NeuroImage **15,** 233–251 (2002)
doi:10.1006/nimg.2001.0977, available online at http://www.idealibrary.com on IDEAL®

Corpus Callosum Shape and Neuropsychological Deficits in Adult Males with Heavy Fetal Alcohol Exposure

Fred L. Bookstein,*[,1] Ann P. Streissguth,† Paul D. Sampson,‡ Paul D. Connor,† and Helen M. Barr†

*Institute of Gerontology, University of Michigan, Ann Arbor, Michigan 48109; and †Department of Psychiatry and ‡Department of Statistics, University of Washington, Seattle, Washington 98105

Received April 8, 2001

Persons with brain damage consequent to prenatal alcohol exposure have typically been diagnosed with either fetal alcohol syndrome (FAS) or fetal alcohol effects (FAE), depending on facial features. There is great variability of behavioral deficits within these groups. We sought to combine neuroanatomical measures with neurocognitive and neuromotor measures in criteria of greater sensitivity over the variety of consequences of alcohol exposure. To this end, midline curves of the corpus callosum were carefully digitized in three dimensions from T1-weighted MR scans of 15 adult males diagnosed with FAS, 15 with FAE, and 15 who were unexposed and clinically normal. From 5 h of neuropsychological testing we extracted 260 scores and ratings pertaining to attention, memory, executive function, fine and gross motor performance, and intelligence. Callosal midline shape was analyzed by new morphometric methods, and the relation of shape to behavior by partial least squares. The FAS and FAE subgroups have strikingly more variability of callosal shape than our normal subjects. With the excess shape variation are associated two different profiles of behavioral deficit unrelated to full-scale IQ or to the FAS/FAE distinction within the exposed subgroup. A relatively thick callosum is associated with a pattern of deficit in executive function; one that is relatively thin, with a deficit in motor function. The two combine in a very promising bipolar discrimination of the exposed from the unexposed in this sample. Thus there is considerable information in callosal form for prognosis of neuropsychological deficits in this frequently encountered birth defect.

Key Words: **corpus callosum; fetal alcohol syndrome; fetal alcohol effects; partial least squares; procrustes shape coordinates; singular warps.**

INTRODUCTION

Alcohol is a teratogen over the course of prenatal human development, and the brain is the fetal organ most sensitive to the damage it induces. Yet contemporary protocols for diagnosis make no reference to brain images, prenatal or postnatal, in classifying persons as damaged or undamaged by the prenatal exposure. This paper demonstrates how an unusual style of morphometric analysis of the adult brain image, in combination with conventional observations of cognitive and motor performance, might lead to a considerable increase in the sensitivity with which alcohol effects can be diagnosed and to a potentially important discrimination among two new subtypes of that diagnosis.

A diagnosis of fetal alcohol syndrome (FAS) entails evidence of a compromised central nervous system (CNS), growth deficiency, and uniquely characteristic facial stigmata subsequent to heavy fetal alcohol exposure (Stratton *et al.*, 1996). The facial features include short palpebral fissures, flat philtrum, thin upper vermilion, and flat midface. Adults with this diagnosis typically have at least a moderate intellectual deficit and a highly variable assortment of deeper specific deficits in attention, memory, motor function, and judgment. The related diagnosis of "fetal alcohol effects" (FAE) has been applied to patients with CNS compromise and a history of exposure without the full set of physical findings. Evidence is accumulating that the sociobehavioral prognosis for patients diagnosed with FAE is as severe as that for FAS. In one large database (Streissguth *et al.*, 1996), the prevalence of "secondary disabilities" (like dropping out of school or going to jail) in patients diagnosed FAE (mean IQ, 90) is no less severe than in those diagnosed FAS (mean IQ, 79). When psychometric scores in the same study were adjusted for IQ, profiles of deficit in the two diagnostic groups were identical. Similar findings have been reported by others (Mattson *et al.*, 1997, 1998). Yet though the clinical behavioral sequelae of the two diagnoses appear indistinguishable, the FAS diagnosis

[1] To whom correspondence should be addressed at Institute of Gerontology, University of Michigan, 300 North Ingalls Building, Ann Arbor, MI 48109-2007. Fax: (734) 936-2116. E-mail: fred@brainmap.med.umich.edu.

1053-8119/02 $35.00

is more distinctive and less controversial than the diagnosis of FAE (Stratton *et al.,* 1996). We recently estimated population prevalence of FAE or alcohol-related neurodevelopmental disorder in Seattle for the cohort born in 1975 (Sampson *et al.,* 1997a) as twice that of FAS (about 6 per thousand, versus 3). It follows that the facial diagnosis of FAS, even though it has 100% specificity, has unacceptable sensitivity for the task of detecting the full range of alcohol teratogenesis.

As alcohol can affect brain development throughout gestation, whereas the period of facial vulnerability is relatively short (Sulik *et al.,* 1981), one might reasonably suspect that behavioral abnormalities would be related to structural brain characteristics more strongly than to facial anomalies. In that case, studies of brains might generate a prognostically useful differential diagnosis not afforded by facial measurements however conscientious. This paper attempts to lay the groundwork for a quantitative approach of far greater sensitivity, without loss of specificity, by combining laboratory tests of neuropsychological function with direct quantitative measurements of three-dimensional MR brain images. For this first analysis we assessed only one structure: the outline of the corpus callosum in its midline. Our interest arises from a consistent finding in the literature associating callosal dysplasia with alcohol exposure. In an early study of this kind (Mattson *et al.,* 1994), of four children diagnosed FAS/"PAE" (our FAE) three showed dysgenesis or agenesis of the corpus callosum. There are now at least 11 reported cases of callosal agenesis in FAS/FAE as viewed in MRI (Mattson and Riley, 1997a; Swayze *et al.,* 1997), and disproportionate reductions are seen in the mean callosa of exposed cases even when grossly normal (Riley *et al.,* 1995; Mattson and Riley, 1997b).

In a thoughtful recent discussion of the joint role of neuroanatomy (possibly quantitative) and neuropsychological testing in psychopathology, James Harris writes:

> The adult neuropsychological database was largely established through the evaluation of adults with known brain damage (strokes, seizures, tumors, penetrating wounds, head trauma). . . . [But lesions of this type] occur far less commonly in children where congenital malformations related to pre- and postnatal insults . . . are more common. A majority of the neuropsychological dysfunctions of early life are not then associated with known brain insults, nor are they associated with lesions demonstrable on routine neuroimaging studies. . . . Neuropsychological testing, [therefore, should] integrate psychiatric and psychological information on behavior and the mind with neurological information on the brain and nervous system. (Harris, 1998, vol. ii, p. 23, p. 20)

The damage caused by prenatal alcohol is best described as diffuse damage that has been hitherto associated with "lesions" in only a small fraction of cases. In this sense, although alcohol is known to be the teratogen in prenatal alcohol exposure, we do not know where the "insult" is. Except for the literature's hint about the corpus callosum, we would know neither where to look in the brain for signs of specific damage nor which behaviors to examine that would assess the function of those specifically damaged systems. In other words, neither approaches from neuroanatomy nor approaches from developmental neuropsychology, separately, promise to be particularly helpful at this time. Progress is likelier by the integration of the two information streams, the behavioral and the neurological.

The present study is intended as a prototype for such investigations, combining a neuroanatomical channel (the corpus callosum) and a neuropsychological channel (a wide variety of neuropsychological laboratory tasks known to be sensitive to prenatal alcohol exposure). Each set of measurements formally serves both to assess group differences in its own domain and to help focus our attention on salient aspects from the other domain. Specifically, we measured callosal outline and neuropsychological function in 45 adult males, 30 of whom were diagnosed as affected by prenatal alcohol exposure. Our hypothesis is that callosal shape will be differently distributed between unexposed and exposed in ways that correlate with clinically meaningful differences in behavior. In fact, the combined data set supports a remarkably specific association of callosal shape variation with a bipolar profile of behavioral anomaly for which the FAS–FAE distinction supplies no information.

The callosal outline data here were previously exploited by Bookstein *et al.* (2001) in combination with data from 33 landmark points in a sample including adult females as well as these males. The present paper extracts much more information from the same outline resource by combining it with neuropsychological measurements. In particular, the major finding of this paper, as conveyed in the final figures, involves the correlation of structural and behavioral information; it would not be accessible from neuroanatomical data alone.

METHODS

Sample

We studied 45 Seattle-area males aged 18 years and over: 15 unexposed normals and 30 with alcohol teratogenesis, collectively referred to here as "the exposed." Prior to enrollment in the Seattle FAS Followup Study database (Streissguth *et al.,* 1996, 1997), these men had been diagnosed as either FAS ($N = 15$) or FAE ($N = 15$). Diagnosis was by D. W. Smith or one of his fellows (usually S. K. Clarren) between 1972 and 1995. A diagnosis of FAS combines findings of maternal alcohol abuse, CNS manifestations, growth deficiency, and specific facial anomalies. Patients with evidence of the first two of these criteria but not all of the physical

findings fall into the FAE group. FAS and FAE diagnoses were similarly ascertained, and each required a maternal alcohol exposure history in accordance with the Institute of Medicine (IoM) description: "a pattern of excessive intake characterized by substantial regular intake, or heavy episodic drinking" (Stratton *et al.*, 1996, p. 6). Diagnostically, all would fall within the IoM's categories of FAS, partial FAS, or alcohol-related neurodevelopmental disorder (ARND). For the actual diagnostic criteria used, see Streissguth *et al.* (1991). Normals were recruited from employees and their children at local health care facilities and educational institutions. The subjects were approximately group-matched for age and ethnicity. All three groups averaged 23 years of age (range, 18–36). Subjects were white except for 4 blacks (2, 2, 0 by diagnostic group) and 13 Native Americans (3, 4, 6). Clinically, all MR images were grossly normal except that our neuroradiologist noted two occurrences of callosal hypoplasia (one an FAS patient, the other an FAE). Subjects were *not* group-matched for IQ, of course, as IQ deficits are entailed in the diagnosis of the alcohol damage itself. Mean IQ scores (normals, FAE, FAS) were 113, 87, and 84. Three patients (2 FAS, 1 FAE) had IQ scores under 70; none were lower than 65. A total of 9 subjects (1 normal, 3 FAE, 5 FAS) were left-handed (measured as in Reitan, 1974). It is not feasible to match for socioeconomic status (SES) in studies of adult patients with fetal alcohol damage, as one characteristic of these patients is the great number of living situations in which they reside over the course of development, likely spanning a wide range of SES levels (Streissguth *et al.*, 1996). We did, however, control the "E" of SES to some extent, by excluding anyone who had completed a college education from the pool of potential unexposed subjects.

MR Images and Derived Data

T1-weighted sagittal SPGR images were acquired over 12 min in a GE 1.5-T Signa scanner at the University of Washington. TE was 8 ms, TR 29 ms, flip angle 45°. The resulting $256^2 \times 124$ arrays of $0.86^2 \times 1.50$ mm^3 voxels were processed by Edgewarp 3D software (Bookstein and Green, 1998; Bookstein *et al.*, 2001). From the MR images were located four landmark points (anterior and posterior commissure, tip of fourth ventricle, and rostrum) along with 39-point semilandmark tracings (see below) of the callosal midline. Morphology was digitized (by one of the authors, F.L.B.) blind to diagnostic group and behavior. Starting at rostrum, callosa were traced as 40-point polygons in space. (That is, digitizing took place in 3D, not in any single plane, with all three coordinates varying together.) A typical set of digitized locations, projected onto a convenient parasagittal image plane, is on the left in Fig. 1. They are spaced roughly inverse to curvature on a reference form (the first one digitized). On the right is a typical digitizing scene: a section perpendicular to the actual callosal midline through a candidate point along the lower border of isthmus. Every point digitized lay precisely on the "vertical" (axis of symmetry) of an image like this at the apparent boundary between callosum and CSF. If the tissue adjacent to callosum was not CSF, perhaps fornix or cavum pellucidum, the point digitized was taken as the (visual) extrapolation of the callosum–CSF boundary on this same axis.

