

Voxel-Based Morphometry of the Human Brain: Methods and Applications

Andrea Mechelli*, Cathy J. Price, Karl J. Friston, John Ashburner

Wellcome Department of Imaging Neuroscience, University College of London, 12 Queen Square, London, WC1N 3BG U.K.

Abstract: In recent years, a whole-brain unbiased objective technique, known as voxel-based morphometry (VBM), has been developed to characterise brain differences *in vivo* using structural magnetic resonance images. The present review provides a brief description of VBM and then focuses on exemplar applications in healthy and diseased subjects. The procedure involves normalising high-resolution structural magnetic resonance images to a standard template in stereotactic space. Normalised images are then segmented into gray and white matter and smoothed using an isotropic Gaussian kernel. Finally, a series of voxel-wise comparisons of gray and white matter in different groups of subjects are performed, using Random Field theory to correct for multiple comparisons. VBM has been useful in characterizing subtle changes in brain structure in a variety of diseases associated with neurological and psychiatric dysfunction. These include schizophrenia, developmental and congenital disorders, temporal lobe epilepsy and even cluster headache. In addition, VBM has been successful in identifying gross structural abnormalities, such as those observed in herpes simplex encephalitis, multiple sclerosis and Alzheimer's disease. Studies of normal subjects, on the other hand, have focussed on the impact of learning and practice on brain structure. These studies have led to the finding that environmental demands may be associated with changes in gray and white matter. For instance, it has been reported that the structure of the brain alters when human beings learn to navigate, read music, speak a second language and even perform a complex motor task such as juggling. We conclude the present review by discussing the potential limitations of the technique.

Keywords: Voxel-based morphometry, magnetic resonance imaging, statistical parametric mapping, gray matter, white matter.

INTRODUCTION

In recent years, a number of unbiased, objective techniques have been developed to characterize neuroanatomical differences *in vivo* using structural Magnetic Resonance images. These techniques can be broadly classified into those that deal with macroscopic differences in brain shape and those that examine the local composition of brain tissue after macroscopic differences have been discounted. The former, which include deformation-based morphometry (DBM), characterize the neuroanatomy of any individual brain in terms of deformation fields that map each brain to a standard reference. The latter, which include voxel-based morphometry (VBM), compare different brains on a voxel-by-voxel basis after the deformation fields have been used to spatially normalize the images [1]. One shared aspect of these techniques is that the entire brain, rather than a particular structure, can be examined in an unbiased and objective manner. The appropriate approach therefore depends on the types of structural difference that are expected among the images. Where there are global patterns of difference, multi-variate approaches such as DBM may be more powerful as they can model covariances between different structures. In contrast, mass uni-variate approaches such as VBM are likely to provide greater sensitivity for localizing small scale, regional differences in gray or white

matter. An additional difference between the two approaches relates to their computational cost. Specifically, DBM requires computationally expensive estimation of high resolution deformation fields that map each individual brain to a standard reference. In contrast, VBM requires the estimation of smooth, low frequency deformation fields and is therefore a simple and pragmatic approach within the capabilities of most research units.

In this review, we present a summary of the VBM methodology and in particular describe the processing steps required to characterize gray and white matter differences in volume and concentration. We then focus on a number of recent applications in diseased and healthy subjects and discuss their contribution to our understanding of brain organization. We conclude the present review by discussing the potential limitations of the technique, which need to be considered when planning an investigation and interpreting the results.

THE METHOD

The aim of VBM is to identify differences in the local composition of brain tissue, while discounting large scale differences in gross anatomy and position. This is achieved by spatially normalising all the structural images to the same stereotactic space, segmenting the normalised images into gray and white matter, smoothing the gray and white matter images and finally performing a statistical analysis to localize significant differences between two or more experimental groups. The output is a statistical parametric map (SPM) showing regions where gray or white matter

*Address correspondence to this author at the Wellcome Department of Imaging Neuroscience, 12 Queen Square London, WC1N 3BG, U.K.; Tel: 0044-(0)20-78337472; Fax: 0044-(0)20-78131420; E-mail: a.mechelli@fil.ion.ucl.ac.uk

differs significantly among the groups. There are several approaches (e.g. RAVENS; [2]) that could be used for the pre-processing of the data, but here we describe the steps that are routinely implemented using the SPM2 software [3].

Spatial Normalisation

Spatial normalisation involves registering the individual MRI images to the same template image. An ideal template consists of the average of a large number of MR images that have been registered in the same stereotactic space. In the SPM2 software, spatial normalisation is achieved in two steps. The first step involves estimating the optimum 12-parameter affine transformation that maps the individual MRI images to the template [4]. Here, a Bayesian framework is used to compute the maximum *a posteriori* estimate of the spatial transformation based on the *a priori* knowledge of the normal brain size variability. The second step accounts for global nonlinear shape differences, which are modeled by a linear combination of smooth spatial basis functions. This step involves estimating the coefficients of the basis functions that minimize the residual squared difference between the image and the template, while simultaneously maximizing the smoothness of the deformations. The ensuing spatially-normalised images should have a relatively high-resolution (1mm or 1.5mm isotropic voxels), so that the segmentation of gray and white matter (described in the next section) is not excessively confounded by partial volume effects, that arise when voxels contain a mixture of different tissue types.

It should be noted that spatial normalisation does not attempt to match every cortical feature exactly, but merely corrects for global brain shape differences. This is because VBM tries to detect differences in the local concentration or volume of gray and white matter having discounted global shape differences. Indeed, if the spatial normalisation was perfectly exact, all the segmented images would appear identical and no significant differences would be detected at a local scale.

