Neuroimage. Author manuscript; available in PMC 2011 June 30.

Published in final edited form as:
Neuroimage. 2010 January 01; 49(1): 587–596. doi:10.1016/j.neuroimage.2009.07.018.

Voxel-based morphometry and automated lobar volumetry: The trade-off between spatial scale and statistical correction

Eduard H.J. Voormolen[a,b], **Corie Wei**[b], **Eva W.C. Chow**[c], **Anne S. Bassett**[c], **David J. Mikulis**[b], and **Adrian P. Crawley**[b,*]

[a] Image Sciences Institute, University Medical Centre Utrecht, Utrecht, The Netherlands [b] Deparment of Medical Imaging, Toronto Western Hospital, University of Toronto, Toronto, Canada [c] Clinical Genetics Research Program, Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada

Abstract

Voxel-based morphometry (VBM) and automated lobar region of interest (ROI) volumetry are comprehensive and fast methods to detect differences in overall brain anatomy on magnetic resonance images. However, VBM and automated lobar ROI volumetry have detected dissimilar gray matter differences within identical image sets in our own experience and in previous reports. To gain more insight into how diverging results arise and to attempt to establish whether one method is superior to the other, we investigated how differences in spatial scale and in the need to statistically correct for multiple spatial comparisons influence the relative sensitivity of either technique to group differences in gray matter volumes. We assessed the performance of both techniques on a small dataset containing simulated gray matter deficits and additionally on a dataset of 22q11-deletion syndrome patients with schizophrenia (22q11DS-SZ) vs. matched controls. VBM was more sensitive to simulated focal deficits compared to automated ROI volumetry, and could detect global cortical deficits equally well. Moreover, theoretical calculations of VBM and ROI detection sensitivities to focal deficits showed that at increasing ROI size, ROI volumetry suffers more from loss in sensitivity than VBM. Furthermore, VBM and automated ROI found corresponding GM deficits in 22q11DS-SZ patients, except in the parietal lobe. Here, automated lobar ROI volumetry found a significant deficit only after a smaller subregion of interest was employed. Thus, sensitivity to focal differences is impaired relatively more by averaging over larger volumes in automated ROI methods than by the correction for multiple comparisons in VBM. These findings indicate that VBM is to be preferred over automated lobar-scale ROI volumetry for assessing gray matter volume differences between groups.

Introduction

Many different techniques exist to measure differences in images of the brain between groups. Traditionally, researchers have manually delineated Regions-Of-Interest (ROI) within images to compare their volumes. Today, several automated alternatives for

*Corresponding author. crawley@uhnres.utoronto.ca (A.P. Crawley).

volumetric comparison exist which are far less labor intensive. Here, we focus on two such methodologies which allow fast and unbiased screening for overall cortical gray matter (GM) volume differences; automated ROI methods and voxel-based morphometry (VBM). Both methods involve spatial normalization of the brain images to a standard stereotactic space by registration to an atlas image.

Automated ROI methods (or brain parcellation) are frequently used to study structural brain anomalies in schizophrenia patients (Lopez-Garcia et al., 2006; Yamasue et al., 2004) and recently are gaining popularity in aging research to demonstrate non-linear relationships between brain volumes and age (Andreescu et al., 2007; Terribilli et al., in press).

An alternative automatic method to detect group differences in brain structure is voxel-based morphometry (VBM). VBM surveys entire brains for volumetric differences and creates a map of statistically significant structural differences at the voxel level. VBM has been applied extensively in studies of schizophrenia and other types of neuropsychiatric patients to detect differences in cortical gray matter (Bonilha et al., 2005; Gogtay et al., 2007; Good et al., 2002; Kubicki et al., 2002; Moorhead et al., 2004).

The motivation for the present work arose from research conducted at our laboratory using Brain Image software (Reiss et al., 1995) to compare gray matter volumes in automatically segmented ROIs corresponding to brain lobes in a group of 22q11 deletion syndrome patients with schizophrenia (22q11DS-SZ) to matched controls (Chow et al., 2002). We wanted to use VBM in this 22q11DS-SZ population and wondered how it would perform. Moreover, another recent study (Cantor et al., 2008) has used VBM and automated ROI at brain lobe scale in conjunction, to study white matter deficiencies in pedophilic men and noted some discrepancies in outcomes between methods.

In this paper, we present a comparison between VBM and automated ROI methods. Ideally, one would attempt to match the spatial scale of the imaging acquisition to the (usually unknown) spatial scale of the effects to be detected. We focus on the precise nature of the trade-off between this spatial scale selection and the demands of statistical correction for multiple comparisons. At the relatively large spatial scale characterized by the use of lobar ROIs, few comparisons are performed, and the applied statistical correction is small. However, at these large scales, brain volume differences may be missed due to volumetric dilution if the effects are focal in nature.

When between-group brain differences are distributed uniformly e.g. over a whole brain lobe, an automated ROI analysis at this spatial scale would be expected to benefit from substantial signal averaging. In contrast, the finer spatial resolution of VBM will limit spatial averaging in this case and require more conservative correction for the much larger number of statistical tests performed. However, VBM should perform relatively better in detecting more focal deficits, since the signal averaging over an ROI will now cause a dilution of the automated ROI measure of the volume difference.

In short, we hypothesize that the relative sensitivity of ROI-based vs. voxel-based methods depends on the relative penalty of volumetric dilution imposed on an ROI analysis vs. the statistical correction penalty imposed on a VBM analysis.

To investigate our hypothesis with minimum bias, pre-processing differences between techniques needed to be eliminated. So, we implemented an automated ROI method operating at lobar scale, but with identical pre-processing as VBM; automated lobar volumetry (ALV). We compared ALV to VBM on a small dataset of images of healthy brains in which we artificially simulated focal and more global GM deficits by manipulating GM voxel intensities. Furthermore, we calculated the theoretical detection sensitivity of ROI analysis and VBM to focal deficits, as a function of ROI size. Moreover, we applied ALV and VBM on our real dataset of 22q11DS-SZ patients and controls (Chow et al., 2002), to assess if differences between techniques led to different results and if the hypothesized mechanism might explain these differences.

Materials and methods

Image sets

We analyzed structural magnetic resonance (MR) images divided in two distinct image sets; image set S (simulated) and image set R (real). *Image set R* consisted of two groups. One group of fourteen scans of 22q11DS-SZ (7 males; mean age 28, standard deviation (SD) 6.4 years) patients and the other group of fourteen age and gender matched healthy control subjects (7 males; mean age 28, SD 6.3 years). However, one subject in the 22q11DS-SZ group had images with motion artefact which precluded tissue segmentation and was excluded from our study.

Additionally, the fourteen controls from image set R were divided into two groups and formed *image set S*. In one group of scans of 7 subjects (4 males; mean age 29, SD 5.7 years) GM deficits were introduced artificially and in the other group of 7 subjects (3 males; mean age 27, SD 7.2 years) no GM deficits were introduced. All 22q11DS-SZ and healthy control images were used in a previous comparison study by Chow et al. (2002), which found significant GM deficits in all lobar areas except for right parietal and occipital lobes.

MRI scan protocol

Structural MR images of all twenty-eight brains were acquired with a GE-Signa 1.5 T scanner (General Electric, Milwaukee, WI). Coronal images were acquired with a three-dimensional volumetric radio frequency spoiled gradient echo (SPGR) using the following scan parameters: flip angle=45°, TR=25 ms, TE=5 ms, matrix size=256×256×124, and voxel dimensions=0.78×0.78×1.5 mm.

