



ELSEVIER

Psychiatry Research: Neuroimaging Section 74 (1997) 1–12

**PSYCHIATRY
RESEARCH
NEUROIMAGING**

Variability of human brain structure size: ages 4–20 years

Nicholas Lange^{a,*}, Jay N. Giedd^b, F. Xavier Castellanos^b,
A. Catherine Vaituzis^b, Judith L. Rapoport^b

^a*Consolidated Department of Psychiatry, Harvard Medical School and Brain Imaging Center, McLean Hospital,
115 Mill St., Belmont, MA 02178, USA*

^b*Child Psychiatry Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA*

Received 21 May 1996; revised 26 November 1996; accepted 20 December 1996

Abstract

Understanding variability of human brain structure sizes during development is important for the design and interpretation of pediatric neuroimaging studies. In this study we analyze the effects of hemisphere, sex and age on size variability of the total cerebrum, cerebellum, lateral ventricles, temporal lobe, amygdala, hippocampus, superior temporal gyrus, corpus callosum, caudate, putamen, and globus pallidus in 115 healthy children and adolescents, ages 4–20 years. Variability differed significantly across structures, with the lateral ventricles demonstrating the highest coefficient of variation and the putamen the lowest. Males varied significantly more than females in the left cerebrum and left superior temporal gyrus, whereas females varied more than males in the right caudate and right putamen. Age effects were seen in increased variability after puberty for the lateral ventricles, hippocampus and superior temporal gyrus. These variances are important determinants of minimum sample sizes required to detect group differences in both cross-sectional and longitudinal studies. © 1997 Elsevier Science Ireland Ltd.

Keywords: Variance; Adolescent; Child; Coefficient of variation; Magnetic resonance imaging; Measurement error models; Morphology; Regression calibration

1. Introduction

Magnetic resonance imaging (MRI), with its lack of ionizing radiation and excellent anatomi-

cal resolution, has provided unprecedented opportunity to examine in vivo brain morphology and has led to a growing number of quantitative neuroanatomic studies of pediatric neuropsychiatric disorders including autism, attention-deficit/hyperactivity disorder (ADHD), childhood-onset schizophrenia, dyslexia, Sydenham's chorea, Tourette's syndrome, Fragile X syndrome, and

* Corresponding author. Tel.: +1 617 8552139; fax: +1 617 8552770; e-mail: lange@emu.mclean.org

William's syndrome (see Giedd, 1996, for review). Understanding the normal variance of brain structure sizes examined in these reports is important for interpreting discordant findings in small sample studies, since group specific variability differences affect statistical tests of mean differences. However, to our knowledge, brain structure size variability has not been examined systematically from a statistical standpoint.

Variability measurements from the adult literature are available in studies whose main aims have been to describe mean differences and general age trends within demographic and clinically defined groups (Pakkenberg and Voight, 1964; Dekaban and Sadowsky, 1978; Holloway, 1980; Ho et al., 1980, 1980; Jerison, 1991; Pfefferbaum et al., 1994; Flaum et al., 1995; Caviness et al., 1996). These studies indicate that male brain volumes are on average larger than female brain volumes and that there is a general male-greater-than-female variability difference in total cerebral volumes.

In this article we provide normative data on 115 healthy children and adolescents and test for significant variability differences by brain structure, sex and age. Quantitative neuroanatomic data from some of these subjects have been reported previously (Giedd et al., 1996a–c). Our goals in the present article were to: (1) analyze differences in variability and covariability with respect to sex and age for eleven brain structure measurements, including total cerebrum, cerebellum, lateral ventricles, temporal lobe, superior temporal gyrus, amygdala, hippocampus, globus pallidus, caudate nucleus, and putamen volumes and midsagittal corpus callosum and cerebellar areas; (2) assess the conformity of these data to classical statistical assumptions and apply alternative methods that accommodate assumption violations when necessary; and (3) provide a graphical method to calculate minimum required sample sizes given variance characteristics and anticipated effect sizes. Due to patterns reported in the adult literature, our hypotheses were that variability would be greater in males vs. females and older vs. younger subjects.

2. Methods

2.1. Subjects

Because of potential over-representation of children with neuropsychiatric disorders in clinically acquired scans, subjects were recruited directly from the community as part of an ongoing study by the Child Psychiatry Branch of the National Institute of Mental Health to assess brain development in healthy and neuropsychiatrically impaired children and adolescents. Of the 748 respondents to local advertisements, 623 were excluded through a three-part screening process (phone interview, questionnaires mailed to parents and teachers, and in-person physical and psychiatric examinations). Exclusion criteria were physical or psychiatric illness in the subjects or their first degree relatives (see Giedd et al., 1996 for further details). Of the final 125 subjects, three were unable to complete scans because of anxiety or claustrophobia and seven were unsatisfactory because of excessive subject motion.

The study was approved by the Institutional Review Board of the National Institute of Mental Health. Assent from the child and written consent from the parents were obtained.

Table 1 presents subject characteristics. There were no significant group differences between males and females on age, height, weight, handedness or IQ subtest scores. As expected from our inclusion criteria, Wechsler Adult Intelligence Scale-Revised (WISC-R) subtest scores were above average, potentially affecting generalizability of findings.

2.2. MRI protocol

All subjects were scanned on the same General Electric 1.5 Tesla Signa scanner using a 3-D SPGR imaging sequence (echo time = 5 ms, time to repeat = 24 ms, flip angle = 45°, acquisition matrix = 192 × 256, number of excitations = 1, field of view = 24 cm). T1-weighted images with slice thickness of 1.5 mm in the axial and sagittal planes and 2.0 mm in the coronal plane were obtained. Head positioning was standardized by

Table 1
Demographic characteristics of healthy children and adolescents ($n = 115$)

	Female	Male
Sample size	50	65
Age (years)	11.2 (4.0)	11.5 (3.6)
Height (in)	57.3 (82.)	58.9 (8.5)
Weight (lb)	89.8 (34.7)	96.4 (37.6)
Tanner stage	2.4 (1.6)	2.4 (1.6)
Vocabulary	12.7 (2.5)	13.8 (3.1)
Block design	12.7 (3.5)	13.4 (3.7)
Digit span	11.4 (2.4)	11.2 (3.1)
% right-handed	90	90

Means are followed by standard deviations in parentheses.

verifying that three vitamin E capsules, one in each auditory meatus and one taped to the lateral aspect of the left inferior orbital ridge, were all visible on the same axial slice and, on that same slice, each subject's nose was aligned at the '12:00' position. No pharmacological sedation was used.

