Title

Pooling morphometric estimates: a statistical equivalence approach

Authors

Heath R. Pardoe[1], Gary R. Cutter[2], Rachel Alter[1], Rebecca Kucharsky Hiess[1], Mira Semmelroch[3], Donna Parker[3], Shawna Farquharson[3], Graeme D. Jackson[3], and Ruben Kuzniecky[1]

[1]Comprehensive Epilepsy Center, Department of Neurology, New York University School of Medicine, New York, USA
[2]School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA
[3]The Florey Institute of Neuroscience and Mental Health, Melbourne, Australia

Corresponding Author

Heath R. Pardoe
Comprehensive Epilepsy Center, NYU School of Medicine
223 East 34th St,
New York, NY, USA, 10016
Phone: +1-646-754-5320
Email: heath.pardoe@nyumc.org

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.808v1 | CC-BY 4.0 Open Access | rec: 28 Jan 2015, publ: 28 Jan 2015

Abstract

Changes in hardware or image processing settings are a common issue for large multi-center studies. In order to pool MRI data acquired under these changed conditions, it is necessary to demonstrate that the changes do not affect MRI-based measurements. In these circumstances classical inference testing is inappropriate because it is designed to detect differences, not prove similarity. We used a method known as statistical equivalence testing to address this limitation.

Equivalence testing was carried out on three datasets: (i) cortical thickness and automated hippocampal volume estimates obtained from 16 healthy individuals imaged different multi-channel head coils; (ii) manual hippocampal volumetry obtained using two readers; and (iii) corpus callosum area estimates obtained using an automated method with manual cleanup carried out by two readers. Equivalence testing was carried out using the "two one-sided tests" approach.

Cortical thickness values were found to be equivalent over 78% of the cortex when different head coils were used ($p = 0.024$). Automated hippocampal volume estimates obtained using the same two coils were statistically equivalent ($p = 4.28 \times 10^{-15}$). Manual hippocampal volume estimates obtained using two readers were not statistically equivalent ($p = 0.97$). The use of different readers to carry out limited correction of automated corpus callosum segmentations yielded equivalent area estimates (1.28×10^{-14}).

We have presented a statistical method for determining if morphometric measures obtained under variable conditions can be pooled. The equivalence testing technique is applicable for analyses in which experimental conditions vary over the course of the study.

Keywords

MRI, statistics, morphometry, volumetrics

1. Introduction

Neuroanatomical changes in disease and normal development are often assessed by measuring morphometric properties of brain structures from an MRI scan. These measurements may be automated or require manual input from a human reader. Common issues with MRI scanning, particularly in large multicenter studies, are changes in the collection or processing conditions over the duration of the study. These changes may include the use of multiple scanners, scanner hardware or software upgrades, or the use of different human readers for manual morphometric measurements such as hippocampal volumes. A number of studies have investigated whether these variable experimental conditions introduce differences, systematic or otherwise, in quantitative measurements [1-6]. The introduction of systematic differences is undesirable because it may reduce the ability to detect differences between groups, or increase the probability of making a false positive or false negative finding.

In this study we apply a statistical method known as equivalence testing to MRI datasets that were acquired or analyzed under changed experimental conditions that may be commonly encountered in neuroimaging studies. Equivalence testing is an inference-based method for determining if measures obtained under variable conditions can be considered ‘equivalent’. The equivalence testing approach addresses a common misinterpretation of classical inference testing that a $p > 0.05$ provides statistical support for the absence of an effect. This interpretation is incorrect. With equivalence testing, the null hypothesis is formulated as there being differences between the groups being compared, and evidence must be used to disprove this hypothesis. If there is enough evidence against the null hypothesis of a difference, we can reasonably conclude that the two groups are equivalent, and we can carry out prospective analyses with a high degree of confidence that the comparisons using pooled data are legitimate. Equivalence testing approaches are becoming increasingly used for a number of biomedical applications, a number of which are described in Walker and Nowacki [7].

