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).

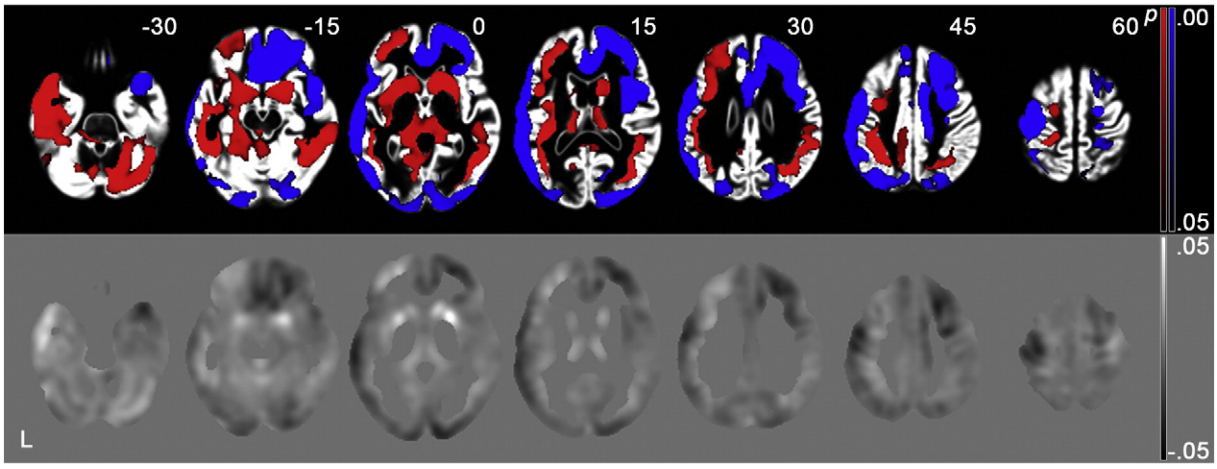

Fig. 1. (A) Voxel-based analysis of the differences in gray matter volume between Scanner 1 and Scanner 2. The color bars represent the *p* value at each voxel (red, Scanner 1 < Scanner 2; blue, Scanner 1 > Scanner 2), corrected for multiple comparisons using the family-wise error (FWE) rate (*t* test,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

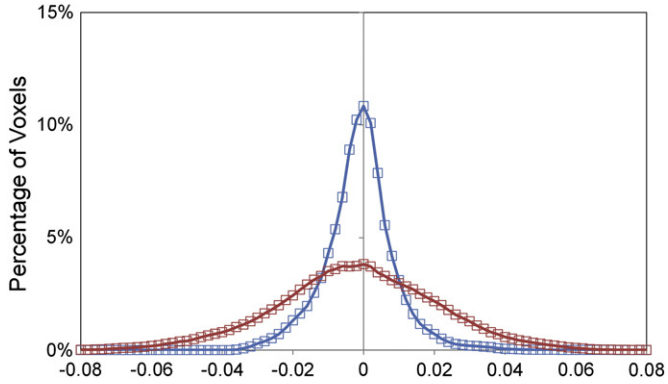

Fig. 2. Histograms of contrast images (blue, differences between Scanner 1 and Scanner 2 [see also Fig. 1B]; red, differences between males and females [see also Fig. 3B]).

The IR-FSPGR images were segmented into gray matter, white matter, and cerebrospinal fluid using an integrated generative model (unified segmentation) (Ashburner and Friston, 2005). The International Consortium for Brain Mapping (ICBM) gray matter, white matter, and cerebrospinal fluid templates were used as priors to segment the images. The Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) algorithm (Ashburner, 2007) was used to spatially normalize the segmented images. The normalized gray matter images were modulated to correct voxel signal intensity for volume displacement during normalization and reflect brain volume (Good et al., 2001b), smoothed using an 8 mm kernel, and normalized using whole gray matter volume.

Statistical analyses

Voxel-based analyses of the processed data were performed using permutation-based, voxel-wise non-parametric testing (Nichols and Holmes, 2002) (as implemented in the randomise tool, part of FSL [FMRIB Software Library 4.1, http://www.fmrib.ox.ac.uk/fsl]).

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 (cases to controls) for each scanner (ratios A–I) (Table 1). The number of subjects for each scanner was fixed at 16. There was a large number of combinations of subjects for each scanner, except for ratios A and I. Of the combinations, 10 were randomly selected for each ratio of males to females (B–H). The numbers of Visits 1 and 2 were fully balanced across scanners and sexes.

