



Virtual brain grafting: Enabling whole brain parcellation in the presence of large lesions

Ahmed M. Radwan^{a,*}, Louise Emsell^{a,b,c}, Jeroen Blommaert^d, Andrey Zhylyka^e, Silvia Kovacs^f, Tom Theys^{c,g}, Nico Sollmann^{h,i}, Patrick Dupont^{c,j}, Stefan Sunaert^{a,c,f}

^a KU Leuven, Department of Imaging and Pathology, Translational MRI, Leuven, Belgium

^b KU Leuven, Department of Geriatric Psychiatry, University Psychiatric Center, Leuven, Belgium

^c KU Leuven, Leuven Brain Institute (LBI), Department of Neurosciences, Leuven, Belgium

^d KU Leuven, Department of Oncology, Leuven, Belgium

^e Department of Biomedical Engineering, Eindhoven University of Technology, Netherlands

^f UZ Leuven, Department of Radiology, Leuven, Belgium

^g KU Leuven, Department of Neurosciences, Research Group Experimental Neurosurgery and Neuroanatomy, Leuven, Belgium

^h Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

ⁱ TUM-Neuroimaging Center, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

^j KU Leuven, Laboratory for Cognitive Neurology, Department of Neurosciences, Leuven, Belgium

ARTICLE INFO

Keywords:

Lesioned brain parcellation
Brain MRI lesion-filling
Brain MRI lesion-inpainting
Gliomas
Clinical imaging

ABSTRACT

Brain atlases and templates are at the heart of neuroimaging analyses, for which they facilitate multimodal registration, enable group comparisons and provide anatomical reference. However, as atlas-based approaches rely on correspondence mapping between images they perform poorly in the presence of structural pathology. Whilst several strategies exist to overcome this problem, their performance is often dependent on the type, size and homogeneity of any lesions present. We therefore propose a new solution, referred to as Virtual Brain Grafting (VBG), which is a fully-automated, open-source workflow to reliably parcellate magnetic resonance imaging (MRI) datasets in the presence of a broad spectrum of focal brain pathologies, including large, bilateral, intra- and extra-axial, heterogeneous lesions with and without mass effect.

The core of the VBG approach is the generation of a lesion-free T1-weighted image, which enables further image processing operations that would otherwise fail. Here we validated our solution based on Freesurfer recon-all parcellation in a group of 10 patients with heterogeneous gliomatous lesions, and a realistic synthetic cohort of glioma patients ($n = 100$) derived from healthy control data and patient data.

We demonstrate that VBG outperforms a non-VBG approach assessed qualitatively by expert neuroradiologists and Mann-Whitney U tests to compare corresponding parcellations (real patients $U(6,6) = 33$, $z = 2.738$, $P < .010$, synthetic-patients $U(48,48) = 2076$, $z = 7.336$, $P < .001$). Results were also quantitatively evaluated by comparing mean dice scores from the synthetic-patients using one-way ANOVA (unilateral VBG = 0.894, bilateral VBG = 0.903, and non-VBG = 0.617, $P < .001$). Additionally, we used linear regression to show the influence of lesion volume, lesion overlap with, and distance from the Freesurfer volumes of interest, on labeling accuracy.

VBG may benefit the neuroimaging community by enabling automated state-of-the-art MRI analyses in clinical populations using methods such as FreeSurfer, CAT12, SPM, Connectome Workbench, as well as structural and functional connectomics. To fully maximize its availability, VBG is provided as open software under a Mozilla 2.0 license (https://github.com/KUL-Radneuron/KUL_VBG).

1. Introduction

Brain mapping is the process of spatially representing known brain structures, e.g., anatomical regions, or functional areas on a coordinate system or space. Attempts to create a map of the human brain predate the 20th century (Moore, 2011; Standring, 2016), initially ex-

isting mostly in anatomical and medical textbooks as sketches of ex-vivo specimens. The earliest modern detailed cytoarchitectonic representation of the human brain is the Brodmann atlas, published in a monograph in 1909 and remains widely used in the field today (Mandal et al., 2012). Thanks to Talairach and Tournoux (Talairach and Tournoux, 1988) and similar coordinate systems as well as other advances in medical imaging and computer science, brain atlases are now widely used in neuroimaging and neurosurgery for structural brain mapping (Tzourio-Mazoyer et al., 2002); (Frazier et al., 2005);

* Corresponding author.

E-mail address: ahmed.radwan@kuleuven.be (A.M. Radwan).

(Desikan et al., 2006); (Makris et al., 2006); (Goldstein et al., 2007); (Nowinski, 2016); (Dickie et al., 2017). Atlas-based brain mapping can be achieved accurately and non-invasively for individual subjects or even groups using brain images acquired from a magnetic resonance imaging (MRI) scanner. Many atlases are available non-commercially as digital image files, representing different brain structures using specific image intensity values (Dickie et al., 2017).

Typically, brain atlas images are defined in the space of a template or reference brain image, which may be a single subject image or an averaged image from a specific cohort, e.g. Montreal Neurological Institute 152 (MNI152) brain template (Mazziotta et al., 1995). Structural brain mapping using an atlas requires it first to be spatially matched to the brain image being mapped, or vice versa. This can be achieved by image registration, a process that maximizes spatial similarity between source and target images, also known as normalization, when a standard brain image such as MNI152 is the target. Many strategies exist, including rigid-body and affine registration, and non-affine approaches such as symmetric diffeomorphic registration (Avants et al., 2008). If successful, the resulting transforms and/or warps can be used to bring the atlas to the patient's brain or, vice versa, achieving a basic brain mapping, which may be efficient but lacking in tissue and subject specificity. More advanced pipelines e.g. Freesurfer (Fischl, 2012) recon-all (FreeSurferWiki, 2020a) leverage several steps e.g. brain extraction, tissue segmentation and surface-based registration to maximize tissue and subject specificity.

Structural brain mapping, also called segmentation or parcellation, can normally be automated, and is available through several programs such as FSL (Jenkinson et al., 2012), SPM (Ashburner et al., 2006), ANTs (Avants et al., 2011a), Freesurfer (Fischl, 2012), AFNI (Cox, 1996) and others <https://github.com/topics/neuroimaging> (accessed 8.2.20), (Hanke and Halchenko, 2011; Jahn, 2020) and <https://www.nitrc.org/>. However, as this process typically relies on image contrast between tissues and prior anatomical knowledge, most methods are optimized for anatomically normal or close to normal brains without major morphologic aberrations. Consequently, brain parcellation remains problematic in the presence of structural brain pathology ((Weiller et al., 1995); (Nachev et al., 2008); (Solodkin et al., 2010); (Ledig et al., 2015)). It is important to realize that brain pathologies are highly heterogeneous, often requiring tailored solutions depending on the nature of the pathology. For example, diffuse pathologies such as Alzheimer's disease can be addressed using more specific or age appropriate priors (Fillmore et al., 2015), as it causes global atrophic changes without distorting the general shape of the brain.

In contrast, for focal lesions, the appropriate solution depends on the type of affected tissue and size of the lesion. For example, small white matter lesions typically only have a minor effect on brain mapping but can bias subsequent analyses such as voxel-based morphometry (VBM) (Guo et al., 2019). Such an impact can be mitigated using lesion filling where the lesioned voxels are replaced with intensity values consistent with neighboring unaffected voxels (Chard et al., 2010); (Battaglini et al., 2012); (Magon et al., 2014); (Popescu et al., 2014); (Griffanti et al., 2016); (Prados et al., 2016); (Guo et al., 2019); (Schmidt et al., 2019)). Larger gray matter lesions (e.g., large focal cortical dysplasia) cannot be addressed using lesion filling unless this is constrained to replicate only healthy gray matter intensities and preserve cortical shape. Pathologies affecting different brain tissues, i.e., not confined to gray matter, white matter or a cerebrospinal fluid (CSF) compartment such as brain tumors, stroke, tumefactive multiple sclerosis and others are more difficult to tackle. Moreover, the presence of pathological mass effect, and perilesional edema adds to this difficulty. Such pathology can potentially degrade the performance of many steps in a brain mapping pipeline, e.g. skull stripping ((Iglesias et al., 2011); (Lutkenhoff et al., 2014); (Isensee et al., 2019)), making it difficult and even impractical to use automated whole-brain structural mapping pipelines for such patients. Freesurfer (Fischl, 2012) recon-all (FreeSurferWiki, 2020a) failure rates can be as high as 30% (Reid et al.,

2016) in moderately pathological cerebral palsy brain images, and even higher in case of brain tumors (Zheng et al., 2017).

A number of methods have been proposed to address the problem of brain mapping in the presence of pathology. However, to date there is no consensus on which are the most optimal, nor is a solution available that can be broadly applied to different types of lesions. For example, in case of small or subtle lesions Freesurfer (Fischl, 2012) recon-all (FreeSurferWiki, 2020a) and similar pipelines may still work. Larger or more obvious lesions may benefit from a cost function masking (CFM) approach (Brett et al., 2001); (Andersen et al., 2010)). This means that the lesion is masked out of the input brain image, and the registration is limited to the non-lesioned brain tissue. CFM has been reported to improve results in some cases ((Kim et al., 2007); (Andersen et al., 2010)), but may also result in a lower quality parcellation in case of large lesions (Nachev et al., 2008). Others resort to functional mapping using resting-state functional MRI to parcellate all healthy gray matter (Wang et al., 2015), or use task-based fMRI for a partial mapping (Reid et al., 2016). Few studies have exclusively investigated this issue in different pathological cohorts ((Weiller et al., 1995); (Nachev et al., 2008); (Solodkin et al., 2010); (Ledig et al., 2015)). Besides the CFM approach, only two of these resulted in dedicated solutions that were made available non-commercially, namely enantiomorphic normalization (Nachev et al., 2008) <http://www.bcblab.com/BCB/normalization.html> and MALP-EM (Ledig et al., 2015) <https://github.com/ledigchr/MALPEM>.

2. Aims of the study

First, we propose a new workflow to reliably segment/label MR images in the presence of a broad spectrum of focal brain pathologies, whether extra-axial or intra-axial, including large, bilateral, heterogeneous lesions with and without mass effect. In doing so, we provide a solution that facilitates state-of-the-art MRI analyses in clinical populations. Subsequently, we test and evaluate its performance in a sample of clinical data and a sample of realistic synthetic data.

3. Material and methods

3.1. Proposed workflow

We present a fully automated open-source image processing pipeline called "Virtual brain grafting" (VBG). Briefly, VBG encompasses two approaches for intra-axial lesions, both of which generate a synthetic template brain image. The first approach is intended for unilateral lesions (uVBG), it generates a donor brain using the native non-lesioned hemisphere and one hemisphere from the synthetic template brain image. The second approach is intended for bilateral lesions (bVBG) here the synthetic template brain image is used directly as the donor brain. Both approaches use the same synthetic template brain, 2 mm full-width half-maximum (FWHM) 3D Gaussian smoothed masks to fill the lesion, and a two-step lesion-filling.