Morphometric Tools

Multivariate analysis of these data was by the newly standardized biometric approach to sets of labeled points in two or three dimensions (Dryden and Mardia, 1998; Bookstein, 1996a,b, 1997a, 1998, 1999). *Shape* is the information left in such a figure after we ignore location, orientation, and scale. For multivariate analysis of shape and its correlates, one uses *Procrustes analysis,* which can be thought of in the following four steps. First, for any single specimen's digitized points (x_i, y_i), $i = 1, \ldots, k$, center them by moving the average location or centroid $(\sum_{i=1}^{k} x_i, \sum_{i=1}^{k} y_i)/k$ to (0, 0), and then, after the specimen has been centered on (0, 0) in this way, divide out the scale factor called *Centroid Size,* square root of the sum of squared distances of the points from (0, 0). After this scaling, the sum of those squared distances is precisely 1.0. Second, for any two sets of points with the same labels that have been centered and scaled in this way, define the *Procrustes distance* between them as the square root of the sum of squared Euclidean distances between corresponding points when one form is rotated with respect to the other just so as to minimize this sum of squared distances. Third, from two or more of these sets of centered and scaled points, compute the *Procrustes average shape* of the sample: the set of points that has the least summed squared Procrustes distance to all the forms of the original sample. The Procrustes average shape is taken as having been centered and scaled already, and by convention it is graphed with its longest axis (larger principal moment of inertia) horizontal. Finally, for each original figure in the data set, superimpose it over the form of the sample average in exactly the position that minimized the sum of squared distances between corresponding points, the sum of squares we already used in the course of determining that this particular average shape minimized their total. The positions at which the original landmark locations arrive after this superposition are called the *Procrustes shape coordinates* of the original figures. For the theorems justifying all of these arbitrary-seeming manipulations—for instance, to understand why there is not a separate

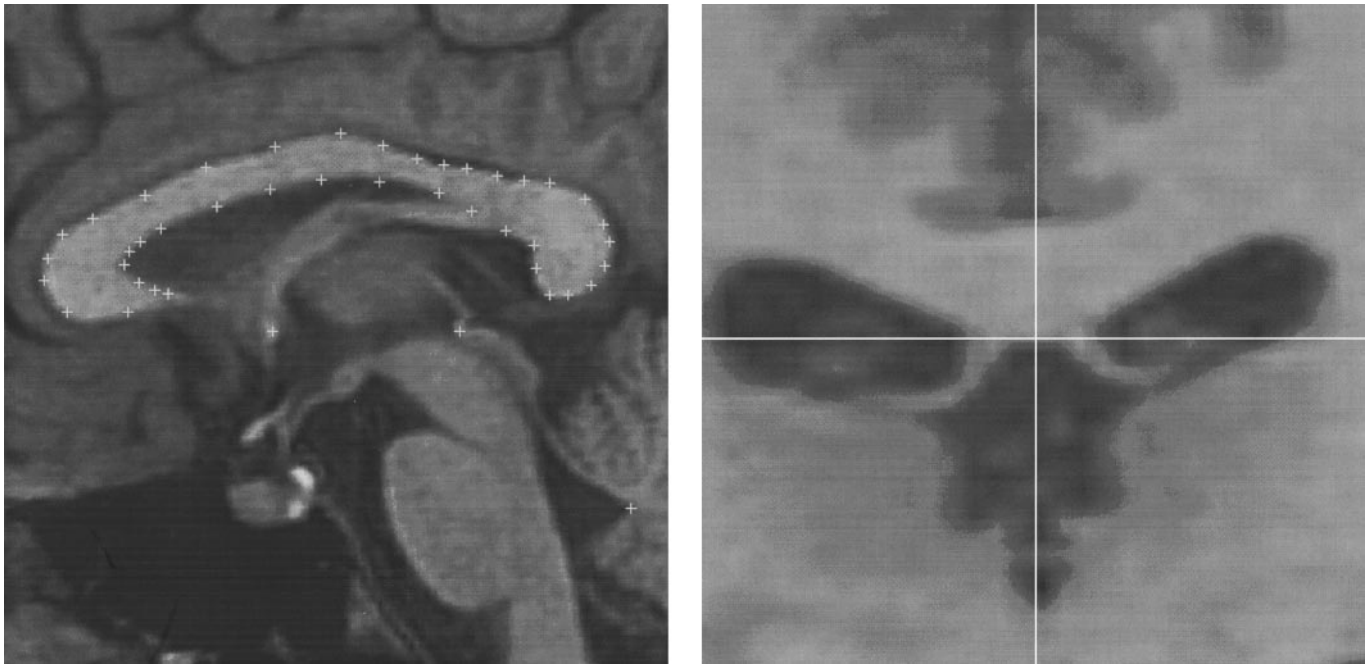

FIG. 1. Aspects of digitizing the corpus callosum in 3D. (Left) Full outline, one subject, as projected onto a near-parasagittal plane of the image. Except for rostrum, all of the points on the outline have been allowed to slip to minimize bending energy with respect to a template form. Other landmarks: 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 Fig. 4) 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, and the point digitized is slipped along that tangent line until it lies on the "midline" located here. Digitizing does not take place in any single plane, but in three dimensions, after a variety of sections such as this one verify that the candidate point is an appropriate one. From Bookstein *et al.* (2001).

centered-and-scaled Procrustes average form for each subgroup—see, for instance, Kent and Mardia (2001).

When empirical curves, such as the callosal outlines here, are digitized as discrete point series, individual points are not claimed to be homologous (to correspond) from subject to subject, and consequently, variability along the tangent direction is not informative. There are various versions of the Procrustes method that slide these points along the tangent direction (line or plane) so as to remove this variation for the purposes of averaging shapes and representing their variation and covariation. For instance, the tangential variation can be removed by minimizing bending energy with respect to an average (Bookstein, 1997b, 1998, 1999) or perhaps by perpendicular projection onto a template (Andresen *et al.*, 2000). Points arising from curves or surfaces by sliding this way are called *semilandmarks*, and analysis of semilandmark configurations is in terms of shape coordinates after the Procrustes superimposition over their average, as before. (The actual algorithm we used alternates the Procrustes averaging with an incremental sliding along the discretized curve.) As the present application deals with an almost planar curve, we projected these points onto a single synthetic midplane prior to further analysis. There resulted a total of 80 coordinates, two (one x-coordinate and one y-coordinate) for each of the 40 points of the outlines, as in Fig. 2a. We will refer to these two-dimensional forms as *midline callosal outlines*.

It is often useful to compute *principal components* of shape when it has been represented by Procrustes shape coordinates in this way. These components should not be thought of as factors somehow causally responsible for the observed variation, but instead as specifying the low-dimensional summaries of the data that, taken together, best predict the individual values of all of the shape coordinates simultaneously. Two other interpretations may be helpful: these principal components are the dimensions that best reproduce the observed (Procrustes) distances between all the forms of the data set using linear combinations of the original coordinates (the interpretation as *principal coordinates*, see Reyment and Jöreskog, 1993); also, they are the linear combinations of the shape coordinates that have the greatest variance in the data compared to the variance they would have if the data arose from a model of pure gaussian digitizing noise of the same small variance at every landmark in every direction (Bookstein, 1997a). The components called for by any

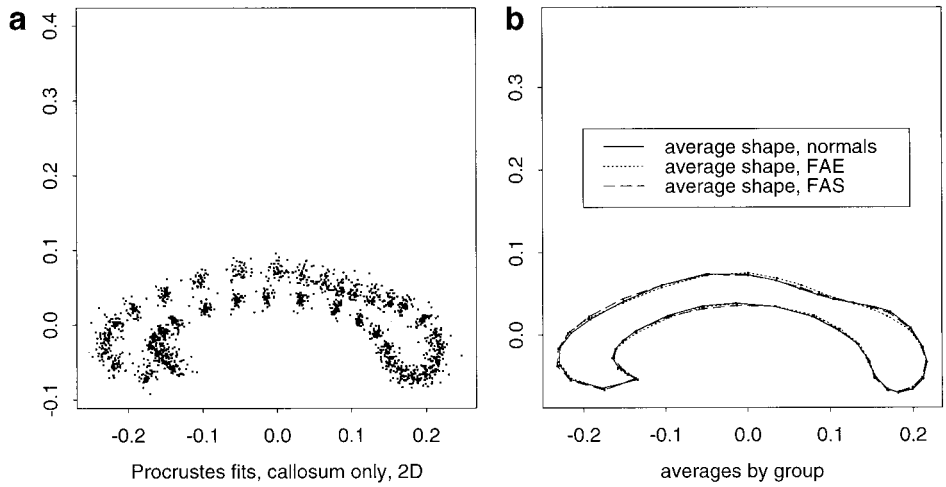

FIG. 2. Procrustes shape coordinates for the callosal outline projected into two dimensions. (a) Full scatter of 80 shape coordinates (*x*- and *y*-coordinates for each of 40 points) for 45 subjects. (b) Average outline shapes for three groups of 15 men each. There are no significant differences of mean shape among these three groups of outlines, although there was a 7% difference in centroid size (not shown) between the exposed and the unexposed.

of these three equivalent definitions are just the ordinary principal components of the *covariance matrix* (not the correlation matrix) of the shape coordinates produced by the Procrustes strategy.

Multivariate findings pertaining to landmark or semilandmark shape can be drawn as deformations of the pictures from which the points were originally extracted. The particular deformation exploited here, the thin-plate spline (Bookstein, 1991, 1999), gives the smoothest diagrams consistent with the actual statistics of the shape coordinates we are using. In this paper it will not particularly matter what we mean by "smoothest" in this characterization; the criterion actually invoked is the bending energy of this thin-plate spline, the integral of summed squared second derivatives of the interpolated mapping over the entire plane.

Significance tests of two principal types will be applied to these outline data: multiple tests of mean shape difference over diagnostic groups, and one test for the additional variance we will notice in the pool of exposed subjects, of whatever diagnosis, with respect to the subsample of the unexposed normals. The shape differences between averages could be at any semilandmark, and likewise the extra variance. Significance levels for complex hypotheses like these should be computed by *permutation testing* (Good, 2000), the same approach often used for tests of functional image signal patterns (Bullmore *et al.*, 1999). In a permutation test, any quantity of interest is assigned a *P* value from the "null" distribution that arises when the underlying association claimed to be real is broken by permuting variable values over cases. In the test of group mean shape difference, for each such permutation one computes *pseudogroup averages* of the cases that were assigned each group label instead of the labels with which they originally entered the data set. Then the distribution of interest is of the distance between pseudogroup averages taken in the direction perpendicular to the sample average outline, the one shape coordinate direction that is meaningful here, and then squared and finally summed over all the semilandmarks. The actual Procrustes distance between the true group averages is referred to that distribution for its *P* value. For the test of localized excess variance, the quantity of interest will be the ratio of variance in the exposed 30 subjects to variance in the unexposed 15, again ignoring any sliding along the curve. In this case the permutation distribution we use for judging the significance of the observed maximum at some particular semilandmark is the maximum of those ratios over all the semilandmarks, not just the one at which we will claim the actual signal is found.

The Neuropsychological Variables

Each neuropsychological data set of 260 scores derives from a 5-h testing session at the University of Washington on the same day as the subject's MRI. For each of 25 different instruments, Table 1 sets out the count and variety of its explicit scores. Either from the sources cited in Table 1 or from unpublished pilot studies, each block was known in advance to show mean shifts between earlier normal and FAS samples. Some of these blocks of tests measured attention, memory, executive function, or information processing. Others, such as motor tasks, balance, and processing speed, focus on a variety of specific brain regions that are suspected of abnormality in the fetal-alcohol-affected adult. To avoid problems of distribution shape, statistical analysis was carried out not on these scores but on

TABLE 1

Neurocognitive and Neuromotor Tests and Types of Scores Used in the Partial Least Squares Analysis

Acronym	Instrument name (domain)	Count	Items
IQ	WAIS-R	12	Full-scale IQ and its 11 subscale scores
WA	Word Attack	1	pronouncing nonsense words
WRAT	WRAT-R	1	standardized arithmetic
TLC	Talland Letter Cancellation (attention)	12	# correct, omissions/commissions/total completed: three conditions (capitals, spaces, capitals and spaces)
CPT	Continuous Performance Task (attention)	13	# errors of omission/commission, mean and SD of reaction time: three conditions (X, AX, and a tonal task)
APT	Attention Process Training	4	# correct, # errors: two conditions (selective and alternating attention)
WCST	Wisconsin Card Sort Test (attention, EF)	28	# errors (perseverative/nonperseverative), conceptual responses, categories completed, total trials administered
SVRT	Spatial Visual Reasoning Test	10	# errors, study and reaction time: 8 trials with 3, 4, or 5 pieces
TR	Trails A & B (EF)	4	# seconds, # errors, 2 conditions
STR	Stroop Color-Word Test (EF)	6	# read, # errors, three conditions (word, color, color-word)
SSM	Stepping Stone Maze (memory)	9	# errors, # seconds: two conditions (learning and criterion trials)
CVLT	California Verbal Learning Test (memory)	19	# correct in learning trials (short and long delay), recall (free and cued), clustering (semantic and serial); No. intrusions, perseverations, false positives
CE	Cognitive Estimation (EF)	15	distance from the median, answers to 15 common knowledge questions
CTT	Consonant Trigrams Test (attention, memory)	4	# correct: 3-, 9-, and 18-s delays
CWT	Controlled Oral Word Association Test (verbal fluency)	5	# animals named; # words named starting with F, A, or S [!]
RFF	Ruff's Figural Fluency (nonverbal fluency)	2	# unique designs and perseverations
DNCT	Denckla Neurological Coordination Test	18	# seconds and difficulty rating on finger tapping (thumb-index), successive finger movements, supination/pronation (alternating arms), toe-tapping, alternating heel-toe tapping: dominant and nondominant
DB	Dynamic Balance	4	# times and total time in balance, and learning over 5 trials
ITRT	Interhemispheric Transfer Reaction Time	22	Mean reaction times and SDs, simple, choice, and reverse conditions: right and left hemispheres
PR	Pursuit Rotor (hand coordination)	13	Time, count, average duration on target, 15, 30, 45, and 60 rpm
HST	Hand Steadiness Test	4	# errors and latencies to correct: dominant and nondominant
FS	Finger Sequencing	11	# correct sequences, # errors in two conditions: 2- and 3-finger sequences (dominant and nondominant)
RCFT	Rey-Osterreith Complex Figure Test (praxis)	1	Accuracy of figure reproduction
HAND	Handedness	1	Binary score
BR	Psychometrist's ratings	41	Ratings of behavior and performance during tests

Note. #, number of; EF, executive function. Except for the CE block, all scores are converted to full-sample ranks prior to multivariate analysis. CE raw scores were "folded" after conversion to ranks: the sample median score is ranked 1, the scores just above it and just below it are ranked 3, and so on until the sample maximum and minimum raw scores are ranked 45.

