Adjustment for Whole Brain and Cranial Size in Volumetric Brain Studies: A Review of Common Adjustment Factors and Statistical Methods

Liam M. O’Brien, PhD, David A. Ziegler, BS, Curtis K. Deutsch, PhD, David N. Kennedy, PhD, Jill M. Goldstein, PhD, Larry J. Seidman, PhD, Steven Hodge, MA, Nikos Makris, MD, PhD, Verne Caviness, MD, DPhil, Jean A. Frazier, MD, Martha R. Herbert, MD, PhD

In this article we address analytic challenges inherent in brain volumetrics (i.e., the study of volumes of brains and brain regions). It has sometimes been assumed in the literature that deviations in regional brain size in clinical samples are directly related to maldevelopment or pathogenesis. However, this assumption may be incorrect; such volume differences may, instead, be wholly or partly attributable to individual differences in overall dimension (e.g., for head, brain, or body size). What quantitative approaches can be used to take these factors into account? Here, we provide a review of volumetric and nonvolumetric adjustment factors. We consider three examples of common statistical methods by which one can adjust for the effects of body, head, or brain size on regional volumetric measures: the analysis of covariance, the proportion, and the residual approaches. While the nature of the adjustment will help dictate which method is most appropriate, the choice is context sensitive, guided by numerous considerations—chiefly the experimental hypotheses, but other factors as well (including characteristic features of the disorder and sample size). These issues come into play in logically framing the assessment of putative abnormalities in regional brain volumes. (Harv Rev Psychiatry 2006;14:141–151.)

Keywords: adjustment factors, allometry, autism, brain morphometry, brain size, head size, macrocephaly, MRI, regional brain volumes, statistical methods, volumetrics

Supported, in part, by the Cure Autism Now Foundation and Colby College Division of Natural Sciences.

Original manuscript submitted 20 June 2005; revised manuscript received 3 November 2005, accepted for publication subject to revision 9 February 2006; final manuscript received 8 March 2006.

Correspondence: Martha Herbert, MD, PhD, Pediatric Neurology and Center for Morphometric Analysis, Massachusetts General Hospital, 149 13th St., 6th fl., Charlestown, MA 02129.

© 2006 President and Fellows of Harvard College

DOI: 10.1080/10673220600784119
INTRODUCTION

The study of the brain basis of neuropsychiatric disorders has been greatly enhanced by the emergence of imaging technologies that have allowed the in vivo study of neuroanatomy. While these developments represent a triumph of medical physics, the challenge of how to interpret the resulting data has fallen onto neuroanatomists, brain morphometers, and statisticians. We focus here upon the particular challenges posed by brain volumetrics, the study of volumes of neuroanatomic structures—volumes derived by measuring the size of anatomically delineated brain regions.∗

Though the literature sometimes attributes deviations in the brain region sizes of clinical samples to pathogenic effects or errors in morphogenesis, these dimensions may simply be related to individual differences in head, overall brain, or body size. How do we most effectively determine the extent to which volume differences are attributable to head-, brain-, and body-size differences, and to what degree there is an additional contribution to volume differences related to maldevelopment or pathogenesis? We consider these questions theoretically and then discuss three common adjustment methods and commonly used adjustment factors.

Neuropathology and Volumetrics

Regional brain volumes are a macroscopic reflection of microscopic tissue properties that may affect brain functioning—such as cell size, cell density, cell organization, extracellular materials, and vasculature, all of which are tightly constrained during development and evolution.1 Interindividual and intergroup volume differences may point to altered organization or mechanisms at the regional and neural-systems levels. Such alterations may yield insight into the brain basis of abnormal function, delineating cellular and molecular mechanisms of potential clinical importance.

Isolated focal lesions are typically not reported for neuropsychiatric disorders, and in cases in which a specific lesion might be identified, it could, of course, be incidental to the condition and not the primary cause of illness. Moreover, the volume differences in both whole-brain and anatomic regions between these types of disorders and controls are often subtle and require careful measurement to be detected; they are not apparent to the unaided eye (e.g., they will not be detected during a neuroradiologist’s clinical reading of a scan. Further, rather than being confined to one region, volumetric alteration may involve multiple structures. Detecting multiple regional alterations constitutes an analytic challenge, and the subtlety of many of these changes makes the results of such analyses highly model dependent. We begin with a review of the historical development of allometry below, and consider how brain volume changes may scale differently with different brain and body measures.

