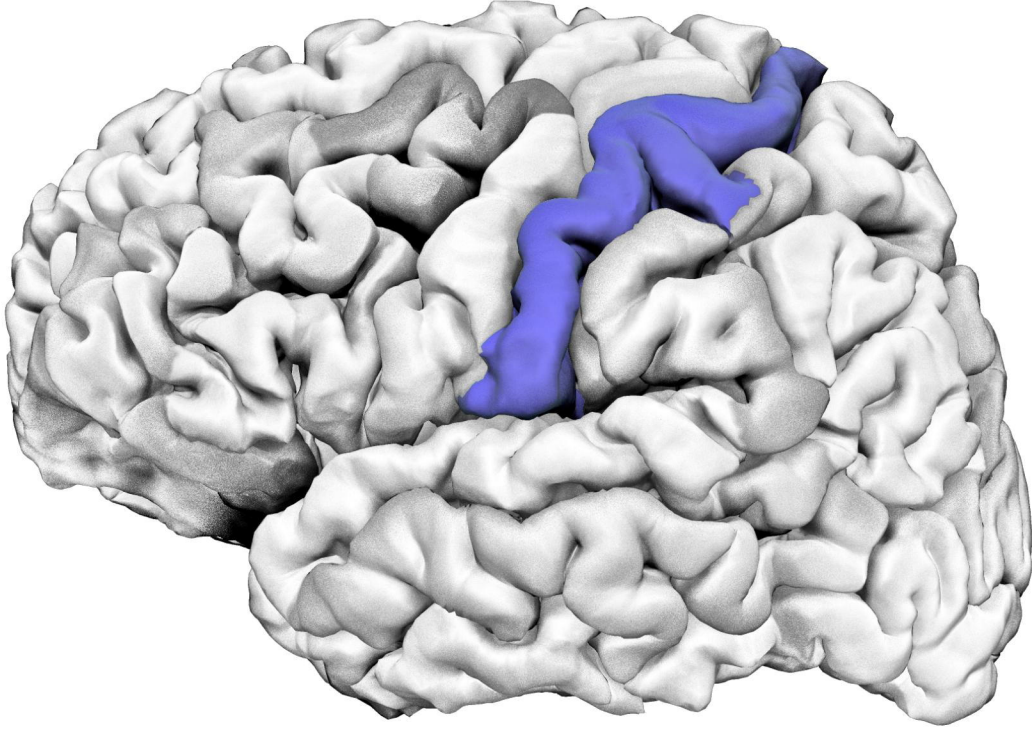
Table 4. Maximal significance of subcortical surface shape inflation clusters in PWE (i.e. contrast: PWE > HC), observed across image types.


In Plane
Through Plane
Coronal
Sagittal
Axial
Reference (MNI)
Sup.
Left
Sup.
Ant.
Ant.
Left
2D (COR)
3D

