Alcohol & Alcoholism Vol. 36, No. 2, pp. 147–159, 2001

MEASURING THE FACIAL PHENOTYPE OF INDIVIDUALS WITH PRENATAL ALCOHOL EXPOSURE: CORRELATIONS WITH BRAIN DYSFUNCTION

SUSAN J. ASTLEY[1,2]* and STERLING K. CLARREN[2]

[1]Department of Epidemiology, School of Public Health and Community Medicine and [2]Department of Pediatrics, School of Medicine, University of Washington, Seattle, WA 98105, USA

(*Received* 10 *August* 2000; *in revised form* 16 *October* 2000; *accepted* 30 *October* 2000)

Abstract — The purpose of this report is to demonstrate how to measure the magnitude of expression of the fetal alcohol syndrome (FAS) facial phenotype using the new 4-Digit Diagnostic Code and the previously developed D-score and to demonstrate how these two measures of the FAS facial phenotype correlate with brain function and structure; correlations that fail to be identified by the older gestalt method of facial measurement. The D-score and the facial component of the 4-Digit Diagnostic Code quantitatively measure the magnitude of expression of the FAS facial phenotype using three facial features (palpebral fissure length, philtrum smoothness, and upper lip thinness). These facial measurement systems were developed by the Washington State FAS Diagnostic and Prevention Network (FAS DPN) of clinics and are used to screen and diagnose the facial component of FAS for all patients evaluated in the network of clinics (1500 to date). The 4-Digit Diagnostic Code is a comprehensive diagnostic system developed by the FAS DPN in 1997 to diagnose the full spectrum of outcomes among patients with prenatal alcohol exposure. The four digits reflect the magnitude of expression of the four key diagnostic features of FAS in the following order: (1) growth deficiency; (2) the FAS facial phenotype; (3) brain dysfunction; (4) gestational alcohol exposure. The 4-Digit Diagnostic Code was developed to overcome the subjective, highly variable gestalt method of diagnosis that has been used as the standard to date, worldwide. Prior to the development of the 4-Digit Diagnostic Code, the first 445 patients evaluated in the FAS DPN were diagnosed using the gestalt method. For research purposes, their gestalt diagnoses were transformed into 4-Digit Diagnostic Codes, presenting a unique opportunity to directly compare the two diagnostic methods. When the facial phenotype was measured using the 4-Digit Diagnostic Code or D-score, the magnitude of expression of the FAS facial phenotype was significantly correlated with structural, neurologic, and functional measures of brain damage, and the phenotype of those receiving a 4-Digit Diagnosis of FAS showed little variability. When the gestalt method of diagnosis was used, the magnitude of expression of the FAS facial phenotype did not correlate with structural, neurologic and functional measures of brain damage, and the facial phenotype of those receiving a gestalt diagnosis of FAS was highly variable. The 4-Digit Diagnostic Code and D-score thus provide more precise and accurate measures of the FAS facial phenotype and reveal important correlations with brain structure and function, suggesting that intermediate expressions of the FAS facial phenotype may serve as important risk factors for brain damage caused by prenatal alcohol exposure.

INTRODUCTION

The fetal alcohol syndrome (FAS) is a permanent birth defect syndrome caused by maternal consumption of alcohol during pregnancy. The syndrome has been broadly characterized by pre- and/or postnatal growth deficiency, a characteristic set of minor facial anomalies, central nervous system (CNS) dysfunction, and prenatal alcohol exposure (Jones and Smith, 1973; Clarren and Smith, 1978; Rosett, 1980; Sokol and Clarren, 1989; Stratton *et al*., 1996). In 1997, a new more objective and comprehensive, case-defined method for diagnosing the full spectrum of outcomes in individuals with prenatal alcohol exposure was created, called the 4-Digit Diagnostic Code (Astley and Clarren, 1997, 1999, 2000). The four digits of the diagnostic code reflect the magnitude of expression of four key diagnostic features of FAS in the following order: (1) growth deficiency; (2) the FAS facial phenotype; (3) brain damage/dysfunction; (4) gestational alcohol exposure. Each is ranked independently on a four-point Likert scale, with 4 reflecting severe expression of the feature and 1 reflecting no expression of the feature. The 4-Digit Diagnostic Code is generated by first recording key clinical data on the standardized FAS diagnostic evaluation form and following specific case-definitions to generate each of the four digits (Astley and Clarren, 1999, 2000). The concept of developing a more objective diagnostic system began with the development of a more objective, empirically derived method for measuring and case-defining the FAS facial phenotype (Astley and Clarren, 1996). A discriminant analysis was used to identify the cluster of minor anomalies and their magnitude of expression that best differentiated 42 individuals with FAS from 84 matched controls without FAS. Three features were identified (reduced palpebral fissure/inner canthal distance (ICD) ratio, smooth philtrum, and a thin upper lip). A discriminant equation was generated from the study demonstrating that, when the magnitude of expression of these features results in a discriminant score (D-score) ≥ 0.80, the facial phenotype was 100% sensitive and specific to FAS. The photographs used to develop this D-score method of facial analysis were obtained from retrospective sources, thus they did not include internal measures of scale for deriving the true palpebral fissure length (PFL) and ICD. The ratio PFL/ICD served as a proxy measure of the true lengths of each feature. Later analyses of direct measures of PFL and ICD among patients seen in the FAS Diagnostic and Prevention Network (FAS DPN) demonstrated that the key differentiating feature was PFL, not ICD. Like the D-score method, the 4-Digit Diagnostic Code uses PFL, philtrum smoothness, and upper lip thinness to define the FAS facial phenotype. Unlike the D-Score, a true measure, rather than a proxy measure, of PFL, is obtained by either direct measurement or placement of an internal measure of scale in the clinical photograph. The 4-Digit Diagnostic Code method for documenting the magnitude of expression of the FAS facial phenotype serves as both a diagnostic and screening tool. The 4-Digit Diagnostic Code facial rank has and

*Author to whom correspondence should be addressed at: Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, CH-47, Seattle, WA 98105, USA.

continues to be used to diagnose the facial component of the syndrome in all patients receiving a FAS diagnostic evaluation in the Washington State FAS DPN of clinics (n = 1500 to date). It is also being used to screen all children entering long-term foster care in King County, WA and all residents in a Washington State juvenile rehabilitation facility.

The present report illustrates how to measure the magnitude of expression of the FAS facial phenotype using the 4-Digit Diagnostic Code and D-score, and demonstrates how the 4-Digit Diagnostic Code and D-score measures of the FAS facial phenotype correlate with brain function and structure; correlations that fail to be identified by the standard gestalt method of diagnosis and facial measurement (Rosett, 1980; Sokol and Clarren, 1989).

SUBJECTS AND METHODS

Study population

Data for this study came from all patients evaluated in the FAS DPN who met the following inclusion criteria: (1) had a confirmed history of prenatal alcohol exposure (Alcohol 4-Digit Diagnostic Code rank = 3 or 4) (Astley and Clarren, 2000) and (2) consented to allow the FAS DPN to use their diagnostic data for research purposes. This study was approved by the University of Washington Human Subjects Division.

FAS facial phenotype

Three features (PFL, philtrum smoothness, and upper lip thinness) are measured to document the magnitude of expression of the FAS facial phenotype (Astley and Clarren, 1996). All other major and minor craniofacial anomalies are measured and recorded for clinical and research purposes, but are not used to rank the magnitude of expression of the FAS facial phenotype. PFL is the distance from the endocanthion to the exocanthion (Fig. 1). The philtrum furrow is the vertical groove extending from the midline of the upper lip to the nose (Fig. 2). The upper lip thinness refers to the area demarcated by the vermilion border (Fig. 2). These three features are measured directly by a physician or measured from a digital photograph using image analysis software. PFL is measured in mm and transformed to a standardized z-score using appropriate published normal anthropometric charts (Iosub *et al*., 1985; Thomas *et al*., 1987; Hall *et al*., 1989). The z-score reflects how many SD above or below the population norm the patient's PFL is, based on the patient's age. The z-score is defined as the patient's PFL minus the mean PFL for the normal population divided by the SD of the mean PFL for the normal population. Philtrum smoothness and upper lip thinness are measured on five-point Likert scales using the pictorial lip–philtrum guide (Fig. 3) (Astley and Clarren, 2000). This method for measuring the facial phenotype directly or photographically is demonstrated on a CD ROM with the aid of animations and video (Astley *et al*., 1999*a*).

Direct measurement of facial features

PFLs are measured to the nearest mm with a clear plastic ruler (1 cm by 14 cm in size) held as closely as possible to the eye without touching the eye or eye lashes (Fig. 1). The FAS DPN chooses not to use calipers, because the patients are often too young and active to co-operate safely. The individual is asked to open his/her eyes widely to allow accurate identification of the endocanthion and exocanthion landmarks. The PFL is compared to the normal PFL for age by using a racially appropriate normal chart for PFL to compute the z-score for the right and left PFLs.