References for tests. IQ, Wechsler 1981, Streissguth 1989. WA, Woodcock 1987, Streissguth 1994b. WRAT, Jastak 1984, Streissguth 1989. TLC, Talland 1965, Streissguth 1994a. CPT, Rosvold 1956, Mirsky 1991, Streissguth 1994a, Connor 1999. APT, Sohlberg 1989, Kerns 1997. WCST, Heaton 1993, Olson 1998. SVRT, Pellegrino 1987, Sampson 1997b. TR, Army 1944, Reitan 1985. STR, Golden 1978, Streissguth 1993. SSM, Milner 1965, Streissguth 1994a. CVLT, Delis 1987, Kerns 1997. CE, Shallice 1991, Kopera-Frye 1996. CTT, Peterson 1959, Kerns 1997. CWT, Benton 1968, Kerns 1997. RFF, Ruff 1986, Kerns 1997. DCNT, Denckla 1974, Streissguth 1993. DB, Bachman 1961, Barr 1990. ITRT, Iacoboni 1995. PR, Heindel 1989, Mattson 1997. HST, Matthews 1978, Barr 1990. FS, Denckla 1974, Streissguth 1993. RCFT, Meyers 1995, Streissguth 1989. HAND, Reitan 1974, Streissguth 1989. BR, Streissguth 1989, 1993.

their rescalings into 1-to-45 sample ranks. This transformation, while somewhat unusual in the field of neuropsychology per se, is familiar in the literature of multivariate permutation tests within which we couch the significance tests to follow; for a discussion of the merits of converting data to ranks prior to significance testing, see, for instance, Good (2000, Section 9.3.2). Ties are scored at the average of the ranks they span. For instance, a binary score split 15:30 would be assigned 15 values of 8 and 30 values of 30.5. Scores missing by virtue of the subjects' inability to perform tasks (e.g., no "trials to criterion" if criterion is never reached) were assigned the poorest rank; other missing data (instrument failures, etc.) were replaced by sam-

ple averages of the valid data. Each score is oriented in the sense of its common clinical scaling, not in the sense of deficit or anticipated alcohol effect.

Partial Least Squares for Shape-Behavior Covariances

The statistical technique of partial least squares (PLS) is useful for representing relationships between high-dimensional measurement blocks (Bookstein *et al.*, 1996; McIntosh *et al.*, 1996). We will apply it to the data set of 80 shape coordinates, X_i, $i = 1, \ldots, 80$, by 260 scaled neuropsychological rank scores Y_j, $j = 1, \ldots, 260$. PLS analysis results in a series of pairs of linear combinations, of which the first two will merit consideration here: the two pairs $(LV_{1X} = \sum_{i=1}^{80} A_{1i}X_i,\ LV_{1Y} = \sum_{j=1}^{260} B_{1j}Y_j)$ and $(LV_{2X} = \sum_{i=1}^{80} A_{2i}X_i,\ LV_{2Y} = \sum_{i=1}^{260} B_{2j}Y_j)$. These formulas, called *latent variables* (LVs), are useful both for their covariances and for their coefficients. The values of these expressions for each case of the analysis are called *callosal shape scores* and *neuropsychological latent* variable scores, respectively. The first pair—that is, the first callosal shape score and the first neuropsychological latent variable score—have the greatest sample covariance of any pair of linear combinations between the blocks for which coefficients have been normalized ($\Sigma A_{1i}^2 = \Sigma B_{1j}^2 = 1$). The geometry of the A_1's specifies the shape variable—the pattern of joint landmark rearrangement—with greatest relevance for neuropsychological functioning as a whole, in this sense of maximizing cross-block covariance. Patterns like this are the shape variables to be presented below in the deformation diagrams. At the same time, the profile given by the B_1's is the neuropsychological profile of greatest relevance for the shape analysis. The second pair of scores (that is, the linear combinations using the A_2's and B_2's) have the greatest sample covariance of any pair with normalized coefficients perpendicular to those of the first pair. All the pairs can be generated at once by the singular-value decomposition (SVD; Mardia *et al.*, 1979) of the entire original 260×80 covariance matrix. Shape variables arrived at from singular-value decomposition and diagrammed as warped grids are called *singular warps*. Here the first pair of latent variables are named CSW1 (callosal singular warp 1) and NLV1 (neuropsychological latent variable 1), and the second pair, CSW2 and NLV2.

The covariances underlying the SVD driving PLS are *pooled over all three groups*: our PLS analysis ignores diagnosis except in the choice of symbols for the scatterplots. This paper reserves the phrase "principal components" for analysis of shape coordinates alone and "latent variables" or "singular warps" for these linear combinations optimally relating shape to behavior. As orthogonal rotations of basis do not matter for an SVD, PLS based on shape coordinates and PLS based on principal components of shape coordinates are identical.

We will emphasize three features of the brain–behavior PLS analysis: an excess of variance among the exposed subsample in the second neurobehavioral latent variable, a correlation between the second pair of latent variables, and the apparent emergence of two clusters of exposed patients differing in both callosal shape and behavioral profile in interpretable ways. The correlation of callosal shape against neurobehavioral profile will be tested by permutation of the callosal shape coordinate vector against the neurobehavioral ranks vector. The finding of excess variance in neurobehavioral profile for the exposed subgroup will be tested by permutation of group along with callosal shape in the same setup. For the final finding to be reported below, the meaningful clusters in the LV–LV scatter that we interpret as subtypes of fetal alcohol damage, we require a confidence interval, not just a significance level; this will be produced by jackknifing the signal strength of the observed difference between the patterns when the entire analysis leading up to the formula for NLV2 is repeated leaving out each of the 45 subjects in turn. All these tests, like those for outline form alone reviewed already, take into account the excess of number of variables over sample size in this data set. That is, all the *P* values in this paper pertain to analysis of all 80 shape coordinates, all 260 outcomes, or their combination, at once; no further adjustment for multiple comparisons is necessary.

RESULTS

Reliability of Curves

As the main study is of structure–behavior associations, not of structural variation per se, we paid only modest attention to digitizing error. Six outlines (two randomly selected from each diagnostic group) were redigitized, from scratch, at an interval of about 12 months. This means that a local axis of symmetry was redetermined independently, and a level of medium gray along it located independently, for rostrum and for each of the 39 sliding points in each of the six cases. In the original Cartesian coordinate system of the MR image—that is, without any of the centering or rescaling involved in the Procrustes analysis to come—each of the redigitized points proves to be close to some segment between consecutive semilandmarks of the originally digitized midsagittal outline polygon. This perpendicular distance is converted to the Procrustes metric by dividing out centroid size. There results a true standard error of measurement of the outline in the informative direction. It is found to be less than 20% of the sample standard deviation at typical points of the figures to come. Thus these curves have ade-

quate reliability for the distinctions we are pursuing, which involve large-scale features of this shape that shift by substantial extents in the Procrustes shape space, as will be shown in the grid images that follow. One of the reasons for this perhaps surprisingly high reliability is that along many arcs of this midline, the intersection of the symmetry line with the callosal surface typically lies near the bottom of a fairly sharp groove that twists out-of-plane in an unpredictable fashion as it winds around the callosum. In these regions, the reliability of the semilandmark representation taken in 3D is likely to be better than that of curves traced in any single plane section of the image, however carefully that section is chosen.

Average Size and Shape

In the Procrustes approach, one analyzes size and shape separately. "Size" is measured as the square root of the sum of squared distances of all the labeled points from their center of gravity case by case, the centroid size introduced above (Bookstein, 1991; Dryden and Mardia, 1998). Callosal outlines of the FAE and FAS groups average 92 and 94% of normal mean size, respectively, when size is quantified in this way prior to any Procrustes analysis. These are not different from each other but, as a pool, are significantly different from normals ($P \sim 0.002$ by t test). For arch-shaped forms like these, the formula for centroid size reduces more or less to the long diameter of the callosal form, the distance from genu to splenium. Midline callosal areas (area of these midline callosal outlines), likewise computed before any scaling, average 91 and 95% of the normal mean for FAE and FAS, respectively, but none of these differences are significant, owing to the fact that callosal area has a larger coefficient of variation than the long diameter of its arch in this sample. For additional discussion of these and other size measures for this data set, see Bookstein *et al.* (2001).

As Figure 2a shows, there is considerable variation in the locations of the semilandmarks within the full sample, but no suggestive or statistically significant differences in the average shapes of the callosum among the three diagnostic groups (Fig. 2b), as computed by any of four separate permutation tests of Procrustes distance: normal versus FAE, normal versus FAS, FAE versus FAS, or normal versus exposed.

Shape Variation

Despite the absence of average shape differences, there is a remarkable difference between exposed and unexposed in *variability* of shape, which can be examined either globally or locally. Globally, we use the version of principal components analysis that applies to labeled point data, as reviewed under Methods. Figure 3 shows the first three principal components for the full 45-subject data set of callosal shapes (eigenvalues 0.031, 0.029, and 0.022 out of a total Procrustes variance of 0.177; note the near circularity of the first two). As deformations, these principal components are large-scale aspects of shape: height–width ratio, splenium height–genu height ratio, and general thinning/thickening of the arc. From the corresponding scatters of scores (Fig. 3, bottom row), there is substantially more variance on principal component 3 for the exposed than for the unexposed ($F_{29,14} = 3.45$, $P \approx 0.01$), but there is no mean difference in scores for any of these principal components among the groups, nor any differential clustering between the FAS and FAE subsets. The fourth component (eigenvalue 0.018) shows neither mean differences nor variance differences, and so we stop this report at the third.

Likewise, local variability of shape is higher in the exposed than in the unexposed. Over all 40 semilandmarks, the greatest ratio of variances between groups in the direction perpendicular to the average curve is 6.73 at semilandmark 28, near the spring of fornix on the isthmus (Fig. 4). Geometrically, the pattern in the figure suggests a considerably more regulated alignment of splenium with isthmus in the unexposed. A discrimination of exposed from unexposed by this single coordinate detects all but 4 of the 30 exposed cases with only 2 false positives. This is a *quadratic* discriminator in that it attends to unsigned (bidirectional) distance from the average. Notice that its distribution in Fig. 4 does not distinguish FAS from FAE. It is this combination of excess shape variation in the exposed both globally and locally that constitutes the structural aspect of our findings.