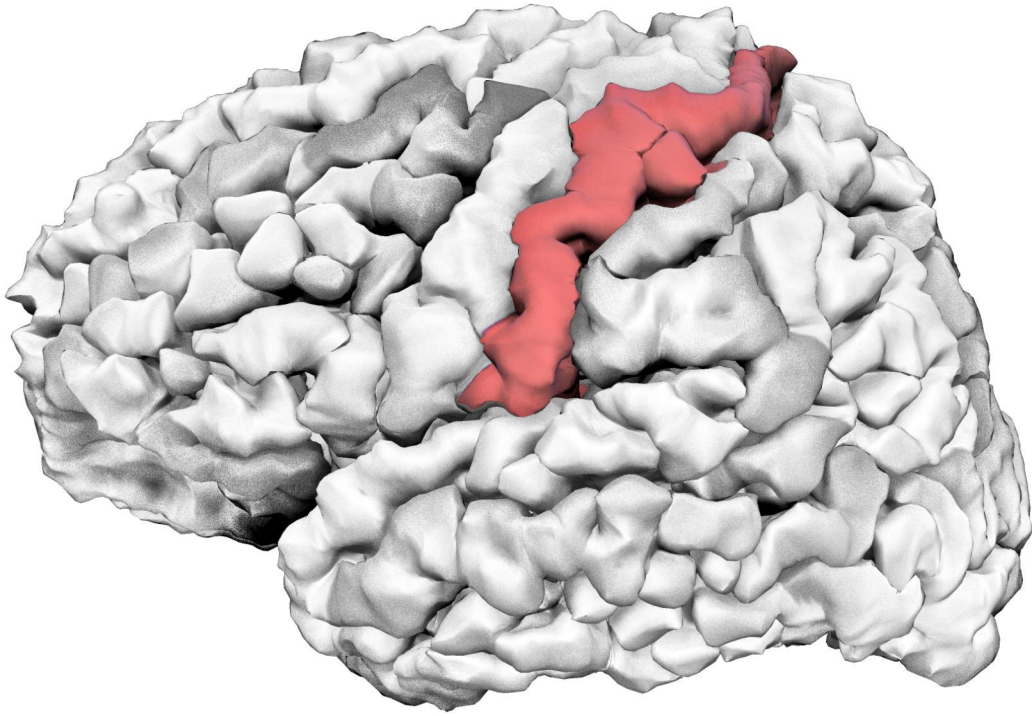

ISO T1-weighted MRI
ANISO T1-weighted MRI
SynthSR: Predict
NiBabel: Processing - Conform
ANTs: N4BiasFieldCorrection
FSL: BET, FLIRT & FIRST Segmentation
FSL: Concat bvars
FSL: Randomise GLMs
Surface Shape Analysis
FreeSurfer: Reconstruct-all
Regional GM Volume and Thickness
Dice Coefficients and Correlations
Morphometry & Volumetry
DeepSCAN
ANTs: DiReCT
DL+DiReCT
A


ANISO T1
ISO T1
ISO Recon-all
ISO Metrics
ANISO Recon-all
ANISO Metrics
NiBabel
Resampled ISO T1
FreeSurfer
Resampled Recon-all
Aparc Stats
Aseg Stats
Resampled Metrics
SynthSR
SynthSR ISO T1
SynthSR Recon-all
SynthSR Metrics
DL+DiReCT
DL+DiReCT Seg
DL+DiReCT Metrics
B