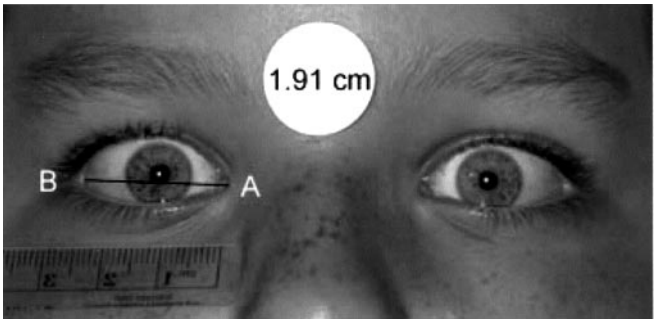

Fig 1. Palpebral fissure length.
Palpebral fissure length (PFL) is measured from the endocanthion (A) to the exocanthion (B). It can be measured directly using a clear plastic centimetre ruler, or it can be measured from a photograph with an internal measure of scale (adhesive paper sticker) placed between the eyebrows or a centimetre ruler placed below the eye.

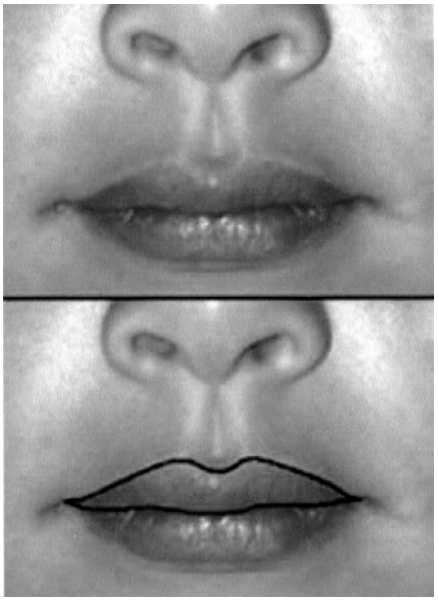

Fig 2. Philtrum and upper lip.
The philtrum is the vertical groove extending between the nose and the vermilion border of the upper lip. The smoothness of the philtrum and thinness of the upper lip (demarcated by its vermilion border) are measured by selecting the photograph from the lip–philtrum guide (Fig. 3) that best matches each feature independently. Upper lip thinness can also be measured from a digital photograph viewed on a computer monitor by tracing the outline of the vermilion border with a mouse and requesting image analysis software like Sigma Scan Pro to compute a measure called circularity (perimeter^2/area) (lower photograph). The thinner the upper lip, the larger the circularity (Fig. 3). The circularity scores of the five lips pictured on the lip–philtrum guide (Fig. 3) assist the physician in selecting the picture that best matches the patient's upper lip thinness. The lip pictured has a circularity of 40.5 and therefore would receive a rank of 1.

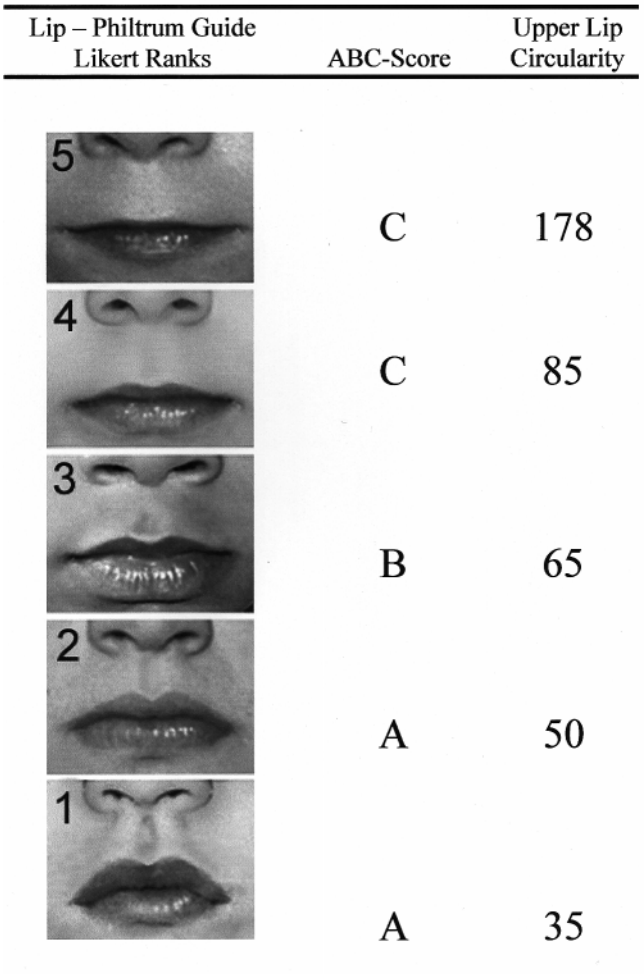

Fig 3. Lip–philtrum guide
Pictorial examples of the five-point Likert scale, upper lip circularity scale, and the ABC scale used to rank upper lip thinness and philtrum smoothness. Circularity (perimeter^2/area) is a continuous measure of upper lip thinness that can be used to facilitate the ranking of upper lip thinness (Fig. 2). It is important that the individual's lips are gently closed with no smile (see Fig. 5).

Philtrum smoothness and upper lip thinness are measured on five-point Likert pictorial scales by holding the lip–philtrum guide next to the patient's face and assigning each feature the Likert rank of the photograph that best matches each feature (Figs 3 and 4). Philtrum smoothness and upper lip thinness are ranked independently of one another. For example, a child could present with a rank 5 philtrum and rank 1 upper lip. The physician's eyes must be aligned in the patient's Frankfort horizontal plane (demarcated by a line drawn through the patient's auditory meatus and the lowest border of the bony orbital rim) (Figs 4 and 6). If the physician's eyes are above or below this plane, the upper lip can appear thinner or thicker respectively than it is. The patient must have a relaxed facial expression with no smile and lips gently closed. A smile can cause the philtrum and upper lip to appear smoother and thinner than they are (Fig. 5).

Although the ICD (right endocanthion to left endocanthion) is not used as a diagnostic feature of the FAS facial phenotype, it is still measured to document the presence of hypo- or hypertelorism, and it is used as the denominator of the proxy measure of PFL (namely PFL/ICD), used to compute the facial D-score when a true measure of PFL is not obtainable. It is measured with a clear plastic ruler and transformed into a *z*-score using a racially appropriate normal chart for ICD.

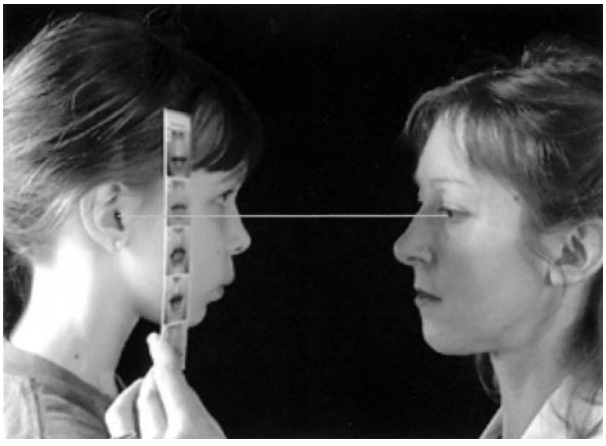

Fig. 4. Facial alignment and expression.
Physician aligned in the patient's Frankfort horizontal plane while using the lip–philtrum guide to rank upper lip thinness and philtrum smoothness. The Frankfort horizontal plane is defined by a line that passes through the patient's auditory meatus (or the upper edge of the tragus when viewed from the front, see Fig. 6) and the lowest border of the bony orbital rim (orbitale). The physician's eyes (or camera lens) should be directly in line with the patient's Frankfort horizontal plane.

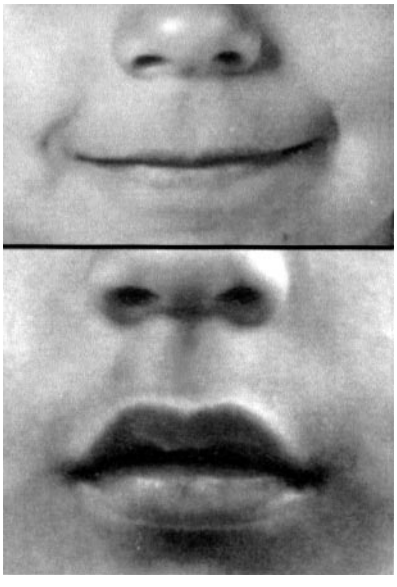

Fig. 5. Impact of a smile on lip and philtrum measures.
This is the same individual with (upper photograph) and without (lower photograph) a smile demonstrating how a smile can erroneously transform a deeply grooved philtrum (Likert rank = 2) and full upper lip (Likert rank = 1, lip circularity = 41) into a smooth philtrum (Likert rank = 4) and thin upper lip (Likert rank 5, lip circularity = 191) (Astley and Clarren, 1996). Circularity (perimeter^2/area) is a continuous measure of upper lip thinness that can be used to facilitate the ranking of upper lip thinness (Figs 2 and 3).

