Geometric Morphometrics of Corpus Callosum and Subcortical Structures in the Fetal-Alcohol-Affected Brain

FRED L. BOOKSTEIN,[1*] **PAUL D. SAMPSON,**[2] **ANN P. STREISSGUTH,**[3] AND **PAUL D. CONNOR**[3]
[1]*Institute of Gerontology, University of Michigan, Ann Arbor, Michigan 48109*
[2]*Department of Statistics, University of Washington, Seattle, Washington 98195*
[3]*Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington 98195*

ABSTRACT

Background: Although experienced clinicians have been diagnosing fetal alcohol syndrome (FAS) for nearly 30 years, the rest of the spectrum of fetal alcohol damage is not being classified effectively. This article describes a quantification of neuroanatomical structure that may supply a useful discriminator of prenatal brain damage from alcohol. It is demonstrated in a data set of adults of both sexes.

Methods: Ninety adults (45 males) were examined by magnetic resonance imaging (MRI). These subjects were group-matched for age and ethnicity across three diagnoses: FAS, fetal alcohol effects (FAE), and normals. All FAS and FAE were heavily alcohol-exposed in utero; normals were not. From T_1-weighted MR brain images, we extracted 3D morphometric representations of shape for 33-landmark point configurations and 40-point outlines of the corpus callosum along its midline (a slightly nonplanar structure).

Results: There are striking differences between exposed and unexposed in the statistical distributions of these two shapes. The differences are better characterized by excess variance in the exposed group than by any change in average landmark or outline shape. For each sex, combining the callosal outline data with the landmark data leads to a powerful quadratic discriminator of exposed from unexposed. The discriminating features include the relationship of brain stem to diencephalon, and localized variabilities of callosal outline shape, but not diagnosis (FAS vs. FAE).

Conclusions: Statistical analysis of brain shape is a powerful new source of information relevant to fetal alcohol spectrum nosology and etiology. Patients with FAS and FAE do not differ in these brain shape features, but both differ from the unexposed. The aspects of brain shape that are especially variable may be entailed in the underlying neuroteratogenetic mechanisms.

Teratology 64:4–32, 2001.   © 2001 Wiley-Liss, Inc.

INTRODUCTION

The teratogenic properties of alcohol were suspected when children with unusual faces, growth deficiency, and a variety of abnormalities were observed among the offspring of alcoholic women (Rouquette, '57; Lemoine et al., '68; Jones et al., '73). Jones and Smith ('73) coined the term "fetal alcohol syndrome" (FAS). Soon afterward, additional groups of children with this diagnosis were reported from France (Dehaene et al., '77), Germany (Majewski et al., '76), Sweden (Olegård et al., '79), and elsewhere. By 1978, after more than 250 published case reports (Clarren and Smith, '78), it was clear that FAS was only one identifiable form of an extended range of disorders associated with maternal alcohol abuse. By 1980, the teratogenic properties of alcohol had been clearly established in animal models (cf. Randall, '77), and neurobehavioral consequences of prenatal alcohol exposure were being discovered that were not necessarily associated with morphologic abnormality or even growth deficiency (Martin et al., '77; Ouelette et al., '77; Sander et al., '77; Landesman-Dwyer et al., '78; Streissguth, '78; Streissguth et al., '80a, b). By the mid-1980s, there was a large body of literature from both animal and human data congruent with the principles of teratology as set out by Wilson and Fraser ('77), showing multiple central nervous system (CNS) effects of prenatal alcohol exposure (West, '86) that depend on the dose, timing, and condition of exposure. Related literature (Riley and Voorhees, '86) enumerated teratogens in addition to alcohol for which brain damage was not necessarily accompanied by morphological abnormalities or growth deficiency.

Yet throughout this period, clinical diagnosis remained focused on FAS. The non-FAS range of the spectrum of fetal alcohol damage has been variously

TABLE 1. Age, race, and IQ by sex and diagnosis

	White	Nat. Am.	Black	Mean age	Age range	Mean IQ	IQ range
Males							
Normals	10	3	2	24.2	19.1–36.9	113	85–137
FAE	9	4	2	23.6	18.6–32.4	87	67–107
FAS	9	6	0	23.9	18.5–36.9	84	65–113
Females							
Normals	9	5	1	23.1	19.0–36.2	114	93–136
FAE	11	4	0	24.9	18.0–37.4	83	75–106
FAS	10	4	1	25.1	18.2–35.9	82	66–102

Nat. Am., Native American; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.

labeled fetal alcohol effects (FAE) (Clarren and Smith, '78; Hanson et al., '78), expanded FAS (Shaywitz et al., '80), alcohol-related birth defects (NIAAA, '83), prenatal exposure to alcohol (Riley et al., '95), partial FAS and alcohol-related neurodevelopmental disorder (ARND) (Stratton et al., '96), or atypical FAS and alcohol encephalopathy (Astley and Clarren, '00). For 20 years, until just recently, there were few clinical protocols for diagnosis within this range in the individual case, despite the overwhelming evidence that alcohol is teratogenic throughout pregnancy (Guerri, '98) at doses and timings of exposure that may not produce observable dysmorphology or growth deficiency. The widespread abuse of alcohol in our society, combined with this persistent nosological confusion regarding "partial manifestations" of the syndrome, have led to major problems identifying and meeting the therapeutic needs of individuals prenatally damaged by alcohol over this extended range of effects.

A quantitative evaluation of brain morphology might improve this diagnostic process. This article is one in a series examining brain morphology and neuropsychological deficit in a balanced sample of 180 subjects equally divided by age (adults and adolescents), sex, and diagnosis (FAS, FAE, and normals for comparison). The present article examines alcohol-related brain damage using data from three-dimensional (3D) analysis of magnetic resonance images (MRI) for the full sample of 90 adult subjects but defers analysis of adolescents and of neuropsychological sequelae at all ages to later manuscripts.

In studies of other severe childhood disorders, such as schizophrenia or autism, subjects are characterized by typical behaviors of unknown etiology. By contrast, in studies of FAS/FAE, patients have all been damaged by a known teratogen, prenatal exposure to ethanol: they represent the spectrum of consequences of a biological process the cause of which is known. Exploiting this knowledge of etiological homogeneity, in recent years several investigators using MRI have reported morphologic abnormalities in patients with FAS/FAE (Mattson et al., '96: diencephalon, cerebellum, and basal ganglia; Riley et al., '95: corpus callosum; Swayze et al., '97: corpus callosum) that arise from the common embryological challenges confronting these patients' brains. The present article shares this thrust, as well as the rich data resources of contemporary MRI, but

exploits a considerably more sophisticated analytic strategy for neuroanatomic data.

The methodology we exploit in this study is landmark-based, as discussed in the section, MR Images and Derived Data. In its handling of size differences, the method is demonstrably more powerful than earlier attempts (e.g., Mattson et al., '96) to "adjust" the size of neuroanatomical components for the microencephaly that is often found to characterize those with the diagnosis of FAS or FAE. The principal sample filter applied in the present study is simply the requirement that the subjects be able to negotiate both the MRI session and the 5-hr neurobehavioral battery. (The neurobehavioral findings will be reported and correlated to the neuroanatomical data in subsequent publications.) The resulting study is the first, we believe, of sufficient sample size and richness of neuroanatomical data structure to develop strategies for individual classification and detection. It is to this elusive question that the present work is addressed, toward the resolution of the conundrum that has driven the entire research program of our group: the prognostically and therapeutically valid classification of patients with brain damage from prenatal alcohol exposure over the full range of forms of damage. Only a broad-spectrum nosology can be expected to drive appropriate service delivery protocols.

SAMPLE

We studied 90 Seattle-area subjects aged 18–37 years, comprising 30 unexposed normals and 60 cases. For brevity, we refer to the pool of all 60 cases as "the exposed," although, of course, each was not only exposed to alcohol prenatally, but also affected, as evidenced by their alcohol-related diagnoses. Thirty had been diagnosed as FAS by a dysmorphologist, 30 as FAE. All groups of 30 were divided equally between males and females, and the subgroups of 15 were group-matched by age and, as far as possible, by ethnicity (Table 1). After giving informed consent, all subjects were examined by the identical protocol. Patient ascertainment was from the Seattle FAS Follow-up Database, accrued over nearly three decades from referrals from dysmorphologists. The diagnosis was made by David W. Smith or one of his fellows or trainees (usually Sterling K. Clarren) after a clinical dys-

morphology examination (Aase, '90). Patients were diagnosed according to the clinical guidelines of the time, for the most part (Jones and Smith, '75; Smith, '76; Clarren and Smith, '78; Smith, '83), before the recent diagnostic changes suggested by the Institute of Medicine (Stratton et al., '96). Those without the full features of FAS were usually classified as FAE, often prefixed by "possible" or "probable." In many ways, they could now be diagnosed as ARND (Stratton et al., '96); however, we have retained the original terminology in this study. Normal subjects were recruited from employees and their children at local health care facilities and educational institutions to match approximately the age and ethnic composition of the exposed group. Potential normal subjects were excluded who had alcohol or drug problems, neurological problems, birth defects, cancer, or human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), who were legally blind, who did not have English as their first language, who had undergone psychological testing during the last year, who had braces, or whose biological mothers had a history of alcohol or drug problems or had ever binged (five drinks or more on an occasion) while pregnant with the subject. Potential exposed subjects were excluded for AIDS, blindness, brain tumor, neurotoxic medications for cancer, or a native language other than English. Exposed subjects were not screened for alcohol abuse, other substance abuse, or multisubstance abuse, as these are known risks secondary to fetal alcohol exposure (Streissguth et al., '96).

The six groups defined by sex and diagnostic category averaged 23–25 years of age (Table 1). The racial/ethnic composition of the sample approximates the composition of the alcohol-affected patient pool. Our neuroradiologist noted two occurrences of "severe callosal abnormality" (one an FAS male, one an FAE male). Although our sample design did not refer to IQ in any explicit way, lowered full-scale IQ is a known consequence of the brain damage that follows fetal alcohol exposure. The exposed fall short of normal mean IQ by almost 2 standard deviations (SD), on average.

MR IMAGES AND DERIVED DATA

T_1-weighted sagittal SPGR images were acquired over a period of 12 min in a GE 1.5T Signa scanner at the University of Washington: TE, 8 msec; TR, 29 msec; and flip angle, 45 degrees. The resulting $256^2 \times 124$ arrays of $0.85^2 \times 1.14$ mm^3 voxels were processed by Edgewarp 3D software (Bookstein and Green, '98).

Most neuroanatomic studies have been based on now conventional volumetric analyses derived typically from detailed manual (possibly computer-aided) identification of neuroanatomic structures and the computation of their volumes. Tissue classification algorithms have also led to analyses of gray and white matter volume. Measurement protocols are being developed for volumetric analyses that we will carry out

later. The current article focuses on biological landmarks, which are named, biologically homologous locations that can be associated with Cartesian coordinates. When landmarks can be identified to characterize structures of interest, the most efficient statistical analyses, as well as the most informative geometric diagrams and biological interpretations of differences or variations in shape, will be based on analyses of landmarks by the best current methods, as reviewed below.

The notion of a landmark is a general one, referring to any geometric locus that might be the target of a label in a textbook illustration. Among the types of landmarks are ordinary geometric points in two or three dimensions, curves in a plane or in space, and 2D surfaces. In this article, we exploit the first two of these types: points and curves.

Landmark points

We began this study believing that other researchers had developed protocols for identification of neuroanatomic landmarks. In particular, we had intended to begin with landmarks suggested by Evans et al. ('91). However, we were unable to find literature citations to any suitable operational definitions for any list of landmarks. (Indeed, many of the points originally proposed by other researchers, including those we proposed to use from the Evans group in Montreal, were not truly point landmarks for which operational definitions could be promulgated.) We therefore conducted an investigation to determine a set of point landmarks that could be reliably identified over a sample of normal or near-normal adult brain images.

To this end, we used the 3D visualization and digitization facilities of the Edgewarp software package from the University of Michigan (Bookstein and Green, '98) to explore operational definitions of landmarks such as those presented in Table 2. In consultation with neuroradiologist David Haynor, University of Washington, we carried out three rounds of training and comparison of digitizations of the point landmarks in Table 2 by our neuropsychologist (P.D.C.) and by an undergraduate research aide. Average inter-observer differences in digitized landmarks (averaging over 5 randomly selected subjects) ranged from ~0.5 mm to <2.0 mm for most landmarks. It was determined, however, that despite generally good agreement, occasional outliers resulting from misidentification of difficult-to-locate landmarks could not be avoided. To ensure reliable landmark location, we adopted the protocol of having all digitizations by the research aide reviewed and adjusted by our neuropsychologist. Our interest was not in establishing reliability levels for identification of these landmarks by other researchers (for which further inter-observer reliability studies would need to be conducted), but simply to provide reliably identified landmarks suitable for analysis in this study.

This investigation resulted in surprisingly few identifiable 3D point landmarks in the human brain, and those mainly subcortical. Table 2 reviews the set of 33

TABLE 2. Landmark points: names and operational definitions

Abbreviation	Name	Sagittal view	Axial view	Coronal view
Tip4[a]	Tip of 4th ventricle	Posterior tip of sharp angle	Most posterior point of CSF	Thin canal of CSF connecting two lateral "pools"
AC[a]	Anterior commissure	Center of small oval of white matter	Center of cylindrical band of white matter between hemispheres	Center of cylindrical band of white matter between hemispheres
PC[a]	Posterior commissure	Middle of vertical arc of little C-shaped appendage to colliculi	Center of cylindrical band of white matter between hemispheres	Center of cylindrical band of white matter between hemispheres
Genu	Genu–CC	Point directly in front of internal genu	Local axis of symmetry of callosum–CSF boundary	Short segment of gray between white matter and CSF
Spl	Splenium–CC	Posteriormost point on CC	Local axis of symmetry of CC–CSF	Short segment of gray between white matter and CSF
Fr–l, Fr–r	Frontal horn of lateral ventricle L, R	Frontalmost point on teardrop shape boundary between brain–CSF	Frontalmost point of boundary between brain–CSF	Slightly off center of small oval of CSF
Cd–l, CD–r	Caudate L, R	Triple point posterior bottom of horn of ventricle at spring of thin white capsule	Triple point of CSF, caudate, and white matter	Beginning of the appearance of gray matter of caudate
Chi	Optic chiasm	Anterior/inferior limit of the chiasm	Front of the chiasm at the midline	Locus where chiasm disappears
SC–l, SC–r	Superior colliculus L, R	Center of hump on outer boundary	Tip of hump	Tip of hump
IC–l, IC–r	Inferior colliculus L, R	Center of hump on outer boundary.	Tip of hump	Center of white oval
PS–l, PS–r	Posterior point of hippocampus L, R	Top of hippocampus below thin strip of white matter below CSF		Cusp of fimbria where turns down to capsule of the hippocampus
AH–l, AH–r	Temporal horn L, R	Anterior limit of CSF	Anterior point of CSF	Center of little spike of CSF lateral to hippocampus
IntG	Interior genu–CC	Point of sharpest curvature on interior border of CC	Local axis of symmetry on boundary of CC	Local axis of symmetry on boundary of CC
Fx–l, Fx–r	Fornix L, R	Center of the spring of fornix emerging from CC	Center of the spring of fornix emerging from CC	Center of the spring of fornix emerging from CC
TpP	Top of pons	Deepest point of crevice of pons	Posteriormost point of wedge of CSF	Center of small part of CSF at back of pons
BtP	Bottom of pons	Deepest point of crevice of pons	Posteriormost point of wedge of CSF	Center of small part of CSF at back of pons
Obex	Obex	Interior corner of shelf of white matter at beginning point of spinal aqueduct	Deepest point of wedge of CSF	Bottom of wedge of CSF
CCS	Central cerebellar sulcus	Hull of cerebellum bisecting the central sulcus	Local midline of cerebellum	Local midline of cerebellum
Pd–l, Pd–r	Cerebellar peduncle L, R	Center of the spring of the peduncle as it emerges from midbrain	Center of peduncle	Center of peduncle
Rost	Rostrum–CC	Sharp corner of CC	Axis of symmetry at narrowest point of CC	Axis of symmetry of last pixels of CC at "drooping moustache" of white matter
FrPo[b]	Frontal pole	Anteriormost point of cortex	Anteriormost point of cortex	First slice in which cortex is visible
OcPo[b]	Occipital pole	Posteriormost point of cortex	Posteriormost point of cortex	Last slice in which cortex is visible
Top[b]	Top of brain	Superiormost aspect of cortex	Superiormost aspect of cortex	Superiormost aspect of cortex
Left[b]	Left side of brain	Last slice in which cortex is visible	Leftmost point of cortex	Leftmost point of cortex
Rght[b]	Right side of brain	Last slice in which cortex is visible	Rightmost point of cortex	Rightmost point of cortex