Segmentation

The spatially normalised images are then segmented into gray matter, white matter, cerebrospinal fluid and three non-brain partitions. This is generally achieved by combining *a priori* probability maps or “Bayesian priors”, which encode the knowledge of the spatial distribution of different tissues in normal subjects, with a mixture model cluster analysis which identifies voxel intensity distributions of particular tissue types. The segmentation step also incorporates an image intensity non-uniformity correction [3] to account for smooth intensity variations caused by different positions of cranial structures within the MRI coil. A further possible step after segmentation would be the binarization of the resulting tissue class images. Tissue classification methods typically produce images where each voxel has an *a posteriori* probability that that voxel should be assigned to a particular tissue type according to the model. These probabilities are values between zero and one. Binarization would involve assigning each voxel to its most probable tissue class.

Smoothing

The segmented gray and white matter images are now smoothed by convolving with an isotropic Gaussian kernel. The size of the smoothing kernel should be comparable to the size of the expected regional differences between the groups of brains, but most studies have employed a 12-mm FWHM kernel. The motivation for smoothing the images before the statistical analysis is three-fold. First, smoothing ensures that each voxel in the images contains the average amount of gray or white matter from around the voxel (where the region around the voxel is defined by the smoothing kernel). Second, the smoothing step has the effect of rendering the data more normally distributed by the central limit theorem, thus increasing the validity of parametric statistical tests. Third, smoothing helps compensate for the inexact nature of the spatial normalisation. Smoothing also has the effect of reducing the effective number of statistical comparisons, thus making the correction for multiple comparisons less severe. However it may also reduce the accuracy of localization, as discussed in the last section of this paper.

Statistical Analysis

Following the pre-processing, the final step of a VBM analysis involves a voxel-wise statistical analysis. This employs the general linear model (GLM), a flexible framework that allows a variety of different statistical tests such as group comparisons and correlations with covariates of interest. The standard parametric procedures (t tests and F tests) used are valid providing that the residuals, after fitting the model, are normally distributed. If the statistical model is appropriate, the residuals are most likely to be normally distributed once the segmented images have been smoothed. The results of these standard parametric procedures are statistical parametric maps [5]. Since a statistical parametric map comprises the results of many voxel-wise statistical tests, it is necessary to correct for multiple comparisons when assessing the significance of an effect in any given voxel. A standard Bonferroni correction for multiple independent comparisons would be inappropriate here, given the smoothing and the fact that gray or white matter in contiguous voxels is highly correlated. Thus, corrections for multiple dependent comparisons are made using the Theory of Random Fields [6–8]. It should be noted that the Random Field correction should be based on the local maxima of the *t* statistic rather than the extent statistic which relates to the size of the clusters. This is because, for a correction based on the extent statistic to be valid, the smoothness of the residuals needs to be spatially invariant throughout the brain. However, this is unlikely to be the case in VBM studies by virtue of the highly non-stationary nature of the underlying neuroanatomy. For example, by chance alone, large size clusters will occur in regions where the images are very smooth and small size clusters will occur in regions where the images are very rough.

Standard & Optimised Pre-Processing

There are cases when structural differences, not related directly to gray or white matter volumes, may be misinterpreted as volumetric differences as a result of

normalisation. One example is when the size of the ventricles differs significantly between two or more experimental groups. If the ventricles of one experimental group are enlarged during normalisation, the surrounding gray and white matter also may be enlarged. This is because the parameters of the normalisation only encode highly smooth, low frequency deformations which may not distinguish between the ventricles and the surrounding tissue. As a result, structural differences pertaining to ventricular volume may show up in a VBM study of gray matter volumes. A way of minimizing this potential source of error is to perform the normalisation using the segmented gray and white matter volumes rather than on the whole-brain images. If all the data entering into the statistical analysis are only derived from gray matter, then any significant differences must be due to gray matter. Likewise, if all the data entering into the statistical analysis are derived only from white matter, then any significant differences must be due to white matter changes. The caveat with this approach, however, would be that the segmentation has to be performed on images in native space. However the Bayesian priors, which encode *a priori* knowledge about the spatial distribution of different tissues in normal subjects, are in stereotactic space. A way of circumventing this problem is to use an iterative version of segmentation and normalisation operators, (see Fig. 1). First, the original structural MRI images in native space are segmented. The resulting gray and white matter images are then spatially normalized to gray and white matter templates respectively to derive the

optimized normalisation parameters. These parameters are then applied to the original, whole-brain structural images in native space prior to a new segmentation. This recursive procedure, also known as “optimized VBM”, has the effect of reducing the misinterpretation of significant differences relative to “standard VBM” [9].

Concentration or Volume Differences?

In both standard and optimized pre-processing, a further processing step can be used to compensate for the effect of spatial normalisation. When warping a series of individual images to match a template, volumetric differences are likely to be introduced. For example, if one subject's temporal lobe has half the volume of that of the template, then its volume will be doubled. As a result, the subject's temporal lobe will comprise twice as many voxels after spatial normalisation and the information about the absolute volume of this region will be lost. In this case, VBM can be thought of as comparing the relative *concentration* of gray or white matter structures in the spatially normalized images (i.e. the proportion of gray or white matter to all tissue types within a region). There are cases, however, when the objective of the study is to identify regional differences in the *volume* of a particular tissue (gray or white matter), which requires the information about absolute volumes to be preserved. Here a further processing step, which is usually referred to as “modulation”, can be incorporated to compensate for the effect of spatial normalisation. This step involves multiplying the spatially normalised gray matter (or other