3D
Postcentral
Gyrus

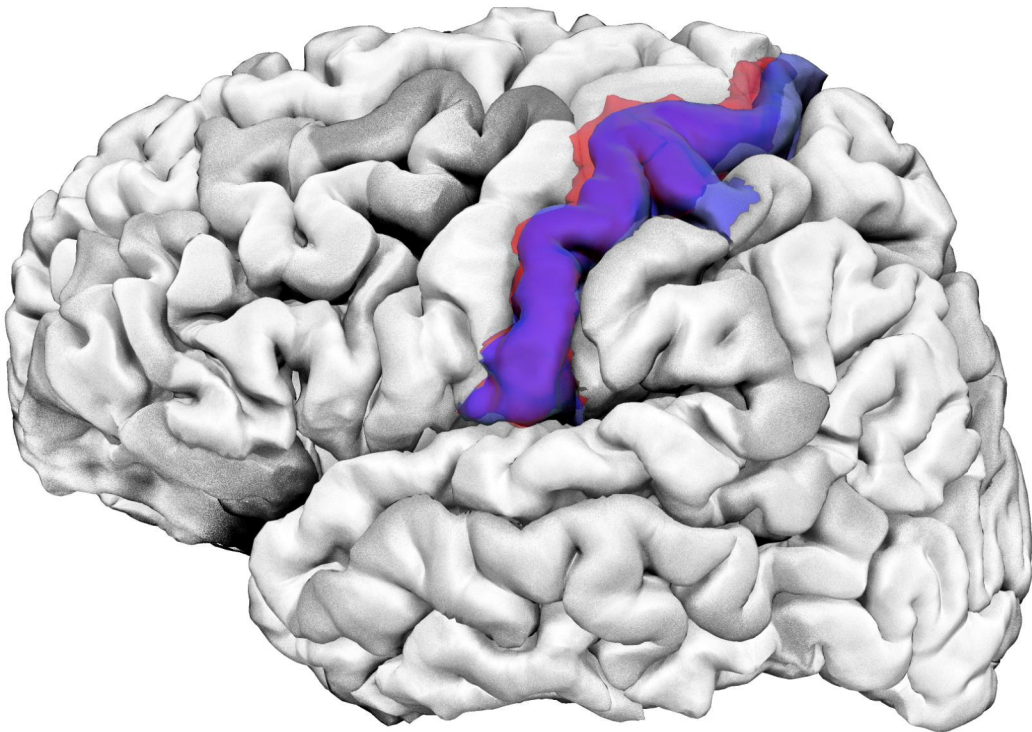


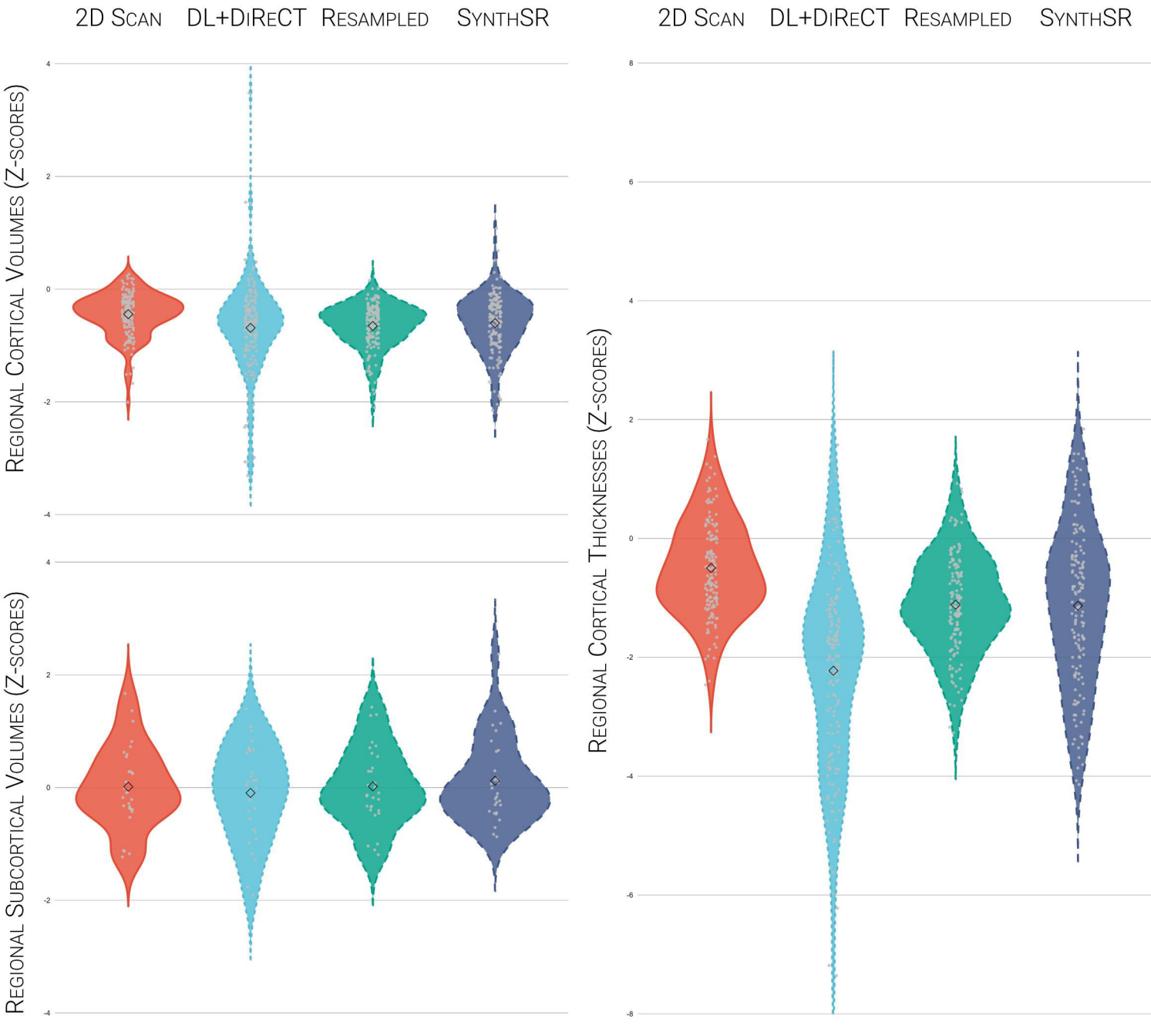
2D
Postcentral
Gyrus

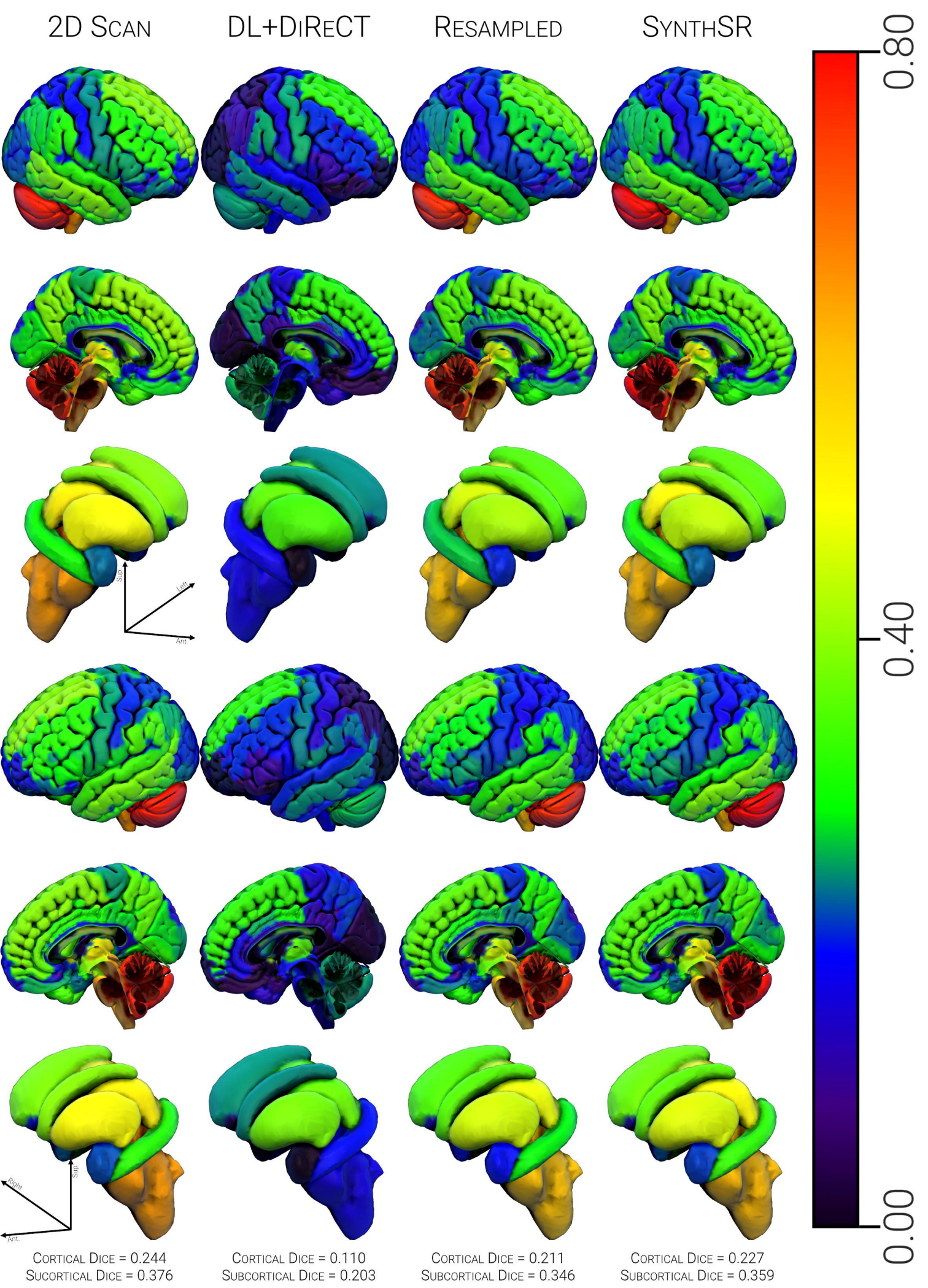


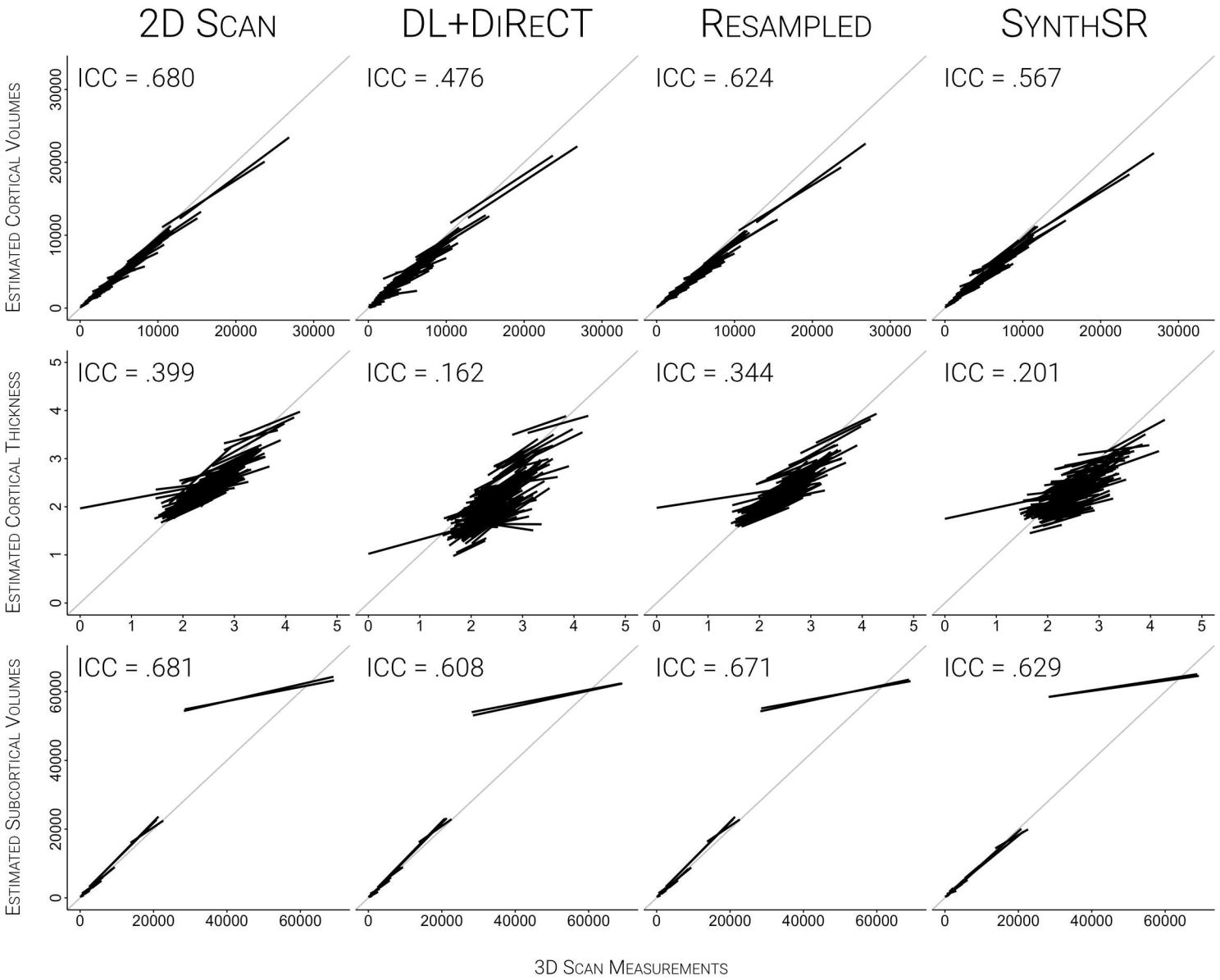
Overlap

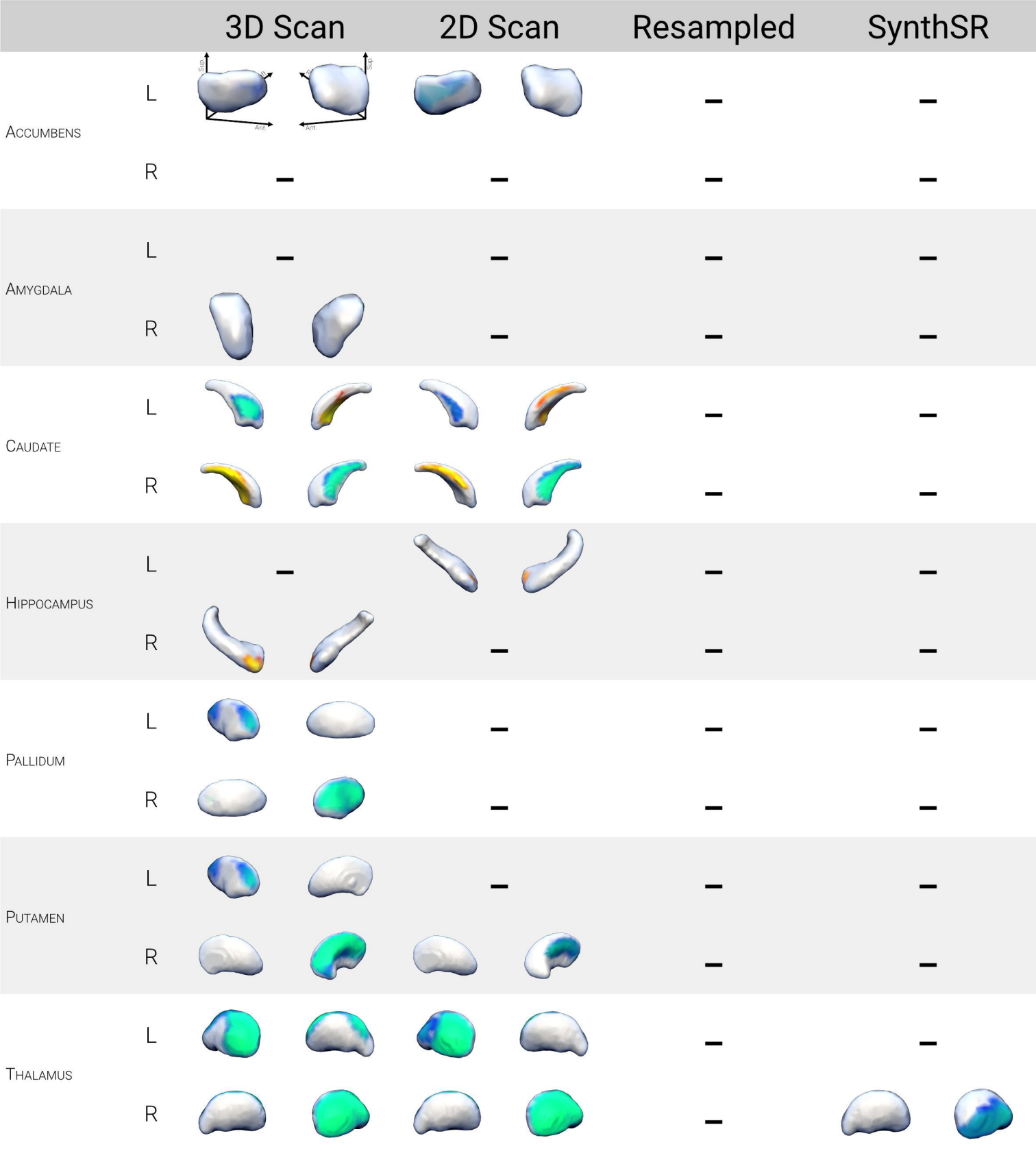
Dice Score = 0.23




2D Scan
DL+DiReCT
Resampled
SynthSR
Regional Cortical Volumes (Z-scores)
Regional Subcortical Volumes (Z-scores)
Regional Cortical Thicknesses (Z-scores)
4
2
0
-2
-4
8
6
-6
-8


2D Scan
DL+DiReCT
Resampled
SynthSR
0.80
0.40