CC, corpus callosum; CSF, cerebrospinal fluid; L, left; R, right.
[a]Three points are used to reorient to a midsagittal plane.
[b]For outer extremes of the brain, cross-hairs are placed on the posterior commissure with one line running through the anterior commissure in a plane perpendicular to the midsagittal. Landmarks are then placed along cross-hair lines.
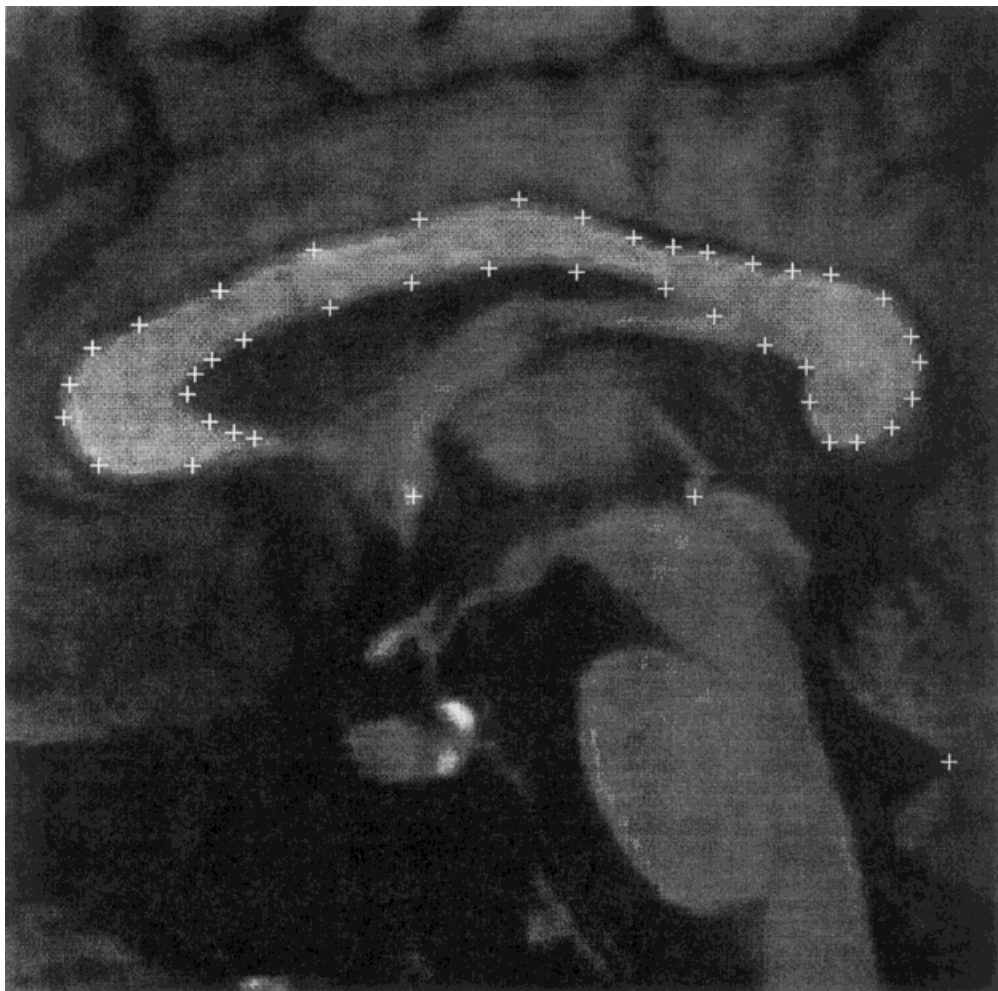
Fig. 1. Aspects of digitizing the corpus callosum in 3D. (**left**) Full outline, one subject, as projected onto a near-parasagittal plane. (The semilandmarks + do not actually lie precisely within this plane or any single plane.) Except for rostrum, all of the points on the outline have been allowed to slip to minimize bending energy with respect to a template form. Other landmarks: anterior commissure, posterior commissure, tip of fourth ventricle, and left and right brain boundaries (not shown) at posterior commissure. (**right**) A typical point of the outline (semilandmark 28, the one used in later figures) for one subject, showing how approximate symmetry is used to determine the point digitized. The section here is perpendicular to the estimated tangent line of the outline in its vicinity. The point digitized is slipped perpendicular to that tangent line until it lies at midgray voxel value on the "midline" located visually by evidence of symmetry, and then is moved along the tangent direction to minimize bending energy of the configuration as a whole.

we ultimately came to use. Twelve are subcortical midline points, 16 come as 8 pairs of bilateral points, and 5 are "extremal landmarks" (Bookstein, '91) at the outer boundary of the cortex with the cranium.

Callosal outlines

There is a small but persuasive literature of alcohol effects on the corpus callosum (e.g., prenatal alcohol exposure seems to be the principal known cause of partial or total callosal agenesis; Riley et al., '95). We therefore determined to represent it more richly than could be managed by its four unpaired landmarks: genu, internal genu, splenium, and rostrum (Table 2). Beginning with the landmark point rostrum already located, we digitized the rest of the callosal midline outline as a 39-point sequence of semilandmarks (land-

marks "slipped" along the outline; see below). Tracing was carried out by one of the authors (F.L.B.), who was blind to the diagnostic group. A typical set of digitized locations, totaling 40 points, is at left in Figure 1. These are spaced roughly inverse to curvature on a reference form (the first one digitized). At right is a typical digitizing scene: a perpendicular section through a candidate point along the lower border of isthmus. Each digitized point lies precisely on the "vertical" (axis of symmetry) of an image like this at the apparent boundary of callosum. The word "vertical" is in inverted commas because anatomically it is oriented perpendicular to the callosal outline, and so lies truly vertically only at a few scattered points (top and bottom of the arch, bottom of the splenium). The cross-hairs in this panel indicate how within any such plane, a plane containing
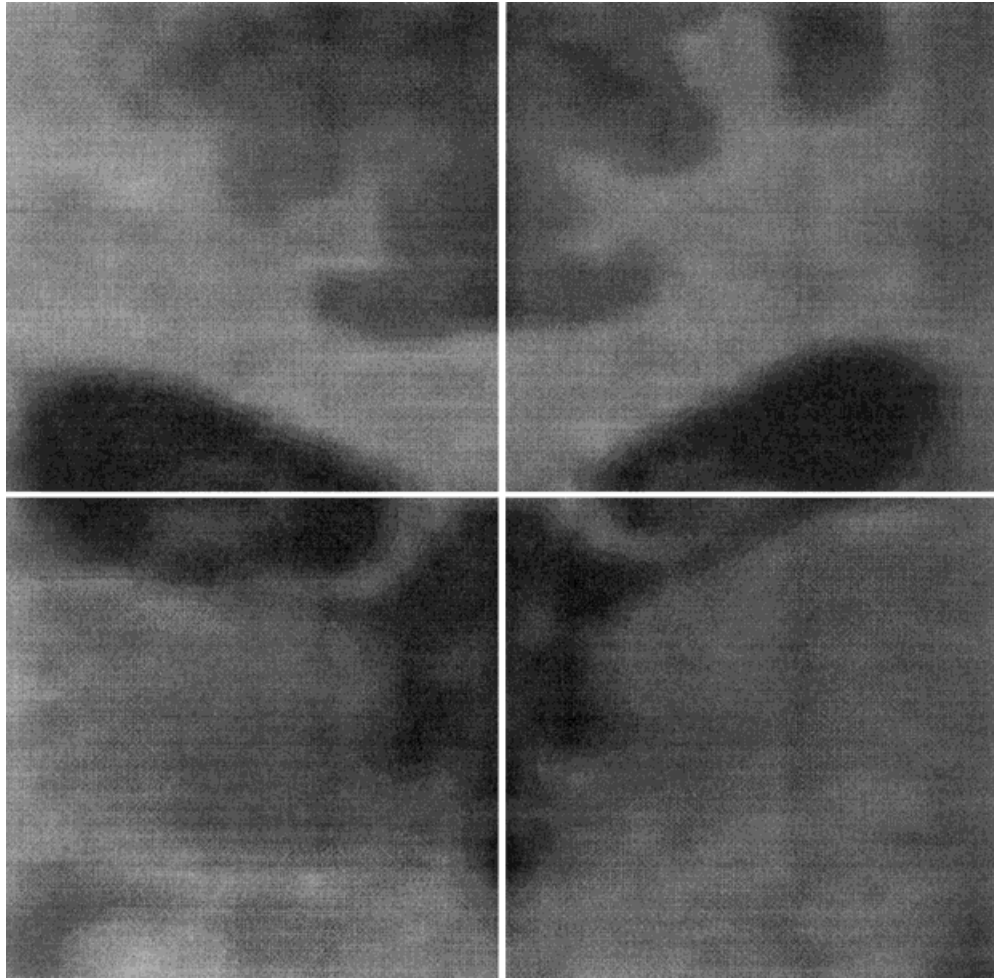
Figure 1. (Continued.)

the normal to the callosal midline somewhere, the particular point we seek is taken by its mediolateral position and its location on the grayscale gradient of the neural tissue. The third coordinate, pointing out of this page, is the coordinate that passes tangentially around the callosum in the left-hand image; it is computed by the sliding algorithm reviewed above, rather than being selected by the digitizing technician.

Even in the normal subgroup, these curves are distinctly nonplanar: they do not lie in any possible "midsagittal plane." For purposes of the statistical analysis to follow, the mediolateral (out-of-plane) coordinate has been suppressed; its distribution shows no differences among the diagnostic subgroups in either mean or variance. Reliability of these outlines over a random subsample of six digitized independently several months apart showed a reproducibility of <0.6 mm in the trajectory of the curve averaged over the whole outline. This is comparable to the standard error of the better landmark points and is considerably smaller than the magnitude of the effects reported in the present analyses (reported in units of Procrustes distance, not mm).

MORPHOMETRIC METHODS

After inspection for errors, the configurations of neuroanatomical landmarks or callosal semilandmarks were analyzed by standard methods of the morphometric synthesis. The entire subcortical region is quantified in a single multivariate analysis that considers not only differences in average size or volume, but also deviations from normal shapes and spatial relations of the different parts of the brain, whether the pattern of damage be found to be gross or localized. The modern statistical toolkit underlying this work has been the subject of a recent textbook (Dryden and Mardia, '98) and several recent review articles (Bookstein, '96, '97a, '98). There is one statistical space for the variation of the shape of the landmark set and another for the variation of the shape of the callosal outline. Thus, in their geometry the landmark data and the callosal outline data have been kept separate. In either of these spaces, the shape of the geometric object in question (a landmark set, or an outline) is represented by one single "observation" per specimen—a rather complicated observation, yes, but one that is treated as a

single algebraic entity for statistical purposes—regardless of the number of points in the original digitized representation. When the spaces are combined (to produce the findings displayed in Fig. 12), it is by statistical, not geometrical, procedures.

Getting from Cartesian coordinates to shape

The construction of a statistical shape space for outline data is a special case of the construction that applies to landmark data. This section provides a quick sketch of the standard method for landmark points and then indicates the nature of the extension to handle smoothly curving forms such as the callosum. The Appendix at the end of this article reviews the standard landmark methods in considerably greater detail.

Shape is the information about landmark configurations that remains unchanged under adjustments of position, orientation, or scale. One straightforward way of representing this information for statistical purposes begins by considering a *shape distance* between any pair of landmark configurations, and then constructing a useful set of geometric coordinates, *shape coordinates,* for which this shape distance is the appropriate "Euclidean" sum of squares. Since the great original paper by Kendall ('84), the distance used for these spaces has always been one or another modification of the following definition of *Procrustes distance.* It is convenient to begin by removing location and size information from each configuration separately, by centering each at its own center of gravity and scaling each to a fixed sum of squares around that center. The scaling factor divided out in this step, called *centroid size,* remains available as a size measure for use at any subsequent stage of analysis. If all forms are standardized in this way, the Procrustes distance between any two is simply the sum of squares of the ordinary Euclidean distances between the matching landmarks of the two landmark configurations when one of them is freely rotated (around the common center of gravity) until this sum of squares is minimized.

Shape averages and shape coordinates

From the shape distance formulation, the rest of the statistical scheme follows very directly. One can define the *average shape* of a set of landmark configurations as the shape from which they have, taken together, the least summed squared Procrustes distance. (This is precisely analogous to the least-squares property of ordinary arithmetic averages: the average of any set of numbers is the value that has the least summed squared difference from the numbers actually being averaged.) After that average has been computed, each original shape of the data set can be superimposed upon it by the rotation described in the preceding paragraph—the rotation that supplies the actual minimum sum of squares serving for its Procrustes distance to the average. The locations at which the original landmarks arrive after this rotation serve us as the shape coordinates of the original landmark configurations with respect to the sample as a whole. For k landmarks, there are $2k$ of these coordinates for 2D data, or $3k$ for three-dimensional data. (But four dimensions of their space (for 2D data) or seven dimensions (for 3D data) necessarily have no variance; instead, they express the constraints on position, orientation, and scale that were imposed during the course of the construction.) These coordinates serve as the set of variables that make possible an analysis of shape by otherwise familiar multivariate procedures. For instance, in Figure 7, each point shown stands for two shape variables—its x-coordinate and its y-coordinate—with group comparisons going forward in terms of the averages, the variances, and the covariances of those variables. (For a general discussion of the relationship between shape coordinates and shape variables, including the most familiar shape variables such as angles or ratios of distances, see Bookstein, '91). All the formulas entailed in this sequence of steps are available in the Dryden textbook and the Bookstein reviews; software is available free of charge, to carry out the computations on most scientific research computer platforms (see the website http://life.bio.sunysb. edu/morph/ maintained by F.J. Rohlf at the State University of New York at Stony Brook). Our analyses were carried out in the Splus statistical system, using functions written by the authors.

Shape coordinates for outline data

A corresponding analysis for the callosal outline data takes into account the indeterminacy of "homologous" points along extended curves like this midline. Bookstein ('97b) suggested that, beginning from any plausible sampling of points along the curves of the sample, a Procrustes average in the sense just reviewed for landmarks alternate with a "sliding" operation that redistributes the semilandmarks of each outline along that outline in such a way as to minimize the "bending energy," quantified in one specific algebraic way, that characterizes the relation of the outline to the full sample average. This approach is preferable to even spacing of points, in permitting coordination of information between top and bottom of the arch, and is preferable to approaches that assign coordinates out of a "center" for the reasons reviewed in Bookstein ('91). For a justification of the relevance of bending energy to this context, see Bookstein ('99). In the callosal data set, our bending energy computation (for sliding) also took account of four conventional landmark points: the commissures, the tip of the fourth ventricle, and the rostrum, the only good point landmark actually located on the callosum. At the convergence of this alternating algorithm, one arrives at a sample average shape and a set of shape coordinates just as before. Points generated by this algorithm are called *semilandmarks.*

Principal components analysis for shape

Once shape coordinates are in hand, a variety of multivariate statistical tools become available that correspond to those that apply to more typical biostatistical data sets. This analysis exploits two of these tools:

principal components analysis and testing statistical significance of group differences in average shape or in the variability of shape.

Whether for landmark data (points) or for semilandmark data (outlines), principal components analysis of shape is carried out by ordinary principal components analysis of the shape coordinates just described. The analysis uses their covariance matrix, not their correlation matrix, in order to preserve the Procrustes geometry through subsequent steps. Within the context of shape analysis, these components are often diagrammed as deformations (warps) of a grand mean (see Fig. 5), and are thus called *relative warps*. (In this article, "deformation" is used only in its mathematical sense, a smooth map from one picture to another.)

For ordinary sets of variables, the first principal component is characterized as the linear composite that has the greatest variance among the set of all possible composites whose coefficients sum in square to 1. Similarly, the first relative warp is the composite shape variable (pattern of joint landmark rearrangement) having the largest variance among all the shape variables of a given Procrustes length. If the concept of the Procrustes length of a variable seems forbidding, there is an exact equivalent that may be more accessible: the Procrustes length of a shape variable is proportional to its variance on a model of "pure digitizing noise," the same small variance in every direction at every landmark. Biological data can often be modeled effectively by this noise distribution, perhaps after systematic factors (e.g., prenatal exposure) have been controlled. The second relative warp is the composite shape variable having greatest variance (per unit Procrustes length) of all those that are uncorrelated with the first relative warp of the sample, and so on. Every subject in the data set has a score (projection) on every relative warp of the data set, and these scores can be scattered to look for patterns or clusters, tested for mean difference between groups, correlated with ostensible causes or effects of shape, and generally treated just like any other set of principal component scores in any other application of multivariate statistics.

Testing statistical hypotheses about shape

Although this morphometric version of principal components analysis is only slightly modified from the standard approach, the way in which one tests hypotheses of mean shape difference changes considerably. In most morphometric data sets, there are more of these shape coordinates (variables) than subjects. Hence our significance tests will usually be permutation tests (Good, '94). In a permutation test, a quantity is selected that captures a scientific question about the relation between two aspects of the data structure. For instance, in studying the association between diagnosis and callosal outline in our 45 males, we might select the Procrustes distance between the average callosal outline shapes of exposed and unexposed as an interesting measure of the scientific signal we are examining. Then the distribution of this quantity is computed

over a very large collection of "pseudo datasets" in which diagnosis (FAS vs. FAE vs. normal) is randomly reassigned over the 45 adult male callosal outlines. The significance level of the empirical association between diagnosis and callosal outline, for the chosen quantity (the mean difference), equals the probability that a random permutation of this type results in a Procrustes difference between averages at least as great as the value actually observed. It was Ronald Fisher himself (the "F" of the F-test) who first noted that this is what we actually mean by a statistical significance level, to which any other quantity deriving from textbook formulas, including his own, merely approximates. For data sets large enough to preclude looking at "all possible permutations" (about 340 billion, for our sample of 30 exposed vs. 15 unexposed males), one looks at an adequately large random sample, here, a few hundred to a thousand or so. Permutation tests are easily performed in any of the standard statistical software packages and, for landmark data such as these, are built into Rohlf's TPSregr program, available free of charge for Windows PCs from the Stony Brook, New York, site.

In the landmark data set, the feature underlying the test for mean shift is Procrustes distance. In the outline data set, it is Procrustes distance in the direction normal to the outline. For demonstrations of hypervariability, it is the ratio of the sum of variances of the first three principal components of shape for the exposed to that for the unexposed.

MAIN MORPHOMETRIC COMPARISONS, BY DIAGNOSIS AND SEX

Our morphometric analyses involve comparisons of both kinds of shape by diagnostic group and sex and principal components of both kinds by sex. There are findings of four different types: size differences, means and principal components of landmark shape, means and principal components of callosal outline shape, and the combined pattern of the two shape components. In this section, each of these is recounted separately, leading to a total of six distinct findings. In general, group differences in shape variability far outweigh group differences in average shape or average size, allowing a startlingly powerful discrimination of the exposed from the unexposed by shape alone, whereas the two exposed subgroups (FAS and FAE) differ in shape hardly at all. The reader who does not wish to wade deeply into the details may choose at this point to turn directly to the section, Summary of the Findings, which highlights the six main findings separately from the supporting computations.

