Contents lists available at ScienceDirect

Psychiatry Research: Neuroimaging

Statistical adjustments for brain size in volumetric neuroimaging studies: Some practical implications in methods

Liam M. O'Brien [a,*], David A. Ziegler [b], Curtis K. Deutsch [c,d,e], Jean A. Frazier [f], Martha R. Herbert [g,h], Joseph J. Locascio [b,h]

[a] Department of Mathematics and Statistics, Colby College, Waterville, ME, USA
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[c] Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[d] Eunice Kennedy Shriver Center, Waltham, MA, USA
[e] McLean Hospital, Belmont, MA, USA
[f] Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, USA
[g] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[h] Department of Neurology, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

ABSTRACT

Volumetric magnetic resonance imaging (MRI) brain data provide a valuable tool for detecting structural differences associated with various neurological and psychiatric disorders. Analysis of such data, however, is not always straightforward, and complications can arise when trying to determine which brain structures are "smaller" or "larger" in light of the high degree of individual variability across the population. Several statistical methods for adjusting for individual differences in overall cranial or brain size have been used in the literature, but critical differences exist between them. Using agreement among those methods as an indication of stronger support of a hypothesis is dangerous given that each requires a different set of assumptions be met. Here we examine the theoretical underpinnings of three of these adjustment methods (proportion, residual, and analysis of covariance) and apply them to a volumetric MRI data set. These three methods used for adjusting for brain size are specific cases of a generalized approach which we propose as a recommended modeling strategy. We assess the level of agreement among methods and provide graphical tools to assist researchers in determining how they differ in the types of relationships they can unmask, and provide a useful method by which researchers may tease out important relationships in volumetric MRI data. We conclude with the recommended procedure involving the use of graphical analyses to help uncover potential relationships the ROI volumes may have with head size and give a generalized modeling strategy by which researchers can make such adjustments that include as special cases the three commonly employed methods mentioned above.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Volumetric magnetic resonance imaging (MRI) studies have been key in identifying structural brain changes associated with many neurological and psychiatric disorders. Such structural changes can manifest in a variety of ways. The challenge of detecting and interpreting these changes has fallen upon neuroanatomists, clinical researchers, and statisticians. The inherently quantitative nature of morphometric MRI data mandates the use of statistical techniques to assess volumetric relationships, often with the goal of detecting subtle differences in regional volumes between or among diagnostic groups.

The questions of primary interest to clinical researchers, however, are not typically as straightforward as simply determining whether a particular brain region is larger or smaller in one group relative to another. In one example from a recent study of microcephaly, despite a decrease in whole-brain volumes only nuclear gray matter was found to differ significantly from controls (Cheong et al., 2008). In studies of autism, macrocephaly is commonly reported in the literature (Lainhart et al., 1997; Fombonne et al., 1999; Fidler et al., 2000; Bolton et al., 2001; McCaffery and Deutsch, 2005; Rice et al., 2005). Most studies in the field have demonstrated that autistic children tend to have larger heads, and MRI studies have found larger brains among children with autism compared with controls. This finding raises an important question: Are brain volumes increased globally in autism (are all structures proportionally bigger?) or locally (is brain overgrowth in autism driven by regionally specific expansion of some brain structures but not others?).

* Corresponding author at: Department of Mathematics and Statistics, Colby College, 5838 Mayflower Hill, Waterville, ME 04901, USA. Tel.: +1 207 859 5838(voice); fax: +1 207 859 5846.
E-mail address: lobrien@colby.edu (L.M. O'Brien).

There is not a single, straightforward approach to addressing these questions. For example, one could make a statement about the overall average white matter volume in autistic children relative to controls. Nevertheless, someone with a larger brain is likely to exhibit increased gray and white matter volumes, although not necessarily according to the same proportions (Zhang and Sejnowski, 2000; Changizi, 2001; Bush and Allman, 2003). We could ask whether the amount of white matter is larger in autistic children after adjusting for total brain volume (TBV) or some other measure of head size (Herbert et al., 2003). The answers to these questions could be quite different depending on the methodology employed.

Considerations are further complicated when one asks what is meant by the phrase "adjusting for" in the previous paragraph. Those familiar with statistical literature are accustomed to seeing this phrase and generally have a preconceived notion of what it means. In the volumetric brain imaging literature, however, there are several ways in which one can assess the relative sizes of volumes of particular regions of interest (ROIs) after "adjusting for" differences in overall head size. It is important to note that although we use the term "head size" for the adjustment factor, different metrics can be used for head size. Total brain volume and intracranial volume (ICV) are two commonly measures, but their correlation generally decreases with increased age (Bartholomeusz et al., 2002). Using TBV may be more appropriate when interest is in how an ROI changes with respect to the brain as a whole. However, if interest is in how ROI volume changes with respect to maximal adult brain size, using ICV may be more appropriate. Other body parameters may also be used as adjustment factors (cf. Peters et al., 1998). It is important to note, however, that the issues discussed and the modeling methods recommended in this paper do not change with the choice of the adjustment variable.

The goal of the present article is to bring to light the origins of the three common adjustment methods, the statistical assumptions that underlie them, and to give examples of common pitfalls that researchers must be wary of when analyzing volumetric MRI brain data. Further, we assess the degree to which prevailing methods are concordant in an example data set, and the degree to which anthropometric dependent measures are interchangeable. We conclude with a generalized strategy which researchers may use when modeling volumetric MRI data.

2. Common methods for adjusting for head size

Here we discuss three common methods for adjusting for head size. While we use the general term "head size," we note that it can be used to refer to various body size measurements — including total brain volume. A more thorough discussion of head/body size parameters and their use in making statistical adjustments is reviewed in O'Brien et al. (2006).

We can generically refer to the three common methods used to adjust for head size as the 1) *proportion*, 2) *analysis of covariance (ANCOVA)*, and 3) *residual* approaches. All three appear in the literature (Goldstein et al., 1999; Sullivan et al., 2001), with the proportion and ANCOVA approaches being the most common. These two methods have been specifically discussed by Seidman et al. (1999), van Petten (2004), Greenberg et al. (2008), Vidal et al. (2008), Chen et al. (2010) and many others. However, the residual approach has also been used in a variety of studies since its introduction by Mathalon et al. (1993). These include applications in schizophrenia (Takayanagi et al., 2010), memory loss in elderly subjects (Mormino et al., 2009), and cognitive and motor decline in alcoholism (Sullivan, 2003). While we give brief descriptions of these methods, along with a general algorithm for the analysis of volumetric MRI data that may be useful in many situations, we caution that there can be no blanket procedure by which all analyses should be prescribed. Our recommendation is to employ a general modeling strategy, of which the

three common methods are special cases. The specific model employed is informed by preliminary graphical methods and a step-down modeling approach to check for possible complex relationships in the data. Note that an illustration of this strategy is given in the Appendix using a real data set.

2.1. The proportion approach

The proportion approach was initially discussed by Arndt et al. (1991) and Mathalon et al. (1993) and has been used extensively in the volumetric brain literature (cf. Seidman et al., 1999; Goldstein et al., 1999). For this approach, a volume of an ROI is taken and divided by a volumetric measure of total brain or intracranial volume (Goldstein et al., 1999; Buckner et al., 2004). This proportionalized outcome is often then regressed on any covariates of interest. It is important to note, however, that this method was originally designed to only test for group differences in proportionalized volumes (Arndt et al., 1991). That is, once the ROI volumes are divided by the measure of head size (such as TBV), a *t*-test is performed to test for equality of group means (or an ANOVA in the case of more than two groups). For simplicity we discuss this method in this light although this assumption is easily relaxed and the statements made here easily generalized to allow for additional covariates.

The proportion method can be viewed as a group-based regression method in which intercepts for the respective within-group regression lines are both assumed to be equal to zero. The difference in the linear relationship between the numerator (ROI volume) and denominator (volumetric head size measure) is then tested for significance (see Supplementary Fig. 1). This is shown algebraically below where *ROI* is the volume of the brain region of interest, and *TBV* is the total brain volume of the subject. Note that other normalizing volumes such as intracranial volume (ICV) can be used as well. Disregarding error, and using *i* to indicate the subject, we have for group 1,

$$E[ROI_{i1}] = \beta_1(TBV_{i1}) \Rightarrow E\left[\frac{ROI_{i1}}{TBV_i}\right] = \beta_1;$$

whereas for group 2 we have,

$$E[ROI_{i2}] = \beta_2(TBV_{i2}) \Rightarrow E\left[\frac{ROI_{i2}}{TBV_{i2}}\right] = \beta_2.$$

We can test the null hypothesis that $\beta_1 = \beta_2$ to formally detect any group differences in proportionalized ROI volumes.