We identified areas of gray matter having significant sex differences using *t* tests with and without scanner as a covariate. Age was included as a covariate. The significance level for *t* tests was set at $p = 0.05$, corrected for multiple comparisons using the family-wise error (FWE) rate. We used the Threshold-Free Cluster Enhancement (TFCE) method (Smith and Nichols, 2009). Two *t* contrasts (female < male, female > male) were computed for *t* tests. The number of permutations was 5000. Only voxels with a volume greater than 0.1 (on all images) were included in analyses.

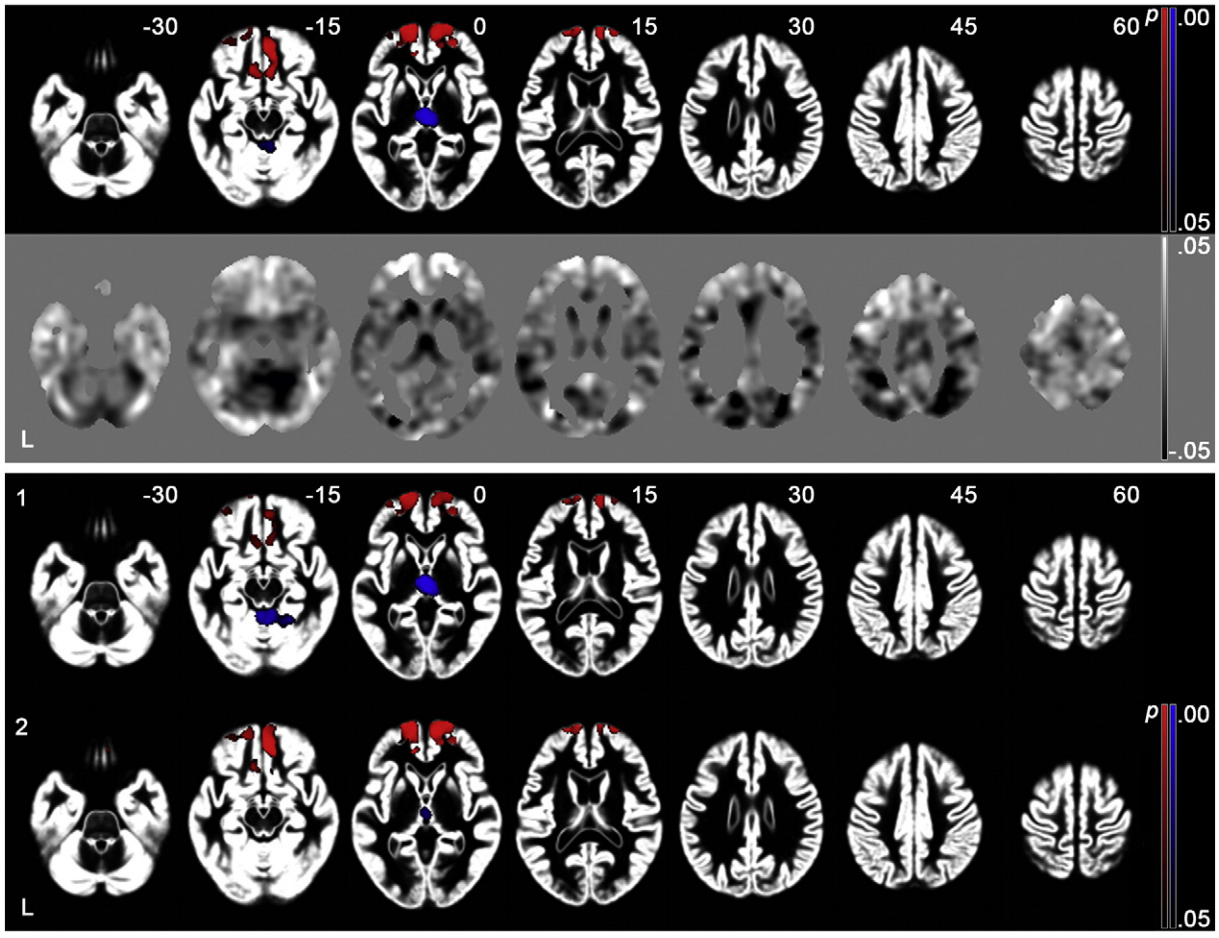

Fig. 3. (A) Voxel-based analysis of sex differences in gray matter volume using mean gray matter images of Scanner 1 and Scanner 2. The color bars represent the *p* value at each voxel (red, female < male; blue, female > male), corrected for multiple comparisons using the family-wise error (FWE) rate (*t* test, Threshold-Free Cluster Enhancement [TFCE]). (B) Contrast images, which represent differences between males and females with the effects of age removed. (C) Voxel-based analysis of sex differences in gray matter volume using images of Scanner 1. (D) Voxel-based analysis of sex differences in gray matter volume using images of Scanner 2.