The first dataset consists of healthy individuals that were scanned in one session using an identical T1-weighted whole brain MPRAGE acquisition with a 20 channel and 32 channel receiver coil. We then used statistical equivalence testing to assess if automated measures of cortical thickness and hippocampal volume estimates from the two coils can be considered equivalent. If morphometric measures estimated from MRI scans obtained using the two coils are shown to be equivalent, future studies can combine MRI scans from each coil with reasonable confidence that no systematic bias has been introduced.

The second dataset addresses inter-rater variability in manual hippocampal volumetry. Differences between readers are typically assessed using descriptive statistics such as percentage volume difference or intra-class correlation, or spatial overlap measures such as the Dice coefficient or Jaccard index. Intuitively people understand that a lower volume difference or higher overlap is better, however these methods are not inference based and so there is no accepted standard for these measures. Statistical equivalence testing is an inference based method and so applies a standard that is accepted by the scientific community. In the case of equivalence testing, this standard is a false positive rate p less than 0.05, where a false positive finding implies that samples are equivalent when in fact they are not.

We will carry out an equivalence analysis on a dataset of manual hippocampal volumes measured using two different readers. It is well known that different readers often obtain hippocampal volume estimates that are systematically different; therefore we may reasonably expect these manual estimates to fail our test for equivalence for this type of difference between readers. Excess variability by one or the other reader can also lead to failure to show equivalence. Finally we will investigate an automated method for estimating corpus callosum area [8] that occasionally requires manual correction for small segmentation errors. Previous experience with the software indicates these errors occur in a small number of cases for images obtained from some MRI scanners.

The following specific hypotheses were tested in this study:

1. Automated vertex-wise cortical thickness measurements and hippocampal volumes measured using MRI data acquired with a 20-channel and 32-channel head coils are statistically equivalent.
2. Manual hippocampal volumes of healthy controls, segmented using two different readers, are statistically equivalent.

3. Corpus callosum area measured using a semi-automated method by two different readers, are statistically equivalent.

Code for carrying out the equivalence analyses presented in this paper is provided at https://sites.google.com/site/hpardoe/equivalence.

2. Methods

A common method for statistical equivalence testing is the two one-sided tests (TOST, described in [9]). The method requires an a-priori definition of an acceptable equivalence margin. The choice of the equivalence margin should be made relative to the expected effect size for the problem at hand, and should be a small fraction of this effect size. While few formal methods exist for determining the equivalence margin, often amounts that are some proportion of a minimally clinically significant difference or so called clinically important difference, such as ¼ or ⅓ of this difference may be used or a similar amount of meaningful biological change as it would be measured over some time interval. For the analyses presented in this study we used an equivalence margin of 5% of the average value for each morphometric measure. More formally, the equivalence margin θ, which is in the same units as the measure of interest, or defined as a suitable proportion of the mean values of the measure of interest, defines the limits of acceptability for defining 'equivalence'.

Hypotheses to be tested using the TOST approach may be stated as follows. These statements follow those described in [9], modified for the morphometric properties investigated in our study: M_1 represents the morphometric parameter of interest (eg. Cortical thickness, hippocampal volume) measured under the first experimental conditions; μ_{m1} is the mean value across subjects; M_2 represents the same morphometric parameter measured under the second experimental conditions; and μ_{m2} is the corresponding mean value.

$$H_0: \mu_{m1} - \mu_{m2} \leq -\theta \text{ or } \mu_{m1} - \mu_{m2} \geq \theta$$

$$H_1: -\theta < \mu_{m1} - \mu_{m2} < \theta$$

The null hypothesis H_0 states that the mean values of the morphometric estimate of interest obtained under different conditions are outside our predefined equivalence margins and therefore are not equivalent. The alternative hypothesis H_1 states that the mean measures obtained under different conditions lie within a margin, such that for practical purposes are equivalent. If we carry out an equivalence test and reject the null hypothesis, which means that we reject both that $\mu_{m1} - \mu_{m2} \leq -\theta$ and that $\mu_{m1} - \mu_{m2} \geq \theta$. We test each of these two null hypotheses at the 5% level and must reject both, thus, we obtain an experimentwise p-value less than each one sided test's alpha (typically equal to 0.05). If both tests are rejected, we conclude that measures obtained under two conditions are equivalent and can be pooled with confidence that we have not introduced systematic differences that have a meaningful impact on our ability to identify our effect of interest.