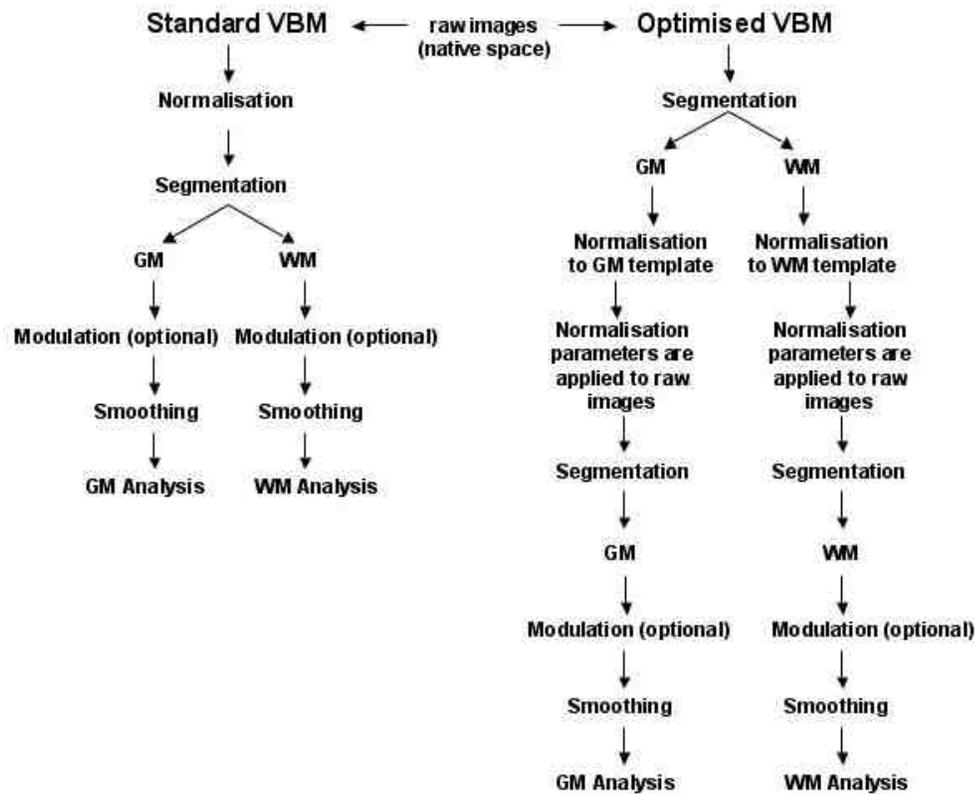


Fig. (1). Flow diagram of the preprocessing steps in standard (left) and optimized (right) VBM. GM = gray matter images; WM = white matter images.

tissue class) by its relative volume before and after spatial normalisation. For instance, if spatial normalisation results in a subject's temporal lobe doubling its volume, then the correction will halve the intensity of the signal in this region. This ensures that the total amount of gray matter in the subject's temporal lobe is the same before and after spatial normalisation.

In short, the multiplication of the spatially normalised gray matter (or other tissue class) by its relative volume before and after warping has critical implications for the interpretation of what VBM is actually testing for. Without this adjustment, VBM can be thought of as comparing the relative concentration of gray or white matter structures in the spatially normalized images. With the adjustment, VBM can be thought of as comparing the absolute volume of gray or white matter structures. The two approaches are known as “non-modulated” and “modulated” VBM, respectively [9].

Global or Localized Inferences?

Uniformly bigger brain will have uniformly higher gray and white matter intensities. The critical implication here is that any detected differences in gray or white matter volume are likely to encompass almost all brain regions. However, in most studies the objective is to identify regionally specific changes in tissue composition that are not confounded by global differences. This can be achieved by modeling “global” measures as confounding effects during the statistical analyses. For instance, when investigating differences in gray or white matter volumes between populations, one could include the total amount of gray or white matter in each brain or the intra-cranial volume of each subject as a covariate of no interest. Previous studies have also used the height or the weight of the subjects as indices of global gray and white matter volumes. Alternatively, one could perform a global scaling of all images prior to the statistical analysis in order to discard overall differences between populations. The different approaches which can be used for modeling “global” measures as confounding effects are likely to account for some of the conflicting results in the literature. In short, inferences about gray and white matter volumes may be dependent on global differences or, alternatively, be based on regionally specific changes. Both types of inferences are valid, and the optimal approach simply depends on the research hypothesis that motivated the study, (see Fig. 2).

To summarize, voxel-based morphometry (VBM) involves a voxel-wise comparison of the local gray and white matter between two groups of subjects. The procedure involves spatially normalising high-resolution MR images, from all subjects, into the same stereotactic space. This is followed by segmenting the gray and white matter volumes from the spatially normalised images, and smoothing those segments. Voxel-wise parametric statistical tests are performed which compare the smoothed gray and white matter images from the groups. Corrections for multiple dependent comparisons are made using the theory of Gaussian random fields. There are several types of analysis that one can perform to address questions pertaining to gray and white matter differences. For instance, “non-modulated” VBM identifies differences in the relative concentration or

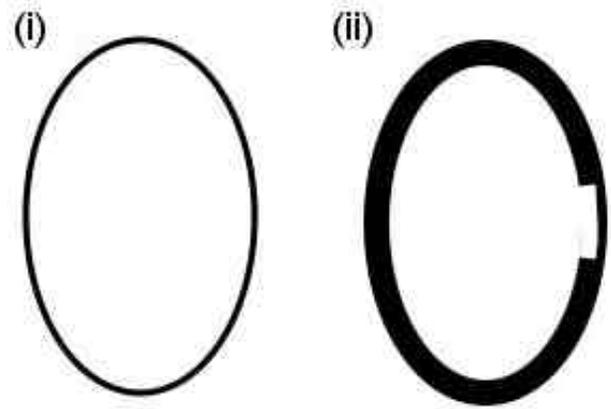


Fig. (2). Global and localized inferences may identify different brain regions. In the example, images (i) and (ii) differ at both global and local levels. Without modelling some kind of “global” value as a confounding effect, the comparison would identify greater volume for image (ii) relative to image (i) in all voxels apart from the thinner area on the right side of image (ii). When global differences are discounted, however, the same comparison will detect greater volume for image (i) relative to image (ii) in the thinner area on the right side of image (ii) only.

density of gray or white matter (i.e., the proportion of gray or white matter relative to other tissue types within a region); “modulated” VBM identifies differences in volume (i.e., the absolute amount of gray or white matter in different regions). The statistical analysis may or may not include some kind of adjustment for global differences, depending on whether one is interested in global or regionally-specific neuroanatomical differences. The bottom line here is that all these approaches are valid, but address different questions and the approach used should be dictated by the experimental question of the researcher.