Recall the first part of this study's hypothesis: that callosal shape is differently distributed between exposed and unexposed subsamples. If the finding concerns a ratio of variances, then we should be testing that ratio by a permutation of group label over the set of 45 outline shapes. The best of these ratios for a single point, 6.73, is nearly double that for the best principal component (the ratio of 3.45 for PC3, Fig. 3, as already noted), and so it is that maximum we test. Although the finding deals only with semilandmark 28, the appropriate distribution is of the maximum of this ratio over all the semilandmarks, not just the 28th. Owing to the obvious presence of a few outlying points in Fig. 2, we work with ranks of the signed deviations (recall that transformation to ranks is a common maneuver in the world of permutation testing generally). In 2000 random permutations of exposure status over the callosal outlines, a ratio greater than that in Fig. 4 is found at any semilandmark only seven times. As the standard error of that count of 7 is almost exactly its square root, and $7 + \sqrt{7} < 10$, we can reasonably take the significance level of this variance ratio finding as $P < (10/2000) = 0.005$. (This simple application of permutation methods to testing differences in variances is appropriate whenever the subsample means

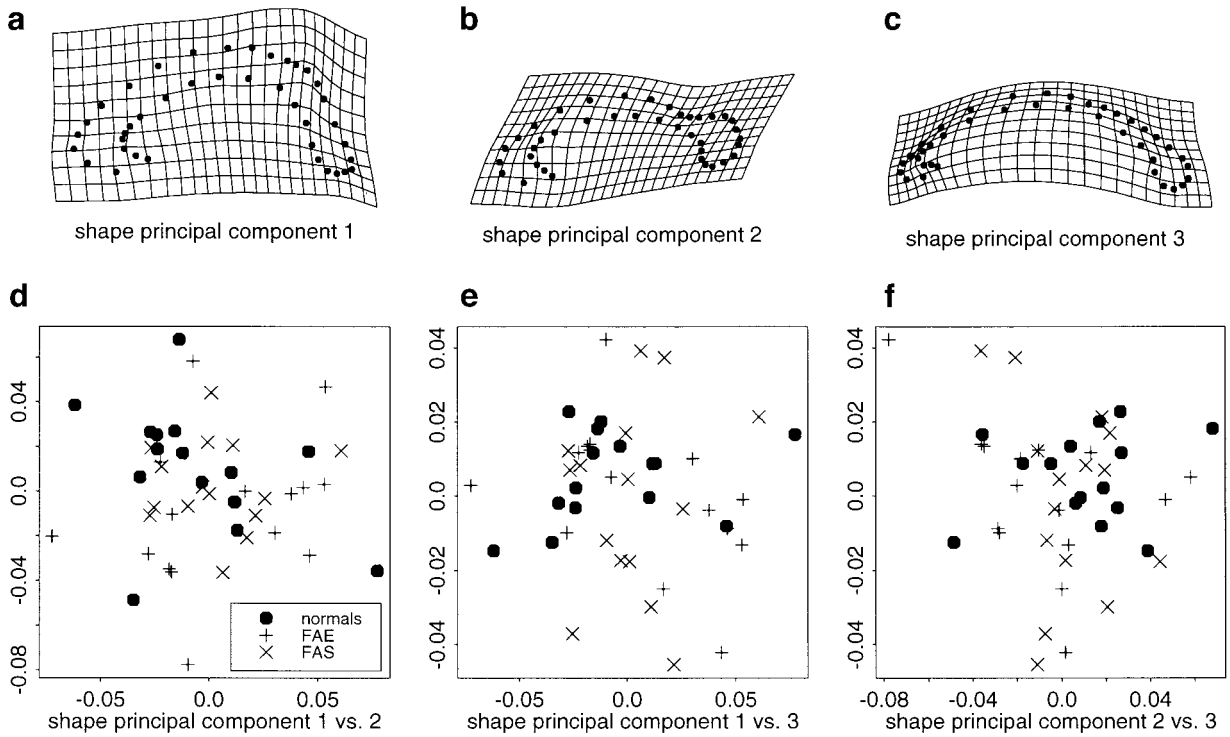

FIG. 3. Shape principal components and scores. (Top row) The first three principal components of the callosal outline shapes in Fig. 2, diagrammed as deformations of the average outline by arbitrary multiples. Eigenvalues: 0.031, 0.029, and 0.022. (Bottom row) Pairwise scatters of the first three principal component scores, showing excess variability of the third in the exposed subgroups. Group is coded by plotting symbol: ●, unexposed; +, FAE; ×, FAS. Scatter of the unexposed subgroup is vertically constrained in (e) and (f).

are "known to be equal" (Good, 2000, p. 40), an assumption that seems justifiable in this case. One reviewer kindly pointed out that this test presumes also that the distributions differ only by a scale factor. By a theorem of Romano, 1990, in any event the resampling distribution of the variance ratio corresponds to a blended distribution. The assumption of proportionality is not obviously violated in the first few principal components of the joint distribution, Fig. 3.)

Brain–Behavior Analysis

The covariance matrix of the ranks of 260 neurocognitive and neuromotor indicators against the shape coordinates of Fig. 2 (a matrix 260 × 80) was subjected to PLS analysis. Singular values were 4.47, 3.03, 2.13, ... Figure 5 shows the first two callosal shape scores and neuropsychological latent variable scores that result, and Fig. 6 shows the first two singular warps (latent variables of shape interpreted as splined deformation grids). Singular warp 1 represents general vertical expansion/compression mildly intensified in the isthmus. Singular warp 2 entails relative thinning/thickening of the callosum almost uniformly along its arc. Note also that the thin arc appears as if it might penetrate further "into the frontal lobe" than the thick. Latent variable pairs beyond the second show no useful patterns by group, and so are neither figured nor scattered here.

Figure 5a shows that the unexposed and the exposed subsamples have the same range of scores on singular warp 1 (vertical expansion/contraction) but differ considerably in average values of the corresponding neuropsychological latent variable, NLV1. This is the pair of normalized dimensions (one of shape and one of neuropsychological performance) that have the highest pooled sample covariance (namely, 4.47), that is, the strongest predictive relationship from form to function. The corresponding correlation is 0.43 (0.30 in the unexposed subsample alone, but 0.53 in the exposed; the covariance is thus not particularly dependent on the mean shift in NLV1 between groups). NLV1 is highly IQ-loaded. Of the 10 items having the highest saliences (coefficients B_1 of the previous equations), 4 are IQ subscales or subsummaries (Picture Assembly, Block Design, the Performance IQ subscale, and Picture Completion), another is Word Attack, 2 are APT success scores (selective and alternating attention, numbers correct), and the rest are Wisconsin Card Sort error scores. The correlation of this summary neuropsychological LV1 with full-scale IQ is −0.743 within the unexposed subsample and, owing to systematic deficits of IQ in the fetal alcohol disorders, a full

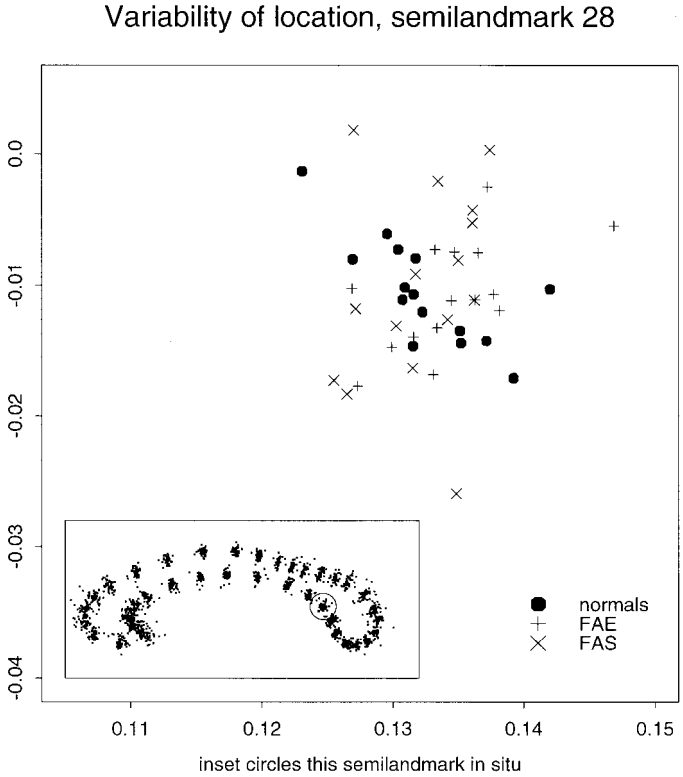

FIG. 4. The exposed subgroups show much greater variability of the location of semilandmark 28, shown here in context between isthmus and splenium (inset) and in magnification. The uniform component of variation from the average callosal shape has been removed from the Procrustes coordinates pictured on the left in Fig. 2. The coordinate perpendicular to the average curve (northeast–southwest direction) is a very effective quadratic discriminator of exposure.

−0.885 in the full sample of 45 men; the 2 exposed subjects who fall clearly among the unexposed (+ and × at lower left in the figure) are the two of highest IQ. The corresponding dimension of shape, which itself correlates −0.292 with IQ, is as shown in Fig. 6c: a lower, flatter arch goes with high IQ. There is no mean difference between FAS and FAE on NLV1 in this sample.

The second latent variable pair, Fig. 5b, displays a huge correlation, more than 0.8, between an aspect of callosal shape (nearly uniform thinning/thickening) and the corresponding neuropsychological latent variable, NLV2. The exposed subgroup (FAS and FAE thoroughly intermingled) is scattered well past the unexposed *on both ends of the distribution*. In the scatter of NLV1 versus NLV2, Fig. 5c, the unexposed are in a fairly tight cluster at left center; the exposed, regardless of diagnosis, are distributed with 8.79 times as much variance, mostly in the vertical band at right. Note that the top 12 points on NLV2 were all exposed (plotted as + or ×) and likewise the bottom 10. The resulting pair of apparent clusters is separated by 2.57 standard deviations of the pooled NLV2 score. This bipolarity in the scatter for the exposed strongly suggests two additional clinical entities that will be characterized separately in the course of the discussion.

The second part of our hypothesis, that callosal shape differences correlate with differences in behavior, can be tested by applying permutation procedures to this PLS analysis. When the matching of images to test scores was randomized 500 times, no correlation among the first three pairs of latent variable scores was ever found larger than 0.45, and an F ratio of greater than 8.79 for the ratio of subgroup variances, exposed over unexposed, was encountered only once. Taken together, the two tests show the PLS analysis of callosal shape against neuropsychological outcome to be highly statistically significant as a whole. In order to estimate sensitivity and specificity of the paired extreme clusters (see Discussion), we need the standard error of the cluster separation emerging in Fig. 5c. For this purpose, we fixed the membership of these clusters (the cases with the top 12 NLV2 scores and the cases with the lowest 10 scores) and then recomputed the formula for NLV2 from first principles for the 45 different data sets of 44 subjects (leaving out each subject in turn). This is called a *jackknifing* of the interesting statistic, the difference between them. In the resulting recomputations, the lower cluster of 10 averaged −1.25 when NLV2 is z-scored, versus −1.38 in the complete analysis, and the upper cluster of 12 averaged 1.26, versus 1.19 in the complete analysis. Their separation thus averages 2.51 units of NLV2, with a jackknifed standard error of 0.44. We conclude that the two poles of NLV2 are stable; they will be characterized in connection with Table 2 in the Discussion.

At this point, then, both clauses of our hypothesis, the difference of distributions of callosal shape between exposed and unexposed and the correlation between callosal shape and neuropsychological performance, have been shown to be highly statistically significant. None of the main multivariate findings can be plausibly attributed to chance, nor can the sorting of most of the exposed cases into the extremes of NLV2 as shown in Fig. 5.