In practice, the two one-sided tests consists of two sequential one-sided tests, with the two null hypotheses being (i) the difference in means is greater than θ and (ii) the difference in means is less than –θ.

$$\text{Test 1. } H_0: \mu_{m1} - \mu_{m2} \geq \theta \qquad H_1: \mu_{m1} - \mu_{m2} < \theta$$

$$\text{Test 2. } H_0: \mu_{m1} - \mu_{m2} \leq -\theta \qquad H_1: \mu_{m1} - \mu_{m2} > -\theta$$

The null hypothesis is only rejected if both null hypotheses are rejected for the separate tests. In our study, following the implementation in [10], the reported p-value for the two one-sided test procedure is the maximum p for the individual tests. As noted in [9], the two one-sided test procedure is the same as determining if the (1 - 2α) confidence interval is within the ±θ equivalence margin.

Graphically this can be seen that we are defining 2 overlapping intervals, such that the overlap defines the equivalence region, but we have shown than it is unlikely that the difference is greater than θ and that the difference is < - θ (Figure 1). If the difference is shown to fall within the equivalence margin, we interpret the differences as not being clinically meaningful.

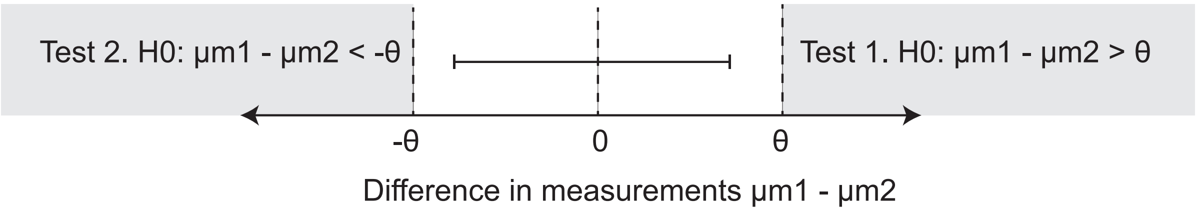
Test 2. H0: µm1 - µm2 < -θ
Test 1. H0: µm1 - µm2 > θ
-θ
0
θ
Difference in measurements µm1 - µm2

Figure 1. Equivalence testing depends on the prior definition of an equivalence margin [-θ, θ]. If

the null hypotheses for Test 1 and Test 2 are rejected, we can conclude that measurements obtained under variable experimental conditions are equivalent. Applying these two one-sided tests are equivalent to determining if the (1 - 2α) confidence interval, shown as the capped line, is within the equivalence margin.

2.1 MRI acquisition and image processing

2.1.1 Cortical thickness estimates from MRI data acquired using different coils.

The first dataset consists of 16 healthy controls (8 female, age 32.4 ± 6.1 years) who were scanned twice on the same MRI scanner in the same imaging session, using (a) a 20-channel receive head coil and (b) a 32-channel receive head coil. Images were acquired on a 3T Siemens Skyra MRI scanner, using a T1-weighted whole-brain 3D MPRAGE acquisition, sagittal slice prescription, 0.9 mm isotropic voxel size, TR = 1900 ms, TE = 2.49 ms, TI = 900 ms, FA = 9°. Pixel bandwidth was 180 Hz/Px for the 20 channel coil and 230 Hz/Px for the 32 channel coil.

MRI scans were processed using Freesurfer version 5.1.0. Vertex-wise cortical thickness and hippocampal volume estimates were derived using the longitudinal processing stream [11]. In brief, this consists of running the standard cross-sectional image processing pipeline, followed by the creation of an unbiased subject specific template from both scans for each subject, and subsequent longitudinal processing of the initial cross-sectional surfaces (more information provided at http://freesurfer.net/fswiki/LongitudinalProcessing at time of publication). Cortical maps were coregistered to the common space “fsaverage” template to allow comparison across subjects. 10 mm full-width half-maximum (FWHM) smoothing was applied to the cortical thickness maps. Thickness values were then read into R using the package “cortex” [12].