APPLICATION TO DISEASED SUBJECTS

In recent years, VBM studies have been successful in characterizing structural brain differences in a variety of diseases including schizophrenia [10], developmental and congenital disorders [11], autism [12], Kallman’s syndrome [13], Klinefelter’s syndrome [14], bipolar disorders [15], cluster headache [16], temporal lobe epilepsy [17], supranuclear palsy [18], Down’s syndrome [19], herpes simplex encephalitis [20], Parkinson’s disease [18], Huntington’s disease [21], Alzheimer’s disease [22] and primary progressive aphasia [23]. Here we review two exemplar studies which provide an opportunity to discuss relevant methodological issues.

While VBM was originally devised to identify subtle neuroanatomical changes associated with neurological and psychiatric dysfunction, the method has also been used to examine gross structural abnormalities. The use of VBM with highly distorted brains presents special challenges however, due to the difficulties that arise during spatial normalisation. Gitelman *et al.* (2001) [20] compared gray matter concentration in five patients with gross structural abnormalities after recovering from herpes simplex encephalitis with five matched control groups. Because

herpes simplex encephalitis is well known to affect specific limbic regions, the study allowed the evaluation of standard VBM with highly atypical brains. VBM identified extensive limbic and paralimbic anatomical changes in patients with herpes simplex encephalitis relative to the controls, consistent with the findings of previous pathology studies [24]. Thus, the method was able to identify anatomical differences between groups even in the context of highly distorted brains. However, after normalisation there were residual macroscopic differences between the two groups and in particular the ventricles were larger in the patients relative to the controls. These differences were not reduced when adding more basis functions and reducing the smoothness of the deformation fields during spatial normalisation. In addition, the segmentation was sub-optimal in the caudate nucleus most probably as a result of group differences in the size of the ventricles after spatial normalisation, (see Fig. 3). Reduced gray matter was detected in the caudate nucleus of patients relative to controls, but it was unclear whether this was due to volumetric differences *per se* or to a displacement of brain structures surrounding the ventricles. The study by Gitelman *et al.* (2001) [20] illustrates the importance of considering the interactions between brain morphology and analytical

methods. In this instance VBM detected caudate differences but the attribution of these changes to gray matter loss *per se* would have been an improper interpretation. The cause of the difference probably lay at a macroscopic level and was expressed, after normalisation and segmentation, in a specific region. Unfortunately, there is no simple metric for evaluating the validity of the normalisation and segmentation procedures. The assessment of these procedures relies on visual inspection after normalisation and segmentation respectively. A number of solutions have been proposed to deal with highly distorted brain. These include “masking out” the abnormal region in the patient’s brain before performing the spatial normalisation based on the remaining structures, and using additional basis functions or minimization algorithms [3, 25, 26].

Of particular interest, from a methodological point of view, is a recent investigation of temporal lobe epilepsy by Keller *et al.* (2004) [17]. The authors used both standard and optimised VBM and furthermore compared “modulated” and “non-modulated” procedures. The impact of the neuroanatomical template was also explored, by using a default template from an independent sample of 148 healthy subjects and a subject-specific, customised template. Because a previous manual volumetric analysis of the same