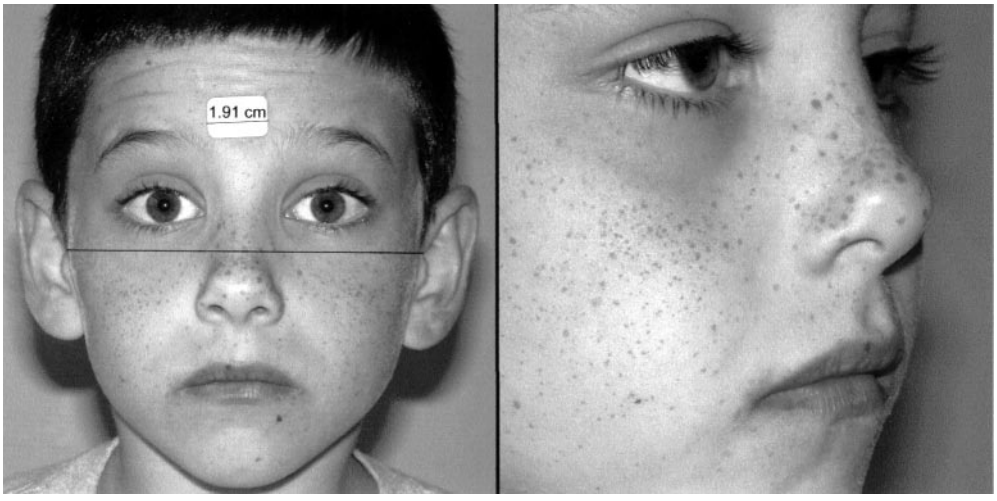

Fig. 6. Standardized facial photographs.

Two standardized facial photographs are obtained (frontal, ¾ view) to measure the facial phenotype of FAS. Eyes should be fully open, no eyeglasses, no smile, lips gently closed, and an internal measure of scale placed between the eyebrows. The right and left ears should be equally visible to ensure accurate measurement of the palpebral fissure lengths and inner canthal distance. An imaginary line drawn from the top of the left and right tragus should fall along the patient's lower bony orbital rims, confirming that the camera is aligned in the patient's Frankfort horizontal plane (see Fig. 4). The ¾ view is obtained to facilitate ranking the philtrum. It is particularly important if the camera has a centrally mounted flash that can diminish the appearance of the philtrum depth in a frontal photograph.

Photographic measurement of facial features

An internal measure of scale is placed on the patient's forehead between the eyebrows (Figs 1 and 6). A small, adhesive paper sticker ½ in. to ¾ in. in size serves well and can be purchased from an office supply store. A frontal and ¾ view photograph of the patient's face is obtained using a digital or 35-mm camera. Polaroid cameras do not provide sufficient image resolution. A close-up photograph is taken, such that the patient's head fills the entire frame (Fig. 6). When using a digital camera, a minimum of 3 megapixel resolution is recommended. The lens of the camera is placed in line with the patient's Frankfort horizontal plane, as described above and illustrated in Fig. 4. To judge the Frankfort horizontal plane when viewing the face through the camera, an imaginary line drawn between the upper border of the left and right tragus should fall across the left and right lower bony orbital rim (Fig. 6). There should also be no left-to-right rotation of the image; both ears should be equally visible in the frontal photograph. The facial expression should be relaxed with no smile, lips gently closed, eyes wide open, and no eyeglasses. The ¾ view is taken to facilitate ranking philtrum smoothness by purposely driving a flash of light across the philtrum to see if a shadow is cast. The ¾ view is particularly important to obtain if the camera has a centrally mounted flash that can diminish the appearance of a grooved philtrum in a frontal photograph. Properly aligned facial photographs are obtained in the FAS DPN clinics with a handheld camera and freestanding patient. Stereotaxic equipment and tripods are not necessary.

The digital image is measured using image analysis software (e.g. Sigma Scan Pro 5, 1999 or FAS DPN software to be distributed in 2001). This software allows one to enlarge the image, enhance the exposure if necessary, make all the necessary measurements, and store the data in an electronic database. If the image is obtained with a 35-mm camera, the slide, print or negative is scanned to generate a digital copy of the image. It is important to note that the resolution (or clarity) of a scanned image as small as a slide or negative may not be sufficiently high. The right and left PFLs are measured by clicking the mouse on the endocanthion and exocanthion landmarks and having Sigma Scan Pro compute the distance between the landmarks in pixels (dots of light on the computer monitor). The length of the internal measure of scale (paper sticker) is also measured in pixels. The real size of the PFL (mm) is computed from the PFL (pixels), the length of the paper sticker (pixels) and the real length of the paper sticker (mm) using the following equation:

$$\text{PFL (mm)} = [(\text{length of sticker, mm}/\text{length of sticker, pixels}) \times (\text{PFL, pixels})] \times 1.07$$

If the image is rotated right or left, insert the mean PFL in pixels into the equation to compute a mean PFL in mm. The margin of error between this mean PFL (mm) and the true mean PFL measured directly with calipers is less than 1%.

The 1.07 adjustment factor is included in the formula to increase the computed PFL by 7% to adjust for the foreshortening effect of measuring a facial feature that is slightly off the midline of the photograph (Farkas, 1994). This adjustment was confirmed to be accurate by comparing computed PFLs from photographs with measures obtained directly from the subjects with calipers. The computed PFL (mm) is transformed into a *z*-score, as described above, to standardize it to the popuation norm. The PFL can also be computed by placing a clear plastic ruler directly under the eye prior to taking the facial photograph (Fig. 1). The actual PFL (mm) would be computed using the equation above without the adjustment factor.

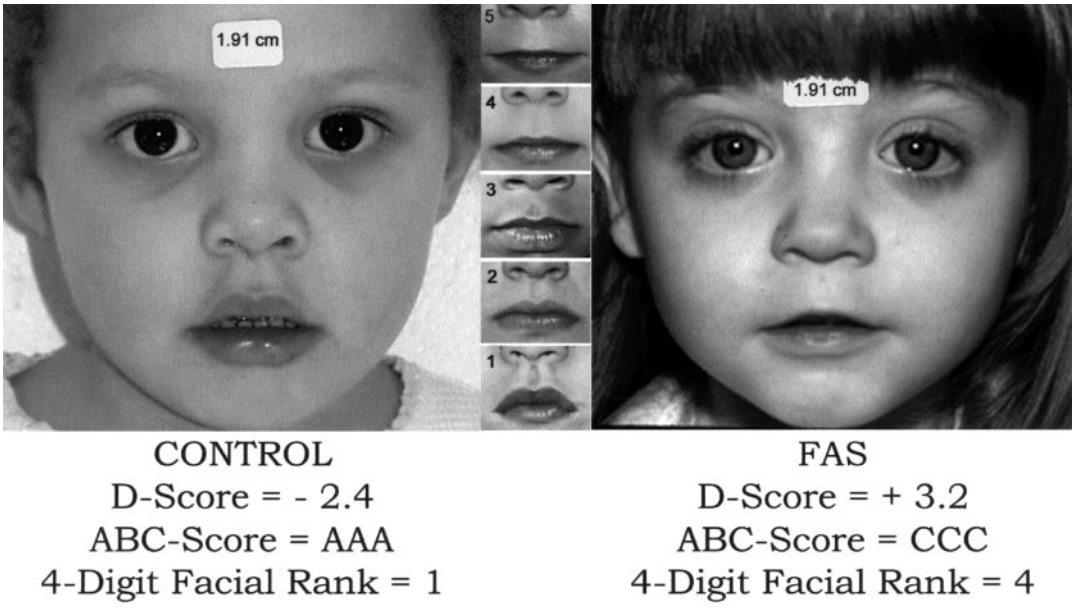

Fig. 7. Facial D-score and 4-Digit Diagnostic Code rank.

Example of the facial D-scores and 4-Digit Diagnostic Code facial ranks of a control child and a child with the facial phenotype of FAS. The facial D-score reflects the magnitude of expression of the FAS facial phenotype and is computed when an internal measure of scale is not in the photograph. A D-score ≥ 0.8 is screen-positive for the face of FAS. The D-score is computed using the palpebral fissure length (PFL)/inner canthal distance ratio and the five-point Likert ranks of philtrum smoothness and upper lip thinness using the lip–philtrum guide pictured in the centre. The 4-Digit Diagnostic Code facial rank also reflects the magnitude of expression of the FAS facial phenotype on a four-point Likert scale. It is computed when the true PFL is available or can be computed. The first step is to generate the ABC-score reflecting the size of the PFLs, philtrum smoothness, and upper lip thinness in that order. Tables 1 and 2 are used to generate the ABC-score and transform it into the 4-Digit Diagnostic facial rank. A facial rank of 4 is a facial phenotype of FAS.

The real size of the ICD (mm) is computed from the ICD (in pixels), the length of the paper sticker (pixels) and the real length of the paper sticker (mm) using the following equation:

$$\text{ICD (mm)} = (\text{length of sticker, mm}/\text{length of sticker, pixels}) \times (\text{ICD, pixels})$$

No adjustment factor is added to the equation, because the sticker and ICD are on the midline, thus there is no foreshortening error.

Philtrum smoothness and upper lip thinness are measured using the lip–philtrum guide described above. Philtrum smoothness is ranked by holding the lip–philtrum guide next to the image on the computer monitor and selecting the picture that best matches the patient's philtrum. Upper lip thinness is measured by tracing the outline of the vermilion border with the mouse and having Sigma Scan Pro compute a measure called circularity (perimeter^2/area) (see Fig. 2). Some image analysis software programs call circularity 'compactness'. Circularity ranges from 12.8 for a circle to infinity as the circle is squashed into a line (or becomes thinner). The thinner the upper lip, the larger the circularity (Fig. 3). The circularity scores of the five lips pictured on the lip–philtrum guide allow the physician to select the picture that best matches the patient's upper lip thinness. The process of taking a facial photograph and measuring the features takes about 10 min.

