Published in final edited form as:

Neuroimage. 2014 October 15; 100: 444–459. doi:10.1016/j.neuroimage.2014.06.029.

An automatic framework for quantitative validation of voxel based morphometry measures of anatomical brain asymmetry

Antonietta Pepe^{a,*}, Ivo Dinov^{b,c}, and Jussi Tohka^a

^aDepartment of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland ^bLaboratory of Neuro Imaging (LONI), The Institute of Neuroimaging and Informatics, Keck School of Medicine, University of Southern California, 2001 North Soto Street, SSB1-102, Los Angeles, CA 90032, USA ^cUniversity of Michigan School of Nursing (UMSN), 400 North Ingalls, Ann Arbor, MI 48109-5482, USA

Abstract

The study of anatomical brain asymmetries has been a topic of great interest in the neuroimaging community in the past decades. However, the accuracy of brain asymmetry measurements has been rarely investigated. In this study, we propose a fully automatic methodology for the quantitative validation of brain tissue asymmetries as measured by Voxel Based Morphometry (VBM) from structural magnetic resonance (MR) images. Starting from a real MR image, the methodology generates simulated 3D MR images with a known and realistic pattern of inter-hemispheric asymmetry that models the left-occipital right-frontal *petalia* of a normal brain and the related rightward bending of the inter-hemispheric fissure. As an example, we generated a dataset of 64 simulated MR images and applied this dataset for the quantitative validation of optimized VBM measures of asymmetries in brain tissue composition. Our results suggested that VBM analysis strongly depended on the spatial normalization of the individual brain images, the selected template space, and the amount of spatial smoothing applied. The most accurate asymmetry detections were achieved by 9-degrees of freedom registration to the symmetrical template space with 4 to 8 mm spatial smoothing.

Keywords

Asymmetry; Magnetic resonance imaging; Gray matter; Quantitative validation; Spatial normalization; Spatial smoothing; Template space; VBM

Introduction

The human brain is anatomically asymmetrical (Toga and Thompson, 2003). During the last decades, thanks to the increased availability of high resolution structural three dimensional (3D) MR images, there has been a considerable number of in vivo studies investigating the anatomical inter-hemispheric brain asymmetry and its links to age (Kovalev et al., 2003),

gender (Kovalev et al., 2003), mental diseases (Pepe et al., 2013; Shenton et al., 2001), and asymmetrical behavioral traits, such as hand and foot preference (Amunts et al., 2000; Beaton, 1997; LeMay, 1977; Moffat et al., 1998), auditory perception (Keenan et al., 2001), and language production (Amunts et al., 1999; Binder, 2000; Dapretto and Bookheimer, 1999; Foundas et al., 1996; Geschwind and Galaburda, 1985).

The earliest in vivo studies of anatomical brain asymmetries from 3D MR images mainly focused on the manual delineation of a specific region-of-interest (ROI) and the analysis of differences in its size across hemispheres. These studies were limited by the burden and the subjectivity of manual ROI delineation and by the use of local and poorly sensitive traditional measures such as ROI's length. More recently, automatic morphometric techniques have been applied to study structural brain asymmetries (Good et al., 2001b; Hopkins et al., 2008; Luders et al., 2004; Pepe et al., 2013). Of particular relevance among them here are the voxel-based morphometry (VBM) (Ashburner and Friston, 2000; Wright et al., 1995) and other VBM techniques (Good et al., 2001a, 2001b; Luders et al., 2004; Mechelli et al., 2005; Sowell et al., 1999; Watkins et al., 2001).

In a VBM based study of neuroanatomical asymmetries, subjects' brain MR images are spatially normalized into a common reference space. Spatially normalized brain images are then partitioned into gray matter (GM), white matter (WM), and cerebro-spinal fluid (CSF) tissue classes, and the resulting brain tissue images are reflected with respect to their planes of symmetry. Next, measures of voxel-level asymmetry are derived from the differences between the original and left-right flipped brain tissue images. Voxel-wise hypothesis tests followed by multiple comparison correction are finally performed on smoothed (tissue-specific) difference images (maps) to establish the statistical significance of asymmetry measures and presented in the form of statistical parametric maps.

VBM analyses of brain asymmetries of normal controls have replicated previously well established postmortem and in vivo brain asymmetries (such as the left-occipital right-frontal brain *petalia* and the *planum temporale* asymmetry (Good et al., 2001b; Watkins et al., 2001)) as well as previously unreported asymmetry findings (such as a pattern of inter-hemispheric asymmetry in the insular cortex found in Watkins et al. (2001)). Even though there are encouraging VBM findings in partial agreement with the neuroanatomical literature (Good et al., 2002; Luders et al., 2004; Sowell et al., 1999) and VBM findings on brain asymmetry confirming existent knowledge on certain disease states and normal conditions (Luders et al., 2004), inconsistent findings have also been reported. For example, Heschl's gyrus asymmetry was observed in the VBM study of Good et al. (2001b) but not in Watkins et al. (2001). Interestingly, Heschl's gyrus and *planum temporale* asymmetries appeared not to be correlated with the hemispheric language dominance in Dorsaint-Pierre et al. (2006) and Keller et al. (2011). See Toga and Thompson (2003) or Jancke and Steinmetz (2003) for a review on brain asymmetry findings.

A crucial intrinsic assumption of VBM methods is that spatial normalization establishes the anatomical correspondence of brain structures at voxel-level while maintaining individual anatomical differences, and that voxel-level statistics can be used to verify specific hypotheses on the data (Good et al., 2001a,b; Salmond et al., 2002). The validity of VBM-

based inferences is affected by spatial normalization inaccuracies (Bookstein, 2001; Davatzikos, 2004; Mechelli et al., 2005; Senjem et al., 2005), the choice of the spatial normalization template and spatial normalization method (Good et al., 2001b; Mechelli et al., 2005; Shen et al., 2007), and by the amount of spatial smoothing applied (Good et al., 2001b; Mechelli et al., 2005). Related to this, the use of customized (tissue type and population specific) templates for spatial normalization, known as optimized VBM, is expected to produce more accurate VBM results (Good et al., 2001b; Mechelli et al., 2005; Shen et al., 2007). Although less obvious and less investigated in the literature, we hypothesize that normal patterns of brain asymmetry in controls might also cause mismatches in asymmetrical brain regions that can thus propagate to other parts of the brain.

Despite its wide use, the validation of VBM is still largely lacking and partially inconclusive due to the difficulties involved in the generation of large datasets of simulated images with a known and realistic inter-hemispheric asymmetry pattern, as well as of ground truth for the validation of automatic morphometric methods. This is true for the VBM in general and specifically for applications of VBM to brain asymmetry studies.

We present here an automatic framework for the quantitative validation of VBM-based measurements of brain asymmetries via construction and analysis of simulated 3D MR images. The main contributions of this study can be summarized as follows.

- i. We propose and implement a method to generate simulated 3D MR images with a known pattern of inter-hemispheric asymmetry based on real MR images via parametric modeling of brain asymmetry. The employed parametric model mimics two of the most consistently reported macroscopic patterns of brain asymmetry in normal human brains, namely the left-occipital right-frontal *petalia* and the related rightward bending of the inter-hemispheric fissure, also referred to as brain torque or Yakovlevian torque.
- ii. We generate a ground truth image of brain asymmetry values using the aforementioned parametric model. This ground truth image can be used, in conjunction with the simulated dataset in (i), for the quantitative validation of methods for the analysis of brain shape asymmetry.
- iii. The choices of spatial normalization, template space, and spatial smoothing might affect the results of VBM analysis (Bookstein, 2001; Davatzikos, 2004; Good et al., 2001a,b; Salmond et al., 2002; Senjem et al., 2005; Shen et al., 2007). We evaluate these effects by performing VBM analysis of asymmetry for a simulated dataset of 64 images and comparing the VBM results to the ground truth values. To the best of our knowledge, there exists no other quantitative validation study evaluating the accuracy of VBM measures of brain asymmetries. More generally, the literature regarding VBM validation studies is scarce and mainly focused on qualitative observations in clinical conditions (Keller et al., 2004; Senjem et al., 2005). As Senjem et al. (2005) recognize, the qualitative visual comparison of VBM results to the previously reported pathology of a certain disease is not the ideal method for comparing various VBM image processing algorithms.