Image pre-processing

This study utilized two different techniques for comparison of brain anatomy between groups; voxel-based morphometry (VBM) and automated lobar volumetry (ALV). To facilitate pre-processing the anatomical MR images were transferred to a Unix workstation and processed using Statistical Parametric Mapping version 5 software (SPM5, Wellcome Department of Imaging Neuroscience, London; available at http://www.fil.ion.ucl.ac.uk/spm) running in MATLAB version 6.5 (The MathWorks, Natick, MA).

In SPM5, tissue classification into GM, WM, and CSF, intensity inhomogeneity correction and normalization of the original anatomical MR images to a Montreal Neurologic Institute (MNI) template (ICBM, NIH P-20 project) approximating Talairach and Tournoux space are combined into a single generative model (Ashburner and Friston, 2005). Spatial normalization was performed by a linear affine transformation followed by non-linear deformations defined by a linear combination of 3D discrete cosine transform basis functions. The normalized images were modulated to compensate for any volumetric differences introduced during normalization. Subsequently, these GM images were smoothed by a 12 mm full width half maximum (FWHM) kernel to render the data normally distributed and to compensate for imperfect spatial normalization, and to look at differences in tissue composition in a local neighbourhood defined by the kernel.

We expanded the standard SPM5 generative model with the code from the VBM5 Toolbox written by Christian Gaser (available at http://dbm.neuro.uni-jena.de/vbm/). This toolbox extends the core segmentation by weighting the prior spatial probability of a given voxel belonging to a particular tissue class on the basis of neighbouring voxel tissue class information through application of a Gaussian Hidden Markov Random Field (GHMRF) model. The prior spatial tissue probability is adjusted according to the number of neighbouring voxels assigned to that tissue class. High numbers increase the probability and low numbers decrease this probability. Effectively, this removes isolated voxels of one tissue class which are unlikely to contain that tissue. Ultimately, this reduces noise and improves segmentation accuracy. Parts of this toolbox are based on an implementation of a GHMRF approach by Cuadra et al (2005).

Automated lobar volumetry

The SPM PickAtlas software toolbox (Maldjian et al., 2003, 2004) was used to create four binary masks of the bilateral frontal, temporal, parietal, and occipital lobes based on the Talairach Daemon database (Lancaster et al., 1997, 2000). Quality control was performed for each subject by inspecting whether the masks projected properly over the corresponding lobes in the image. Subsequently, these binary lobe-shaped masks were multiplied with segmented unsmoothed GM images to yield lobe-specific GM images for the frontal, temporal, parietal and occipital lobes for each subject. Note that only for the analysis in image set R did we use an extra sub-mask of the left supramarginal gyrus to assess GM differences in that area (please see the Results section for a more detailed explanation).

These images were converted to lobar GM volumes in cm^3 by a script written by J. Ashburner (Pernet, 2005). Figs. 1a–d shows the four brain lobe masks used in ALV pre-processing. Note that the previously published study by Chow et al (2002) conducted on this image set indicated little lateral asymmetry amongst lobar pairs (Chow et al., 2002). Consequently, bilateral lobar pairs were used instead of single lobes.

GM deficit simulations

We generated a range of GM deficits in seven structural MR scans of subjects in image set S. We simulated both focal GM deficits and more global (i.e. lobar) GM deficits. Focal deficits were introduced in the form of cubes centered around a GM voxel in the left superior

temporal gyrus (STG). The left STG was chosen because previous reports have consistently found GM deficits here in schizophrenia patients(Honea et al., 2005). These focal cubical GM deficits were generated by registering the original structural images to MNI space and subsequently multiplying the registered images with images containing cube-shaped masks sized 6 mm^3, 12 mm^3 and 20 mm^3 of intensity value 0.1, centered on Talairach coordinates (x=58, y=−32, z=10). Effectively, the voxel intensity value within the cubical area in the STG was reduced by 90%. Please see Figs. 2a–c for examples of image slices containing the artificial cubical deficits.

In a similar fashion global GM deficits were simulated in the left superior temporal gyrus (STG; see Fig. 2d) and left temporal lobe (LTL; see Fig. 2h). These deficits were created by first multiplying the registered unsmoothed GM images by a binary PickAtlas mask of the STG or LTL. A suitable GM mask was defined as voxels where the GM image values exceeded 0.001. Throughout this mask, the original registered anatomical image intensities were multiplied by 0.1 to reduce signal in those areas by 90%. This procedure was performed for each subject separately.

Furthermore, we also eroded the cortical GM of the whole left temporal lobe (LTL) by varying amounts. For each subject, the GM and white matter (WM) probability maps were added to each other and subsequently imported into AFNI version 2.52 h (available at http://afni.nimh.nih.gov/). These WM+GM images were eroded by the AFNI function “3Dmerge”. In order to create simulated deficit masks to be applied to the anatomical images as before, the eroded images were subtracted from un-eroded images which resulted in images containing a rim of GM voxels at the cortical edge and a rim of WM voxels at the ventricular borders. Next, we multiplied these rim images with a binary subject-specific LTL GM image to remove the ventricular WM borders and obtain a pure cortical rim image of the LTL for each subject. As before, voxel intensities in the registered anatomical image that corresponded to voxels in the deficit mask were reduced by 90%. These cortical GM erosions were performed successively three times to create GM deficits with widths averaging one, two and three voxels respectively (see Figs. 2e–g).

After introducing the GM deficits in the images they were warped back to original space using inverse deformation maps obtained during the original registrations. This operation involved trilinear interpolation which caused these images to differ slightly from the original structural images. To adjust for this, we also subjected the MR scans of the control group to registration and inverse registration. In effect, registration was performed twice, once to generate the simulated deficits and subsequently to facilitate statistical analyses.

Theoretical detection sensitivity as function of ROI size

A computer program was written to calculate the theoretical sensitivity of a relatively focal effect (comprising m “signal” voxels, i.e. voxels with GM deficit) detected via the mean of an ROI measurement over N independent voxels, compared to the maximum effect obtained at any voxel within this ROI corresponding to a peak voxel VBM measurement. Signal dilution in the ROI mean was modelled as the voxel signal-to-noise (z_{vox}) multiplied by $m/\sqrt{N}$. The detection sensitivity for the ROI method was calculated at p<0.05 as the statistical power for a Gaussian probability density function centered at ($m/\sqrt{N}$) z_{vox}. For VBM, a

stringent Bonferroni correction $p_{fwe} = p / N$ was modelled. Since this tightly controls the false positive rate, any peak voxel that survives this threshold is highly likely to come from the modelled distribution of deficit voxels. Therefore the distribution for the peak voxel z-statistic was calculated as the theoretical signal distribution obtained by picking the maximum out of the set of m signal voxels. The cumulative distribution function (CDF) for this distribution is given by $CDF_{max}(x)=[CDF(x)]^m$, which gives the following probability density function: $PDF_{max}(x)= m\,[CDF(x)]^{m-1}PDF(x)$, where we assume $PDF(x)$ to be Gaussian z_{vox} scores. VBM detection sensitivity was calculated as the statistical power obtained by integrating $PDF_{max}(x)$ above a threshold corresponding to $p_{fwe} < 0.05$.