2.3. Image analysis

A clinical neuroradiologist reported increased T2 signal intensities in the right parietal lobe in one subject and left semiovale in another subject. These findings were clinically insignificant and the images were retained in the analysis. No other gross abnormalities were reported for any subject. Details of the quantification of various structure sizes are described elsewhere (Giedd et al., 1996a–c). Briefly, total cerebral volume was quantified by a deformable model in which an elastic template was conformed to each individual brain through an iterative energy minimization algorithm (Snell et al., 1995). An advantage of this approach was that it allowed a priori knowledge of brain anatomy to supplement the sometimes ambiguous boundaries of MR images. Program output was edited manually to remove segmentation artifacts when required. Ventricular volume was quantified using a thresholding function available in NIH Image (Rasband, 1993). Volumes of the caudate, putamen, globus pal-

lidus, amygdala, hippocampus, and temporal lobes and midsagittal area of the cerebellum were quantified using a manual tracing feature. Midsagittal area of the corpus callosum and seven subdivisions were quantified using a semi-automated program written in C and available upon request. Inter-rater correlation coefficient measurement reliabilities for the quantified structures were as follows: total cerebrum, 0.99; cerebellum, 0.88; lateral ventricles, 0.99; temporal lobe, 0.98; amygdala, 0.86; hippocampus, 0.87; caudate, 0.88; putamen, 0.84; globus pallidus, 0.82; corpus callosum, 0.92.

2.4. Statistical analyses

We employed mean structure size and coefficient of variation (CV) by sex and hemisphere as parametric data summaries. The CV is defined as the standard deviation expressed as a percentage of the mean and is a unitless quantity preferred to sample variances when comparing different brain structures because variance may be determined in part by structure size. While the CV is by itself inadequate for quantifying and comparing brain structure size and variability differences, we nonetheless find it a useful, albeit insufficient, summary statistic. Statistical tests of significant differences between CVs employed asymptotic standard errors equal to the CV divided by the square root of twice the sample size (see, for instance, Sokal and Rohlf, 1981).

We analyzed differences in residual variabilities in brain structure sizes by sex, age and hemisphere both with and without adjustment for effects of age, height, weight, and total cerebral volume. The relationship between brain size and body size in humans is surprisingly poor (Harvey and Krebs, 1990). In contrast to the relative stability of brain weight after childhood, body weight varies widely among individuals and can vary substantially for the same individual over time. Height is also a poor indicator of brain size, as is implied by contrasting the notable increase in height from ages 4 to 20 years with the lack of corresponding increase in brain size. This general

trend for the young to have disproportionately large head-to-height ratios compared to adults is widely observed throughout the mammalian species. These considerations led us to try several different adjustment options, including no adjustment whatsoever, since current neuroimaging literature is neither clear on what variables are important for adjustment nor on how adjustments should be made once the variables have been chosen.

Table 2 lists the variety of models used for expressing mean function relationships and from which the residuals were obtained. Our adjustment models encompass a very wide range of options to test sex- and/or age-dependent variability differences after removing mean size differences and effects of other potential covariates. Each model is fit separately for the males and for the females; we do not use pooled estimates of a common variance as in a two-sample *t*-test. Separate estimates of variance are computed for each group from the residuals of the fitted models. All models that include total cerebral volume accommodate measurement errors in this variable by regression calibration; see Eq. (1) and accompanying descriptions. Model U (Unadjusted) simply compares the variances of male and females residuals after subtracting the mean of each group. Model SL (Simple Linear) makes linear adjustments to the outcomes for the effects of total cerebral volume and age (X, z_1). Model ML (Multi-Linear) makes linear adjustments to the outcomes for the effects of total cerebral volume

and age, height and weight (X, Z). Because assumptions of parametric methods, such as SL, ML, analysis of variance (ANOVA), analysis of covariance (ANCOVA) and multivariate analysis of variance (MANOVA) were not always supported by our data, we also employed flexible regression methods (Hastie and Tibshirani, 1990). Models SF (Simple Flexible) and MF (Multi-Flexible) make no straight-line assumptions about the relationships between these variables but instead employ locally adaptive functions that capture relationships between variables by non-parametric regression. As in models SL and ML, total cerebral volume and (age, height, weight) are treated by separate functions. Transforming outcomes Y into ratios Y/X has been shown to be flawed (Arndt et al., 1991) and hence ratio adjustment models are not considered here.

We take two approaches to testing for significant differences between sex and age dependent variances. The first is a classical variance-ratio *F*-test, known to be quite sensitive to non-Gaussian data. The effects of non-Gaussian data on the *F*-test, or on the related tests of Bartlett, Hartley and Cochran, are catastrophic (Pearson, 1931; Miller, 1986), yielding very unreliable results. Since we are not always willing to make the parametric assumption that the outcomes derive from the infinitely smooth bell-shaped probability curve required for a reliable *F*-test, we also use a computer-intense Monte Carlo simulation method. This method involves repeated sampling of the observed data to yield an assumption-free, empirical significance test. Non-Gaussian errors, unequal variances and non-linear functional relationships, all present to some degree in our sample, demand tests whose results are free of classical assumptions. Resampling the observed data without replacement, so that each subject appears only once, yields a non-parametric randomization or permutation test; see for instance Manly, 1991. Resampling the observed data with replacement, so that each subject can appear more than once, yields a non-parametric bootstrap test; see Efron, 1982. In testing for variance differences, improvements over the classical *F*-test in non-standard situations such as ours have been demonstrated through sampling without replacement (Bailer,

Table 2
Different forms of adjustment to brain structure volumes (Y) for covariates measured with error (X) and without error (Z)

Model ^a	Label	Equation
Unadjusted	U	$Y = \mu + \epsilon$
Linear	SL	$Y = X\beta + z_1\gamma + \epsilon$
	ML	$Y = X\beta + Z\gamma + \epsilon$
Flexible	SF	$Y = f(X) + g_1(z_1) + \epsilon$
	MF	$Y = f(X) + \sum_k g_k(z_k) + \epsilon$

^aEach model is fit separately for males and for females. X = total cerebral volume; $Z = (z_1, z_2, z_3)$ = (age, height, weight).