DISCUSSION

Midline Curves as Data

The findings here in respect of size have been anticipated in earlier studies by other methods. Our finding of 7% diminution in midline callosal area among the fetal-alcohol-exposed is in the direction expected from earlier reports in humans (Mattson *et al.*, 1994; Riley *et al.*, 1995; Swayze *et al.*, 1997, Sowell *et al.*, 2001a) and is in keeping with volumetric studies of fetal alcohol effects upon other regions of the brain (Archibald *et al.*, 2001; Sowell *et al.*, 2001b). Other methods of studying the information content of solid brain images, such as

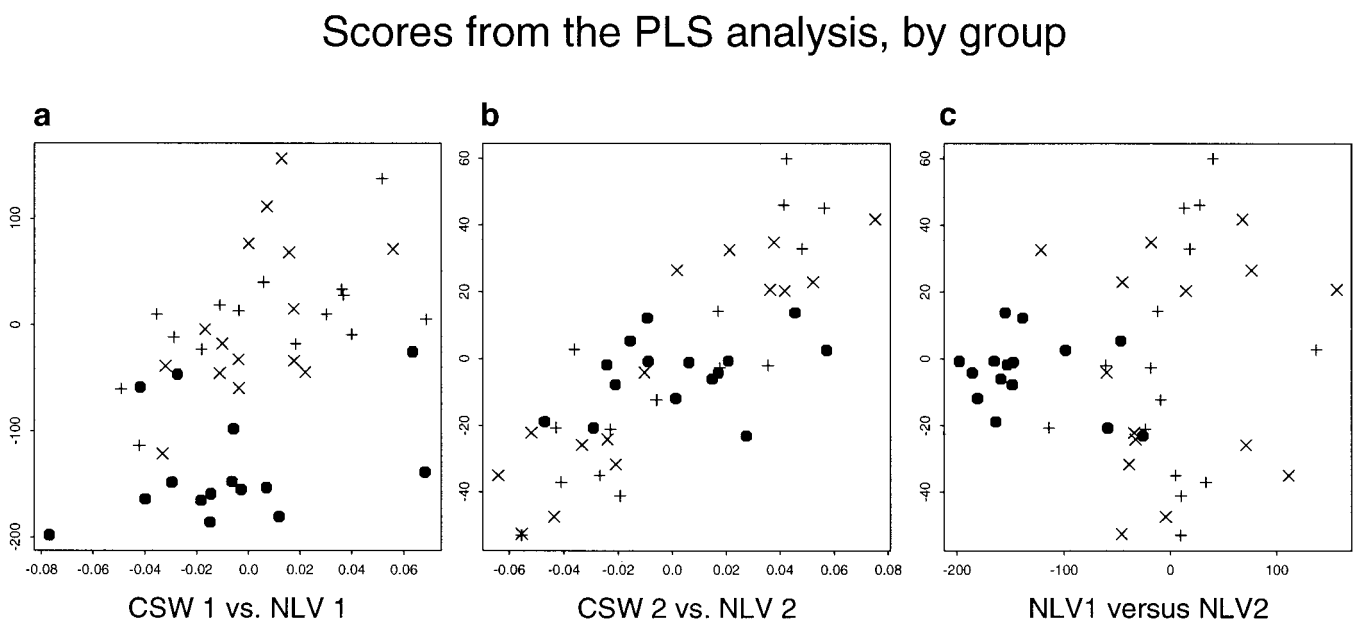

FIG. 5. Partial least squares (PLS) analysis of callosum shape against 260 ranked neuropsychological scores: scatters for the first two latent variables, in pairs. Group is coded as in Fig. 3. CSW, callosal singular warp; NLV, neuropsychological latent variable. (a and b) Callosal singular warp scores against corresponding neuropsychological LV scores for LVs 1 and 2. Singular values (covariances of LV pairs), 4.47, 3.03; correlations, 0.43, 0.81. (c) Scatter of the first two neuropsychological LV scores.

the configurations of 33 landmark points that were also studied in these same 45 subjects (Bookstein *et al.*, 2001), confirm the size difference. But the finding of shape hypervariability in the exposed subsample appears to be novel and may owe to the nature of the morphometric data representation we have pursued here.

The data in Fig. 2, while two-dimensional, do not arise from any geometrical plane within the original three-dimensional data volumes, nor (after the fashion of sulcal tracings) as actual curves in the three-dimensional space of actual human brains. Instead, they are an abstraction exploiting the near symmetry of brain form in our species. Many small structures (anterior and posterior commissure, fourth ventricle) have their own planes of symmetry that are very closely aligned with one another, and other paired structures, like the colliculi, lie almost equidistant from these planes on either side. When these structures are represented as points, statistical methods are well-developed (e.g., Mardia *et al.*, 2000). Beyond these punctate loci, the midplane includes at least one important curving form: "the outline" of the corpus callosum, the bundle of neural matter connecting the two cerebral hemispheres. Call it a *symmetry curve:* a locus having an implied *symmetry normal* as well as a tangent vector. The symmetry normal is not the principal normal to the curve (which lies in the mirroring plane), nor is it the binormal, but is instead the normal to the mirroring structure in which, locally, elements of the curve are presumed to lie. The points of this symmetry curve, in other words, are characterized as lying on the (approximate) line of symmetry of the image cut by any plane through the symmetry normal at that point. One of these cut planes includes the tangent to the curve itself and another the actual (geometric) normal to the curve. Figure 1 (right) is an instance of this second type of cut plane verification. The symmetry normal is the horizontal crosshair in the figure, the normal to the curve is the vertical crosshair, and the tangent to the curve is the vector into/out of the page.

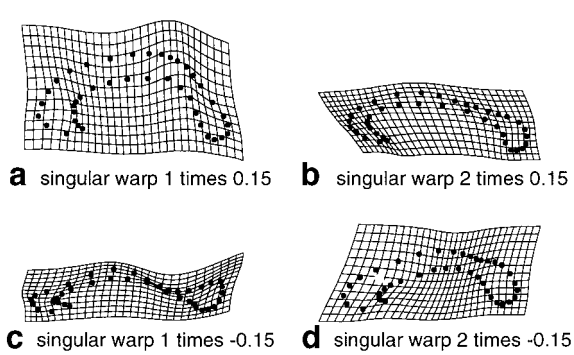

FIG. 6. PLS analysis, continued: geometry of the singular warps. Panels (a and c) and (b and d) represent the average callosal shape plus or minus 0.15 times the first two singular warps (shape variables) as deformations of the average. These exaggerate the shape trends associated with the clusters at the extremes of NLV2 (Fig. 5b). The amplitude of 0.15 is for convenience of visualization only; it lies beyond the range of variation in the singular warp scores shown horizontally in Figs. 5a and 5b.

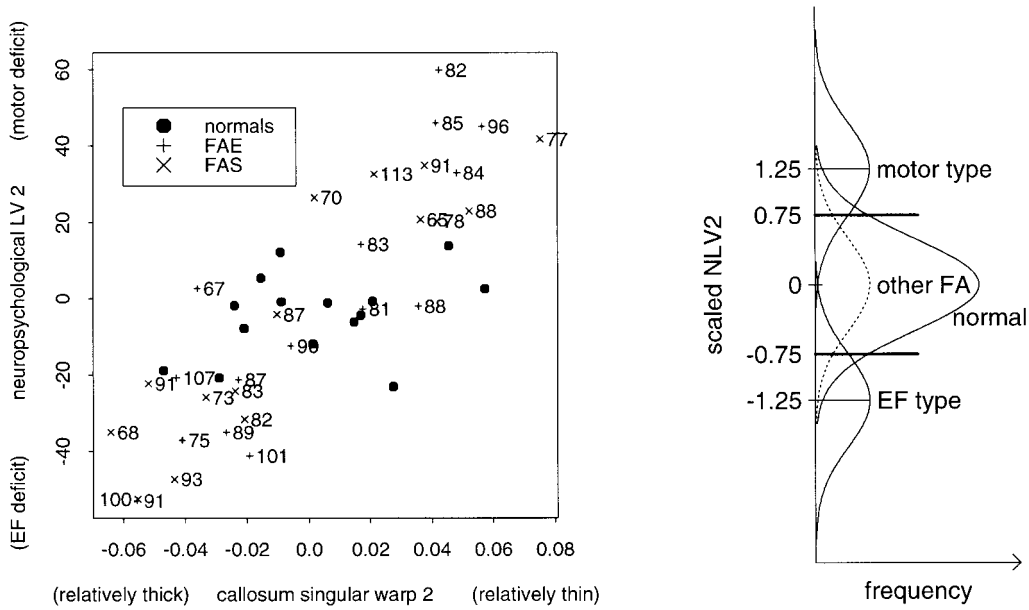

FIG. 7. From PLS to a possible diagnostic protocol. (Left) Enhancement of Fig. 5b: callosal LV2 versus neuropsychological LV2 by group, with full-scale IQ indicated for the exposed. IQ does not confound the obvious excess of variability of this neuropsychological profile in the exposed subsample. The F ratio for difference in variance of neuropsychological LV 2 between the groups is significant at about 0.004 by permutation test, whereas the correlation of 0.81 between the two LV scores is significant beyond any level that can be computed by this approach (see text). EF, executive function. (Right) Schematic of a discrimination based on the extreme clusters of NLV2. The four curves model a high-risk screening population, half unexposed and half exposed, as a mixture of four gaussian subpopulations each with standard deviation 0.5: a normal subsample with mean 0 and three fetal-alcohol-affected subpopulations at means −1.25, 0, and 1.25. Diagnosis by $|NLV2| > 0.75$ (thick lines) correctly classifies most of the "EF-type" and "motor type" as affected but misses most of the "other FA" (dotted histogram). Estimated specificity, 86%; sensitivity, 60% (see text).

The morphometric portion of our data set consists of 40 points on this curve for each of 45 adult male subjects: a total of 3600 Procrustes coordinates. This is a vanishingly small fraction of the full information content in 45 8-megabyte MR images, nor does it sample any aspect of callosal form away from this synthetic midplane. It is therefore particularly startling that so much power for the specific hypotheses at issue here—differences in brain form by diagnostic group and covariation of form with behavior—derives from so low-dimensional a representation. Yet, far from being simplistic, the reduction of all the information of the MR image into this single abstraction of "midline callosal shape" actually embodies a reasonable image-analytic strategy in light of the clear suggestion in prior literature that callosal form is obviously abnormal in many of these patients (cf. Swayze *et al.*, 1997). The finding in Fig. 5b suggests that we have quantified this qualitative difference in a way that might apply to the majority of fetal alcohol cases, not just the extremes.

The Behavioral Profiles That Covary with Callosal Form

The variables emphasized by NLV1, as reviewed in the previous section, tend to be the customary IQ-loaded psychometric measurements—the component IQ scale scores themselves, IQ-mediated measures of executive function, and the like. Centroid size correlates −0.41 with this neuropsychological score and −0.25 with the corresponding callosal shape score. (Correlations with midline callosal area are −0.11 and 0.18, respectively.) The callosa corresponding to the group in behavioral deficit on NLV1 tend to be anteroposteriorly shortened (lower centroid size) and their tendency to excess relative arch height was diagrammed in Fig. 6a.

Each of the second pair of latent variables correlates about −0.40 with callosal area; otherwise, the interpretation of these two scores is not usefully pursued in terms of their correlations. At left in Fig. 7 is an enhancement of Fig. 5b indicating full-scale WAIS-R IQ score at each point for an exposed subject. The 12 subjects highest on NLV2 (upper right corner of plot) were all exposed: 5 are FAE, and 7 are FAS. Their IQs range from 65 to 113. But 10 patients have scores lower than every unexposed subject on this same neuropsychological score, and their IQs range almost as widely (from 68 to 101). In neither cluster, furthermore, is IQ correlated with position in the scatterplot. Hence this dimension of excess variation is picking out extreme neuropsychological profiles independent of both diagnosis and IQ within the exposed group, yet strongly correlated with callosal shape. On the corresponding

singular warp, Figs. 6b–6d, to be either abnormally thin or abnormally thick is equally pathognomonic. The excess of variability in the direction of thinning is consistent with the existing literature of partial or total callosal agenesis. In the other direction, excess relative thickness has not previously been reported in humans.

It turns out that the association of NLV2 with specific neurobehavioral scores is different at its two poles. This is shown in Table 2, which lists the outcome variables, winnowed from the full roster of 260, that show unusually low or unusually high tendencies within the exposed subgroups of lowest and highest NLV2 scores. The first three columns of values are mean raw scores; the last three, mean ranks—it is on these that the entries have been sorted. In reading Table 2, bear in mind that as there are 45 cases the mean rank for the full sample is necessarily 23 on each of these variables and that the direction of deficit may be different for different outcomes, as suggested in their labels (for instance, poor performance is represented by either high error counts or low counts of correct responses).

For the subgroup of 10 exposed subjects with the lowest NLV2 scores, including the three callosa of greatest relative thickness, a variety of CVLT scores have unusually low average values, and a variety of WCST scores have unusually high average values, all in the direction of deficit (Table 2A, column 4) (see Table 1 for all acronyms.) The CVLT tasks assess organization of verbal memory; the WCST, visually mediated executive attention. This profile of damage thus might be characterized as centered about "executive function" (EF), a generalized deficit in judgment, problem-solving, and reasoning (Denckla, 1996; Eslinger, 1996; Dugbartey *et al.*, 1999) often used to describe alcohol-affected people (Connor *et al.*, 2000; Kodituwakku *et al.*, 1995; Mattson *et al.*, 1999). That the extent of this EF-related deficit fluctuates widely at these generally lower levels of IQ (Fig. 5c) is consistent with the "complex overlap" between these constructs noted earlier by Denckla. Notice, too, that the relative callosal thickening associated with the EF-related deficits is also associated with good performance on a simple motor coordination task (Table 2A, entry for PR).