In standard SPM terminology, a “resel” is an effective resolution element with a volume that is considerably larger than the original voxel volume, resulting from the user-specified image smoothing operation. It can be thought of as the effective ROI size that is used by VBM. Here, the theoretical detection sensitivity as a function of the number of independent resels investigated was calculated for deficits characterized by resel z-scores of 3.0 over a spatial extent of 10 resels.

Statistical analyses

Statistical tests were performed on gray matter data generated from images from both image sets R and S. The statistical input data was volumetric for ALV or voxel-wise for VBM. The volumetric data was analyzed with an analysis of co-variance (ANCOVA) in which we corrected for total intracranial volume variance (TIV). TIV was calculated by adding segmented gray matter, white matter and cerebrospinal fluid volumes. A Bonferroni correction was applied because multiple statistical comparisons were performed on the same image sets. The correction entailed that our threshold probability of p=0.05 was divided by the number of comparisons. For ALV we performed four comparisons (in the frontal, temporal, parietal, and occipital lobes) so our cut-off p-value was set at p_{bon} =0.0125.

The voxel-wise data was subjected to statistical analyses with SPM5. SPM5 was used to carry out an ANCOVA corrected for TIV. The ANCOVA yielded 3D statistical parametric T-score maps (SPM-Ts) in which voxel clusters that exhibited significant GM differences were identified by thresholding at T-scores corresponding to a family-wise error corrected probability (p_{fwe}) threshold of 0.05 because of multiple comparisons over the whole brain. We chose to use family-wise error correction because it is conservative and akin to Bonferroni correction. MRIcro software (version 1.4; available at http://www.sph.sc.edu/comd/rorden/mricro.html) was used to project these SPMs over the normalized modulated GM image of a single subject to associate GM group differences with corresponding brain anatomy visually. Talairach Daemon software was used to relate voxel coordinates with anatomical areas such as brain lobe, gyrus and Brodmann area (Lancaster et al., 2000).

Results

Gray matter deficit simulations

VBM analysis of images in image set S (7 controls with deficits vs. 7 normal controls) with cubical deficit sizes of 20 mm^3 showed a single voxel cluster with a significant group

difference in the middle of the cubic GM deficit in the left STG (see Fig. 3a). However, VBM analysis of images with cubical deficits of 6 mm^3 and 12 mm^3 did not show any significant voxel clusters. Analysis of the STG deficit yielded an extensive supra-threshold cluster over the whole of the STG (see Fig. 3b).

Moreover, VBM was able to identify global GM thinning deficits of an extent of 2 voxels and beyond in the LTL cortex. Interestingly, at 2 voxels of lobar cortical thinning the resulting thresholded SPM-T depicted a single supra-threshold cluster in the area of the LTL similar to the results obtained in the focal cubical GM deficit case (see Fig. 3c). At 3 voxels of cortical thinning two separated clusters appeared (see Fig. 3d). When total loss of cortex was simulated a more global cluster pattern emerged (see Fig. 3e) reflecting the true nature of the underlying cortical deficit. VBM did not give rise to false positive outcomes when no GM deficits had been introduced, with SPMs thresholded at *T*-scores corresponding to p_{fwe} =0.05.

Additionally, in the same image set S, ALV analysis with a temporal lobe mask was performed. ALV correctly identified significant group differences in LTL cortical thinning simulations of 2 voxels and beyond. However, it was unable to identify any cubical or STG GM deficit. ALV yielded no false positive results in the frontal, parietal and occipital lobes.

We compared the relative detection sensitivity of ALV and VBM to the different types of GM deficits by plotting the *T*-scores of both techniques in Fig. 4. For VBM, the *T*-score of the most significantly different voxel was obtained from the different SPM-Ts. The axes of Fig. 4 were positioned at the *T*-score cut-off value at which group differences became significant. For VBM analyses a T_{VBM} of 18.9 was obtained. For ALV, the cut-off value was at T_{ALV} =3.1. These correspond to a family-wise error corrected $p<0.05$ in both cases. Fig. 4 divides into four quadrants. Results that lie in the top left quadrant represent VBM true positives that are ALV false negatives. VBM scored false negatives in 3 out of 8 deficit cases (sensitivity: 62.5%), compared to 5 cases out of 8 (sensitivity: 38%) for ALV. Note the excellent linear correlation (Pearson correlation=0.99) between the VBM and ALV *T*-scores for the 4 lobar cortical thinning simulations and for the cubical GM deficits (Pearson correlation=0.99), both shown as dashed lines in Fig. 4.

Theoretical detection sensitivity to focal deficits in ROI and VBM methods

The calculated sensitivities of detecting a focal deficit at different ROI sizes are shown in Fig. 5 for both ROI-based and VBM analyses. At a small ROI size, equal to the size of the deficit, VBM and ROI volumetry have similar detection sensitivities of 100%. However, as ROI size increases, the theoretical detection sensitivity of ROI volumetry decreases far more rapidly than the sensitivity of VBM.

Gray matter deficits of 22q11DS-SZ patients compared to healthy controls

VBM analysis (SPMs were thresholded at a *T*-score corresponding to a p_{fwe} of 0.05) of image set R (14 patients vs. 14 controls) showed that the 22q11DS-SZ group exhibited statistically significant GM deficits in the frontal, temporal, parietal and occipital lobes. (Additionally, GM deficits were found in the limbic lobe and cerebellum, but these regions

were not investigated by ALV.) No areas where 22q11DS-SZ patients demonstrated excess GM were found. These results are shown in Fig. 6.

Similarly, ALV analysis found significant 22q11DS-SZ GM deficits in the frontal, temporal and occipital lobes. However, ALV data did not yield a statistically significant GM deficit in the parietal lobes. Because of this discrepancy, we conducted further investigations. More specifically, VBM analysis found a significant difference in the left supramarginal gyrus. Initially, unilateral ALV analysis of the left parietal lobe still could not find a significant group difference. However, when volumes of the left supramarginal gyrus were compared using a sub-mask, significant differences were found (T=2.66, $p_{bon} < 0.0125$; data not shown). Table 1 summarizes the results obtained for image set R using VBM and ALV analyses.

Discussion

We investigated the influence of the spatial scale trade-off between averaging over larger volumes with the risk of diluting small GM deficits in lobar-scale automated ROI methods, and the stringent statistical corrections for multiple comparisons in VBM. To the best of our knowledge, we are the first to show that VBM is able to outperform large scale automated ROI methods at detecting GM volume differences.

Our simulations indicate that VBM is similarly sensitive to lobar cortical thinning deficits compared to ALV. However, ALV is relatively insensitive to focal deficits as compared to voxel-wise analyses, due to the small volume of the deficit with respect to the total volume of the lobar ROI. This trade-off phenomenon is nicely illustrated in Fig. 4.

In Fig. 4, there is a fixed relationship between VBM and ALV T-scores for a given simulated deficit type, as seen by the steep fit line passing through the most focal cubical deficit points and the less inclined fit line passing through the more global lobar deficit points. As the simulated deficit involves more voxels, the ALV method gains an advantage due to spatial averaging. The VBM method also benefits from a modest increase in extent, since the probability of the highest T-score voxel exceeding the threshold increases with the number of deficit voxels. But deficits that affect a relatively small proportion of a lobe are hard to detect with the ALV method, since the effect of the deficit is diluted by measurements over the rest of the unaffected lobe, which contribute only noise to the overall lobar volume estimate. Because the resulting ALV T-scores are so much lower than the VBM T-scores we expect VBM to outperform ALV for focal deficits (such as the simulated STG deficit), despite the fact that VBM needs more stringent statistical correction. In terms of statistical power, the distribution of mean ROI scores is shifted down towards the standard statistical threshold due to signal dilution as the size of the ROI becomes much larger than the spatial extent of the signal. We find theoretically (see Fig. 5) that this shift is generally much larger than the upward shift of the VBM threshold that is required by a Bonferroni-type correction as the number of voxels increases.

