NeuroImage 84 (2014) 133–140

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Effects of study design in multi-scanner voxel-based morphometry studies

Hidemasa Takao [a,*], Naoto Hayashi [b], Kuni Ohtomo [a]

[a] Department of Radiology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
[b] Department of Computational Diagnostic Radiology and Preventive Medicine, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

ARTICLE INFO

Article history:
Accepted 19 August 2013
Available online 29 August 2013

Keywords:
Magnetic resonance imaging
Multi-center
Multi-site
Reliability
Reproducibility
Variability

ABSTRACT

Interest has recently grown in multi-center studies, which have more power than smaller studies in conducting sophisticated evaluations of basic neuroanatomy and neurodegenerative disorders. The large number of subjects that result from pooling multi-center datasets increases sensitivity, but also introduces a between-center variance component. Taking sex differences as an example, we examined the effects of different ratios of cases to controls (males to females) between scanners in multi-scanner morphometric studies, using voxel-based morphometry and data obtained on two scanners of the exact same model. Each subject was scanned twice with both scanners. Using the image obtained on either of the two scanners for each subject, voxel-based analyses were repeated with different ratios of males to females for each scanner. As the ratio of males to females became more imbalanced between the scanners, the differences between the two scanners more strongly affected the results of analyses of sex differences. When the ratio of males to females was balanced, the inclusion of scanner as a covariate in the statistical analysis had almost no influence on the results of analyses of sex differences. When the ratio of males to females was ill-balanced, the inclusion of scanner as a covariate suppressed scanner effects on the results, but made sex differences less likely to become significant. The present results suggest that as long as the ratio of cases to controls is well-balanced across different scanners, it is not always necessary to include scanner as a covariate in the statistical analysis, and that when the ratio of cases to controls is ill-balanced across scanners, the inclusion of scanner as a covariate in the statistical analysis can suppress scanner effects, but may make differences less likely to be detected.

Introduction

Multi-center studies have more power than smaller studies in conducting sophisticated evaluations of basic neuroanatomy and neurodegenerative disorders. There has been growing interest in multi-center studies such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Jack et al., 2008), which is a multi-center observational study of healthy elders, mild cognitive impairment, and Alzheimer's disease. Such studies provide researchers with larger datasets by pooling data from different sites and hence improving the statistical power. The large number of subjects that results from pooling multi-center datasets increases sensitivity, thus allowing detection of subtle effects, and offers increased reliability and confidence regarding effect size by averaging out unforeseen confounds. However, multi-center studies also introduce a between-center variance component. One important confound of combining images obtained from different scanners is the potential for scanner effects (e.g., scanner-dependent geometrical inaccuracies, image intensity variability) to introduce systematic error, thus complicating the interpretation of results.

Many previous studies have evaluated the effect of using different scanners on cross-sectional or longitudinal morphometric results (Briellmann et al., 2001; Dickerson et al., 2008; Ewers et al., 2006; Fennema-Notestine et al., 2007; Focke et al., 2011; Gunter et al., 2009; Han et al., 2006; Ho et al., 2010; Huppertz et al., 2010; Jovicich et al., 2009; Kempton et al., 2011; Kruggel et al., 2010; Moorhead et al., 2009; Pardoe et al., 2008; Pfefferbaum et al., 2012; Schnack et al., 2004; Stonnington et al., 2008; Suckling et al., 2010; Takao et al., 2011, 2013). Regarding volumetric measurement, there is generally greater inter-scanner than intra-scanner variability. Scanner effects are inevitable to a greater or lesser extent, and are impossible to eliminate completely.

In cross-sectional morphometric studies that consider data obtained on multiple scanners, the ratio of cases to controls is often different between scanners. Some studies analyze data without considering scanner effects, whereas others analyze data while including scanner as a covariate in the statistical analysis. Previous studies have combined and analyzed data obtained on multiple scanners and examined the validity of combining multi-scanner datasets (Fennema-Notestine et al., 2007;

* Corresponding author at: Department of Radiology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Fax: +81 3 5800 8935.
E-mail address: takaoh-tky@umin.ac.jp (H. Takao).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.08.046

Table 1
Ratios (A–I) of males to females for each scanner.

	Scanner 1 (n = 16)		Scanner 2 (n = 16)	
	Female	Male	Female	Male
A	0 (0%)	16 (100%)	16 (100%)	0 (0%)
B	2 (12.5%)	14 (87.5%)	14 (87.5%)	2 (12.5%)
C	4 (25%)	12 (75%)	12 (75%)	4 (25%)
D	6 (37.5%)	10 (62.5%)	10 (62.5%)	6 (37.5%)
E	8 (50%)	8 (50%)	8 (50%)	8 (50%)
F	10 (62.5%)	6 (37.5%)	6 (37.5%)	10 (62.5%)
G	12 (75%)	4 (25%)	4 (25%)	12 (75%)
H	14 (87.5%)	2 (12.5%)	2 (12.5%)	14 (87.5%)
I	16 (100%)	0 (0%)	0 (0%)	16 (100%)

Meda et al., 2008; Pardoe et al., 2008; Segall et al., 2009; Stonnington et al., 2008). Most of these studies demonstrated that scanner-related differences were much smaller than group differences or that consistent patterns of structural change were found across sites. They concluded that the results were not confounded by scanner differences and that it was possible to pool data obtained on multiple different scanners, with a caveat about the need to have balanced comparison groups for each scanner. To our knowledge, however, it has not been fully evaluated to what extent different ratios of cases to controls between scanners affect morphometric results, and whether the inclusion of scanner as a covariate in the statistical analysis can adequately eliminate scanner effects from morphometric results.