Paired TOST inference tests of cortical thickness data were carried out vertex-wise with alpha = 0.05, equivalence margin of 5% of vertex-wise mean cortical thickness. P-values for each vertex were mapped and thresholded at $p < 0.05$ to allow visualization of regions, which may reasonably be concluded to be equivalent. False discovery rate thresholding was used to correct for multiple comparisons. In addition to the TOST testing procedure, vertex-wise paired t-tests were carried out using the t.test function provided as part of the R stats package [13]. P-values were recorded and mapped in a similar manner to the TOST procedure.

The results of the cortical thickness analyses were summarized by recording the number of suprathreshold vertices as a percentage of the total number of vertices for both the equivalence testing procedure and vertex-wise paired t-tests. Non-cortex vertices along the medial hemispheric surfaces were excluded using the fsaverage “lh.cortex.label” and “rh.cortex.label” files provided with the Freesurfer distribution. Uncorrected ($p < 0.05$) and FDR corrected ($q < 0.05$) measures of coverage were measured.

2.1.2 Automated hippocampal volumes estimated from MRI data acquired using different coils.

A similar analysis was carried out with hippocampal volume estimates from the automated subcortical segmentations provided with Freesurfer. The longitudinal processing stream was used to estimate hippocampal volumes for images acquired using each coil, and paired TOST tests were carried out with alpha = 0.05, and an equivalence margin = 5% of the average hippocampal volume. Left and right hippocampal volumes were concatenated. Paired T-tests were performed with alpha = 0.05. Average percentage difference in hippocampal volumes was measured, which was calculated as the mean of $100*\text{abs}(HV_{coil01} – HV_{coil02})/\text{mean}(HV_{coil01},HV_{coil02})$.

2.1.3 Manual hippocampal segmentations measured using different readers.

The second dataset consisted of hippocampal volumes measured from manual segmentations carried out by two readers. The dataset consists of 40 healthy controls (20 female, age 30.5 ± 8.8 years). Images were acquired on a 3T Siemens TIM Trio MRI scanner, using a T1-weighted whole-brain 3D MPRAGE acquisition, 0.9 mm isotropic voxel size, TR = 1900 ms, TE = 2.6 ms, TI = 900 ms, flip angle = 9°.

Paired TOST analyses were carried out with alpha = 0.05 and an equivalence margin of 5% of the overall average hippocampal volume (across left and right hippocampi and both readers). Left and right hippocampal volumes were concatenated. Paired T-tests were performed with alpha = 0.05. Average percentage difference in hippocampal volumes was measured, which was calculated as the mean of $100*\text{abs}(HV_{reader01} – HV_{reader02})/\text{mean}(HV_{reader01},HV_{reader02})$.

2.1.4 Semi-automated corpus callosum segmentations measured using different readers.

The third dataset consists of corpus callosum area measurements measured using an automated software package “yuki” developed by Ardekani et al [8]. There are occasionally minor segmentation errors that require manual editing to obtain accurate estimates of corpus callosum area. In this study we compare corpus callosum area measurements assessed using two readers for a single site from the Autism Brain Imaging Data Exchange (ABIDE) study (http://fcon_1000.projects.nitrc.org/indi/abide/, site University of Michigan Sample 1 [14, 15]), in order to determine if we could pool corpus callosum area measurements obtained using different readers to assess the entire ABIDE dataset (consisting of 1000+ MRI scans). Whole brain T1-weighted SPGR MRI was obtained on a 3 T GE Signa scanner, 1.2 mm slice thickness, 1 mm^2 in-plane resolution, TE = 1.8 ms, Prep Time = 500 ms, flip angle = 15 degrees.

Paired TOST analyses were carried out with alpha = 0.05 and an equivalence margin of 5% of the average corpus callosum area. Paired T-tests were performed with alpha = 0.05. Average percentage difference in corpus callosum area was measured, which was calculated as the mean of $100*abs(CC.area_{reader01} – CC.area_{reader02})/mean(CC.area_{reader01},CC.area_{reader02})$.