Computing the facial D-score

The facial D-score is computed when a true measure of PFL cannot be obtained (e.g. home photographs or retrospective photograph sets that did not contain an internal measure of scale). The facial D-score is computed using the equation:

$$\text{D-score} = 0.7408 - [5.7337 \times (\text{PFL/ICD})] + (1.1677 \times \text{philtrum five-point Likert rank}) + (0.1587 \times \text{upper lip five-point Likert rank})$$

A facial phenotype with a D-score ≥ 0.8 is classified as screen-positive for the facial phenotype of FAS (Fig. 7). This discriminant function and cut off value differentiated 42 patients with FAS from 84 controls with 100% sensitivity and specificity in an earlier study (Astley and Clarren, 1996).

Examples of facial D-scores for a control child and child with the FAS facial phenotype are presented in Fig. 7. The D-score for the 2.8-year-old control child was –2.4 = 0.748 –[5.7337 × (117 pixels/150 pixels)] + (1.1677 × 1) + (0.1587 × 1). The D-score for the 2.1 year-old child with the FAS face was + 3.2 = 0.748 –[5.7337 × (105 pixels/143 pixels)] + (1.1677 × 5) + (0.1587 × 5).

Computing the facial 4-Digit Diagnostic Code rank

The facial 4-Digit Diagnostic Code rank is the most accurate diagnostic measure of the magnitude of expression of

Table 1. 4-Digit Diagnostic Code for facial phenotype rank[a]

(A) Five-point Likert Scale for philtrum and lip	*z*-score for mean PFL	Circle the ABC-scores for: Palpebral fissure	Philtrum	Upper lip
4 or 5	≤ –2 SD	**C**	C	C
3	> –2 SD and ≤ –1 SD	B	B	**B**
1 or 2	> –1 SD	A	**A**	A

(B) 4-Digit Diagnostic Code rank	Level of expression of FAS facial phenotype	Palpebral fissure–philtrum–lip ABC-score combinations
4	Severe	CCC
3	Moderate	CCB, CBC, BCC
		CCA, CAC, CBB, CBA, **CAB**, CAA
2	**Mild**	BCB, BCA, BBC, BAC
		ACC, ACB, ACA, ABC, AAC
1	Absent	BBB, BBA, BAB, BAA
		ABB, ABA, AAB, AAA

[a]Case definitions used to define the 4-Digit Diagnostic Code ranks for FAS facial phenotype (Astley and Clarren, 2000).
(A) The first step in deriving the 4-Digit Diagnostic Code rank for the facial phenotype is to derive the facial ABC score. For example, if a patient's palpebral fissure lengths were > 2 SD below the norm and their philtrum and upper lip received Likert scores of 2 and 3 respectively (see Fig. 3), the facial phenotype would receive an ABC score of **CAB**.
(B) The final step is to convert the ABC score for facial phenotype to a 4-Digit Diagnostic Code rank. A **CAB** score translates into a 4-Digit Diagnostic Code rank of **2**. This rank would serve as the second digit in the 4-Digit Diagnostic Code.
PFL, palpebral fissure length.

the FAS facial phenotype, because it uses the actual PFL, rather than the proxy measure (PFL/ICD) used by the D-score. It can be computed from direct measures or from photographs that contain internal measures of scale. The first step in deriving the facial 4-Digit Diagnostic Code rank is to derive the facial ABC-score. The magnitude of palpebral fissure length deficiency, philtrum smoothness, and upper lip thinness are ranked by circling A, B, or C in each column in the ABC-score table (Table 1A). The facial ABC-score is converted to the facial 4-Digit Diagnostic Code rank using Table 1B.

Examples of facial 4-Digit Diagnostic Code ranks for a control child and child with the FAS facial phenotype are presented in Fig. 7. The control child had PFLs equal to 25 mm (*z*-score = 0), a rank 1 philtrum and a rank 1 upper lip. These measures result in an ABC-score of AAA and a 4-Digit Diagnostic Code facial rank of 1 (normal). The child with the FAS facial phenotype had PFLs equal to 18 mm (*z*-score = –4.7), a rank 5 philtrum and rank 5 upper lip. These measures result in an ABC-score of CCC and a 4-Digit Diagnostic Code facial rank of 4 (severe). The control child's true PFLs (mm) were computed from the photograph with the aid of the internal measure of scale (25 mm = [(19.1 mm/97 pixels) × (118 pixels)] × 1.07. The PFL *z*-score = 0 = [(25 mm – 25 mm)/1.31 mm] (Hall *et al.*, 1989).

The gestalt FAS facial phenotype

Prior to the development of the 4-Digit Diagnostic Code, all patients (n = 462) seen in the FAS DPN were diagnosed using the typical gestalt (Sokol and Clarren, 1989) method of diagnosis. The gestalt method uses a less specific qualitative definition for the FAS facial phenotype and records the outcome on a dichotomous scale (present/absent). As reported by Sokol and Clarren (1989), 'A characteristic face is currently qualitatively described as including short palpebral fissures, an elongated midface, a long and flattened philtrum, thin upper lip and flattened maxilla. The specific clinical features will vary with patient age'. It is rare to find documentation in a patient's medical record or even in the medical literature as to what facial features were present when a diagnosis of FAS was given, thus, if an individual received a gestalt diagnosis of FAS, one can only infer that the FAS facial phenotype described by Sokol and Clarren (1989) was present.

Measures of brain function and structure

Structural (occipital frontal circumference (OFC) magnetic resonance imaging/computed tomography/positron emission tomography), neurological (seizures, cerebral palsy, etc) and functional (standardized psychometric tests of intellect, achievement, adaptation, language, neuropsychological performance, development, and behaviour) measures of the brain are assessed during the FAS DPN diagnostic evaluation. Many of these measures are obtained from the patient's school and medical records; others are collected at the time of the patient's diagnostic evaluation. A few examples of the types of standardized psychometric tests most frequently obtained within each domain include: *Intelligence*: Wechsler Intelligence Scale for Children — 3rd Edition (Wechsler, 1996), Wechsler Adult Intelligence Scale — Revised (Wechsler, 1981), Test of Nonverbal Intelligence (Martin *et al*., 1990); *Achievement*: Woodcock-Johnson Psycho-educational Battery (Woodcock and Johnson, 1990), Wide Range Achievement Test (Wilkinson, 1994); *Adaptation*: Vineland Scales of Adaptive Behaviour Survey (Sparrow *et al*., 1984); *Language*: Test of Word Knowledge (Wiig and Secord, 1992), Test of Auditory Comprehension of Language — Revised (Carrow-Woolfolk, 1985), Peabody Picture Vocabulary Test — Revised (Dunn and Dunn, 1981), Clinical Evaluation of Language Function (Semel *et al*., 2000), Test of Language Development — P:3

(Newcomer and Hammill, 2000); *Neuropsychological*: Rey Complex Figure Test (Spreen and Strauss, 1998), Tests of Visual–Motor Integration (Berry, 1989), Wide Range Memory and Learning Test (Adams and Sheslow, 2000), California Verbal Learning Test–C (Delis *et al*., 1994); *Infant Development*: Bayley Scales of Infant Development (Bayley, 1969), Battelle Developmental Inventory (Newborg *et al*., 1984); *Behaviour*: Child Behavior Check List (Achenbach, 1991), Conners Parent Rating Scale (unpublished, Children's Hospital National Medical Center, Washington DC). Due to the age range of the patients and the multiple sources of data, no two patients have an identical, comprehensive set of data. To assess the correlation between the facial phenotype and brain structure and function, three types of brain outcome measures were generated from the FAS DPN clinic database. (1) When a sufficient number of patients had the same standardized assessment performed [e.g. OFC centile, full-scale intelligence quotient (FSIQ), Quick Neurologic Screening Test (QNST; a test of soft neurologic signs), visual motor integration], the standardized scores from these assessments served as outcome measures. (2) More typically, the clinical data set included a broad array of standardized assessments within and across one or more of the following domains: intelligence, achievement, adaptation, language, sensory processing integration, motor skills, behavioural regulation, memory, and infant development. The patient's performance across all tests in each domain was ranked on a four-point Likert scale. The ranks were defined as follows: 0 (no tests conducted, most likely because child was too young) 1 (all test outcomes were in the normal range; no test score was lower than 0.9 SD below the norm), 2 (one or more test outcomes were in the borderline range, between 1.0 SD and 1.9 SD below the norm, but no test was 2 or more SD below the norm) and 3 (one or more tests were below normal, defined as 2 or more SD below the norm). (3) Finally, the four-point Likert scale used by the 4-Digit Diagnostic Code to rank evidence of organic brain damage was used as a global composite measure of brain structure and function. The case definitions and clinical names applied to each rank are: rank 4 (microcephaly or abnormalities on brain images or evidence of persistent neurologic findings or an IQ $\leq$ 60); rank 3 (performance on standardized psychometric tests > 2 SD below the norm across three or more of the following areas: sensory processing/integration, motor skills, behavioural regulation, adaptive behaviour, memory language, achievement, intelligence); rank 2 (observational data strongly suggest the possibility of brain damage, but data do not permit a rank 3 or 4 classification); rank 1 (no evidence of problems likely to reflect brain damage). The FAS DPN assigns the clinical term static encephalopathy to brain rank 3 and 4 and neurobehavioural disorder to brain rank 2. More detailed definitions of these terms are presented in Astley and Clarren (1999, 2000).