0.00
Sup.
Left
Ant.
Sup.
Right
Ant.
Cortical Dice = 0.244
Sucortical Dice = 0.376
Cortical Dice = 0.110
Subcortical Dice = 0.203
Cortical Dice = 0.211
Subcortical Dice = 0.346
Cortical Dice = 0.227
Subcortical Dice = 0.359


2D Scan
DL+DiReCT
Resampled
SynthSR
ICC = .680
ICC = .476
ICC = .624
ICC = .567
ICC = .399
ICC = .162
ICC = .344
ICC = .201
ICC = .681
ICC = .608
ICC = .671
ICC = .629
Estimated Cortical Volumes
Estimated Cortical Thickness
Estimated Subcortical Volumes
3D Scan Measurements


3D Scan
2D Scan
Resampled
SynthSR
Accumbens
Amygdala
Caudate
Hippocampus
Pallidum
Putamen
Thalamus
L
R
Sup.
Ant.

		3D Scan	2D Scan	DL+DiReCT	Resampled	SynthSR
Cortical Regions	Dice score (mean)	-	0.244	0.110	0.211	0.227
	Volume z-score (mean)	-	-0.443	-0.687	-0.652	-0.604
	Volume ICC	-	0.680	0.476	0.624	0.567
	Thickness z-score (mean)	-	-0.499	-2.227	-1.117	-1.133
	Thickness ICC	-	0.399	0.162	0.344	0.201
Subcortical Regions	Dice score (mean)	-	0.376	0.203	0.346	0.359
	Volume z-score (mean)	-	0.015	-0.096	0.020	0.121