Size

A morphometric analysis of either landmarks or semilandmarks should begin by considering centroid size, the scaling factor divided out in the course of producing shape coordinates. Means and standard deviations of this descriptor are presented in Table 3 for

TABLE 3. Size measures by diagnosis and sex*

Group	Landmark CS	Callosal CS	Callosal area
Normal M	248.9 ± 4.8	26.25 ± 1.72	1159 ± 155
FAE M	238.2 ± 9.1	24.22 ± 1.67	1057 ± 216
FAS M	240.5 ± 7.7	24.63 ± 1.99	1103 ± 227
Normal F	241.7 ± 7.3	25.52 ± 1.80	1199 ± 165
FAE F	233.8 ± 6.9	24.73 ± 1.29	1154 ± 181
FAS F	224.6 ± 9.4	23.15 ± 1.10	928 ± 162

CS, centroid size; M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Means $\pm$ SD in digitizing units of mm, except for area, which is in units of mm^2.

the two different data structures of this study, the landmark configuration and the callosal outline configuration. For the landmark data in the males, both exposed groups differ by about 4% from the normal mean. The females diagnosed with FAE differ from the normal females by about the same amount, but the FAS females were about 8% in deficit. Student's t-tests for these comparisons are all significant at $P < 0.01$. For this landmark configuration, centroid size is approximately proportional to "shoebox size," the diagonal of a rectangular shoebox around the landmark configuration as a whole. For the centroid size of the callosal outline, the difference of 7% between the normal mean and the average for the exposed pool is significant at $P \sim 0.002$ by t-test. For forms that are this long and narrow, centroid size is approximately proportional to the diameter of the callosum itself, the distance from genu to splenium. For the females, the FAS mean is clearly different from both of the others. In terms of callosal area, the conventional alternative to centroid size, the male diagnostic groups likewise show a 7% shortfall for the exposed pool, but none of these differences is significant, as the within-group coefficient of variation for area is far greater than that for length. For the females, mean callosal area in the FAS subgroup is again clearly and significantly different from the other two ($P < 0.002$).

Our shape findings are reviewed in three subsections: landmark shape variation, callosal outline shape variation, and the combination.

Landmark shape variability

Figure 2 illustrates the mean landmark shape configurations according to their three cardinal views (from the front, the top, and the side). The axes are in units of Procrustes coordinates, which are dimensionless. In the top row, the landmarks are named at their grand mean locations; in the bottom row, the sex- and diagnosis-specific means are shown, exaggerated fivefold away from the grand means for visibility. The Procrustes distances among these six means are shown above the diagonal in Table 4, and their significance levels (by permutation test using 500 permutations) below the diagonal. In general, no differences attributable to syndrome are as large as the difference between the male and female normals, and within sex no mean difference between any pair of the three diagnostic groupings is significant. Male and female normals, as

well as male and female FAE, are very significantly different, whereas male and female FAS are not too different. However, such contrasts are not the primary interest of this study. Although the groups are known to differ in centroid size, and although centroid size is a covariate of shape variation in general human populations, it is not appropriate to "correct" any of these comparisons for size difference, as brain size is determined by the same prenatal dose that led to the diagnostic assignments of these subjects in the first place.

We learn a great deal more from the examination of relative warps than we could glean from these mean comparisons alone. Figure 3 shows the scatter of the first two relative warps separately by sex. Recall that these are the patterns of relatively greatest shape variation, each one standing for a set of correlated shifts of all the landmarks jointly. Like any other principal component, the pattern they delineate pertains to the pool of all the male subjects, or all the females, and not to any single subject. In either sex, the scatter of the normal sample (filled circles) is much less than that of the $+$ and $\times$ symbols standing for the exposed. The original analysis was of the first three relative warps, and tests were carried out using that slightly larger subspace. For the males, the significance level of the extra variance in the exposed subsample, summed over the first three relative warps, is $P \sim 0.01$; for the females, it is $P \sim 0.025$. Most of the excess of variance in the relative warps can be captured by projecting onto the directions shown (separately by sex) in Figure 3. The short axes of the normal subsamples after outliers (one for the males, two for the females) are sequestered. (That is, once those outliers are sequestered, the center of gravity of the remaining control data in each panel is the midpoint of the little segment there, and the best-fitting ellipse to the covariance pattern of these points has the direction indicated as its shortest diameter.) Separately by sex, these patterns of joint landmark rearrangement have considerably more variance in the alcohol-exposed subsample than in the unexposed sample.

Visualization of systematic aspects of 3D shape variation, such as these particular linear combinations, goes forward best via dynamic (tumbling) 3D displays. For publication, one selects a number of still images from these displays. Figure 4 shows the effect of an arbitrary multiple of these changes at all landmarks simultaneously, in the usual three views. In both sexes,
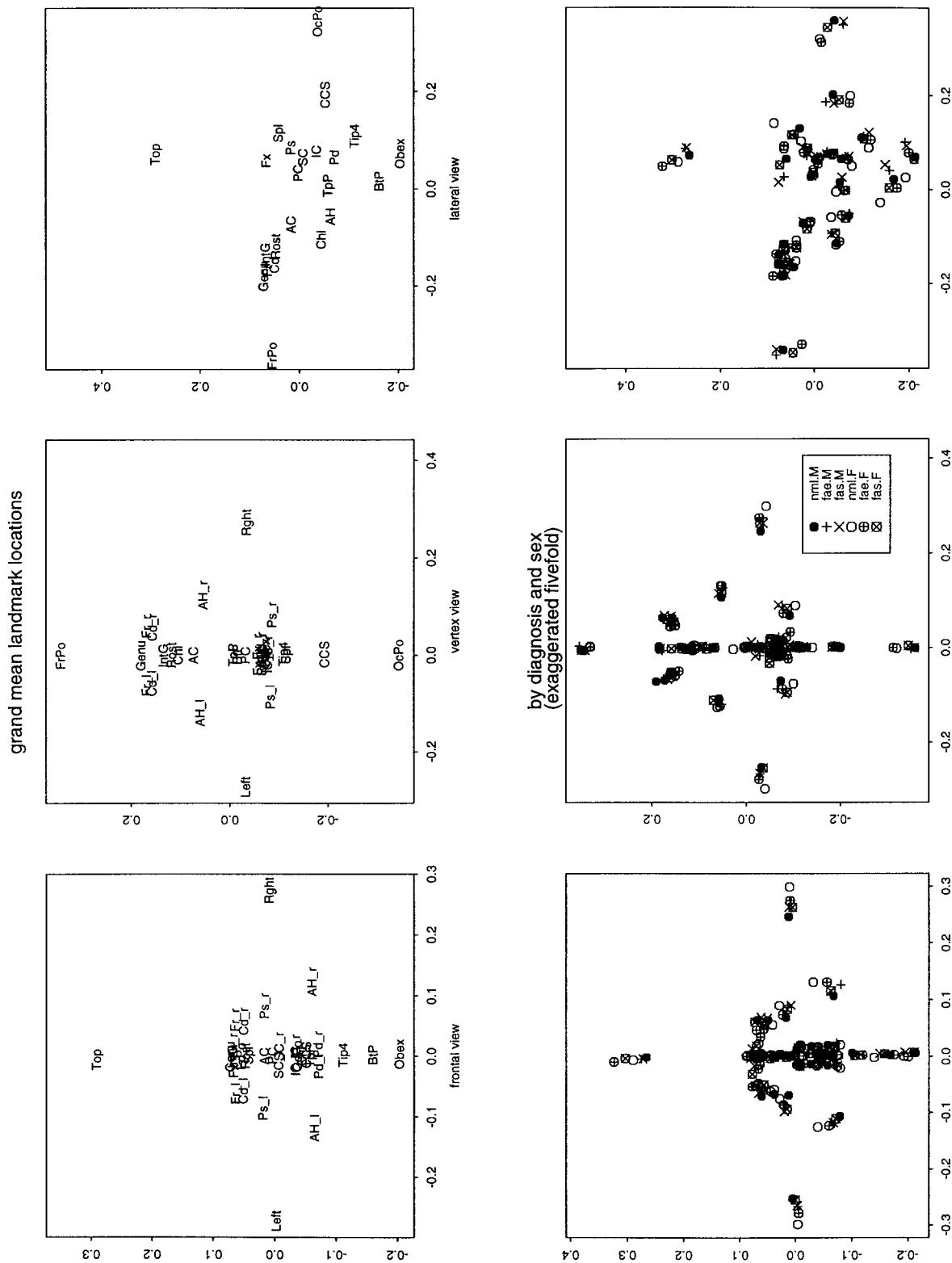
Fig. 2. Averages of Procrustes shape coordinates for the 33-landmark configuration. Upper row, grand means of all 90 cases; lower row, means by diagnostic group and sex as indicated in the legend, represented after fivefold exaggeration of their separation from the grand mean. None of the diagnostic group differences here are statistically significant. Left column, frontal view; center column, vertex view; right column, lateral view. In the lateral view, paired landmarks are represented by their bilateral averages, and left side and right side points are omitted. Axes are in units of Procrustes distance throughout. For full names and operational definitions of all landmarks, see Table 2.

TABLE 4. Procrustes distances between diagnostic groups: landmark point data*

	Male			Female		
	Normal	FAE	FAS	Normal	FAE	FAS
Normal M	—	61	72	138	104	58
FAE M	0.222	—	18	194	129	64
FAS M	0.176	>0.5	—	206	149	71
Normal F	0.002	—	—	—	77	94
FAE F	—	0.002	—	0.052	—	50
FAS F	—	—	0.295	0.323	>0.5	—

M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Entries above upper left–lower right diagonal: squared Procrustes distances between group mean shapes, multiplied by 10^5. Below diagonal: significance levels, according to 500 permutations of diagnostic label over landmark configuration.
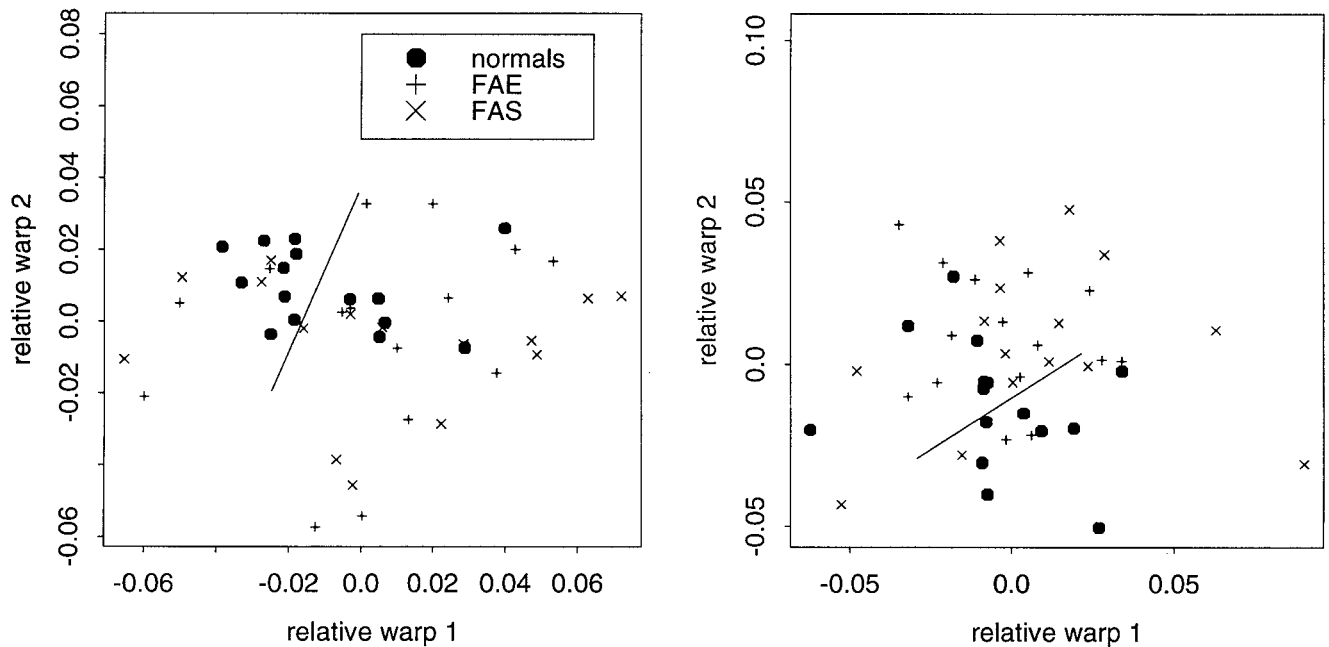
Fig. 3. Relative warps analysis (principal components analysis of shape coordinates) for the landmark data set. **Left,** males; **right,** females. Scatters of relative warp scores 1 versus 2 are in units of Procrustes distance. Short segments in each panel, indicating the dimension of least variance of the normals, are taken along the shorter principal axis for the subscatters of the apparently typical normals (14 males or 13 females). Outliers excluded for this purpose (but included in all statistical testing): for the males, the rightmost normal point; for the females, the leftmost and rightmost normal points.

very little of this correlated hypervariation lies in the mediolateral direction; thus, we can profitably restrict our attention to the lateral view, which nullifies this dimension. In that view, Figure 5 demonstrates the effect of an arbitrary amount of change in the transversal dimensions of Figure 3 by displacement vectors (top) and by splined deformations (bottom). The landmark key is repeated for convenience at lower right. The thin-plate spline, a standard tool of the new morphometrics, depicts a landmark rearrangement as a deformation of the diagram plane in which the landmarks lie. Of all the grids that could be drawn for this

purpose, the spline is the smoothest—the one that has the least extent of local bending in one specific algebraic sense. For an extended explanation and justification, see Bookstein ('96, '97a, '98).