A third approach, inspired by the recent work of Hou et al. (Hou et al., 2020), is included for extra-axial lesions, where the lesion is zero-filled, though this is not based on virtual grafting. VBG uses various open-source Unix-based image processing programs in a fully automated pipeline, including ANTs (Avants et al., 2011a) v2.3.0, FSL (Jenkinson et al., 2012) v6.0, MRtrix3 (Tourneris et al., 2019) v3.0_RC3, and Freesurfer (Fischl, 2012) v6.0. However, the steps outlined below may be executed with different programs, VBG as described here is fully automated, currently available for use without the need for any closed-source software. Required inputs are a 3D T1-weighted image and a binary integer format lesion mask covering the entire pathology in the same space, which can be obtained using any manual or semi-automated segmentation tool-box (McCarthy, 2020; Yushkevich et al., 2006). This mask should cover the entire lesion and include perilesional edema if any.

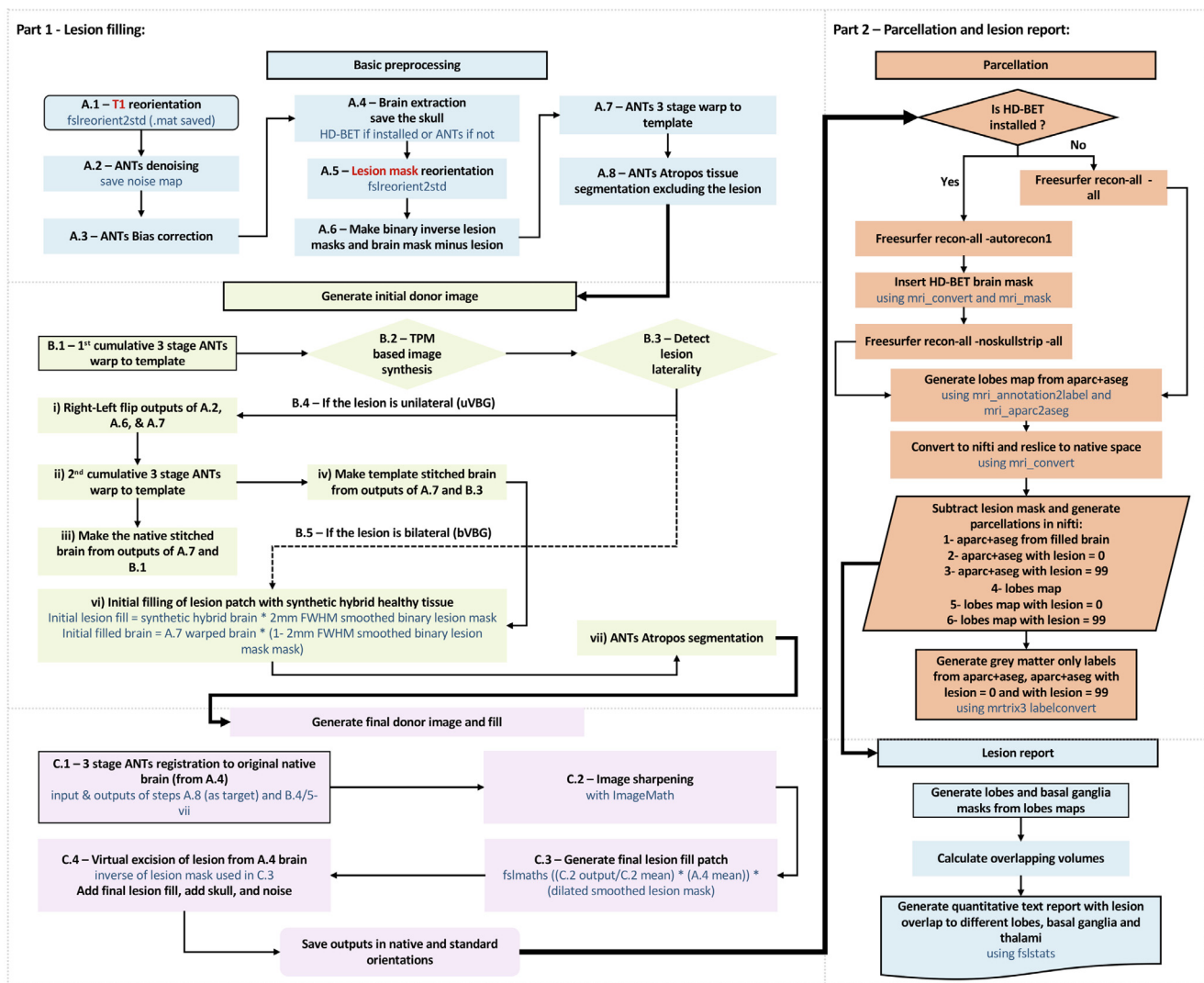


Fig. 1. Schematic representation of VBG: Part 1 generates the lesion-free image. It starts in the top left with basic processing and Atropos segmentation (light blue), donor image generation and initial filling (yellow), then final lesion filling (pink). Part 2 is concerned with parcellation (orange). Lastly a text report is generated detailing lesion overlap with various labelled brain structures.

VBG can be split into two main parts, i.e. (1) lesion filling and (2) whole-brain parcellation of the lesion-free output image using Freesurfer (Fischl, 2012) recon-all (FreeSurferWiki, 2020a). The different steps are explained broadly below, illustrated in Fig. 1 and with additional detail provided in the supplementary material.

3.1.1. Part 1, lesion filling; this generates the lesion-free T1-weighted images through the following 3 stages

A. Basic preprocessing. This stage applies basic preprocessing to the input T1-weighted image and lesion mask. It starts with image reorientation in FSLpy (McCarthy et al., 2020), denoising (Manjón et al., 2010) and bias correction (Tustison et al., 2010) and denoising (Manjón et al., 2010) in ANTs (Avants et al., 2011a), and brain extraction using HD-BET (Isensee et al., 2019) or ANTs (Avants et al., 2011a). The brain is then warped using cost-function masking in ANTs (Avants et al., 2011a) to the VBG template brain in MNI space. Finally, ANTs (Avants et al., 2011a) Atropos (Avants et al., 2011b) is applied for segmenting the warped brain while excluding the lesion using the brain mask with the lesion subtracted.

B. Initial donor image generation. The second stage involves flipping the brain along the right-left axis, iterative deformation to match the tem-

plate brain using ANTs (Avants et al., 2011a), synthesizing an initial donor image from the inverse warped template and tissue probability maps (TPMs), and selecting the appropriate pipeline for unilateral or bilateral lesions.

TPM-based T1-weighted image synthesis (supplementary figure 1) uses the inverse warped template, prior TPMs, subject's brain image and TPMs with the lesion excluded. First the input brain images are normalized by a two-pass mean division, and the template (source) TPMs are binarized at a lower threshold of 0.1 and used as masks to isolate each tissue from the source image. The target TPMs are binarized at 95% probability and used to calculate target tissue specific mean intensity with mrstats (Tournier et al., 2019). Each source tissue map is multiplied by the corresponding target tissue mean, and a forced correction of CSF signal intensity is used to scale its maximum to 0.2 of the gray matter mean signal intensity. Finally, all tissues are combined using scalar addition with ImageMath "addtozero" (Avants et al., 2011a).

Automated lesion laterality detection (supplementary figure 2) sets the workflow to be followed for the rest of the script. It uses the lesion mask in MNI space, and binary masks of the template's right and left hemispheres.

If the lesion is unilateral, the lesioned hemisphere is replaced with a synthetic hemisphere retaining the non-lesioned native hemisphere. The

resulting image is called a stitched brain. An initial filled brain image is generated by replacing the lesion with healthy tissue from this stitched brain. Next, the stitched brain is further deformed to match the initial filled one, then segmented into different tissues with ANTs (Avants et al., 2011a) Atropos (Avants et al., 2011b). If the lesion is bilateral the synthetic brain is directly used to derive the initial filled brain, which is then segmented with Atropos (Avants et al., 2011b).

C. Final donor image generation and fill. This constitutes the last stage in lesion filling. Here the initial filled brain is warped back to native space and sharpened with ANTs (Avants et al., 2011a) to create the final donor brain image. The lesion replacement graft is harvested using a 2 mm FWHM smoothed mask and inserted into the recipient image, similar to (Chard et al., 2010). The skull and noise are added back to generate a realistic lesion-free whole head T1-weighted image. Finally, the initial transformation is reversed to generate the image in native orientation. Whole head, brain extracted, and brain mask images are saved in original and standard orientations.

3.1.2. Part 2, brain parcellation

This involves running recon-all (FreeSurferWiki, 2020a) on the output lesion-free image in native orientation. First recon-all (FreeSurferWiki, 2020a) is run until the brain extraction stage, then the VBG generated brain mask is applied to the recon-all (FreeSurferWiki, 2020a) results, after which it is restarted and run to the end. Finally, parcellation maps are generated with an additional lesion label, and with a zero-filled lesion.

3.2. VBG testing and evaluation

3.2.1. General outline

Dice similarity coefficient (DSC) calculations represent one of the mainstays for quantitative analyses of segmentation accuracy in the presence of a ground truth. In the case of brain pathology, a manual delineation is typically used as the ground truth. However, for the whole brain this process would be highly time consuming and was not feasible for the current work. Thus, we resorted to two approaches: first, we processed and parcellated T1-weighted images from 10 clinical patient participants (real-patients) with gliomas, with and without the proposed method. We included patients with gliomatous lesions of different sizes and locations. A group of healthy control (HC) volunteers ($N = 10$) was also included. We generated a synthetic cohort ($N = 200$), consisting of two groups. First, a lesion-free group ($N = 100$), which was created by non-linear deformation of the HC images to match the mass effect of the patients, referred to henceforth as the synthetic-mass-effect group. Second, a synthetic-patients group ($N = 100$) were generated containing both the mass effect and the lesions. All synthetic-mass-effect images were parcellated with recon-all (FreeSurferWiki, 2020a) and the synthetic-patients images were parcellated after VBG filling. We also attempted to parcellate the real-patients and synthetic-patients' images without VBG. We assessed the parcellation accuracy of all volumes-of-interest (VOIs) outside the lesion patch both qualitatively and quantitatively. Fig. 2 shows a schematic of both the image processing steps and the evaluation procedures.

All completed non-VBG parcellations (HC, real-patients, and synthetic-mass-effect), real-patients' non-VBG and uVBG, synthetic-patients non-VBG and corresponding synthetic-patients' uVBG parcellations were qualitatively evaluated (Section 3.4.1.). For the synthetic-mass-effect parcellations this was used to confirm acceptable quality prior to further analyses. None of these were excluded. Consequently, all completed synthetic-patients' parcellations (non-VBG, uVBG and bVBG) could be quantitatively evaluated with DSC using the corresponding synthetic-mass-effect parcellations as the ground truth (Section 3.4.2.).