- iv. The dataset in (i), the ground truth in (ii), and the whole automatic framework in (i)–(iii) are to be made freely available to promote further quantitative studies aiming at the validation of morphometric methods for the analysis of neuroanatomical asymmetries. This is important, e.g., to evaluate the accuracy and robustness of novel and existing morphometric methods of brain asymmetry, and possibly to clarify contradicting findings on anatomical brain asymmetries in normal and clinical conditions.

Materials and methods

Participants, MR image acquisition and pre-processing

64 healthy right-handed subjects (16 males, 48 females) in the adult lifespan (aged from 18 to 93; mean age = 39.4 years; SD = 18.8 years) were selected from the OASIS dataset, a larger and freely available sample which has been described in detail elsewhere (Marcus et al., 2007; <http://www.oasis-brains.org>). All participants were non-demented as assessed by clinical evaluations (Marcus et al., 2007) and by the Clinical Dementia Rating scale (Morris, 1993; Morris et al., 2001), and had no major gross anatomical abnormalities or known history of head trauma, stroke, use of psychoactive drugs, and neurological or psychiatric illness.

Enrolled subjects were imaged either 3 or 4 times within a single imaging session. Imaging was performed using a 1.5-T Vision scanner (Siemens, Erlangen, Germany) with a T1-weighted magnetization prepared rapid gradient-echo sequence (MP-RAGE) and the following image acquisition parameters: TR = 9.7 ms, TE = 4.0 ms, flip angle = 10°, TI = 20 ms, TD = 200 ms, matrix size = 128 × 256 × 256 voxels, voxel sizes = 1.25 × 1 × 1 mm³.

To correct for motion artifacts and thus enhance the signal to noise ratio, co-registered images of the individual scans in the native acquisition space were averaged, interpolated to isotropic voxel sizes (matrix size = 160 × 256 × 256 voxels, voxel size = 1 × 1 × 1 mm³), and provided for download by the OASIS distribution. These co-registered and averaged images were screened for artifacts and then used as the sample of this study.

Methods overview

This work introduces a novel and fully automatic framework for the quantitative validation of VBM and its applications for the analysis of inter-hemispheric asymmetries. First, based on each of the 64 brain MR images included in this study, we generated a simulated MR image with perfect inter-hemispheric anatomical symmetry modeling the extreme case of a brain with low or no inter-hemispheric asymmetry. Then, based on this simulated and symmetrical image, we generated a simulated MR image with a parametrically known pattern of inter-hemispheric asymmetry that mimicked the brain *petalia* and the Yakovlevian torque as often found in normal conditions. The workflow of dataset generation is presented in Fig. 1 and described in detail in the Synthetic dataset generation section. Second, an optimized VBM analysis of voxel-level inter-hemispheric asymmetries in brain tissue composition was performed, in both sets consisting of symmetrical and asymmetrical images (see an example in Fig. 3), and applied for the quantification of detected

asymmetries compared with ground truth values, and its relations to the spatial normalization scheme, template space, and amount of smoothing used. The applied VBM analysis is presented in detail in the VBM analysis section and overview of it can be seen in Fig. 3. For the sake of brevity, we consider only GM asymmetries.

Synthetic dataset generation

Generation of synthetic symmetrical MR images in native space—The images were automatically corrected for intensity non-uniformities (Manjón et al., 2007) and denoised (Manjón et al., 2010). Denoised images were then processed to generate simulated symmetrical images in native space by merging the right portion (voxels on the right hand side of the mid-sagittal plane) of the denoised images with their flipped versions with respect to the mid-sagittal plane (Pepe and Tohka, 2012), see Fig. 1. Next, Rician noise (Gudbjartsson and Patz, 1995) with realistic intensity was added to them so as to obtain simulated symmetrical images with non-symmetrical noise realizations, here denoted as $\tilde{s}$.

Generation of synthetic asymmetrical MR images in stereotaxic space—Each noisy simulated symmetrical image $\tilde{s}$ was spatially normalized to the MNI152 stereotaxic space using the FLIRT tool (Jenkinson and Smith, 2001; Jenkinson et al., 2002) of the FMRIB FSL Software Library (<http://www.fmrib.ox.ac.uk/fsl/>). Since the registration of skull stripped volumes in FLIRT is to be preferred, images $\tilde{s}$ were skull stripped using the BET (Brain Extraction Tool) tool (Smith, 2002) of FSL with the default parameter values (fractional intensity 0.5, vertical gradient in fractional intensity threshold 0). Next, the 7-Degrees of Freedom (DoF) affine transformation parameters T that spatially normalized the images into the MNI152 stereotaxic space were estimated using normalized correlation as a cost function, and were then applied to the corresponding non-skull stripped images $\tilde{s}$ using trilinear interpolation (see Fig. 1). The resulting spatially normalized images are here denoted as $T(\tilde{s})$. For the spatial normalization, customized subject-specific and skull stripped symmetrical template images in stereotaxic space were used to improve the spatial normalization accuracy, while the 7-DoF affine registration guaranteed that the symmetry and other shape related information were preserved. The customized subject-specific and skull stripped symmetrical template images were calculated (details not depicted in Fig. 1) by first computing the left–right flipped version across the mid-sagittal plane of each individual skull stripped image. Next, flipped and un-flipped versions were averaged. The resulting subject-specific averaged template images were spatially re-normalized to the stereotaxic space using a 7-DoF transformation to ascertain that the co-registration to the stereotaxic space was not lost due to the averaging process.

Once in the stereotaxic space, each simulated symmetrical (non-skull stripped) image $T(\tilde{s})$ was spatially deformed (see Fig. 1) to generate simulated non-skull stripped images $\tilde{a}$ with a known pattern of inter-hemispheric asymmetry. For this purpose, the Rescaled Adaptive Modified Global Linear Bending (ramGLB) parametric model introduced in Pepe and Tohka (2012) was used with the following model parameters: $k_1 = 0.00003$, $k_2 = -0.0001$, $y_A = -90$, $y_1 = -22$, $y_2 = 62$, $y_C = 91$, $z_1 = 74$, $z_2 = 76$, $z_{min} = 45$ and $z_{max} = 105$, $n_{max} = 10$, and n constant in blocks of 3 consecutive slices along the z axis (see Appendix A for more details). These parameters were found experimentally in Pepe and Tohka (2012) to provide, if

applied to MR images in the stereotaxic space, a realistic (in intensity, location, and direction) modeling of the left-occipital right-frontal *petalia* and the related rightward bending of the inter-hemispheric fissure as often found in the healthy human brain. In addition to this, the fact of the bending deformation was applied to the spatially normalized images in the stereotaxic space assured that the parametrically modeled inter-hemispheric asymmetry was consistently applied (in similar locations and with similar magnitude) across subjects despite of possible differences in image or voxel sizes or head positioning during the scanning (Pepe and Tohka, 2012).

Back to the native space—The aforementioned subject-specific 7-DoF affine registration transformations T were inverted (T^{-1}) and applied to the corresponding ramGLB deformed images $\tilde{a}$ to map them back into their respective native spaces (see Fig. 1). This was performed in FLIRT with trilinear interpolation. The resulting set $A = \{a\}$ of simulated asymmetrical images (in native space) can be used to evaluate the performances of VBM-methods for asymmetry detection.

To avoid systematic differences among the individual symmetrical images s (in the native space) and the asymmetrical images a (in the native space) due to the trilinear interpolation involved with the repeated spatial normalization of the latter, the spatially normalized simulated images $T(s)$ were mapped back to their corresponding native spaces using the same inverted 7-DoF affine registration transformations (T^{-1}) that was applied (in FLIRT with trilinear interpolation) to the asymmetrical images $\tilde{a}$ (see Fig. 1). The resulting set $S = \{s\}$ of symmetrical images in native space can be used to evaluate the performances of VBM-methods for asymmetry detection in the specific case of no asymmetry.

The simulated symmetrical s and asymmetrical a images displayed realistic noise realizations and had a known pattern of anatomical inter-hemispherical brain asymmetry: none in the former (see an example in Fig. 2b, and also supplementary Figs. S11 panels b, e and h), and as described by the ramGLB model in the latter (see an example in Fig. 2c, and supplementary Figs. S11 panels c, f and i). Due to the method design, the sets S and A retained most of the subject-specific features of brain anatomy and thus individual variability as in the original images. Moreover, since the simulated images were all obtained from images scanned with a similar device with identical imaging parameters, and since these images were screened for artifacts in image acquisition and reconstruction, possible effects deriving from these two potential confounding factors were excluded from the following VBM analysis. The simulated images are available for download at <https://sites.google.com/site/brainmorphorg/data-release>.