1989) and sampling with replacement (Boos and Brownie, 1989). In the present analysis, a permutation test is preferred to a bootstrap test since we want to retain the observed collections of covariates. For each simulation, we permute subjects at random across the fixed male/female group boundary, changing their sex label while keeping the overall group sizes fixed, and calculate a test statistic for each such permutation. Since the total number of possible permutations is quite large, we draw a large sample of size 999 from the permutation distribution of the data using a random number generator. We calculate and rank the simulated test statistics together with the original observed value to obtain an empirical P -value for each. For instance, if the original test statistic ranks as the 39th largest amongst the 1000 (999 + 1) realizations, its empirical P -value is 0.039. The entire procedure requires a rather lengthy calculation in which 105 models are fit to each permuted data set (8 volumes \times 3 (left hemisphere, right hemisphere, total) \times 11 areas) \times 5 types of adjustments, 1000 permutations, for a total of over one-half million simulations. In general, we found that parametric F -tests yielded smaller P -values than non-parametric permutation tests. However, in order to make substantive conclusions that depend minimally on modeling assumptions, variance differences were deemed significant only when the null hypothesis was rejected by parametric tests for all models at level 0.05 and was also rejected by our non-parametric tests for one or more of the models at this same level. In such cases, we report the largest P -value of tests and models. We took this conservative approach not only to account for multiplicity of statistical tests but also to address potential criticisms of parametric assumptions and of adjustment method used (including none) for significance testing.

Adjustments for total cerebral volume, which differed significantly between males and females, employed measurement error models and regression calibration methods (Fuller, 1987; Carroll et al., 1995). For instance, the Multi-Linear model of Table 2 is augmented to account for measurement error in total cerebral volume by the fol-

lowing coupled equations:

$$\begin{cases} Y = X\beta + Z\gamma + \epsilon \\ W = X + U \end{cases} \quad (1)$$

In Eq. (1), all symbols occurring previously in Table 2 are interpreted as before: Y is a brain structure size measurement, X is total cerebral volume, Z is a vector of covariates (age, height and weight), and ϵ is a random error in the linear equation relating Y to X and Z . In addition, for new symbols, W is the observed value of total cerebral volume with true unobserved value X , and U is a random error made in observing X . Unlike covariates such as sex, age, height and weight that can usually be assumed error-free, covariates such as total cerebral volume and other brain segmentation outcomes that require subjective judgments are measured with some degree of uncertainty. Regression calibration replaces the observed X in the first equation with its conditional expected value given W and Z . Without regression calibration, a naive simple least-squares approach would yield biased estimates and inflated residual variances. Degree of bias is determined by a reliability ratio which is equivalent to Fisher's intraclass correlation coefficient (Fisher, 1946; Winer, 1971; Lange and Ryan, 1989). The reliability ratio is beginning to appear in the neuroimaging literature (Arndt et al., 1991; Bartzokis et al., 1993; Mathalon et al., 1993), as are several ad hoc attempts to correct for measurement error effects, and goes under several names in the econometric, psychometric and genetics literature. In genetics, for instance, the reliability ratio is heritability, W is phenotype, X is genotype and measurement error U is the environmental effect on phenotype (Fuller, 1987, p. 3). For more statistical detail on the treatment of measurement error in brain imaging studies, see Lange et al. (1996). Regression calibration is used here to enable reliable variability comparisons of residuals from a variety of regression analyses.

3. Results

Table 3 gives sample means and CVs for the

Table 3

Summary statistics for unadjusted brain structure sizes by sex and hemisphere for 115 healthy subjects, ages 4-20 years

	Total (n = 115)			Female (n = 50)			Male (n = 65)		
	T	RH	LH	T	RH	LH	T	RH	LH
(a) Means (mm)									
Total cerebrum ^{a,d}	1127.54	568.88	558.66	1072.82	541.40	531.42	1169.63	590.02	579.60
Superior temporal gyrus	51.16	26.46	24.70	49.87	25.57	24.30	52.25	27.20	25.05
Putamen	10.66	5.22	5.44	10.25	5.01	5.24	11.00	5.39	5.61
Caudate ^{b,c}	10.24	5.20	5.04	10.35	5.27	5.09	10.15	5.14	5.00
Lateral ventricles ^{a,d}	10.08	4.81	5.27	9.28	4.52	4.75	10.68	5.02	5.66
Hippocampus	9.20	4.73	4.47	9.05	4.69	4.35	9.32	4.75	4.57
Amygdala	4.66	2.41	2.25	4.39	2.23	2.16	4.89	2.56	2.33
Globus pallidus ^{a,c}	2.38	1.19	1.19	2.21	1.10	1.11	2.50	1.26	1.24
Cerebellum ^e	1154.34	—	—	1125.49	—	—	1174.70	—	—
Corpus callosum ^e	619.77	—	—	604.86	—	—	631.48	—	—
(b) Coefficients of variation (S.D. expressed as a percentage of the mean) × 100									
Total cerebrum ^{f,d}	11.3	11.3	11.5	9.6	9.8	9.6	11.0	10.9	11.3
Superior temporal gyrus ^{f,d}	11.5	12.8	12.9	8.7	11.5	9.9	12.9	13.1	14.8
Putamen ^{g,c}	9.4	9.7	9.6	10.4	10.8	10.3	7.6	7.6	8.0
Caudate ^{g,c}	12.7	12.9	12.7	14.5	14.9	14.2	10.9	10.9	11.3
Lateral ventricles	63.5	67.3	64.6	71.1	75.8	72.0	58.1	61.5	59.4
Hippocampus	10.3	11.7	11.2	10.7	11.7	11.8	10.0	11.8	10.2
Amygdala	19.3	22.2	22.7	17.4	22.1	20.9	19.4	20.6	23.6
Globus pallidus	15.4	16.0	16.8	15.0	15.6	16.0	14.6	14.5	16.9
Cerebellum	10.7	—	—	10.9	—	—	10.2	—	—
Corpus callosum	14.7	—	—	14.5	—	—	14.7	—	—

^a Male > Female.^b Female > Male.^c RH > LH.^d LH > RH.^e Midsagittal area (ml²).^f Male > Female, LH.^g Female > Male, RH.