Statistical tests used the results of both evaluation approaches to investigate the accuracy of VBG-driven parcellations. Exploratory analyses

using linear regression were also used to investigate the effects of the lesions on each VOI's DSC in the VBG-driven parcellations (Section 3.3.4). Fig. 3 shows representative slices from the patients and uVBG-generated lesion-free output.

3.2.1.1. Participants. This study included ten adult surgery naïve patients with gliomas (3 females, 7 males, age range 19 – 61 years, median 36.5), and 10 adult healthy participants (8 females, 2 males, age range 18 – 66 years, median 35) who were free of any neurological disease. This study was approved by the local research ethics committee at UZ/KU Leuven, Leuven, Belgium, study number S61759, and conducted in accordance with the Declaration of Helsinki. A written informed consent was acquired from each participant prior to scanning. Table 1 lists lesion features, and MR acquisition.

3.2.1.2. Medical image acquisition. MRI acquisition used two 3-Tesla scanners (Achieva dStream and Ingenia-X, Philips Medical Systems, Best, The Netherlands), both with 32 channel phased array receive head coils. The following images were acquired for the patients: sagittal 3D Turbo-Field Echo T1-weighted images pre and post intravenous gadolinium injection (TR/TE/FA: 9 ms/4.1 ms/8°, voxel size: $0.9 \times 0.88 \times 0.88$ mm, matrix: $200 \times 288 \times 288$), sagittal 3D Turbo-Spin-Echo (TSE) T2-weighted images (TR/TE/FA: 3000 ms/280 ms/90°, voxel size: $0.9 \times 0.98 \times 0.98$ mm, matrix: $220 \times 256 \times 256$), and sagittal 3D TSE T2 fluid attenuation inversion recovery (FLAIR)-weighted images, (TR/TE/TI/FA: 2800 ms/340 ms/1650 ms/90°, voxel size: $1.0 \times 0.98 \times 0.98$ mm, matrix: $200 \times 256 \times 256$). We used all available modalities for lesion segmentation and only the pre-contrast T1-weighted images for the rest of this study. The healthy participants were scanned on the same scanners with the same non-contrast enhanced T1-weighted scan but reconstructed at $0.6 \times 0.6 \times 0.6$ mm voxel size and matrix size of $228 \times 384 \times 384$. One patient (PAT004) was included but had only post-contrast T1-weighted images acquired with the same protocol as the healthy participants.

3.3. Image processing

3.3.1. Lesion segmentation

Lesion masks were generated by a neuro-radiologist (AR) in ITK-snap (Yushkevich et al., 2006) v3.8.0 in Mac OSX 10.13.6. We followed the protocol described in (Yushkevich et al., 2019) for multimodality semi-automated lesion segmentation. However, we aggregated the different lesion tissue components into a single binary mask, which is required for VBG. The remainder of the work used the original unprocessed pre-contrast T1-weighted images. The pathological mass effect of the real patients was subjectively rated (by AR) into none, mild, moderate, and severe for descriptive purposes.

3.3.2. Initial VBG application

Following lesion segmentation, the uVBG approach was applied to generate lesion-free T1 brain images from the real-patients' data. None were excluded upon visual inspection. All images were included with their original resolution since VBG is designed to accommodate varying spatial resolutions. The lesion-free images generated here were used to create the synthetic cohort.

3.3.3. Synthetic cohort creation

3.3.3.1. The synthetic-mass-effect group. This was generated using ANTs (Avants et al., 2011a) nonlinear warping of each HC T1 brain image to each real-patient's lesion-free uVBG output image, then applying the generated transforms and warps to the whole head T1-weighted images resulting in 100 synthetic subjects. All 10 HCs were registered to each real-patients image mimicking the pathological mass effect of each patient in the 10 HCs. Resulting images were used as the synthetic ground truth for quantitative evaluation. We hypothesized that in the absence of a focal pathology, FreeSurfer (Fischl, 2012) recon-all

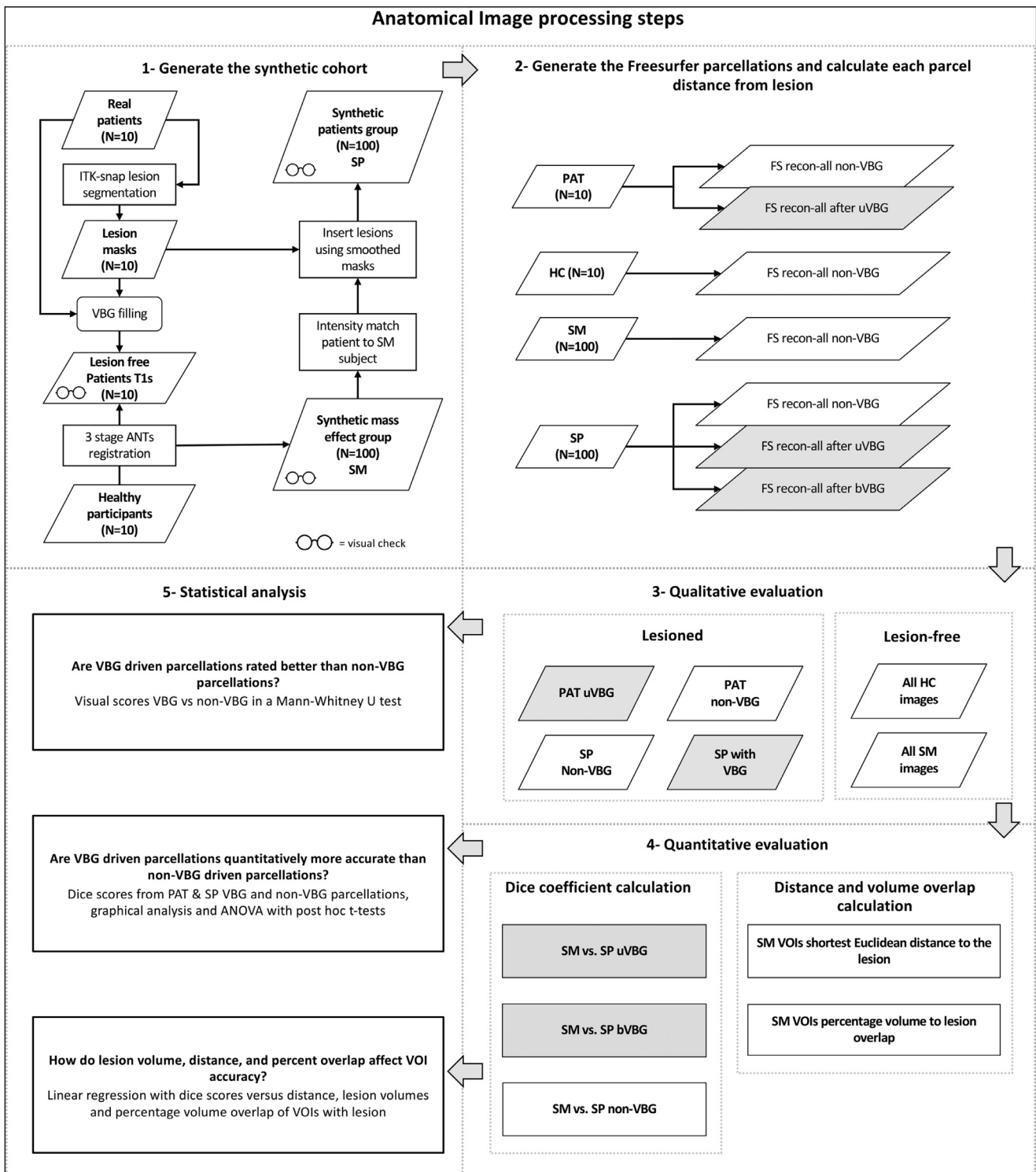


Fig. 2. A schematic representation of the VBG validation process. Top: image processing steps applied to generate the Freesurfer parcellations used for evaluation, 1–2. Bottom: qualitative and quantitative evaluation and statistical testing, 3–5. (PAT = real patients, HC = healthy controls, SM = synthetic-mass-effect group, SP = synthetic-patients group, VBG = virtual brain grafting, uVBG = unilateral VBG, bVBG = bilateral VBG), VBG derived datasets are highlighted in gray.

(FreeSurferWiki, 2020a) would be able to run without failure and accurately represent the mimicked pathological mass effect in the parcellations.

3.3.3.2. The synthetic-patients group. These images were generated as follows: first the intensity histogram of each real patient’s original T1 brain image was matched to the target synthetic-mass-effect image with

mrhistmatch (Tournier et al., 2019). The lesion patch and edema were isolated from the patient’s image and inserted into the synthetic-mass-effect brain using a smoothed lesion mask (2 mm FWHM 3D gaussian kernel) to avoid a sharp interface with the recipient image. This expanded our test group to 100 simulated patients, accurately represent-