Each scene divides into two regions: structures on or adjacent to the pons, and rearrangements in the vicinity of the corpus callosum. In both sexes, the segment connecting fornix (Fx) to bottom of pons (BtP) is rotated strikingly, in accord with the realignment of grid lines from vertical throughout the middle of both grids. The feature in question varies bidirectionally; for example, the rotation of that vertical grid line could have just as
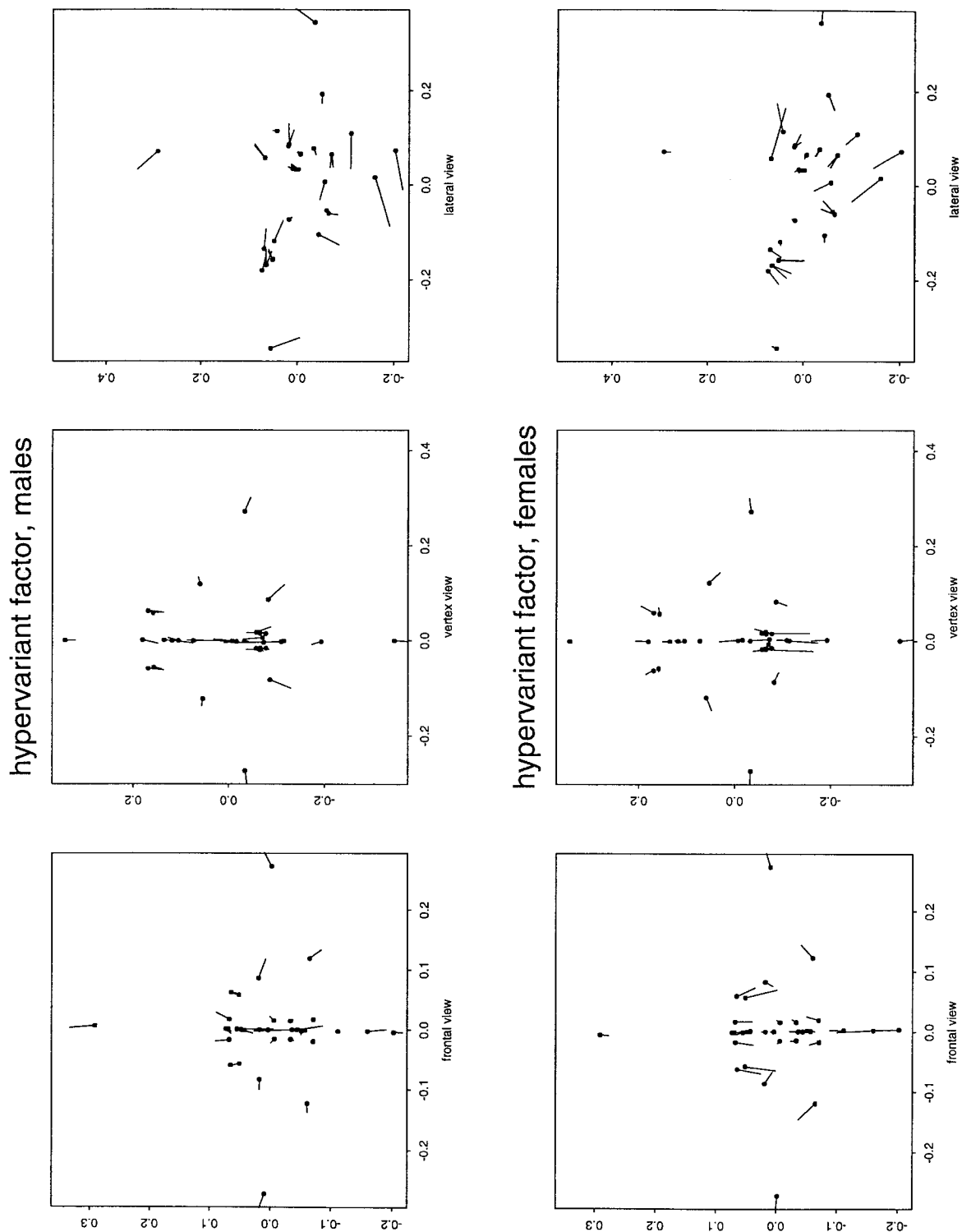
Fig. 4. The segments shown in Fig. 3 correspond to vectors that shift the grand mean landmark shape by correlated changes at all landmarks. **Top row,** an arbitrary multiple of the indicated direction of shape change for the males; **bottom row,** for females. Dots: grand mean configuration. The meaning would be the same if each small segment were rotated 180 degrees around the dots to which it pertains. Most of the information in these vectors pertains to the lateral view.
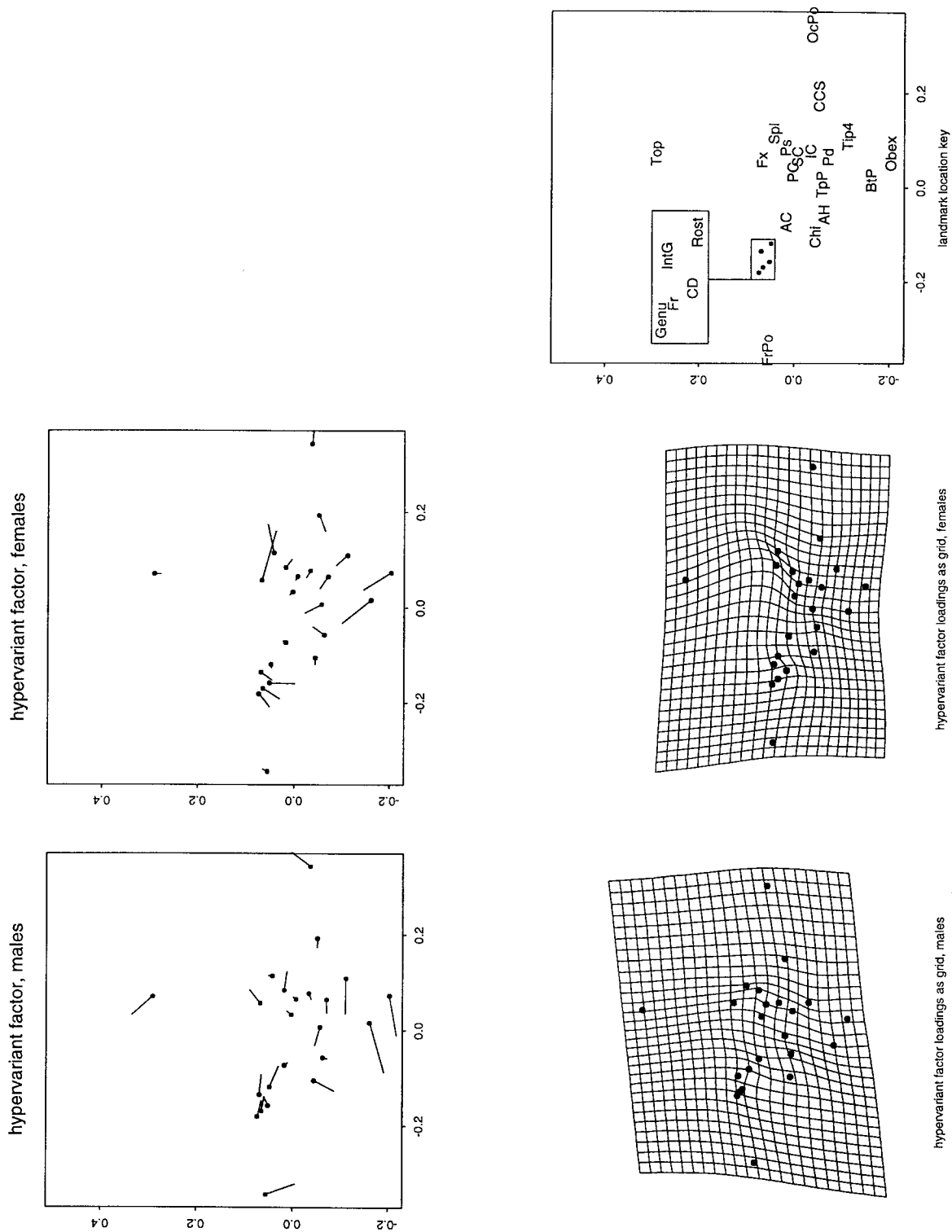
Fig. 5. Landmark hypervariability factors, continued. **Top row,** lateral panels from Fig. 4, with bilateral landmarks averaged and left side and right side landmarks deleted. **Bottom row,** the same shifts diagrammed as thin-plate spline transformation grids. Notice the concentration of the hypervariability at the front of the callosum and along the posterior vertical from splenium through pons.
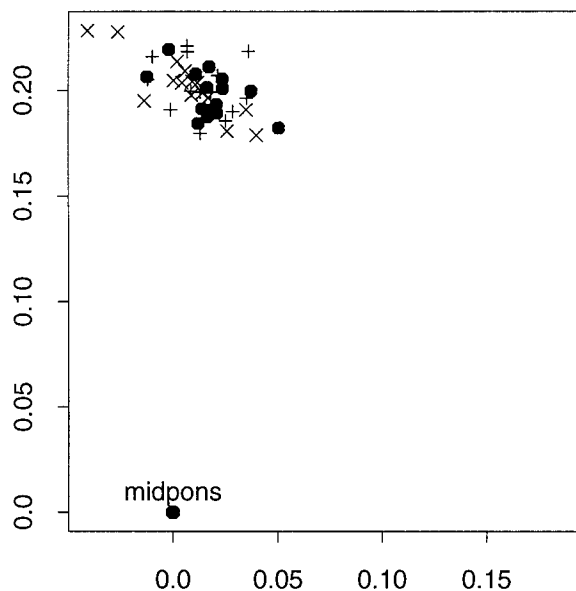
fornix-to-midpons vector, males
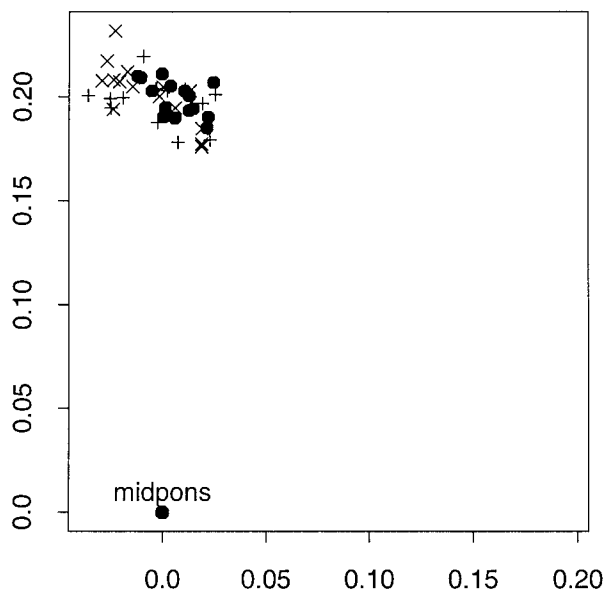
fornix-to-midpons vector, females

Fig. 6. Interpretation of landmark hypervariance factors by localization to a vector from midpons to midfornix in the lateral view. The vectors are plotted as if all images were registered to a common midpons point, as shown. Legend: ●, normals; +, FAE; ×, FAS. "Midpons" is the centroid of the quadrilateral TpP, BtP, Obex, and Tip4.

well been drawn as counterclockwise, by reversing all the vectors in the top panels. In the males, this factor of hypervariability moves BtP, Obex, Tip4, and TpP strongly anteriorly (respectively posteriorly) with respect to the commissures, whereas all the points of the callosum shift posteriorly (respectively anteriorly) except for Spl, the posteriormost point itself. In the females, this brain stem shift is rather more upward (respectively downward) than forward (respectively backward), the principal shift in the posterior callosum is at the fornices, and the front of the callosum shifts downward (respectively upward) rather than backward (respectively forward) with respect to the commissures. Also in both sexes, the central cerebellar sulcus (CCS) has moved distinctly anterior (respectively posterior) with respect to the occipital pole OcPo. Keep in mind that each of these is to be read as a pattern of coordinated shifts in all landmarks, and that what we are looking at is not the shift of an average shape among the diagnostic groups (none of those was significant), but rather the extra variation apparent in the pool of diagnostic groups by comparison with the normals, separately by sex.

These patterns supply tantalizing hints of localization of the excess variation for confirmation in other samples. For example, as shown in Figure 6, the within-sample variance of the vector from "midpons" (BtP, Obex, Tip4, and TpP) to fornix is 0.00049 in the male exposed pool and 0.00051 in the female, versus 0.00020 in the normal males and 0.00032 in the normal females; this variance ratio is significant at $P \sim .005$ for

the males, and at 0.07 for the females, by ordinary F-test. (The variance of a vector is the sum of the variances of its two components separately; this sum is independent of the orientation of the coordinate system in which the components are specified.) The vector may be imagined as a sort of axis for the relationship of brain stem to diencephalon. (In particular, the notion of a "midpons" has no biological reality but is just a convenient way of summarizing the common shift pattern of the four points in its vicinity with respect to the shifts of the two fornix points, which go the other way according to this same pattern.) In this sense, there is considerably more variance in the exposed than in the normals in the angle between brain stem and diencephalon; the hypervariation factors of Figure 4 show this variation along with the displacements of other landmarks, such as those near genu, that happen to be associated with it.

Callosal outline shape variability

The data set of callosal outline shape offers the richer findings here. Figure 7 (left) shows the complete sample scatter of 40 points (39 semilandmarks, together with rostrum) for the complete data set of 90 adults. This variability is commensurate with that of the landmark points analyzed in the previous section. Figure 7 (right) shows the six diagnosis- and sex-specific means. In pairwise mean comparisons (by permutation test), none of the differences among the male diagnostic subgroups is significant, but all pairwise comparisons among averages for the females are nominally signifi-
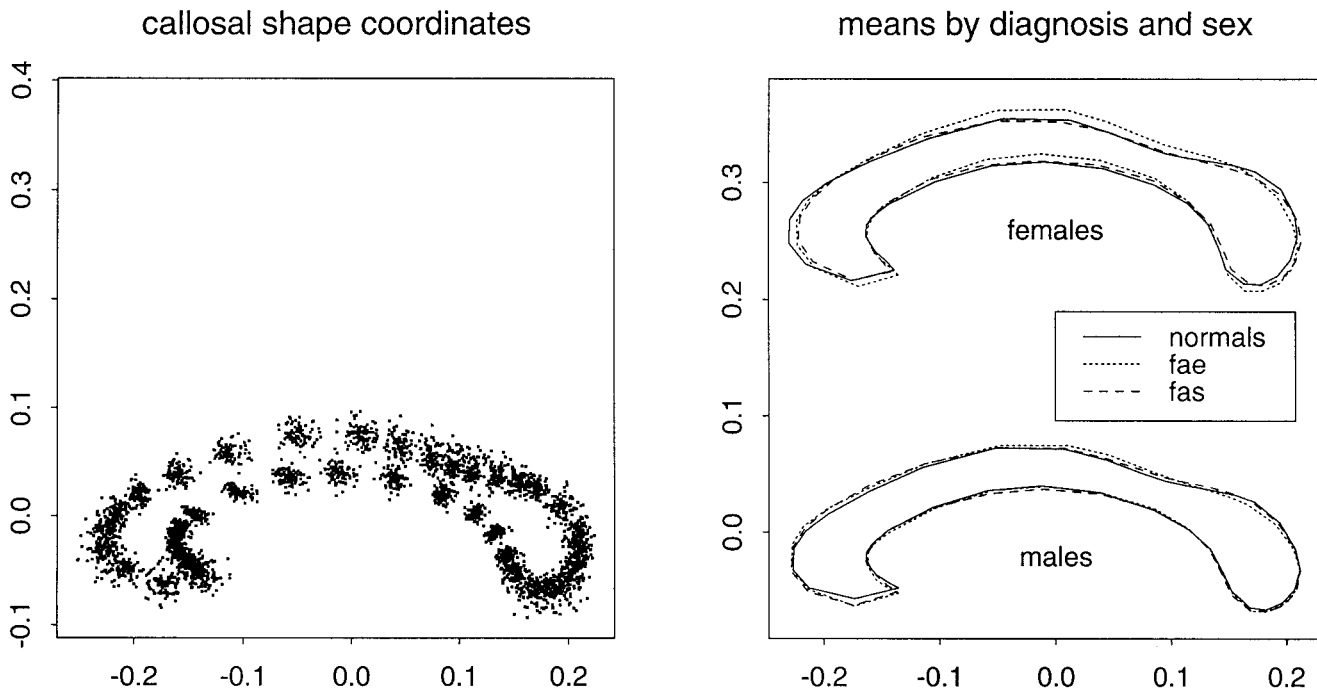
callosal shape coordinates

means by diagnosis and sex

Fig. 7. Procrustes shape coordinates for the projected callosal outlines. **Left,** complete sample scatter of rostrum and 39 semilandmarks for all 90 subjects. **Right,** means by diagnosis and sex.

TABLE 5. Procrustes distances between diagnostic groups: callosal outline data*

	Male			Female		
	Normal	FAE	FAS	Normal	FAE	FAS
Normal M	—	53	34	65	162	63
FAE M	0.425	—	20	71	99	77
FAS M	>0.5	>0.5	—	58	108	57
Normal F	0.126	—	—	—	157	102
FAE F	—	0.120	—	0.006	—	118
FAS F	—	—	0.323	0.040	0.044	—

M, male; F, female; FAE, fetal alcohol effects; FAS, fetal alcohol syndrome.
*Entries above upper left–lower right diagonal: squared Procrustes distances between group mean shapes, multiplied by 10^5. Below diagonal: significance levels, according to 500 permutations of diagnostic label over callosal shape.

cant. Distances and significance levels (by permutation tests of 500 runs) are presented in Table 5. Figure 8 sets out their shape differences pairwise as thin-plate splines, exaggerated threefold for legibility. The normal mean clearly bears a differently shaped splenium from either diagnostic group, and a thicker genu. The isthmus is thinner in the FAS subgroup, with the arch relatively higher in the FAE subgroup. As the top row of Figure 8 suggests, the splenium anomaly for both exposed groups is similar to the representation of the mean shape for normal males as a deformation of that for the normal females.

Even more than the landmarks, our callosal data bespeak a strong excess variability in the exposed subsample. Figure 9 shows the subscatter of its 45 outline shapes for each sex (as in Fig. 7). When the normal subsamples are indicated by connecting the dots, it becomes apparent that the semilandmarks of the syndromal subsample lie well outside the envelope of the outlines for the normal subsamples at several sites. We can test this impression rigorously by applying the same permutation test (variance of relative warps between exposed and unexposed) that was used for the landmark data. Summed over the three callosal shape dimensions of largest Procrustes variance, the net shape variability within the exposed pool is significantly larger than that within the normal pool at $P \sim 0.01$ for the males, $P \sim 0.025$ for the females. Figure 10 localizes this difference by ordinary F-tests (Bartlett's test for difference of variances) at each semilandmark separately. Because variation along the direction of the average curve has already been adjusted out
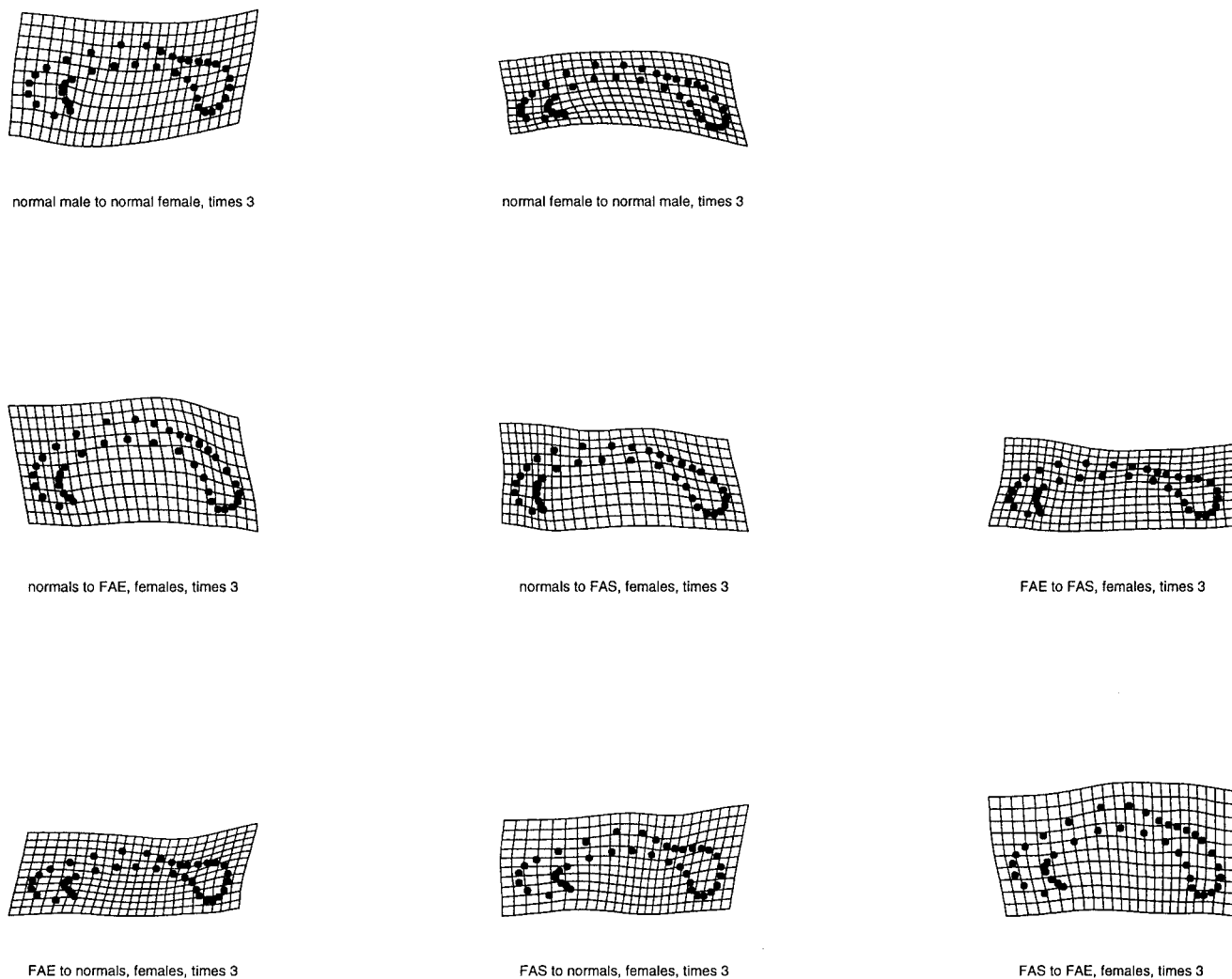
normal male to normal female, times 3

normal female to normal male, times 3

normals to FAE, females, times 3

normals to FAS, females, times 3

FAE to FAS, females, times 3

FAE to normals, females, times 3

FAS to normals, females, times 3

FAS to FAE, females, times 3

Fig. 8. Thin-plate splines for comparison of normal males and females and also for all pairwise comparisons of the female averages from Fig. 7. Each difference is exaggerated threefold for legibility.

during the course of producing these Procrustes coordinates (see section, MR Images and Derived Data), the test is of variance in one direction only, perpendicular to the average curves of Figure 7. For at least one semilandmark in each sex, the significance of this variance discrepancy is $P < 0.05/39 \sim 0.001$, so that the set of multiple comparisons is significant according to the conventional (Bonferroni-corrected) inference as well. The arcs of greatest discrepancy are quite differently situated in the two sexes: for males, anterocaudal to isthmus; for females, in the arch.