T, total; RH, right hemisphere; LH, left hemisphere.

unadjusted brain structure volumes. Significant mean differences are indicated in Table 3a, as in Giedd (1996). Sample variances can be obtained by multiplying the CVs in Table 3b by their corresponding values in Table 3a, dividing by 100 and squaring the result. Note the considerable degrees of variability and the differences in variability across structures. For instance, although the variance of the total cerebrum was considerable, it constituted only about 11% of its mean. Contrast this with the lateral ventricles whose variability was roughly 64% of their mean volume. Variances of the lateral ventricles, amygdala and globus pallidus differed from each other and from the other structures. The putamen and caudate nucleus were comparably variable, as were the

superior temporal gyrus, cerebrum and hippocampus.

Omnibus MANOVA tests indicated global differences in male and female volumes for total cerebrum and lateral ventricles (Wilks' $\Lambda = 0.844$, $P = 0.0009$), for temporal lobe structures ($\Lambda = 0.906$, $P = 0.02$) and for the basal ganglia ($\Lambda = 0.755$, $P < 0.0001$). Significant global differences between the sexes persisted when differences in cerebral hemispheres were tested separately.

Fig. 1 gives boxplots for the four significant differences in variability by sex. Each boxplot has four components: centerline (median, or 50th percentile), box (25th through 75th percentiles), whiskers (10th and 90th percentiles) and extreme points (outliers) in some cases; see, for instance,

Andreasen et al., 1986, for their use in structural brain imaging analysis. Male-greater-than-female variance for total cerebrum was due primarily to significant variance differences in the left hemisphere (model ML; classical $P = 0.017$; permutation $P = 0.04$). Volumes of the superior temporal gyrus also varied more for males, again on the left (model MF; classical $P = 0.012$, permutation $P = 0.018$). We found two sex-dependent variabilities in the right hemisphere: female volumes varied more than male volumes of the caudate nucleus (model SF; classical $P = 0.001$, permutation $P = 0.04$) and the putamen (model U; classical $P = 0.010$, permutation $P = 0.039$).

Fig. 2 gives a parametric summary of results for

total cerebrum and the caudate, for which Gaussian assumptions apply. As indicated previously in Fig. 1, females had significantly greater variance than males for the caudate, yet increased variance was not associated with increased hemispheric asymmetry, defined as $2 \times (\text{right} - \text{left}) / (\text{right} + \text{left})$. Figs. 2c,d show that variance of asymmetry measurements did not differ significantly by sex in spite of sex differences in variance about the means; classical and permutation tests did not consistently reject the null hypothesis of asymmetry variance equality for either structure.

There were no significant sex dependent variance differences for any of the remaining brain

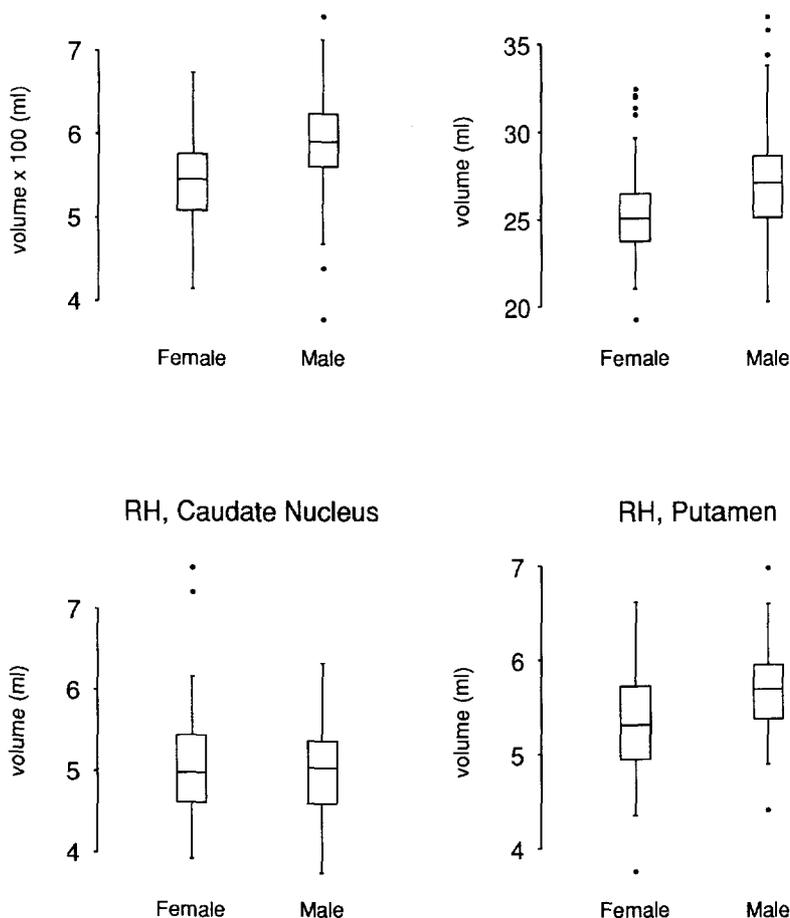


Fig. 1. Boxplots of unadjusted brain structure volumes exhibiting significant sex-dependent variability differences (50 females, 65 males): RH, right hemisphere; LH, left hemisphere; see text.