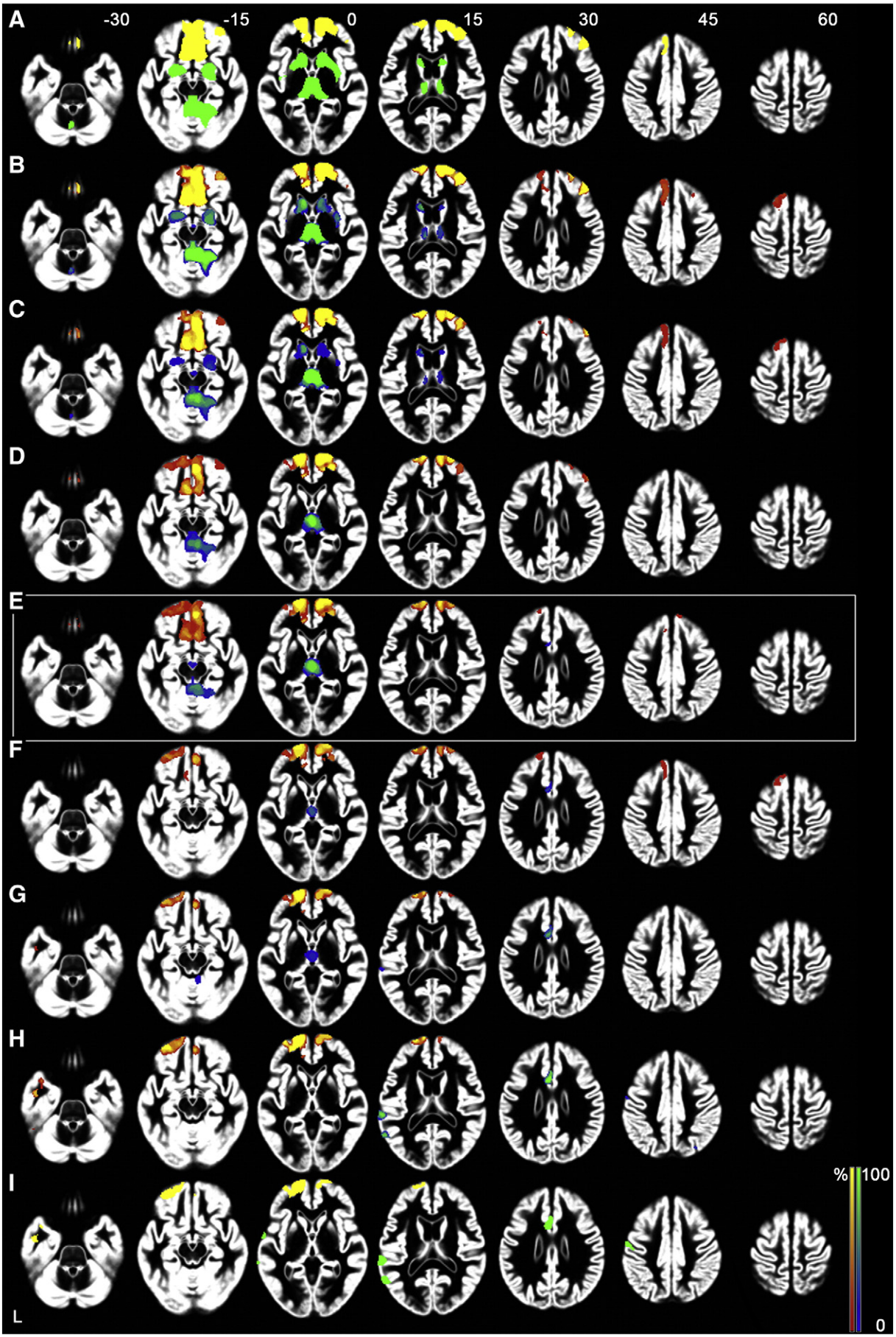
A
-30
-15
0
15
30
45
60
B
C
D
E
F
G
H
I
%
100
0
L

Table 2
Numbers of voxels for each percentage of analyses showing significant sex differences in voxel-based analyses repeated with different ratios (A–I) of males to females for each scanner, without scanner as a covariate.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
In regions with significant sex differences in voxel-based analysis of mean images											
A	1128	–	–	–	–	–	–	–	–	–	10,093
B	764	79	73	52	57	43	57	73	109	307	9607
C	772	50	52	68	100	105	134	275	521	816	8328
D	140	309	339	408	792	866	1246	1060	1022	1522	3517
E	2	135	580	1166	1389	1616	1168	1230	1268	1029	1638
F	3028	1251	993	878	1296	671	673	528	400	1347	156
G	4667	1379	685	509	419	351	462	586	1144	374	645
H	5509	445	592	504	276	380	336	325	725	266	1863
I	7680	–	–	–	–	–	–	–	–	–	3541
In regions without significant sex differences in voxel-based analysis of mean images											
A	387,618	–	–	–	–	–	–	–	–	–	28,488
B	385,940	4341	3358	2517	1696	1616	2802	2707	1734	2023	7372
C	393,374	8345	3233	2074	1671	1533	1598	1539	1388	964	387
D	405,390	5028	2536	1596	1166	262	104	24	0	0	0
E	406,615	4827	2582	1755	270	56	1	0	0	0	0
F	413,079	2503	370	149	5	0	0	0	0	0	0
G	413,415	1651	344	301	213	58	31	42	21	30	0
H	410,870	776	641	538	710	707	626	507	248	223	260
I	411,304	–	–	–	–	–	–	–	–	–	4802