Figure 11 focuses on the best three of these points, for each sex, to illustrate what the F-test is detecting. Point by point, the normals' semilandmark coordinates concentrate themselves about the average outline at least as well as did the relative warp scores for the whole set of landmarks (Fig. 3). We thus harvest a whole new discriminator to augment the pair suggested in Figure 3. For the males, the point of greatest apparent information content, semilandmark 28, is concentrated indeed, and so we take its coordinate

perpendicular to the outline as the simplest useful scalar for adducing the hypervariance of the exposed. For the females, the semilandmarks atop the arch show the greatest variance ratios. A summary score that weights upper and lower arcs equally takes the average vertical coordinate of the set of all four.

Combining the shape spaces

Figure 12 combines the findings for landmarks and callosal outlines in one composite display. Separately by sex, the dimension of sharpest increase in variance from the landmark relative warp analysis (Fig. 4) is plotted against the shape coordinate of greatest hypervariance from the outline analysis (Fig. 11). Clearly, the two analyses, by landmark and by callosal outline, are complementary. If, as shown, boxes are taken tightly around the cores of the scatters for the normals, a classification rule emerges that calls subjects alcohol-affected just when they lie outside the boxes. The rule has putative sensitivity and specificity of 0.90 and 0.93 for males, 0.97 and 0.87 for females—enormously
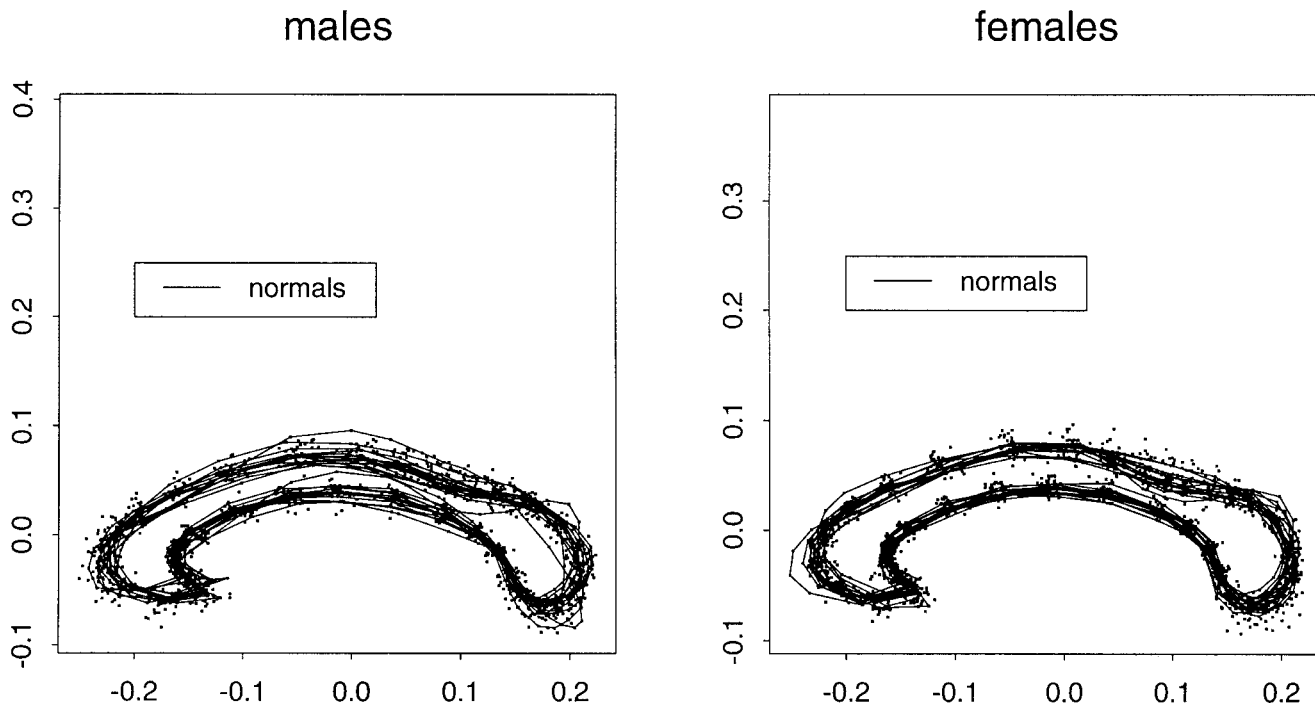
Fig. 9. Procrustes shape coordinates for the outlines shown in Fig. 7, separately by sex, with the outlines of the normal subjects connected. The variation of the exposed shapes outside the range of the normal is clear at several arcs around the circumference. The overall ratio of within-group variances, exposed vs. unexposed, is significant by the permutation test described in the text.

greater than anything hitherto reported from neuroanatomical data, with or without knowledge of size. These classifiers are uncorrelated; thus, each channel of measurement detects some exposed cases as hypervariant that are not detected by the other. But, of course, as the findings are not homologous between the sexes, each needs to be replicated in its own additional sample. Replications using 45 male and 45 female adolescents are in progress.

As the earliest FAS patients were often mentally retarded, it is important to think about IQ as a relevant dimension of these subject populations; nevertheless, because lowered IQ is itself a consequence of the brain damage that is our primary dependent variable, we did not impose any IQ selection criterion during the course of assembling our samples. Figure 13 annotates Figure 12 with the full-scale IQ scores of the alcohol-affected subgroup. The plot is entirely consistent with our emerging awareness that facial stigmata, intellectual deficit, and actual neuroanatomical abnormalities are fairly independent within the relatively high-performing end of the exposed range. There is no apparent pattern of full-scale IQ deficit by position in this plot in either sex, nor is the FAE subgroup "intermediate" in any useful sense between the normal subsample and the subjects diagnosed with the full FAS. As this is an important finding, albeit a negative one, we present it more explicitly in Figure 14, which makes explicit how IQ varies with "net severity" of brain damage, measured as distance from the average of the typical normal subgroups boxed in Figure 12. The line on each

graph is a standard scatterplot smoother applied to the exposed subsample only. Clearly, there is no tendency for IQ to decrease with distance from the neuroanatomically normal in this composite shape space. Within the exposed group there is also no association of these shape discriminators with centroid size.

SUMMARY OF THE FINDINGS

The preceding discussion was a lengthy recount of analyses that are likely to be relatively unfamiliar to the reader. It may be helpful to summarize these findings before discussing their implications. We analyzed variations of shape, by sex, among normal subjects and subjects diagnosed with an alcohol-related disorder, either FAS or FAE. Two data structures were extracted from the same MRI: one of 33-point landmark locations and the other of 40-point representations of a callosal midline curve. Our principal findings are the following:

1. For both sexes, and for both shape representations, the exposed group has distinctly more shape variability than the normal group.
2. In the landmark point data, the excess variability seems to involve the brain-stem-to-diencephalon axis.
3. In the callosal outline data, both sexes have more variance among the exposed than among the normals, but the site of particular hypervariation differs by sex. The additional variance seems concentrated under the isthmus in males, but in the height of the arch for females.
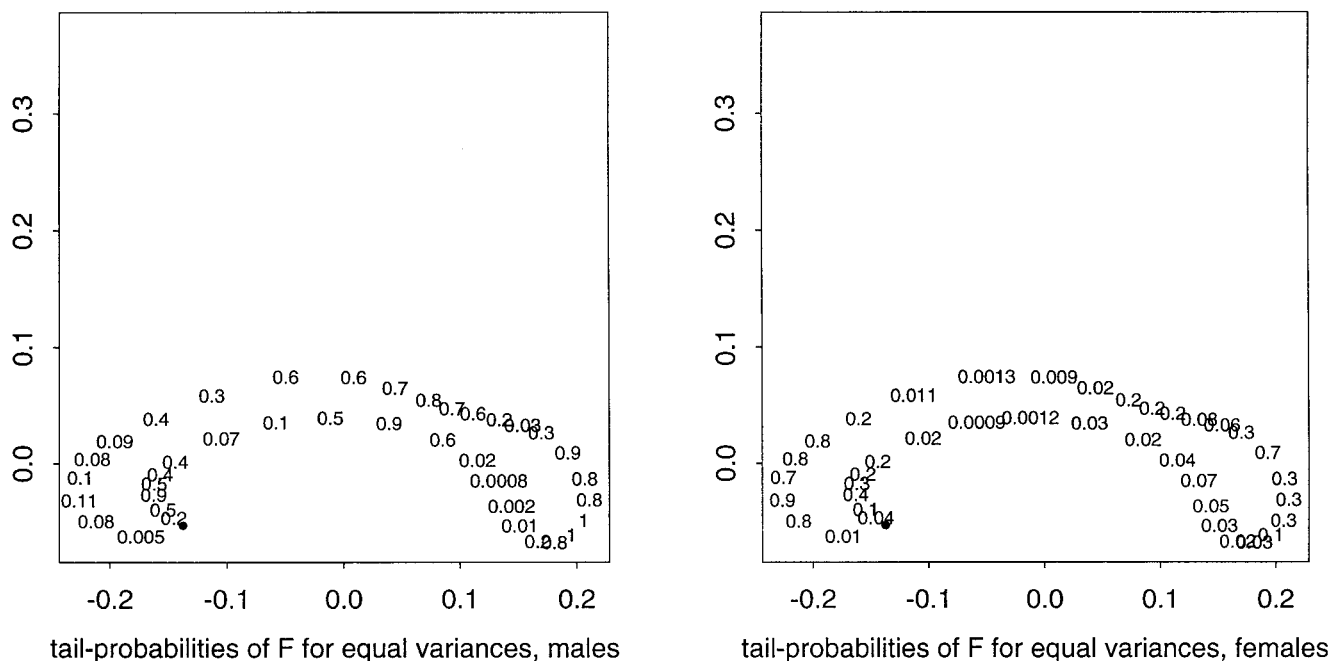
tail-probabilities of F for equal variances, males   tail-probabilities of F for equal variances, females

Fig. 10. *P*-values of F-tests for excess variance in the exposed subsample perpendicular to the average outline, semilandmark by semilandmark. For the males, the signal shown concentrates in the isthmus; for the females, in the arch.

4. The combination of these two localizations of variability supports a discrimination between the normals and the alcohol-affected with startlingly high sensitivity and specificity.

5. For both sexes, FAS and FAE each differ from the normal mean, but in neither sex do they differ much from each other. In particular, the discrimination of normals from exposed is not improved by knowledge of subdiagnosis.

6. Nor is the discrimination of normals from exposed improved by knowledge of IQ within this sample population.

DISCUSSION

Our alcohol-exposed subjects had been diagnosed before the onset of the study. Each was assigned either a diagnosis of FAS or what was then called FAE after examination by a dysmorphologist experienced in FAS. We combined these historical diagnostic records with data collected under a new methodology for quantifying brain differences between groups. Two types of data were involved: (1) landmark point configurations, restricted to brain regions where landmarks could easily be located (i.e., subcortical structures); and (2) closed curves in space, required for the analysis of the corpus callosum. The statistical analysis of their variability went forward separately, using closely related algebraic tools, and was then fused at the final stage of the analysis, the discrimination step.

The methodology weaving these data together involves three separate strategic decisions, each more or less unusual within our literature. The principal findings (as reviewed above) are expressed in terms of variances, not mean differences; they refer to shape variables instead of size variables; and the scientific goal they pursue is a discrimination, not a description. Interrelations among these thrusts are built into the methodology of shape coordinate analysis on which we have relied. For instance, the initial decision to examine shape variables more intensively than size variables made possible the important finding of the excess variability among the exposed—the shape features involved are not easily summarized in conventional measures of distance or area within parts of the form. In turn, the strength of the hypervariation underlies the surprisingly effective separation of affected subjects from normals conveyed in Figure 12, the sensitivity and specificity of which are so unexpected. Using either landmark point shape or outline shape, whichever is more practical for the tissue(s) at hand, the new toolkit supports studies of the affected parts of the brain and of the relationship among those parts at the same time and in the same computations. And however widely distributed the data in space, their statistical summary remains a unitary multivariate computation. We highlight the implications of these findings under five headings: the biotheoretical meaning of findings dealing with variability, the corpus callosum in ethanol teratogenesis, the literature of dysregulated brain stem/diencephalon relationships, the growing suspicion that FAS and FAE are the same clinical entities, and some methodological details.
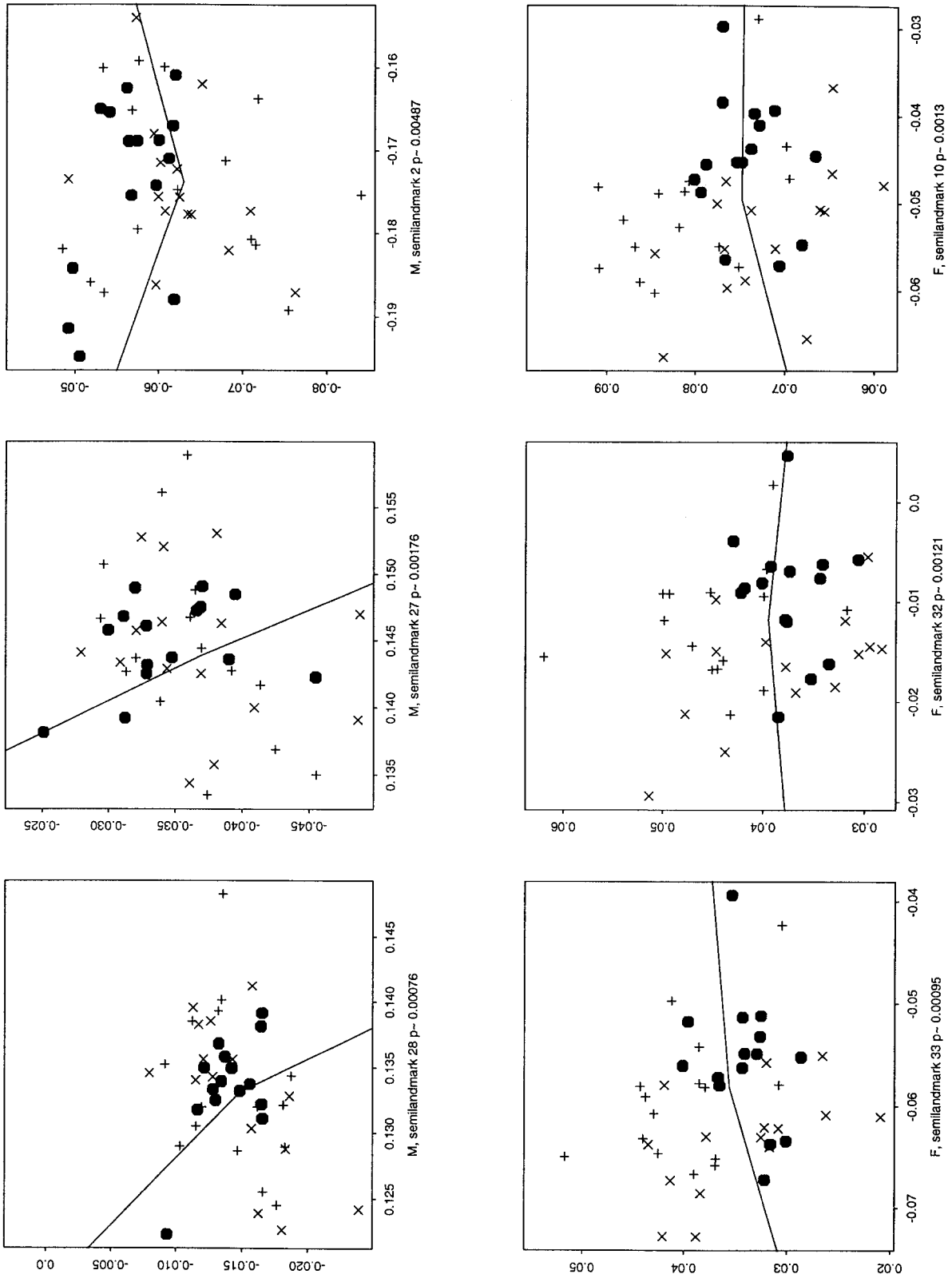
Fig. 11. Explicit scatters of the three most hypervariant semilandmarks, with diagnosis indicated. Legend: ●, normals; +, FAE; ×, FAS.
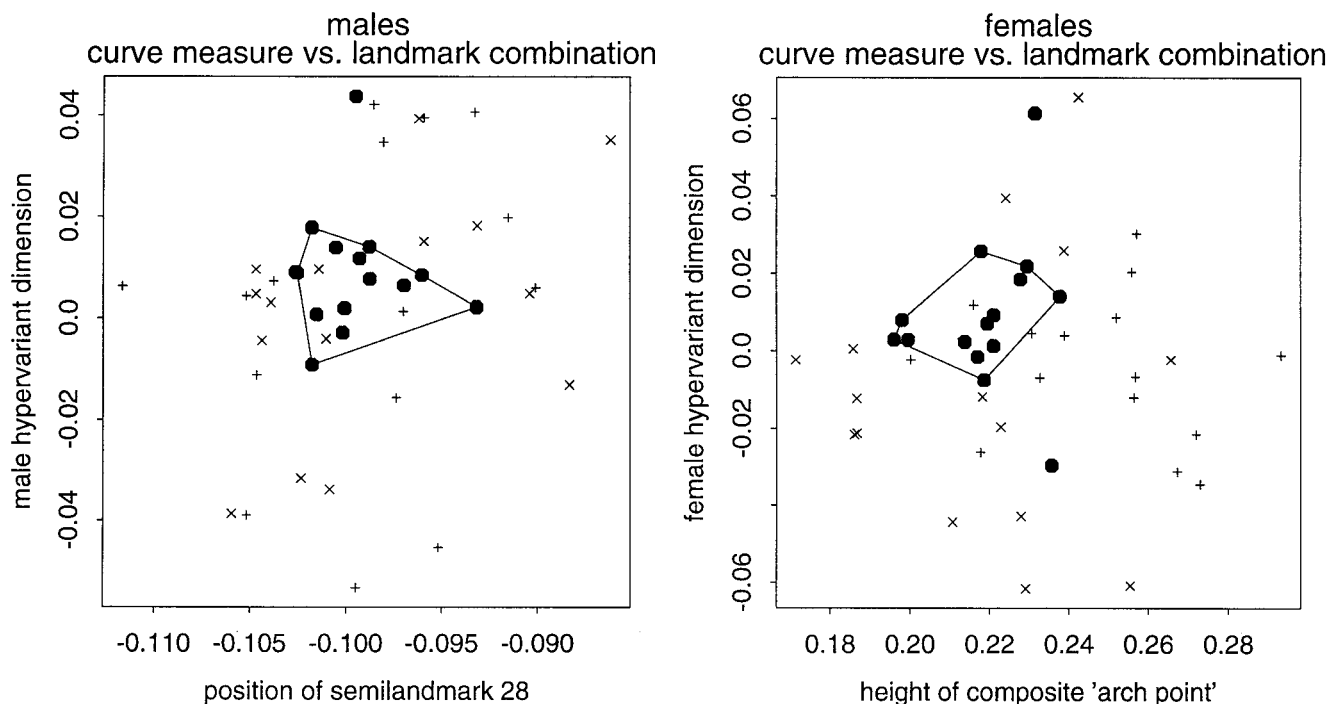
males
curve measure vs. landmark combination