Allometry

A key challenge for brain volumetric methodology is sorting out the sources of brain volume variability between individuals, between groups, and between species. One key source of variability is size; in their 1936 article, Huxley and Teissier2 popularized the term allometry to refer to the modeling of changes in dimension or shape as a function of size. Evolutionary biologists, anthropologists, and artists have long grappled with how to accurately scale the relative dimensions of anatomic features. A classic example of an allometric relationship is the proportion between arm spread and height—a correspondence illustrated by Leonardo da Vinci.3,4 In more recent evolutionary studies of brain allometry, interest has extended beyond macroscopic volumetric data to explorations of how microscopic properties of the brain, such as synapse density, neuron number and density, and axonal radius, scale with overall brain size.5

Of particular relevance to the present study are discussions in comparative and evolutionary studies of brain volumes across species, which are relevant to the scaling issues addressed by this review.6 A primary focus of this literature is the debate as to whether all brain structures and regions increase in proportion to each other, the “global” model (i.e., evolution of brain size occurs in a highly constrained fashion),7 or whether selection can act in a “mosaic” manner,8 altering the size of some brain structures without affecting others (i.e., structures that are tightly linked to each other scale uniformly but independently of volume changes in less related structures). There is strong evidence that evolution of brain size occurs globally, through a coordinated and highly constrained alteration of developmental programs, with all structures exhibiting proportional volumetric changes.9 This model is illustrated in Figure 1A, which uses the example of proportions among white matter, gray matter, and the whole brain. That is, the size of a particular brain structure in any given species can easily be predicted given two variables: (1) the structure’s developmental “birthday” and (2) absolute brain size.7,9,10 Nevertheless, other comparative studies suggest that individual structures can indeed grow larger (or smaller, for that matter) without scaling to volumetric changes in the rest of the

∗Brain volumes are commonly derived by separating (“segmenting”) gray from white matter, and then identifying the boundaries of specific brain regions (e.g., thalamus, hippocampus); volumes are then typically derived by a count of voxels. The discussion that follows thus does not directly bear on brain structure measurement techniques (e.g., voxel-based morphometry) that do not specifically yield measurements of regional volumes.
FIGURE 1. Uniform or proportional, versus nonuniform or disproportionate, regional enlargement. In A, the larger brain shows a uniform or proportional increase over the smaller one, whereas in B, the increase is nonuniform, with white matter showing a disproportionate increase compared to gray matter.

<table>
<thead>
<tr>
<th></th>
<th>Total Brain</th>
<th>White Matter</th>
<th>Gray Matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Increase</td>
<td>Proportional increase</td>
<td>Proportional increase</td>
</tr>
<tr>
<td>B</td>
<td>Increase</td>
<td>Disproportionate increase</td>
<td>Relative decrease</td>
</tr>
</tbody>
</table>

brain, as illustrated in Figure 1B. Proponents of mosaic theories of brain evolution suggest that selectively altering the size of brain regions particularly relevant to the ecological demands facing a species might confer some niche- or sensory-specific competitive advantage. This debate in evolutionary allometry is analogous to the issues in volumetrics that we are discussing; that is, do volumes of individual brain structures scale uniformly or nonuniformly to whole brain size? There are different ways to frame this question. First, regarding total brain volume one may take into account its relationship to body size, head size, or both. If a person has a larger than average brain, this volume increase might be explained entirely in relation to body size and not in relation to altered brain functioning; correlations between intracranial volume and height have been found to be around 0.5. But second, volumes of specific regions may differ as a function of neuroanatomical abnormalities and not body size. Here one can draw an analogy with interspecies comparisons: while regional volume differences between species may be related to differences in ecological or sensory niches, volume differences in neuropsychiatric disorders may relate to differences in functional capacities based in disorder-related abnormalities.

Yet there is also a third category of total brain volume difference that influences the interpretation of regional brain volume differences. In some disorders there appears to be an overall shift in mean brain volume that goes beyond what might be expected from body-size considerations. For example, in autistic children, whole brains are, on average, larger than in matched, typically developing individuals.
(although microcephaly is also sometimes observed, if infrequently, in autistic children).

In contrast, it has been reported that children with attention-deficit/hyperactivity disorder have smaller whole brain volumes than normal, though this finding is controversial.