Results

Differences between Scanner 1 and Scanner 2

Voxel-based analysis of subtraction images, produced by subtracting the gray matter images of Scanner 1 from those of Scanner 2 for each subject, revealed a number of regions with significant differences between the two scanners (Fig. 1A). Fig. 1B shows contrast images obtained from the above analysis, which represent actual differences between the two scanners. Fig. 2 shows a histogram of the contrast images (histogram bin width, 0.002; range, −0.1 to 0.1). Approximately 6% of voxels showed a difference of less than −0.02 (3.2%) or more than 0.02 (3.0%).

Differences between males and females using mean images of Scanner 1 and Scanner 2

Voxel-based analysis of mean images, produced by averaging the gray matter images of Scanner 1 and those of Scanner 2 for each subject, revealed several gray matter regions with significant differences between males and females (Fig. 3A). Fig. 3B shows contrast images obtained from the above analysis, which represent differences between males and females with the effects of age removed. Fig. 2 shows a histogram of the contrast images (histogram bin width, 0.002; range, −0.1 to 0.1).

Figs. 3C and D show the results of voxel-based analyses of sex differences in gray matter volume using images of Scanner 1 and Scanner 2, respectively; these results are almost the same as those using mean images of Scanner 1 and Scanner 2.

Effects of different ratios of males to females between Scanner 1 and Scanner 2 in terms of the Results of Sex Differences

Fig. 4 and Table 2 show the results of voxel-based analyses repeated with different ratios (A–I) of males to females for each scanner, without scanner as a covariate. The figure displays the percentage of analyses for each ratio of males to females that showed significant sex differences. The greater the imbalance in the ratio of males to females between the two scanners (E → D/F → C/G → B/H → A/I), the greater the differences between the two scanners affected the results of analyses of sex differences.

Fig. 5 and Table 3 show the results of voxel-based analyses repeated with different ratios (B–H) of males to females for each scanner, with scanner as a covariate. The figure displays the percentage of analyses for each ratio of males to females that showed significant sex differences. In the case of an even ratio of males to females between the two scanners (E), the results of analyses with scanner as a covariate (Fig. 5) were similar to those of analyses without scanner as a covariate (Fig. 4). In the case of imbalanced ratios of males to females between the two scanners, the inclusion of scanner as a covariate in statistical analyses suppressed the effects of scanner on the results of analyses of sex differences; however, the more imbalanced the ratio of males to females between the two scanners (E → D/F → C/G → B/H), the less likely that sex differences were significant.

Discussion

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 data obtained on two scanners of the exact same model. The greater the imbalance in the ratio of males to females between the two scanners, the more strongly the differences between the two scanners affected the results of analyses of sex differences.

The inclusion of scanner as a covariate in the statistical analysis barely influenced the results of analyses of sex differences in the case of an even ratio of males to females between the two scanners. In the case of imbalanced ratios of males to females, the inclusion of scanner as a covariate suppressed scanner effects, but sex differences were less likely to be significant. The present results suggest that as long as the ratio of cases to controls is well-balanced across 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 result in differences being less likely to be detected. This situation may arise because it becomes more difficult to differentiate between group effects and scanner effects when the ratio of cases to controls is ill-balanced across scanners. The

Fig. 4. Voxel-based analyses repeated with different ratios (A–I) of males to females for each scanner, without scanner as a covariate. The color bars represent the percentage of analyses showing significant sex differences (red-yellow, female < male; blue-green, female > male) at each voxel (*t* test, Threshold-Free Cluster Enhancement [TFCE], $p < 0.05$ corrected for multiple comparisons using the family-wise error [FWE] rate).