3. Results

Cortical thickness values measured using two different coils were determined to be equivalent over 73.4% of the left and 83.4% of the right hemisphere cortical surfaces (Table 1, $q < 0.05$ FDR correction for multiple comparisons, and Figure 2). Few vertices were determined to have significant differences in cortical thickness due to the two coils, with 7.4% left hemisphere vertices and 0.3% right hemisphere vertices having suprathreshold p-values.

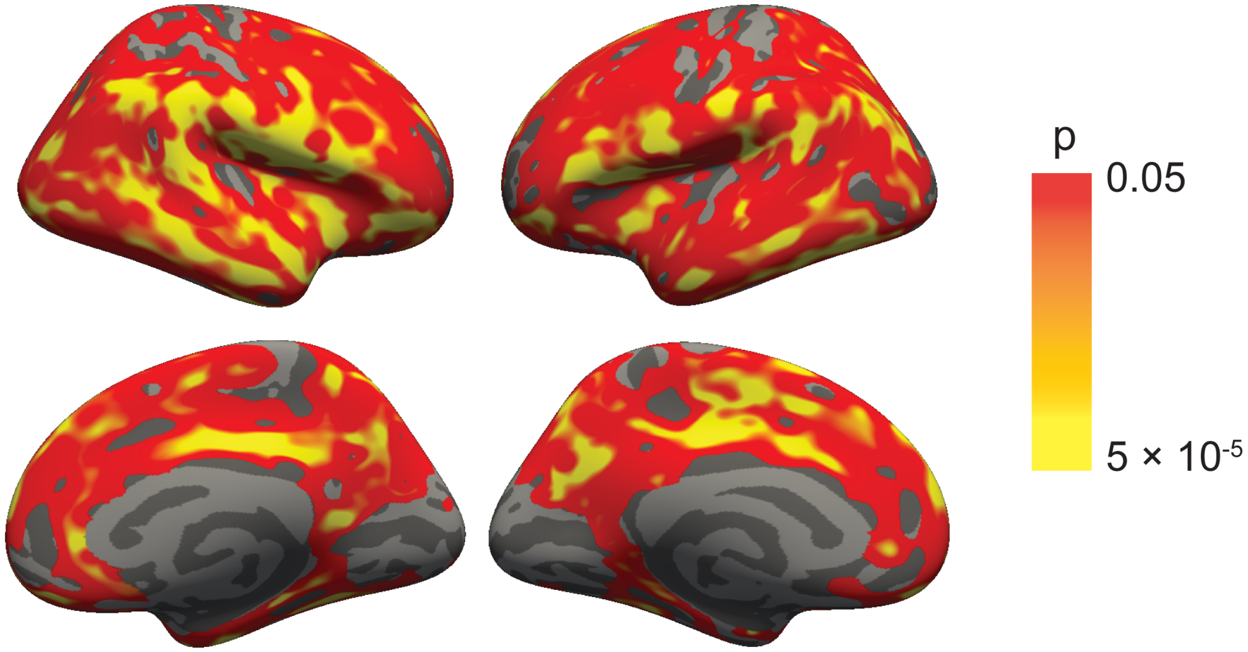

Figure 2. Cortical thickness measurements obtained using MRI scans from a 20 channel and 32 channel coils are equivalent over 78% of the cortex (73.4% left hemisphere, 83.4% right hemisphere, TOST $p < 0.05$, equivalence margin 5%).

Automated hippocampal volume estimates derived from MRI scans acquired on different coils were determined to be equivalent using the TOST approach ($p = 4.28 \times 10^{-15}$). Paired T-tests did not show significant differences ($p = 0.61$) , and the difference in hippocampal volume estimates derived from the two coils was small (0.38%). In contrast to this finding, hippocampal volumes measured using manual segmentation with two different readers did not pass the equivalency test (TOST $p = 0.97$), and systematic differences between these two estimates were identified (7.49% volume difference, T-test $p = 3.42 \times 10^{-8}$).