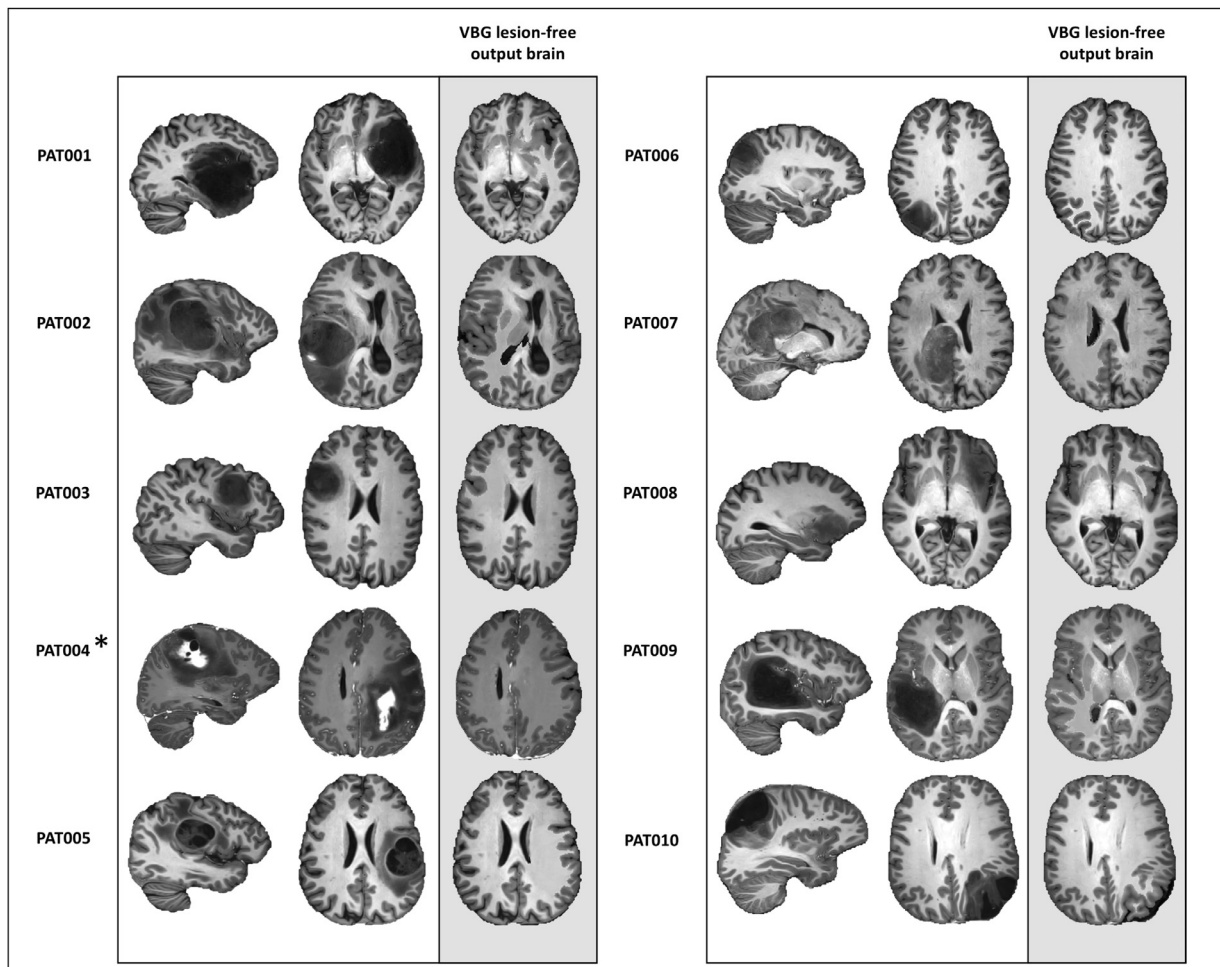


Fig. 3. Representative sagittal and axial slices from each patient and the results of uVBG lesion filling demonstrated in axial. Asterisk indicates post-contrast T1-weighted image used as input. (VBG = virtual brain grafting, PAT = real patients).

Table 1
Lesion features and MR acquisition.

Patient ID	Lesion location	Lesion grade (WHO)	Lesion + edema mask volume in cm ³	Mass effect severity	MR images acquired
PAT001	Left temporal	Low grade	109.75	Moderate	T1 / T2 / FLAIR / T1 +c
PAT002	Right temporo-parietal	High grade	190.48	Severe	T1 / T2 / FLAIR / T1+C
PAT003	Right frontal	Low grade	38.98	None	T1 / T2 / FLAIR / T1+C
PAT004	Left fronto-parietal	High grade	171.71	Moderate	T2 / FLAIR / T1+C
PAT005	Left fronto-parietal	High grade	70.41	Mild	T1 / T2 / FLAIR / T1+C
PAT006	Right parietal	Low grade	21.41	None	T1 / T2 / FLAIR / T1+C
PAT007	Right midline parietal	High grade	77.44	Moderate	T1 / T2 / FLAIR / T1+C
PAT008	Left frontal	Low grade	17.51	None	T1 / T2 / FLAIR / T1+C
PAT009	Right temporal	Low grade	101.56	Mild	T1 / T2 / FLAIR / T1+C
PAT010	Left parieto-occipital	High grade	70.35	Mild	T1 / T2 / FLAIR / T1+C

PAT = real patients, cm³ = cubic centimeter, FLAIR = fluid attenuation inversion recovery, +C = contrast enhanced, WHO = world health organization, WHO grades I&II = low-grade glioma, and WHO grades III&IV = high-grade glioma.

ing each patient's pathology in each of the 10 healthy volunteers (*supplementary figure 3*).

3.3.4. Parcellation

All VBG lesion filling and recon-all ([FreeSurferWiki, 2020a](#)) parcellations were done on a dedicated compute node of the Vlaams (Flemish) Supercomputer Center (VSC) with two Intel Xeon Gold 6240 2.6 GHz CPUs, 36 cores in CentOS 7.8.2003. The HC ($N = 10$) and synthetic-mass-effect ($N = 100$) data were parcellated using recon-all ([FreeSurferWiki, 2020a](#)) with an HD-BET ([Isensee et al., 2019](#)) brain mask inserted, as done in VBG. The real-patients ($N = 10$) and synthetic-

patients ($N = 100$) data were parcellated with uVBG and attempted after only zero-filling the lesion patch similar to CFM, and also without VBG filling (non-VBG). bVBG was applied to a subset of synthetic-patients ($N = 25$), since there were no true bilateral lesions in our sample. Non-VBG real-patients and synthetic-patients parcellations were used as control parcellations to compare recon-all ([FreeSurferWiki, 2020a](#)) success/failure rates and parcellation quality with and without VBG. A total of 455 recon-all ([FreeSurferWiki, 2020a](#)) analyses were attempted, each with a runtime cap of 8 h using GNU timeout. Those that quit with an error or exceeded the timeout duration were considered to have failed, and all parcellations were allowed two attempts in case of failure. We

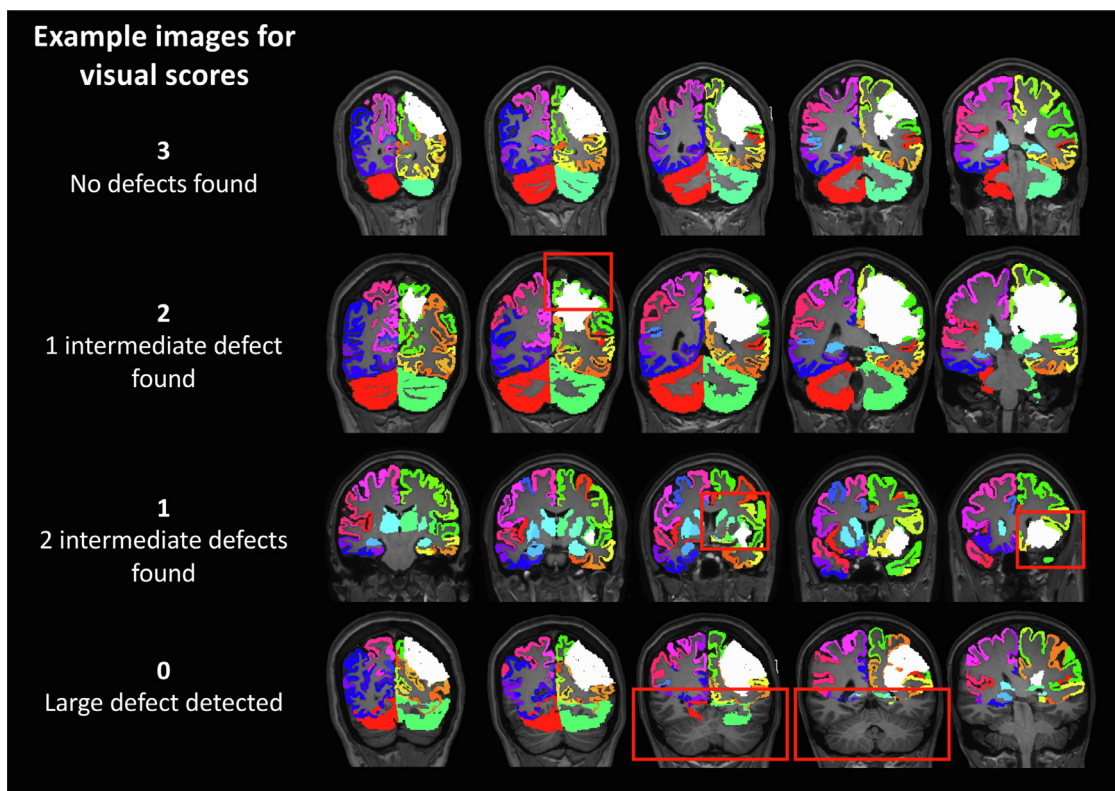


Fig. 4. Representative coronal slices for each qualitative score (Red boxes indicate identified defects). Score 3 indicates the best quality and score 0 indicates the worst.

imposed the 8-hour time limit after an initial test of non-VBG recon-all (FreeSurferWiki, 2020a) on the first 18 synthetic-patients' images using 2 cores per subject continued for 24 h with less than 50% completion rate, while VBG driven recon-all tended to finish in under 7 h.

3.4. Evaluation

3.4.1. Qualitative evaluation

Two independent experienced neuro-radiologists (NS & SS) visually evaluated the parcellations and assigned a quality score. We attempted to minimize rater bias, particularly for the evaluation of lesioned brain parcellations. More specifically, our main concern was blinding the raters to whether the parcellation came from VBG or not, and whether the source was real or synthetic data. To this end we resorted to (a) coding and mixing datasets together, and (b) standardizing the parcellations by using the lesion free T1-weighted images as underlays and subtracting the lesion patch from all parcellations except the HC group, which we used to imply a gold standard. This allowed us to blind our expert raters to the source image of all parcellations of interest. Images were given to the raters as high-resolution multi-frame panels in axial, coronal and sagittal with 10 mm interslice gap, generated using fsleyes render (McCarthy, 2020). Example coronal slices from each qualitative score are shown in Fig. 4.

The qualitative evaluation protocol we used can be described as follows:

A defect (error) may be minor, intermediate or major. A minor defect was defined as an unlabeled cluster of voxels (e.g., 10) belonging to gray matter, or non-brain tissue e.g., dura labelled as gray matter, or a more subtle focal underestimation of the cortical ribbon thickness. An intermediate defect/error was a larger falsely labelled or unlabeled area on the scale of the inferior frontal gyrus pars orbitalis, or anterior temporal pole. Finally, a major defect was any defect on the level of a whole gyrus or larger. We defined four categories for the quality of a parcella-

tion: 3 – Good, up to 3 minor unconnected defects, which can happen with structurally normal data (FreeSurferWiki, 2020b; Guenette et al., 2018; Klein et al., 2017, 2005; McCarthy et al., 2015). 2 – Acceptable, up to 5 minor, unconnected, or 1 intermediate scale defect/error. 1 – Fair, between 5 and 7 unconnected minor defects, or 2 intermediate unconnected defects. 0 – Poor, > 7 unconnected minor defects, > 2 intermediate defects or a major defect.

Experts rated all completed non-VBG synthetic-patients, real-patients parcellations, corresponding uVBG driven ones, all the synthetic mass-effect, and HC parcellations. We also added a group of a-posteriori warped HC parcellations (HC*), where each HC parcellation was warped to a patient using ANTs (Avants et al., 2011a) and nearest-neighbor interpolation.