Prenatal alcohol exposure

All patients in this study had a confirmed history of prenatal alcohol exposure (4-Digit Diagnostic Code alcohol rank = 3 or 4) (Astley and Clarren, 1999, 2000). A history was considered confirmed if the birth mother reported consumption of alcohol during pregnancy, another individual directly observed the birth mother drinking during pregnancy and/or there was information available in the medical records that confirmed that the birth mother had been drinking during pregnancy (e.g. blood-alcohol concentrations, reported intoxicated at the time of delivery, etc). During the diagnostic evaluation, the following maternal alcohol use information is recorded on a standardized diagnostic evaluation form: average and maximum number of drinks/drinking occasion just before and during pregnancy, average number of days she drank/week just before and during pregnancy, type of alcohol consumed, trimester(s) in which alcohol was consumed, was she ever diagnosed with alcoholism, did she ever receive treatment for alcoholism and finally, what was the source and reliability of the above reported information.

Statistical analyses

Descriptive statistics were used to summarize the profile of the study population. Pearson correlation coefficients were computed to assess correlations between outcomes recorded on continuous scales. Regression analysis was used to determine if significant Pearson correlations were influenced by covariates, such as age and gender. χ^2-Tests were used to assess trends between outcomes recorded on nominal and ordinal scales. One-way ANOVA with post-hoc tests for linear trends was used to compare mean outcomes across three or more groups. Stepwise discriminant analysis (maximizing Wilk's λ) was used to identify the facial feature(s) that best differentiated patients with and without FAS diagnosed using the gestalt and 4-Digit Diagnostic Code methods. Prior probability of FAS was set equal to the prevalence in the study samples. The probability of F to enter was 0.05, and F to remove was 0.10. The unstandardized canonical discriminant function coefficients were computed to derive the discriminant equation for calculation of each subject's discriminant score. The discriminant score was used to predict each subject's diagnostic classification (FAS, not FAS). The predicted diagnoses were compared to the actual diagnoses to compute sensitivity and specificity.

RESULTS

Sociodemographic profile of study population

Of the 1130 patients evaluated in the FAS DPN clinics through 1999, 952 (84%) had a confirmed history of prenatal alcohol exposure (4-Digit Diagnostic Code alcohol rank = 3 or 4). All had given consent to use their data for research. A brief sociodemographic profile of this study population is presented in Table 2. The population was 49% Caucasian, 44% female with an average age (±SD) of 9.2 ± 6.7 years. Using the 4-Digit Diagnostic Code, 76 (8%) had a diagnosis of FAS or Atypical FAS (FAS without growth deficiency) and 767 (81%) had a diagnosis of static encephalopathy or neurobehavioural disorder without the full physical features (growth deficiency and/or facial phenotype) of FAS. A subset of 462 patients received a FAS diagnostic evaluation using the gestalt method prior to the development of the 4-Digit Diagnostic Code. The sociodemographic profile of this subgroup of 462 is comparable to the entire study population of 952 patients. The gestalt method of diagnosis had been carried out by one of three dysmorphologically trained paediatricians [a dysmorphologist (S.K.C.), a geneticist, and a developmental paediatrician].

Table 2. Sociodemographic profile of patient population with confirmed prenatal alcohol exposure

Characteristics	Entire study sample (n = 952)	Gestalt subset (n = 462)
Age (years), mean (SD), range	9.2 (6.7) 0.2 to 50.8	9.4 (6.6) 0.2 to 50.8
Race, *n* (%)		
Caucasian	463 (48.5)	231 (50.0)
African American	120 (12.6)	44 (9.5)
Native American/Alaskan/Canadian	229 (24.1)	113 (24.5)
Other	141 (14.8)	74 (16.0)
Gender, *n* (%)		
Female	418 (43.9)	194 (42.0)
4-Digit Diagnostic Code diagnostic category[a], *n* (%)		
FAS[1]	28 (2.9)	10 (2.2)
Atypical FAS[2]	48 (5.0)	17 (3.7)
Static encephalopathy, not FAS[3]	295 (31.0)	139 (28.1)
Neurobehavioural disorder, not FAS[4]	482 (50.7)	256 (55.5)
Other[5]	99 (10.4)	49 (10.6)
Facial phenotype		
D-score, mean (SD)	–0.9 (1.7)	–1.2 (1.8)
4-Digit Diagnostic Code, *n* (%)		
Absent (1)	215 (22.6)	93 (20.1)
Mild (2)	544 (57.1)	285 (61.7)
Moderate (3)	100 (10.5)	45 (9.7)
Severe (4)	93 (9.8)	39 (8.4)
Reported prenatal alcohol exposure:		
Just prior to pregnancy:		
Average no. of drinks[b] per occasion, mean (SD)	9.9 (10.9)	9.5 (6.8)
No. of drinking days/week, mean (SD)	4.6 (2.2)	4.5 (2.3)
During pregnancy:		
Average no. of drinks[b]/occasion, mean (SD)	8.5 (10.2)	8.0 (6.8)
No. of drinking days/week, mean (SD)	4.5 (2.4)	4.4 (2.4)

[a]Astley and Clarren, 2000: (1) 4-Digit Diagnostic Code category A; (2) 4-Digit Diagnostic Code category C; (3) 4-Digit Diagnostic Code categories E and F; (4) 4-Digit Diagnostic Code categories G and H; (5) 4-Digit Diagnostic Code categories I and J. [b]A drink equals 0.5 fluid oz of absolute alcohol.

Correlations between FAS facial phenotype and brain structure/function

The magnitude of expression of the FAS facial phenotype, when measured using the D-score and the 4-Digit Diagnostic Code, correlated significantly with structural, neurological, and functional measures of brain damage (Tables 3 and 4). When the magnitude of expression of the FAS facial phenotype increased, OFC percentile decreased, the QNST standard score increased (a high score reflects neurologic dysfunction), the FSIQ decreased, and composite measures of language and early childhood development were more dysfunctional or delayed. The 4-Digit Diagnostic Code ranks for the magnitude of expression of the facial phenotype and evidence of brain damage (measured independently during the diagnostic evaluation) were also correlated. As the magnitude of expression of the FAS facial phenotype increased from 1 (normal) to 4 (severe FAS), the proportion of patients with evidence of organic brain damage (structural, neurological and/or functional) increased significantly. When the four-point Likert scale for brain is collapsed into the three clinical categories (rank 1, no evidence of brain damage; rank 2, neurobehavioural disorder, and ranks 3 and 4, static encephalopathy), the correlation between the magnitude of expression of the facial phenotype and brain dysfunction increased. The 4-Digit Diagnostic Code and D-score measures of the magnitude of expression of the FAS facial phenotype were not influenced by age, race or gender. In contrast to the 4-Digit Diagnostic Code and D-score measures of the FAS facial phenotype, the gestalt measure of the FAS facial phenotype did not correlate with any measures of brain structure or function in this study population.

Correlations between FAS facial phenotype and alcohol exposure

The magnitude of expression of the FAS facial phenotype measured by the D-score and 4-Digit Diagnostic Code facial rank increased significantly with increasing number of days of maternal drinking/week, both before and during pregnancy. For example, as the 4-Digit Diagnostic Code facial rank increased from 1 (normal) to 4 (severe FAS), the mean number of days/week the birth mother drank during pregnancy increased from 4.0 to 4.4 to 4.9 to 4.8 respectively (F = weighted linear term, P = 0.006). The Pearson correlation coefficient between the facial D-score and the number of days/week the birth mother drank during pregnancy was +0.11, P = 0 009.

Variability of the FAS facial phenotype

When patients were diagnosed using the gestalt method, the facial phenotype of those receiving a gestalt diagnosis of FAS was highly variable (Table 5). In contrast, when the same patients were diagnosed using the 4-Digit Diagnostic Code, the facial phenotype among those receiving a 4-Digit diagnosis of FAS showed little variability. Of the 462 patients who received diagnostic evaluations using both the gestalt and 4-Digit Diagnostic Code methods, 445 had sufficiently complete data sets for inclusion in the following descriptive comparison of the gestalt and 4-Digit Diagnostic Code

Table 3. Correlations between 4-Digit Diagnostic Code rank of FAS facial phenotype and brain structure/function

Brain structure and function	4-Digit Diagnostic Code rank of FAS facial phenotype				*P*-value
	1: Normal	2: Mild	3: Moderate	4: Severe	
OFC centile: mean (SD)	51.2 (4.4)	51.9 (4.0)	50.5 (3.0)	50.3 (4.4)	**
FSIQ mean (SD)	89.9 (15.7)	85.4 (15.6)	84.0 (17.8)	79.5 (14.9)	**
QNST standard score: mean (SD)	26.9 (17.9)	33.0 (15.3)	42.2 (14.0)	39.0 (19.7)	*
CBCL external T: mean (SD)	72.2 (10.1)	70.4 (10.7)	70.4 (8.1)	65.3 (11.8)	*
Language: *n* (%)					
Normal (above –1 SD)	75 (49.7)	147 (34.9)	28 (35.0)	15 (23.4)	***
Borderline (–1.0 SD to –1.9 SD)	38 (25.2)	120 (28.5)	28 (35.0)	22 (34.4)	
Clinical (≥ 2 SD below norm)	38 (25.2)	154 (36.6)	24 (30.0)	27 (42.2)	
Infant development: *n* (%)					
Normal (above –1 SD)	25 (47.2)	43 (26.9)	7 (19.4)	8 (22.2)	**
Borderline (–1.0 SD to –1.9 SD)	10 (18.9)	55 (34.4)	12 (33.3)	11 (30.6)	
Clinical (≥ 2 SD below norm)	18 (34.0)	62 (38.8)	17 (47.2)	17 (47.2)	
Brain damage: *n* (%)					
4-Digit Diagnostic rank: likelihood (evidence)					
1: Unlikely	45 (20.9)	42 (7.7)	6 (6.0)	4 (4.3)	***
2: Possible (care-giver report)	116 (54.0)	291 (53.7)	51 (51.0)	24 (25.8)	
3: Probable (psychometric)	31 (14.4)	110 (20.3)	15 (15.0)	29 (31.2)	
4: Definite (struct./neurologic)	23 (10.7)	99 (18.3)	28 (28.0)	36 (38.7)	
Brain damage: *n* (%)					
4-Digit Diagnostic Code rank: diagnostic name					
1: Normal	45 (20.9)	42 (7.7)	6 (6.0)	4 (4.3)	***
2: Neurobehavioural disorder	116 (54.0)	291 (53.7)	51 (51.0)	24 (25.8)	
3 and 4: Static encephalopathy	54 (25.1)	209 (38.6)	43 (43.0)	65 (69.9)	
Age (years) at diagnosis: mean (SD)	8.3 (6.7)	10.1 (6.4)	8.6 (7.9)	7.6 (5.5)	