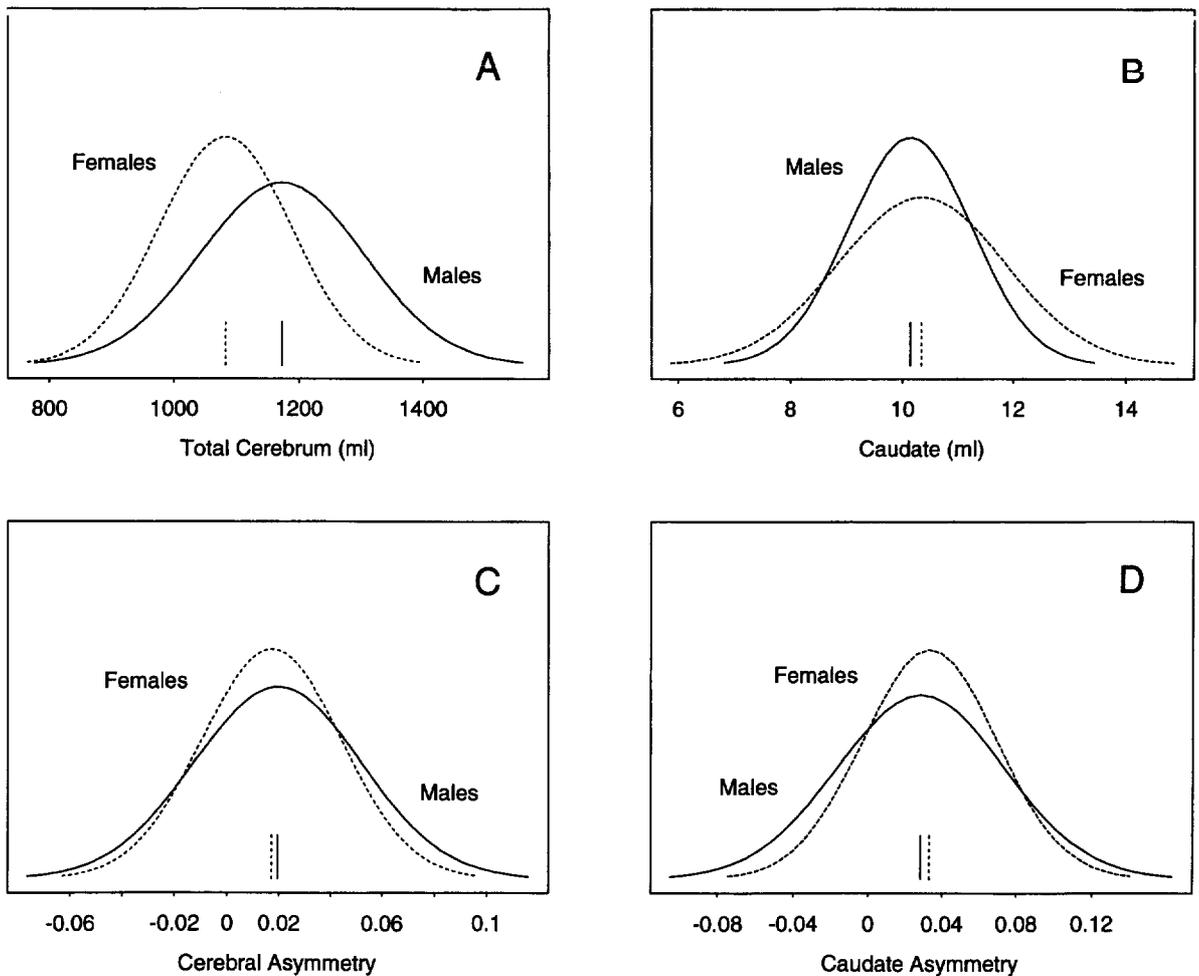


Fig. 2. Gaussian probability densities for healthy males and females, with means and variances equal to the sample values: A, total cerebrum; B, caudate; C, cerebral asymmetry; D, caudate asymmetry.

structure volumes (lateral ventricles, amygdala, hippocampus) and areas (corpus callosum and cerebellum).

Defining Tanner stage 1 as 'pre-puberty' and stages 2–5 as 'puberty', we found significant age or puberty-related increases in variances for the lateral ventricles in the left (model MF; classical $P = 0.0001$, permutation $P = 0.02$) and right hemispheres (model U; classical $P < 0.0001$, permutation $P = 0.013$), the left superior temporal gyrus left (model SF; classical $P = 0.015$, permutation $P = 0.044$) and the left hippocampus left (model MF; classical $P = 0.0002$, permutation P

$= 0.010$) (see Fig. 3). These age-related variability increases applied for both sexes. However, we found that males exhibited significantly greater variability than females in the left superior temporal gyrus ($P < 0.01$) for pubertal subjects only, indicating a sex by age interaction.

4. Discussion

In our sample of 115 healthy children and adolescents, variability of brain structure sizes differed by structure, sex, and age. Ventricular volume was the structure with the highest coef-