3.4.2. Quantitative evaluation

We relied on DSC as a quantitative measure of parcellation quality. For reference, DSC is a measure of spatial similarity between two binary images, ranging between 0 and 1, with 0 meaning no similarity and 1 meaning perfect agreement. It can be expressed as follows (if X represents the synthetic-mass-effect VOI and Y the synthetic-patient VOI):

$$DSC = \frac{2 * (X \cap Y)}{(X \cup Y + X \cap Y)}$$

In Matlab r2018a we calculated the center of mass (CoM) of each label in the synthetic-mass-effect parcellations, then the shortest Euclidean distance from each VOI's CoM to the lesion mask. This is referred to simply as "distance". Lastly, we calculated the percent of volume by which each VOI overlaps with the lesion mask, referred to for simplicity as "percent overlap", another factor we hypothesized would influence accuracy.

Table 2

Parcellation outcomes, inclusions for visual QC, DSC calculations, and inter-rater reliability results.

Parcellations	Lesioned data (PAT + SP)			Lesion-free data (HC + SM)
	Non-VBG	uVBG	bVBG	Non-VBG
Attempted	10 PAT + 100 SP	10 PAT + 100 SP	25 SP	10 HC + 100 SM
Completed	6 PAT + 48 SP	10 PAT + 100 SP	25 SP	all
Failed 1st attempt	6 PAT + 56 SP (zF)	1 PAT*	none	none
Failed 2nd attempt	4 PAT + 52 SP (non-VBG)	none	none	none
Visual QC per expert	6 PAT (common) + 34 SP (24 unique)	10 PAT (common) + 34 SP corresponding to non-VBG SP (24 unique and 10 common)	none	10 HC (common) + 10 HC*(common) + 61 SM (39 unique + 22 common)
DSC	48 SP non-VBG vs. corresponding SM	100 SP vs. corresponding SM	25 SP vs. corresponding SM	100 SM (as ground truth)
Gwet's AC2 (Γ) scores	PAT non-VBG = 0.910	PAT uVBG = 0.932 - SP uVBG = 0.849	-	All = 0.967 - SM only = 0.969 - HC only = 0.988 - HC* only = 0.936

SP = synthetic patient, PAT = real patient, SM = synthetic mass effect group, HC = healthy control, HC* = α -posteriori warped HC parcellations, uVBG = unilateral VBG, bVBG = bilateral VBG, DSC = dice similarity coefficient, uVBG* = uVBG corresponding to completed non-VBG parcellations, ANOVA = analysis of variance, df = degrees of freedom, Std dev = standard deviation, Stat = statistic.

3.5. Statistical analyses

All statistical testing was done in Matlab r2018a, results were considered significant at $P \leq .05$. For inter-observer agreement we compared the visual scores of the parcellations evaluated by both experts using Gwet's AC2 (gamma) (Gwet, 2016) in mReliability tools (Girard, 2016). To ensure the validity of the synthetic-mass-effect group parcellations we first compared their visual scores to those of the HCs plus the α -posteriori warped HC parcellations using an unpaired Mann-Whitney U (MWU) test. Secondly, we asked "Are VBG driven parcellations qualitatively rated higher than non-VBG parcellations?" To answer this question, we compared the ratings of the real-patients and synthetic-patients non-VBG driven parcellations to corresponding unilateral VBG driven parcellations using a paired Mann-Whitney U test.

The remaining analyses focused on the synthetic-patients quantitative evaluation results. Here, we asked "Do VBG driven parcellations have higher dice scores than non-VBG driven parcellations?" All dice values from the three parcellation types were plotted for visual comparison. Next, we compared the mean dice values from the three parcellation methods. First, we compared completed non-VBG parcellations to the corresponding uVBG ($N = 48$ non-VBG and $N = 48$ uVBG) and bVBG ($N = 25$ non-VBG and $N = 25$ bVBG) ones separately using paired two-sample t-tests. Then the common parcellations from the three approaches in a one-way ANOVA and post hoc paired t-tests ($N = 20$)

Additionally, we explored the effect of the lesion on uVBG and bVBG DSC with Spearman correlations between average DSC per subject and the lesion volume. Each 10 synthetic-patients with the same lesion were assigned the lesion volume as calculated from their source real-patient. Next, we applied linear regression to investigate the effect of different lesion properties on uVBG and bVBG DSC, namely lesion volume, distance of the VOI from the lesion in mm, and percent overlap between the lesion and the VOI. We hypothesized that larger lesions would result in a global reduction of DSC scores, and that VOIs closer to the lesion and/or with higher lesion overlap would have a lower DSC score than those further away and/or without overlap. The parameters explaining most of the variance were used in a stepwise multiple linear regression model, with the pooled VOI DSC scores from all subjects in the synthetic-patients group as dependent variable.

4. Results

First, the parcellations derived from the synthetic-mass-effect group were comparable in visual quality scores (88 scored 3, 11 scored 2,

and 1 scored 1) to the HC group and α -posteriori warped HC parcellations (15 scored 3, and 5 scored 2). A two-tailed MWU test comparing both groups without pairing indicated no significant differences, $U(100,20) = 1127.5$, $z = 1.486$, $P = .137$. Secondly, Gwet's AC2 (Gwet, 2016) was used to measure inter-observer reliability. This showed a good overall agreement between both experts using all common parcellations (gamma = 0.943, observed agreement = 0.960, chance agreement = 0.318). Thirdly, we report on the results of parcellation, qualitative and quantitative evaluations, as well as the paired statistical comparison of parcellation methods and exploratory analyses (see supplementary information for results of the unpaired comparison). Table 2 lists the attempted, completed and failed parcellations, inclusions for visual QC, and DSC calculations, as well as Gwet's AC2 results for inter-rater reliability.

We included the results from the second attempt (non-VBG, $N = 48/52$ completed/failed, grand average DSC = 0.675, and standard deviation = 0.281) due to higher DSC values and more completed parcellations (see supplementary information for results of the first attempt). All synthetic-patients uVBG parcellations succeeded on the first attempt. Only 1/10 Patient's uVBG parcellation failed initially, but completed on the second attempt, compared to 6/10 failed after zero-filling and 4/10 failed non-VBG patients parcellations. We used the minimum of the visual scores assigned for the common parcellations by both experts, and the remaining unique scores from each expert.

Qualitative evaluation showed real-patients uVBG parcellations scored significantly higher than corresponding non-VBG ones, $U(6,6) = 33$, $z = 2.738$, $P < .01$. uVBG also outperformed non-VBG parcellations in the synthetic-patients group. This was confirmed on a paired MWU-test comparing corresponding parcellations from both methods $U(48,48) = 2076$, $z = 7.336$, $P < .001$.

Quantitative analysis also confirmed this impression using the synthetic-patients average DSC scores, uVBG out-scored non-VBG in the paired two-sample t-test ($N = 48$ each, t-stat = 3.011, $P < .01$, t-critical two-tailed = 2.684). One-way ANOVA and post hoc paired t-tests comparing the common parcellations from the three approaches (uVBG, bVBG and non-VBG, $N = 20$ each) showed significant differences between the three groups, with both VBG approaches scoring significantly higher than non-VBG. Table 3 lists the visual scores summary statistics, and paired DSC comparison results.

Fig. 5 Shows plotted visual scores from all synthetic-patients non-VBG and uVBG parcellations plus the average DSC scores from the synthetic-patients common to the three parcellation approaches. Unpaired analysis results are detailed in supplementary information.

Table 3
Visual scores summary statistics and paired DSC comparison results.

(A) Visual scores summary statistics							
Parcellation	Count	Completed/failed	Visual QC	Score 0 (worst)	Score 1	Score 2	Score 3 (best)
uVGB PAT	10	10/0	10	0	1	1	8
non-VGB PAT	10	6/4	6	0	1	4	1
uVGB SP	100	100/0	48	0	9	15	76
bVGB SP	25	25/0	none	-	-	-	-
Non-VGB SP	100	48/52	48	18	10	15	7
Non-VGB SM	100	100/0	100	0	1	12	87
Non-VGB HCs	10	10/0	10	0	0	1	9
Non-VGB HCs*	10	10/0	10	0	0	4	6

(B) Paired DSC comparison results (matching the three approaches)				
Summary statistics				
Groups	Count	Average	Std dev	
uVGB	20	0.894	0.020	
bVGB	20	0.903	0.017	
Non-VGB	20	0.617	0.286	

ANOVA: Single factor				
Source of variation	df	F	P-value	F critical
Between groups	2	19.188	< 0.0001	3.158
Within groups	57			
Total	59			

Post hoc t-tests: Paired two-sample assuming unequal variances				
	df	t stat	P (two-tail)	t critical (two-tail)
uVGB vs non-VGB	19	2.871	< 0.01	2.093
bVGB vs non-VGB	19	3.008	< 0.01	2.093

SP = synthetic patient, PAT = patient participant, SM = synthetic mass effect group, HCs = healthy control participants, HCs* = α -posterior warped HCs, uVGB = unilateral VGB, bVGB = bilateral VGB, DSC = dice similarity coefficient, uVGB* = uVGB corresponding to completed non-VGB parcellations, ANOVA = analysis of variance, df = degrees of freedom, Std dev = standard deviation, Stat = statistic.

Table 4
Statistical results exploring the relation between DSC and lesion characteristics.

	Statistical test	Independent variables	Results (significance accepted at $P < .05$)				
uVGB	SC	LeV	$r_s = -0.862$		$P < .001$		
			Adjusted R²	Beta	P value	RMS Error	
		SLR	Log ₁₀ D	0.467	0.130	< 0.001	0.064
		SLR	%ov	0.542	-0.004	< 0.001	0.059
		MLR	Log ₁₀ D	0.576	Log ₁₀ D: 0.050	< 0.001	0.057
		LeV	LeV: -4.6e-05				
		%ov	%ov: -0.003				
bVGB	SC	LeV	$r_s = -0.894$		$P < .001$		
			Adjusted R²	Beta	P value	RMS Error	
		SLR	Log ₁₀ D	0.409	0.120	< 0.001	0.064
		SLR	%ov	0.598	-0.004	< 0.001	0.060
		MLR	Log ₁₀ D	0.605	Log ₁₀ D: 0.02	< 0.001	0.060
				LeV	LeV: -7.0e-05		
		%ov	%ov: -0.004				

uVGB = unilateral VGB, bVGB = bilateral VGB, RMS = root mean squared, LeV = lesion volume in milliliters, Log_{10D} = log₁₀ shortest Euclidean distance in millimeters, %ov = percent label to lesion volume overlap, SC = spearman correlation, SLR = simple linear regression, MLR = multiple linear regression.

We found a strong logarithmic relationship between distance and DSC, thus log₁₀ scaled distance measure was used for exploratory analyses. Results from the statistical tests used to explore the lesion's influence on DSC are listed in Table 4.

Briefly, we found a strong significant inverse association using Spearman correlation (uVGB ($N = 100$): $r_s = -0.863$, $P < .001$; bVGB ($N = 25$): $r_s = -0.89$, $P < .001$), between the average dice scores per subject and lesion volumes. Simple linear regression (SLR) revealed a significant inverse relation between the DSC per VOI and log₁₀ scale distance, and a significant inverse relation to percent overlap. Results of the SLR tests for uVGB and bVGB dice scores versus log₁₀ distance are shown in Fig. 6, along with those of the non-VGB parcellations for comparison. In a mul-

tiply linear regression model, these variables together explained 58% of the variance in dice scores. Results using the shortest Euclidean distance between each VOI CoM and Hausdorff distances as measures of similarity are provided in the supplementary information.