It thus appears that in neuropsychiatric volumetrics we face a dual allometric challenge: our efforts to discern diagnosis-related regional differences may be confounded not only by brain or body size, but also by pathology-related overall brain volume differences that are not all attributable to body-size differences. In order to distinguish between these possibilities, it is critical to make sure that the methods we use to adjust for allometric contributors to volume differences are optimized. This concern motivates our review of common methods that adjust for allometric contributions to brain volumes. First, we take up the question of which covariates or adjustment factors can be used to parcel out the variability due to head, brain, and body size. We then describe statistical methods that have commonly been used to make such adjustments.

**POTENTIAL ADJUSTMENT FACTORS FOR ALLOMETRIC CONTRIBUTORS TO VOLUME DIFFERENCES**

There are numerous measures one might choose as adjustment factors to clarify how much overall growth or size contributes to regional size in brain volumetric studies. Deciding on the most appropriate measure is sometimes difficult since brain size correlates with multiple other anatomical features. We will focus on the issues related to choosing among the anthropometric measures of head circumference, body size, and brain volume as adjustment factors.

**Head Circumference**

The correlation between head circumference and brain volume appears to be conditioned on age. Bartholomeusz and colleagues studied the relationship between head circumference and brain volume in male autistic and typically developing subjects ranging in age from 1.7 to 42 years. They found a significant correlation in both groups between head circumference and brain volume was found to have a somewhat quadratic relationship. One interprets multiple R values in basically the same way as a standard correlation, however. One reason why the strength of the relationship depends on age is that head growth is related to brain growth early in life, but later loss of brain volume does not result in a diminution of head size. This later loss of brain volume has been noted in autism—with children before adolescence, but not adults, having larger brain volumes relative to age-matched controls. There are also gender-associated differences in head circumference. Thus, age and gender, among other factors, would be important to control for in heterogeneous populations, either as covariates or as factors conditioned upon within normative databases.

**Body Parameters: Height and Weight**

The relationship of both height and weight to brain size can also be confounded by age. For example, Pakkenberg and Voigt used a large sample to investigate the relationship between height and brain weight, and found that the relationship between these variables depends on both gender and age. Passingham found a significant correlation between height and brain weight, whereas Jerison found none. Each of these studies used a subset of the data obtained by Pakkenberg and Voigt. Peters and colleagues noted that discrepancies in results are likely due to the different age ranges used in these studies—which again illustrates that age may modify or confound the relationship between height and brain weight.

Given the strong correlation between height and weight, these measures could conceivably be interchangeable for relating body size and head size. However, Schoenemann found that brain size correlates more highly with lean body mass than with total body mass in mammals, indicating that weight may have a different relationship to brain size depending on the body composition. Obviously, an individual who gains a substantial amount of weight is not going to experience a concomitant brain volume increase. Others have investigated the utilization of combined height and weight as covariates to predict brain size. For example, body mass index (BMI) has been found to bear a stronger relationship to brain weight than height alone (see Skellurud). Peters and colleagues, however, noted that this relationship held only within ethnicities. For example, certain ethnic groups that are typically tall and slender have a different relationship between BMI and brain weight than ethnic groups that are more heavily built. The relationship between weight and brain size is confounded not only by ethnicity, but also by gender and age. It appears that, unless the subjects in a

---

*The Passingham study used an age range of 18 to 45, whereas the Pakkenberg and Voigt study used an age range of 28 to 41.*
particular study are similar with respect to demographics such as age, ethnicity, and gender, one cannot assume that the relationships among body parameters and brain size will be constant.

**Brain Volume**

With the advent of MRI came the opportunity to use total brain volume (TBV) or total intracranial volume (ICV) as an adjustment factor, in lieu of a proxy such as head circumference. Given the changes that are seen in brain volume across the lifespan, using TBV as an adjustment factor for regional brain volumes should, in theory, yield results that may differ as a function of developmental or maturational epoch. Differences in TBV related to stages of life, TBV tends to provide an accurate measure of maximum volume, the approach of old age is associated with decreased brain volume (as well as increased cortical thickness, increased ventricular size, and sulcal widening). In aging populations, ICV or head circumference tend to better reflect maximum brain volume reached earlier in life. This relationship is underscored by the maturational differences in the magnitude of correlation between head circumference and TBV, with the two measures showing stronger correlations during childhood and decreased correlation with further age. Thus, at later stages of life, TBV is a less accurate index of maximal brain volume achieved across the entire lifespan; and ICV, which is uncorrelated with age in adults, is thought to serve as a better index. If, however, the question of interest is framed in terms of how much a particular region or structure contributes to overall brain volume at the time of study—rather than at maximal lifetime volume—then TBV may be a more appropriate adjustment factor (see Herbert et al.). Because correlations among head circumference, ICV, and TBV are expected to be high for preadolescent subjects, adjusted regional brain volumes should not vary substantially based on the type of measure used for adjustment. After adolescence, correlation between TBV and ICV diminishes with increasing age.