2.2. The analysis of covariance (ANCOVA) approach

The ANCOVA approach can also be viewed as a regression-based method. This procedure falls into the broader category of the generalized linear model framework. These models can handle many different types of outcomes (e.g., binary, count) although in this paper we focus on continuous, normally distributed data in which case the model can be termed the "general linear model" (GLM). Standard parametric significance tests assume residuals from ANCOVA/GLM models are normally distributed and if they are not, transformations such as the log or square root (for positive skewed distributions) or power transformations (for negative skewed) can often produce approximately normal residuals. (Nonparametric tests allow for non-normality but tend to be less versatile and powerful than parametric models.) The outcome is the raw ROI volume, and the predictors are typically a diagnostic group indicator(s) and the head size (e.g. TBV) measure. Other predictors of interest that may be important predictors or confounders can also easily be included in the model. We use the term ANCOVA because one will generally have both a categorical (group indicator) and continuous (total brain

volume) predictor, although in a strict sense it could be argued that we are performing a multiple regression analysis.

The simple ANCOVA method assumes the groups have the same slope for their respective regression lines, and the analysis tests if the intercepts for their regression lines differ. See Supplementary Fig. 2 for an illustration. The ANCOVA model can be written using standard regression notation and disregarding random error as,

$$E(ROI_i) = \beta_0 + \beta_1 I_{i,Group=2} + \beta_2(TBV_i),$$

where ROI_i is the volume of the ROI of interest for subject i, $I_{i, Group=2}$ is an indicator function that equals 1 if subject i is a member of "group 2" and 0 otherwise, and TBV_i is the total brain volume of subject i. To better illustrate this we consider the equation above for each group separately. For group 1 we have,

$$E(ROI_i) = \beta_0 + \beta_2(TBV_i),$$

And for group 2 we have,

$$E(ROI_i) = (\beta_0 + \beta_1) + \beta_2(TBV_i).$$

We can test for a difference between the groups by testing the null hypothesis $H_0 : \beta_1 = 0$. That is, we are testing whether the groups differ by formally testing a difference in intercepts. Thus the ANCOVA model assumes that the same linear relation of ROI volumes to TBV holds in each group, except that the ROI volumes in one group are allowed to be augmented by a constant amount, β_1, relative to those in the other group.

It is possible for unadjusted group means to differ, but for the ANCOVA adjusted means not to be significantly different (as shown in Supplementary Fig. 5a). The converse situation can also occur in which there is no difference in unadjusted means but there is a difference in ANCOVA-adjusted means. The ordinal relation of the groups can even reverse after ANCOVA adjustment (i.e., one group is larger than the other in terms of unadjusted means but smaller than the other with respect to the ANCOVA-adjusted means; Supplementary Fig. 5b).

There are two situations in which the ANCOVA method does *not* provide an adjustment to the group means. If the groups do not differ in terms of their mean TBV there is no adjustment as one may expect. This situation is illustrated graphically in Supplementary Fig. 6a. There is also no adjustment made if there is no within-group correlation between the dependent variable (ROI volume) and the covariate (TBV) as illustrated in Supplementary Fig. 6b. In the first case, ANCOVA may still be advantageous if precision is increased due to the covariate removing enough error variance to offset the loss of a degree(s) of freedom for the test of group effect(s). An example of the first case might be adjusting for head circumference between older- and middle-aged groups of subjects (head circumference does not decline with age as brain volume does). An example of the second situation might occur when adjusting groups based on height or weight. Since these measures often do not correlate highly with ROI volumes, unadjusted results would be similar to adjusted results (see O'Brien et al., 2006, for a demonstration of this point).

If the two groups differ in the association between ROI volume and head size, a regression analysis with an interaction term of group status with head size will elucidate the effect. Results are more complicated then, and may indicate that one group's ROI volumes are larger than the other's within a particular range of head size, but lower or not different in some other range of head size. An example of this was noted by Ueda et al. (2010), whereby an interaction between group status and ICV was found for some gray matter regions in schizophrenia patients. Further, quadratic terms for head size can determine whether single-bend curves describe the relationship between ROI volumes and head size well and whether accelerating or decelerating relationships between them are present in the data.

An interaction of group status with a quadratic head size term indicates that the curvature differs between the groups, or that curvature is present in one group but not the other. Thus, the ANCOVA approach accommodates a wide variety of possible relationships among ROI volume, head size, and group status.

2.3. The residual approach

The residual approach was discussed in detail by Mathalon et al. (1993) and is a more difficult technique to implement than the proportion or GLM/ANCOVA methods. To perform this method, one must take data *from the control group only* and run a regression model that regresses the ROI volume on the total brain volume as well as any other covariates of interest. This is essentially the ANCOVA model described above using only the data from the control group.

For simplicity, we will first restrict our discussion of this method to the situation where there are no predictors other than diagnostic group status and a measure of total head/brain size. Residuals are generated for all the subjects using the estimates obtained from the model for the control group only. Thus, each residual represents the deviation of each subject's observed ROI volume from what would be expected of a control subject with the same total brain volume.

Obtain estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$ using the controls only from a model similar to the one used in the ANCOVA method,

$$E(ROI_i) = \beta_{0(controls)} + \beta_{1(controls)} TBV_i.$$

Residuals are generated using these estimates for all subjects,

$$residual_i = ROI_i - E(ROI_i),$$

where ROI_i is the observed ROI volume for subject i, and $E(ROI_i)$ is the expected ROI volume for a control subject with total brain volume equal to the total brain volume of subject i. The residuals can then be tested between or among groups.

3. Comparing methods

3.1. Comparison of ANCOVA and proportion methods

The proportion and ANCOVA methods detect different types of group effects in the data. Which method one should use depends on which of the underlying models discussed above is the true one (see Supplementary Fig. 3). A good first step to determine which of these scenarios is likely to be tenable is to plot the data using a different plotting symbol for each group. If the groups are generally parallel with a vertical offset (i.e., when the ANCOVA method would be appropriate), use of the proportion method of analysis may result in no group difference even if the vertical offset between the groups is statistically significant. It may also show a significant group difference when there actually is none. Conversely, if the proportion model is correct (i.e., the regression lines for both for the groups have a zero intercept) and the ANCOVA is run, it will correctly indicate no group difference when there is none. If there is a true group effect under the proportion model, ANCOVA may or may not detect it. Even if it does, however, its nature may be misunderstood. It is important to also note that the assumptions of the ANCOVA method are violated since the within-group regression lines are not parallel — and it is this inequality of slopes that the proportion method is detecting when finding a group effect.

Although the proportion method may be more substantively and theoretically meaningful than the ANCOVA method in certain situations, the ANCOVA method has a number of statistical advantages. First, in ANCOVA the variable being adjusted for does not have to be in the same units as the variable being adjusted. Second, it is possible to adjust for more than one variable simultaneously when

using the ANCOVA method. For example, one could easily adjust for height and weight in addition to TBV, although in such situations one should be aware of the issue of multi-collinearity (Kutner et al., 2005). Multiple predictors can increase predictive accuracy, especially if nonvolumetric measures are the only variables available for adjustment (e.g., head circumference, height, or weight). Further, the partialed relations of these predictors individually with the ROI volume, the shapes of various relationships between the ROI volumes and the total brain volumes, and the best predicting linear combination of covariates may be interesting by-products of the ANCOVA. With the proportion method, if one wishes to adjust for a variable it must be related to the proportionalized outcomes using a regression model (e.g., GLM). In that case, one cannot say whether any significant relationships detected are due to a relationship with the ROI volume, the TBV, or a combination of both. A third reason ANCOVA may be advantageous is that curvilinear relations can be adjusted for using quadratic or higher-order polynomial terms. Fourth, the proportion method is much more restrictive in its zero-intercept assumption than ANCOVA. Fifth, possible interactive effects resulting from a different relationship between the ROI and head size depending on diagnostic group can be handled by the ANCOVA method through the inclusion of an interaction term in the model.