5. Discussion

The first aim of this work was to propose and explain VGB as a workflow for heterogeneous brain lesion filling and optional subsequent structural mapping using Freesurfer (Fischl, 2012) recon-all (FreeSurferWiki, 2020a). Our second aim for this study was to test and evaluate the quality and accuracy of the VGB driven whole brain parcel-

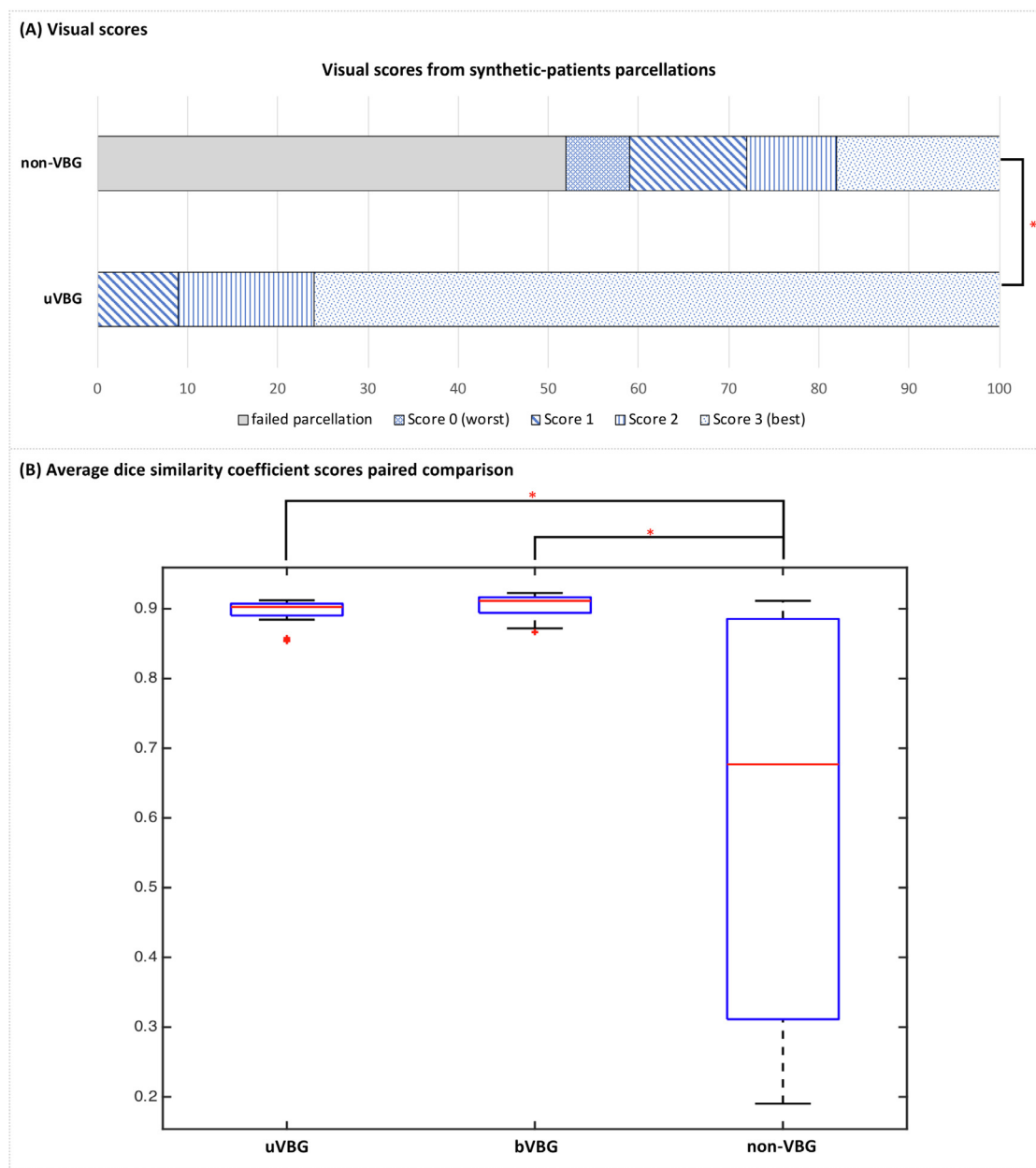


Fig. 5. (A) shows synthetic-patients (SP) visual scores for uVBG & non-VBG parcellations. (B) shows a box-plot comparison of grand average dice scores from all VOIs outside the lesion mask of all subjects common to the three parcellation approaches used (uVBG, bVBG and non-VBG, $N = 20$ each), asterisks indicate statistical significance at $P < 0.05$ (< 0.01), error bars indicate standard error. (VBG = virtual brain grafting, uVBG = unilateral VBG, bVBG = bilateral VBG).

lations. We chose a test sample of preoperative patients with gliomas for this work as they provide a variety of lesion sizes, locations, and mass effect.

Our evaluation shows a significant benefit from using VBG in this group, both qualitatively and quantitatively in the real and synthetic cohorts. Put simply, VBG allows an accurate parcellation for patients where recon-all (FreeSurferWiki, 2020a) would otherwise fail to complete. Our exploratory analysis partially explained the variation in VOI dice values in the VBG driven parcellations of the synthetic cohort. In intuitive terms, our analysis showed that VOIs lost 0.004 DSC for every 100 mL increase in lesion, 0.05 DSC was gained for a 10-fold increase in distance to the lesion, and 0.003 DSC is lost for every 1% label volume lost to overlap with the lesion. However, there is an inherent multicollinearity between the three parameters as in the case of larger lesions

there is less probability for VOIs to be more distant, and more chance for overlap with the lesion mask.

Only 1 of 135 VBG recon-all (FreeSurferWiki, 2020a) runs failed, a real patient's postcontrast T1 weighted image that was appropriately filled by VBG but lagged in the automated topographical error correction (FreeSurferWiki, 2020c) of recon-all (FreeSurferWiki, 2020a). We hypothesized that the cause was the gadolinium signal confounding cortical surface morphology. This was confirmed upon inspecting recon-all (FreeSurferWiki, 2020a) inflated surfaces, which showed major defects in the vicinity of the bright blood vessels (supplementary figure 4). Thus, we repeated recon-all (FreeSurferWiki, 2020a) without a time-limit. This finished without error after 18 h and the dataset was included. Such cases would be better addressed using another modality to assist with the cortical surface reconstruction, e.g., T2 or FLAIR. Alternatively, a different surface reconstruction approach, e.g. FastSurfer

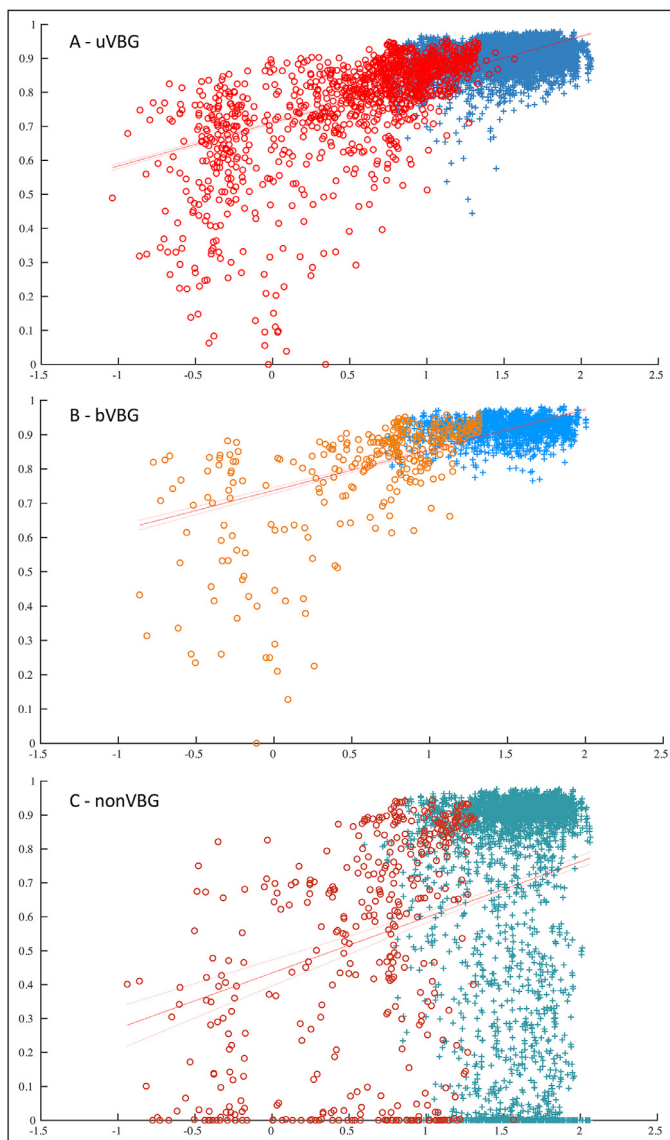


Fig. 6. Results of simple linear regression for VOI DSC on the (Y – axis) against the log10 distance between each VOI and lesion mask in mm (X – axis). Red line and parallel dotted range represent the intercept and 95% confidence interval. (+) represents VOIs overlapping with lesion mask by 1% or less and (o) represents VOIs overlapping with lesion mask by 1% or more. Shown in navy and red for uVBG (A), blue and orange for bVBG (B), and teal and maroon for nonVBG (C) (VBG = virtual brain grafting, uVBG = unilateral VBG, bVBG = bilateral VBG).

(Henschel et al., 2020) could be used, which may be adopted in the next version of VBG as this also promises to significantly shorten runtime.

Adequate lesion filling must conform to the healthy brain tissue morphology and image intensity pattern. Furthermore, sharp interfaces between the lesion fill graft and surrounding recipient brain image must be avoided. VBG achieves this by generating a synthetic donor image that matches the input brain image to ensure the filling tissue does not carry concomitant pathologies or parts of the main lesion from the relatively non-lesioned side. In bilateral lesions, there is no healthy hemisphere to rely on, and the unilateral approach would simply duplicate the whole lesion into the filling patch (*supplementary figure 5*). Thus, the synthetic images are used directly for the initial and final lesion filling. Here, we did not observe a significant difference in performance between unilateral and bilateral approaches, as quantified with mean DSC.