At the other end of the NLV2 scale (Table 2B, column 6), the 12 highest NLV2 scores characterize a distinctly different profile of alcohol effects, involving poor balance (DB), poor motor coordination (PR 15 and 30: seldom on target and unable to stay on target), more errors and slower to correct errors in hand steadiness (HST), and poor spatial learning (SSM). This motor-centered profile of damage corresponds to the relatively thin callosal shape (Fig. 6b). By contrast, the EF scores for this poor-motor subgroup (Table 2A, column 6) are not particularly deficient: they are close to the full sample average rank of 23.

Most of the subjects in this subgroup of affected have a callosal outline that is thinner (higher on CSW2) than average. This combination of poor motor coordination with thinner CC could stem from the loss of some white matter pathways important for that performance, those projecting from parietal sensory inputs, caudate nucleus, and cerebellum through the CC to motor and premotor centers. Among the motor scores found characteristic of this subgroup of affected subjects (Table 2B), pursuit rotor deficits have been found associated with caudate damage in Huntington's disease (Heindel *et al.*, 1989), dynamic balance with poor cerebellum integrity (Baloh *et al.*, 1998), and bimanual coordinated hand movements with CC integrity itself (Eliassen *et al.*, 2000). Sowell *et al.* (2001b) have found that patients damaged by prenatal alcohol exposure have significantly less white matter than unaffected subjects in the parietal and temporal regions, regions that project through the isthmus of the CC (Aboitiz *et al.*, 1992). As Fig. 6b indicates, the isthmus does participate in the generalized thinning associated with this motor type of deficit. Regarding the EF type of deficit, recall the hint in Fig. 6 that the thick callosum (Fig. 6d) fails to extend as far into the frontal lobes as the thinner callosum (Fig. 6b). Executive function deficits have frequently been found to be associated with frontal lobe damage (Shallice and Burgess, 1991; Arnett *et al.*, 1994), and so the failure of the CC to extend far into the frontal lobes in this subgroup of patients may indicate some derangement of those white matter fibers as they cross through the CC.

The presence or absence of "the Face"—the distinction between FAS and FAE—seems to make no difference for this clinically crucial profiling of neuropsychological characteristics as related to callosal shape, and both subtypes have the same average IQ deficit (NLV1). It is as if we have two kinds of patients here, one kind averaging poor IQ with poor performance on certain EF-related tasks and the other kind averaging poor IQ with poor motor performance. Both kinds mix FAS and FAE, while covering a wide range of IQs around that 2-standard-deviation mean deficit. Notice, too, that inasmuch as these two groups have the same average IQ, IQ cannot be responsible for the contrast between poles of NLV2.

In a differently designed study of fetal-alcohol-affected individuals, Sowell *et al.* (2001a) find a significant correlation between our CVLT total correct score (Table 2A, line 5) and an aspect of the *position* of the corpus callosum, specifically its "anterior displacement," in a space standardized to the entire brain. Their methodology of assessing position may well incorporate certain aspects of the reshaping of the anterior callosum analogous to what is encoded in our second singular warp score (Fig. 6). Otherwise, they find

TABLE 2

Neuropsychological Scores Characterizing the Two Extremes of NLV2

		Scores			Ranks		
Task[a]	Variable description	Low 10[b]	Middle	High 12	Low 10	Middle	High 12
	A. Variables on which the 10 subjects with lowest NLV2 scores perform most poorly						
Lowest mean ranks (column 4)							
CVLT	# correct, semantic clustering, long delay free recall	1.9	6.6	4.5	**10.6**	28.6	20.6
CVLT	# correct, semantic clustering, short delay free recall	1.6	6.0	4.0	**11.4**	28.5	20.3
CVLT	# correct, semantic clustering, all 5 learning trials	8.9	19.2	15.5	**12.6**	26.9	22.3
CVLT	Total correct, short delay free recall	7.5	11.6	9.5	**12.6**	27.8	20.6
CVLT	Total correct, all 5 learning trials	41.3	53.6	48.5	**12.7**	27.4	21.3
BR	Behavior rating: good endurance	5.0	6.4	5.5	**12.7**	29.3	19.5
CVLT	Total correct, long delay free recall	8.1	11.8	9.6	**12.8**	27.5	20.9
CVLT	Total correct, short delay cued recall	9.2	12.4	11.0	**13.0**	27.3	21.2
Highest mean ranks (column 4)							
WCST	# perseverations	22.0	9.2	19.4	**31.8**	16.4	28.3
PR[c]	Seconds on target, 15 rpm	18.5	17.5	14.5	**31.9**	25.7	10.5
WCST	# errors, perseverative	20.0	8.7	17.6	**32.2**	16.2	28.3
WCST	# trials administered	115.7	96.1	109.2	**32.2**	17.6	25.6
WCST	% errors, all types	38.9	19.4	33.2	**32.8**	16.6	27.2
WCST	# errors, all types	47.3	19.5	38.7	**32.8**	16.5	27.2
WCST	# errors, nonperseverative	27.3	10.8	21.1	**33.3**	17.0	25.9
WCST	% errors, nonperseverative	22.4	10.6	18.0	**33.4**	17.4	25.0
	B. Variables on which the 12 subjects with highest NLV2 scores perform most poorly						
Lowest mean ranks (column 6)							
PR	Seconds on target, 15 rpm	18.5	17.5	14.5	31.9	25.7	**10.5**
DB	Seconds balanced, all 5 trials	37.0	44.2	29.1	21.2	29.0	**11.2**
PR	Average seconds per time on target, 15 rpm	9.2	7.1	2.0	31.7	25.2	**11.6**
PR	Time on target, 30 rpm	14.2	15.4	11.6	22.8	28.5	**12.6**
BR	Behavior rating: superior, overall	2.5	4.2	2.2	17.0	31.0	**12.7**
DNCT	Difficulty, supination/pronation, alternating arms	5.8	5.9	3.9	26.2	26.9	**12.8**
BR	Behavior rating: superior mental development	2.5	4.2	2.2	16.2	31.3	**12.9**
TLC	# correct, capitals and spaces	50.8	70.6	45.2	17.4	30.7	**13.0**
Highest mean ranks (column 6)							
HST	# errors, dominant hand	88.4	56.4	104.7	26.2	16.3	**31.3**
HST	Sum, latencies to correct, Nondominant hand	7.2	5.9	12.5	22.0	17.6	**32.2**
SSM	Total time, learning trials	349.0	170.6	424.3	30.1	15.0	**32.3**
SSM	# of trials to first error-free trial	9.1	5.7	13.9	22.0	16.5	**32.5**
CPT	SD of reaction time, AX task	10.7	7.3	12.7	26.2	15.6	**32.7**
HST	Sum, latencies to correct, dominant hand	6.8	3.8	10.4	24.4	16.0	**33.4**
PR	# times on target, 15 rpm	3.2	5.7	10.3	15.0	21.0	**33.5**
SSM	# errors, learning trials	38.5	19.5	68.3	27.8	15.1	**34.2**

Note. Scores are included to permit comparisons across studies; extraction of variables here was actually in terms of ranks within the present study. The column in sort is indicated in boldface.

[a] See Table 1 for the expansions of these acronyms.

[b] Subsamples having the lowest 10, middle 23, and highest 12 NLV2 scores.

[c] Exception: not in the direction of deficit.

no signal corresponding to our correlations between outline shape and behavioral profiles, no analogue of our finding of hypervariability and meaningful subtypes among the fetal-alcohol-affected, and no analogue of our structural–behavioral finding in respect of IQ, the relation between CSW1 and NLV1. We cannot

tell to what extent this divergence of findings owes to differences in image-analytic or biostatistical methodology or neuropsychological breadth and to what extent it expresses true differences between their sample and ours. (The Sowell *et al.* sample averages 13 years of age and is almost equally divided between males and females, the outline analyzed was taken from one flat plane section of the volume, and there is no formal representation of its "shape.") To the extent that the finding of a correlation between callosal anatomy and CVLT is held in common between the two studies, it suggests particular attention be paid to alcohol effects on the brain regions crucial to the behaviors tapped by the CVLT task. Of the seven scores appearing in Table 2A, three relate to semantic clustering, a working memory task often associated with frontal lobe integrity, and four relate to simple learning or recall, which is associated with perisylvian and temporal lobe integrity.

Clinical Implications

The Institute of Medicine report (Stratton *et al.*, 1996) proposed replacement of present diagnostic systems for fetal alcohol effects by a fivefold categorization: FAS with confirmed exposure, FAS without confirmed exposure, partial FAS with confirmed exposure, alcohol-related birth defects (ARBD), and alcohol-related neurodevelopmental disorder (ARND). The proposed diagnoses of partial FAS and ARND are intended to apply to patients who show particular behavioral or cognitive abnormalities or delays (e.g., learning difficulties or poor metacognition) that cannot be explained by environment or family background alone. However, the effects of environment or family background are quite difficult to assess. We propose instead that the behavioral or cognitive abnormalities scrutinized for this diagnostic purpose be just those found to be linked, by an analysis such as ours, to the crucial intervening variable for any behavioral teratology study, viz., the trace of prenatal brain damage.

In Harris's language, as excerpted in our Introduction, we have produced strongly suggestive evidence that callosal shape is a crucial aspect of the "neurological information" against which information on behavior and the mind is to be integrated for this class of patients. This might, indeed, supply the localized "lesion" required if quantitative aspects of the image are to abet the diagnostic process. For the brain image to become ancillary to diagnosis in this way, the findings in Figs. 3 through 6 need to be converted to practical decision rules. One discrete shape feature that emerges from a closer study of Fig. 6, for instance, is the difference in shape of the region of genu as a whole. The cases that proved extreme on NLV2 tended to be extreme on CSW2 as well, which entails a vertex of curvature interior to the arch at genu that is strikingly too close to or too far from the line connecting rostrum to the front of the callosal curve. The development of this line of investigation into a formal decision rule for detecting that neuroanatomical "lesion" is in preparation.

While there seems to be only one such focus for the telling "lesion" in this analysis, there seem to be two (the two subclusters of Fig. 5c or Table 2) for the corresponding behavioral profile. It is this pair of possible subdiagnoses that, if confirmed, will lead to the more important follow-on hypotheses combining structural imaging, functional imaging, and behavioral assessment. Callosal shape variability may actually be expressing a range of different, ad-hoc developmental accommodations to the direct impact of alcohol neurotoxicity per se. In that event, the distinction here between our "EF-related" and "motor" types—that is, between "thick" and "thin" callosa—will need to be be confirmed by structural or functional studies of the specific brain systems distant from the callosum itself, the systems that link through the arcs of the callosum entailed in the patterns of Fig. 6. To the extent that there are two types of fetal alcohol damage, such studies need to use neurobehavioral profile as a covariate of shape difference; studies averaging over the two clusters in Fig. 7 will lose power just as studies of schizophrenia lose power when they average over the clinical subtypes so well-known there.

How good a detection protocol might we hope for? At the right in Fig. 7 is a schematic of a plausible underlying model for the observed distribution on NLV2: a mixture of four subpopulations. Assume that the population is half fetal-alcohol-affected and half unaffected and that the affected group is divided evenly in thirds over the three component distributions of "EF-deficit," "motor deficit," and "other" alcohol effects. At the top, at mean 1.25, is the motor type of fetal alcohol patient; at the bottom, at mean -1.25, the EF-deficit type. In the middle, at mean 0, are the rest of the fetal alcohol cases in this sample, those who were not extreme on NLV2 (dotted curve), together with all the normals (tall solid curve). The jackknifed quantities reported earlier are consistent with the common standard deviation of about 0.50 for all three of these distributions that is drawn. Consider, for instance, the "detection rule" that assigns values of NLV2 above 0.75 ($+1.5$ normal standard deviation) to the motor type of fetal alcohol deficit and those below -0.75 to the EF type and calls everybody else unexposed. Given the sample proportions here, we would detect all but 16% of the two extreme types of fetal alcohol deficit, while misclassifying only 14% of the unexposed. Unfortunately, a full 86% of the nonextreme type of cases here, those commingling with the normals in Fig. 7, would likewise be classified as unexposed. The resulting "diagnosis" would have specificity 86% but sensitivity only 60%.