Supplementary Fig. 3 illustrates examples of when the ANCOVA method and the proportion method give different results. Which method is correct depends on the research question and assumptions the data satisfy. For example, in Supplementary Fig. 3a, ROI volumes may be growing at a proportionally slower rate than the TBV, resulting in a smaller proportion of ROI volume to TBV as TBV increases. The ANCOVA method is correctly modeling this phenomenon and adjusting for it appropriately, whereas the proportion method is not. The proportion method is thus not sensitive to the fact that the group with smaller ROI volumes has a smaller mean ROI volume than its relationship with TBV would otherwise predict it to have. Presumably something, perhaps pathological, about that group is causing it to have the lower than expected ROI volume (See Locascio and Cordray, 1983 for a related problem of contradictory results from an ANCOVA and "Gain Score Analysis" of the same data — often referred to as "Lord's Paradox"). Supplementary Fig. 3b illustrates the opposite situation in which the proportion method detects a group difference, but ANCOVA does not.

3.2. Comparison of the ANCOVA and residual methods

Both the ANCOVA and residual methods are based on regression analyses. The ANCOVA method obtains estimates of the model parameters using all available data. The groups are thus compared utilizing information from all subjects. The residual method obtains estimates of the model parameters based on information from the control group only. The residuals computed using these estimates are done for the entire sample (cases and controls alike). These residuals represent the difference between the observed ROI volume and the predicted ROI volume for a *control* subject with the same predictor pattern.

The residuals should be compared between or among groups. Note that the mean of the residuals for the controls will always be 0 (because the model was generated using these data), and one is essentially testing whether the mean of the comparison group's residuals is different from 0. An important point not made by Mathalon et al. (1993) concerning the residual method is that there should be a reduction in the error degrees of freedom of the *t*-test (or *F*-test) when comparing the residuals between (or among) groups. This reduction should be equal to the number of covariates included in the regression model. In ANCOVA, the analogous adjustment to the degrees of freedom is included as an integral part of the algorithm and is the default in standard statistical software. In both ANCOVA and the residual methods, residuals from a nonlinear curve can be analyzed

relating the ROI volume and TBV. Supplementary Fig. 4 compares the ANCOVA and residual models in terms of schematic graphs corresponding to each.

If the assumptions underlying the ANCOVA model are tenable, then the ANCOVA method would generally be superior to the residual method in that the head size adjustment is made using all the data rather than just the data from controls. One important aspect of the residual method not mentioned in Mathalon et al. (1993) is that although homogeneity of regression line slopes across groups is not an assumption, if heterogeneity exists, heterogeneity of group variances may also result. This heteroscedasticity violates assumptions of the significance test of group differences in the residuals, and a data transformation or nonparametric test might be necessary.

Below, we examine the agreement in a particular example among these adjustment methods using data collected from bipolar and psychotic children and normal controls. We also examine the degree to which anthropometric-size-adjustment-dependent measures are interchangeable.

4. A quantitative example illustrating agreement among methods

4.1. Subjects

Structural MRI data were collected from 83 subjects (35 males and 48 females) ranging in age from 6.2 to 16.9 years. The mean age was 11.4 ± 2.78 years and did not differ significantly between the two diagnostic groups. The sample was ethnically homogeneous with 77 of the 83 subjects being Caucasian. The subjects were taken from a study of psychosis and bipolar disorder, but for the purposes of this exercise, subjects were either considered to be "patient" or "normal controls" with the patient group consisting of the pooled bipolar and psychosis groups. All subjects' guardians gave written informed consent for a protocol approved by the McLean Hospital Institutional Review Board, and the subjects themselves gave written ascent.

We initially controlled for sex in all analyses, but the analyses reported here do not control for sex. The results did not differ appreciably between sex (and sex distributions did not vary significantly across diagnostic groups), and this is done only to make the presentation of the methods clearer. We would like to emphasize that, in general, controlling for effects such as age and sex is often useful since they may be important confounders. A recent study, however, found that many regional volumetric differences could be attributed to individual differences in cerebral volume rather than to sex (Leonard et al., 2008).

Volumes were calculated for segmentation (Filipek et al., 1994) and cortical parcellation units (Rademacher et al., 1992; Caviness et al., 1996) derived from MRI imaging analysis of the brains of these subjects. Table 1 shows how age, head circumference, height, weight, and brain volumes differ between these two groups.

4.2. Imaging acquisition and analysis

Structural imaging was performed at the McLean Hospital Brain Imaging Center on a 1.5 T Scanner (Signa; GE Medical Systems, Milwaukee, WI, USA). Acquisitions included a conventional T_1-weighted sagittal scout series (20 slices), a proton density/T_2-weighted interleaved double-echo axial series [120 slices, slice thickness = 3 mm, field of view (FOV) = 24 cm², TR = 3 s, TE = 30/80 ms, acquisition matrix = 256–192, number of excitations = 0.5], and a three-dimensional inversion recovery-prepped spoiled gradient recalled echo coronal series which was used for structural analysis (124 slices, prep = 300 ms, TE = 1 min, flip angle = 25°, FOV = 24 cm², slice thickness = 1.5 mm, acquisition matrix = 256-by-192, number of excitations = 2). All scans were reviewed by a clinical neuroradiologist to rule out gross pathology.

Table 1
Demographic and volumetric summary statistics.

Measure	"Control" group		"Patient" group		p
	Mean	S.D.	Mean	S.D.	
Age	11.05	2.67	11.64	2.83	0.355
Head circumference	54.72	1.78	54.45	1.62	0.486
Height	56.86	6.49	57.82	5.56	0.499
Weight	94.67	35.20	112.19	35.10	0.040
Total brain volume	1382.81	83.25	1323.21	114.06	0.015
Cerebral cortex	726.63	53.01	684.70	67.44	0.005
Cerebral white matter	408.81	41.98	397.48	47.08	0.282
Cerebellar white matter	28.82	3.23	28.21	3.27	0.419
Cerebellar cortex	121.36	11.70	118.14	11.02	0.219
Hippocampus	7.79	0.78	7.38	0.87	0.039
Amygdala	3.58	0.56	3.40	0.49	0.125
Caudate	8.59	0.85	8.36	0.96	0.290
Thalamus	16.36	1.21	15.72	1.13	0.019
Putamen	11.10	1.03	10.84	1.12	0.319
Nucleus accumbens	1.38	0.26	1.39	0.28	0.916
Temporal pole	25.71	4.00	25.26	5.21	0.750
Precentral gyrus	39.64	6.26	38.38	5.62	0.468
Ant. parahippocampal gyrus	6.11	1.02	6.74	1.33	0.089
Post. superior temporal gyrus	11.74	1.86	10.09	2.36	0.014
Superior frontal gyrus	33.29	6.22	30.22	5.70	0.082
Paracingulate gyrus	14.97	1.86	14.66	2.28	0.626
Frontal pole	80.11	7.79	76.00	11.48	0.182
Post. cingulate gyrus	13.73	2.94	13.93	2.63	0.798
Planum polare	3.90	0.81	3.51	0.64	0.068
Insula	17.57	1.33	16.61	2.03	0.076

We utilized a method for comprehensive volumetric profiling developed at the Center for Morphometric Analysis at the Massachusetts General Hospital and described in detail elsewhere (Rademacher et al., 1992; Filipek et al., 1994; Caviness et al., 1996). This method has been applied to the study of brain volumes in a variety of neuropsychiatric disorders (Rauch et al., 2001; Herbert et al., 2003, 2004; Takeoka et al., 2004; Frazier et al., 2005; Makris et al., 2008).

Briefly, major gray and white matter regions are manually segmented in coronal slices throughout an unwarped and untransformed brain. Then, the cerebral cortex is parcellated through manual delineation of a canonical set of sulci and gyri that are then used to semi-automatically parcellate the entire cortex. The comprehensive set of analytical units thus produced is well suited for a comparative assessment of methods for adjusting for head, brain or body size.