Our analysis with real data comparing 22q11DS-SZ patients to healthy controls suggests that false negative results can indeed occur when large scale ROI methods are used and the

underlying deficit is focal. This was demonstrated by the left supramarginal gyrus deficit, which was detected by VBM, but required a smaller ROI than the bilateral parietal lobe ROI used by ALV. Thus, the analyses on simulated data, our theoretical calculations and the analysis on real data point toward the same conclusion: our trade-off phenomenon entails a relative disadvantage for automated ROI compared to VBM because of a ‘dilution of effect’ in small scale GM deficits.

In this paper, we have focused on the relative benefit of VBM over ALV for the detection of fairly robust volume differences restricted to a relatively small spatial extent within a large ROI. Conversely, ALV would be expected to outperform VBM in the case of volume differences that extend fairly uniformly throughout a lobe but correspond to a small fractional difference in total volume, so that spatial averaging gives ALV a crucial advantage over sub-threshold VBM results. But above some reasonable fractional volume difference, even VBM can gain detection power from a spatially extended deficit. We have noted above that the distribution of the expected maximum *T*-score shifts to higher *T*-values with increasing spatial extent. Although we have considered FWE correction for VBM without the use of any additional spatial extent threshold, in practice such spatial cluster thresholding is often employed to improve detection of these kinds of deficits. And while more difficult to compare directly with statistical thresholding of ALV results, the increasingly popular method of controlling the false discovery rate within a large number of statistical tests should yield an improved sensitivity for VBM when there are a reasonable number of deficit voxels. So, while VBM with its smoothing kernel can be viewed as a multiple small ROI method, it is more sensitive to volume deficits that extend well beyond these kernel-defined ROIs than might be expected.

Since the ALV method does not provide information concerning spatial patterns of volume differences detected within each lobe, we have restricted our comparison to the relative sensitivity of ALV and VBM to detecting the presence of a deficit within a large pre-specified region. While VBM does give additional spatial information, it is important to note that, due to imperfect sensitivity at the voxel level, it can give quite a misleading picture of the true spatial extent, as demonstrated in Fig. 3c. However, performing VBM cluster-analysis with a reduced height threshold would improve basic detection and accuracy of spatial mapping.

Evaluating VBM and ALV on the basis of simulated deficits, as we have done here, has the advantage of giving ‘ground truth’ information about underlying deficits. Therefore, the relative sensitivity of both techniques can be compared meaningfully. This was our primary goal. However, the disadvantage is that simulations bring some form of artificiality to the analysis. This is a real limitation of our study.

Our simulations were realistic in the sense that normal sample variances were preserved. Deficits contained remnants of the signal intensity variance present in the original anatomical images. Moreover, the simulated deficits had realistic spatial variance because the inter-individual variability in cortical shapes was preserved. By splitting our control group into two groups for the simulation study, each of these groups only contained 7 subjects. This would be inadequate for a conventional study, but for the purposes of this

simulation study, we only needed to create some realistic inter-individual variability, since we could vary the statistical power over a wide range by introducing deficits of different geometry and spatial extent. However, the geometry of the cube-shaped simulated GM deficits did not represent realistic disease states. Note that we only use cortical GM images for analyses, so making a cube sized deficit in the pre-segmented images is similar to performing total cortical erosion in the area constrained by the cube.

Compared to the cubical deficits, the geometries of the STG shaped cortical deficit and lobar LTL cortical thinning deficits were more analogous to real anatomical anomalies found in neuropsychiatric patients. Indeed, several studies discovered cortical thinning of approximately 0.5–3 mm in schizophrenia patients in post-mortem studies (Benes et al., 2001) and in-vivo with MR-based cortical thickness measurement techniques (Gogtay et al., 2007; Kuperberg et al., 2003). This seems in good correspondence to our erosions of 1–3 voxels spanning a range of 0.78–4.5 mm.

Conversely, VBM's sensitivity to focal deficits relies on those deficits being located in similar regions of the brain. However, in patients, there may be variability in location of deficits due to variability in functional localization between different subjects. Our simulations did not incorporate such functional variability.

It has been suggested previously that as a general rule, VBM results should be compared to a reference standard, i.e. manually traced ROI volumetry measurements (Kubicki et al., 2002). Therefore, numerous studies have attempted to validate VBM with manual ROI volumetry. On several occasions, VBM showed discrepant results in terms of magnitude or location of significant inter-group brain differences compared to manual ROI techniques (Giuliani et al., 2005; Good et al., 2002; Kennedy et al., 2009; Kubicki et al., 2002; Testa et al., 2004; van Amelsvoort et al., 2004).

VBM entails normalization to a standard brain space, while manual ROI methods do not. Therefore, up until now, most researchers have argued that VBM is always inferior to manual ROI methods and that differences are due to errors in VBM; either in VBM pre-processing, such as systematic registration errors introduced during normalization (Ashburner and Friston, 2001; Bookstein, 2001). or the relative insensitivity of VBM to detect differences in brain regions with high local variability (Bookstein, 2001; Tisserand et al., 2002). While the overall degree of concordance is encouraging given the inherent differences between the two methods, the historic attribution of the cause of discrepant results to VBM errors is the root of remaining concerns whether VBM is an adequate method for assessing brain anatomy.

The present study was not designed to add to such comparative studies. However, our study does indicate that significant sensitivity differences should be expected between VBM and ROI studies that have used relatively large ROIs such as lobar volumes, which may provide an alternative explanation for some of the inconsistencies observed in these comparative studies.

The VBM method was originally designed with smoothing as an essential step, which automatically incorporates an effective small ROI procedure centered on each voxel. While

this cannot correct for misregistered gyral patterns as well as a carefully executed manual ROI study may be able to, with a sufficiently large smoothing kernel the procedure does account for spatial variation in deficit location in essentially the same manner as an ROI study. We also note that in conventional ROI studies, deficits of relatively large spatial extent may be split between non-overlapping ROIs, giving rise to the signal dilution problem. VBM, however, can improve the detection sensitivity by its effective use of many overlapping ROIs.

Some of the remaining concerns about misregistration issues are being addressed by methods that attempt to register cortical areas across subjects onto standard flattened cortical sheets (Dale et al., 1999; Fischl et al., 1999). Given the desire to map cortical volume differences at an anatomically meaningful spatial resolution, whether on a standard 3D brain or a cortical sheet representation, the spatial scale trade-off with statistical correction remains an important issue.

Conclusion

Our results show that VBM is generally more sensitive than ALV at identifying focal cortical gray matter differences, in the regime of voxel signal-to-noise and spatial extent normally studied. This is explained by the relatively more deleterious effects of signal dilution (partial volume effect) when the ROI volume is considerably larger than the spatial extent of the signal, as compared to even the most stringent requirements of Bonferroni-type statistical correction for multiple voxel-wise tests in VBM. Moreover, VBM appears just as sensitive to global gray matter differences as ALV. This suggests that VBM is a better technique for evaluating structural brain differences compared to large scale automated ROI methods.