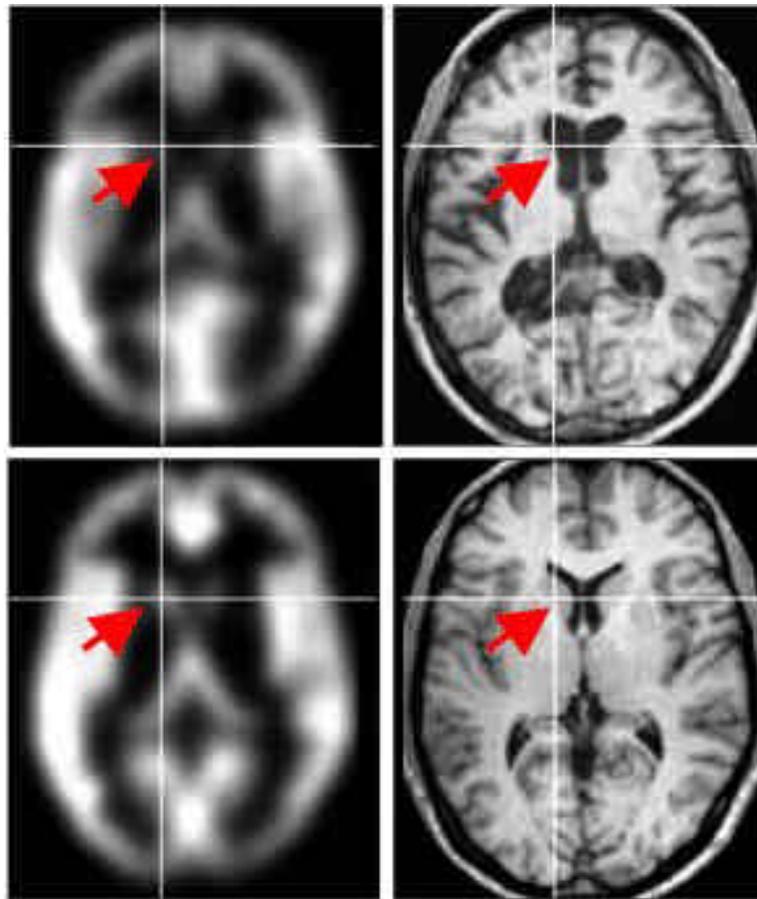


Fig. (3). Smoothed gray matter segments and corresponding anatomical slices from one patient (top) and one control subject (bottom). Note that the estimation of the left caudate gray matter is reduced in the patient as indicated by fewer white pixels at the arrow in the upper left compared to the lower left image. In this case, slight mismatches between the position of the caudate in the normalized patient image and its position in the probability interfered with segmentation. Modified from [20].

group of patients revealed hippocampal atrophy, the study provided an opportunity to examine the anatomical face validity of alternative VBM approaches. All analyses correctly detected hippocampal atrophy with a statistical threshold of $p < 0.05$ (corrected for multiple comparisons). However, the distribution of both hippocampal and extrahippocampal gray matter loss differed significantly between standard and optimised VBM and between “modulated” and “non-modulated” VBM, irrespective of the neuroanatomical template (default vs. customised). Standard VBM without modulation identified reduced gray matter in a distributed set of extrahippocampal regions that were not classified as abnormal in previous region-of-interest and histopathological studies. In contrast, optimised VBM with modulation detected extrahippocampal abnormalities that replicated previous findings. Finally, gray matter differences identified using optimised VBM without modulation were confined to sub-regions of the hippocampus. These results illustrate three important methodological points. First, in analyses of gray matter *concentration*, standard VBM may be more sensitive to subtle brain abnormalities than optimised VBM. This is due to the use of tissue-specific templates in optimised VBM, which ensures that the individual gray and white matter images are more closely matched to their templates. As a result, differences between patient and control groups are minimised without any correction for nonlinear warping (i.e. “modulation”). Second, non-modulated and modulated VBM detect differences in concentration and volume respectively and, therefore, may identify significant effects in different regions. Third, the difference between standard and optimised VBM, and between “modulated” and “non-modulated” procedures, may account for some of the inconsistencies in the literature.

APPLICATION TO HEALTHY SUBJECTS

Although VBM was originally devised to examine structural abnormalities in patients, the technique can also be used with healthy subjects. Here, we focus on some of the recent VBM studies that investigated the impact of learning and practice on brain structure [27-31]. These studies were motivated by reports that experience-related structural changes may occur in the brain of small mammals and birds [32-36].

Maguire *et al.* (2000) [27] first investigated the impact of learning and practice on brain structure. The authors used VBM to test whether structural changes could be detected in the brain of London taxi drivers as a result of extensive experience of spatial navigation. The posterior hippocampi of taxi drivers were significantly larger relative to those of healthy controls, while the anterior hippocampi were larger in controls than taxi drivers. Furthermore, hippocampal volume correlated with the amount of time spent as a taxi driver (positively in the posterior and negatively in the anterior hippocampus). These results suggest that the posterior hippocampi expand regionally in individuals who have extensive experience of spatial navigation. An alternative hypothesis, however, is that the difference in the hippocampal volume is associated with innate navigational expertise, leading to an increased likelihood of becoming a taxi driver.

A recent study by Draganski *et al.* (2004) [30] sheds light on this issue, by demonstrating that the acquisition of new skills may indeed change neuroanatomy. Brain scans were acquired from healthy subjects before they learnt a classic three-ball cascade juggling routine and 3 months later when they had become skilled performers. The comparison of the scans acquired before and after practice revealed an expansion in gray matter in bilateral mid-temporal areas and left posterior intra-parietal sulcus. These findings were specific to the training stimulus, as a group of controls showed no changes in gray matter over the same period. Because jugglers and controls were randomly divided by the experimenters, the results can be confidently interpreted in terms of structural changes induced by training rather than genetic predisposition. The authors also report a decrease in the expansion of gray matter once the jugglers stopped practicing for 3 months. This suggests that the effect of training on brain structure may be transient.

A recent study extends these findings by showing that structural changes occur in the human brain in response to second language acquisition, and that the degree of structural reorganization depends on the age of acquisition and the proficiency attained [31]. To test differences in gray and white matter volume between bilingual and monolingual English subjects, we recruited 25 monolinguals; 25 early bilinguals who had learnt a second European language before the age of 5; and 33 late bilinguals who had learnt a second European language between the ages of 10 and 15. Increased gray matter volume was identified in the left inferior parietal cortex of bilinguals relative to monolinguals, with greater effects in early bilinguals than in late bilinguals, (see Fig. 4a). We next investigated the relationship between brain structure and second language proficiency and age of acquisition. For this purpose, we recruited 22 medically normal right-handed native Italian speakers who had learnt English as a second language between the ages of 2 and 34 years. Remarkably, gray matter volume correlated with proficiency in exactly the same left inferior parietal region identified in the first study, (see Fig. 4b). Furthermore, gray matter volume correlated negatively with the age of acquisition of the second language in the same region, (see Fig. 4c). These effects could result from structural reorganization induced by second language acquisition, or from genetic predisposition. However, early bilinguals are more likely to have learnt a second language due to social circumstances rather than genetic predisposition. Thus, we interpret our findings in terms of structural plasticity in the bilingual brain which is dependent on age of acquisition and the proficiency attained. This study illustrates two complementary applications of VBM, which can be used to compare two or more experimental groups and to investigate the dependency between brain structure and one or more variables of interest.