TBV (excluding ventricles) has often been considered an appropriate adjustment factor when studying volumes of major gray and white matter brain structures. However, when using brain volume measures as an adjustment factor, a choice can be made between using TBV or using volumes of large constituent brain structures (e.g., cerebrum or cerebral cortex). Total cerebral cortex volume bears a closer relationship than TBV to the volume of cortical subdivisions, a situation in which one would not want the normalizing factor to include subcortical structures not under consideration. An important consideration is whether the volumetric measure used as the normalizing factor is considerably larger than the region-of-interest (ROI) volume, since certain measures derived from these normalizing factors (such as the proportionalized measure discussed in the next section) would otherwise be greatly reduced. This reduction in variability would result from the strong positive correlation that an ROI volume would have with a structure that is very similar in size. This reduced variability in the outcome measure would complicate statistical analyses, since it might lower the power to detect group differences.

**STATISTICAL METHODS FOR ALLOMETRIC ADJUSTMENT: THREE COMMON METHODS**

We have outlined adjustment factors and now turn our attention to statistical methods used to make these allometry-related adjustments. Three methods in the literature have most commonly been used to adjust for size when analyzing structural magnetic resonance imaging (MRI) data (see Sullivan et al.): the proportion, general linear models–analysis of covariance (GLM/ANCOVA), and residual approaches. The first two have been more widely used than the third. There are several other, more complicated techniques, including varying applications of the GLM that can accommodate a variety of different situations. In this review, however, we focus on the three commonly used methods outlined in Table 1.

**Proportion Approach**

The proportion approach uses as its numerator the volume of an ROI for an individual and as its denominator a volumetric measure of brain size of that individual (e.g., volume of total brain or of some large structure of which the ROIs are components). Here, volume is expressed not as a quantity (e.g., cubic centimeters), but as a ratio, fraction, percentage, or proportion. An ANOVA (or for two-samples, a t-test) can then be used to determine whether proportionalized volumes differ by diagnostic group. To control for other covariates, a linear regression model can be used, with the proportionalized volume as the outcome (dependent) variable, and the group indicator and any other covariates (e.g., age) as the predictor (independent) variables.

Brain volume measures have typically been used as normalizing factors in the proportion approach (see Seidman et al.). The fact that the proportion method uses a ratio constrains the adjustment factor that is used since nonvolumetric measures of head size differ in their units of measurement. A ratio obtained using a numerator and denominator with different units would not be unit-less, making interpretation of the results less straightforward.

**General Linear Models–Analysis of Covariance Approach**

The GLM/ANCOVA approach adopts the raw volume of an ROI as the outcome variable and analyzes it via a linear
### TABLE 1. Comparison of Statistical Methods I

<table>
<thead>
<tr>
<th>Steps to generate outcome</th>
<th>Proportion</th>
<th>GLM/ANCOVA</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome measure</td>
<td>Proportionalized volumes</td>
<td>ROI volume</td>
<td>Residualized volumes</td>
</tr>
<tr>
<td>Analytical model</td>
<td>Linear regression with the proportionalized volume as the outcome and predictors as covariates; does not include the measure of head size</td>
<td>Linear regression model, including head-size measure and all other relevant predictors as covariates</td>
<td>ANOVA or two-sample t-test comparing residuals for the controls to the diagnostic group</td>
</tr>
<tr>
<td>Accommodates nonvolumetric head-size measures?</td>
<td>No</td>
<td>Yes</td>
<td>Yes, but volumetric measures are typically used</td>
</tr>
</tbody>
</table>

regression model using relevant covariates (e.g., age, gender, diagnostic status) as predictors; the ROI volume is the outcome variable. The measure chosen to adjust for head size is included as a covariate along with the group indicator and any other covariates of interest. Therefore, the model includes a continuous outcome measure and both categorical and continuous predictors. Alternative GLM models can accommodate a variety of designs. For example, they could accommodate models with no categorical predictors (i.e., no group indicators), and they would be equally capable of handling models that have only categorical predictors (e.g., a model with just a group indicator as a predictor).