Acknowledgments

Financial support for this study was provided by the Canadian Institutes of Health Research (MOP-74631 and MOP-97800).

References

Andreescu C, Butters MA, Begley A, Rajji T, Wu M, Meltzer CC, Reynolds CF III, Aizenstein H. Gray matter changes in late life depression—a structural MRI analysis. Neuropsychopharmacology. 2007; 33:2566–2572. [PubMed: 18075490]

Ashburner J, Friston KJ. Why voxel-based morphometry should be used. NeuroImage. 2001; 14:1238–1243. [PubMed: 11707080]

Ashburner J, Friston KJ. Unified segmentation. NeuroImage. 2005; 26:839–851. [PubMed: 15955494]

Benes FM, Vincent SL, Todtenkopf M. The density of pyramidal and nonpyramidal neurons in anterior cingulate cortex of schizophrenic and bipolar subjects. Biol Psychiatry. 2001; 50:395–406. [PubMed: 11566156]

Bonilha L, Rorden C, Castellano G, Cendes F, Li LM. Voxel-based morphometry of the thalamus in patients with refractory medial temporal lobe epilepsy. NeuroImage. 2005; 25:1016–1021. [PubMed: 15809001]

Bookstein FL. “Voxel-based morphometry” should not be used with imperfectly registered images. NeuroImage. 2001; 14:1454–1462. [PubMed: 11707101]

Cantor JM, Kabani N, Christensen BK, Zipursky RB, Barbaree HE, Dickey R, Klassen PE, Mikulis DJ, Kuban ME, Blak T, Richards BA, Hanratty MK, Blanchard R. Cerebral white matter deficiencies in pedophilic men. J Psychiatr Res. 2008; 42:167–183. [PubMed: 18039544]

Chow EWC, Zipursky RB, Mikulis DJ, Bassett AS. Structural brain abnormalities in patients with schizophrenia and 22q11 deletion syndrome. Biol Psychiatry. 2002; 51:208–215. [PubMed: 11839363]

Cuadra MB, Cammoun L, Butz T, Cuisenaire O, Thiran JP. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Trans Med Imaging. 2005:1548–1565. [PubMed: 16350916]

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. segmentation and surface reconstruction. NeuroImage. 1999; 9:179–194. [PubMed: 9931268]

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: ii: inflation, flattening, and a surface-based coordinate system. NeuroImage. 1999; 9:195–207. [PubMed: 9931269]

Giuliani NR, Calhoun VD, Pearlson GD, Francis A, Buchanan RW. Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. Schizophr Res. 2005; 74:135–147. [PubMed: 15721994]

Gogtay N, Greenstein D, Lenane M, Clasen L, Sharp W, Gochman P, Butler P, Evans A, Rapoport J. Cortical brain development in nonpsychotic siblings of patients with childhood-onset schizophrenia. Arch Gen Psychiatry. 2007; 64:772–780. [PubMed: 17606811]

Good CD, Scahill RI, Fox NC, Ashburner J, Friston KJ, Chan D, Crum WR, Rossor MN, Frackowiak RSJ. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. NeuroImage. 2002; 17:29–46. [PubMed: 12482066]

Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. Am J Psychiatr. 2005; 162:2233–2245. [PubMed: 16330585]

Kennedy KM, Erickson KI, Rodrigue KM, Voss MW, Colcombe SJ, Kramer AF, Acker JD, Raz N. Age-related differences in regional brain volumes: A comparison of optimized voxel-based morphometry to manual volumetry. Neurobiology of Aging. 2009; 30 (10):1657–1676. [PubMed: 18276037]

Kubicki M, Shenton ME, Salisbury DF, Hirayasu Y, Kasai K, Kikinis R, Jolesz FA, McCarley RW. Voxel-based morphometric analysis of gray matter in first episode schizophrenia. NeuroImage. 2002; 17:1711–1719. [PubMed: 12498745]

Kuperberg GR, Broome MR, McGuire PK, David AS, Eddy M, Ozawa F, Goff D, West WC, Williams SCR, van der Kouwe AJW, Salat DH, Dale AM, Fischl B. Regionally localized thinning of the cerebral cortex in schizophrenia. Arch Gen Psychiatry. 2003; 60:878–888. [PubMed: 12963669]

Lancaster JL, Summerln JL, Rainey L, Freitas CS, Fox PT. The Talairach Daemon, a database server for Talairach atlas labels. NeuroImage. 1997:s633.

Lancaster JL, Woldorff MG, Parsons LM, et al. Automated Talairach atlas labels for functional brain mapping. Hum Brain Mapp. 2000; 10:120–131. [PubMed: 10912591]

Lopez-Garcia P, Aizenstein HJ, Snitz BE, Walter RP, Carter CS. Automated ROI-based brain parcellation analysis of frontal and temporal brain volumes in schizophrenia. Psychiatry Res Neuroimaging. 2006; 147:153–161. [PubMed: 16949259]

Maldjian J, Laurienti P, Burdette J, Kraft RA. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage. 2003:1233–1239. [PubMed: 12880848]

Maldjian J, Laurienti P, Burdette J. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. NeuroImage. 2004; 21:450–455. [PubMed: 14741682]

Moorhead TW, Job DE, Whalley HC, Sanderson TL, Johnstone EC, Lawrie SM. Voxel-based morphometry of comorbid schizophrenia and learning disability: analyses in normalized and native spaces using parametric and nonparametric statistical methods. NeuroImage. 2004; 22:188–202. [PubMed: 15110009]

Pernet C. 2005Re: GM/WM/CSF volume computation in SPM5b: http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind05&L=SPM&P=R387304&I=-3

Reiss, AL., Hennessey, J., Subramanium, B., Rubin, M., Beach, L. Brain Image. Kennedy Kruger Institute Neuroimaging Center; Baltimore: 1995.

Terribilli D, Schaufelberger MS, Duran FLS, Zanetti MV, Curiati PK, Menezes PR, Scazufca M, Amaro J, Leite CC, Busatto GF. Age-related gray matter volume changes in the brain during non-elderly adulthood. Neurobiology of Aging. in press.

Testa C, Laakso MP, Sabattoli F, Rossi R, Beltramello A, Soininen H, Frisoni GB. A comparison between the accuracy of voxel-based morphometry and hippocampal volumetry in Alzheimer's disease. J Magn Reson Imaging. 2004; 19:274–282. [PubMed: 14994294]

Tisserand DJ, Pruessner JC, Sanz Arigita EJ, van Boxtel MPJ, Evans AC, Jolles J, Uylings HBM. Regional frontal cortical volumes decrease differentially in aging: an MRI study to compare volumetric approaches and voxel-based morphometry. NeuroImage. 2002; 17:657–669. [PubMed: 12377141]

van Amelsvoort T, Daly E, Henry J, Robertson D, Ng V, Owen M, Murphy KC, Murphy DGM. Brain anatomy in adults with velocardiofacial syndrome with and without schizophrenia: preliminary results of a structural magnetic resonance imaging study. Arch Gen Psychiatry. 2004; 61:1085–1096. [PubMed: 15520356]