To summarize, VBM studies of healthy subjects have challenged the traditional view that the acquisition of new skills only changes the way the brain functions, by showing structural changes at the macroscopic level. These studies have also suggested that the gray matter changes induced by training may be transient and dependent on the age of the subjects and the performance achieved. Future studies will

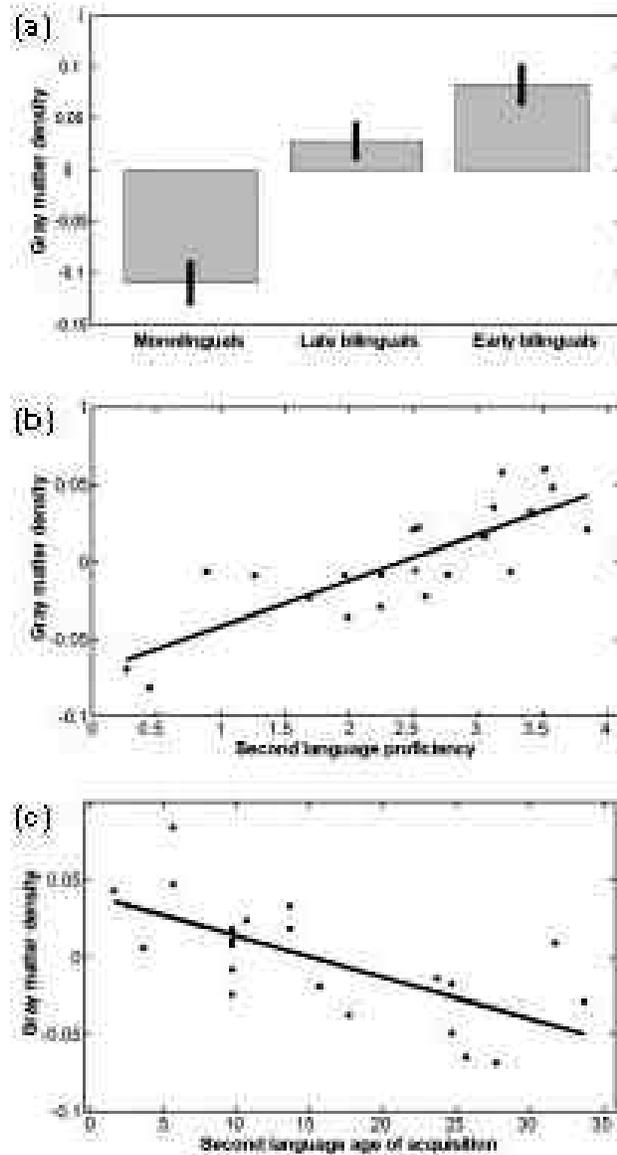


Fig. (4). Gray matter density, measured as cubic millimeters of gray matter per voxel, in the left inferior parietal cortex (co-ordinates: $x = -45$ $y = -59$ $z = 48$). (a) Gray matter density in monolinguals, late bilinguals and early bilinguals. Error bars indicate standard deviation. (b) Gray matter density as a function of second language proficiency. (c) Gray matter density as a function of second language age of acquisition. All values are mean corrected. Modified from [31].

need to better characterize the relationship between brain structure, age of acquisition, performance and amount of practice, across verbal and non-verbal learning domains. Furthermore, methods other than whole-brain MRI will be required to investigate whether the gray-matter increases induced by the acquisition of new skills are related to changes in neuropil, neuronal size, dendritic or axonal arborisation.

LIMITATIONS OF VBM

The exemplar studies discussed above suggest that there are potential confounds and limitations to be considered when planning a study and interpreting the results. For instance, all raw original structural images need to be acquired on the same scanner with identical imaging parameters. This is because any significant effects might be attributable to scanner or MR sequence differences rather than to the subjects themselves. Other potential issues relate to the difficulty of spatially normalising atypical brains, the robustness of standard parametric tests and the interpretation of the results. We discuss these and related issues below.

Normalisation and Segmentation Confounds

An objection to VBM is that specific patterns of abnormal anatomy may result in group-specific misregistration. As a result, VBM will be sensitive to systematic shape differences attributable to misregistration from the spatial normalisation step [37]. An example is given by the effects in the caudate nucleus reported by Gitelman *et al.* (2001) [20] that may result from group differences in the size of the ventricles. Furthermore, a particular experimental group may move more in the scanner, so the resulting images contain motion artifact. This motion may interact with the segmentation to produce systematic classification differences. These are some of several potential systematic differences that can arise between patients and controls that may be detected by VBM [38]. They reflect a real difference between data obtained from different populations, but may not necessarily be due to changes in gray or white matter.

Pre-processing of Atypical Brains

The investigation of brains containing severe pathologies may also be problematic, due to the difficulties that arise when spatially normalizing and segmenting atypical brains. For example, if a brain contains features that are not present in the template image, then an accurate match can not be achieved. The effects of this mismatch may also propagate to other brain regions because of the inherent smoothness of the deformation fields. In addition, the segmentation model only allows for gray matter, white matter and cerebrospinal fluid within the brain and can not model atypical tissue types found in pathology, such as tumors and artero-venous malformations. Possible solutions here consists in (i) "masking out" the abnormal region in the patient's brain before spatial normalisation, (ii) smoothing the nonlinear deformation fields to minimize displacements, (iii) using signal from the skull and other tissues outside the brain to constrain the boundaries of normalisation and (iv) employing additional basis functions or minimization algorithms. Further work is necessary in order to develop warping methods that can model the various forms of severe pathology that may be encountered.