VBM analysis

Skull stripping—The skull, scalp and other non-brain tissues in the simulated image sets S and A were removed using the BET tool of FSL with default parameters. BET often slightly overestimates the brain/non-brain boundary, and it might thus retain some non-brain voxels (Smith, 2002; Wallis et al., 2006). Non-brain voxels remaining from the skull stripping might be mis-classified as brain tissues, and this could in turn affect the results of VBM analysis. For this reason, skull stripped images were first processed via iterative

morphological operations to disconnect and remove possible remaining non-brain voxels, and then visually screened for errors.

Brain tissue classification—The skull stripped brain images were partial volume voxel classified in the native space (Tohka, 2013; Tohka et al., 2004). As a result, GM, WM and CSF maps were estimated from each image. The primary benefit of using this method (Tohka, 2013; Tohka et al., 2004) was that it does not require prior knowledge on the spatial distribution of brain tissues in a standard reference space, thus avoiding the iterative segmentation–registration–segmentation work-flow of several existing VBM implementations.

Spatial normalization—Customized, tissue-type specific template images for the spatial normalization were generated, separately for both image sets *S* and *A*, as described next (details not depicted in Fig. 3). First, a symmetrical skull stripped template image in the stereotaxic space was calculated. This was obtained by averaging all the corrected skull stripped images of one set, symmetrizing them by averaging the template image obtained from one set with its left-right flipped versions across the mid-sagittal plane, and then spatially smoothing the resulting average symmetrical template image with the 8-mm FWHM Gaussian kernel. The skull stripped images were spatially normalized to the just calculated skull stripped whole brain template image using 12-DoF affine registration in FLIRT with trilinear interpolation and normalized correlation as a cost function. The estimated transformation parameters were then applied to the corresponding GM tissue maps to establish a coarse alignment. Once the aforementioned coarse alignment of the GM tissue maps was achieved, the tissue type-specific template images in the stereotaxic space were calculated, symmetrized as described above, and then spatially smoothed using an 8 mm FWHM Gaussian kernel. Next, the original GM images were spatially normalized using these template images, thus avoiding any contribution of non-brain voxels (Good et al., 2001a, 2001b). Notice that group- and tissue type-specific template images (only GM addressed here) were computed both in the stereotaxic space (asymmetrical) and symmetrized version of it.

The spatial normalization of each GM tissue image was performed in FSL (version 4.1.4), using 7, 9, and 12 DoF affine spatial normalization with trilinear interpolation and normalized correlation as a cost function (see Table 1), and in SPM (version 8, <http://www.fil.ion.ucl.ac.uk/spm/>) with trilinear interpolation, with and without modulation, and with both 0 or 16 non-linear iterations (see Table 1 for details). For each of these 7 distinct registration schemes, the spatial normalization was performed independently for both symmetrical and asymmetrical set-specific and tissue-specific template images, and for both image sets *A* and *S*. It should be noted at this point that the above spatial normalization routines did not attempt to match every cortical feature exactly, but merely corrected for global shape differences, and thus regional asymmetries were preserved (Good et al., 2001a, 2001b).

Asymmetry measure—For each spatially normalized GM map, a difference image *DI* was calculated replicating Luders et al. (2004)

$$DI=2(u-f)/(u+f) \quad (1)$$

where f and u denote the left–right flipped and un-flipped versions of the spatially normalized GM image, respectively. The image DI encoded the voxel-level inter-hemispheric asymmetry in an image u : positive values in the right hand-side of image DI indicated higher intensities in the right hemisphere compared with the left one; while positive values in the left hand-side of image DI indicated higher intensities in the left hemisphere compared with the right one (Luders et al., 2004).

Spatial smoothing—The difference images DI were smoothed using 4, 8, and 12 mm FWHM isotropic Gaussian kernels. This conditioned the residuals to conform more closely to the Gaussian random field model underlying the statistical process used for adjusting t-statistics values (Friston et al., 1995; Good et al., 2001b; Worsley et al., 1996). For a comparison, non-smoothed DI measures were also considered.

Statistical analysis—Voxel-level t-tests on the shape asymmetry indexes DI were calculated using the General Linear Model (GLM) estimation tool of SPM (Friston et al., 1995), and the consequent multiple comparisons problem was addressed using the intensity based Random Field theory (Worsley et al., 1996) at α - level = 0.05. This was performed separately for the 2 case studies (with the set S and the set A), and then repeated for each of the 7 spatial normalization schemes, each of the 2 study-specific template space selections (symmetrical and asymmetrical), and each of the 4 levels of spatial smoothing.

In principle, it would be possible to consider also the inter-group study between S and A . However, in between-groups settings, this would be confounded by the fact that both groups contained one simulated image per one original subject, thus being a poor model for between-group asymmetry studies. While there are obvious work-arounds for the problem, we decided not to consider between-group settings in this study.

Ground truth

Ground truth measures of VBM inter-hemispheric asymmetries were computed using the ramGLB parametric model with the same model parameters as used in the Synthetic dataset generation section. Each voxel of the ground truth corresponded to the displacement, expressed in millimeters, between the original location of that voxel to the new voxel location after that the bending deformation was applied to it (see Fig. 4).

Quantitative validation

Different quantitative validation measures are appropriate for the two sets (S and A) considered. Within the set S , images display no asymmetry, and therefore every detection of asymmetry is false. In addition to the number of voxels detected as asymmetrical (false positives), the variance and maximum of the uncorrected t-statistic values were computed to quantify the test statistic values rather than thresholded asymmetry detections. Within the set A , with images displaying a known asymmetry pattern, the sensitivity and specificity of detections were also computed. A voxel was defined to display true asymmetry if its displacement was over 1 mm in the ground truth. To study if the amount of true

displacement had an effect to the detection, other millimeter thresholds were considered as well. In addition to this, the Pearson's correlation coefficient between uncorrected voxel-wise (absolute) tstatistic values and the (absolute) true displacements in millimeters was computed to address the accuracy of the test statistics. Note that while the true displacements have dimensionality (millimeters), the tstatistic values are dimensionless and have a different interpretation than the true displacements. Although, too strong conclusions based on these correlation values should be avoided, this measure is a useful supplement to the sensitivity and specificity values that concern the thresholded statistic-maps. Notice also how the Pearson correlations, sensitivities, and specificities were computed over the whole brain rather than limiting the considerations to a pre-defined GM mask. This choice is motivated by the fact that there is a certain level of arbitrariness in the selection of a proper GM mask. This is especially true for considerations relating to the spatial smoothing where different levels of smoothing would require different GM masks, that would in turn complicate the interpretation of the results.

The whole framework (including the tools for simulated dataset and ground truth generation, VBM analysis, and quantitative VBM validation) will be made freely available for download at <https://sites.google.com/site/brainmorphorg/tools-release>.¹

Implementation and computational cost

The developed automatic framework runs on Linux, and it has been optimized to work for SGE based cluster grids. It integrates FSL's (version 4.1.4) and SPM's (version 8) spatial normalization scheme. Model estimation was performed using the GLM tool of SPM. All the other image processings were performed in Matlab (<http://www.mathworks.com/>).

The experiments described in this work were computationally costly. This was due to the long image processing pipeline required for the synthesis and analysis of the data, and to the number of images (64) and cases (56 study cases $\times$ 2 sets = 112 VBM analyses) being studied. The whole study took several CPU months to compute on the Cranium cluster (http://www.loni.usc.edu/about_loni/resources/ComputingResources.php), a grid engine that uses Rocks Cluster 4.5 and Sun Grid Engine 6.1u4 to accept, schedule, distribute and manage distributed execution of submitted jobs across a total of 1152 available processors (9 TB Ram, over 4500 slots) running CentOS release 4.5 as the operating system.

Results

Examples on individual DI maps (see Eq. (1)) and corrected group-wise asymmetry maps are presented in Fig. 5. Based on Fig. 5, it is clear that the methods which used asymmetrical template space were overly sensitive to the GM boundaries thus leading to a large number of false positives (see panels j and l which show the corrected t-statistics on the symmetrical set S , where there should be no detected asymmetries). This phenomenon can be seen already on the individual DI images (see panels b and d), where DI values in the set S were high along GM boundaries albeit the images were anatomically symmetric. The use of a symmetrical template space largely solved the problem, in the sense that the number of false

¹Source code will be made available shortly.

positives (see panels i and k) and individual DI values (see panels a and c) was greatly reduced while the true asymmetries were adequately detected (see panels m and o; note that the VBM analyses consider only GM asymmetries, and as such they are not designed to detect the whole asymmetry pattern). Clear differences between different registration techniques were evident based on Fig. 5. It is particularly interesting to note the different asymmetry patterns detected with the 9 DoF affine registration and non-linear registrations. All approaches produced false asymmetries near ventricles, which cannot be explained by the ‘signal spread’ originating from the regions of correctly detected asymmetries (in the frontal and occipital lobes). These results confirmed our initial hypothesis that normal patterns of brain asymmetry in controls, as the ones we observed in the proximity of the brain ventricles, might cause mismatches that in turn might affect the validity of VBM measures.