Yamasue H, Iwanami A, Hirayasu Y, Yamada H, Abe O, Kuroki N, Fukuda R, Tsujii K, Aoki S, Ohtomo K, Kato N, Kasai K. Localized volume reduction in prefrontal, temporolimbic, and paralimbic regions in schizophrenia: an MRI parcellation study. Psychiatry Res: Neuroimaging. 2004; 131:195–207. [PubMed: 15465289]

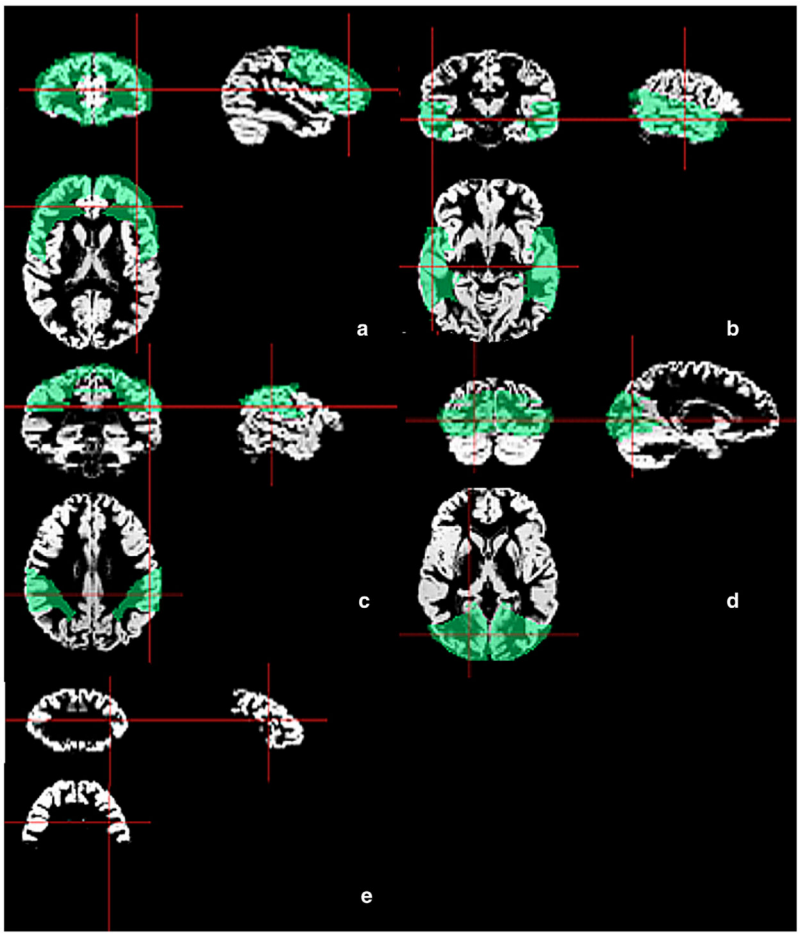

Fig. 1.

(a–d) Binary lobar masks (in green) used in automated lobar volumetry projected over the segmented gray matter images of a healthy subject. These masks were used in the ALV pre-processing to obtain lobar gray matter volumes of the frontal lobes (a), temporal lobes (b), parietal lobes (c) and occipital lobes (d). (e) Lobar gray matter image of a control subject obtained after multiplication of the subject's gray matter image with the frontal lobes binary mask.

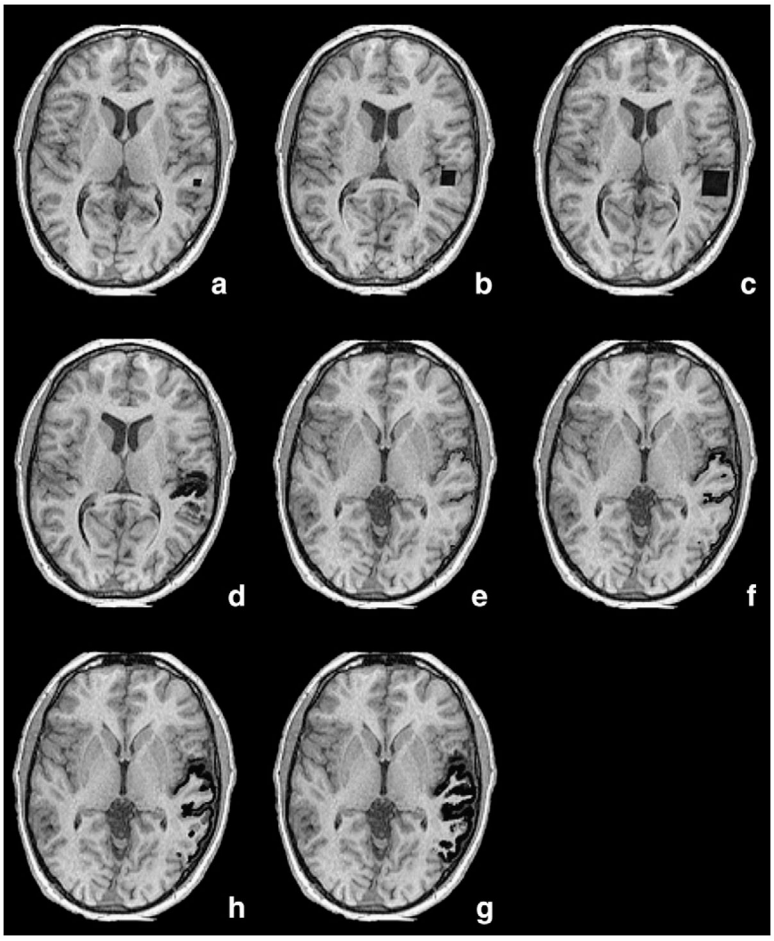

Fig. 2.

Anatomical MR images of one healthy subject in image set B warped back to original space after gray matter deficits were artificially introduced. (a–c) Cubical gray matter deficits of 6 mm^3 (a), 12 mm^3 (b) and 20 mm^3 (c) centered around voxel coordinates in the left superior temporal gyrus. (d) Cortical gray matter loss of the whole left superior temporal gyrus. (e–h) Left temporal lobe cortical thinning of one gray matter voxel (e), two gray matter voxels (f), of three gray matter voxels (g) and complete loss of cortical gray matter (h).