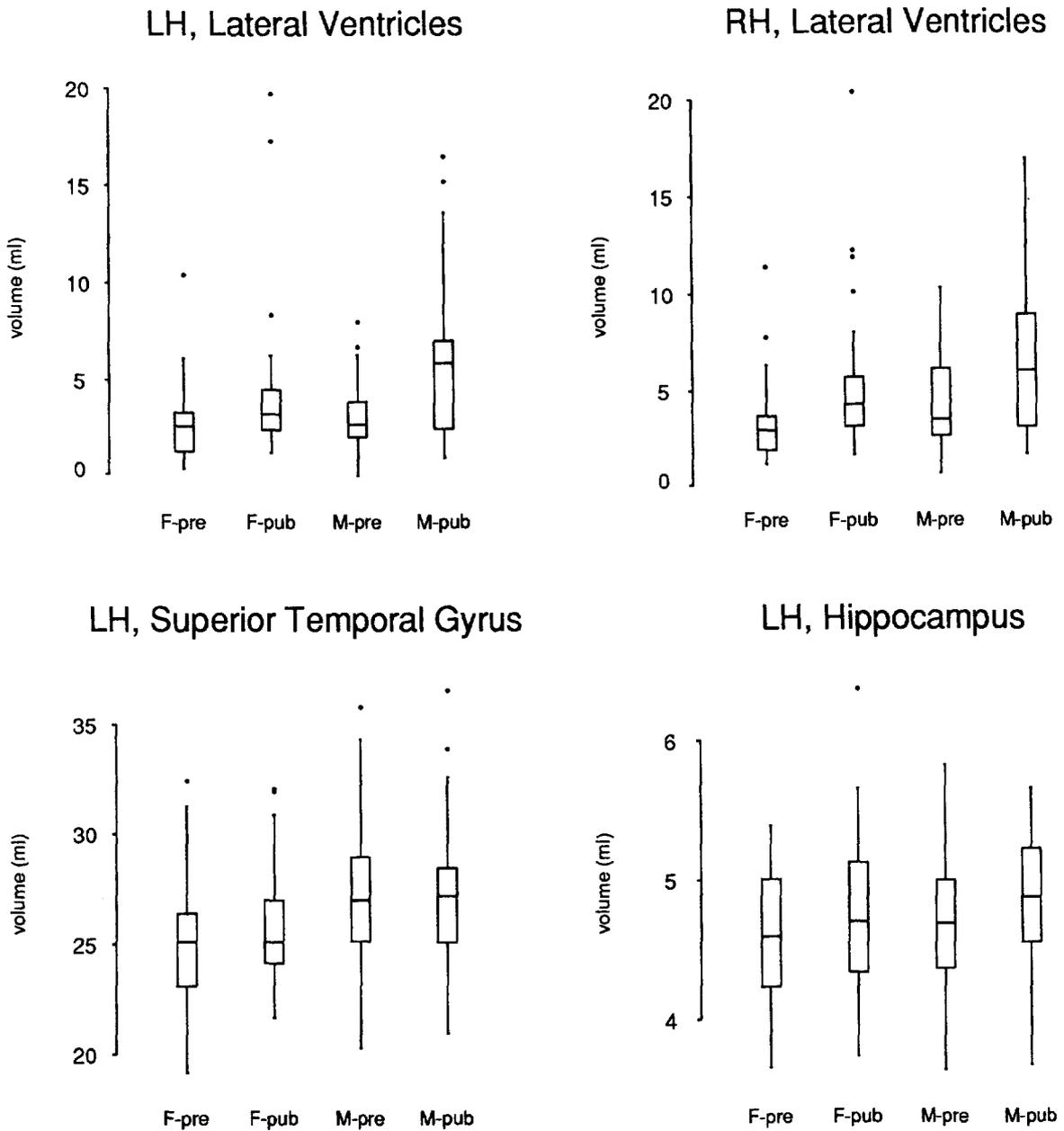


Fig. 3. Boxplots of unadjusted brain structure volumes exhibiting significant age-dependent variability differences (50 females, 65 males): RH, right hemisphere; LH, left hemisphere; F, Female; M, Male; pre, 'pre-puberty'; Tanner stage 1; pub, 'puberty'; Tanner stage 2–5; see text.

ficient of variation, due possibly to accumulated variability of the many surrounding structures that define its shape. Males varied significantly more

than females in the left hemisphere for the total cerebrum and the superior temporal gyrus, whereas females varied significantly more than

males in the right hemisphere for the caudate and the putamen. Age-dependent variability was also noted, being significantly greater for pubertal subjects in the lateral ventricles, the hippocampus and the left superior temporal gyrus. In addition, for pubertal subjects only, males varied significantly more than females in the left superior temporal gyrus.

Variability of brain structure size in normal development is important for the design and interpretation of pediatric neuroimaging studies. Minimum sample size required to detect a change in mean structure size will depend in part on CV. Simple formulas can be developed for sample size and power calculations for cross-sectional and longitudinal studies of brain structure sizes that involve CV, as follows. The routine sample size calculation for the two-sample problem for independent samples in cross-sectional studies is

$$n = (\sigma_1^2 + \sigma_2^2) / (\mu_1 - \mu_2)^2 [z(1 - \alpha/2) + z(1 - \beta)]^2 = f(\eta), \quad (2)$$

where $\eta = (\sigma_1^2, \sigma_2^2, \mu_1, \mu_2, \alpha, \beta)$ and $z(p)$ is the p th quantile from the standard Gaussian distribution. In Eq. (2), n is the minimum number of subjects needed per group in order to detect a difference in group means of $\mu_1 - \mu_2$ with power $1 - \beta$ and Type I error α when the independent groups have true means and variances (μ_1, μ_2) and (σ_1^2, σ_2^2) respectively. For comparing brain structure sizes of two groups, the formula may be inadequate. Tables 1 and 3 show that a more germane formulation of the problem involves CV. Assuming that each group has the same coefficient of variation $\nu = \sigma_1/\mu_1 = \sigma_2/\mu_2$ for a particular brain structure in question, a better sample size calculation (van Belle and Martin, 1993) is

$$n = \nu^2(\theta^2 + 1) / (\theta - 1)^2 [z(1 - \alpha/2) + z(1 - \beta)]^2, \quad (3)$$

where $\theta = \mu_1/\mu_2$ is the mean ratio. The formula applies when the minimum detectable difference is phrased in terms of a percentage change in the means, which is θ , along with percent variability, which is ν .

Fig. 4 uses Eq. (3) in a graphical description of the relationship between sample size, CV, and the ratio of means in a two-group cross-sectional study with 5% Type I error and 80% power. Group variances need not be equal; only the ratio of the standard deviation to the mean is assumed identical for the two groups. For instance, from Table 3b, the CV of the globus pallidus for females is 15%. This implies that about 95% of the observations are within 30% of their mean assuming Gaussian data ($0.15 \times 1.96 \approx 0.30$). Using Fig. 4, one finds that in order to detect a 10% change in mean volume $n = 38$ females per group are required.