More qualitative results are shown in supplemental figures from Figs. S3 to S10, where the asymmetries detected by all 7 different approaches (using symmetrical templates) are depicted for each of the 4 levels of spatial smoothing applied in both sets *S* and *A*. Based on these figures, it can be observed that 7, 9, and 12 DoF affine registration to a symmetrical template space (denoted as FSL7S, FSL9S, and FSL12S, respectively) reached a good compromise between the specificity (relatively low number of false positive voxels) and the sensitivity (relatively high number of true detections). In figures from Figs. S3 to S10, it is difficult to pinpoint the differences between different registration schemes, which highlight the importance of the quantitative metrics introduced in the Quantitative validation section.

Quantitative results are presented from Figs. 6 to 10. Taking FSL9S with 4 mm smoothing as an example, the variance of (uncorrected) t-statistic values was 0.3042 (see Fig. 6) and the maximum t-statistic value was 14.1123 (see Fig. 7) with the set *S* where no asymmetry was present. As can be seen in Table 2, the multiple comparison correction effectively limited the number of false asymmetry detections.

With the set *A*, the Pearson's correlation between the t-statistic maps and the true displacement values was 0.3642 (see Fig. 8). This value might seem low, but note that t-statistic maps do not try to capture the magnitude of displacements, and that t-statistic computation was limited to GM while the correlation was computed across the whole brain. The asymmetries were detected with the specificity of 99.24% (see Fig. 9) and the sensitivity of 26.06% (see Fig. 10) with peak based multiple comparisons correction (the family-wise error was controlled at α - level = 0.05 using Random Field theory). As was the case with Pearson's correlations, the sensitivity for asymmetry detection might seem low, but recall that the asymmetry detection concerned only GM whereas ground truth was computed over the whole brain.

The results in Figs. 6 and 7 suggested that the use of asymmetrical templates should be avoided, especially when coupled with a high level of spatial smoothing. In more detail, for the set *S* and any fixed level of spatial smoothing and spatial normalization routine, the use of asymmetrical templates resulted consistently in higher variances (Fig. 6) and maximum values (Fig. 7) of the t-statistic values than the use of symmetrical templates. Since the same set of simulated and symmetrical images were spatially normalized into different template

spaces while leaving everything else unchanged, the excess variance and larger maximum of the t -statistics for the asymmetrical templates was due to the template space selection. The t -statistics should have close to zero value for the set S , since the images in this set feature no asymmetry. With the set A , the correlation coefficients between t -statistics and the ground truth displacements were higher when the symmetrical template was employed (see Fig. 8). The hypothesis test was clearly more specific and slightly less sensitive when a symmetrical template was used (see Figs. 9 and 10).

The variance of t -statistics increased with the width of the smoothing kernel for the image set S (see Fig. 6) and any fixed spatial normalization scheme and template space employed. The specificity of the detected asymmetries decreased with the width of the smoothing kernel in the set A (Fig. 9). As expected, the sensitivity of the detected asymmetries improved with the kernel width (Fig. 10). This leads to the expected conclusion that there was a trade-off between sensitivity and specificity in terms of kernel width, which can also be seen in Fig. 11 and supplemental figures from Figs. S3 to S10, where the detected asymmetries are visualized. The Pearson's correlation can be seen as a summary measure taking into account both specificity and sensitivity. Using the Pearson's correlation to mediate this trade-off showed that the optimum kernel width depended on the particular registration algorithm: considering now only symmetrical templates, the optimum filter width was 8 mm for all SPM-based methods while several FSL-based methods favored smaller kernel widths (see Fig. 8).

Interestingly, the specificity–sensitivity trade-off was also apparent with respect to the degrees of freedom and the spatial normalization algorithm employed (see Fig. 8). The best (highest) Pearson's correlation coefficients were obtained in SPM for nonlinear spatial normalization (SPMnS and SPMnSm) with any fixed selection of the smoothing kernel. A similar pattern of Pearson's correlation coefficients was also observed with FSL-based spatial normalization methods except for the fact that the 9-DoF registration (FSL9S) was slightly better than the 12-DoF registration (FSL12S) for any fixed selection of the smoothing kernel. This supports the widely held opinion that 9 DoF registrations are the best choice for asymmetry studies probably as the additional shears of the 12-DoF registration would reduce the positional asymmetries.

Related to this, it is of interest to notice that SPM and FSL's implementations of the spatial normalization are different, and this caused different performances of the two. As an example, it is of interest to compare the FSL12S to SPM1S as these methods differ only based on the used registration algorithm. FSL12S had smaller t -statistic variance and smaller maximum t -statistic value than SPM1S with the symmetrical set S . With the set A and 1 mm ground truth threshold, FSL12S was typically slightly more sensitive and specific for low levels of smoothing (0 or 4 mm) while for high levels of spatial smoothing (8 and 12 mm FWHM) SPM12S was more specific and FSL12S was more sensitive. Interestingly, with the higher ground truth thresholds and high levels of smoothing, SPM12S outperformed FSL12S also in terms of sensitivity (see Figs. 9 and 10). The modulation had only minor effects to SPM results, either with linear or non-linear registrations.

Discussion

Objective

This work introduced a novel and fully automatic framework for the quantitative validation of brain tissue asymmetries as measured by VBM. The framework is based on a parametric model for the inter-hemispheric bending of the human brain. This was used for generating simulated 3DMR images for which the brain asymmetry was known at the voxel level. Starting from a dataset of T1-weighted 3D MR images, two datasets of simulated images were generated. Due to the framework design, the obtained sets of simulated images retained most of the normal inter-subject variability and morphological features as present in the original images, and had a known pattern of inter-hemispheric asymmetry: none in the symmetrical images; and of intensity, direction, and location as described by the ramGLB model in the asymmetrical images. This framework was applied to 64 MR images of healthy right-handed subjects selected from the OASIS dataset. The resulting 64 simulated MR images were then used to validate VBM based analysis of whole brain asymmetry (only GM results presented). VBM analysis was performed separately for 7 different spatial normalization methods with different number of DoF, with both symmetrical and asymmetrical customized template images (2 cases for template space selection), and with progressively increasing kernel sizes for the spatial smoothing (0, 4, 8, and 12 mm FWHM: 4 cases for spatial smoothing). Altogether, 56 different cases were studied with each of the two sets of images. The developed automatic framework, along with the simulated MR images and the corresponding ground truth values of brain asymmetries in a stereotaxic space, were made freely available to promote further quantitative studies aiming at the validation of new and existing morphometric methods for the analysis of neuroanatomical asymmetries.

Simulated image construction

In our earlier work (Pepe and Tohka, 2012), we developed a realistic parametric model for the brain *petalia* and the rightward bending of the inter-hemispheric fissure that is an essential component of the current work. For example, in the current work, taking only the results concerning the set *S* (when there is no effect) into account would lead to the (incorrect) conclusion that the use of spatial smoothing would be detrimental to any of the VBM methods studied (see Figs. 6, 7, from Figs. S3 to S6, and Table 2). There exists instead a sensitivity–specificity trade-off with the level of spatial smoothing that is well visible in the results concerning set *A* (figures from Figs. 8 to 11, and from Figs. S7 to S10): larger filters led to increased sensitivity at a price of the decreased specificity.

As detailed in Appendix A, the simulated asymmetry pattern modeled in ramGLB is obtained with a combination of positional and rotational deformations of the image which mimicked normal patterns of brain *petalia* and related inter-hemispheric fissure bending as often found in normal subjects. Although positional and rotational asymmetries were the main components of the ramGLB model, a slight volumetric deformation was also added to the model to ensure realism of the simulated asymmetry, see Pepe and Tohka (2012) for more details. The ramGLB model is therefore different from models of purely volumetric deformation, such as the volumetric loss modeled in Shen et al. (2007) and the bilateral

reductions in cortical thickness modeled in Radua et al. (2014), aiming at the validation of VBM for the studies of brain atrophy.