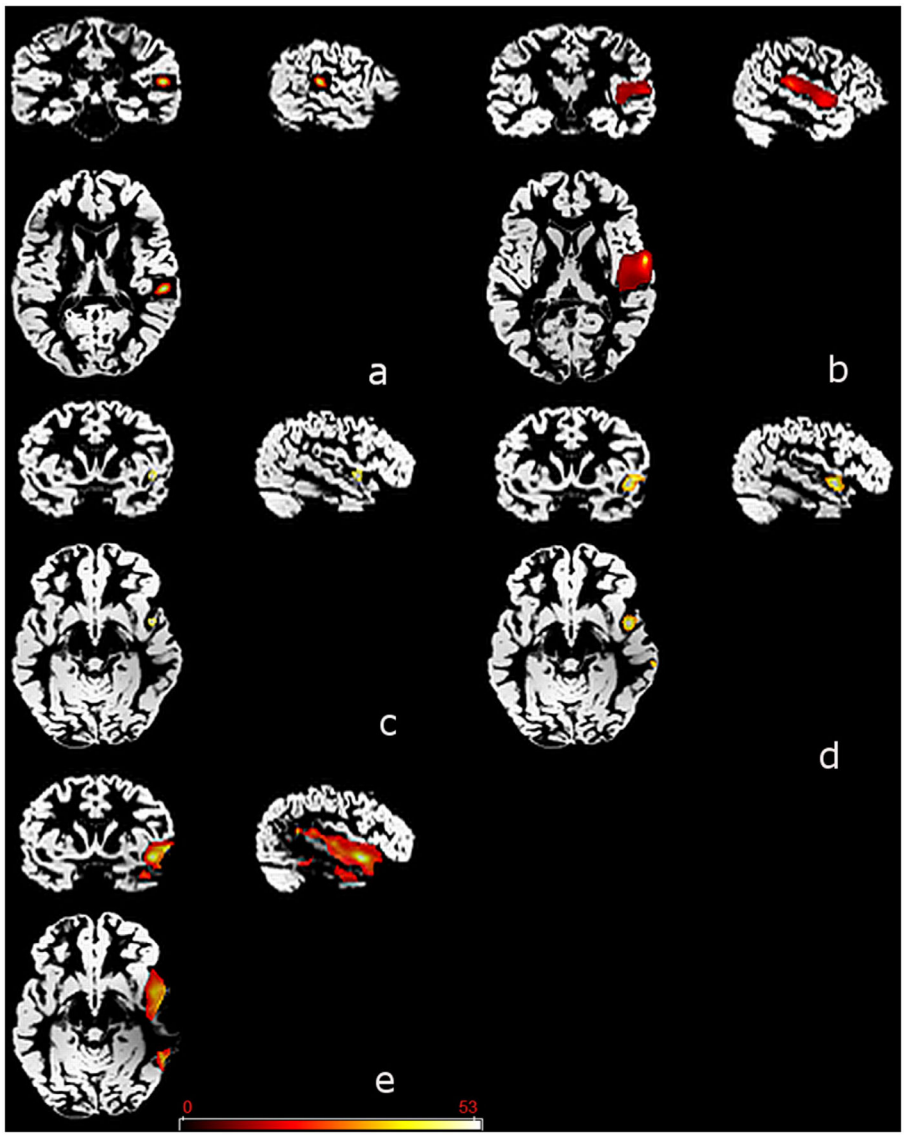

Fig. 3.

Thresholded (p_{fwe} =0.05) statistical parametric *T*-score maps generated by VBM analysis of image set S projected over a gray matter image of one of the subjects with artificial gray matter deficits. Brains are oriented with their right side depicted on the left side of each figure. (a) VBM accurately detected the simulated cubical 20 mm^3 GM deficit. (b) VBM analysis yielded a large significant cluster covering the left STG. (c) VBM showed the 2 voxel left temporal lobe cortical thinning as a focal cluster of GM deficit. (d) Two distinct clusters became present when the cortex was eroded by 3 voxels. (e) VBM generated clusters covering large parts of the left temporal lobe in the case all cortical GM was removed. Note the cluster voxels are color coded, red corresponding to lower *T*-scores (closer to the fwe-threshold) and yellow and white colors corresponding to higher *T*-scores.

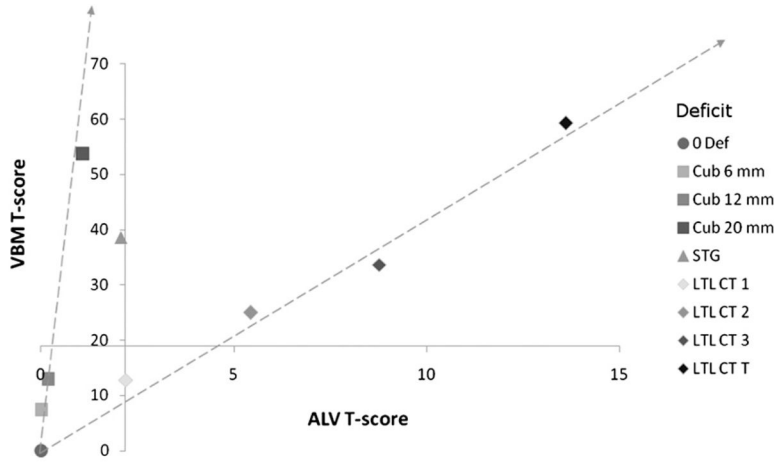

Fig. 4.

Graph of VBM *T*-scores plotted against ALV *T*-scores which were obtained during statistical analyses of the different types of simulated gray matter deficits in image set S. Abbreviations: 0 Def=no deficit; Cub=cubical GM deficits STG=superior temporal gyrus cortical deficit; LTLCT 1/2/3/T=left temporal lobe cortical thinning of 1,2,3 voxels or T=total GM deficit.

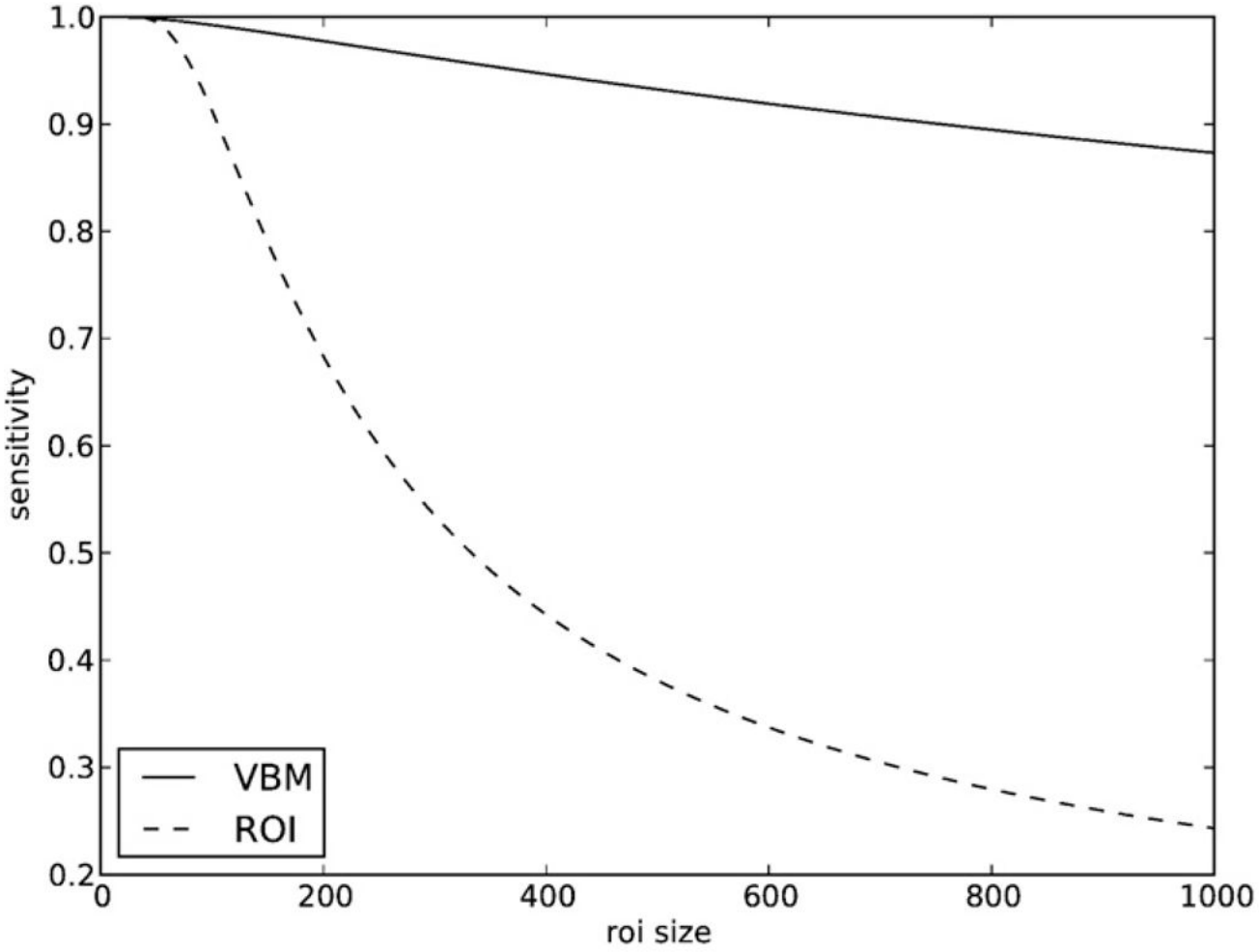

Fig. 5.