	Left Hemisphere	Right Hemisphere
TOST null rejected (%)	73.4 (76.8)	83.4 (81.7)
T-test null rejected (%)	7.3 (26.5)	0.8 (16.5)

Table 1. Equivalence analysis applied to cortical thickness estimation using two different receive coils indicates that most of the cortex can be considered equivalent (FDR correction for multiple comparisons, $q < 0.05$). Uncorrected values are provided in brackets ($p < 0.05$).

The ABIDE study University of Michigan sample 1 consisted of 110 structural MRI scans (55 autism cases and 55 controls). Eleven corpus callosum segmentations required manual edits. Paired TOST analysis of the entire dataset indicated that corpus callosum areas obtained using both readers were equivalent ($p = 1.28 \times 10^{-14}$). No significant differences were observed between estimates from both readers (area difference = 2.1%, T-test p = 0.71).

Method	Variable Condition	TOST p-value	T-test p-value	Difference in morphometric estimates (%)
1 a. Cortical thickness	Coils	0.024*	0.33*	2.84*
1 b. Hipp. volume (automated)	Coils	4.28×10^{-15}	0.61	0.38
2. Hippocampal volume (manual)	Readers	0.97	3.42×10^{-8}	11.21
3. Corpus callosum area	Readers	1.28×10^{-14}	0.71	2.1

Table 2. Equivalence testing p-values and comparative paired T-test values for morphometric parameters derived under variable experimental conditions. In all analyses the equivalence margin was set to 5% of mean estimate of interest. *Median value of p-values/thickness difference measured over the cortical surface. See Figure 1 for images showing distribution of TOST p values over the cortical surface.

4. Discussion

In this study we have applied a statistical inference method to morphometric estimates obtained under variable conditions to determine if these measures are equivalent. If measures are demonstrated to be equivalent within a predetermined equivalence margin, measurements obtained under these variable conditions may be pooled. The specific outcomes of our experiments indicated that cortical thickness measurements obtained from MRI scans acquired using different coils are statistically equivalent to within 5% of mean values over most of the cortex. Similarly hippocampal volumes obtained using the Freesurfer automated subcortical segmentation algorithms are equivalent to within 5% of mean values when obtained from a 20 channel and 32 channel head coil. Based on these analyses, we would infer that studies may reasonably pool MRI data acquired from these two coils for morphometric analyses of cortical thickness or hippocampal volume, as long as researchers are not aiming to detect an effect size close to 5%. If researchers are hoping to detect effects of the order of 5%, scans should be limited to a single coil or approximately equal numbers of controls and subjects of interest are obtained on both coils and the coil is modeled as a potential confounding factor in statistical analyses.

With regard to vertex- or voxel-wise measures, our results indicate that measures may be equivalent only across regions of the cortex; in the case of the cortical thickness estimates obtained from different head coils, approximately 73 – 83% of vertices were equivalent (Table 1). We therefore recommend a two-stage testing process for these kinds of analyses. The first step would involve carrying out equivalence tests of the whole brain average measure (eg. cortical thickness). If the p-value for this test is greater than 0.05, it indicates that most vertices are not equivalent and so the researcher can assume non-equivalence and proceed accordingly. If the whole brain average value passes this initial test ($p < 0.05$), we then recommend carrying out a vertex- or voxel-wise equivalence test in order to determine the regional variability of equivalence over the brain; this will allow the researcher to be aware of any potential areas that are not equivalent in subsequent analyses. In the case of our cortical thickness example, these 'danger regions' are primarily located in the insula and pre-central sulcus regions (Figure 1).

An important consideration for these analyses is whether the analysis has included enough participants for the study to be adequately powered. As with classical hypothesis testing, the power of a study is improved by the inclusion of more participants. If a small number of participants are included in an equivalence analysis, there is a greater danger of making a false negative finding, ie. failing to demonstrate equivalence when the measures are equivalent. In the case of our vertex-wise comparison of cortical thickness estimated using two coils, 16 subjects were imaged. One may find that more of the cortex can be considered equivalent if more participants were included in the study. This fact demonstrates an important distinction with the erroneous interpretation of a classical hypothesis test of difference (eg. Student's t-test) with a $p > 0.05$ as implying equivalence; using this incorrect interpretation, a $p > 0.05$ is more likely if the study is underpowered, which is a clear contradiction.