Our results also help in the design of longitudinal studies. In such cases, one does not have independent samples because subjects are observed repeatedly over time and present values may depend on past values. For sample size calculations, two additional variables are required: the number of observations per subject, t , and the correlation between repeated observations for the same subject, ρ . The development here covers the simple intraclass correlation model described in the methods section. Assume that a simple linear relationship exists between brain structure size and a single fixed covariate, c_j , also measured repeatedly at the same times for all subjects, $j = 1, \dots, t$. For instance, if subjects are scanned at baseline and every year for 4 years, then $t = 5$ and $c_1 = 0, \dots, c_5 = 4$. To compare two groups of subjects with different linear relationships, a longitudinal counterpart to the cross-sectional formula is simply

$$n = f(\eta)(1 - \rho) / (t\sigma_c^2), \quad (4)$$

where σ_c^2 is the common within-subject variance of the c_j 's.

As an example, suppose that one is to plan a developmental study of decreases in caudate size for pre-pubescent males. Further suppose that one baseline and four repeated measurements of each subject are taken at yearly intervals, a serial correlation of 0.65 and a total estimated variance of 0.90 from the simple linear adjustment model corrected for measurement error in total cerebral

volume (0.90 is the actual value from our sample). In order to detect a minimum decrease in volume of 0.14 ml/year, the formula shows that at least $n = 25$ subjects are required.

In addition to guiding study design, brain morphometric variance characteristics may them-

selves be important discriminating features in comparisons between control and clinical groups. Studies are underway to examine whether variability is altered in pediatric neuropsychiatric disorders such as ADHD and childhood-onset schizophrenia.

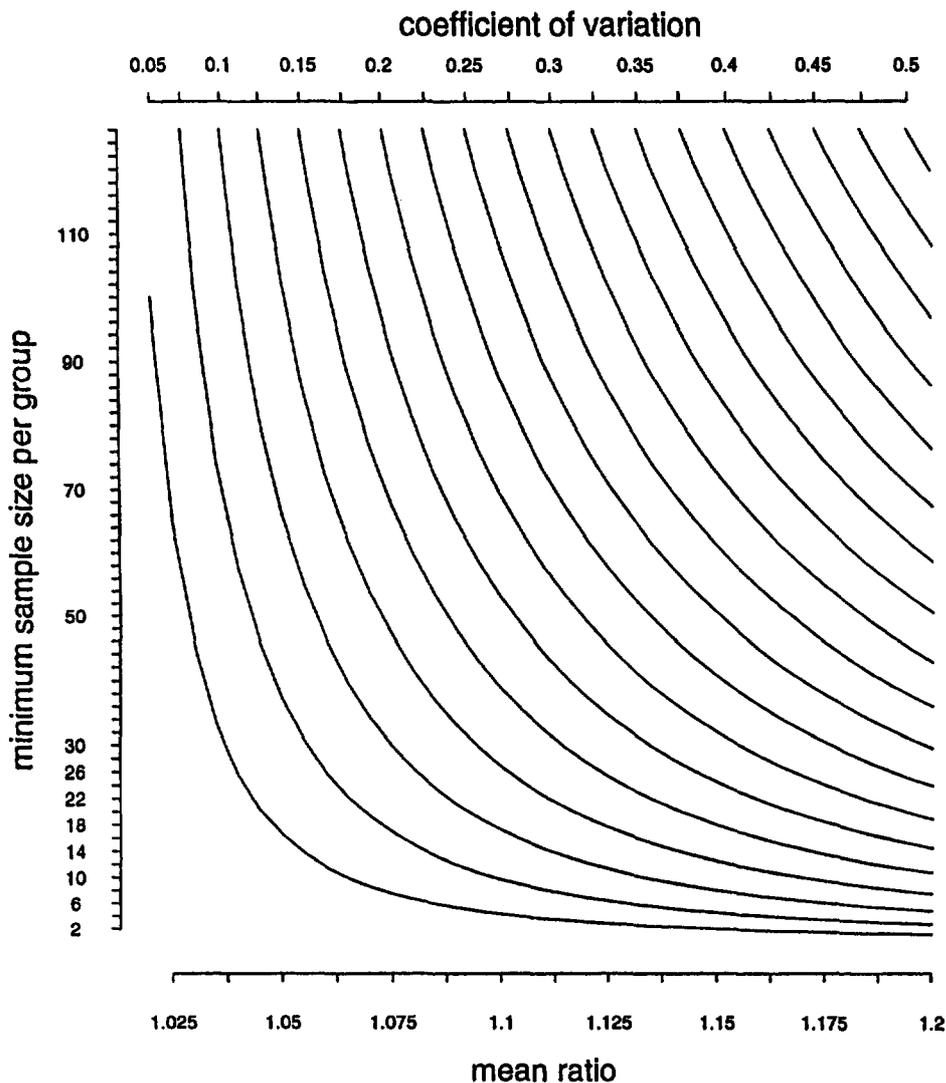


Fig. 4. Minimum sample sizes per group, N , with Type I error $\alpha = 0.05$ and power $1 - \beta = 0.80$. The following example assumes Gaussian data. Example: Find the sample size per group required to detect a 10% change in mean volume of the globus pallidus when 95% of the observations in each group are within 30% of their group means. These conditions imply mean ratio $M1/M2 = 1.1$ and coefficient of variation $SD1/M1 = SD2/M2 = 0.30/1.96 \approx 0.15$, the coefficient of variation for total volume of the globus pallidus for females in Table 2b. Place a straight edge parallel to the vertical axis at the point 1.1 on the horizontal axis. Mark where the straight edge crosses the curve for 0.15, the fifth curve from the left. Place the straight edge parallel to the horizontal axis at this point, to read $N = 38$ subjects per group.