females
curve measure vs. landmark combination

Fig. 12. Scatter of the most hypervariant landmark shape dimension against the selected hypervariant callosal shape coordinate, separately by sex. The great concentration of the normal subgroups (filled circles) is evident. +, FAE; ×, FAS.

Variability as a finding

It is not that the idea of considering differences of variance instead of differences in mean is unfamiliar to the quantitative biologist. It is present in most applied statistics texts under the heading of "verifying the assumptions" (of equal variances) required for conventional analyses of variance. In its essentials the comparison of two variances is a matter of the same F-ratios (variance ratios) that already underlie the usual tests for mean differences. What makes them unusually apposite for teratological applications such as these is their direct application to a specific high-dimensional vector of shape coordinates. In teratology, it is the malformations of shape, not those of size, that are inherently of greatest interest, and so a difference in variances need not be considered a nuisance interfering with inferences about means, but rather can be a powerful finding in its own right.

This finding is in accord with the now rather old literature concerning the origins of normal variation, the denominators of all the variance ratios we have been finding significant in the present study. For a classic overview of this topic, see Chapter 3, "Anatomical Variations—Significance," in Williams ('56). This chapter reviews the literature (through 1956) on quantitative and qualitative variations of structure in animals and humans, concluding, unsurprisingly, that great variation from one normal individual to another can be found in structures in all of the body's physiological systems, and that, in particular, "the brain is extremely variable in every character that has been

subject to measurement" (an observation the author attributes originally to Karl Lashley). Williams goes on: "Few studies are available concerning the structural variations in human brain tissue, considering the possible importance of their relation to behavior. Virtually nothing is known about disharmonies of development in the central nervous system except for very gross deficiencies." We know of no review over the intervening 44 years that would substantially alter this summary; the last monograph on methods for the multivariate study of such variations as might arise, Olson and Miller's *Morphological Integration* ('58), also dates from the 1950s. Regarding current methodological fashions, it is worth noting that the index of one popular current reference on psychopathology (Harris, '98) does not include any citations to the use of the term "variability" at all.

All the more startling, therefore, that our study of one abnormal developmental pathway, having the known cause of alcohol teratogenesis, has uncovered two distinct neuroanatomical configurations—the brain stem/diencephalon axis and the quantitative shape of the midline corpus callosum—for which the "disharmonies of development" prove so spatially specific. The finding is a variance ratio; thus, the signal it betokens might pertain either to its denominator (unusual invariance of form in the normally developing embryo) or to its numerator (a disregulation of the normal process in the same embryo that has been alcohol-affected). The former interpretation, the narrowing of range of the most crucial aspects of embryogen-
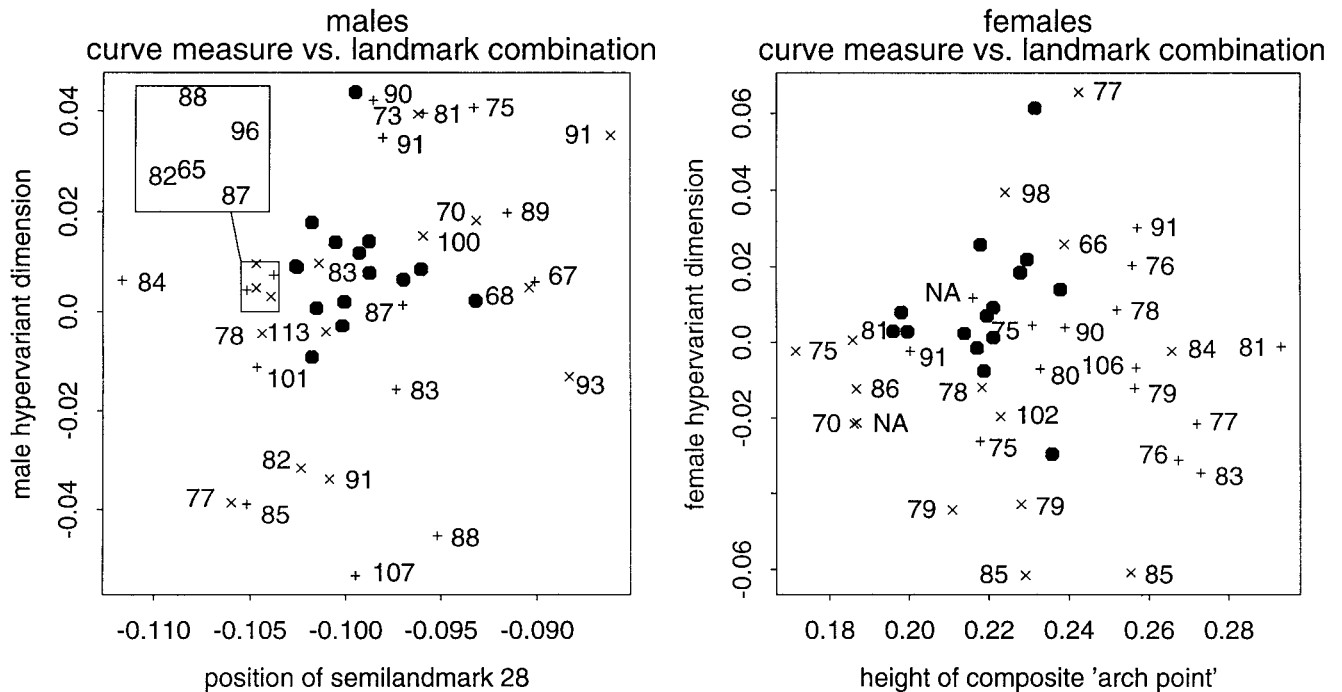
Fig. 13. Enhancement of Fig. 12 by full-scale IQ for the exposed subsamples. NA, score missing (unobtainable).

esis, is the more fruitful theoretically, as it corresponds to Waddington's classic notion of *developmental canalization,* whereby individuals attain a common endpoint of anatomical form despite variation in ontogenetic conditions. The callosum finding has precisely the logical structure of a disrupted canalization. Rhetorically, a "disruption" lacks the implication of a polarity that would otherwise be borne in words such as "deficit"; indeed, in the hypervariation characterizing the present findings, there is no particular direction in which the exposed group has been substantially shifted with respect to the controls—no "direction of deficit," such as would underlie a linear discriminant function; instead, there is a calibration of similarity versus dissimilarity to the normal, a contrast of typical with atypical in all directions of potential shape defect, not just one preferred direction.

The corpus callosum finding

Development of the human corpus callosum begins on about the 39th postconception day with differentiation of the commissural plate, and callosal fibers differentiate from that plate at about the 74th day, achieving adult morphology by day 115 (Loeser and Alvord, '68). It is therefore reasonable, that damage secondary to prenatal alcohol exposure would result principally from exposure within the first trimester of pregnancy, as is the case for the general "latent brain damage" observed by detailed measurement of behavior in large human samples (Streissguth et al., '93). Indeed, partial or total agenesis of the corpus callosum has been noted before in subjects diagnosed with FAS

(Riley et al., '95; Swayze et al., '97)—this observation was the reason we chose to measure callosal outline for the purposes of the present study. We had no cases of diagnosable agenesis in this adult sample. If agenesis itself is a direct consequence of exposure, it is likely to be engendered only at levels of exposure that are higher than those that typify the subjects of the present study (who, the reader will recall, had to be capable of undergoing a 5-hr battery of neurobehavioral tests in addition to the MR acquisition generating the data analyzed in the present study). The callosum typically develops in a rostrocaudal direction (Schaefer et al., '90), and partial agenesis is usually observed posteriorly, but we do not see any concentration of size difference or of variance differences at either end of the callosum (cf. Fig. 7).

Inasmuch as all the patients in this study were diagnosed with alcohol damage, the data set affords no contrasts speaking to any specificity of the findings. Indeed, more serious prenatal insults, such as spina bifida, can entail partial callosal agenesis at considerably greater rates than are found in fetal alcohol populations. But the literature suggests no reason to suspect hypervariance of callosal form in any milder syndrome. We would welcome comparative studies on this theme, especially in other patient groups characterized by attention disorders (cf. Banich, '98).

Although Riley et al. ('95) make no explicit reference to callosal shape variability per se, their Figure 4 includes standard deviations of "proportional area" (areas of five sectors of the callosal outline as a percentage of their total), and thus lets us interpret their data set
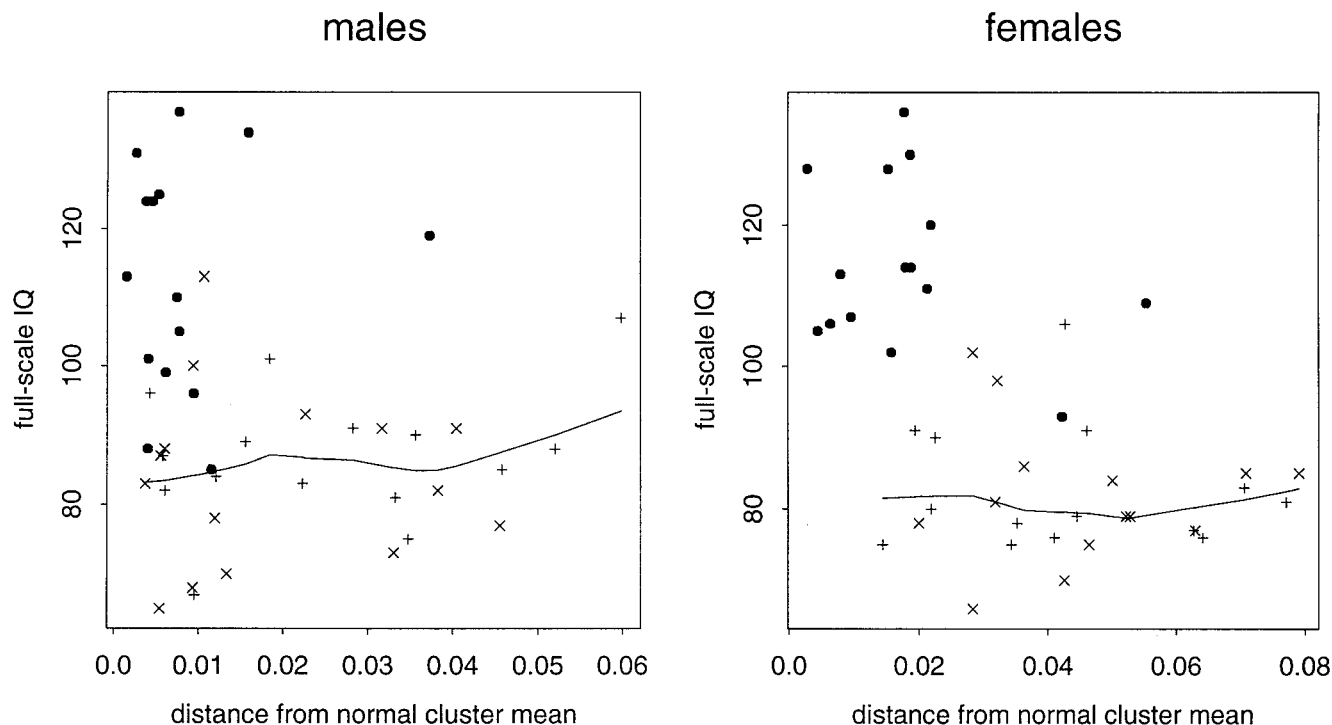
Fig. 14. Within this sample of exposed subjects, none profoundly retarded, atypicality of neuroanatomical form greatly distinguishes them from the normal subgroup but nevertheless is not associated with either the facial features of the full FAS syndrome or the measured full-scale IQ. Broken line, lowess scatterplot smoother (for the exposed subsample only). ●, normals; +, FAE; ×, FAS.

in the light of our present findings. The coefficient of variation of all five of the proportional areas is obviously significantly larger in their exposed subgroup than their unexposed subgroup, "preconfirming" our analysis. Apparently more discrepant is one recent study reporting an increase in callosal size consequent to exposure. Using a binge model in macaques, Miller et al. ('99) found that the rostral half (only) of the exposed callosa increased in size by comparison with unexposed controls. Operationally, the separation between rostral and caudal halves of the callosum is at the divergence of the fornix. This is perhaps an unfortunate strategy for group comparisons in which that point itself may have shifted along the callosal outline. In our human sample, hypervariance is concentrated at that point, especially in females (Fig. 4). Note also that, according to Table 3 of their publication, callosal size is clearly hypervariable in the exposed groups. Hence the Miller finding, based as it is on a total of only 15 animals over three groups, could well be an expression of hypervariance rather than the mean shift reported by those authors.

The brain stem/diencephalon finding

Brain stem anomalies have been found in a number of developmental syndromes, often teratogenetic. In one autopsy of an autistic patient, a shortening of the brain stem was noted, an outcome arising in animal models including Hoxa-I gene knockout and exposure to antimitotics or to thalidomide (Rodier et al., '96, '97).

A mouse model of holoprosencephaly (Lanoue et al., '97) shows abnormalities of mid- and hindbrain structure. Retinoic acid early in pregnancy in mice produces Arnold-Chiari malformations (Alles and Sulik, '92), including herniation of the hindbrain, owing to the primary damage to the neural crest and rhombencephalon. Similar brain stem dysmorphology is often observed after alcohol exposure in laboratory animals (Maier et al., '99; Thomas et al., '96). Studies of prenatally alcohol-exposed rats have linked brain stem damage of this sort to some of the behavioral deficits that characterize FAS/FAE in humans: decreased brain stem weight with motor deficits (Thomas et al., '96), or brain stem damage with auditory processing (Church et al., '96)—for the human analogue, see Pettigrew and Hutchinson ('84). But these studies did not pursue the issue of geometric shape as it has been formalized in the analyses presented in this article.

FAS and FAE do not seem to differ neurologically

Studies of neurobehavioral teratology in animals have demonstrated repeatedly that both brain dysmorphogenesis and behavioral dysfunction occur in offspring prenatally exposed to alcohol even in the absence of dysmorphic facial features, limb anomalies, or growth deficiency (Goodlett and West, '91; Riley et al., '90; Means et al., '88). For more than a decade, first qualitative and, recently, quantitative evidence for the absence of clear behavioral distinctions between pa-

tients with FAS and those with FAE has been accruing from several research sites (Streissguth et al., '91; Kodituwakku et al., '95; Aronson, '97; Mattson et al., '97; '98; Mattson et al., '99; Autti-Rämö, '00). Previous efforts to detect brain anomalies in patients with FAS compared with FAE (Clark et al., '00; Mattson et al., '97) have been straitjacketed by their reliance on size and volume assessments, which are irrevocably confounded with IQ levels, and thus with diagnosis. (Alcohol-exposed children who are mentally retarded are far more likely to be diagnosed with FAS than are those who are not; see Sampson et al., '00). The sharp focus on shape variation here more effectively quantitates the intricacies of brain maldevelopment observed qualitatively by teratologists and embryologists for decades. That the diagnosis of FAS cannot be distinguished from that of FAE from solid brain MRI (as quantified in the way we have done in the present study) bears enormous implications for the clinical course of these patients, inasmuch as brain dysmorphology is the center of the prenatal teratogenetic effect of alcohol. Diagnoses that can take these into consideration are thus likely to improve the delivery of appropriate social services, as they will be more closely tailored to the actual neuroteratological basis for prognosis in this class of patients.