Accuracy of Localization

VBM uses a relatively simple warping method, which only attempts to register the brain images "globally". However, one of the main disadvantages of resorting to a method that uses a less precise image registration, is that any regional volumetric differences cannot be localized

accurately. Furthermore, with the current spatial normalisation, a large anatomical difference in one part of the brain can cause a systematic relative translation in another part. When precise localization is critical, therefore, nonlinear registration methods that estimate an exact match between brain structures using a very high number of parameters may be preferable. Fully optimizing a model of this order, however, takes a very long time, and is susceptible to local minima, which, in turn, depend upon starting estimates. Smoothing can also affect the accuracy of localization, by shifting the peak of the SPM slightly in a direction toward regions of lower variance. This shift may sometimes lead to significant effects in voxels outside the brain, where the variance tends to be lowest.

Validity of Statistical Inferences

The standard parametric procedures (t tests and F tests) typically used in VBM are valid providing that the residuals, after fitting the model, are normally distributed. This raises the possibility that non-normality in the error terms can make statistical inferences invalid in some VBM studies. Salmond *et al.* (2002) [39] have shown that in balanced designs, provided the data are smoothed with a 4 mm FWHM kernel, any non-normality in the data is sufficiently attenuated to render the tests valid. However in unbalanced designs, such as the comparison of a single patient with a control group, a significant number of false positives arise when smoothing at 4 and 8 mm. Thus, in addition to being generally less sensitive, unbalanced designs appear to be less robust to violations of normality than balanced design. One possible solution here involves smoothing the data to a greater extent (e.g. 12 mm) in order to ensure the residuals to conform to the normality assumption by central limit theorem. Alternatively, it may be necessary to resort to non-parametric methods which do not require the residuals to be normally distributed.

Sensitivity to Distributed, Non-linear Differences

As recently pointed out by Davatzikos (2004) [40], mass uni-variate approaches such as VBM are biased toward group differences that are localized in space. However, pathology and experience-related changes can also be expressed, anatomically, in a spatially complex and subtle fashion. For instance, their expression in one part of the brain may depend on their expression elsewhere. The

characterization of these inter-regional dependencies requires a multi-variate approach. VBM is not a surrogate for multi-variate volumetric analyses that have been specifically developed to characterize highly distributed, non-linear changes. However, as discussed by Friston and Ashburner (2004) [41], there are ways of using VBM to assess inter-regional dependencies and accommodate nonlinearities.

Interpretation.

The nature of gray and white matter changes identified with VBM is still poorly understood especially in healthy subjects. In studies of Alzheimer's dementia and other degenerative disorders, a neuronal loss is likely to be the primary cause of the gray and white matter changes observed in VBM. However, the interpretation of gray and white matter differences may be problematic when subtle changes in brain structure are observed in association with neurological and psychiatric dysfunction. For instance, the average amount of gray or white matter from around the voxel is often referred to as gray or white matter "density" or "concentration", but should not be confused with cell packing density measured cytoarchitectonically. Furthermore, it is unclear whether the experience-dependent changes in gray and white matter observed in healthy subjects are related to changes in neuropil, neuronal size, dendritic or axonal arborisation. In the neocortex, increased gray matter volume could also result from more folding as well as thicker gray matter, (see Fig. 5). This issue is of great interest given the growing number of VBM studies being published but can only be addressed by methods other than MRI.

We have presented a summary of the VBM methodology and discussed a number of potential limitations which need to be considered when planning an investigation and interpreting the results. VBM studies of structural abnormalities may inform the interpretation of data acquired with other modalities such as Positron Emission Tomography (PET) and Single Photon-Emission Tomography (SPECT). If the amount of uptake per unit of gray matter is the same in the patient and the control group, then the differences observed may not be directly related to glucose utilization per unit of gray matter but, rather, units of gray matter *per se* [42]. Thus, some of the PET/SPECT studies comparing patients and controls may simply be



Fig. (5). Differences in gray matter volume may result from more folding or thicker gray matter. In this example, exactly the same differences would be detected when comparing images of thin, un-folded cortex (left circle) against thin, folded cortex (middle circle) and thick, unfolded cortex (right circle).

tapping into the same differences that a VBM investigation would detect. We envisage that, in future years, VBM will be combined with functional approaches such as functional MRI and EEG/MEG to better characterize brain function-structure relationships in health and disease.

ACKNOWLEDGEMENTS

AM is supported by MH64445 from the National Institutes of Health (USA). CP, KJF and JA are supported by The Wellcome Trust.