OFC, occipital frontal circumference; FSIQ, full-scale intelligence quotient; QNST, Quick Neurological Screen Test; CBCL, Child Behaviour Check List.
* *P*-value < 0.05; ** *P*-value < 0.01; *** *P*-value < 0.001.

Table 4. Correlations between FAS facial D-score and brain structure/function

Brain structure/function	Pearson Correlation Coefficient		*n*	*P*-value
OFC centile:	–0.19		901	***
FSIQ	–0.18		405	***
Verbal IQ	–0.13		296	*
Performance IQ	–0.23		300	***
QNST	+0.42		64	***
	Mean D-Score[a]	(SD) D-Score	*n*	*P*-value
Language				
Normal (> –1 SD)	–1.1	(1.6)	260	*
Borderline (–1.0 SD to –1.9 SD)	–0.7	(1.7)	199	
Clinical (≥ 2 SD below norm)	–0.7	(1.8)	235	
Early infant development				
Normal (> –1 SD)	–0.8	(1.6)	78	*
Borderline (–1.0 SD to –1.9 SD)	–0.5	(1.8)	86	
Clinical (≥ 2 SD below norm)	–0.3	(1.8)	109	
Brain damage/dysfunction				
4-Digit Diagnostic Code rank: likelihood (source of evidence)				
1: Unlikely	–1.5	(1.6)	88	***
2: Possible (caregiver report)	–1.1	(1.6)	466	
3: Probable (psychometric)	–0.7	(1.7)	181	
4: Definite (structural/neurologic)	–0.1	(1.9)	174	
Brain damage/dysfunction				
4-Digit Diagnostic Code rank: diagnostic name				
1: Normal	–1.5	(1.6)	88	***
2: Neurobehavioural disorder	–1.1	(1.6)	466	
3 and 4: Static encephalopathy	–0.4	(1.9)	355	

OFC, occipital frontal circumference; FSIQ, full-scale intelligence quotient; QNST, Quick Neurological Screen Test. * *P*-value < 0.05; ** *P*-value < 0.01; *** *P*-value < 0.001.
[a]The higher the D-Score, the more FAS-like the facial phenotype.

Table 5. Comparison of gestalt and 4-Digit Diagnostic Code FAS facial phenotype classifications among the 445 patients who received both a gestalt and a 4-Digit Diagnostic Code diagnosis

Feature	Gestalt diagnosis		4-Digit Diagnostic Code diagnosis	
	FAS (*n* = 52)	Not FAS (*n* = 393)	FAS[a] (*n* = 10)	Not FAS[b] (*n* = 435)
PFL *z*-score: *n* (%)				
Normal (> –2 SD)	5 (9.6)	143 (36.4)	0 (0.0)	148 (34.0)
Small (–2 SD to –2.99 SD)	8 (15.4)	98 (24.9)	0 (0.0)	106 (24.4)
Very small (≤ –3 SD)	39 (75.0)	152 (38.7)	10 (100.0)	181 (41.6)
Upper lip thinness 4-Digit Diagnostic Code rank[c]: *n* (%)				
1. Very thick	5 (9.8)	137 (35.5)	0 (0.0)	142 (33.2)
2. Moderately thick	9 (17.6)	47 (12.2)	0 (0.0)	56 (13.1)
3. Average	14 (27.5)	76 (19.7)	0 (0.0)	90 (21.0)
4. Moderately thin	9 (17.6)	83 (21.5)	2 (20.0)	90 (21.0)
5. Very thin	14 (27.5)	43 (11.1)	8 (80.0)	50 (11.7)
Philtrum smoothness 4-Digit Diagnostic Code rank[c]: *n* (%)				
1. Deeply grooved	8 (15.7)	197 (51.0)	0 (0.0)	205 (47.9)
2. Moderately grooved	7 (13.7)	82 (21.2)	0 (0.0)	89 (20.8)
3. Average	14 (27.5)	67 (17.4)	0 (0.0)	81 (18.9)
4. Moderately smooth	8 (15.7)	31 (8.0)	1 (10.0)	38 (8.9)
5. Very smooth	14 (27.5)	9 (2.3)	9 (90.0)	15 (3.5)
FAS facial 4-Digit Diagnostic Code rank: *n* (%)				
1. No FAS features	1 (1.9)	92 (23.4)	0 (0.0)	93 (21.4)
2. Mild FAS features	27 (51.9)	258 (65.6)	0 (0.0)	285 (65.5)
3. Moderate FAS features	10 (19.2)	29 (7.4)	0 (0.0)	39 (9.0)
4. Severe FAS features	14 (26.9)	14 (3.6)[d]	10 (100.0)	18 (4.1)[e]
FAS Facial D-score[f]: mean (SD)	0.6 (1.8)	–1.5 (1.5)	2.9 (0.7)	–1.4 (1.6)
Epicanthal folds[g]: *n* (%)	11 (26.2)	75 (28.5)	2 (28.6)	84 (28.9)
Hypertelorism: *n* (%)	0 (0.0)	13 (3.4)	0 (0.0)	14 (3.3)
Hypotelorism: *n* (%)	1 (2.0)	28 (7.3)	0 (0.0)	29 (6.8)
Clown eyebrows: *n* (%)	6 (15.0)	20 (8.2)	2 (20.0)	7 (2.5)
Flat nasal bridge: *n* (%)	3 (7.1)	12 (4.7)	2 (20.0)	13 (4.5)
Ptosis: *n* (%)	9 (22.0)	30 (12.0)	2 (20.0)	6 (1.1)
Flat hypoplastic midface: *n* (%)	8 (19.0)	17 (7.6)	0 (0.0)	25 (9.7)
Age (years) at diagnosis: mean (SD)	9.7 (6.2)	9.4 (6.6)	7.4 (5.6)	9.5 (6.6)

[a]4-Digit Diagnostic Code categories A and B (Astley and Clarren, 2000); [b]4-Digit Diagnostic Code categories E–V (Astley and Clarren, 2000); [c]see Fig. 3; [d]9/14 patients ≤ 5 years old; [e]all patients ≤ 5 years old, thus too young to confirm brain dysfunction and diagnose as FAS yet; [f]D-score ≥ 0.8 = FAS facial phenotype; [g]increasing severity of epicanthal folds are significantly associated with decreasing age: $F = 8.2$, $P = 0.004$.
PFL, palpebral fissure length.

methods of diagnosis. When the gestalt method was used, 52 of the 445 patients (11.7%) received a diagnosis of FAS. When the 4-Digit Diagnostic Code method was used, 10 of the 445 patients (2.2%) received a diagnosis of FAS. Of the 52 patients who received a gestalt diagnosis of FAS, only 34% had growth deficiency (height and weight below the 10th percentile), only 27% had the full FAS facial phenotype (as defined by rank 4 in the 4-Digit Diagnostic Code) and only 52% had psychometric, structural and/or neurological evidence of brain damage. In contrast, 100% of the 10 patients with a 4-Digit Diagnostic Code of FAS had growth deficiency, the full FAS facial phenotype and evidence of brain damage as defined in the sentence above. The magnitude and frequency of expression of nine minor facial anomalies frequently reported to be associated with the gestalt FAS facial phenotype were compared between the patients who did and those who did not receive a diagnosis of FAS using the two diagnostic methods (Table 5). The prevalence of all other minor anomalies was relatively low. Hypertelorism (an ICD greater than 2 SD above the norm), often referred to in the literature as a diagnostic feature of FAS, was not observed in any of the 52 patients with either a gestalt or 4-Digit Diagnostic Code of FAS. The most prevalent minor anomaly in the gestalt group was small palpebral fissure lengths. When the same patients were diagnosed using the 4-Digit Diagnostic Code, the facial phenotype of the patients who received a diagnosis of FAS did not vary from patient to patient. All patients diagnosed with FAS had small palpebral fissures, a smooth philtrum (rank 4 or 5) and a thin upper lip (rank 4 or 5).