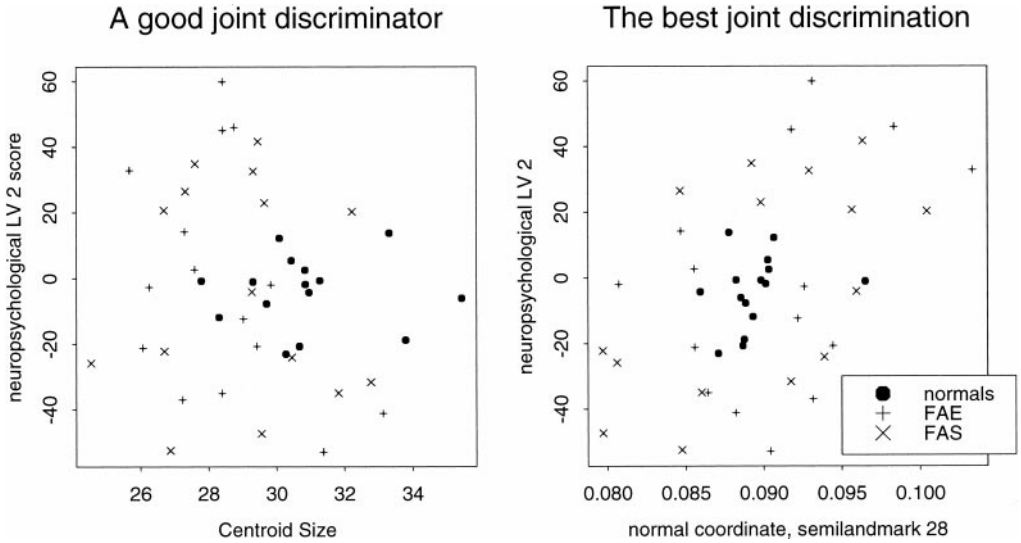

FIG. 8. Better detection rules combining morphometric and neuropsychological information. (Left) A promising combination of morphometric and neurobehavioral scores, centroid size by NLV2, permits the detection of half the exposed cases overlooked by the neurobehavioral categorization alone (see text). (Right) A very lucky discriminator with 100% sensitivity and 93% specificity for exposure: the interior and boundary of the normal cluster (solid circles) here. The horizontal axis is the perpendicular coordinate from semilandmark 28, Fig. 4, and the vertical the neuropsychological score from Fig. 7. The cluster in the middle of the plot incorporates 14 of the 15 unexposed subjects without a single exposed subject intruding.

We can do better by bringing back in information from callosal shape. Even though CSW2 is enormously correlated with the neurobehavioral profile to which we are attending, other dimensions of callosal form need not be. On the left in Fig. 8, for instance, we have scattered the simplest callosal measure of all, centroid size (in effect, its length), against this same bipolar neurobehavioral discriminator and declared all forms abnormal that have either an extreme NLV2 score or else a centroid size less than the minimum of the centroid sizes for the unexposed subsample, regardless of NLV2 score. The number of false negatives immediately drops by half, leaving only four classification errors (three $+$'s and one $\times$ inside the cluster of normals) instead of the eight when we attended to the vertical axis alone. In effect, small size detects a handful of the abnormal callosa that are otherwise normal in neuropsychological profile even as that profile casts suspicion on a great number of subjects having normal callosal length. (An analogous scatter using callosal area in place of centroid size shows no such improvement.)

The shape coordinate incorporated in Fig. 4 turns out to be another such uncorrelated shape variable, one showing quite a bit more effect of exposure than does centroid size. The possible improvement such a focus might afford for differential diagnosis is suggested in the scatter at right in Fig. 8, where this coordinate is combined with the same NLV2 score. All the exposed, along with just one unexposed subject, lie in a ring outside the central cluster of 14 unexposed. The discrimination rule that says "diagnose everything outside this normal cluster as alcohol-damaged" thus nominally has 93% specificity, there being only one "false positive," but 100% sensitivity as well, unconfounded by IQ.

Of course this nearly perfect separation cannot be expected to persevere into additional samples. The "hole" in the middle of the distribution for the 30 patients is surely a fortuitous accident that cannot be expected to recur. Also, the shape coordinate here is not a local measure of the isthmus outline on which it lies; its quantification required the detailed digitizing of the entire 40-point polygon. If we could find a simple shape descriptor uncorrelated with NLV2 that has a variance ratio of 4:1 between the groups (instead of the very helpful ratio of 6:7:1 in the lucky example here), even absent any mean differences by exposure group a further thresholding of the shape measure at ± 1.5 of the standard deviation for the unexposed would reduce the false-negative count by nearly half, raising sensitivity to 78% (versus 60% in Fig. 7) while dropping specificity only to 71% (versus 78% in Fig. 7). (The discrimination in Fig. 8a seems to be running at a somewhat higher level, owing to the nongaussianity of the apparent joint distribution here, but we cannot model these deviations effectively in a sample this small.) A supplemental morphometric descriptor like this need not have derived from callosal outline alone, but could involve quantification of the same MR image from any part of the brain supplying the white-matter pathways here. Separately, the profile-relevant parts (Table 2) of the full battery in Table 1 might be winnowed into a simpler net psychometric protocol.

In any event, explicit crossvalidation of these and other detection rules will require additional samples, not reuse of this one. Three additional waves of 45 subjects each (adult females, adolescent males, and adolescent females) are available for these callosal data, but the neurobehavioral battery is not ready yet. When it is, we will carefully replicate this entire structural–behavioral analysis. For the anatomical data alone, the callosal outline form here, the information content relevant to diagnosis appears to be of equivalent amplitude between adult males and adult females (Bookstein *et al.*, 2001), but the pertinent discriminators arise from different subarcs of the callosum. We do not know yet how specific these discriminators may prove against syndromes owing to other causes than prenatal alcohol exposure (Coles *et al.*, 1997; Jacobson, 1998), such as Down syndrome, Asperger's disorder, or fragile-X. In respect of unselected populations, however, our earlier experience with checklists specific to FAS/FAE behavioral eccentricities (Streissguth *et al.*, 1998) is cause for optimism.

Historically, it was the facial pattern of FAS that brought the reality of alcohol teratogenesis to our attention. If our finding is valid, the face does not tap the dimensions that calibrate the effect of prenatal alcohol exposure on brain and behavior, at least for the type of exposed subjects studied here, who are diagnosed with FAS or FAE, but are not frankly mentally retarded. It is the brain image, not the portrait, that is the more informative for these patients. Syndromology aside, then, our finding forces us to face a terrible antinomy having severe societal consequences. Current systems of social service in the United States are keyed to the dysmorphological diagnosis of FAS or else to mental retardation (Sampson *et al.*, 2000). If major dimensions of the neuropsychological damage consequent upon prenatal exposure, however detected, are independent of both those criteria, then it behooves us to pursue alternative diagnostic protocols equally specific to alcohol damage but enormously more sensitive. The findings reported here have approached that most humane end via careful examination of covariation between brain structure and behavior over the full range of neuropsychological alcohol effects. In particular, the specific executive function deficit suggested here as pertaining to some but not all of these patients comprises the aspect of behavioral deficit that, in practice, induces the greatest range of secondary disabilities (Streissguth *et al.*, 1996; Connor *et al.*, 2000). If this profile is indeed predictable by brain imaging at earlier ages, the implications for clinic and community are as substantial as those for developmental neuroscience.

ACKNOWLEDGMENTS

We thank David Haynor, M.D., for clinical readings of all the MRI scans; psychometrist Michael Hampton; Julia Kogan, for her heroic efforts at assembling a balanced patient sample; and Sterling K. Clarren, M.D., who provided most of the diagnoses within the exposed group. Our digitizing used the dig3 module of Bill Green's Edgewarp morphometrics package, which is available for Silicon Graphics workstations and PCs running Linux by anonymous FTP from ftp://brainmap.med.umich.edu/pub/edgewarp3.2. Other graphics were produced in Splus, a commercial statistics package. Functions for duplicating the analyses here are available from the senior author. Four anonymous reviewers of an earlier version helped greatly with the flow of the argument. The research on which this paper is based was supported by USPHS Grants AA-11037 to A.P.S. and GM-37251 to F.L.B.

REFERENCES

Note: References listed here that are not cited in the main text are cited in Table 1 as sources for elements of the behavioral measurement battery.

Aboitiz, F., Scheibel, A. B., and Zaidel, E. 1992. Morphometry of the Sylvian fissure and the corpus callosum, with emphasis on sex differences. *Brain* **115:** 1521–1541.

Andresen, P. R., Bookstein, F. L., Conradsen, K., Ersbøll, B. Marsh, J., and Kreiborg, S. 2000. Surface-bounded growth modeling applied to human mandibles. *IEEE Trans. Med. Imag.* **19:** 1053–1063.

Archibald, S. L., Fennema-Notestine, C., Gamst, A., Riley, E. P., Mattson, S. N., and Jernigan, T. L. 2001. Brain dysmorphology in individuals with severe prenatal alcohol exposure. *Dev. Med. Child Neurol.* **43:** 148–154.

Army Individual Test Battery. 1944. *Manual of Directions and Scoring*. War Department, Adjutant General's Office, Washington, DC.

Arnett, P. A., Rao, S. M., Bernardin, L., Grafman, J., Yetkin, F. Z., and Lobeck L. 1994. Relationship between frontal lobe lesions and Wisconsin card sorting test performance in patients with multiple sclerosis. *Neurology* **44:** 420–425.

Bachman, J. C. 1961. Specificity vs. generality in learning and performing two large muscle motor tasks. *Res. Q.,* **32**(1): 3–11.

Baloh, R. W., Jacobson, K. M., Beykirch, K., and Honrubia, V. 1998. Static and dynamic posturography in patients with vestibular and cerebellar lesions. *Arch. Neurol.* **55:** 649–654.

Barr, H. M., Streissguth, A. P., Darby, B. L., and Sampson, P. D. 1990. Prenatal exposure to alcohol, caffeine, tobacco, and aspirin: Effects on fine and gross motor performance in four-year-old children. *Dev. Psychol.* **26:** 339–348.

Benton, A. L. 1968. Differential behavioral effects in frontal lobe disease. *Neuropsychologia* **6:** 53–60.

Bookstein, F. L. 1991. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge Univ. Press, New York.

Bookstein, F. L. 1996a. Biometrics, biomathematics, and the morphometric synthesis. *Bull. Math. Biol.* **58:** 313–365.

Bookstein, F. L. 1996b. Combining the tools of geometric morphometrics. In *Advances in Morphometrics* (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, D. E. Slice, Eds.), NATO ASI Series A: Life Sciences, Vol. 284, pp. 131–151. Plenum, New York.

Bookstein, F. L. 1996c. Morphometrics. *Math Horizons* **February:** 28–31.

Bookstein, F. L. 1997a. Shape and the information in medical images: A decade of the morphometric synthesis. *Comput. Vis. Image Understand.* **66:** 97–118.

Bookstein, F. L. 1997b. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *Med. Image Anal.* **1:** 225–243.

Bookstein, F. L. 1998. A hundred years of morphometrics. *Acta Zool.* **44:** 7–59.

Bookstein, F. L. 1999. Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability. In *Brain Warping* (A. W. Toga, Ed.), pp. 157–182. Academic Press, San Diego.

Bookstein, F. L., Streissguth A. P., Sampson, P. D., and Barr, H. M. 1996. Exploiting redundant measurement of dose and behavioral outcome: New methods from the behavioral teratology of alcohol. *Dev. Psychol.* **32:** 404–415.

Bookstein, F. L., Green, W. D. K. 1998. *Edgewarp 3D: A Preliminary Manual.* Posted to the internet as ftp://brainmap.med.umich.edu/pub/edgewarp3.1/manual.html, August, 1998.

Bookstein, F. L., Sampson, P. D., Streissguth, A. P., and Connor, P. D. 2001. Geometric morphometrics of corpus callosum and subcortical structures in the fetal-alcohol-affected brain. *Teratology* **64:** 4–32.

Bullmore, E. T., Suckling, O., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. 1999. Global, voxel, and cluster tests, by theory and permutation for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imag.* **18:** 32–42.

Coles, C. D., Platzman, K. A., Raskind-Hood, C. L., Brown, R. T., Falek, A., and Smith, I. E. 1997. A comparison of children affected by prenatal alcohol exposure and attention deficit, hyperactivity disorder. *Alcoholism: Clin. Exp. Res.* **21:** 150–161.

Connor, P. D., Streissguth, A. P., Sampson, P. D., Bookstein, F. L., and Barr, H. M. 1999. Auditory and visual attention in adults with and without fetal alcohol syndrome and fetal alcohol effects. *Alcoholism: Clin. Exp. Res.* **23:** 1395–1402.

Connor, P. D., Sampson, P. D., Bookstein, F. L., Barr, H. A., and Streissguth, A. P. 2000. Direct and indirect effects of prenatal alcohol damage on executive function. *Dev. Neuropsychol.* **18:** 331–354.

Delis D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. 1987. *California Verbal Learning Test*. Psychological Corp., San Antonio, TX.

Denckla, M. B. 1974. Development of motor co-ordination in normal children. *Dev. Med. Child Neurol.* **16:** 729–741.