In order to assess the relationship between the three common methods of adjusting for head, brain, or body size in volumetric MRI studies, we analyzed data from ten segmentation structures and ten cortical parcellation units. The segmentation structures considered were cerebral cortex, cerebral white matter, cerebellar cortex, cerebellar white matter, hippocampus, amygdala, caudate, thalamus, putamen and nucleus accumbens. The parcellation units (PUs) considered were temporal pole, precentral gyrus, anterior parahippocampal gyrus, posterior superior temporal gyrus, superior frontal gyrus, paracingulate cortex, frontal pole, posterior cingulate gyrus, planum polare, and insula. These parcellation units were selected based on results from a previous inter-rater reliability study (Caviness et al., 1996; Kennedy et al., 1998) and coefficients of variation (CV = [standard deviation/mean]*100). All PUs included in this study had an intra class correlation coefficient of at least 0.8 and a CV below 20. All residuals generated from the segmentation and parcellation units were found to have an approximate normal distribution.

4.3. Adjustment factors

When adjusting the ROIs using purely brain volumetric measures (as opposed to head circumference, height, or weight), the adjustment factors were assigned as follows: total brain volume (not including ventricles) was used when examining the segmentation structures throughout the brain, while total cerebral cortex volume was used when examining cortical subdivisions or parcellation units. The goal was to maintain an acceptable level of variance in the adjusted measures. This was accomplished by ensuring that the adjustment factors were reasonably larger than the ROIs for which they are adjusting. In the case of the cortical parcellation units, the cerebral cortex was chosen as the adjustment factor to minimize the impact of non-cortical structures on the cortex-specific comparisons.

4.4. Software

For the proportion and ANCOVA procedures, we utilized the GLM procedure in SAS 9.1.3 (SAS Institute, Cary, NC). We used the open source package R 2.5.1 for the residual procedure (http://www.r-project.org), and Stata/SE 11 for all graphics (Stata Corporation, College Station, TX).

4.5. Concordance among methods

There is no formal way to compare the main effect of group across methods since they are modeled so differently, as described previously. However, one can assess the degree of agreement across them in terms of ordinal ranking, and to do so we used Kendall's coefficient of concordance (also known as Kendall's W). Kendall's W ranks the magnitudes of the group effects for the ROIs, within each of the three methods, and compares the ranks across methods. If the ordering of the magnitudes of the group effects across ROIs is identical for each method, then Kendall's W is equal to 1. In this case, we say that there is perfect rank concordance; if there is no concordance at all (i.e., each ROI has a completely different rank depending on method) then Kendall's W is equal to 0. The p-value associated with Kendall's W corresponds to a test of the null hypothesis of no agreement. Thus, a significant p-value indicates that the value of Kendall's W is significantly larger than 0. This procedure is a nonparametric method commonly used when distributional assumptions cannot be made and a comparison between more than two dependent groups is to be performed. When comparing only two methods, we used Spearman's *rho*. Spearman's *rho* is a nonparametric rank correlation procedure for paired data that gives values between −1 and 1, indicating perfect negative and positive rank correlation, respectively. Note that both of these nonparametric procedures indicate the degree to which two or more sets of ranks have similar orderings. Thus, we can infer whether the group effects for the ROIs tend to increase in similar ways across methods using these measures, but they do not provide a way to formally compare the magnitudes of these group effects.

5. Results

5.1. Assessment of the agreement between statistical methods

5.1.1. Assessment of concordance among methods

The group effect p-values associated with each of the methods we assessed are tabulated for the segmentation structures in Table 2 and for the parcellation units in Table 3. P-values are reported in place of t-statistics or test statistics; because the degrees of freedom differ depending on which method is used (due to inherent differences among the tests), the test statistics are not directly comparable between methods. However, the p-values are, in a sense, standardized versions of these test statistics and can be compared across adjustment methods. Kendall's W for the adjusted segmented volumes was 0.666 ($p = 0.04$) indicating a high level of concordance among the three adjustment methods. Similarly, a high degree of concordance was found for the parcellation units with a Kendall's W of 0.795 ($p = 0.01$). This indicates that the ordering of the magnitude of group effects by ROI was significantly similar among adjustment methods. However, the magnitude of these effects was different depending on analytic method, with the ANCOVA approach providing somewhat smaller group effects

Table 2
P-values for TBV adjustment of segmentation structures by method.

p-value (rank)				
Segmentation structures	Raw	ANCOVA	Proportion	Residual
Cerebral cortex	0.0048 (10)	0.1525 (10)	0.1121 (10)	0.0760 (10)
Cerebral white matter	0.4193 (2)	0.6689 (4)	0.2813 (6)	0.2780 (7)
Cerebellar white matter	0.2815 (5)	0.2331 (8)	0.3526 (5)	0.1790 (8)
Cerebellar cortex	0.2187 (6)	0.6301 (5)	0.2683 (9)	0.4400 (4)
Hippocampus	0.0390 (8)	0.3628 (6)	0.6478 (3)	0.1650 (9)
Amygdala	0.1249 (7)	0.7676 (2)	0.7643 (1)	0.7460 (1)
Caudate	0.2904 (4)	0.7510 (3)	0.3952 (4)	0.6760 (2)
Thalamus	0.0192 (9)	0.3261 (7)	0.6493 (2)	0.5900 (3)
Putamen	0.3187 (3)	0.9629 (1)	0.2748 (7)	0.4370 (5)
Nucleus accumbens	0.9163 (1)	0.2152 (9)	0.2699 (8)	0.3540 (6)

overall. This finding is consistent with previous reports in the MRI literature (cf. Seidman et al., 1999).

5.1.2. Pairwise comparison of adjustment methods

In order to compare pairs of adjustment methods separately, Spearman's rho rank correlations were calculated using the p-value ranks for each pair of adjustment methods for the segmentation structures and parcellation units. We also compared each adjustment method to the group effects obtained from unadjusted analyses (i.e., an ANCOVA with the ROI volume as the dependent variable and age and group as independent variables, but without a measurement of head or body size in the model). Results of analogous comparisons are reported for the parcellation units. A moderate degree of correlation was found between the ANCOVA and residual method for both the segmentation structures ($r = 0.62$, $p = 0.06$) and parcellation units ($r = 0.53$, $p = 0.12$). The proportion method did not correlate highly with either the residual ($r = 0.49$, $p = 0.15$) or ANCOVA ($r = 0.38$, $p = 0.27$) adjustment methods for the segmentation structures. However, the residual ($r = 0.73$, $p = 0.02$) and ANCOVA ($r = 0.82$, $p < 0.01$) adjustment methods did have significant agreement with the proportion method for the parcellation units.

5.1.3. Pairwise comparison of adjusted and unadjusted results

There was little correlation among the adjusted and unadjusted results, regardless of the adjustment method used, for both the segmentation structures and parcellation units (all p-values greater than 0.25).

5.2. Assessment of the exchangeability of head, brain and body size measures used for adjustment

5.2.1. Correlation of head, brain and body size measures

In order to assess whether the choice of the adjustment measure one uses for addressing the influence of head, brain, or body size

Table 3
P-values for TCCX adjustment of parcellation units by method.

p-value (rank)				
Parcellation units	Raw	ANCOVA	Proportion	Residual
Temporal pole	0.7504 (2)	0.2641 (8)	0.5589 (4)	0.6740 (2)
Precentral gyrus	0.4679 (4)	0.6859 (2)	0.6116 (3)	0.5500 (4)
Ant. parahippocampal gyrus	0.0892 (6)	0.0589 (9)	0.0170 (10)	0.1990 (8)
Post. superior temporal gyrus	0.0139 (10)	0.0467 (10)	0.0886 (9)	0.0110 (10)
Superior frontal gyrus	0.0816 (7)	0.3297 (6)	0.3784 (6)	0.4410 (6)
Paracingulate gyrus	0.6256 (3)	0.4491 (4)	0.4273 (5)	0.6580 (3)
Frontal pole	0.1815 (5)	0.9361 (1)	0.9528 (1)	0.9720 (1)
Post. cingulate gyrus	0.7982 (1)	0.3371 (5)	0.2034 (8)	0.1080 (9)
Planum polare	0.0676 (9)	0.2647 (7)	0.3192 (7)	0.4430 (5)
Insula	0.0763 (8)	0.5398 (3)	0.7803 (2)	0.2160 (7)

yields comparable results in these data, we also considered head circumference, height, and weight as adjustment factors (in addition to the volumetric measures) when using the ANCOVA method. In our data, we saw a varying amount of correlation among these measures and the volumetric measures of cerebral cortex or total brain volume. Table 4 illustrates the correlations among these measures, with bold values indicating correlations that are larger than 0.5. These results indicate that height and weight are not closely related to the volumetric measures we assessed (i.e., either to cerebral cortex or to total brain volume). However, head circumference, height, and weight were all highly correlated with each other. Thus there is a reasonable degree of overlap in the information provided by these three measures of head or body size, but not a high degree of overlap with the volumetric measures of brain size.