**Residual Approach**

The residual approach was first applied in the imaging literature by Arndt and colleagues and Mathalon and colleagues. This method, which is more computationally intensive than the others, controls for the effect of head size by using the data from the control group only if the data were derived from a study distinguishing between features of those with a diagnosis and features of controls. It aims to find the relationship between the ROI volume and each of the covariates (for a given set of all other predictors) for subjects in the control group. In this regression-based method, the raw ROI volume is regressed on the size measure (as well as on any other covariates of interest) using only data from the control group. This computation yields values for the regression coefficients that relate each of the covariates to the ROI volume for the control subjects. Thus, using information for any given set of covariates, a prediction can be made for the ROI volume expected for a control subject. This information is then used to obtain predicted ROI volumes for the entire set of subjects. These predicted volumes are subtracted, in turn, from the observed volumes for each subject in the data set. The resulting values are the residuals for each subject. A residual represents the deviation of an individual subject’s volume from the ROI volume, as predicted for a control subject, using the subject's specific values for each covariate (e.g., age, gender, head size). The group effect is often assessed by performing an ANOVA or t-test of the standardized residuals (with the standardization achieved by dividing the residuals by an estimate of their standard deviation). We note that such tests are valid only if distributional assumptions hold (typically, that the data are distributed normally). If these assumptions do not hold, it may be possible to transform the data to achieve the desired distribution, which may facilitate a more satisfactory outcome in addressing allometric aspects of the data interpretation.

**SIMILARITIES AND DIFFERENCES AMONG METHODS**

These three contrasting approaches are summarized in Table 2.
TABLE 2. Comparison of Statistical Methods II

<table>
<thead>
<tr>
<th>Analytical model</th>
<th>Proportion</th>
<th>GLM/ANCOVA</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of size measure</td>
<td>Regression based</td>
<td>Regression based</td>
<td>$r$-test (although outcomes are generated via a regression approach)</td>
</tr>
<tr>
<td>Level at which adjustment is made</td>
<td>Ratio (outcome)</td>
<td>Covariate</td>
<td>Covariate</td>
</tr>
<tr>
<td>Source of data for adjustment</td>
<td>Individual</td>
<td>Group or individual</td>
<td>Group</td>
</tr>
<tr>
<td>Impact on measurement error variance</td>
<td>Individual</td>
<td>All subjects or individual</td>
<td>Controls only</td>
</tr>
<tr>
<td>possibly increased, due to presence of measurement error in both numerator and denominator; proportionalized outcomes are correlated with the volumetric measure used to make adjustments</td>
<td>Measurement error present in outcomes remains unchanged; however, outcomes are correlated with the volumetric measure used to make adjustments</td>
<td>Measurement error present in outcomes remains unchanged; residualized outcomes are uncorrelated with the volumetric measure used to make adjustments</td>
<td></td>
</tr>
<tr>
<td>Types of group effects</td>
<td>Unconditional on the volumetric measure of head size used</td>
<td>Conditional on the value of head size if adjustment is done at group level</td>
<td>Conditional on the value of head size</td>
</tr>
</tbody>
</table>

Units Used

Whereas the residual and GLM/ANCOVA methods use a head-, brain-, or body-size measure as a covariate, the proportion method uses it to form a ratio. Unlike the proportion approach, in which the numerator and denominator should be measured in similar units, the GLM/ANCOVA and residual approaches more easily accommodate different types of measures useful for size adjustment.

Group Versus Individual Adjustment

When the GLM/ANCOVA method makes its adjustment for head, brain, or body size on a group level, it uses information from all subjects in the control and diagnostic groups. It is important to note, however, that adjustments within the GLM/ANCOVA method can also be computed on the individual level (see the description of the proportion method below).

The residual method adjustment is made at the group level, basing that adjustment, in most cases, upon the relationship between the ROI volume and all other covariates only for the control group; however, for comparisons that are not based on distinctions between control and diagnostic status, the entire data set may be used to generate residuals.54 Also note that the variability among the controls may be smaller than the variability among the diagnostic group, which may improve the statistical power of the residual method to detect group differences.

