Imperial College London

Removing inter-subject technical variability in multi-site magnetic resonance imaging studies

Archie Atkinson, Nicholas Brown | Dr Anna Barnes (Supervisor) | Centre for Medical Imaging - UCL

1) Motivation

Magnetic resonance (MR) images suffer from several sources of technical variation- non-biologically relevant variation in the intensity and structure of tissues. This restricts the medical community's ability to detect imaging **biomarkers**- measures of a biological state which differentiates healthy development from disease (e.g. a cancerous lesion in a patient's lung).

Technical variability arises from several sources including:

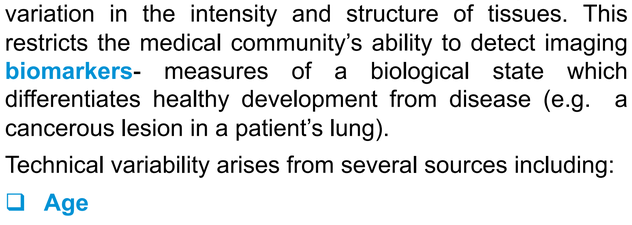
- ❑ **Age**
- ❑ **Gender**
- ❑ **Scanner manufacturer**

Both biomarkers and technical variations exist on the mm to cm scale meaning that the progression of biomarkers may be underestimated or missed entirely in **multi-site** (multiple imaging centres) and **inter-subject** (multiple patients) studies [1]. Additionally, as images from multiple centres often use varying imaging protocols and scanner types, technical variabilities are all the more likely to occur.

Brain images provide favourable conditions as they are relatively static and so are less affected by motion artifacts (see Figure 1). Moreover, inter-patient brain structure and shape is more uniform compared to other regions. Our project looks into the impact of this variability on **abdominal images**.

Our aim

To use **Voxel-Based-Morphometry (VBM)**, a statistical technique comparing individual voxels (3D pixels), to remove technical variability in **abdominal images**.



Patient 1 **Patient 2**

Figure 1. This Sagittal section shows obvious evidence of breathing motion in Patient 2 compared to Patient 1 occurring during the scan. In practice, patients are encouraged to try and take shallow breaths as this imaging sequence is 2.5 minutes long.

2) Methodology

To perform statistical tests, **images are required to be in a standard space**. Table 1 shows this process. The **General Linear Model (GLM)** quantifies the impact of technical variability in each voxel and from statistical tests comparing similar voxels across the dataset a **statistical parametric map (SPM)** is formed.

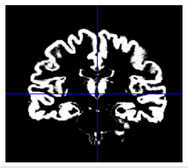
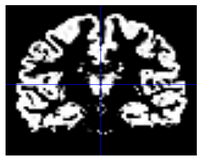
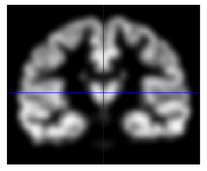
Segmentation	Normalisation	Smoothing
Purpose: Decompose a T_1 image into its **constituent tissue types**.	*Purpose:* Put segmented images into a **standard space** to enable comparative statistics.	*Purpose:* Suppress the noise of the images and to make the data more **normally distributed**.
Process: 1. Using tissue probability maps to inform the likelihood of a voxel belonging to a tissue type, images of each tissue can be made.	*Process:* 1. Segmented images are registered with a template image (atlas). 2. Non-linear (smooth) variations are removed from the images.	*Process:* 1. The image is convolved with a gaussian field, subjecting each voxel to the average of its neighbouring voxels.
Difficulties: • Large number of tissue types in the body. • There are currently no comprehensive tissue probability maps for the abdomen.	*Difficulties:* • The variation in fields of view (FOV) across a dataset may lead to an incorrect transformation to the standard space [1]. • No current atlases for the abdomen.	*Difficulties:* • Reduces the accuracy of technical variability localisation [1].
Steps taken to limit impact of error: • Determine tissue type through voxel intensity to find regions that correspond to the same tissue type. Cerebrospinal fluid (CSF) is used as a control for the intensity.	*Steps taken to limit impact of error:* • Align images and consider smaller regions to allow for correct transformation.	*Steps taken to limit impact of error:* • Increase the significance of statistical tests to filter for highly affected areas [1]. • Adjust the FWHM of Gaussians to be like expected regional differences.
		

Table 1: Voxel-Based Morphometry (VBM) is a voxel-by-voxel statistical analysis whereby a SPM is produced. To perform the analysis, smoothed images are required; the table above outlines the required processing pipeline, and the images are the progression of a grey matter image.

The General Linear Model (GLM):

The GLM is a linear multiple regression model that quantifies technical variability and is expressed by,

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{N1} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{NN} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

Y: Image Data
X: Explanatory variables- scanner type
β: Unknown parameters to be determined
ϵ: Residual error

Assumptions:

- **Y is univariate- each variable is independent.**
- **ϵ is normally distributed [2].**

The comparison of relevant values of β across the images allows for the statistical significance of desired explanatory variables to be measured and localised.

Statistical Parametric Map (SPM):

Figure 2 shows how statistical t-tests are performed over the same regions on a voxel-wise basis to determine regions of interest. These tests accept or reject our null hypothesis, H_0 , that no technical variability is related to scanner type. Age is a nuisance variable and so its effect refines the GLM but is not ultimately considered when quantifying the impact of scanner type. The level of significance for our tests is set to $p < 0.001$ ($T_{critical} = 3.53$) to localise regions (clusters) of greater significance [1].

***Figure 2.** Comparative two tailed t-test across the same region of smoothed images*

3) Preliminary Results

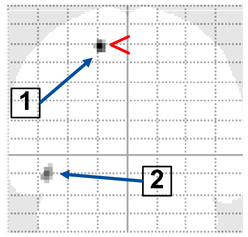


Figure 3: The SPM of brain data comparing age.

Considering figure 3, the maximally correlated voxel T-values uncorrected for multiple statistical tests are:

- ❑ Cluster 1: $T_{max} = 5.28$
- ❑ Cluster 2: $T_{max} = 4.46$,

Thus, **rejecting H_0 in these regions**. Ongoing analysis is being completed for abdominal images, as there are image normalisation complications due to the lack of abdominal atlases.

Findings in abdominal images

Images of the abdomen do not always cover the same "common areas", complicating the normalisation process as more complicated transformations are needed. Hence, incorrect normalisation has arisen due to factors such as:

- ❑ **Patient position**
- ❑ **Scanner set-up**
- ❑ **Physiological variability (for example, body fat)**

4) Implications

Multi-site studies have demonstrated that scanner variability is significant in the brain, however the complications in the normalisation of abdominal images in our project outlines the need for further research into the construction of abdominal atlases in order to quantity technical variability [3].

Probing the smaller "common areas" to assess the effect of scanner variability is still currently being performed. However, our recommendations for further research and experimental design include:

- ❑ **Introducing stringent patient alignment measures**
- ❑ **Developing atlases of the abdomen**

These recommendations could enhance the scientific community's ability to remove technical variability in order to more effectively study imaging biomarkers.

References

[1]: Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

[2]: Mechelli, A., Price, C., Friston, K., & Ashburner, J. (2005). Voxel-Based Morphometry of the Human Brain: Methods and Applications. *Current Medical Imaging Reviews*, *1*(2). https://doi.org/10.2174/1573405054038726

[3]: Stonnington C.M., *et al*.. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. *Neuroimage*. 39(3): pg.1180-5 (2018)