5.2.2. Comparison of head, brain, and body size measures

To investigate the exchangeability of head circumference, height, and weight (i.e., non-volumetric measures) with a volumetric measure of brain size (either total brain volume or total cerebral cortex volume), we considered the ANCOVA method for our ten segmentation structures and ten parcellation units using each measure as the adjustment factor. Tables 5 and 6 show group effect p-values for the analyses of unadjusted volumes and of volumes adjusted for size by volume (total brain for the segmentation structures, and cerebral cortex for the parcellation units), and by non-volume measures (head circumference, height, and weight). Again using Kendall's coefficient of concordance, the ranks of group effect p-values obtained using different adjustment factors showed a significant amount of agreement for segmented structures (Kendall's $W = 0.727$, $p = 0.002$) and parcellation units (Kendall's $W = 0.721$, $p = 0.002$).

5.2.3. Pairwise comparisons of adjustment measures: non-volumetric measures

In order to compare pairs of adjustment measures separately, Spearman's rho rank correlations were calculated for p-values obtained using different adjustment measures in the ANCOVA method, as well as p-values from unadjusted results. For the segmented structures, there was a high amount of rank correlation between the unadjusted group effects and the group effects obtained from the ANCOVA method in which non-volumetric measures of head or body size were used for adjustment (head circumference, $r = 1$, $p < 0.0001$; height, $r = 0.96$, $p < 0.0001$; weight, $r = 0.96$, $p < 0.0001$). For parcellation units, a similarly high level of rank agreement was found comparing the unadjusted effects with those from analyses using head circumference ($r = 0.92$, $p < 0.001$), height ($r = 0.98$, $p < 0.0001$) and weight ($r = 0.95$, $p < 0.0001$). When the group effects from analyses using non-volumetric measures were compared to each other, high rank correlations were found for both segmentation structures (all $r > 0.96$, $p < 0.0001$) and parcellation units (all $r > 0.90$, $p < 0.001$).

Table 4
Correlations among cranial size adjustment measures. Bold values indicate $p < 0.05$.

	Head circumference	Height	Weight	Total brain volume	Cerebral cortex volume
Head circumference	1.000				
Height	**0.591**	1.000			
Weight	**0.673**	**0.709**	1.000		
Total brain volume	0.409	−0.122	−0.150	1.000	
Cerebral cortex volume	0.292	−0.300	−0.231	**0.882**	1.000

Table 5
Comparison of cranial measures used for adjustment of segmentation structures. Adjustments made using TBV.

p-value (rank)					
Segmentation structures	Raw	Volume	Head circumference	Height	Weight
Cerebral cortex	0.0048 (10)	0.1525 (10)	0.0066 (10)	0.0028 (10)	0.0069 (10)
Cerebral white matter	0.4193 (2)	0.6689 (4)	0.5576 (2)	0.2482 (4)	0.2258 (3)
Cerebellar white matter	0.2815 (5)	0.2331 (8)	0.3735 (5)	0.1524 (5)	0.1831 (5)
Cerebellar cortex	0.2187 (6)	0.6301 (5)	0.2982 (6)	0.1451 (6)	0.1748 (6)
Hippocampus	0.0390 (8)	0.3628 (6)	0.0519 (8)	0.0291 (8)	0.0392 (8)
Amygdala	0.1249 (7)	0.7676 (2)	0.1511 (7)	0.0798 (7)	0.0890 (7)
Caudate	0.2904 (4)	0.7510 (3)	0.3742 (4)	0.2627 (3)	0.2582 (2)
Thalamus	0.0192 (9)	0.3261 (7)	0.0262 (9)	0.0086 (9)	0.0141 (9)
Putamen	0.3187 (3)	0.9629 (1)	0.4238 (3)	0.3182 (2)	0.1875 (4)
Nucleus accumbens	0.9163 (1)	0.2152 (9)	0.8185 (1)	0.8693 (1)	0.8449 (1)

5.2.4. Pairwise comparisons of adjustment measures: volumetric vs. non-volumetric measures

When comparing the segmentation structure results from the ANCOVA in which total brain volume was used for adjustment to an ANCOVA in which a non-volumetric measure of head or body size was used for adjustment, the correlations were weaker (all $r<0.35$). The same was true for parcellation unit effects when comparing results from an ANCOVA using total cerebral cortex volume to results from ANCOVAs that used non-volumetric adjustment measures (all $r<0.42$). This indicates that, while there is a high degree of similarity between the results of the unadjusted analyses on the one hand and analyses adjusted using head circumference, height, and weight on the other, the group effects obtained from the ANCOVA method adjusted using volumetric adjustment measures resulted in markedly different results.

Tables 5 and 6 also indicate that not only were the rankings of the group effects similar between the unadjusted results and the results from the adjusted analyses using a non-volumetric measure, but the magnitudes of the p-values were also quite similar. However, the ANCOVA results obtained from using a volumetric measure of head size for adjustment had noticeably different group effect p-value magnitudes, indicating a substantial difference in the ANCOVA results depending on the measure of head size used for adjustment.

6. Discussion

The methods illustrated in this paper are widely applicable to a variety of volumetric imaging situations. Although we used a pediatric sample (in which TBV and ICV would be expected to be highly correlated) where the relationship between head size and ROI volume is linear, the strategy implemented is generalizable to work with nonlinear associations with head size as well. This includes situations in which there may be a differential relationship between ROI volume and head size depending on group membership (Ueda et al., 2010). We also note that consideration of head size, and possible adjustment for it, are also recommended when using voxel-based morphology (Ridgway, et al., 2010).

6.1. Considerations in adopting statistical adjustment methods

In considering the three statistical adjustment methods, an investigator might consider the following factors:

(1) One should ask *why* linear, curvilinear, intercept, no intercept, homogeneous and inhomogeneous regression slopes and variance effects such as those described earlier might be occurring. For example, a curvilinear relationship between an ROI volume and head size might hold in a patient group because the ROI volume increases in tandem with the head size to a point. However, as a result of the illness that affects a single primary ROI, at some stage the ROI volume begins to level off or even decrease while head size steadily increases.

(2) Further, one should consider why group differences in head size and/or other covariates might be obtained. For example, if two groups differ in their mean TBV, once needs to as whether this effect is just an irrelevant, nuisance, or chance difference to be adjusted for statistically and forgotten, or does it mean something substantively important (e.g., the disease is causing the TBV to atrophy in the patient group)? If the latter applies, then the question remains as to whether group differences in ROI volumes should be "adjusted for" the TBV difference or not. Perhaps neither the unadjusted nor adjusted ROI volume methods is "wrong" and *both* should be reported because they answer different, complementary questions.