Denckla, M. B. 1996. A theory and model of executive function: A neuropsychological perspective. In *Attention, Memory, and Executive Function* (G. R. Lyon and N. A. Krasnegor, Eds.), pp. 263–278. Brookes, Baltimore.

Dryden, I. L., and Mardia K. V. 1998. *Statistical Shape Analysis.* Wiley, Chichester, UK.

Dugbartey, A. T., Rosenbaum, J. G., Sanchez, P. N., and Townes, B. D. 1999. Neuropsychological assessment of executive functions. *Semin. Clin. Neuropsychiatry* **4:** 5–12.

Eliassen, J. C., Baynes, K., and Gazzaniga, M. S. 2000. Anterior and posterior callosal contributions to simultaneous bimanual movements of hands and fingers. *Brain* **123:** 2501–2511.

Eslinger, P. J. 1996. Conceptualizing, describing, and measuring components of executive function: A summary. In *Attention, Memory, and Executive Function* (G. R. Lyon, and N. A. Krasnegor, Eds.), pp. 367–395. Brookes, Baltimore.

Golden, G. J. 1978. *Stroop Color and Word Test*. Stoelting, Chicago.

Good, P. 2000. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses,* 2nd ed. Springer-Verlag, New York.

Harris, J. C. 1998. *Developmental Neuropsychiatry*. Oxford Univ. Press, New York.

Heaton, R. K., Curtiss, G., and Tuttle, K. 1993. *Wisconsin Card Sorting Test: Computer Version-2*. Psychological Assessment Resources, Lutz, FL.

Heindel, W. C., Salmon, D. P., Shults, C. W., Walicke, P. A., and Butters, N. 1989. A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *J. Neurosci.* **9:** 582–587.

Iacoboni, M., Zaidel, E. 1995. Channels of the corpus callosum: Evidence from simple reaction times to lateralized flashes in the normal and the split brain. *Brain* **118:** 779–788.

Jacobson, S. W. 1998. Specificity of neurobehavioral outcomes associated with prenatal alcohol exposure. *Alcoholism: Clin. Exp. Res.* **22:** 313–320.

Jastak, S., and Wilkinson, G. S. 1984. *Manual for the Wide Range Achievement Test Revised*. Jastak Associates, Wilmington, DE.

Kent, J. T., and Mardia, K. V. 2001. Shape, procrustes tangent projections and bilateral symmetry. *Biometrika* **88:** 469–485.

Kerns, K. A., Don, A., Mateer, C. A., and Streissguth, A. P. 1997. Cognitive deficits in nonretarded adults with fetal alcohol syndrome. *J. Learn. Disabil.* **30:** 685–693.

Kodituwakku, P. W., Handmaker, N. S., Cutler, S. K., Weathersby, E. K., and Handmaker, S. D. 1995. Specific impairments in self-regulation in children exposed to alcohol prenatally. *Alcoholism: Clin. Exp. Res.* **19:** 1558–1564.

Kopera-Frye, K., Dehaene, S., and Streissguth, A. P. 1996. Impairments of number processing induced by prenatal alcohol exposure. *Neuropsychologia* **34:** 1187–1196.

Mardia, K. V., Bookstein, F. L., and Moreton, I. J. 2000. Statistical assessment of bilateral symmetry of shapes. *Biometrika* **87:** 285–300.

Mardia, K. V., Kent, J., and Bibby J. 1979. *Multivariate Analysis.* Academic Press, London.

Matthews, C. G., and Klove, H. 1978. *Wisconsin Motor Steadiness Battery: Administration Manual for Child Neuropsychology Battery*. University of Wisconsin Medical School, Neuropsychology Laboratory, Madison, WI.

Mattson, S. N., Jernigan, T. L., and Riley, E. P. 1994. MRI and prenatal alcohol exposure: Images provide insight into FAS. *Alcohol Health Res. World* **18:** 49–52.

Mattson, S. N., Riley, E. P., Gramling, L. S., Delis, D. C., and Jones, K. L. 1997. Heavy prenatal alcohol exposure with or without physical features of fetal alcohol syndrome leads to IQ deficits. *J. Pediatr.* **131:** 718–721.

Mattson, S. N., and Riley, E. P. 1997a. Brain anomalies in fetal alcohol syndrome. In *Fetal Alcohol Syndrome: From Mechanisms to Prevention* (E. L. Abel, Ed.). CRC Press, Boca Raton, FL.

Mattson, S. N., and Riley, E. P. 1997b. Neurobehavioral and neuroanatomical effects of heavy prenatal exposure to alcohol. In *The Challenge of Fetal Alcohol Syndrome: Overcoming Secondary Disabilities* (A. P. Streissguth and J. Kanter, Eds.), pp. 3–14. Univ. of Washington Press, Seattle.

Mattson, S. N., Riley, E. P., Gramling, L. S., Delis, D. C., and Jones, K. L. 1998. Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. *Neuropsychology* **12:** 146–153.

Mattson, S. N., Goodman, A. M., Caine, C., Delis, D. C., and Riley, E. P. 1999. Executive functioning in children with heavy prenatal alcohol exposure. *Alcoholism: Clin. Exp. Res.* **23:** 1808–1815.

McIntosh, A. R., Bookstein, F. L., Haxby, J., and Grady, C. 1996. Multivariate analysis of functional brain images using partial least squares. *NeuroImage* **3:** 143–157.

Meyers, J. E., and Meyers, K. R. 1995. *Rey Complex Figure Test and Recognition Trial: Professional Manual.* Psychological Assessment Resources, Odessa, FL.

Milner, B. 1965. Visually-guided maze learning in man: Effects of bilateral hippocampal, bilateral frontal, and unilateral cerebral lesions. *Neuropsychologia* **3:** 317–338.

Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., and Kellam, S. G. 1991. Analysis of the elements of attention: A neuropsychological approach. *Neuropsychol. Rev.* **2:** 109–145.

Olson, H. C., Feldman, J., Streissguth, A. P., Sampson, P. D., and Bookstein, F. L. 1998. Neuropsychological deficits in adolescents with fetal alcohol syndrome: Clinical findings. *Alcoholism: Clin. Exp. Res.* **22:** 1998–2012.

Pellegrino, J. W., Hunt, E., Abate, R., and Farr, S. 1987. A computer-based battery for the assessment of static and dynamic spatial reasoning abilities. *Behav. Res. Meth. Inst. Comp.* **19:** 231–236.

Peterson, L. R., and Peterson, M. J. 1959. Short-term retention of individual verbal items. *J. Exp. Psychol.* **58:** 193–198.

Reitan, R. M., and Davison, L. A. 1974. *Clinical Neuropsychology: Current Status and Applications*. Winston/Wiley, New York.

Reitan, R. M., and Wolfson, D. 1985. *The Halstead–Reitan Neuropsychological Test Battery*. Neuropsychology Press, Tucson.

Reyment, R. A., and Jöreskog, K. G. 1993. *Applied Factor Analysis in the Natural Sciences*. Cambridge Univ. Press, Cambridge.

Riley, E. P., Mattson, S. N., Sowell, E. R., Jernigan, T. L., Sobel, D. F., and Jones, K. L. 1995. Abnormalities of the corpus callosum in children prenatally exposed to alcohol. *Alcoholism: Clin. Exp. Res.* **19:** 1198–1202.

Romano, J. P. 1990. On the behavior of randomization tests without a group invariance assumption. *J. Am. Stat. Assoc.* **55:** 686–692.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., and Beck, L. N. 1956. A continuous performance test of brain damage. *J. Consult. Psychiatry* **20:** 343–350.

Ruff, R. M., Evans, R., and Marshall, L. F. 1986. Impaired verbal and figural fluency after head injury. *Arch. Clin. Neuropsychol.* **1:** 87–101.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., Little, R. E., Clarren, S. K., Dehaene, P., Hanson, J. W., and Graham, J. M. 1997a. Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology* **56:** 317–326.

Sampson, P. D., Kerr, F. B., Olson, H. C., Streissguth, A. P., Hunt, E., Barr, H. M., Bookstein, F. L., and Thiede, K. 1997b. The effects of prenatal alcohol exposure on adolescent cognitive performance: A speed–accuracy tradeoff. *Intelligence* **24:** 329–353.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., and Barr, H. M. 2000. On categorizations in analyses of alcohol teratogenesis. *Environ. Health Perspect.* **108**(Suppl.3): 421–428.

Shallice, T., and Burgess, P. W. 1991. Deficits in strategy application following frontal lobe damage in man. *Brain* **114:** 727–741.

Sohlberg, M. M., and Mateer, C. M. 1989. *Introduction to Cognitive Rehabilitation: Theory and Practice*. Guilford, New York.

Sowell, E. R., Mattson, S. N., Thompson, P. M., Jernigan, T. L., Riley, E. P., and Toga, A. W. 2001a. Mapping callosal morphology and cognitive correlates. *Neurology* **57:** 235–244.

Sowell, E. R., Thompson, P. M., Mattson, S. N., Tessner, K. D., Jernigan, T. L., Riley, E. P., and Toga, A. W. 2001b. Voxel-based morphometric analyses of the brain in children and adolescents prenatally exposed to alcohol. *Cogn. Neurosci. Neuropsychol.* **12:** 515–523.

Stratton, K., Howe, C., and Battaglia, F., (Eds.) 1996. *Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment*. Institute of Medicine, U.S. National Academy of Sciences, National Academy Press, Washington, DC.

Streissguth, A. P., Barr, H. M., Sampson, P. D., Bookstein, F. L., and Darby, B. L. 1989. Neurobehavioral effects of prenatal alcohol (three parts.) *Neurotoxicol. Teratol.* **11:** 461–507.

Streissguth, A. P., Aase, F. M., Clarren, S. K., Randels, S. P., LaDue, R. A., and Smith, D. F. 1991. Fetal alcohol syndrome in adolescents and adults. *J. Am. Med. Assoc.* **265:** 1961–1967.

Streissguth, A. P., Bookstein, F. L., Sampson, P. D., and Barr, H. M. 1993. *The Enduring Effects of Prenatal Alcohol Exposure on Child Development, Birth through Seven Years: A Partial Least Squares Solution*. University of Michigan Press, Ann Arbor.

Streissguth, A. P., Sampson, P. D., Olson, H. C., Bookstein, F. L., Barr, H. M., Scott, M., Feldman, J., and Mirsky, A. F. 1994a. Maternal drinking during pregnancy: Attention and short-term memory in 14-year-old offspring—A longitudinal prospective study. *Alcoholism: Clin. Exp. Res.* **18:** 202–218.

Streissguth, A. P., Barr, H. M., Olson, H. C., Sampson, P. D., Bookstein, F. L., and Burgess, D. M. 1994b. Drinking during pregnancy decreases Word Attack and Arithmetic scores on standardized tests: Adolescent data from a population-based prospective study. *Alcoholism: Clin. Exp. Res.* **18:** 248–254.

Streissguth, A. P., Barr, H. M., Kogan, J., and Bookstein, F. L. 1996. *Understanding the Occurrence of Secondary Disabilities in Clients with Fetal Alcohol Syndrome (FAS) and Fetal Alcohol Effects (FAE)*, final report, CDC Grant CCR008515. Fetal Alcohol and Drug Unit, University of Washington School of Medicine, Seattle, WA.

Streissguth, A. P., Barr, H. M., Kogan, J., and Bookstein, F. L. 1997. Secondary disabilities in people with FAS and FAE. In *The Challenge of Fetal Alcohol Syndrome: Overcoming Secondary Disabilities* (A. P. Streissguth and J. Kanter, Eds.), pp. 203–221. University of Washington Press, Seattle.

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Press, S., and Sampson, P. D. 1998. A fetal alcohol behavior scale. *Alcoholism: Clin. Exp. Res.* **22:** 325–333.

Sulik, K. K., Johnson, M. C., and Webb, M. A. 1981. Fetal alcohol syndrome: Embryogenesis in a mouse model. *Science* **214:** 936–938.

Swayze, V. W., II, Johnson, V. P., Hanson, J. W., Piven, J., Sato, Y., Giedd, J. N., Mosnik, D., and Andreasen, N. C. 1997. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. *Pediatrics* **99:** 232–240.

Talland, G. A. 1965. *Deranged Memory*. Academic Press, New York.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., and Toga, A. W. 2000. Growth patterns in the developing human brain detected by using continuum-mechanical tensor maps. *Nature* **404:** 190–193.

Wechsler, D. 1981. *Wechsler Adult Intelligence Scale—Revised*. Psychological Corp., San Antonio, TX.

Woodcock, R. W. 1987. *Woodcock Reading Mastery Tests Revised*. American Guidance Service, Circle Pines, NM.