Theoretical sensitivities of detecting a focal deficit as a function of ROI size are shown for both ROI (dashed line) and VBM analyses (line) as calculated by our computer program. Sensitivities were calculated for deficits characterized by resel (SPM resolution elements; see Methods section for explanation) *z*-scores of 3.0 over a spatial extent of 10 resels.

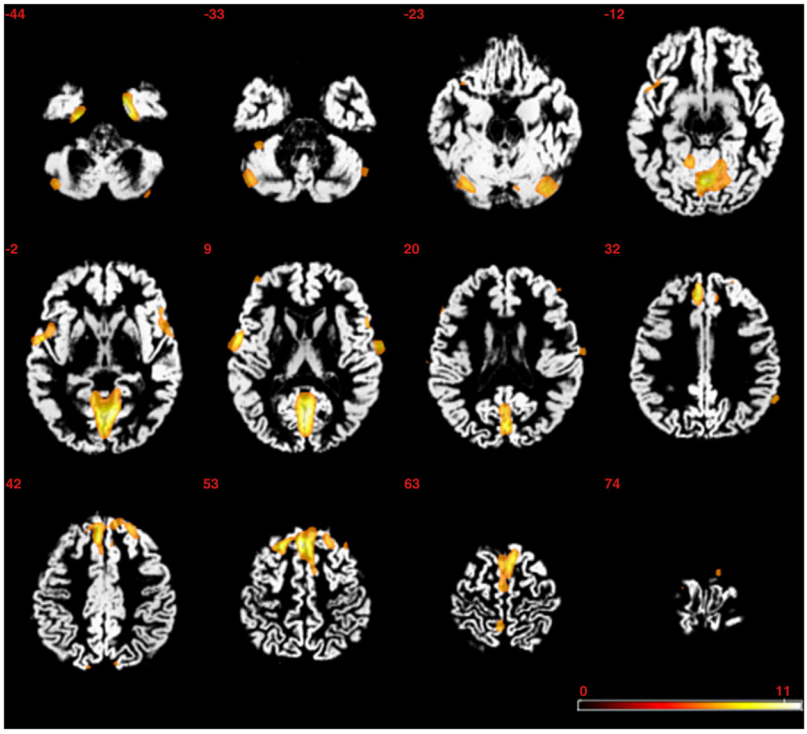

Fig. 6.

Graphic overview of the results from image set R comparing 22q11DS-SZ patients with healthy controls using VBM. A thresholded SPM-T is projected over a GM image of a healthy control showing relative GM deficit clusters in the 22q11DS-SZ group. In the left upper corner of each panel the *z*-coordinate (in mm) of each axial section is given in standard MNI space. The brains are oriented with their right side depicted on the left side of each panel. The colored bar in the left lower corner of the figure shows the color code for voxel *T*-score values. In panel −44 the GM deficits in the left and right limbic lobe and cerebellar posterior lobe deficits are shown. Panels −33, −22 and −12 show posterior and anterior lobe cerebellar deficits. In panels −2, 9 and 20 the deficits in the occipital lobe and left and right temporal lobes are shown. Also minor frontal lobe deficits are depicted. In panel 32 frontal lobe and left parietal lobe deficits are shown. Panels 42 to 74 show more frontal lobe deficits.

Table 1

Gray matter differences detected by VBM and ALV in image set R.

Brain	Lobe	Hemisphere	Gyrus	Brodmann	Tal-*x*	Tal-*y*	Tal-*z*	*T*-score VBM	*T*-score ALV
Cerebrum	Frontal lobe	L	Inferior frontal gyrus	47	−51	24	−1	7.65	3.46
		L	Precentral gyrus	44	−54	13	5	8.46	
		L	Superior frontal gyrus	6	−6	7	62	10.83	
		R	Inferior frontal gyrus	47	41	17	−13	6.93	
		R	–	45	59	21	13	7.08	
		R	–	45	55	26	20	6.91	
		R	Medial frontal gyrus	6	4	−7	63	10.35	
		R	Middle frontal gyrus	10	45	54	3	8.04	
		R	Paracentral lobule	5	5	−44	60	8.11	
		R	Precentral gyrus	6	58	−1	7	10.1	
		R	Superior frontal gyrus	8	3	29	48	11.44	
	Temporal lobe	L	Superior temporal gyrus	22	−67	−6	11	9.23	2.70
		L	–	22	−53	9	−5	7.59	
		R	Superior temporal gyrus	22	56	0	−3	8.85	
	Parietal lobe	L	Supramarginal gyrus	40	−61	−52	32	7.87	1.73
	Occipital lobe	L	Lingual gyrus	19	−9	−67	1	10.66	3.55
	Limbic lobe	L	Posterior cingulate	30	−4	−64	9	11.37	?
		L	Uncus	36	−20	−8	−37	11.86	
		R	Uncus	36	23	−14	−36	9.15	
Cerebellum	Posterior lobe	L	Inferior semi-lunar lobule	*	−48	−70	−39	8.27	?
		R	Declive	*	34	−75	−13	8.92	
		R	Inferior semi-lunar lobule	*	45	−73	−36	8.32	
		R	–	*	9	−71	−44	7.44	
		R	Tuber	*	49	−71	−23	7.65	
		R	–	*	51	−59	−27	7.15	
	Anterior lobe	L	Culmen	*	−1	−69	−4	10.71	?
		R	–	*	42	−37	−28	7.77	

Column “Brain” indicates whether the significant voxel is located in cerebrum or cerebellum. The column “Lobe” specifies in which brain lobe the significant voxel is located. Column “Hemisphere” indicates whether the cluster is in the right (“R”) or left (“L”) hemisphere. Column “Gyrus” states in which gyrus the voxel is located. Column “Brodmann” specifies the number of the Brodmann area the supra-threshold voxel is located in. The columns “Tal-*x*”, “Tal-*y*” and “Tal-*z*” indicate the Talairach coordinates of significant voxels. “VBM *T*-score” gives the *T*-score obtained in VBM analysis per significant voxel. “ALV *T*-score” gives the *T*-scores obtained in ALV analysis for the four brain lobes. Note an underlined *T*-score indicates that no significant difference between groups was found. Asterisks (“*”) mean no Brodmann areas that corresponded to the voxel coordinates could be identified. Question marks (“?”) indicate that the technique was unable to evaluate group differences in that specific anatomical area.

Neuroimage. Author manuscript; available in PMC 2011 June 30.