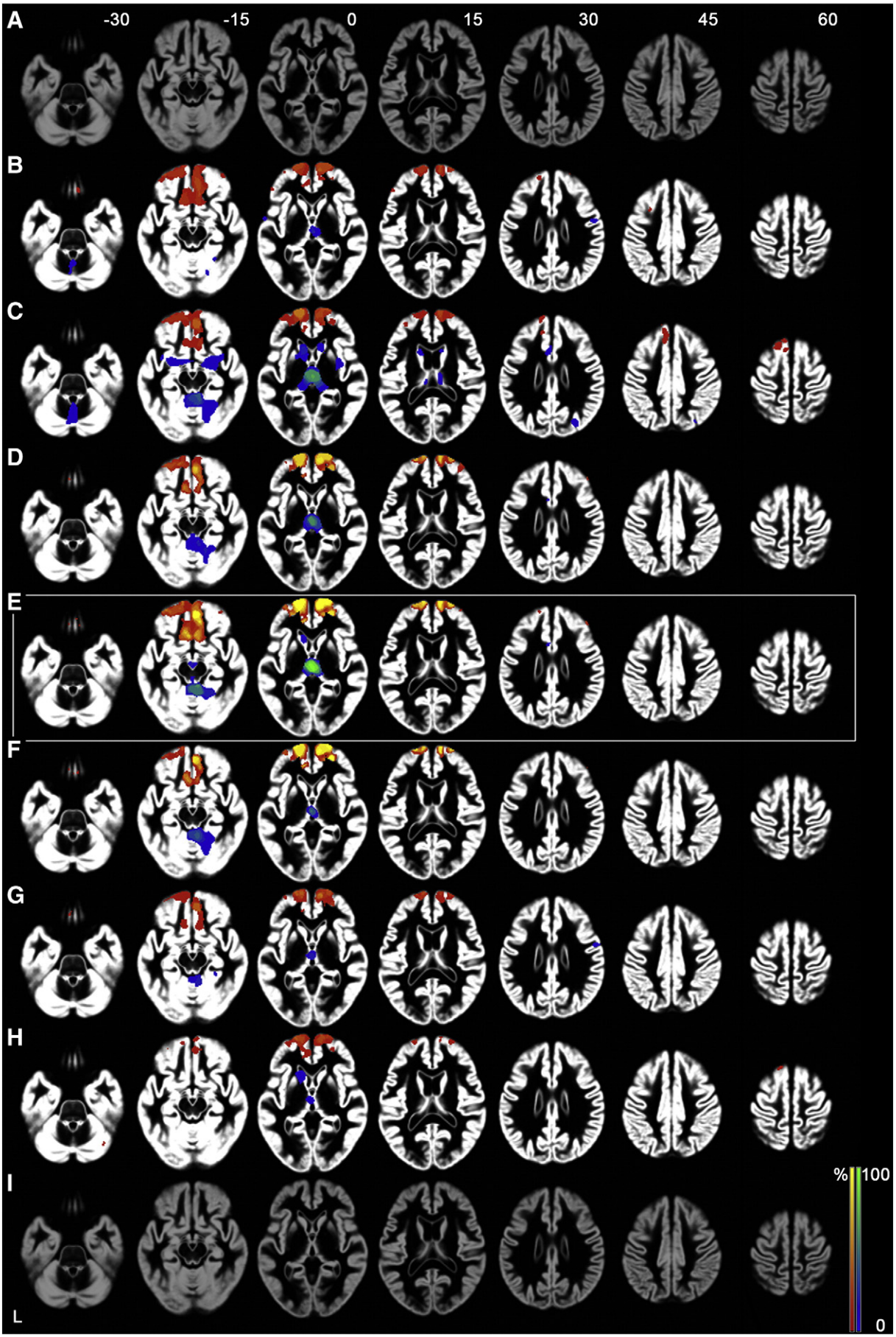
A
-30
-15
0
15
30
45
60
B
C
D
E
F
G
H
I
%
100
0
L

Table 3
Numbers of voxels for each percentage of analyses showing significant sex differences in voxel-based analyses repeated with different ratios (B–H) of males to females for each scanner, with scanner as a covariate.

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
In regions with significant sex differences in voxel-based analysis of mean images											
B	3156	2327	4013	1590	135	0	0	0	0	0	0
C	1410	3295	2684	1445	822	892	557	116	0	0	0
D	427	1713	1794	1393	1086	998	1208	951	1295	356	0
E	0	14	194	838	1460	1656	1178	1273	1110	1281	2217
F	1207	1448	1191	1360	862	767	751	734	576	1795	530
G	3263	4043	2180	1092	517	87	14	25	0	0	0
H	6051	3326	1340	487	17	0	0	0	0	0	0
In regions without significant sex differences in voxel-based analysis of mean images											
B	409,350	5893	825	38	0	0	0	0	0	0	0
C	395,076	18,337	2116	412	104	53	8	0	0	0	0
D	411,490	3919	660	37	0	0	0	0	0	0	0
E	405,750	4658	2890	2122	585	101	0	0	0	0	0
F	411,106	3839	746	344	71	0	0	0	0	0	0
G	412,766	3307	33	0	0	0	0	0	0	0	0
H	414,177	1903	26	0	0	0	0	0	0	0	0

extent of the influence may depend not only on the difference in the ratio of cases to controls between scanners, but also on the extent of differences between scanners, and on differences between cases and controls. In the present study, we took sex differences as an example; however, the extent of the influence may differ in other situations with more subtle differences between cases and controls, and is difficult to fully estimate from the present dataset. Of course, it would be ideal to balance the ratio of cases to controls across scanners as much as possible, to diminish scanner effects as much as possible, and to include an adequate number of subjects.