The proportion method makes its adjustment at the individual level by using each subject's ROI volume and dividing it by a measure of that subject's total brain volume. If one wishes to control for several covariates, however, then a GLM logistic regression model is indicated. Thus, the GLM may be used to make adjustments at either the individual or group level (as is the case for GLM/ANCOVA).

Since the proportion method generally adjusts for size at the level of the individual subject, it is often used when the level of inference is on the individual, rather than group, level—which may be the case in small samples or when volumetric data are extensive enough to be used for individual predictions (such as diagnosis). That is, since each individual's ROI volume has been divided by that individual's total brain volume, the head-size adjustment is individual specific.

Whether the method adjusts at the group or individual level affects the level of the potential inferences. Information obtained from group-level analyses should not be used to make individual-level predictions. Doing so can lead to erroneous conclusions since the relationships between the outcome and the predictors are often inflated at the group level.

Variance Considerations

Mathalon and colleagues58 argued that the residual method is superior to the proportion method. They reasoned that
the residual method allows one to construct residualized outcomes that are uncorrelated with the volumetric measure used to make adjustments, suggesting that the residual method may result in a reduction of the amount of “true” variance in the residualized outcome measure. The observed variance in a measure consists of two components: the “true” variance and the measurement-error variance. Both Arndt and colleagues\(^57\) and Mathalon and colleagues\(^58\) weighed the relative advantages of the types of variance modeled in each approach and the implications, in turn, for the reliability of each method. How a possible decrease in “true variance” may differentially affect analyses using the residual method versus the proportion method is discussed more fully below.

**Conditional Versus Unconditional Group Effects**

Each adjustment method produces a different type of group effect. As the regression coefficient associated with the indicator of diagnostic status, the “group effect” is generally the coefficient of primary interest in MRI analyses. Strictly speaking, it represents the average increase (or decrease) in predicted ROI volume for a subject in the diagnostic group versus a subject in the control group (assuming all other predictors are the same between the two subjects). Whereas an unconditional group effect indicates the average volume difference in the diagnostic group as compared to the control group, a group effect that is conditional on total volume indicates the average volume difference between the diagnostic and control groups specifically for a given value of the total volume. That is, a conditional group effect can be interpreted only when the value of the adjustment factor is held constant. An unconditional group effect can be interpreted without the adjustment factor being held constant.

This distinction relates to a fundamental difference in interpretation between the proportion and ANCOVA/residual approaches (if ANCOVA adjustment has been made on the group level). In the proportion approach, the coefficient for the group effect has an interpretation that is not conditioned on the volumetric measure used for head, brain, or body size, although it is conditioned on the other covariates in the model. By contrast, in the GLM/ANCOVA and the residual approaches, the group-effect coefficient has an interpretation that is conditioned on the value of head size (as well as any other covariates in the model). This distinction is important when presenting results since erroneous conclusions can be reached if the unconditional or conditional group effects are confused.\(^60\) For instance, one inconsistency could occur if a significant diagnostic effect was found not when the using an unconditional group effect, but only when using a conditional group effect. Such results would indicate that there is no difference in the average ROI volume when comparing all subjects in the diagnostic group to all subjects in the control group. There would be a significant diagnostic effect, however, when comparing subjects with a given value of head size in the diagnostic group to subjects with the same given value of head size in the control group. Formally, neither conclusion is incorrect, but the interpretation of the group effect differs in the two cases.

**Reliability and Validity**

Arndt and colleagues\(^57\) suggested that the adjusted volumes obtained from using the proportion method were inherently less reliable than the raw volumes because of the measurement error present in both the ROI volume and size measure (the numerator and denominator, respectively). Mathalon and colleagues\(^58\) confirmed this result and demonstrated that the reliability worsened as the correlation between the numerator and denominator used in the proportion method increased. Just how to interpret this reduced reliability is not straightforward, however. Mathalon and colleagues\(^58\) suggested that this reduction in reliability could be due to a decrease in the “true” variance of the proportionalized outcome (in addition to its already increased measurement error variance), which may result in improved criterion validity. This improved validity would presumably enhance the detection of group effects associated with volumetric differences not attributable to differences in head size. This bias has been previously reported in the structural MRI literature.\(^49\)