(3) Moreover, it is important to consider how the research question is framed. Clarifying the question often leads one to the appropriate method that addresses that particular question in the best way. For example, is it important that one group has a lower unadjusted mean volume for an ROI than another group, regardless of whether this is "due to" its smaller TBV or not, or is only the ROI's relative size compared to (i.e., adjusted for) the TBV of any consequence? One might also ask whether an ROI's size differs relative to head circumference, which does not change much after middle age, or instead, whether the

Table 6
Comparison of cranial measures used for adjustment of parcellation units. Adjustments made using total cerebral cortex volume.

p-value (rank)					
Parcellation units	Raw	Volume	Head circumference	Height	Weight
Temporal pole	0.7504 (2)	0.2641 (8)	0.7863 (1)	0.7966 (2)	0.9310 (1)
Precentral gyrus	0.4679 (4)	0.6859 (2)	0.5010 (4)	0.3381 (4)	0.4612 (4)
Ant. parahippocampal gyrus	0.0892 (6)	0.0589 (9)	0.0646 (9)	0.1553 (7)	0.1812 (6)
Post. superior temporal gyrus	0.0139 (10)	0.0467 (10)	0.0150 (10)	0.0085 (10)	0.0105 (10)
Superior frontal gyrus	0.0816 (7)	0.3297 (6)	0.0962 (6)	0.1760 (6)	0.0848 (7)
Paracingulate gyrus	0.6256 (3)	0.4491 (4)	0.6355 (3)	0.4905 (3)	0.7660 (2)
Frontal pole	0.1815 (5)	0.9361 (1)	0.2165 (5)	0.2046 (5)	0.1874 (5)
Post. cingulate gyrus	0.7982 (1)	0.3371 (5)	0.7402 (2)	0.9797 (1)	0.7190 (3)
Planum polare	0.0676 (9)	0.2647 (7)	0.0805 (8)	0.1104 (8)	0.0818 (8)
Insula	0.0763 (8)	0.5398 (3)	0.0909 (7)	0.0668 (9)	0.0794 (9)

ROI's size differs relative to TBV. These issues imply different analyses to be employed.

6.2. GLM and graphical tools in decision making

One helpful preliminary step in decision making is to perform graphical and GLM analyses of the data, and then make an informed judgment about the best fitting model. We provide an example with a single ROI from our data summarized in the Appendix. It is important to examine scatterplots that have symbols to denote group membership with within-group regression lines and/or quadratic curves overlaid. The best method, in general, is to run a GLM on the data with multiple runs testing different models of varying complexity with all analyses done in tandem with graphical analysis. The GLM strategy is essentially a multiple regression (or ANCOVA) model that detects possible group effects using one or more group indicator variables. It should include as many meaningful covariates as is necessary and reasonable and as the sample size permits. Quadratic (or cubic, logarithmic) terms for covariates suspected (or noticed via graphical tools) to have curvilinear relations with the dependent variable also need to be assessed along with interactions of the group indicator(s) with these covariates. All relevant demographic and diagnostic covariates also need to be included. Covariates that are not significant to the model can be removed though a backward stepwise model building strategy. Factorial ANOVA designs including interactions can be embedded into the model. The proportion approach, classic ANCOVA, and the residual method will fall out as special cases of the GLM if appropriate. A test for a nonzero intercept would be relevant to the appropriateness of the ratio method, whereas tests of group-by-covariate interactions assess the appropriateness of conventional ANCOVA with its homogeneous slope assumptions. After determining optimal fitting and statistically significant relations, a graph of predicted values within the range of the head size measure in the data at representative covariate values can be helpful to interpretation, especially when complex group interactions or curvilinear relations are present.

If there are a large number of ROIs to analyze, this elaborate multi-stage analysis may not be practically feasible, and reasonable assumptions based on substantive theory and previous research may be employed to reduce the number of analyses. Also, if there are many ROIs to analyze, the possibility of chance significant effects due to multiple tests becomes a concern and should be addressed. There are many methods to perform such corrections (e.g., the Benjamini–Hochberg False Discovery rate, Sidak, step-down methods, and permutation and bootstrap resampling methods) (Tobias, 2000). Similarly, multivariate regression modeling techniques provide omnibus tests for differences in groups of ROIs that help control for the problem of multiple tests (Cnaan et al., 1997; Herbert et al., 2003; Goldstein et al., 2007).

6.3. Conclusions

We have brought to light a number of important considerations that researchers should evaluate when analyzing volumetric MRI data. Although agreement among methods has generally been seen as indicating stronger support of a research hypothesis, it is crucial that one understand the appropriateness of the models being applied. Our recommendation is to gain an initial understanding of the relationship that the ROI volume has with head size via the use of graphical methods allowing one to detect possible non-linear or interactive effects within and between study groups. A step down GLM approach can then be used to eliminate predictors which are not important in the model (although any variables, such as age and gender, that are thought to be confounders should not be regardless of the statistical significance). Although the example we use is pediatric consisting of a mixture of bipolar and psychotic subjects with controls, the results are generalizable to any study group. One should be careful to use an appropriate measure of head size noting that the correlation between

TBV and ICV generally decreases with age. The methods advocated here do not change with the choice of the head size proxy. An additional advantage to the GLM approach is that it provides great generality in that it can be used for studies that do not involve groups (e.g., asking whether an ROI volume correlates with numeric scores on a memory test within a single sample of healthy normal people, or adjusting for a relation of the ROI volume to TBV and other confounding covariates such as age), and is thus robust to a wide variety of applications.

Acknowledgements

Appendix A. Worked example using a generalized linear modeling approach

Here we consider total anterior hippocampal gyrus to illustrate the method of analysis that we suggest as a guide to the analysis of structural MRI data. We note that no one method can be considered a gold standard and only suggest this as a method that will help tease out group differences within the data and guide analysis.

We recommend always beginning with graphical analyses of the data. Fig. 1 shows anterior parahippocampal gyrus volumes plotted against cerebral cortex volume with each group (control and diagnostic) indicated with a different plotting symbol. In examining this plot, one can begin to investigate linear and non-linear relationships that may be present in the data. In our data, we see that there seems be to be a linear relationship, but not a quadratic (or higher order polynomial) relationship. However, we will still investigate this in the analysis for illustration.

As a first modeling step, we use a generalized linear model with anterior parahippocampal gyrus as the response, and with group status and cerebral cortex volume as predictors (an ANCOVA model). Initially, we also investigate the significance of a quadratic term for cerebral cortex volume and the interaction of group status with both the linear and quadratic terms for cerebral cortex. Neither the coefficient for the squared cerebral cortex volume nor for the interaction with the squared cerebral cortex volume was significantly different from zero (as we suspected from the lack of a visually
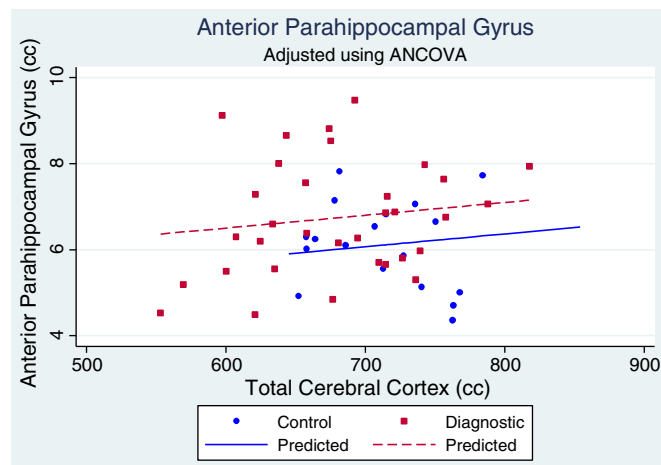

Fig. 1. Scatterplot of anterior parahippocampal volumes (PHA) vs. total cerebral cortex volume. ANCOVA adjusted PHA means are regression line intercepts.
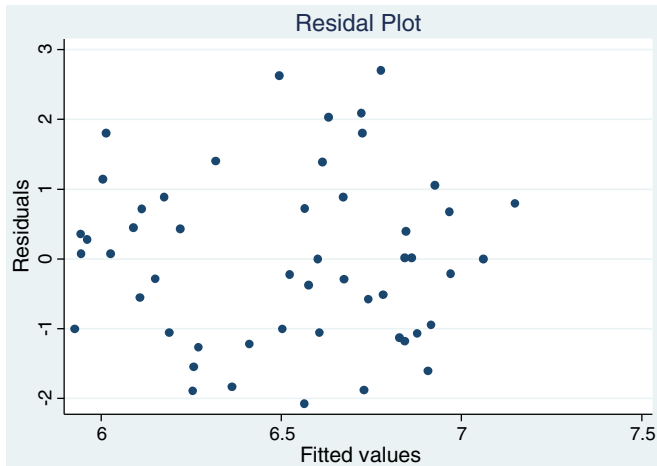
Fig. 2. Residual plot for PHA.

apparent quadratic relationship in Fig. 1). This leaves us with the following model,

$$E[\text{PHA}] = \beta_0 + \beta_1 I_{group = diagnostic} + \beta_2(\text{CCTX}) + \beta_3 \left(I_{group = diagnostic} \right)(\text{CCTX}),$$

where PHA is the anterior parahippocampal gyrus volume, CCTX is the cerebral cortex volume and $I_{group = diagnostic}$ is an indicator function that equals 1 if the observation comes from a subject in the diagnostic group and 0 otherwise.