VBG can be used to achieve an adequate whole brain parcellation in the presence of a wide variety of lesions, both unilateral and bilateral, as opposed to previously described lesion filling-based approaches, which were more suited for smaller unilateral lesions. For example, enantiomorphic normalization (Nachev et al., 2008) relies on the native healthy hemisphere for patching the lesion prior to normalizing to the MNI152 template brain. This is available non-commercially (<http://www.bcblab.com/BCB/normalization.html>) but reportedly performs best on unilateral lesions with no significant midline shift (Solodkin et al., 2010) and aims only to achieve a better normalization. Another approach called virtual brain transplantation was described in 2010 (Solodkin et al., 2010). This also relied on the native healthy hemisphere for lesion filling, and evaluated its impact on the accuracy of recon-all (FreeSurferWiki, 2020a) parcellations. The authors reported encouraging results in a sample of patients with relatively large stroke lesions. However, to our knowledge this method has not yet been made available on any open-source platforms.

MALP-EM (Ledig et al., 2015) has shown encouraging results in a sample of patients with traumatic brain injury (Ledig et al., 2015) and was validated in patients with ventriculomegaly (Carass et al., 2017). In contrast to the previously mentioned methods and to VBG, MALP-EM (Ledig et al., 2015) does not need a lesion mask, and does not rely on lesion filling prior to segmentation. It uses multi-atlas matching and expectation maximization for volumetric segmentation. Incorporating MALP-EM (Ledig et al., 2015) with VBG could be mutually beneficial for the results of both approaches, and ultimately improve the accuracy of structural brain mapping. VBG initially focuses on FreeSurfer (Fischl, 2012) as this enables the standardization of analyses in pathological data to match state-of-the-art methods used in normal data e.g., CIFTIFY (Dickie et al., 2019) and the HCP processing pipeline (Glasser et al., 2013). Alternatively, the lesion-free output of VBG could also be used to improve results of normalization or other structural mapping pipelines such as MALP-EM (Ledig et al., 2015), ANTs MALF (Wang et al., 2011), CAT12 (Gaser and Dahnke, 2012), CARET (Van Essen et al., 2013), connectome workbench (Marcus et al., 2011) or for lesion-symptom mapping (Foulon et al., 2018).

VBG can benefit from a reduction in runtime, and optimizing workflow details such as image synthesis, which can occasionally expand some structures within the filling patch in very large lesions (>150 ml), e.g., insula, putamen (*supplementary figure 6*). The gray - white matter interface may also appear sharper than normal. We tested the bilateral approach (bVBG) in a subset of synthetic data only, as none of the included patients had truly bilateral lesions, only midline shift due to mass effect, or slight contralateral extension through white matter structures (e.g., PAT004). Future validation studies would benefit from a larger sample with more varying lesion and types of pathology, e.g., stroke, cerebral palsy, and cortical dysplasia. As well as the use of more advanced statistical tests employing nonlinear models and/or linear mixed models for evaluating parcellation performance. Further development will adopt full BIDS (Gorgolewski et al., 2016) compliance, release a VBG Docker (Merkel, 2014) image, fine-tune donor image synthesis, reduce run-time, and potentially include different MR modalities, e.g. T2 and FLAIR, from which a T1 image could be synthesized.

Future studies addressing the issue of structural mapping of brain images with large lesions could rely on deep learning solutions, for example a convolutional neural network (CNN) e.g. DeepMedic (Kamnitsas et al., 2017) may be used to segment the lesion, a generative adversarial network can generate a lesion-free version of the input brain, and the last CNN would segment the remaining healthy tissue, e.g. FastSurfer (Henschel et al., 2020), or DeepNat (Wachinger et al., 2017). In this context VBG could benefit deep learning solutions by providing the required ground truth and training data, after sufficient quality checking. This would ideally be undertaken in a multidisciplinary collaborative setup and use different MR modalities from different scanners for maximum yield.

Conclusions

In this work we have proposed VBG, a new workflow for reliable structural mapping of T1 images using FreeSurfer (Fischl, 2012) in the presence of heterogeneous, and large pathologies. Our testing and evaluation show it to be a reliable approach in this group of patients with heterogeneous gliomas, and similarly in the synthetic testing dataset. To conclude, VBG achieves a realistic lesion filling and enables an accurate whole brain parcellation as tested with FreeSurfer (Fischl, 2012). It is an open-source workflow, currently requiring no GPU processing, and is available under a Mozilla 2.0 license, via (https://github.com/KUL-Radneuron/KUL_VBG).

Author contributions

Ahmed M. Radwan: Conceptualization, Methodology, Software, Data curation, Investigation, Visualization, Writing-Original draft preparation, Writing - Review & Editing

Louise Emsell: Writing-Original draft preparation, Critical revision, Visualization, Statistical analysis, Writing - Review & Editing

Jeroen Blommaert: Methodology, Writing-Original draft preparation, Visualization, Statistical analysis, Writing - Review & Editing

Andrey Zhyhka: Testing, Critical revision.

Silvia Kovacs: Data acquisition, Data curation, Testing, Critical revision.

Tom Theys: Testing, Critical revision, Writing-Original draft preparation, Supervision.

Nico Sollmann: Testing, Visualization, Qualitative assessment, Critical revision, Writing - Review & Editing.

Patrick Dupont: Methodology, Statistical analysis, Critical revision, Visualization, Supervision.

Stefan Sunaert: Conceptualization, Methodology, Qualitative assessment, Critical revision, Data acquisition and curation, Writing - Review & Editing, Supervision.

Acknowledgements

The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government – department EWI. The authors also thank Prof. Josien Pluim, Prof. Steven De Vleeschouwer, Dr. Laura Michiels for their support, as well as the healthy volunteers and patients for their participation. LE is supported by the research foundation Flanders (FWO, Grant No. GOC0319N) and KU Leuven Sequoia Fund. JB is an aspirant researcher for the FWO (Grant No. 11B9919N).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2021.117731](https://doi.org/10.1016/j.neuroimage.2021.117731).

References

Andersen, S.M., Rapcsak, S.Z., Beeson, P.M., 2010. Cost function masking during normalization of brains with focal lesions: still a necessity? *Neuroimage* 53, 78–84. doi:10.1016/j.neuroimage.2010.06.003.

Ashburner, J., Trujillo-Barreto, N., Brett, M., Büchel, C., Flandin, O.D.G., Friston, K., Gitelman, D., Glaser, D., Glauche, V., Harrison, L., Henson, R., Holmes, A., Kiebel, S., Kilner, J., Mattout, J., Nichols, T., Penny, W., Phillips, C., Poline, J.-B., Stephan, K., 2006. Statistical parametric mapping: the analysis of functional brain images, 1st Edition. ed.

Avants, B., Epstein, C., Grossman, M., Gee, J., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi:10.1016/j.media.2007.06.004.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., Gee, C., 2011a. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi:10.1016/j.neuroimage.2010.09.025.A.

Avants, B.B., Tustison, N.J., Wu, J., Cook, P.A., Gee, J.C., 2011b. An open source multivariate framework for N-tissue segmentation with evaluation on public data. *Neuroinformatics* 9, 381–400. doi:10.1007/s12021-011-9109-y.

Battaglini, M., Jenkinson, M., De Stefano, N., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum. Brain Mapp.* 33, 2062–2071. doi:10.1002/hbm.21344.

Brett, M., Leff, A.P., Rorden, C., Ashburner, J., 2001. Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage* 14, 486–500. doi:10.1006/nimg.2001.0845.

Carass, A., Shao, M., Li, X., Dewey, B.E., Blitz, A.M., Roy, S., Pham, D.L., Prince, J.L., Ellingsen, L.M., 2017. Whole brain parcellation with pathology: validation on ventriculomegaly patients. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 20–28. doi:10.1007/978-3-319-67434-6_3.

Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A.M., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *J. Magn. Reson. Imaging* 32, 223–228. doi:10.1002/jmri.22214.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi:10.1006/cbmr.1996.0014.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.

Dickie, D.A., Shenkin, S.D., Anblagan, D., Lee, J., Cabeza, M.B., Rodriguez, D., Boardman, J.P., Waldman, A., Job, D.E., Wardlaw, J.M., 2017. Whole brain magnetic resonance image atlases: a systematic review of existing atlases and caveats for use in population imaging. *Front. Neuroinf.* doi:10.3389/fninf.2017.00001.

Dickie, E.W., Anticevic, A., Smith, D.E., Coalson, T.S., Manogaran, M., Calarco, N., Viviano, J.D., Glasser, M.F., Van Essen, D.C., Voineskos, A.N., 2019. Ciftify: a framework for surface-based analysis of legacy MR acquisitions. *Neuroimage* 197, 818–826. doi:10.1016/j.neuroimage.2019.04.078.

Fillmore, P.T., Phillips-Meek, M.C., Richards, J.E., 2015. Age-specific MRI brain and head templates for healthy adults from 20 through 89 years of age. *Front. Aging Neurosci.* 7, 44. doi:10.3389/fnagi.2015.00044.

Fischl, B., 2012. FreeSurfer. *Neuroimage* 62, 774–781. doi:10.1016/j.neuroimage.2012.01.021.

Foulon, C., Cerliani, L., Kinkingnéhun, S., Levy, R., Rosso, C., Urbanski, M., Volle, E., Thiebaut de Schotten, M., 2018. Advanced lesion symptom mapping analyses and implementation as BCBtoolkit. *Gigascience* 7, 1–17. doi:10.1093/gigascience/gyi004.

Frazier, J.A., Chiu, S., Breeze, J.L., Makris, N., Lange, N., Kennedy, D.N., Herbert, M.R., Bent, E.K., Koneru, V.K., Dieterich, M.E., Hodge, S.M., Rauch, S.L., Grant, P.E., Cohen, B.M., Seidman, L.J., Caviness, V.S., Biederman, J., 2005. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am. J. Psychiatry* 162, 1256–1265. doi:10.1176/appi.ajp.162.7.1256.

FreeSurferWiki, 2020a. FreeSurferMethodsCitation - Free Surfer Wiki [WWW Document]. URL <https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferMethodsCitation> (accessed 8.2.20).

FreeSurferWiki, 2020b. FsTutorial/TroubleshootingDataV6.0 - Free Surfer Wiki [WWW Document]. URL <https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingDataV6.0> (accessed 7.19.20).

FreeSurferWiki, 2020c. mris_fix_topology - Free Surfer Wiki [WWW Document]. URL https://surfer.nmr.mgh.harvard.edu/fswiki/mris_fix_topology#Author.2BAC8-s (accessed 7.30.20).

Gaser, C., Dahnke, R., 2012. CAT - A computational anatomy toolbox for the analysis of structural MRI data. *HBM Conf.* 32, 7743.

Girard, J.M., 2016. mReliability: Reliability Analysis in MATLAB [WWW Document]. Github.com. URL <https://github.com/jmgirard/mReliability> (accessed 8.19.20).

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal pre-processing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

Goldstein, J.M., Seidman, L.J., Makris, N., Ahern, T., O'Brien, L.M., Caviness, V.S., Kennedy, D.N., Faraone, S.V., Tsuang, M.T., O'Brien, L.M., Caviness, V.S., Kennedy, D.N., Faraone, S.V., Tsuang, M.T., 2007. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol. Psychiatry* 61, 935–945. doi:10.1016/j.biopsych.2006.06.027.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3, 1–9. doi:10.1038/sdata.2016.44.