The issue of reliability in delineating ROI volumes is of crucial importance. Lange and colleagues\(^44\) used coupled regression equations (or regression calibration) to account for the inherent uncertainty in volumetric estimates. This technique replaces the value of TBV used as a predictor in the GLM/ANCOVA method with the expected value of total brain size for a given observed value of brain volume and the given values of any other predictors (e.g., age) that one wishes to use. This expected value is determined via a second (coupled) equation that relates the observed volume to the “true” volume plus random error. This procedure diminishes bias in parameter estimates caused by measurement error within volumes. The question of whether methods such as spatial stereotaxic and deformable templates,\(^61,62\) which are more recently developed tools for generating volume data, would have an impact on measurement error is not a primary concern when discussing the methodologies reviewed in this article.

**DISCUSSION**

**Are Normalizing Statistical Methods Equivalent, or Is One Method or Measure Superior?**

Using the residual method, Mathalon and colleagues\(^59\) noted improvement in the detection of group effects when evaluating gray matter ROIs, but not when considering white
matter or cerebrospinal fluid. We may thus expect to see enhancement in statistically significant group effects in gray matter when using the residual method, as opposed to using the proportion method, if such effects are present in the data under consideration; however, other studies have seen the proportion method detect such differences more easily than the residual method.\textsuperscript{55}

To what extent do these methods agree? There is at least a moderate amount of agreement between the proportion and GLM/ANCOVA approaches\textsuperscript{55} or among all three adjustment procedures\textsuperscript{54,63} in adult samples. No comparable work exists, however, in pediatric samples. There has also been little investigation into the use of different adjustment factors.

Depending on one's particular situation, other applications of the GLM can be used. For example, the GLM can be used to assess differences among multiple ROIs simultaneously. The proportion method detect such differences more easily than the proportional or residual methods. However, a strict enhancement in statistically significant group effects in gray matter when using the residual method, as opposed to using the proportion method, if such effects are present in the data under consideration; however, other studies have seen the proportion method detect such differences more easily than the residual method.\textsuperscript{55}

To what extent do these methods agree? There is at least a moderate amount of agreement between the proportion and GLM/ANCOVA approaches\textsuperscript{55} or among all three adjustment procedures\textsuperscript{54,63} in adult samples. No comparable work exists, however, in pediatric samples. There has also been little investigation into the use of different adjustment factors.

Depending on one's particular situation, other applications of the GLM can be used. For example, the GLM can be used to assess differences among multiple ROIs simultaneously.\textsuperscript{56} More information on the statistical merits of the GLM can be found in McCullagh and Nelder.\textsuperscript{56}

**When Adjustment May Not Be Indicated**

We have considered several adjustment factors and statistical methods in common use. However, one can envision alternative scenarios in which one might choose not to make an adjustment at all, or might choose a scaling measure other than those cited above. Those decisions should be guided by informed judgment rather than deference to a strict set of rules.

If an exploratory analysis were performed with no prior knowledge of how total brain volumes might be affected, it would be most illuminating to perform analyses using both regional measures and measures of total brain volume. If pervasive volumetric changes were found, it may be fruitful to compare the results of unadjusted with adjusted volumes; this comparison would permit evaluation of whether widespread changes are uniform or nonuniform in their effects, through a determination of whether the volumetric changes in individual regions are proportional or disproportional to the overall volume alteration.

**SUMMARY**

The investigator studying volumetric brain measures is faced with an abundance of available methods and a welter of methodological detail. We have described here three commonly used techniques to scale and analyze group differences in these measurements, pointing out relative advantages of these methods, as well as some potential pitfalls. It is important to note that the flexibility of the multiple linear regression models makes them more widely applicable than the proportional or residual methods. However, a strict formulary for size correction is not recommended. Thus, the choice of method should be guided more by how the hypotheses are framed, the characteristics of the disorder, the sample size, and the measures that are available than by an assessment as to which method is the “correct” one.

**REFERENCES**


52. Piven J, Bailey J, Ranson BJ, Arndt S. No difference in hippocampal volume detected on magnetic resonance imaging in autistic boys. Brain 2003;126(pt 5):1182–92.


63. Goldstein JM, Goodman JM, Seidman LJ, et al. Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. Arch Gen Psychiatry 1999;56:537–47.