It is of interest to note that this model incorporates both the proportional method and ANCOVA method described earlier. If β_3 is not significantly different from zero, then the relationship between PHA and CCTX is the same for each group (i.e., the within-group regression lines would be parallel). If both β_0 and β_1 are not significantly different from 0, then the zero-intercept assumption of the proportion method is satisfied, and group differences are determined by the magnitude of β_1. In our data, the β_0 and β_1 coefficients are marginally significantly different from 0. Thus, the zero-intercept assumption is not strictly satisfied. Nevertheless, we should investigate the significance of the group-by-cerebral cortex volume interaction to determine if the

relationship between anterior parahippocampal gyrus and cerebral cortex volume differs according to diagnostic group status. In our data, this interaction is not significantly different from 0 indicating no difference in this relationship.

We are now left with a generalized linear model that is the ANCOVA model described in this article:

$$E[\text{PHA}] = \beta_0 + \beta_1 I_{group = diagnostic} + \beta_2(\text{CCTX}).$$

The β_1 coefficient is often of primary interest in this model, and in our data it is marginally significantly different from zero ($p = 0.059$). Thus, the adjusted mean anterior parahippocampal gyrus volumes differ by group but fail to reach significance at a strict 5% level. However, it is still useful for illustration purposes. This is shown in Fig. 1 as the vertical distance between the two regression lines. Note that the β_2 term was not significantly different from zero and, therefore, not strictly necessary to include. However, the sample size was large enough to permit leaving it in for any small adjustment it provides and reduction of error variance.

Note that we should check the homogeneity of variance assumption inherent in ordinary least squares regression analysis using residual plots. Fig. 2 shows the residuals plotted against the predicted values of anterior parahippocampal gyrus. This plot should show no obvious patterns and does not have an increase (or decrease) in the vertical spread of the residuals as one looks across the x-axis. Our plot looks satisfactory and we can assume the homogeneity of variance assumption holds.

Since we are performing parametric testing of regression parameters we also need to make sure the residuals after the model is fit are approximately normally distributed. This can be examined graphically through the use of simple histograms, or through examination of normal quantile plots. We also recommend the use of hypothesis testing to establish the tenability of the normal distributional assumption such as the Shapiro–Wilk test (although in large samples such tests may show departure from normality that is statistically significant, but trivial in magnitude and ignorable). For the anterior hippocampal gyrus, the distribution of the residuals is found to not differ from the normal (Shapiro–Wilk $p = 0.47$) and we may use the generalized linear modeling techniques described in this paper. A histogram and normal quantile plot appear in Fig. 3.

While this analysis did not take into account other predictors such as age and gender, the inclusion of them is straightforward when utilizing
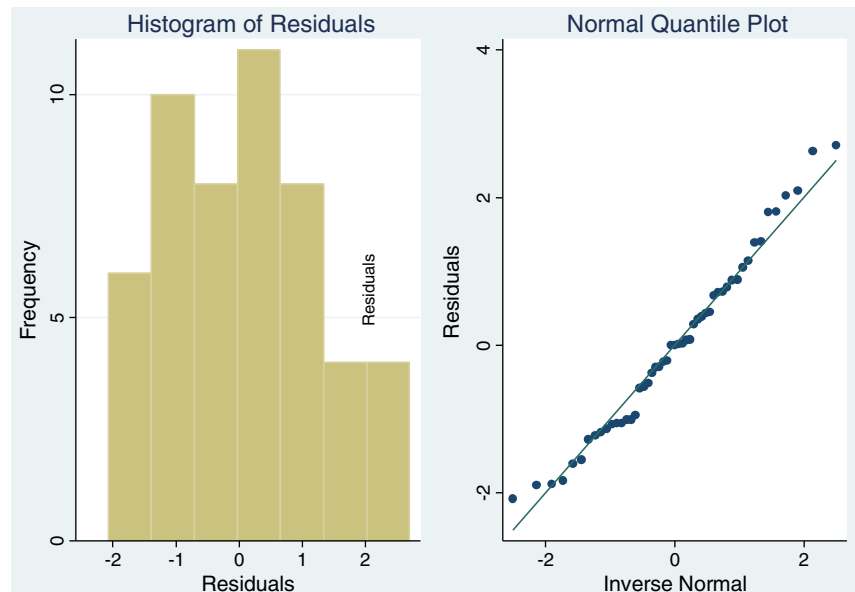

Fig. 3. Distribution of anterior parahippocampal gyrus residuals.

the generalized linear modeling approach. We can simply include them as predictors and the coefficient estimates of interest (generally those that include the diagnostic group indicator) are adjusted accordingly. This modeling approach is more robust than a "blind" ANCOVA approach in which one does not consider non-linear relationships or interactions among predictors within the data. By using the GLM method, one can explore many different relationships that the ROI may have with the measure of head size, including the possibility that those relationships interact in a possibly complex way with other covariates. Both the proportion and ANCOVA methods are special cases of the procedure described here and we thus feel it to be a very useful approach.

Appendix B. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.pscychresns.2011.01.007.