Stepwise discriminant analyses performed on the subset of patients who received both gestalt and 4-Digit Diagnostic Code evaluations further confirmed that the FAS facial phenotype was highly variable when the gestalt method was used, and showed little variability when the 4-Digit Diagnostic Code method was used. The following facial features were made available to the stepwise discriminant analyses: mean PFL *z*-score, ICD *z*-score, mean PFL/ICD ratio, lip thinness measured on a five-point Likert scale, philtrum smoothness measured on a five-point Likert rank, epicanthal folds, flat nasal bridge, hypoplastic midface, ptosis, clown eyebrows, and nose length to midface height ratio. Only patients who had all of these facial descriptors measured in their data sets were included in the analyses. Among the 431 patients who received a gestalt diagnostic evaluation, the stepwise discriminant analysis was unable to identify a pattern of facial anomalies that accurately differentiated the 52 patients who received,

from the 379 who did not receive, a gestalt diagnosis of FAS. Two features did meet the stepwise entry criteria for inclusion into the discriminant equation: the PFL/ICD ratio and philtrum smoothness. These two features, however, were only able to differentiate the 52 with FAS from the 379 patients without FAS with 97.6% specificity (364 of 379 without FAS were correctly classified as not having FAS) and 37.3% sensitivity (only 19 of the 52 with FAS were correctly classified as having FAS). In contrast when the same patients were diagnosed using the 4-Digit Diagnostic Code, the discriminant analysis identified three facial features [PFL *z*-score, philtrum smoothness, and upper lip thinness (both measured on the five-point Likert scale from the lip–philtrum guide)] as the features that differentiated the 10 patients with a 4-Digit Diagnostic Code diagnosis of FAS from the 411 patients who did not receive a 4-Digit Diagnostic Code diagnosis of FAS, with 100% sensitivity and specificity.

DISCUSSION

The 4-Digit Diagnostic Code and D-score methods for measuring and reporting the magnitude of expression of the FAS facial phenotype offer many advantages over the gestalt method. The use of a specifically case-defined diagnostic method that relies on objective, quantitative, higher level measurement scales: (1) facilitates the collection of more accurate and precise outcome measures by a broader array of medical professionals, (2) establishes a common descriptive language for more clearly communicating outcomes in medical records and in the medical literature; (3) provides more power to detect clinically important associations that are at risk of being missed when more subjective, qualitative, nominal measurement scales are used. Noted experts in dysmorphology and anthropometry have long stressed the importance of collecting more accurate, objective measures of facial anomalies in syndrome identification (Feingold, 1975; Farkas, 1994).

The facial anomalies used to generate the 4-Digit Diagnostic Code and D-score measures of the FAS facial phenotype were identified by multivariate discriminant analyses, and found to be highly sensitive and specific to FAS and prenatal alcohol exposure (Astley and Clarren, 1996). In contrast, the gestalt approach relies on anomaly checklists that purportedly characterize the FAS facial phenotype, leaving it up to the physician or researcher to select arbitrarily which anomalies define the phenotype, how many must be present, and how severe they must be expressed (Rosett, 1980; Sokol and Clarren, 1989; Wiedemann *et al*., 1989; Gorlin *et al*., 1990; Jones, 1997). This approach has led to highly variable outcomes with no documented sensitivity or specificity to prenatal alcohol exposure (Centers for Disease Control, 1993; Floyd *et al*., 1994). Consider the following series of studies that utilized anomaly checklists to address an important diagnostic question ‘Does the FAS facial phenotype diminish with age?’. In a follow-up study of 54 patients, Spohr and Steinhausen (1987) reported a statistically significant reduction in facial features defined as ‘characterizing the craniofacial dysmorphology’ of FAS (epicanthal folds, blepharophimosis, ptosis, short upturned nose, high arched palate/cleft palate, and retrognathia). The one feature that did not change with age was a thin upper vermilion. PFL and philtrum smoothness were not measured. In a retrospective study of 200 alcohol-exposed children, Majewski (1993) reported that, in older cases, the nose was no longer short and upturned, the lips were no longer thin, and the chin often became rather prominent. The one feature that did not change with age was short PFLs. Finally, in a 10-year follow-up study of eight of the first 11 children to be diagnosed with FAS, Streissguth *et al*. (1985) reported that, while some craniofacial features changed with age (nasal bridges became more prominent and mandibles became relatively prognathic), others did not change with age (palpebral fissures remained short, philtrums remained hypoplastic, the vermilion border of the upper lip remained thin, and the midface remained flat). From these and similar studies, the 1996 report by the Institute of Medicine concluded ‘that some FAS craniofacial anomalies may be less evident at birth, become more conspicuous during early infancy and childhood, and often diminish or even disappear during adolescence and adulthood’ (Stratton *et al*., 1996). But most of the features that were reported to diminish with age: (1) have never been confirmed to be sensitive or specific to prenatal alcohol exposure; and (2) are remarkably consistent with descriptions of normal facial growth. Enlow and Hans (1996) reported that, when one compares the face of a normal child to that of a normal adult, the child’s nose is short and upturned, the nasal bridge is low and the mandible is small and retrusively placed. Interestingly, the features that were least likely to change with age (short PFL, smooth philtrum, and a thin upper lip) are the only features confirmed to be sensitive and specific to prenatal alcohol exposure in our previous (Astley and Clarren, 1996) and current studies and match the features originally identified as defining the face of FAS by David Smith back in 1979. As stated by Smith (1979), ‘as far as the diagnosis is concerned, perhaps the most important point to emerge in the last few years is that the facial abnormalities seen in affected infants are the key cluster of features that tend to make FAS a clinically discernible entity. Many disorders result in mental and growth deficiency, but in FAS the deficiencies are typically present in a patient whose face has short palpebral fissures, a hypoplastic upper lip with a thinned, vermilion border and a smoothed or absent philtrum. Up to now, the descriptions of the facial features of FAS that have appeared in the literature have not always emphasized the same abnormalities. This has led to some confusion, but inspection of the photographs accompanying these reports leaves no doubt about the facial similarities of FAS patients’. While clinical judgement plays an important role in the initial identification and definition of a new syndrome, more analytical approaches to pattern recognition, such as discriminant analysis, supported by objective, quantitative measures of outcome, can and should be used to hone the definition. The match between the facial features identified by our discriminant analysis and reported by Smith (1979) further demonstrates that the analytical approach used by the FAS DPN has succeeded in objectively case-defining, not redefining, the original FAS facial phenotype.

Correlations between face and brain

The correlations observed between the magnitude of expression of the FAS facial phenotype and brain structure and function: (1) further validate that short PFLs, a smooth philtrum and a thin upper lip are key diagnostic facial features; (2) are

consistent with the clinical literature that midline defects can *predict* underlying brain dysfunction (DeMeyer, 1975; Astley *et al*., 1999b); (3) provide evidence that an intermediate expression of the FAS facial phenotype may serve as an important clinical risk factor for brain damage caused by prenatal alcohol exposure. The FAS facial features (short PFLs, a smooth philtrum, and a thin upper lip) selected by the discriminant analyses in this study and the previous study (Astley and Clarren, 1996) are midline anomalies derived from the anterior frontal neural crest primordia of the early forebrain (Johnston, 1975). Deficiencies in the numbers of crest cells most frequently affect development of the frontonasal derivatives and are usually associated with defective forebrain and eye development (Johnston, 1975). It has long been speculated that some extreme forms of midline facial anomalies (i.e. cyclopia, holoprosencephaly, arhinencephaly) are pathognomonic of brain malformation (DeMeyer, 1975). This speculation was further supported by the presence of a proportional increase in midventral forebrain deficiencies and the severity of facial dysmorphia in mice and a non-human primate with holoposencephaly, all of which were exposed to ethanol early in gestation (Sulik and Johnston, 1982, 1983; Sulik, 1984; Seibert *et al*., 1991). Now, two additional studies have demonstrated that much more subtle midline facial anomalies (craniofacial bony alterations in non-human primates and soft-tissue facial anomalies in this current human clinical population) appear to be pathognomonic of brain malformation/dysfunction (Astley and Clarren, 1999). Smith (1979) reported similar findings: 'the severity of dysmorphic features appears to be related to the degree of mental deficiency'. The dysmorphic features he was referring to were small palpebral fissures, a smooth philtrum, and a thin upper lip. No other studies, to our knowledge, have reported significant linear correlations between the magnitude of expression of the FAS facial phenotype and cognitive impairment among individuals with prenatal alcohol exposure. Other clinical research teams have reported correlations between the number of physical anomalies observed over the entire body and brain dysfunction in individuals with prenatal alcohol exposure, although not all were reported to be statistically significant (Majewski, 1993; Spohr *et al*., 1993). No correlations were observed between the gestalt FAS facial phenotype and brain dysfunction in this study. Failure to detect statistically significant correlations between face and brain, when a gestalt approach to diagnosis was used, has also been reported by others (Graham *et al*., 1988; Spohr *et al*., 1993).

In summary, thousands of individuals with FAS have been identified and thousands of laboratory, clinical and population-based studies have been conducted. While these studies have greatly advanced our understanding of alcohol's teratogenic potential, advancements in the clinical and public health arenas are less impressive. To date, we still cannot derive an accurate estimate of the prevalence of FAS (Floyd *et al*., 1994) nor can we document success in preventing FAS. Advancements in these two arenas are contingent upon physicians making accurate diagnoses. Accurate diagnoses require specific and objective case definitions that document the full range of outcomes associated with prenatal alcohol exposure. These definitions should be continually honed to incorporate the latest technological advances (e.g. magnetic resonance spectroscopy and functional magnetic resonance imaging, digital image analysis etc.) and should be guided by more sophisticated, multivariate, analytical approaches to pattern recognition.