Previous studies have analyzed data obtained on multiple scanners and evaluated the validity of combining multi-scanner datasets (Fennema-Notestine et al., 2007; Meda et al., 2008; Pardoe et al., 2008; Segall et al., 2009; Stonnington et al., 2008). In the present study, we evaluated the effects of different ratios of cases to controls between scanners in terms of morphometric results, taking sex differences as an example. Sex differences themselves were, however, outside the scope of the present study because the sample size was relatively small, thus limiting the generalizability of the findings. It is well known that brain volume is smaller in women than in men, which is explained in part by the smaller body dimensions of women. Gray and white matter volumes also vary by sex, and women have smaller gray matter and white matter volumes than men (Luders and Toga, 2010). When tissue measurements were adjusted for brain size, some studies found a higher percentage of gray matter in women (Gur et al., 1999; Luders et al., 2002), while others failed to detect any sex differences (Nopoulos et al., 2000) or observed a higher percentage of gray matter in men (Good et al., 2001a). Several studies have examined regional sex differences in gray matter volume using voxel-based morphometry, with somewhat mixed results (Barnes et al., 2010; Chen et al., 2007; Cheng et al., 2009; Good et al., 2001a; Lentini et al., 2012; Lord et al., 2010; Pell et al., 2008; Takahashi et al., 2011; Van Laere and Dierckx, 2001; Witte et al., 2010). Among these studies, inconsistencies regarding regional sex differences in gray matter volume may arise due to heterogeneity of the populations and differences in methods of image analysis including global normalization (proportional scaling vs. residual method; total intracranial volume vs. whole brain volume vs. total gray matter volume).

In the present study, we used grad_unwarp and N3 to correct for geometric distortion due to gradient non-linearity and intensity non-uniformity, which is caused by factors such as inhomogeneous radiofrequency fields, inhomogeneous reception sensitivity, and electromagnetic interaction with the object being scanned (Vovk et al., 2007), respectively. Grad_unwarp corrects for geometric distortion caused by gradient non-linearity but not that caused by scanner-dependent geometrical inaccuracies. In the present study, scanner-dependent geometrical differences and residual intensity non-uniformity may be the causes of the observed differences between the two scanners.

Voxel-based morphometry is a mass univariate approach for comparing the volume of tissue among populations of subjects, free of rater bias, and provides objective information regarding brain structure (Ashburner and Friston, 2000). Previous studies that compared the results of voxel-based morphometry analyses with those from manual measurements found relatively good correspondence (Takao et al., 2010b). Voxel-based morphometry involves segmenting the images into gray matter, white matter, and cerebrospinal fluid; warping these tissue maps into standard space; and smoothing the resulting spatially normalized tissue maps. Voxel-by-voxel statistical analysis is performed and the statistical measures are corrected for multiple comparisons. Because voxel-based morphometry is one of the most widely used objective techniques for investigating brain morphometry, we used this method in the present study. The present results are, however, not limited to voxel-based morphometry: they are also applicable to other morphometric techniques including the region-of-interest method and cortical thickness analysis.

In conclusion, we examined the effects of different ratios of cases to controls between scanners in multi-scanner morphometric studies using data obtained on two scanners of the exact same model, taking sex differences as an example. The results indicate that as long as the ratio of cases to controls is well-balanced across 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 make differences less likely to be detected.

Acknowledgments

We thank Dr. Eriko Maeda and Dr. Takeharu Yoshikawa for their help in collecting data. This work was supported by a Grant-in-Aid for Young Scientists (A) 24689047 from the Japan Society for the Promotion of Science.

Fig. 5. Voxel-based analyses repeated with different ratios (B–H) of males to females for each scanner, with scanner as a covariate. The color bars represent the percentage of analyses showing significant sex differences (red-yellow, female < male; blue-green, female > male) at each voxel (*t* test, Threshold-Free Cluster Enhancement [TFCE], $p < 0.05$ corrected for multiple comparisons using the family-wise error [FWE] rate).

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. Neuroimage 11, 805–821.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 53, 1244–1255.