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., Schulz, U.G., Kuker, W., Battaglini, M., Rothwell, P.M., Jenkinson, M., 2016. BIANCA (Brain AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. *Neuroimage* 141, 191–205.

Guenette, J.P., Stern, R.A., Tripodis, Y., Chua, A.S., Schultz, V., Sydnor, V.J., Somes, N., Karmacharya, S., Lepage, C., Wrobel, P., Allosco, M.L., Martin, B.M., Chaisson, C.E., Coleman, M.J., Lin, A.P., Pasternak, O., Makris, N., Shenton, M.E., Koerte, I.K., 2018. Automated versus manual segmentation of brain region volumes in former football players. *NeuroImage: Clin.* 18, 888–896. doi:10.1016/j.nicl.2018.03.026.

Guo, C., Ferreira, D., Fink, K., Westman, E., Granberg, T., 2019. Repeatability and reproducibility of FreeSurfer, FSL-SIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. *Eur. Radiol.* 29, 1355–1364. doi:10.1007/s00330-018-5710-x.

Gwet, K.L., 2016. Testing the difference of correlated agreement coefficients for statistical significance. *Educ. Psychol. Meas.* 76, 609–637. doi:10.1177/0013164415596420.

- Hanke, M., Halchenko, Y.O., 2011. Neuroscience runs on GNU/Linux. *Front. Neuroinf.* 5, 8. doi:10.3389/fninf.2011.00008.
- Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M., 2020. FastSurfer - A fast and accurate deep learning based neuroimaging pipeline. *Neuroimage* 219, 117012. doi:10.1016/j.neuroimage.2020.117012.
- Hou, X., Yang, D., Li, D., Liu, M., Zhou, Y., Shi, M., 2020. A new simple brain segmentation method for extracerebral intracranial tumors. *PLoS ONE* 15. doi:10.1371/journal.pone.0230754.
- Iglesias, J.E., Liu, C.Y., Thompson, P.M., Tu, Z., 2011. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans. Med. Imaging* 30, 1617–1634. doi:10.1109/TMI.2011.2138152.
- Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., Bendzus, M., Maier-Hein, K.H., Kickingereder, P., 2019. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* 40, 4952–4964. doi:10.1002/hbm.24750.
- Jahn, A., 2020. Welcome to Andy's Brain Book! — Andy's Brain Book 1.0 documentation [WWW Document]. URL <https://andysbrainbook.readthedocs.io/en/latest/> (accessed 8.2.20).
- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015.
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi:10.1016/j.media.2016.10.004.
- Kim, J., Avants, B., Patel, S., Whyte, J., 2007. Spatial normalization of injured brains for neuroimaging research: An illustrative introduction of available options. *NCRRN methodology papers - mri.org*.
- Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., Keshavan, A., 2017. Mindboggling morphometry of human brains. *PLoS Comput. Biol.* 13. doi:10.1371/journal.pcbi.1005350.
- Klein, A., Mensh, B., Ghosh, S., Tourville, J., Hirsch, J., 2005. Mindboggle: automated brain labeling with multiple atlases. *J. Clin. Neurosci.* 8, 5–7. doi:10.1186/1471-2342-5-7.
- Ledig, C., Heckemann, R.A., Hammers, A., Lopez, J.C., Newcombe, V.F.J., Makropoulos, A., Lötjönen, J., Menon, D.K., Rueckert, D., 2015. Robust whole-brain segmentation: application to traumatic brain injury. *Med. Image Anal.* 21, 40–58. doi:10.1016/j.media.2014.12.003.
- Lutkenhoff, E.S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J.D., Owen, A.M., Monti, M.M., 2014. Optimized brain extraction for pathological brains (optiBET). *PLoS ONE* 9, 1–13. doi:10.1371/journal.pone.0115551.
- Magon, S., Gaetano, L., Chakravarty, M.M., Lerch, J.P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., Sprenger, T., 2014. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci.* 15, 106. doi:10.1186/1471-2202-15-106.
- Makris, N., Goldstein, J.M., Kennedy, D., Hodge, S.M., Caviness, V.S., Faraone, S.V., Tsuang, M.T., Seidman, L.J., 2006. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr. Res.* 83, 155–171. doi:10.1016/j.schres.2005.11.020.
- Mandal, P.K., Mahajan, R., Dinov, I.D., 2012. Structural brain atlases: design, rationale, and applications in normal and pathological cohorts. *J. Alzheimer's Dis.* doi:10.3233/JAD-2012-120412.
- Manjón, J.V., Coupé, P., Martí-Bonmati, L., Collins, D.L., Robles, M., 2010. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J. Magn. Reson. Imaging* 31, 192–203. doi:10.1002/jmri.22003.
- Marcus, D.S., Harwell, J., Olsen, T., Hodge, M., Glasser, M.F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S.W., Van Essen, D.C., 2011. Informatics and data mining tools and strategies for the human connectome project. *Front. Neuroinform.* 5. doi:10.3389/fninf.2011.00004.
- Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* 2, 89–101. doi:10.1006/nimg.1995.1012.
- McCarthy, C.S., Ramprasad, A., Thompson, C., Botti, J.A., Coman, I.L., Kates, W.R., 2015. A comparison of FreeSurfer-generated data with and without manual intervention. *Front. Neurosci.* 9, 379. doi:10.3389/fnins.2015.00379.
- McCarthy, P., 2020. FSLeves. <https://doi.org/10.5281/ZENODO.3937147>
- McCarthy, P., Cottaar, M., Webster, M., Fitzgibbon, S., Craig, M., 2020. fslyp. <https://doi.org/10.5281/ZENODO.3890969>
- Merkel, D., 2014. Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* 2014 (239). [WWW Document]. online. URL <https://dl.acm.org/doi/10.5555/2600239.2600241>. (accessed 8.19.20).
- Moore, W., 2011. The Edwin Smith papyrus. *BMJ* 342, d1598. doi:10.1136/bmj.d1598.
- Nachev, P., Coulthard, E., Jäger, H.R., Kennard, C., Husain, M., 2008. Enantiomorphic normalization of focally lesioned brains. *Neuroimage* 39, 1215–1226. doi:10.1016/j.neuroimage.2007.10.002.
- Nowinski, W.L., 2016. Usefulness of brain atlases in neuroradiology: current status and future potential. *Neuroradiol.* J. 29, 260–268. doi:10.1177/1971400916648338.
- Popescu, V., Ran, N.C.G., Barkhof, F., Chard, D.T., Wheeler-Kingshott, C.A., Vrenken, H., 2014. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. *Neuroimage: Clin.* 4, 366–373. doi:10.1016/j.nicl.2014.01.004.
- Prados, F., Cardoso, M.J., Kanber, B., Ciccirelli, O., Kapoor, R., Gandini Wheeler-Kingshott, C.A.M., Ourselin, S., 2016. A multi-time-point modality-agnostic patch-based method for lesion filling in multiple sclerosis. *Neuroimage* 139, 376–384. doi:10.1016/j.neuroimage.2016.06.053.
- Reid, L.B., Cunnington, R., Boyd, R.N., Rose, S.E., 2016. Correction: surface-based fMRI-driven diffusion tractography in the presence of significant brain pathology: a study linking structure and function in cerebral palsy. *PLoS ONE* 11, e0162271. doi:10.1371/journal.pone.0162271.
- Schmidt, P., Pongratz, V., Küster, P., Meier, D., Wuerfel, J., Lukas, C., Bellenberg, B., Zipp, F., Groppa, S., Sämann, P.G., Weber, F., Gaser, C., Franke, T., Bussas, M., Kirschke, J., Zimmer, C., Hemmer, B., Mühlau, M., 2019. Automated segmentation of changes in FLAIR-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging. *NeuroImage: Clin.* 23, 101849. doi:10.1016/j.nicl.2019.101849.
- Solodkin, A., Hasson, U., Siugzdaitė, R., Schiel, M., Chen, E.E., Kotter, R., Small, S.L., 2010. Virtual brain transplantation (VBT): a method for accurate image registration and parcellation in large cortical stroke. *Arch. Ital. Biol.* 148, 219–241. doi:10.4449/aib.v148i3.1221.
- Standing, S., 2016. A brief history of topographical anatomy. *J. Anat.* 229, 32–62. doi:10.1111/joa.12473.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Thieme Medical Publishers, Inc., New York.
- Tournier, J.-D., Smith, R.E., Raffelt, D.A., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* 202, 116137. doi:10.1016/j.neuroimage.2019.116137.
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Yuanjie, Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi:10.1109/TMI.2010.2046908.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi:10.1006/nimg.2001.0978.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., Consortium, WU-Minn HCP, for the, W.-M.H., 2013. The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.
- Wachinger, C., Reuter, M., Klein, T., 2017. DeepNAT: deep convolutional neural network for segmenting neuroanatomy. *Neuroimage* 170, 434–445. doi:10.1016/j.neuroimage.2017.02.035.
- Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., Baker, J.T., Stufflebeam, S.M., Wang, K., Wang, X., Hong, B., Liu, H., 2015. Parcellating cortical functional networks in individuals. *Nat. Neurosci.* doi:10.1038/nn.4164.
- Wang, H., Das, S.R., Suh, J.W., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P.A., 2011. A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation. *Neuroimage* 55, 968–985. doi:10.1016/j.neuroimage.2011.01.006.
- Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Müller, S., Bier, D., Dutschka, K., Woods, R.P., Noth, J., Diener, H.C., 1995. Recovery from wernicke's aphasia: a positron emission tomographic study. *Ann. Neurol.* 37, 723–732. doi:10.1002/ana.410370605.
- Yushkevich, P.A., Pashchinskiy, A., Oguz, I., Mohan, S., Schmitt, J.E., Stein, J.M., Zukić, D., Vicory, J., McCormick, M., Yushkevich, N., Schwartz, N., Gao, Y., Gerig, G., 2019. User-Guided Segmentation of Multi-modality Medical Imaging Datasets with ITK-SNAP. *Neuroinformatics* 17, 83–102. doi:10.1007/s12021-018-9385-x.
- Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31, 1116–1128. doi:10.1016/j.neuroimage.2006.01.015.
- Zhang, F., Kahali, P., Suter, Y., Norton, I., Rigolo, L., Savadjiev, P., Song, Y., Rathi, Y., Cai, W., Wells, W.M., Golby, A.J., O'Donnell, L.J., 2017. Automated connectivity-based groupwise cortical atlas generation: application to data of neurosurgical patients with brain tumors for cortical parcellation prediction. In: *Proceedings - International Symposium on Biomedical Imaging*. IEEE Computer Society, pp. 774–777. doi:10.1109/ISBI.2017.7950633.