References

- Andreasen, N.A., Nasrallah, H.A., Dunn, V., Olson, S.C., Grove, W.M., Erhardt, J.C., Coffman, J.A. and Casset, J.H.W. (1986) Structural abnormalities in the frontal system in schizophrenia. *Arch Gen Psychiatry* 43, 136–144.
- Arndt, S., Cohen, G., Alliger, R.J., Swayze, V.W. and Andreasen, N.C. (1991) Problems with ratio and proportion measures of imaged cerebral structures. *Psychiatry Res Neuroimag* 40, 79–90.
- Bailer, A.J. (1989) Testing variance equality with randomization tests. *J Stat Comput Simul* 31, 1–8.
- Bartzokis, G., Mintz, J., Marx, P., Osborn, D., Gutkind, D., Chiang, F., Phelan, C.K. and Marder, S.R. (1993) Reliability of in vivo volume measures of hippocampus and other brain structures using MRI. *Magn Reson Imag* 11, 993–1006.
- Boos, D.D. and Brownie, C. (1989) Bootstrap methods for testing homogeneity of variances. *Technometrics* 31, 69–82.
- Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995) *Measurement Error in Nonlinear Models*. Chapman and Hall, London.
- Caviness, V.S., Jr., Kennedy, D.N., Richelme, C., Rademacher, J. and Filipek, P.A. (1996) The human brain age 7–11 years: a volumetric analysis based on magnetic resonance images. *Cereb Cortex* 6, 726–736.
- Dekaban, A.S. and Sadowsky, D. (1978) Changes in brain weight during the span of human life: relation of brain weights to body heights and body weights. *Ann Neurol* 4, 345–356.
- Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- Fisher, R.A. (1946) *Statistical Methods for Research Workers*. Oliver and Boyd, London.
- Fuller, W.A. (1987) *Measurement Error Models*. John Wiley and Sons, New York.
- Flaum, M., Swayze, V.W. II, O'Leary, D.S., Yuh, W.T., Ehrhardt, J.C., Arndt, S.V. and Andreasen, N.C. (1995) Effects of diagnosis, laterality, and gender on brain morphology in schizophrenia. *Am J Psychiatry* 152, 704–714.
- Giedd, J.N. (1996) Developmental disorders. In: Krishnan, K.R.R. and Doraiswamy, M.P. (Eds.), *Brain Imaging in Clinical Psychiatry*, Marcel Dekker, New York (in press).
- Giedd, J.N., Rumsey, J.M., Castellanos, F.X., Rajapakse, J.C., Kaysen, D., Vaituzis, A.C., Vauss, Y.C., Hamburger, S.D. and Rapoport, J.L. (1996a) A quantitative MRI study of the corpus callosum in children and adolescents. *Brain Res Dev Brain Res* 91, 274–280.
- Giedd, J.N., Snell, J.W., Lange, N., Rajapakse, J.C., Kaysen, D., Vaituzis, A.C., Vauss, Y.C., Hamburger, S.D., Kozuch, P.L. and Rapoport, J.L. (1996b) Quantitative magnetic resonance imaging of human brain development: ages 4–18. *Cereb Cortex* 6, 551–560.
- Giedd, J.N., Vaituzis, A.C., Hamburger, S.D., Lange, N., Rajapakse, J.C., Kaysen, D., Vauss, Y.C. and Rapoport, J.L. (1996c) Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4–18. *J Comp Neurol* 366, 223–230.
- Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*. Chapman and Hall, London.
- Harvey, P.H. and Krebs, J.R. (1990) Comparing brains. *Science*, 249, 140–146.
- Ho, K.C., Roessmann, U., Straumfjord, J.V. and Monroe G. (1980a) Analysis of brain weight: I. Adult brain weight in relation to sex, race and age. *Arch Pathol Lab Med* 104, 635–639.
- Ho, K.C., Roessmann, U., Straumfjord, J.V. and Monroe G. (1980b) Analysis of brain weight: II. Adult brain weight in relation to body height, weight and surface. *Arch Pathol Lab Med* 104, 640–645.
- Holloway, R.L. (1980) Within-species brain-body weight variability: a reexamination of the Danish data and other primate species. *Am J Phys Anthropol* 53, 109–121.
- Jerison, H.J. (1991) Brain size and the evolution of mind. *Fifty-Ninth James Arthur Lecture on the Evolution of the Human Brain*. American Museum of Natural History, New York.
- Lange, N. and Ryan, L. (1989) Assessing normality in random effects models. *Ann Stat* 17, 624–642.
- Lange, N., Giedd, J.N. and Rapoport, J.L. (1996) Regression calibration for in vivo brain structure size estimation by MRI. Technical report, NIMH, Bethesda, MD.
- Manly, B.F.J. (1991) *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- Mathalon, D.H., Sullivan, E.V., Rawles, J.M. and Pfefferbaum, A. (1993) Correction for head size in brain-imaging measurements. *Psychiatry Res Neuroimag* 50, 121–139.
- Miller, R.G., Jr. (1986) *Beyond ANOVA: Basics of Applied Statistics*. John Wiley and Sons, New York, p. 264.
- Pakkenberg, H. and Voight, J. (1964) Brain weight of the Danes. *Acta Anat* 56, 297–307.
- Pearson, E.S. (1931) The analysis of variance in cases of non-normal variation. *Biometrika* 23, 114–133.
- Pfefferbaum, A., Mathalon, D.H., Sullivan, E.V., Rawles, J.M., Zipursky, R.B. and Lim, K.O. (1994) A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Arch Neurol* 51, 874–887.
- Rasband, W. (1993) NIH Image, ver.1.6. Bethesda, MD. National Institutes of Health (Public Domain).
- Snell, J.W., Merickel, M.B., Ortega, J.M., Goble, J.C., Brooker, J.R. and Kassell, N.F. (1995) Boundary estimation of complex objects using hierarchical active surface templates. *Pattern Recognition* 28(10), 1599–1609.
- Sokal, R.R. and Rohlf, F.J. (1981) *Biometry*, 2nd edn. W.H. Freeman, New York, p. 139.
- van Belle, G. and Martin, D.C. (1993) Sample size as a function of coefficient of variation and ratio of means. *Am Stat* 47, 165–167.
- Winer, B.J. (1971) *Statistical Principles in Experimental Design*, 2nd edn. McGraw-Hill, New York.