Briellmann, R.S., Syngeniotis, A., Jackson, G.D., 2001. Comparison of hippocampal volumetry at 1.5 tesla and at 3 tesla. Epilepsia 42, 1021–1024.

Chen, X., Sachdev, P.S., Wen, W., Anstey, K.J., 2007. Sex differences in regional gray matter in healthy individuals aged 44–48 years: a voxel-based morphometric study. Neuroimage 36, 691–699.

Cheng, Y., Chou, K.H., Decety, J., Chen, I.Y., Hung, D., Tzeng, O.J., Lin, C.P., 2009. Sex differences in the neuroanatomy of human mirror-neuron system: a voxel-based morphometric investigation. Neuroscience 158, 713–720.

Dickerson, B.C., Fenstermacher, E., Salat, D.H., Wolk, D.A., Maguire, R.P., Desikan, R., Pacheco, J., Quinn, B.T., Van der Kouwe, A., Greve, D.N., Blacker, D., Albert, M.S., Killiany, R.J., Fischl, B., 2008. Detection of cortical thickness correlates of cognitive performance: reliability across MRI scan sessions, scanners, and field strengths. Neuroimage 39, 10–18.

Ewers, M., Teipel, S.J., Dietrich, O., Schonberg, S.O., Jessen, F., Heun, R., Scheltens, P., van de Pol, L., Freymann, N.R., Moeller, H.J., Hampel, H., 2006. Multicenter assessment of reliability of cranial MRI. Neurobiol. Aging 27, 1051–1059.

Fazekas, F., Chawluk, J.B., Alavi, A., Hurtig, H.I., Zimmerman, R.A., 1987. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJ. Am. J. Roentgenol. 149, 351–356.

Fennema-Notestine, C., Gamst, A.C., Quinn, B.T., Pacheco, J., Jernigan, T.L., Thal, L., Buckner, R., Killiany, R., Blacker, D., Dale, A.M., Fischl, B., Dickerson, B., Gollub, R.L., 2007. Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. Neuroinformatics 5, 235–245.

Focke, N.K., Helms, G., Kaspar, S., Diederich, C., Toth, V., Dechent, P., Mohr, A., Paulus, W., 2011. Multi-site voxel-based morphometry—not quite there yet. Neuroimage 56, 1164–1170.

Good, C.D., Johnsrude, I., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001a. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. Neuroimage 14, 685–700.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001b. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 14, 21–36.

Gunter, J.L., Bernstein, M.A., Borowski, B.J., Ward, C.P., Britson, P.J., Felmlee, J.P., Schuff, N., Weiner, M., Jack, C.R., 2009. Measurement of MRI scanner performance with the ADNI phantom. Med. Phys. 36, 2193–2205.

Gur, R.C., Turetsky, B.I., Matsui, M., Yan, M., Bilker, W., Hughett, P., Gur, R.E., 1999. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. J. Neurosci. 19, 4065–4072.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 32, 180–194.

Ho, A.J., Hua, X., Lee, S., Leow, A.D., Yanovsky, I., Gutman, B., Dinov, I.D., Lepore, N., Stein, J.L., Toga, A.W., Jack Jr., C.R., Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2010. Comparing 3 T and 1.5 T MRI for tracking Alzheimer's disease progression with tensor-based morphometry. Hum. Brain Mapp. 31, 499–514.

Huppertz, H.J., Kroll-Seger, J., Kloppel, S., Ganz, R.E., Kassubek, J., 2010. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. Neuroimage 49, 2216–2224.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46, 177–192.

Kempton, M.J., Underwood, T.S., Brunton, S., Stylios, F., Schmechtig, A., Ettinger, U., Smith, M.S., Lovestone, S., Crum, W.R., Frangou, S., Williams, S.C., Simmons, A., 2011. A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: evaluation of a novel lateral ventricle segmentation method. Neuroimage 58, 1051–1059.

Kruggel, F., Turner, J., Muftuler, L.T., 2010. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49, 2123–2133.

Lentini, E., Kasahara, M., Arver, S., Savic, I., 2012. Sex differences in the human brain and the impact of sex chromosomes and sex hormones. Cereb. Cortex (in press).

Lord, C., Engert, V., Lupien, S.J., Pruessner, J.C., 2010. Effect of sex and estrogen therapy on the aging brain: a voxel-based morphometry study. Menopause 17, 846–851.

Luders, E., Toga, A.W., 2010. Sex differences in brain anatomy. Prog. Brain Res. 186, 3–12.

Luders, E., Steinmetz, H., Jancke, L., 2002. Brain size and grey matter volume in the healthy human brain. Neuroreport 13, 2371–2374.