	Volume ICC	-	0.681	0.608	0.671	0.629

		HC (n = 39)	PWE (n = 31)	*p* =
Age at Scan	Mean (SD)	32.38 (8.65)	32.16 (14.20)	0.936
Sex	F	23 (58.97)	17 (54.84)	0.917
	M	16 (41.03)	14 (45.16)	

		3D Scan	2D Scan	Resampled	SynthSR
Accumbens	L	.034	.021	.534	.068
	R	.404	.089	.289	.129
Amydala	L	.670	.130	.222	.268
	R	.013	.360	.319	.383
Caudate	L	.012	.065	.879	.110
	R	.017	.005	.746	.422
Hippocampus	L	.195	.117	.583	.088
	R	.238	.158	.090	.447
Pallidum	L	.019	.247	.514	.232
	R	.003	.109	.199	.505
Putamen	L	.039	.140	.832	.266
	R	.003	.030	.718	.351
Thalamus	L	.001	.002	.316	.077
	R	.002	.000	.100	.028

		3D Scan	2D Scan	Resampled	SynthSR
Accumbens	L	.127	.210	.771	.471
	R	.563	.413	.605	.703
Amydala	L	.782	.495	.754	.806
	R	.481	.501	.827	.949
Caudate	L	.000	.016	.779	.417
	R	.004	.011	.127	.267
Hippocampus	L	.176	.035	.861	.233
	R	.012	.051	.585	.398
Pallidum	L	.826	.446	.863	.522
	R	.177	.622	.904	.785
Putamen	L	.766	.416	.114	.999
	R	.397	.421	.698	.793
Thalamus	L	.482	.774	.677	.539
	R	.456	.372	.821	.217