In the present study, taking sex differences as an example, we evaluated how different ratios of cases to controls (males to females) between scanners affected morphometric results, using voxel-based morphometry (VBM) and data obtained on two scanners of the exact same model. We also investigated whether the inclusion of scanner as a covariate in the statistical analysis can eliminate scanner effects from morphometric results when the ratio of cases to controls is ill-balanced across scanners.

Materials and methods

Subjects

A total of 32 normal subjects (16 females and 16 males, mean age = 58 ± 9 years [female: 58 ± 9 years, male: 58 ± 9 years], age range = 45–72 years [female: 45–72 years, male: 45–72 years]) were included in this study. None of the subjects had a history of neuropsychiatric disorder including serious head trauma, psychiatric disorder, or alcohol/substance abuse or dependence. The mean Mini-Mental State Examination (MMSE) score was 29.6 ± 0.7 (range = 27–30). A board-certified radiologist reviewed all scans (including T1-weighted and T2-weighted images) and found no gross abnormalities such as infarct, hemorrhage, or brain tumor in any of the subjects. Fazekas score (range, 0–3) was 0 (absence) or 1 (caps, pencil-thin lining and/or punctuate foci) (Fazekas et al., 1987). The Ethics Committee of the University of Tokyo Hospital approved the study. After a complete explanation of the study to each subject, written informed consent was obtained.

Acquisition of imaging data

MR data were obtained on two 3.0-T Signa HDx scanners (GE Medical Systems, Milwaukee, WI) with an 8-channel phased-array head coil. Both scanners were the exact same model. Each subject was scanned twice, at an interval of about 1 year (mean interval = 0.9 ± 0.1 years, range = 0.6–1.2 years). Of the 32 subjects, 16 (8 females and 8 males, mean age = 58 ± 8 years, age range = 45–72 years) were first scanned with Scanner 1 (Visit 1) and then with Scanner 2 (Visit 2), and the remaining 16 (8 females and 8 males, mean age = 59 ± 11 years, age range = 45–72 years) were first scanned with Scanner 2 (Visit 1) and then with Scanner 1 (Visit 2).

T1-weighted images were acquired using three-dimensional (3D) inversion recovery prepared fast spoiled gradient recalled acquisition in the steady state (IR-FSPGR) in 176 sagittal slices (repetition time = 5.3–5.4 ms; echo time = 1.7 ms; inversion time = 450 ms; flip angle = 15°; field of view = 250 mm; slice thickness = 1.0 mm with no gap; acquisition matrix = 256 × 256; number of excitations = 0.5; image matrix = 256 × 256). Parallel imaging (ASSET; Array Spatial Sensitivity Encoding Technique) was used with an acceleration factor of 2.0. The images were corrected for spatial distortion due to gradient non-linearity using 'grad_unwarp' (Jovicich et al., 2006; Takao et al., 2010a, 2010b) and for intensity non-uniformity using the nonparametric non-uniform intensity normalization algorithm N3 (Sled et al., 1998; Takao et al., 2010a, 2010b).

Image processing

Images were processed mainly using statistical parametric mapping (SPM) 8 software (http://www.fil.ion.ucl.ac.uk/spm) developed in the Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, running in MATLAB 7.13.0 (Mathworks, Sherborn, MA).

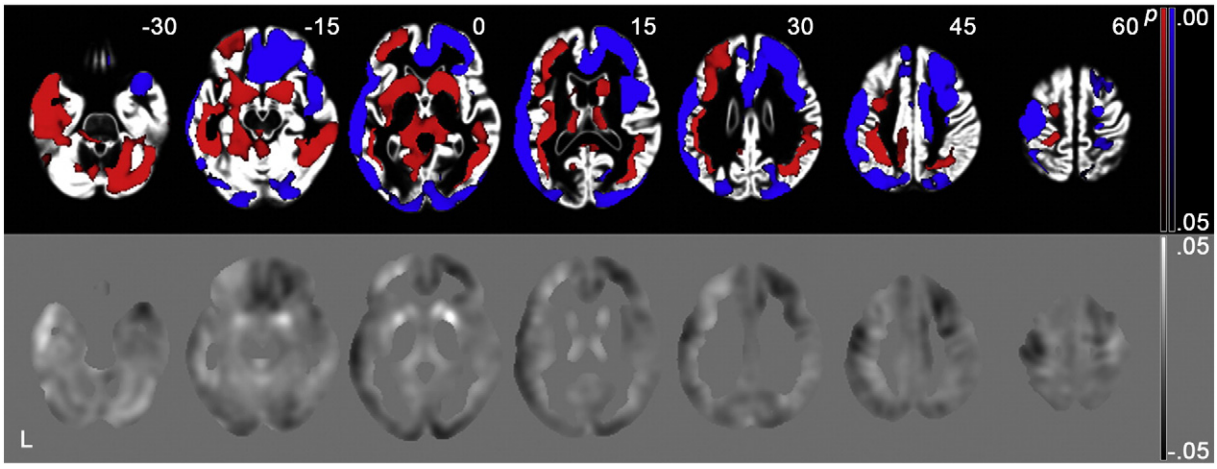

Fig. 1. (A) Voxel-based analysis of the differences in gray matter volume between Scanner 1 and Scanner 2. The color bars represent the *p* value at each voxel (red, Scanner 1 < Scanner 2; blue, Scanner 1 > Scanner 2), corrected for multiple comparisons using the family-wise error (FWE) rate (*t* test, Threshold-Free Cluster Enhancement [TFCE]). (B) Contrast images, which represent actual differences between the two scanners.