References

Arndt, S., Cohen, G., Alliger, R.J., Swayze, V.W., Andreasen, N.C., 1991. Problems with ratio and proportion measures of imaged cerebral structures. Psychiatry Research: Neuroimaging 40, 79–89.
Bartholomeusz, H.H., Courchesne, E., Karns, C.M., 2002. Relationship between head circumference and brain volume in healthy normal toddlers, children, and adults. Neuropediatrics 33, 239–241.
Bolton, P.F., Roobol, M., Allsopp, L., Pickles, A., 2001. Association between idiopathic infantile macrocephaly and autism spectrum disorders. Lancet 358, 726–727.
Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 23, 724–738.
Bush, E.C., Allman, J.M., 2003. The scaling of white matter to gray matter in cerebellum and neocortex. Brain, Behavior and Evolution 61, 1–5.
Caviness Jr., V.S., Meyer, J., Makris, N., Kennedy, D.N., 1996. MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate reliability. Journal of Cognitive Neuroscience 8, 566–587.
Changizi, M.A., 2001. Principles underlying mammalian neocortical scaling. Biological Cybernetics 84, 207–215.
Chen, K.H.M., Chuah, L.Y.M., Sim, S.K.Y., Chee, W.L., 2010. Hippocampal region-specific contributions to memory performance in normal elderly. Brain and Cognition 72, 400–407.
Cheong, J.L.Y., Hunt, R.W., Anderson, P.J., Howard, K., Thompson, D.K., Wang, H.X., Bear, M.J., Inder, T.E., Dotle, L.W., 2008. Head growth in preterm infants: correlation with magnetic resonance imaging and neurodevelopmental outcome. Pediatrics 121, E1534–E1540.
Cnaan, A., Laird, N.M., Slasor, P., 1997. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Statistics in Medicine 16, 2349–2380.
Fidler, D.J., Bailey, J.N., Smalley, S.L., 2000. Macrocephaly in autism and other pervasive developmental disorders. Developmental Medicine and Child Neurology 42, 737–740.
Filipek, P.A., Richelme, C., Kennedy, D.N., Caviness, V.S., 1994. The young adult human brain: an MRI-based morphometric analysis. Cerebral Cortex 4, 344–360.
Fombonne, E., Roge, B., Claverie, J., Courty, S., Fremolle, J., 1999. Microcephaly and macrocephaly in autism. Journal of Autism and Developmental Disorders 29, 113–119.
Frazier, J.A., Breeze, J.L., Makris, N., Giuliano, A.S., Herbert, M.R., Seidman, L., Biederman, J., Hodge, S.M., Dieterich, M.E., Gerstein, E., Kennedy, D.N., Rauch, S.L., Cohen, B., Caviness, V.S., 2005. Cortical gray matter differences identified by structural magnetic resonance imaging in pediatric bipolar disorder. Bipolar Disorders 7, 555–569.
Goldstein, J.M., Goodman, J.M., Seidman, L.J., Kennedy, D.N., Makris, N., Lee, H., Tourville, J., Caviness, V.S., Faraone, S.V., Tsuang, M.T., 1999. Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. Archives of General Psychiatry 56, 537–547.
Goldstein, J.M., Seidman, L.J., Makris, N., Ahern, T., O'Brien, L.M., Caviness Jr., V.S., Kennedy, D.N., Faraone, S.V., Tsuang, M.T., 2007. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. Biological Psychiatry 61, 935–945.
Greenberg, D.L., Messer, D.F., Payne, M.E., MacFall, J.R., Provenzale, J.M., Steffens, D.C., Krishman, R.R., 2008. Aging, gender, and the elderly adult brain: an examination of analytical strategies. Neurobiology of Aging 29, 290–302.
Herbert, M.R., Ziegler, D.A., Deutsch, C.K., O'Brien, L.M., Lange, N., Bakardjiev, A., Hodgson, J., Adrien, K.T., Steele, S., Makris, N., Kennedy, D., Harris, G.J., Caviness Jr., V.S., 2003. Dissociations of cerebral cortex, subcortical and cerebral white matter volumes in autistic boys. Brain 126, 1182–1192.
Herbert, M.R., Ziegler, D.A., Makris, N., Filipek, P.A., Kemper, T.L., Normandin, J.J., Sanders, H.A., Kennedy, D.N., Caviness, V.S., 2004. Localization of white matter volume increase in autism and developmental language disorder. Annals of Neurology 55, 530–540.
Kennedy, D.N., Lange, N., Makris, N., Bates, J., Meyer, J., Caviness Jr., V.S., 1998. Gyri of the human neocortex: an MRI-based analysis of volume and variance. Cerebral Cortex 8, 372–384.
Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied Linear Statistical Models. McGraw-Hill, New York, pp. 278–289.
Lainhart, J.E., Piven, J., Wzorek, M., Landa, R., Santangelo, S.L., Coon, H., Folstein, S.E., 1997. Macrocephaly in children and adults with autism. Journal of the American Academy of Child and Adolescent Psychiatry 36, 282–290.
Leonard, C.M., Towler, S., Welcome, S., Halderman, L.K., Otto, R., Eckhart, M.A., Chiarello, C., 2008. Size matters: cerebral volume influences sex differences in neuroanatomy. Cerebral Cortex 18, 2920–2931.
Locascio, J.J., Cordray, D.S., 1983. A reanalysis of Lord paradox. Educational and Psychological Measurement 43, 115–126.
Makris, N., Oscar-Berman, M., Jaffin, S.K., Hodge, S.M., Kennedy, D.N., Caviness, V.S., Marinkovic, K., Breiter, H.C., Gasic, G.P., Harris, G.J., 2008. Biological Psychiatry 64, 192–202.
Mathalon, D.H., Sullivan, E.V., Rawles, J.M., Pfefferbaum, A., 1993. Correction for head size in brain-imaging measurements. Psychiatry Research: Neuroimaging 50, 121–139.
McCaffery, P., Deutsch, C.K., 2005. Macrocephaly and the control of brain growth in autistic disorders. Progress in Neurobiology 77, 38–56.
Mormino, E.C., Kluth, J.T., Madison, C.M., Rabinovici, G.D., Baker, S.L., Miller, B.L., Koeppe, R.A., Mathis, C.A., Weiner, M.W., Jagust, W.J., Initiative, Alzheimer's Disease Neuroimaging, 2009. Episodic memory loss is related to hippocampal-mediated beta-amyloid deposition in elderly subjects. Brain 132, 1310–1323.
O'Brien, L.M., Ziegler, D.A., Deutsch, C.K., Kennedy, D.N., Goldstein, J.M., Seidman, L.J., Hodge, S., Makris, N., Caviness, V., Frazier, J.A., Herbert, M.R., 2006. Adjusting whole brain and cranial size in volumetric brain studies: a review of common adjustment factors and statistical methods. Harvard Review of Psychiatry 14, 141–151.
Peters, M., Jancke, L., Staiger, J.F., Schlaug, G., Huang, Y., Steinmetz, H., 1998. Unresolved problems in comparing brain sizes in Homo Sapiens. Brain and Cognition 37, 254–285.
Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness Jr., V.S., 1992. Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. Journal of Cognitive Neuroscience 4, 352–374.
Rauch, S.L., Makris, N., Cosgrove, G.R., Kim, H., Cassem, E.H., Price, B.H., Baer, L., Savage, C.R., Caviness Jr., V.S., Jenike, M.A., Kennedy, D.N., 2001. A magnetic resonance imaging study of regional cortical volumes following stereotactic anterior cingulotomy. CNS Spectrums 6, 214–222.
Rice, S.A., Bigler, E.D., Cleavinger, H.B., Tate, D.F., Sayer, J., McMahon, W., Ozonoff, S., Lu, J., Lainhart, J.E., 2005. Macrocephaly, corpus callosum morphology, and autism. Journal of Child Neurology 20, 34–41.
Ridgway, B.J., Henley, S.M., Lehman, M., Hobbs, N., Clarkson, M.J., Macmanus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age, and gender adjustments in MRI studies: a necessary nuisance? Neuroimage 53, 1244–1255.
Seidman, L.J., Faraone, S.V., Goldstein, J.M., Goodman, J.M., Kremen, W.S., Toomey, R., Tourville, J., Kennedy, D., Makris, N., Caviness, V.S., Tsuang, M.T., 1999. Thalamic and amygdala–hippocampal volume reductions in first-degree relatives of patients with schizophrenia: an MRI-based morphometric analysis. Biological Psychiatry 46, 941–954.
Sullivan, E.V., 2003. Compromised pontocerebellar and cerebellothalamocortical systems: speculations on their contributions to cognitive and motor impairment in nonammesic alcoholism. Alcocholism: Clinical and Experimental Research 27, 1409–1419.
Sullivan, E.V., Rosenbloom, M.J., Desmond, J.E., S.V., Pfefferbaum, A., 2001. Sex differences in corpus callosum size: relationship to age and intracranial size. Neurobiology of Aging 22, 603–611.
Takayanagi, Y., Kawasaki, Y., Nakamura, K., Takahasi, T., Orikabe, L., Toyoda, E., Mozue, Y., Sato, Y., Itokawa, M., Yamasue, H., Kasai, K., Kurachi, M., Okazaki, Y., Matsushita, M., Suzuki, M., 2010. Differentiation of first-episode schizophrenia patients from healthy controls using ROI-based multiple structural brain variables. Progress in Neuro-Psychopharmacology & Biological Psychiatry 34, 10–17.
Takeoka, M., Riviello, J.J., Duffy, F.H., Kim, F., Kennedy, D.N., Makris, N., Caviness, V.S., Holmes, G.L., 2004. Bilateral volume reduction of the superior temporal areas in Landau–Kleffner syndrome. Neurology 63, 289–292.
Tobias, R.D., 2000. Multiple Comparisons and Multiple Tests: Using the SAS System Workbook. SAS Publishing, Cary, NC.
Ueda, K., Fujiwara, H., Miyata, J., Hirao, K., Saze, T., Kawada, R., Fujimoto, S., Tanaka, Y., Sawamoto, N., Fukuyama, H., Murai, T., 2010. Investigating association of brain volumes with intracranial capacity in schizophrenia. Neuroimage 49, 2503–2508.
Van Petten, C., 2004. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia 42, 1394–1413.
Vidal, C.N., Nicholson, R., Boire, J.-Y., Barra, V., DeVito, T.J., Hayashi, K.M., Geaga, J.A., Drost, D.J., Williamson, P.C., Nagalingam, R., Toga, A.W., Thompson, P.M., 2008. Three-dimensional mapping of the lateral ventricles in autism. Psychiatry Research: Neuroimaging 163, 106–115.
Zhang, K., Sejnowski, T.J., 2000. A universal scaling law between gray matter and white matter of cerebral cortex. Proceedings of the National Academy of Sciences of the United States of America 97, 5621–5626.