Morphometric data

Of the two morphometric data channels combined in this article, that of landmark point locations is the less unfamiliar. The general issue of homology among points of the cortex proper (e.g., points on gyri or sulci, points on the gray–white boundary) is currently the subject of intense debate among many research groups (see, e.g., Toga, '99). Thus far there is no methodology for judging the comparative merits or demerits of suggestions from this class: for instance, there is not yet any general agreement that Talairach's ('88) original clinical suggestion of an AC–PC registration is demonstrably wrong on scientific grounds, let alone agreement on the reasons why it should be considered wrong (cf. Toga, '99). The registration here, using the machinery of Procrustes shape coordinates, arises from a justification that is mathematical, not necessarily empirical: it is the optimal way of visualizing shape variations of landmark configurations in general. In empirical applications, analyses that assign correspondences to points stand or fall on the covariations of the ensuing shapes or sizes with known causes or consequences of form, such as the fetal alcohol spectrum diagnoses in the present study.

In this connection, we had originally included, very tentatively, three other landmarks that did not make it into the final list of 33 (Table 2). The "bottom of the top of the cingulate sulcus"—a midline landmark locating the base of the cingulate sulcus where it turns laterally along the superior aspect of the brain—proved intolerably unreliable across raters. It is typical, we now believe, of landmark structures that are truly curves in space, to be represented as such for statistical purposes, but we restricted the present investigation to just one such curve, the callosum, because of the clear implication of prenatal alcohol in callosal anomalies already noted in the literature.

The posterior analogues of landmarks Fr–r and Fr–l, tips of the frontal horns of the ventricle, would be the two matching tips of the occipital horns. These could indeed be localized reliably on the individual subject, usually by following the narrow crevices of CSF posteriorly to some extent. However, the points thus arrived at appear to be bimodally distributed, with one tentative location about 2 cm posterior to the other. The appearance of the landmark in one or the other of these positions proved not to be associated with diagnosis, sex, or, in many cases, the contralateral position. In fact, this "tip" is a pair of landmarks, one of which was missing on each side for each subject. We omitted these landmarks because, taken individually, they were systematically missing in this way. Otherwise, the reader may have noted (by the absence of a footnote to the contrary in Table 2) that there are no missing landmark data across the full collection of 33 points here for 90 subjects. We would welcome the attempts of others to extend Table 2 to include cortical points that could be characterized on the same principles.

The corpus callosum outline data set here, which follows the midline of that structure as it gently twists in the vicinity of the putative midsagittal plane, is not itself a plane curve. As compared with the more usual technique of callosal visualization, which extracts an outline from one single image plane, the twisting leaves measures of area, both total and sectoral, relatively unaltered, but greatly affects the assessment of variability of actual locations at small scale. As we have seen, the areal finding is weaker than that for curving outline shape as regards mean differences by diagnosis, and offers no equivalent of the techniques for localizing shape hypervariability that constitute the principal signal we have found. (For more on contemporary strategies for callosal morphometrics, see Bookstein, '00.)

This Appendix expands on the terse summary of multivariate morphometric tools sketched under the Morphometric Methods section. Four main notions are explained in detail: Procrustes shape distance, Procrustes shape averages, Procrustes shape coordinates, and relative warps. The presentation is limited to the simplest case, landmark points in two dimensions, but the extensions to 3D data and to semilandmarks (points on curves) are straightforward. Sources that treat all this material at greater length include Bookstein ('96, '97a, '98) and Dryden and Mardia ('98). These particular characterizations of the main technical terms were worked out during the mid-1990s. A useful on-line glossary (Slice et al., '95) expands on their interrelationships, which represent a consensus among a community of several toolbuilders, along with a variety of other terms central to earlier applications of these methods.

Procrustes shape distance

To carry out multivariate analysis of the "shape" of a data set of k-landmark point configurations, it is sufficient to have a distance measure between the two shapes that obeys the usual rules. Suppose we have two landmark configurations, that is to say, two sets X_1, X_2 of k points with coordinates (x_{1i}, y_{1i}), $i = 1, \ldots, k$, for the first form and (x_{2i}, y_{2i}) for the second. If we were talking about "location" rather than shape, a reasonable notion of squared distance between the two would just be the usual Pythagorean sum

$$\sum_{i=1}^{k} [(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2]$$

of all squared coordinate differences between the positions of corresponding landmarks in the two forms.

We need to adapt this formula so that it gives the same answer whenever either of the two shapes is moved, rotated, or rescaled: then it will be talking about shape, as we want it to, rather than merely about locations in the original digitizing planes. It turns out best if we reformulate the problem in a way that turns out to reduce to just this Euclidean formula under certain conditions. We circumvent the problem of change of position by not allowing position to vary. Each form X_1 or X_2 has to be put down with its coordinates centered at $(0, 0)$—that is, we subtract $\sum_{i=1}^{k} x_{1i}/k$ from each x_{1i}, and similarly for y_{1i}, x_{2i}, and y_{2i}. Geometrically, the effect of this is just a shift of the origin of coordinates of each form, leaving its shape, as well as its size, alone. Also, we circumvent change of scale by likewise not allowing scale to vary: replacing each set of centered coordinates with a new set chosen so that the sum of their squared distances from the origin $(0, 0)$ of coordinates, which is now also their centroid, is exactly 1. We do this by dividing each centered form by a suitable scale factor, namely, the square root of whatever that sum of squares was before this operation. This factor, called centroid size, is examined by group in Table 3 (see text).

With position and scale both standardized, that leaves rotation. Just as we have repositioned each form independently, and rescaled each form independently, we could imagine having to rotate each form independently to some arbitrary horizontal or vertical. But that is not the way the method actually goes. It turns out a much better idea to state instead that we would like the result of adjusting rotation to give us back the Euclidean sum of squares corresponding to the rotated points. When two sets of points are both rotated by the same angle, the sum of squared distances doesn't change at all, and so what is needed is the relative rotation of one form with respect to the other that leads to the least Euclidean sum of squares out of all possible relative rotations. One version of Procrustes shape distance is then defined as the square root of this sum of squares when the relative rotation is chosen appropriately. Other versions of the definition differ from this one only by inconsequential adjustments, like using the arcsine of a small number in place of the number.

The steps in this computation follow down the rows of Figure A1. At the top are two quadrilaterals of landmarks presumed to arise from real images. Erase whatever outline information goes with these landmarks, but treat them purely as configurations of disconnected points. Then connect each landmark to the centroid of its own form. Its centroid size is the square root of the sum of squares of these lengths. For each form, rescale the sum of squares of the distances shown to unity (second row) by dividing by this centroid size. Next (third row, left) translate one of the forms so that its centroid directly overlies the centroid of the other form. Finally, identify the rotation (third row right) that minimizes the sum of squares of the residual distances between matched landmarks. The squared Procrustes distance between the forms (fourth row, redrawn with their own outlines back in) is (to a very good approximation) the sum of squares of those residuals at this minimum: total area of the circles at lower right, divided by π.

Procrustes average shape

To this point, there is now a formalism for computing a shape distance measure between any two landmark configurations, but not, as yet, any way to add or subtract them, so that we can't define the average of a shape data set as its sum divided by its count, the way we do for vectors—at least, not yet. Turn instead to a different characterization of the same idea: the average shape as the shape about which the specimens of a data set have the least sum of squared Procrustes distances. Although it may seem to you that this definition is somehow both vacuous and circular, in fact it is perfectly rigorous mathematically. Furthermore, for any landmark data set you will ever encounter in practice, the (unique) average shape can be computed by the
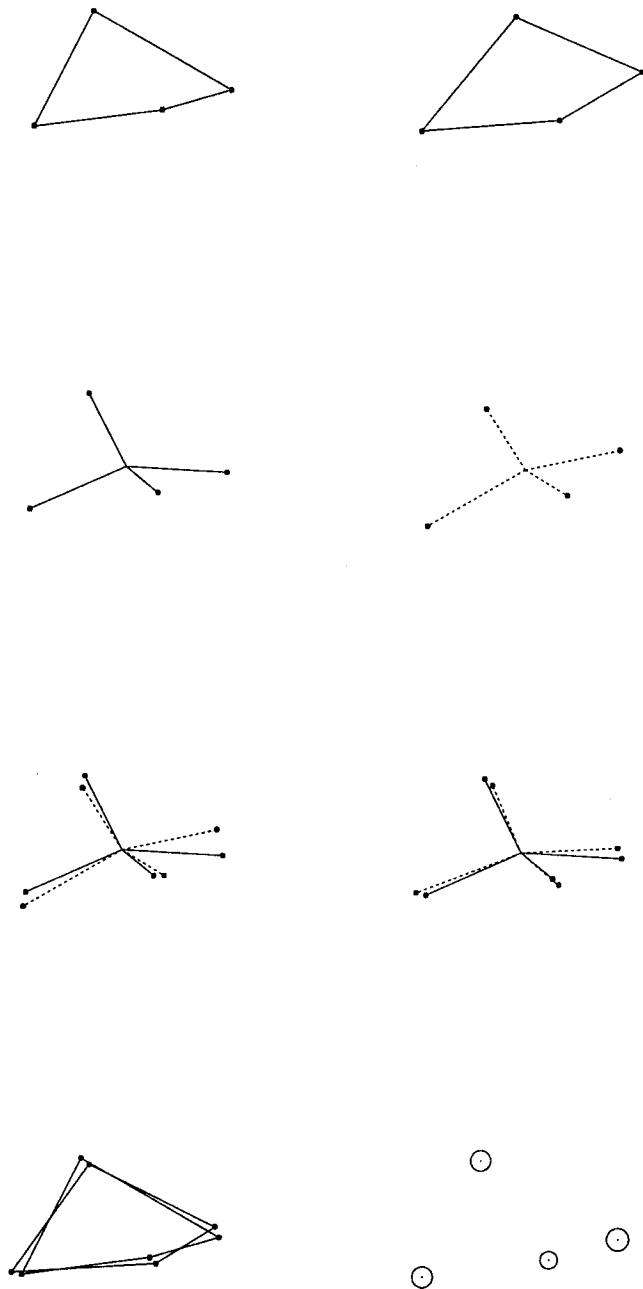
Fig. A1. The Procrustes superposition for a pair of forms. (**top row**) Two forms of four homologous landmarks. (**second row**) Each form is rescaled so that the sum of squares of the distances to the centroid of its four landmarks is 1. This is the sum of squares of the four lines shown. (**third row**) The centroids are superposed, and then one form is rotated over the other so that the sum of squared distances between corresponding landmarks is a minimum. (**fourth row**) With the construction lines erased, the squared Procrustes distance between the pair of forms is that sum of squared distances. It is proportional to the total area of the circles drawn at lower right.

iterative algorithm sketched in Figure A2 for a sample of four four-landmark shapes.

The top row displays some "raw data": four quadrilaterals of similar but not identical shapes. Guess at the value of the average shape—not a wild guess, but something in the vicinity of the real data: for instance,

guess that the average shape is the same as the shape of the first specimen. Then (second row) superimpose each of the original forms (including this form) over this guessed average in exactly the posture implied by Figure A1—the position in which that Euclidean sum of squared distances is smallest.

The next guess for the average shape is made up of the averages of the positions of the landmarks, one by one, after they are superimposed over the previous guess in this way. The third row of Figure A2 shows what happens when you update the candidate for average shape and go through the cycle of fits again: you get the same average, to the accuracy of the printer's dots in Figure A2. In nearly every real data set, this straightforward iterative algorithm converges to adequate precision by the end of the second iteration. To keep the figures from getting smaller row after row, it is also customary to rescale each candidate average to centroid size 1 before diving back into the second row, the refitting procedure.

Procrustes shape coordinates

The last panel in Figure A2 shows two concepts: the "Procrustes average shape" we sought, and copies of all the original forms scattered around it, each Procrustes-fitted to their common average. This composite image is the crux of the value of the whole Procrustes toolkit. By definition of Procrustes distance, the sum of squared distances of the shape coordinates of each original shape from those of the Procrustes average shape is its squared Procrustes distance from that average. But also, the sum of squared distances between the final positions of the landmarks of any two of the original forms is also *their* squared Procrustes distance.

For Figure A3, erase everything except the little scatters around the average at lower right in Figure A2, and put a separate coordinate system down centered at each averaged landmark position in turn. We thereby arrive at an exact analogue of one of the two great customary ways of setting up a multivariate statistical analysis, the approach usually called principal coordinates analysis, beginning with sums of squared distances instead of values of variables. Readers familiar with factor analysis will recognize this under the name of R-mode analysis. Beginning with distances, we have arrived (in an essentially unique way) at an equivalent set of $2k$ ordinary variables.

These are the Procrustes shape coordinates, which represent all the information in the shape of the original sets of landmarks for any linear multivariate statistical purpose. Any question about the correlation of shape with its causes or effects can be answered by using this single set of coordinates as a "vector of shape variables" in the corresponding standard multivariate procedure. For instance, to talk about averages of these coordinates by subgroups of the sample (e.g., the three diagnostic groups of this paper), it is sufficient to average their Procrustes shape coordinates, which produced the locations plotted in Figures 2 and 7. Our permutation tests for significance of these differences
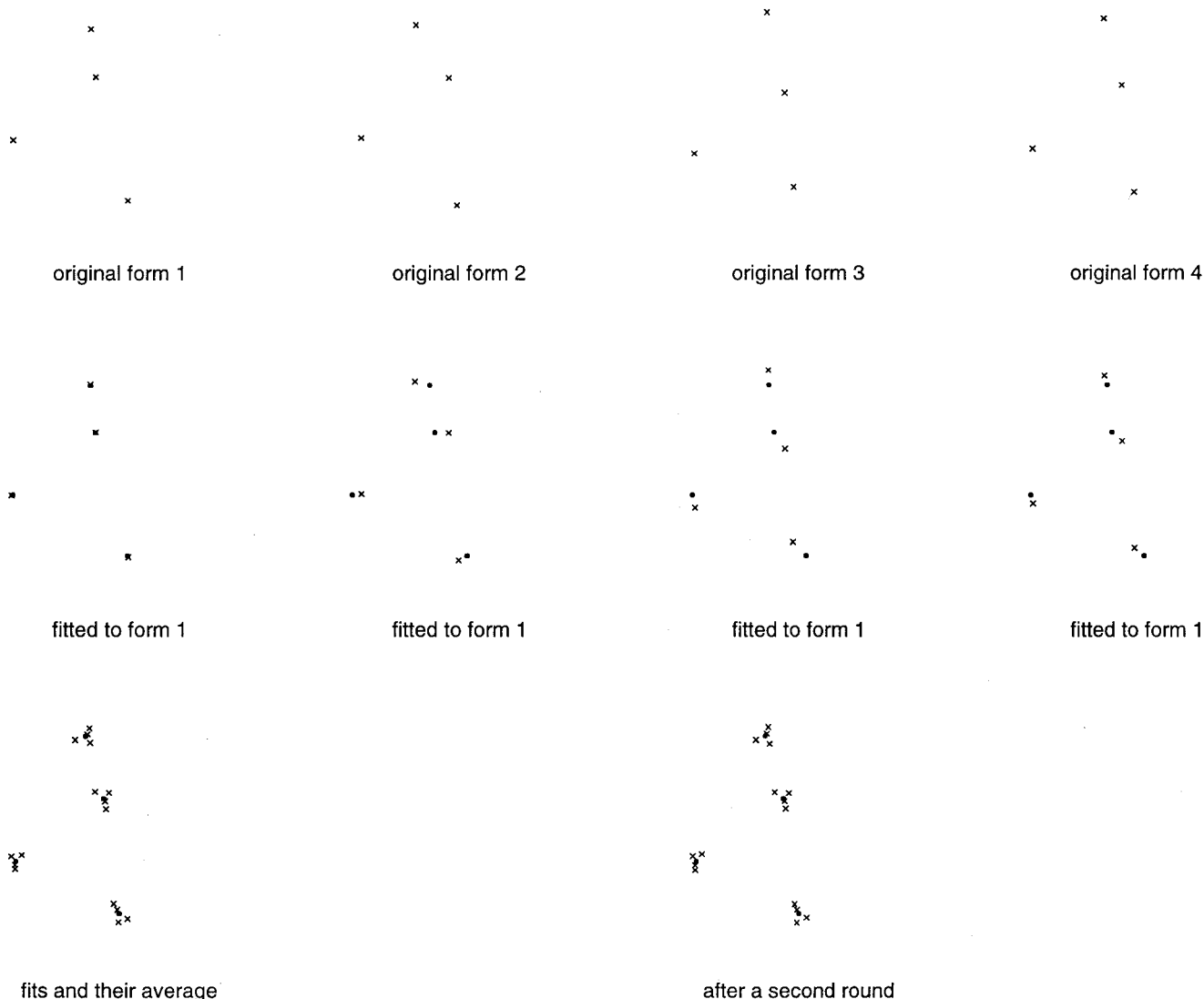
original form 1   original form 2   original form 3   original form 4

fitted to form 1   fitted to form 1   fitted to form 1   fitted to form 1

fits and their average   after a second round

Fig. A2. Procrustes averaging and Procrustes shape coordinates. **Top row:** four forms of four landmarks. **Middle row:** Procrustes fit of each (X's) to an arbitrary starting guess (dots: the first form). **Bottom left:** the next estimate of the average (dots) is the average of the fitted locations from the previous step. **Bottom right:** a second round of fits and averages changes it hardly at all—the algorithm seems to have converged already.

computed exactly analogous averages over pseudogroups rather than real groups. Correlations of shape with its causes or consequences, such as IQ, likewise proceed coordinate by coordinate in this representation, and are tested by permutation procedures using explained and unexplained squared Procrustes distances just as for variables arrived at by direct observation in the ordinary way.