REFERENCES

- [1] Wright IC, McGuire PK, Poline JB, *et al.* A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage* 1995; 2: 244–252.
- [2] Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* 2001; 14: 1361–1369.
- [3] Ashburner J, Friston KJ. Voxel-based morphometry: the Methods. *NeuroImage* 2000; 11: 805–821.
- [4] Ashburner J, Friston KJ. Multimodal image coregistration and partitioning – A unified framework. *NeuroImage* 1997; 6: 209–217.
- [5] Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 1995; 2: 189–210.
- [6] Friston KJ, Holmes AP, Poline JB, Price CJ, Frith CD. Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage* 1995; 4: 223–235.
- [7] Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant voxels in images of cerebral activation. *Hum Brain Mapp* 1996; 4: 58–73.
- [8] Worsley KJ, Andermann M, Koulis T, MacDonald D, Evans AC. Detecting changes in non-isotropic images. *Hum Brain Mapp* 1999; 8: 98–101.
- [9] Good CD, Johnsrude I, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage* 2001; 14: 21–36.
- [10] Kubicki M, Shenton ME, Salisbury DF, *et al.* Voxel-Based Morphometric Analysis of Gray Matter in First Episode Schizophrenia. *NeuroImage* 2002; 17: 1711–1719.
- [11] Abell F, Krams M, Ashburner J, *et al.* The neuroanatomy of autism: a voxel-based whole brain analysis of structural scans. *NeuroReport* 1999; 10: 1647–1651.
- [12] Boddaert N, Chabane N, Gervais H, *et al.* Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *NeuroImage* 2004; 23: 364–369.
- [13] Krams M, Quinton R, Ashburner J, *et al.* Kallman's syndrome: Mirror movements associated with bilateral corticospinal tract hypertrophy. *Neurology* 1999; 52: 816–822.
- [14] Shen D, Liu D, Liu H, Clasen L, Giedd J, Davatzikos C. Automated morphometric study of brain variation in XXY males. *NeuroImage* 2004; 23: 648–653.
- [15] Lyoo IK, Kim MJ, Stoll AL, *et al.* Frontal lobe gray matter density decreases in bipolar I disorder. *Biol Psychiatry* 2004; 55: 648–51.
- [16] May A, Ashburner A, Buchel C, *et al.* Correlation between structural and functional changes in brain in an idiopathic headache syndrome. *Nat Med* 1999; 5: 836–838.
- [17] Keller SS, Wilke M, Wiesmann 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.
- [18] Price S, Paviour D, Scahill R, *et al.* Voxel-based morphometry detects patterns of atrophy that help differentiate progressive supranuclear palsy and Parkinson's disease. *NeuroImage* 2004; 23: 663–669.
- [19] White NB, Alkire MT, Haier RJ. A voxel-based morphometric study of non-demented adults with Down syndrome. *NeuroImage* 2003; 20: 393–403.
- [20] Gitelman DR, Ashburner J, Friston KJ, Tyler LK, Price CJ. Voxel-Based Morphometry of Herpes Simplex Encephalitis. *NeuroImage* 2001; 13: 623–631.
- [21] Thieben MJ, Duggins AJ, Good CD, *et al.* The distribution of structural neuropathology in pre-clinical Huntington's disease. *Brain* 2002; 125: 1815–28.
- [22] Karas GB, Scheltens P, Rombouts SARB, *et al.* Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *NeuroImage* 2004; 23: 708–716.
- [23] Gorno-Tempini ML, Dronkers NF, Rankin KP, *et al.* Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol* 2004; 55: 335–46.
- [24] Damasio AR, Van Hoesen GW. The limbic system and the localisation of herpes simplex encephalitis. *J Neurol Neurosurg Psychiatr* 1985; 48: 297–301.
- [25] Fox NC, Warrington EK, Rossor MN. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease. *Lancet* 1999; 353: 2125.
- [26] Thompson PM, MacDonald D, Mega MS, Holmes CJ, Evans AC, Toga AW. Detection and mapping of abnormal brain structure with a probabilistic atlas of cortical surfaces. *J Comput Assisted Tomogr* 1997; 21: 567–581.
- [27] Maguire EA, Gadian DG, Johnsrude IS, *et al.* Navigation-related structural change in the hippocampi of taxi drivers. *Proc Natl Acad Sci USA*. 2000; 97: 4398–4403.
- [28] Maguire EA, Spiers HJ, Good CD, Hartley T, Frackowiak RS, Burgess N. Navigation expertise and the human hippocampus: a structural brain imaging analysis. *Hippocampus* 2003; 13: 250–259.
- [29] Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. *J Neurosci* 2003; 23: 9240–9245.
- [30] Draganski B, Gaser C, Busch V, Schuierer G, Bogdahn U, May A. Neuroplasticity: changes in grey matter induced by training. *Nature* 2004; 427: 311–2.
- [31] Mechelli A, Crinion JT, Noppeney U, *et al.* Neurolinguistics: structural plasticity in the bilingual brain. *Nature* 2004; 431: 757.
- [32] Kempermann G, Gast D, Gage FH. Neuroplasticity in old age: sustained fivefold induction of hippocampal neurogenesis by long-term environmental enrichment. *Ann Neurol*. 2002; 52: 135–143.
- [33] Trachtenberg JT, Chen BE, Knott GW, *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* 2002; 420: 788–794.
- [34] Grutzendler J, Kasthuri N, Gan WB. Long-term dendritic spine stability in the adult cortex. *Nature* 2002; 420: 812–6.
- [35] Lee DW, Miyasato LE, Clayton NS. Neurobiological bases of spatial learning in the natural environment: neurogenesis and growth in the avian and mammalian hippocampus. *NeuroReport* 1998; 9: 15–27.
- [36] Smulders TV, Sasson AD, DeVoogd TJ. Seasonal variation in hippocampal volume in a food-storing bird, the black-capped chickadee. *J Neurobiol* 1995; 27: 15–25.
- [37] Bookstein FL. "Voxel-based morphometry" should not be used with imperfectly registered images. *NeuroImage* 2001; 14: 1454–1462.
- [38] Ashburner J, Friston KJ. Why Voxel-Based Morphometry Should Be Used. *NeuroImage* 2001; 14: 1238–1243.
- [39] Salmond CH, Ashburner J, Vargha-Khadem F, Connelly A, Gadian DG, Friston KJ. Distributional Assumptions in Voxel-Based Morphometry. *NeuroImage* 2002; 17: 1027–1030.
- [40] Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 2004; 23: 17–20.
- [41] Friston KJ, Ashburner J. Generative and recognition models for neuroanatomy. *NeuroImage* 2004; 23: 21–24.
- [42] Richardson MP, Friston KJ, Sisodiya SM, *et al.* Cortical grey matter and benzodiazepine receptors in malformations of cortical development. A voxel-based comparison of structural and functional imaging data. *Brain* 1997; 120: 1961–1973.