Meda, S.A., Giuliani, N.R., Calhoun, V.D., Jagannathan, K., Schretlen, D.J., Pulver, A., Cascella, N., Keshavan, M., Kates, W., Buchanan, R., Sharma, T., Pearlson, G.D., 2008. A large scale (N = 400) investigation of gray matter differences in schizophrenia using optimized voxel-based morphometry. Schizophr. Res. 101, 95–105.

Moorhead, T.W., Gountouna, V.E., Job, D.E., McIntosh, A.M., Romaniuk, L., Lymer, G.K., Whalley, H.C., Waiter, G.D., Brennan, D., Ahearn, T.S., Cavanagh, J., Condon, B., Steele, J.D., Wardlaw, J.M., Lawrie, S.M., 2009. Prospective multi-centre voxel based morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. BMC Med. Imaging 9, 8.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Nopoulos, P., Flaum, M., O'Leary, D., Andreasen, N.C., 2000. Sexual dimorphism in the human brain: evaluation of tissue volume, tissue composition and surface anatomy using magnetic resonance imaging. Psychiatry Res. 98, 1–13.

Pardoe, H., Pell, G.S., Abbott, D.F., Berg, A.T., Jackson, G.D., 2008. Multi-site voxel-based morphometry: methods and a feasibility demonstration with childhood absence epilepsy. Neuroimage 42, 611–616.

Pell, G.S., Briellmann, R.S., Chan, C.H., Pardoe, H., Abbott, D.F., Jackson, G.D., 2008. Selection of the control group for VBM analysis: influence of covariates, matching and sample size. Neuroimage 41, 1324–1335.

Pfefferbaum, A., Rohlfing, T., Rosenbloom, M.J., Sullivan, E.V., 2012. Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths. Neuroimage 60, 940–951.

Schnack, H.G., van Haren, N.E., Hulshoff Pol, H.E., Picchioni, M., Weisbrod, M., Sauer, H., Cannon, T., Huttunen, M., Murray, R., Kahn, R.S., 2004. Reliability of brain volumes from multicenter MRI acquisition: a calibration study. Hum. Brain Mapp. 22, 312–320.

Segall, J.M., Turner, J.A., van Erp, T.G., White, T., Bockholt, H.J., Gollub, R.L., Ho, B.C., Magnotta, V., Jung, R.E., McCarley, R.W., Schulz, S.C., Lauriello, J., Clark, V.P., Voyvodic, J.T., Diaz, M.T., Calhoun, V.D., 2009. Voxel-based morphometric multisite collaborative study on schizophrenia. Schizophr. Bull. 35, 82–95.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98.

Stonnington, C.M., Tan, G., Kloppel, S., Chu, C., Draganski, B., Jack Jr., C.R., Chen, K., Ashburner, J., Frackowiak, R.S., 2008. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. Neuroimage 39, 1180–1185.

Suckling, J., Barnes, A., Job, D., Brenan, D., Lymer, K., Dazzan, P., Marques, T.R., MacKay, C., McKie, S., Williams, S.R., Williams, S.C., Lawrie, S., Deakin, B., 2010. Power calculations for multicenter imaging studies controlled by the false discovery rate. Hum. Brain Mapp. 31, 1183–1195.

Takahashi, R., Ishii, K., Kakigi, T., Yokoyama, K., 2011. Gender and age differences in normal adult human brain: voxel-based morphometric study. Hum. Brain Mapp. 32, 1050–1058.

Takao, H., Abe, O., Hayashi, N., Kabasawa, H., Ohtomo, K., 2010a. Effects of gradient nonlinearity correction and intensity non-uniformity correction in longitudinal studies using structural image evaluation using normalization of atrophy (SIENA). J. Magn. Reson. Imaging 32, 489–492.

Takao, H., Abe, O., Ohtomo, K., 2010b. Computational analysis of cerebral cortex. Neuroradiology 52, 691–698.

Takao, H., Hayashi, N., Ohtomo, K., 2011. Effect of scanner in longitudinal studies of brain volume changes. J. Magn. Reson. Imaging 34, 438–444.

Takao, H., Hayashi, N., Ohtomo, K., 2013. Effects of the use of multiple scanners and of scanner upgrade in longitudinal voxel-based morphometry studies. J. Magn. Reson. Imaging. http://dx.doi.org/10.1002/jmri.24038.

Van Laere, K.J., Dierckx, R.A., 2001. Brain perfusion SPECT: age- and sex-related effects correlated with voxel-based morphometric findings in healthy adults. Radiology 221, 810–817.

Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans. Med. Imaging 26, 405–421.

Witte, A.V., Savli, M., Holik, A., Kasper, S., Lanzenberger, R., 2010. Regional sex differences in grey matter volume are associated with sex hormones in the young adult human brain. Neuroimage 49, 1205–1212.