Relative warps

Finally, a useful factor analysis of shapes is one version of ordinary principal components of these same Procrustes shape coordinates. Instead of starting from their correlation matrix, the usual procedure in most branches of applied biometrics, one works with their covariance matrix (an option available under the name "unscaled" in most packages). The reason for this is, at root, a magnificent mathematical elegance underlying the entire Procrustes toolkit. When data arise on a model of wholly random digitizing error around the same true landmark locations—digitizing error that is the same at every landmark and in every direction, what is called isotropic noise—the resulting distribution of Procrustes shape coordinates has extraordinary symmetries, regardless of what that average form was. Specifically, under that strong isotropic assumption (which is not far from applying to many real data sets once systematic factors of form are regressed out), the theoretical distribution of Procrustes shape coordinates necessarily has $2k - 4$ dimensions of exactly the same variance, and a final 4 dimensions of no variance at all, regardless of the average shape.
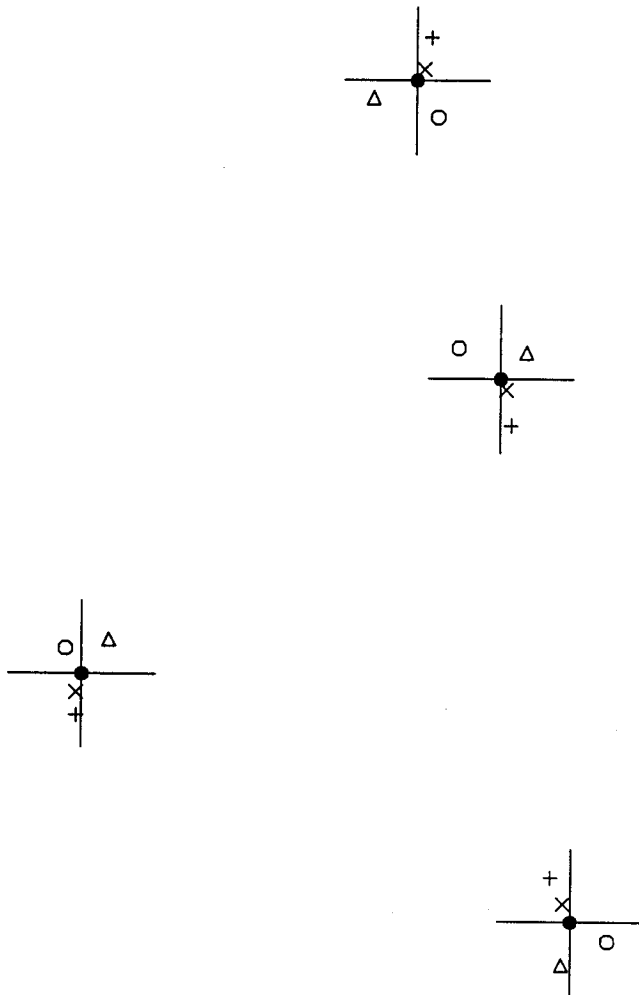
Fig. A3. Procrustes shape coordinates (now indicated in different symbols for each form) are the deviations of fitted landmarks from the average shape (dots) at the convergence of the algorithm in Fig. A2. There are a total of eight of these coordinates: two each for the four landmarks of the data set. The sum of squared differences between the values of the coordinates in any pair of specimens equals the squared Procrustes shape distance between the two specimens as originally digitized; that is, these coordinates serve as a set of principal coordinates for the Procrustes shape distance data (see text).

Because this is important, we say it another way as well. No matter what the average form looks like, if data arise from it by uninformative noise, the probability distribution of all those Procrustes shape coordinates is pretty much proportional to e^{-cPD^2}, where c is a suitable precision-like constant that takes into account the centroid size of the "true picture" as well as the amplitude of digitizing noise, and PD^2 is the squared Procrustes distance of any digitized form from the true average. If a spherical covariance matrix stands for no information, one that is not spherical stands for exactly the kind of information at which a principal components analysis is aimed. Principal components of Procrustes shape coordinates (under the covariance-matrix option) represent precisely the dimensions of shape variability that have the highest variance "per unit Procrustes distance" just as principal components of ordinary lists of variables have the highest variance "for unit sum of squared coefficients," and those dimensions of extra variance help us ordinate data distributions with the greatest efficiency just as do scatterplots of factor scores in most other applications.

You are probably used to seeing such components emerge from packages only in tabular form—columns of coefficients, one for each component, headed by its eigenvalue ("explained variance"). For shape coordinates, the corresponding tables are immediately converted to geometric diagrams showing how the points move away from the average shape, landmark by landmark, in strongly or weakly correlated ways (depending on the magnitude of the analogous eigenvalues). Specialized principal components of this sort, restricted to the covariance matrix of shape coordinates, are called relative warps because these displacements are usually drawn out in turn by images of deformed (warped) Cartesian grids. Each of these graphical styles is exploited in Figure 5.

LITERATURE CITED

Aase JM. 1990. Diagnostic dysmorphology. New York: Plenum Press.

Alles AJ, Sulik KK. 1992. Pathogenesis of retinoid-induced hindbrain malformations in an experimental model. Clin Dysmorphol 1:187–200.

Aronson M. 1997. Children of alcoholic mothers: results from Göteborg, Sweden. In: Streissguth AP, Kanter J, editors. The challenge of fetal alcohol syndrome: Overcoming secondary disabilities. Seattle, WA: University of Washington Press. pp. 15–24.

Astley SJ, Clarren SK. 2000. Diagnosing the full spectrum of fetal alcohol exposed individuals: introducing the 4-Digit Diagnostic Code. Alcohol Alcohol 35:400–410.

Autti-Rämö I. 2000. Twelve-year follow-up of children exposed to alcohol in utero. Dev Med Child Neurol 42:406–411.

Banich MT. 1998. The missing link: The role of interhemispheric interaction in attentional processing. Brain Cognition 36:128–157.

Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. New York: Cambridge University Press. 455 p.

Bookstein FL. 1996. Biometrics, biomathematics, and the morphometric synthesis. Bull Math Biol 58:313–365.

Bookstein FL. 1997a. Shape and the information in medical images: a decade of the morphometric synthesis. Computer Vision Image Understand 66:97–118.

Bookstein FL. 1997b. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med Image Anal 1:225–243.

Bookstein FL. 1998. A hundred years of morphometrics. Acta Zool Hung 44:7–59.

Bookstein FL. 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In: A. Toga, editor. Brain warping. San Diego, CA: Academic Press. pp. 157–181.

Bookstein FL. 2000. Morphometrics for callosal shape studies. In: Zaidel E, Iacoboni M, editors. The parallel brain: cognitive neuroscience of the corpus callosum. Cambridge, MA: MIT Press. In press.

Bookstein FL, Green WDK. 1998. Edgewarp 3D: A preliminary manual. Posted to the internet as ftp://brainmap.med.umich.edu/pub/edgewarp3.1/manual.html.

Church MW, Abel EL, Kaltenbach JA, Overbeck GW. 1996. Effects of prenatal alcohol exposure and aging on auditory function in the rat: preliminary results. Alcoholism 20:172–179.

Clark CM, Li D, Conry J, Conry R, Loock C. 2000. Structural and functional brain integrity of Fetal Alcohol Syndrome in nonretarded cases. Pediatrics 105:1096–1099.

Clarren SK, Smith DW. 1978. The fetal alcohol syndrome. N Engl J Med 198:1063–1067.

Dehaene P, Titran M, Samaille-Villette C, Samaille P, Crépin G, Delahousse G, Walbaum R, Fasquelle P. 1977. Fréquence du syndrome d'alcoolisme foetal. Nouv Presse Med 6:1763.

Dehaene P, Walbaum R, Titran M, Samaille-Villette Ch, Samaille PP, Crépin G, Delahousse G, Decocq J, Delcroix M, Caquant F, Querleu D. 1977. La déscendance des mères alcooliques chroniques: À propos de 16 cas d'alcoolisme foetal. Rev Fr Gynecol Obstet 72:491–498.

Dryden IL, Mardia KV. 1998. Statistical shape analysis. New York: John Wiley & Sons.

Evans AC, Dai W, Collins L, Neelin P, Marrett T. 1991. Warping of a computerized 3-D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. SPIE Proc 1445: 236–247.

Good PI. 1994. Permutation tests. New York: Springer-Verlag.

Goodlett CR, West JR. 1992. Fetal alcohol effects: rat model of alcohol exposure during the brain growth spurt. In: Zagon IS, Slotkin TA, editors. Maternal substance abuse and the developing nervous system. San Diego, CA: Academic Press. pp. 45–75.

Guerri C. 1998. Neuroanatomical and neurophysiological mechanisms involved in central nervous system dysfunctions induced by prenatal alcohol exposure. Alcoholism 22:304–312.

Hanson JW, Streissguth AP, Smith DW. 1978. The effects of moderate alcohol consumption during pregnancy on fetal growth and morphogenesis. J Pediatr 92:457–460.

Harris JC. 1998. Developmental neuropsychiatry. New York: Oxford University Press.

Jones KL. 1986. Fetal alcohol syndrome. Pediatr Rev 8:122–126.

Jones KL, Smith DW. 1973. Recognition of the fetal alcohol syndrome in early infancy. Lancet 2:999–1001.

Jones KL, Smith DW. 1975. The fetal alcohol syndrome. Teratology 12:1–10.

Jones KL, Smith DW, Ulleland CN, Streissguth AP. 1973. Pattern of malformation in offspring of chronic alcoholic mothers. Lancet 1:1267–1271.

Kendall DG. 1984. Shape-manifolds, procrustean metrics, and complex projective spaces. Bull Lond Math Soc 16:81–121.

Kodituwakku PW, Handmaker NS, Cutler SK, Weathersby EK, Handmaker SD. 1995. Specific impairments in self-regulation in children exposed to alcohol prenatally. Alcoholism 19:1558–1564.

Landesman-Dwyer S, Keller LS, Streissguth AP. 1978. Naturalistic observations of newborns: effects of maternal alcohol intake. Alcoholism 2:171–177.

Lanoue L, Dehart DB, Hindsale ME, Maeda N, Tint GS, Sulik KK. 1997. Limb, genital, CNS, and facial malformations result from gene/environment-induced cholesterol deficiency: further evidence for a link to sonic hedgehog. Am J Med Genet 73:24–31.

Lemoine P, Harousseau H, Borteyru J-P, Menuet J-C. 1968. Les enfants de parents alcooliques: anomalies observées, à propos de 127 cas. Ouest Med (Paris) 21:476–482.

Loeser JD, Alvord EC. 1968. Agenesis of the corpus callosum. Brain 91:553–570.

Maier SE, Miller JA, West JR. 1999. Prenatal binge-like alcohol exposure in the rat results in region-specific deficits in brain growth. Neurotoxicol Teratol 21:285–291.

Majewski F, Bierich JR, Löser H, Michaelis R, Leiber B, Bettecken F. 1976. Zur Klinik und Pathogenese der Alkohol-Embryopathie: Bericht über 68 Fälle. Munch Med Wochenschr 118:1635–1642.

Martin JC, Martin DC, Lund CA, Streissguth AP. 1977. Maternal alcohol ingestion and cigarette smoking and their effects on newborn conditioning. Alcoholism 1:243–247.

Mattson SN, Riley EP, Sowell ER, Jernigan DF, Sobel DF, Jones KL. 1996. A decrease in size of the basal ganglia in children with fetal alcohol syndrome. Alcoholism 20:1088–1093.

Mattson SN, Riley EP, Gramling L, Delis DC, Jones KL. 1997. Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. J Pediatr 131:718–721.

Mattson SN, Riley EP, Gramling L, Delis DC, Jones KL. 1998. Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. Neuropsychology 12:146–153.

Mattson SN, Goodman AM, Canine C, Delis DC, Riley EP. 1999. Executive functioning in children with heavy prenatal alcohol exposure. Alcoholism 23:1808–1815.

Means LW, Burnette MA, Pennington SN. 1988. The effect of embryonic ethanol exposure on detour learning in the chick. Alcohol 5:305–308.

Miller MW, Astley SJ, Clarren SK. 1999. Number of axons in the corpus callosum of the mature macaca nemestrina: increases caused by prenatal exposure to alcohol. J Comp Neurol 412:123–131.

NIAAA (National Institute on Alcohol Abuse and Alcoholism). 1983. Fifth Special Report to the U.S. Congress on Alcohol and Health. U.S. Department of Health and Human Services, Public Health Service, ADAMHA (Alcohol, Drug Abuse, and Mental Health Administration), NIAAA.

Olegård R, Sabel KG, Aronsson M, Sandin B, Johansson PR, Carlsson C, Kyllerman M, Iversen K, Hrbek A. 1979. Effects on the child of alcohol abuse during pregnancy. Acta Paediatr Scand (Stock) 275(suppl):112–121.

Olson EC, Miller RL. 1958. Morphological integration. Chicago, IL: University of Chicago Press.

Ouellette EM, Rosett HL, Rosman NP, Weiner L. 1977. Adverse effects on offspring of maternal alcohol abuse during pregnancy. N Engl J Med 297:528–530.

Pettigrew AG, Hutchinson I. 1984. Effects of alcohol on functional development of the auditory pathway in the brainstem of infants and chick embryos. Ciba Found Symp 105:26–46.

Randall CL. 1977. Teratogenic effects of in utero ethanol exposure. In: Blum K, Bord D, Hamilton M, editors. Alcohol and opiates: neurochemical and behavioral mechanisms. New York: Academic Press. pp 91–107.

Riley EP. 1990. The long-term behavioral effects of prenatal alcohol exposure in rats. Alcoholism 14:670–673.

Riley EP, Vorhees CV. 1986. Handbook of behavioral teratology. New York: Plenum Press.

Riley EP, Mattson SN, Sowell ER, Jernigan TL, Sobel DF, Jones KL. 1995. Abnormalities of the corpus callosum in children prenatally exposed to alcohol. Alcoholism 19:1198–1202.

Rodier PM, Ingram JL, Tisdale B, Croog VJ. 1997. Linking etiologies in humans and animal models: studies of autism. Reprod Toxicol 11:417–422.

Rodier PM, Ingram JL, Tisdale B, Nelson S, Romano J. 1996. Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. J Comp Neurol 370:247–261.

Roebuck TM, Mattson SN, Riley EP. 1998. A review of the neuroanatomical findings in children with fetal alcohol syndrome or prenatal exposure to alcohol. Alcoholism 22:339–344.

Rouquette J. 1957. Influence de l'intoxication alcoolique parentele sur le développement physique et psychique des jeunes enfants. Thesis University of Paris, Paris, France.

Sampson PD, Streissguth A, Bookstein FL, Barr HM. 2000. On categorizations in analyses of alcohol teratogenesis. Environ Health Persp 108(suppl 3):421–428.

Sander LW, Snyder PA, Rosett HL, Lee A, Gould JB, Ouellette E. 1977. Effects of alcohol intake during pregnancy on newborn state regulation: a progress report. Alcoholism 1:233–241.

Schaefer GB, Shuman RM, Wilson DA, Saleeb S, Domek DB, Johnson SF, Bodensteiner JB. 1990. Partial agenesis of the anterior corpus callosum: correlation between appearance, imaging, and neuropathology. Pediatr Neurol 7:39–44.

Shaywitz SE, Cohen DJ, Shaywitz BA. 1980. Behavior and learning difficulties in children of normal intelligence born to alcoholic mothers. J Pediatr 96:978–982.

Slice D, Bookstein FL, Marcus LF, Rohlf FJ. 1995. A glossary for geometric morphometrics. Electronic publication: http://life.bio.sunysb.edu/morph/gloss1.html.

Smith DW. 1976, 1983. Recognizable patterns of human malformation; genetic, embryologic, and clinical aspects, 2nd, 3rd ed. Philadelphia: WB Saunders.

Stratton KR, Howe CJ, Battaglia FC, editors. 1996. Fetal alcohol syndrome: diagnosis, epidemiology, prevention, and treatment. Washington, DC: Institute of Medicine, National Academy Press.

Streissguth AP. 1978. Fetal alcohol syndrome: an epidemiologic perspective. Am J Epidemiol 107:467–478.

Streissguth AP, O'Malley K. 2000. Neuropsychiatric implications and long term consequences of fetal alcohol spectrum disorders. Semin Clin Neuropsychiatry 5:177–190.

Streissguth AP, Barr HM, Martin DC, Herman CS. 1980a. Effects of maternal alcohol, nicotine and caffeine use during pregnancy on infant mental and motor development at 8 months. Alcoholism 4:152–164.

Streissguth AP, Landesman-Dwyer S, Martin JC, Smith DW. 1980b. Teratogenic effects of alcohol in humans and laboratory animals. Science 209:353–361.

Streissguth AP, Aase JM, Clarren SK, Randels SP, LaDue RA, Smith DF. 1991. Fetal alcohol syndrome in adolescents and adults. JAMA 265:1961–1967.

Streissguth AP, Bookstein FL, Sampson PD, Barr HM. 1993. The enduring effects of prenatal alcohol exposure on child development, birth through seven years. Ann Arbor, MI: University of Michigan Press.

Streissguth AP, Barr HM, Kogan JA, Bookstein FL. 1996. Understanding the occurrence of secondary disabilities in clients with Fetal Alcohol Syndrome and Fetal Alcohol Effects. Final report, CDC grant CCR008515. Fetal Alcohol and Drug Unit, University of Washington School of Medicine.

Swayze VW II, Johnson VP, Hanson JW, Piven J, Sato Y, Giedd JN, Mosnik D, Andreasen NC. 1997. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. Pediatrics 99:232–240.

Talairach J, Tournoux P. 1988. Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging. Stuttgart: Thieme Verlag.

Thomas JD, Wasserman EA, West JR, Goodlett CR. 1996. Behavioral deficits induced by bingelike exposure to alcohol in neonatal rats: importance of developmental timing and number of episodes. Dev Psychobiol 29:433–452.

Toga AW, editor. Brain warping. New York: Academic Press.

West JR, editor. 1986. Alcohol and brain development. New York: Oxford University Press.

Williams RJ. 1956. Biochemical individuality. New York: John Wiley & Sons.

Wilson JG, Fraser FC, editors. 1977. Handbook of teratology. Vol. 1: General principles, etiology. New York: Plenum Press.