Although different levels of brain *petalia* and bending of the inter-hemispheric fissure can be obtained with the ramGLB model by varying its parameters (see Appendix A), we chose to keep the model parameters identical across the images. This was done to ascertain that a realistic pattern of brain asymmetry was consistently applied (in similar locations and with similar magnitudes and directions) to all the images so as to model a realistic shape change (Gao and Bouix, 2012). The ramGLB model was applied in the stereotaxic space while the inter-individual variation of asymmetry patterns was automatically introduced in the native space. Naturally, it is difficult to give precise statements about the realism of this variation other than the qualitative demonstration that the simulated asymmetry matched well with the observed gross-level asymmetry in the applied sample (see Figs. 2 and S11).

All the OASIS images included in this study were acquired on the same scanner with identical imaging parameters, thus preventing possible systematic variations in the image acquisition system to be erroneously interpreted as group variations (Good et al., 2001b; Mechelli et al., 2005; Tardif et al., 2010). In particular, it has been shown that the acquisition protocol and the magnetic field strength have a substantial impact on the image quality characteristics that in turn affect the sensitivity and specificity of VBM measures (Tardif et al., 2010). Although not tested, we have no reason to believe that our automatic framework cannot be applied for validating the accuracy of VBM measures of brain asymmetry for a 3 T image acquisition system.

We performed a careful screening of the OASIS images to be included in this study. This strict quality control resulted in a number of images being discarded if these had even just a mild image distortion, acquisition problem, reconstruction error, poor head positioning, skull stripping error, or any other artifact or processing error that could potentially challenge the image processing involved in this study and thus affect the outcome of our VBM validation study. This was due to the fact that our study aimed at evaluating the intrinsic performances and limitations of the VBM methodology per se rather than testing the robustness against irregularities in the data of a particular VBM implementation and/or image processing routines implemented therein. We also discarded the images exhibiting an appreciable rightward asymmetry as they would result in a non-realistic double inter-hemispheric fissure artifact in regions of the simulated images corresponding to the locations of higher degree of hemispheric protrusion in the original image (Pepe and Tohka, 2012).

VBM methodology

We performed the VBM analysis while varying DoF in the spatial normalization, spatial normalization algorithms, and the amount of spatial smoothing applied to the *DI* maps. We did not vary the specific algorithm for brain tissue classification to avoid the exponential increase of the number of cases needed to consider. Our particular choice of the brain tissue classification algorithm (Tohka, 2013; Tohka et al., 2004) does not need prior knowledge on the spatial distribution of brain tissues in a stereotaxic space, thus avoiding the iterative segmentation–registration–segmentation work-flow of several existing VBM implementations (Good et al., 2001b), which is a known caveat of optimized VBM.

We used peak based (intensity based) correction for multiple comparisons as opposed to the extent based (cluster based) correction. This choice was made since the use of the extent based correction relying on the stationary Random Field theory is not recommended in VBM studies due to the highly non-stationary nature of the brain anatomy (Hayasaka et al., 2004; Mechelli et al., 2005; Moorhead et al., 2005; Radua et al., 2014; Worsley et al., 1999). Moreover, as demonstrated in Radua et al. (2014), the use of large smoothing kernel might connect small clusters, and this might affect the null distribution of cluster sizes, and thus compromising the validity of the extent based correction and complicating the interpretation of the results. In addition, extent based corrections would require the cluster defining threshold to be set resulting in one extra parameter to evaluate.

Findings

VBM accuracy was quantified in this study in terms of various metrics including sensitivity, specificity, and correlation of the VBM measures of group-wise asymmetry and the ground truth asymmetry values. Results indicated that if a symmetrical (customized) template space is used, the asymmetry was localized with reasonable accuracy in most but not all registration schemes. Low DoF spatial normalization in FSL such as 9 DoF affine registration (here referred to as FSL9S) appeared to reach a good compromise between sensitivity (relative amount of the correct detections) and specificity (relative amount of correctly undetected voxels). This finding is in agreement with Salmond et al. (2002), suggesting that very high DoF warping might be not the best option for a VBM analysis. While a spatial normalization with sufficiently high number of DoF (non-linearities) would in principle be able to properly account for the normal inter-subject variability, this does not guarantee that important group-characterizing effects would be retained, or accurately localized.

There were an excess number of false positive asymmetries detected when the asymmetrical template was employed. This clearly indicated that the use of asymmetrical templates should be avoided in asymmetry studies and quantitatively confirmed the correctness of symmetrical template spaces as used in previous VBM studies of brain asymmetry (Good et al., 2001a; Luders et al., 2004; Watkins et al., 2001).

As expected, the number of true and false VBM asymmetry detections was modulated by the level of spatial smoothing. This exemplifies the importance of making a proper choice of kernel size while designing a VBM study. Depending on the applied spatial normalization method, 4 mm or 8 mm FWHM Gaussian kernel smoothing was found to provide a good trade-off between sensitivity and specificity. This makes perfect sense due to the intensity and distribution of asymmetry introduced by the ramGLB model with parameters as described in the Synthetic dataset generation section. In our case, we introduced a pattern of inter-hemispheric asymmetry that ranged between a maximum of 8 mm displacement in a few voxels in the occipital pole, to a minimum of 1 mm displacement in more anterior voxels, with an averaged displacement of approximately 4 mm.

Relating findings to other VBM evaluations

An outcome of the recent scientific debate on the use of VBM is that the interpretation of VBM results should take into account not only the accuracy of the image pre-processing, especially segmentation and spatial normalization, but also the template space used, the amount of non-stationary residual variance in the data, and the kernel size used for the Gaussian spatial smoothing (Senjem et al., 2005; Thacker, 2008). As already mentioned, the literature regarding VBM analysis validation, in general applications and specifically for the analysis of brain asymmetries, is scarce and mainly focused on qualitative observations (Keller et al., 2004; Senjem et al., 2005) concluding that 1) inter-group morphometric comparisons are highly affected by the implementation of VBM used for the analysis, and 2) the optimized VBM outperforms the standard VBM. For example, Senjem et al. (2005) extrapolated qualitative observations about VBM from tests done on a group of elderly controls and patients with probable Alzheimer's disease based on visual comparisons of the obtained VBM results to the known pathology of Alzheimer's disease. Keller et al. (2004) performed VBM validation experiments with images of subjects suffering from the temporal lobe epilepsy and interpreted them on the basis of their agreement to previous volumetric findings on hippocampal and extrahippocampal abnormalities in temporal lobe epilepsy.

It is interesting to make a few observations regarding the two quantitative VBM validation works of which we are aware, Shen et al. (2007) and Radua et al. (2014), although they do not directly address VBM for brain asymmetries. Shen et al. (2007) tested standard, customized or ad hoc template spaces for the validation of VBM in simulated clinical populations and found that in simulated images presenting stroke-like lesions (2 images) and brain atrophy (50 images), the likelihood of detecting between group differences with VBM was highly dependent on the VBM design, and that the choice of template space was highly important. The quantitative validation was done in terms of percentage differences between the detected size of atrophic regions and the simulated size of the atrophic regions. The problem with this validation measure is that it does not take into account the location where a difference was detected. Thus, a perfect validation score might be obtained in case of a completely erroneous localization of the atrophy pattern that is (by chance) of correct size. In addition, while Shen et al. (2007) simulated the data modeling a volumetric loss of GM, asymmetries modeled in this work were mainly positional and rotational (see Appendix A), and the outcomes of two studies are therefore not directly comparable one to the other. Radua et al. (2014) evaluated the effects of the modulation combined with high DoF diffeomorphic spatial normalization algorithms. They found that in simulated images presenting abnormal patterns of bilateral cortical thinning and reductions of the volumes of subcortical nuclei, the modulation was associated to a substantial increase in variance and thus to a substantial decrease of the t-statistic values and sensitivity of VBM detections, especially after applying high DoF spatial normalization algorithms. Related to this, we found that 9 DoF registrations provided a better sensitivity–specificity trade-off as compared to 12-DoF or higher DoF spatial normalization suggesting that 9-DoF might better preserve the amount of asymmetries as introduced by the ramGLB model. On the other hand, we observed a relatively small effect of the modulation step that was in line with Good et al. (2001b) and Keller et al. (2004) but in contradiction with Radua et al. (2014). This was probably due to the fact that in our work, as well as in Good et al. (2001b) and in Keller et

al. (2004), lower DoF spatial normalization scheme was employed (as compared to the few thousands DoF of the spatial normalization employed in Radua et al. (2014)) leading to a lesser “multiplicative noise” effect as discussed in Radua et al. (2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Academy of Finland grants 130275 and 263785, and the Tampere Doctoral Programme in Information Science and Engineering (TISE). This work was also partially supported by NSF grants 0716055 and 1023115. Data used in this study was provided by the Open Access Series of Imaging Studies (OASIS) project, and made available by the Washington University Alzheimer's Disease Research Center, Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Research Network (BIRN). The OASIS project was supported by grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