Applying equivalence analysis to manual hippocampal volume estimates indicates that measures obtained using two different readers are not equivalent. Therefore these measurements should not be pooled without including readers as a factor, as well as requiring each reader to segment a balanced number of cases and controls. The presence of systematic reader-specific differences in manual hippocampal volumes is well known. However, it is important to note that equivalence analyses do not investigate the sensitivity of a particular method to detecting effects of interest. For example, although hippocampal volume measurements obtained manually are not equivalent between readers, they may still be a preferable approach compared with automated measurements because they are more sensitive to detecting disease related hippocampal volume changes. We recently demonstrated the improved sensitivity and specificity of manual hippocampal volumetry over automated methods in individuals with temporal lobe epilepsy [16].

Finally we demonstrated that corpus callosum area estimates, obtained with occasional manual input to correct minor segmentation errors, may be considered statistically equivalent with a 5% equivalence margin. This allows us to be confident that the use of different readers for the error correction process does not introduce systematic differences in the measured corpus callosum area.

The size of imaging studies is increasing, as is the use of multi-site designs. Typically differences between sites are measured empirically [2-4]. The method presented in this paper is a useful inference-based technique for determining if subtle changes in experimental conditions in a study can influence quantitative measurements derived from MRI, and will allow for improved design of large neuroimaging studies. The statistical equivalence testing technique will likely be useful in other MRI modalities such as diffusion imaging and for metrics derived from functional MRI analyses. There may be additional useful advantages for the approach described in this paper; for example these methods might be adapted to form more homogeneous groupings of individuals based on having “equivalent” brain parameters.

Acknowledgements

Supported by Amazon Web Services Education in Research grants.

References

1. Dewey, J., et al., *Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study*. Neuroimage, 2010. **51**(4): p. 1334-44.
2. Han, X., et al., *Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer*. Neuroimage, 2006. **32**(1): p. 180-94.
3. Jovicich, J., et al., *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, 2009. **46**(1): p. 177-92.
4. Jovicich, J., et al., *Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations*. Neuroimage, 2013. **83**: p. 472-84.
5. Nugent, A.C., et al., *Automated subcortical segmentation using FIRST: test-retest reliability, interscanner reliability, and comparison to manual segmentation*. Hum Brain Mapp, 2013. **34**(9): p. 2313-29.
6. Schuff, N., et al., *MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers*. Brain, 2009. **132**(Pt 4): p. 1067-77.
7. Walker, E. and A.S. Nowacki, *Understanding equivalence and noninferiority testing*. J Gen Intern Med, 2011. **26**(2): p. 192-6.
8. Ardekani, B.A., *yuki module of the Automatic Registration Toolbox (ART) for corpus callosum segmentation*. 2013. http://www.nitrc.org/projects/art
9. Schuirmann, D.J., *A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability*. J Pharmacokinet Biopharm, 1987. **15**(6): p. 657-80.
10. Robinson, A., *equivalence: provides tests and graphics for assessing tests of equivalence*. 2010. p. R package version 0.5.6.
11. Reuter, M., et al., *Within-subject template estimation for unbiased longitudinal image analysis*. Neuroimage, 2012. **61**(4): p. 1402-18.
12. Pardoe, H., *cortex: Sample size estimates for well-powered cortical thickness studies*. 2012. http://www.nitrc.org/projects/cortex/
13. Team, R.C., *R: A language and environment for statistical computing*. 2013, R Foundation for Statistical Computing: Vienna, Austria.
14. Monk, C.S., et al., *Abnormalities of intrinsic functional connectivity in autism spectrum disorders*. Neuroimage, 2009. **47**(2): p. 764-72.
15. Di Martino, A., et al., *The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism*. Mol Psychiatry, 2013.
16. Pardoe, H.R. and G.D. Jackson, *Manual hippocampal volumetry is a better detector of hippocampal sclerosis than current automated hippocampal volumetric methods*. AJNR Am J Neuroradiol, 2013. **34**(10): p. E114-5.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.808v1 | CC-BY 4.0 Open Access | rec: 28 Jan 2015, publ: 28 Jan 2015