Appendix A. Brain torque modeling

In short, in the Rescaled Adaptive Modified Global Linear Bending (ramGLB) parametric model, the space is bent along a line parallel to one of the axis by a piece-wise constant deformation that includes several bending regions and non-bending regions. In the bending regions, the bending deformation is approximated by simultaneous rotations and translations of two components of each point around the third one (Barr, 1984; Pepe and Tohka, 2012). In the non-bending regions, the space is rigidly rotated and translated (Barr, 1984; Pepe and Tohka, 2012). The ramGLB model also includes constraints for realistically approximating the occipital rightward and frontal leftward bending of the inter-hemispheric fissure. These additional constraints maintained the simplicity of the ramGLB bending deformation while increasing its modeling capabilities as we have demonstrated in Pepe and Tohka (2012).

Let (x, y, z) and (X, Y, Z) denote the original (undeformed) and deformed x -, y - and z -coordinates, respectively. The ramGLB bending along a centerline parallel to the y axis is defined as follows:

$$X(z) = \begin{cases} r_x^z [\cos(\theta^z) \left(x - \frac{1}{k_1} \right) + \frac{1}{k_1} + \sin(\theta^z)(y - y_A)] & \text{if } y < y_A \\ r_x^z [\cos(\theta^z) \left(x - \frac{1}{k_1} \right) + \frac{1}{k_1}] & \text{if } y_A \leq y < y_1^z \\ r_x^z [\cos(\theta^z) \left(x - \frac{1}{k_1} \right) + \frac{1}{k_1} + \sin(\theta^z)(y - y_1^z)] & \text{if } y_1^z \leq y \leq y_2^z \\ r_x^z [\cos(\theta^z) \left(x - \frac{1}{k_2} \right) + \frac{1}{k_2}] & \text{if } y_2^z < y \leq y_C \\ r_x^z [\cos(\theta^z) \left(x - \frac{1}{k_2} \right) + \frac{1}{k_2} + \sin(\theta^z)(y - y_C)] & \text{if } y > y_C \end{cases} \quad (2)$$

$$Y(z) = \begin{cases} r_y^z [-\sin(\theta^z) \left(x - \frac{1}{k_1}\right) + \cos(\theta^z)(y - y_A)] + y_1^z & \text{if } y < y_A \\ r_y^z [-\sin(\theta^z) \left(x - \frac{1}{k_1}\right)] + y_1^z & \text{if } y_A \leq y < y_1^z \\ r_y^z [-\sin(\theta^z) \left(x - \frac{1}{k_1}\right) + \cos(\theta^z)(y - y_1(z))] + y_1^z & \text{if } y_1^z \leq y \leq y_2^z \\ r_y^z [-\sin(\theta^z) \left(x - \frac{1}{k_2}\right)] + y_2^z & \text{if } y_2^z < y \leq y_C \\ r_y^z [-\sin(\theta^z) \left(x - \frac{1}{k_2}\right) + \cos(\theta^z)(y - y_C)] + y_2^z & \text{if } y > y_C \end{cases} \quad (3)$$

$$Z(z) = z \quad (4)$$

where $k_1 \in \mathbb{R}$ and $k_2 \in \mathbb{R}$ are the constants bending rates measured in radian per unit length; where θ^z is the bending angle

$$\theta^z = \theta(z) = \begin{cases} n_1^z k_1 (y_A - y_1^z) & \text{if } y < y_A \\ n_1^z k_1 (y - y_1^z) & \text{if } y_A \leq y < y_1^z \\ 0 & \text{if } y_1^z < y \leq y_2^z \\ n_2^z k_2 (y - y_2^z) & \text{if } y_2^z < y \leq y_C \\ n_2^z k_2 (y_C - y_2^z) & \text{if } y > y_C \end{cases} \quad (5)$$

where $n_1 \in \mathbb{R}^+$ and $n_2 \in \mathbb{R}^+$ are the constant factors of bending amplification; n_1^z and n_2^z are the factors of bending amplification, in general different for each value of z ; y_1 and y_2 are the centers of the deformations; and where $y_A \leq y_1^z \leq y_2^z \leq y_C$ and $z_{min} \leq z_1 \leq z_2 \leq z_{max}$ define the bending regions in the way described in Pepe and Tohka (2012). The dependence from z of the model parameters is detailed in Table A.1. Notice how, for clarity of notation, the dependence of the ramGLB model from its 12 model parameters ($k_1, k_2, n_1, n_2, y_1, y_2, y_A, y_C, z_1, z_2, z_{min}$ and z_{max}) is omitted throughout this manuscript.

Appendix B. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2014.06.029>.

References

- Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 1999; 412:319–341. [PubMed: 10441759]
- Amunts K, Jancke L, Mohlberg H, Steinmetz H, Zilles K. Interhemispheric asymmetry of the human motor cortex related to handedness and gender. *Neuropsychologia.* 2000; 38:304–312. [PubMed: 10678696]
- Ashburner J, Friston KJ. Voxel-based morphometry: the methods. *NeuroImage.* 2000; 11:805–821. [PubMed: 10860804]
- Barr A. Global and local deformations of solid primitives. *SIGGRAPH Comput. Graph.* 1984; 18:21–30.
- Beaton AA. The relation of planum temporale asymmetry and morphology of the corpus callosum to handedness, gender and dyslexia: a review of the evidence. *Brain Lang.* 1997; 60:255–322. [PubMed: 9344480]
- Binder J. The new neuroanatomy of speech perception. *Brain.* 2000; 123:2371–2372. [PubMed: 11099441]

- Bookstein F. Voxel-based morphometry should not be used with imperfectly registered images. *NeuroImage*. 2001; 14:1454–1462. [PubMed: 11707101]
- Dapretto M, Bookheimer SY. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron*. 1999; 24:427–432. [PubMed: 10571235]
- Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*. 2004; 23:17–20. [PubMed: 15325347]
- Dorsaint-Pierre R, Penhune VB, Watkins K, Neelin P, Lerch J, Bouffard M, Zatorre RJ. Asymmetries of the planum temporale and heschl's gyrus: relationship to language lateralization. *Brain*. 2006; 129:1164–1176. [PubMed: 16537567]
- Foundas AL, Leonard CM, Gilmore RL, Fennell EB, Heilman KM. Pars triangularis asymmetry and language dominance. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:719–722. [PubMed: 8570622]
- Friston K, Holmes A, Worsley K, Poline J, Frith C, Frackowiak R. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 1995; 2:189–210.
- Gao Y, Bouix S. Synthesis of realistic subcortical anatomy with known surface deformations. *Mesh Processing in Medical Image Analysis 2012 Lecture Notes in Computer Science*. 2012:80–88.
- Geschwind N, Galaburda AM. Cerebral lateralization: biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Arch. Neurol.* 1985; 42:428–459. [PubMed: 3994562]
- Good CD, Johnsrude I, Ashburner J, Henson R, Friston K, Frackowiak R. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage*. 2001a; 14:685–700. [PubMed: 11506541]
- Good CD, Johnsrude I, Ashburner J, Henson R, Friston K, Frackowiak R. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*. 2001b; 14:21–36. [PubMed: 11525331]
- Good CD, Scahill RI, Fox NC, Ashburner J, Friston KJ, Chan D, Crum WR, Rossor MN, Frackowiak RS. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. *NeuroImage*. 2002; 17:29–46. [PubMed: 12482066]
- Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn. Reson. Med.* 1995; 34:910–914. [PubMed: 8598820]
- Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE. Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage*. 2004; 22:676–687. [PubMed: 15193596]
- Hopkins W, et al. Gray matter asymmetries in chimpanzees as revealed by voxel-based morphometry. *NeuroImage*. 2008; 42:491–497. [PubMed: 18586523]
- Jancke, L.; Steinmetz, H. Anatomical brain asymmetries and their relevance for functional asymmetries. In: Hugdahl, K.; Davidson, RJ., editors. *The Anatomical Brain Asymmetry*. Cambridge, MA: The MIT Press; 2003. p. 187-230.
- Jenkinson M, Smith SM. A global optimisation method for robust affine registration of brain image. *Med. Image Anal.* 2001; 5:143–156. [PubMed: 11516708]
- Jenkinson M, Bannister P, Brady J, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002; 17(2):825–841. [PubMed: 12377157]
- Keenan JP, Thangaraj V, Halpern AR, Schlaug G. Absolute pitch and planum temporale. *NeuroImage*. 2001; 14:1402–1408. [PubMed: 11707095]
- Keller SS, Wilke M, Wieshmann UC, Sluming VA, Roberts N. Comparison of standard and optimized voxel-based morphometry for analysis of brain changes associated with temporal lobe epilepsy. *NeuroImage*. 2004; 23:860–868. [PubMed: 15528086]
- Keller SS, Roberts N, Garcia-Finana M, Siawoosh M, Ringelstein EB, Knecht S, Deppe M. Can the language-dominant hemisphere be predicted by brain anatomy? *J. Cogn. Neurosci.* 2011; 23:2013–2029. [PubMed: 20807056]
- Kovalev V, Vassili A, Frithjof KD, Yves von C. Gender and age effects in structural brain asymmetry as measured by MRI texture analysis. *NeuroImage*. 2003; 19:895–905. [PubMed: 12880818]
- LeMay M. Asymmetries of the skull and handedness. *J. Neurol. Sci.* 1977; 32:243–253. [PubMed: 874523]

- Luders E, Gaser C, Schlaug G. A voxel-based approach to gray matter asymmetries. *NeuroImage*. 2004; 22:656–664. [PubMed: 15193594]
- Manjón J, Lull JJ, Carbonell-Caballero J, García-Martí G, Martí-Bonmati L, Robles M. A nonparametric MRI inhomogeneity correction method. *Med. Image Anal.* 2007; 11:336–345. [PubMed: 17467325]
- Manjón J, Coupé P, Martí-bonmati L, Robles M, Collins D. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J. Magn. Reson. Imaging*. 2010; 31:192–203. [PubMed: 20027588]
- Marcus D, Wang T, Parker J, Csernansky J, Morris J, Buckner R. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented and demented older adults. *J. Cogn. Neurosci.* 2007; 19:1498–1507. [PubMed: 17714011]
- Mechelli A, Price CJ, Friston KJ, Ashburner J. Voxel-based morphometry of the human brain: methods and applications. *Curr. Med. Imaging Rev.* 2005; 1:105–113.
- Moffat SD, Hampson E, Lee DH. Morphology of the planum temporale and corpus callosum in left handers with evidence of left and right hemisphere speech representation. *Brain*. 1998; 121:2369–2379. [PubMed: 9874487]
- Moorhead TWJ, Job DE, Spencer MD, Whalley HC, Johnstone EC, Stephen ML. Empirical comparison of maximal voxel and non-isotropic adjusted cluster extent results in a voxel-based morphometry study of comorbid learning disability with schizophrenia. *NeuroImage*. 2005; 28:544–552. [PubMed: 16085427]
- Morris J. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. 1993; 43:2412–2414. [PubMed: 8232972]
- Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch. Neurol.* 2001; 58:397–405. [PubMed: 11255443]
- Pepe A, Tohka J. 3D bending of surfaces and volumes with an application to brain torque modeling. *ICPRAM*. 2012; 12(2):411–418.
- Pepe A, Zhao L, Koikkalainen J, Hietala J, Ruotsalainen U, Tohka J. Automatic statistical shape analysis of cerebral asymmetry in 3D T1-weighted magnetic resonance images at vertex-level: application to neuroleptic-naïve schizophrenia. *Magn. Reson. Imaging*. 2013; 31:676–687. [PubMed: 23337078]
- Radua J, Canales-Rodriguez EJ, Pomarol-Clotet E, Salvador R. Validity of modulation and optimal settings for advanced voxel-based morphometry. *NeuroImage*. 2014; 86:8–90.
- Salmond C, Ashburner J, Vargha-Khadem F, Connelly A, Gadian D, Friston KJ. The precision of anatomical normalization in the medial temporal lobe using spatial basis functions. *NeuroImage*. 2002; 17:507–512. [PubMed: 12482103]
- Senjem M, Gunter J, Shiung M, Petersen R, Jack CJ. Comparison of different methodological implementations of voxel-based morphometry in neurodegenerative disease. *NeuroImage*. 2005; 26:600–608. [PubMed: 15907317]
- Shen S, Szameitat A, Sterr A. VBM lesion detection depends on the normalization template: a study using simulated atrophy. *Magn. Reson. Imaging*. 2007; 25:1385–1396. [PubMed: 17467945]
- Shenton M, Dickey C, Frumin M, McCarley R. A review of MRI findings in schizophrenia. *Schizophr. Res.* 2001; 49:1–52. [PubMed: 11343862]
- Smith S. Fast robust automated brain extraction. *Hum. Brain Mapp.* 2002; 17(3):143–155. [PubMed: 12391568]
- Sowell ER, Thompson PM, Holmes CJ, Batth R, Jernigan TL, Toga AW. Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage*. 1999; 9:587–597. [PubMed: 10334902]
- Tardif CL, Collins DL, Pike GB. Regional impact of field strength on voxel-based morphometry results. *Hum. Brain Mapp.* 2010; 7:943–957. [PubMed: 19862698]
- Thacker NA. Tutorial: a critique analysis of voxel based morphometry (VBM). *Tina Memo No.* 2003-011, Internal Memo. 2008:1–10.
- Toga AW, Thompson PM. Mapping brain asymmetry. *Nat. Rev. Neurosci.* 2003; 4:37–48. [PubMed: 12511860]

- Tohka J. FAST-PVE: extremely fast Markov random field based brain MRI tissue classification. SCIA 2013, Scandinavian Conference on Image Analysis, Helsinki, Finland, 2013. Lect. Notes Comput. Sci. 2013; 7944:266–276.
- Tohka J, Zijdenbos A, Evans A. Fast and robust parameter estimation for statistical partial volume models in brain mri. *NeuroImage*. 2004; 23(1):84–97. [PubMed: 15325355]
- Wallis LI, Widjaja E, Wignall EL, Wilkinson ID, Griffiths PD. Misrepresentation of surface rendering of pediatric brain malformations performed following spatial normalization. *Acta Radiol*. 2006 Dec.47:1094–1099. [PubMed: 17135016]
- Watkins KE, Paus T, Lerch JP, Zijdenbos A, Collins DL, Neelin P, Taylor J, Worsley KJ, Evans AC. Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. *Cereb. Cortex*. 2001; 11:868–877. [PubMed: 11532891]
- Worsley K, Marrett S, Neelin P, Vandal A, Friston K, Evans A. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp*. 1996; 4:58–73. [PubMed: 20408186]
- Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC. Detecting changes in nonisotropic images. *Hum. Brain Mapp*. 1999; 8:98–101. [PubMed: 10524599]
- Wright IC, McGuire PK, Poline JB, Travers JM, Murray RM, Frith CD, Frackowiak RSJ, Friston KJ. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage*. 1995; 2:244–252. [PubMed: 9343609]

Fig. 1.

Synthetic dataset generation work-flow. An original image i is pre-processed to correct for intensity non-uniformities, denoised, artificially symmetrized (by merging the right portion of the denoised images with their flipped versions with respect to the mid-sagittal plane), and then corrupted with a Rician noise distribution. The obtained simulated symmetrical images $\tilde{s}$ are mapped into the MNI152 space via a spatial normalization transformation T estimated from the skull stripped version of $\tilde{s}$. Next, a parametrically known pattern of inter-hemispheric asymmetry is introduced via ramGLB model to the spatially normalized symmetrical image $T(\tilde{s})$ before mapping it back to its respective native space to obtain the simulated asymmetrical image a . Notice how, to ensure that left and right portions of the artificially symmetrized images $\tilde{s}$ would not exhibit a symmetrical realization of the noise process, the original images had to be first denoised, then symmetrized, and finally a realization of Rician distributed randomnoise could be added to them. Notice also that, to avoid systematic differences among the individual symmetrical $\tilde{s}$ (in native space) and asymmetrical a (in native space) images due to the double trilinear interpolation involved with the repeated spatial normalization of the latter, the spatially normalized simulated images $T(\tilde{s})$ were re-mapped back to their corresponding native spaces and used as the control set of this study.

Fig. 2.

Example of an original image (panel a), and the corresponding simulated symmetrical (panel b) and asymmetrical (panel c) images generated with the method proposed in this study.

Fig. 3.

VBM analysis work-flow consisting of skull stripping, tissue classification, spatial normalization, computation of *DI* images (encoding the tissue specific inter-hemispheric asymmetry), spatial smoothing, univariate statistical analysis of voxel-level asymmetries, and correction for the multiple comparisons. The work-flow is depicted for the study of GM asymmetry, for both densitometric (depicted with a green arrow, where the VBM analysis does not involve a modulation stage) and volumetric (depicted with blue arrows, where the VBM analysis involves a modulation stage by the Jacobian determinants of the deformation field) VBM analyses. The process is done separately for the image sets *S* and *A*.

Fig. 4.

In panel a, a “grid-image” (black and white grid) overlaid to an individual GM contour (green) is shown. The space in panel a is deformed via the ramGLB model with the parameters as described in the Generation for synthetic asymmetrical MR images in stereotaxic space section and the deformed space is depicted in panel b. Panel b shows also the ground truth image in the stereotaxic space color coded from 1 mm displacement (dark red) up to 6 mm displacement (yellow) with 1 mm step increments. Higher displacements, up to 8 mm, occur at different axial sections of the image. In panel c, the ground truth is visualized as a deformation field overlaid on a brain surface.

Fig. 5.

Effects of spatial normalization with 4 mm FWHM smoothing. Examples of an individual DI image (obtained from the same simulated image as in Fig. 2, panels b and c) spatially normalized with FSL9S (first column), FSL9A (second column), SPMnS (third column), and SPMnA (fourth column), for the set S (first row, panels a–d) and the set A (second row, panels e–h). Positive values of DI are color coded in red and negative DI values are color coded in blue. DI values are dimensionless. In the third and fourth rows, the corrected asymmetry detections obtained with FSL9S (first column), FSL9A (second column),

SPMnS (third column), and SPMnA (fourth column) are visualized overlaid on the ground truth displacements (in millimeters), for the set S (third row, panels i–l) and the set tA (fourth row, panels m–p). The red and blue colors denote positive and negative asymmetries surviving the multiple comparisons correction. The ground truth displacements expressed in millimeters are color coded in a gray scale starting at 0 mm (white) up to the maximum displacement value of 8 mm (black). In panels i–l, the ground truth for the set A is visualized to provide a reference of anatomic locations.

Fig. 6. Variance of the t-statistic values in the set S obtained in case of 4 mm (depicted in pink), 8 mm (depicted in green), and 12 mm (depicted in red) FWHM Gaussian kernels smoothing as well as in the case of no smoothing (depicted in blue). The images are anatomically symmetrical, and thus ideally, the variance should have value zero.

Fig. 7.
Maximum t-statistic values in set *S*. The images are anatomically symmetrical, and thus ideally, the maximum t-statistics should have value zero.

Fig. 8.
Pearson's correlation coefficients in the set A. For any given Gaussian kernel size, maximal correlation coefficients were obtained with the FSL9S spatial normalization.

Fig. 9.

Specificity of the detected asymmetry by VBM (with the peak based correction at $p = 0.05$) in the set A calculated for different hard thresholdings of the ground truth (GT), see the Ground truth and Quantitative validation sections. The bar graph represents the specificity values when the ground truth displacement threshold was 1 mm, i.e., if the displacement of the voxel was greater or equal than 1 mm, it was considered to display asymmetry.

Fig. 10.

Sensitivity of the detected asymmetry by VBM (with the peak based correction at $p = 0.05$) in the set A calculated for different hard thresholdings of the ground truth (GT), see the Ground truth and Quantitative validation sections. The bar graph represents the sensitivity values when the ground truth displacement threshold was 1 mm.

Fig. 11.

Effects of spatial smoothing. In the first row, examples of an individual (same as in Figs. 2, panels b and c, and 5, panel e) DI image are depicted in case of no smoothing (panel a), 4 mm FWHM (panel b), 8 mm FWHM (panel b), and 12 mm FWHM (panel d) Gaussian kernels smoothing. Positive values of DI are color coded in red, while negative DI values are color coded in blue. DI values are dimensionless. In the second row, multiple comparison corrected t-statistics obtained with no smoothing (panel e), 4 mm FWHM (panel f), 8 mm FWHM (panel g), and 12 mm FWHM (panel h) Gaussian kernel spatial smoothing are visualized overlaid on ground truth values of displacement expressed in millimeters. The red color denotes positive t-statistic values and the blue color negative t-statistic values that survived the correction for multiple comparisons. The ground truth displacements expressed

in millimeters are color coded in a gray scale starting at 0 mm (white) up to 8 mm (black).
All depicted cases refer to FSL9S spatial normalization and the set A.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Different cases are referred to as [“environment”, “DoF”, “template”, “modulation”], where “environment” can be either be “FSL” or “SPM”; “DoF” refers to the number of DoF used, it can be either “7”, “9”, or “12” for 7-, 9-, or 12-DoF spatial normalization in FSL, and “i” or “n” for 12-DoF spatial normalization in SPM followed by either none or 16 non-linear iterations, respectively; “template” refers to the template space used, it can either be symmetrical “S” or asymmetrical “A”; “modulation” refers to the presence or absence of modulation by the Jacobian of the spatial normalization transformation, it can be either “m” if amounts are preserved (PA), or none if concentrations are preserved (PC). In SPM, the default value of 25 DCT bases for cutoff was used.

Case	Registration scheme	Template type	Modulation	Cost function	Interpolation	Implementation
FSL7S	7 DoF affine	Symmetric	No (PC)	Normal correlation	Trilinear	FSL
FSL9S	9 DoF affine	Symmetric	No (PC)	Normal correlation	Trilinear	FSL
FSL12S	12 DoF affine	Symmetric	No (PC)	Normal correlation	Trilinear	FSL
SPMIS	12 DoF affine + 0 non-linear iterations	Symmetric	No (PC)	Variance	Trilinear	SPM
SPMISm	12 DoF affine + 0 non-linear iterations	Symmetric	Yes (PA)	Variance	Trilinear	SPM
SPMnS	12 DoF affine + 16 non-linear iterations	Symmetric	No (PC)	Variance	Trilinear	SPM
SPMnSm	12 DoF affine + 16 non-linear iterations	Symmetric	Yes (PA)	Variance	Trilinear	SPM
FSL7A	7 DoF affine	Asymmetric	No (PC)	Normal correlation	Trilinear	FSL
FSL9A	9 DoF affine	Asymmetric	No (PC)	Normal correlation	Trilinear	FSL
FSL12A	12 DoF affine	Asymmetric	No (PC)	Normal correlation	Trilinear	FSL
SPMIA	12 DoF affine + 0 non-linear iterations	Asymmetric	No (PC)	Variance	Trilinear	SPM
SPMIAm	12 DoF affine + 0 non-linear iterations	Asymmetric	Yes (PA)	Variance	Trilinear	SPM
SPMnA	12 DoF affine + 16 non-linear iterations	Asymmetric	No (PC)	Variance	Trilinear	SPM
SPMnAm	12 DoF affine + 16 non-linear iterations	Asymmetric	Yes (PA)	Variance	Trilinear	SPM

Table 2

The number of false positives in the set S obtained in the case of no smoothing, and in the cases of 4 mm, 8 mm, and 12 mm FWHM-Gaussian smoothing.

	FSL7S	FSL9S	FSL12S	SPMIS	SPMISm	SPMInS	SPMnSm	FSL7A	FSL9A	FSL12A	SPMIA	SPMIAm	SPMnA	SPMnAm
No smooth	140	132	330	1842	1814	278	184	142,506	176,490	187,166	49,994	49,780	154,202	148,150
4 mm FWHM	4876	4280	5306	12,116	12,092	3520	3986	747,948	801,214	826,728	419,868	420,028	611,694	604,900
8 mm FWHM	15,764	13,990	16,502	55,378	55,884	13,722	16,570	1,674,218	1,753,968	1,770,550	1,143,508	1,143,314	1,506,112	1,489,600
12 mm FWHM	28,664	26,920	31,606	134,552	135,570	21,348	25,270	2,342,680	2,410,998	2,417,062	1,748,510	1,747,006	2,266,900	2,214,536