REFERENCES

Achenbach, T. M. (1991) *Manual for the Child Behavior Checklist/ 4–18 and 1991 Profile*. University of Vermont Department of Psychiatry, Burlington, VT.

Adams, W. and Sheslow, D. (2000) *Wide Range Memory and Learning Test*. Psychological Assessment Resources, Inc., Odessa, FL.

Astley, S. J. and Clarren, S. K. (1996) A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome. *Journal of Pediatrics* **129**, 33–41.

Astley, S. J. and Clarren, S. K. 1997) *Diagnostic Guide for Fetal Alcohol Syndrome and Related Conditions: The 4-Digit Diagnostic Code*. University of Washington Publication Services, Seattle, WA.

Astley, S. J. and Clarren, S. K. (1999) *Diagnostic Guide for Fetal Alcohol Syndrome and Related Conditions: The 4-Digit Diagnostic Code*, 2nd edn. University of Washington Publication Services, Seattle, WA.

Astley, S. J. and Clarren, S. K. (2000) Diagnosing the full spectrum of fetal alcohol-exposed individuals: introducing the 4-Digit Diagnostic Code. *Alcohol and Alcoholism* **35**, 400–410.

Astley, S. J., Clarren, S. K., Gratzer, M., Orkand, A. and Astion, M. (1999*a*) *Fetal Alcohol Syndrome Tutor™ Medical Training Software* CD ROM, March of Dimes, New York.

Astley, S. J., Magnuson, S. I., Omnell, L. and Clarren, S. K. (1999*b*) Fetal alcohol syndrome: changes in craniofacial form with age, cognition, and timing of ethanol exposure in the macaque. *Teratology* **59**, 163–172.

Bayley, N. (1969) *Bayley Scales of Infant Development: Birth to Two Years*. The Psychological Corporation, San Antonio, TX.

Berry, K. E. (1989) *Developmental Test of Visual–Motor Integration*. Modern Curriculum Press, Cleveland, OH.

Carrow-Woolfolk, E. (1985) *Test for Auditory Comprehension of Language-Revised*. DLM Teaching Resources, Allen, TX.

Centers for Disease Control (1993) Linking multiple data sources in fetal alcohol syndrome surveillance — Alaska *Morbidity and Mortality Weekly Report* **42**, 3312–3314.

Clarren, S. K. and Smith, D. W. (1978) Fetal alcohol syndrome. *New England Journal of Medicine* **298**, 1063–1067.

Delis, C., Kramer, J. H., Kaplan, E. and Ober, B. A. (1994) *California Verbal Learning Test — C Children's Version*. The Psychological Corporation, San Antonio, TX.

DeMeyer, W. (1975) Median facial malformations and their implications for brain malformations. In *Morphogenesis and Malformation of Face and Brain. The National Foundation–March of Dimes Birth Defects: Original Article Series*, Bergsma, D. ed., **11**(7), 155–181. March of Dimes, New York.

Dunn, L. M. and Dunn E. S. (1981) *Peabody Picture Vocabulary Test — Revised*. American Guidance Service, Circle Pines, MN.

Enlow, D. H. and Hans, M. G., eds (1996) Facial form and pattern. In *Essentials of Facial Growth*, pp. 122–145, W. B. Saunders Company, Philadelphia, PA.

Farkas, L. G. (1994) *Anthropometry of the Head and Face*, 2nd edn, p. 445. Raven Press, New York.

Feingold, M. (1975) The face and syndrome identification. In *Morphogenesis and Malformation of Face and Brain, The National Foundation–March of Dimes Birth Defects*: *Original Article Series* Bergsma, D. ed., **11**(7), 213–215. March of Dimes, New York.

Floyd, R. L., Davis, M. K., Martin, M. L., Hungerford, D and Hymbaugh, K. J. (1994) Fetal alcohol syndrome In *From Data to Action*, Wilcox, L. S., Marks, J. S. eds, pp. 343–350. US Public Health Service, Centers for Disease Control, Atlanta, GA.

Gorlin, R. J., Cohen, M. M. and Levin, L. S. (1990) Fetal alcohol syndrome. In *Syndromes of the Head and Neck*, pp. 16, 17. Oxford University Press, New York.

Graham, J. M., Hanson, J. W., Darby, B. L., Barr, H. M. and Steissguth, A. P. (1988) Independent dysmorphology evaluations at birth and 4 years of age for children exposed to varying amounts of alcohol *in utero*. *Pediatrics* **81**, 772–778.

Hall, J. G., Froster-Iskenius, U. G. and Allanson, J. E. (1989) *Handbook of Normal Physical Measurements*. Oxford University Press, Oxford.

Iosub, S., Fuchs, M., Bingol, N., Stone, R. K., Gromisch, D. S. and Wasserman, E. (1985) Palpebral fissure length in black and hispanic children: correlation with head circumference. *Pediatrics* **75**, 318–320.

Johnston, M. C. (1975) The neural crest in abnormalities of the face and brain. In *Morphogenesis and Malformation of Face and Brain. The National Foundation — March of Dimes Birth Defects: Original Article Series*, Bergsma, D. ed., **11**(7), 1–18. March of Dimes, New York.

Jones, K. L. (1997) Fetal alcohol syndrome. In *Smith's Recognizable Patterns of Human Malformation*, 5th edn, pp. 555–558. W. B. Saunders, Philadelphia, PA.

Jones, K. L. and Smith, D. W. (1973) Recognition of the fetal alcohol syndrome in early infancy. *Lancet* **ii**, 999–1001.

Majewski, F. (1993) Alcohol embyropathy: experience in 200 patients. *Developmental Brain Dysfunction* **6**, 248–265.

Martin, J. D., Blair, G. E. and Bledsoe, J. R. (1990) Measure of concurrent validity and alternate-form reliability of the test of non-verbal intelligence. *Psychological Reports* **66**, 503–508.

Newborg, J., Stock, J. R. and Wnek, L. (1984) *Battelle Developmental Inventory*. DSM Teaching Resources, Allen, TX.

Newcomer, P. L. and Hammill, D. D. (2000) *Test of Language Development-P:3*. Psychological Assessment Resources, Inc., Odessa, FL.

Rosett, H. L. (1980) A clinical perspective of the fetal alcohol syndrome. *Alcoholism: Clinical and Experimental Research* **4**, 119–122.

Seibert, J. R., Astley, S. J. and Clarren, S. K. (1991) Holoprosencephaly in a fetal macaque (*Macaca nemestrina*) following weekly exposure to ethanol. *Teratology* **44**, 29–36.

Semel, E., Wiig, E. H. and Secord, W. A. (2000) *Clinical Evaluation of Language Function-3*. Psychological Corporation, San Antonio, TX.

SigmaScan Pro Image Analysis (1999) Version 5.0, SPSS Inc.

Smith, D. W. (1979) The fetal alcohol syndrome. *Hospital Practice*, October, 121–128.

Sokol, R. J. and Clarren, S. K. (1989) Guidelines for use of terminology describing the impact of prenatal alcohol on the offspring. *Alcoholism: Clinical and Experimental Research* **13**, 597–598.

Sparrow, S. S., Balla, D. A. and Cicchetti, D. V. (1984) *Vineland Adaptive Behavior Scales: Interview Edition Survey Form Manual*. American Guidance Service, Circle Pines, MN.

Spohr, H. L. and Steinhausen, H. C. (1987) Follow-up studies of children with fetal alcohol syndrome. *Neuropediatrics* **18**, 13–17.

Spohr, H. L., Willms, J. and Steinhausen, H. C. (1993) Prenatal alcohol exposure and long-term developmental consequences. *Lancet* **341**, 907–910.

Spreen, O. and Strauss, E. (1998) A compendium of neuropsychological tests. *Administration, Norms and Commentary*, 2nd edn. Oxford University Press, New York.

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

Streissguth, A. P., Clarren, S. K. and Jones, K. L. (1985) Natural history of the fetal alcohol syndrome: a 10-year follow-up of 11 patients. *Lancet* **ii**, 85–91.

Sulik, K. K. (1984) Critical periods for alcohol teratogenesis in mice, with special reference to the gastrulation stage of embryogenesis. In *Mechanisms of Alcohol Damage in utero, Ciba Foundation Symposium*, 105, pp. 124–141. Pitman, London.

Sulik, K. K. and Johnston, M. C. (1982) Embryonic origin of holoprosencephaly: interrelationship of the developing brain and face. *Scanning Electron Microscopy* **1**, 309–322.

Sulik, K. K. and Johnston, M. C. (1983) Sequence of developmental alterations following acute ethanol exposure in mice: craniofacial features of the fetal alcohol syndrome. *American Journal of Anatomy* **166**, 257–269.

Thomas, I. T., Gaitantzis, Y. A. and Frias, J. L. (1987) Palpebral fissure length from 29 weeks gestation to 14 years. *Journal of Pediatrics* **111**, 267–268.

Wechsler, D. (1981) *Manual for the Wechsler Adult Intelligence Scale — Revised*. The Psychological Corporation, San Antonio, TX.

Wechsler, D. (1996) *Manual for the Wechsler Intelligence Scale for Children*. Psychological Corporation, San Antonio, TX.

Wiedemann, H. R., Junze, J. and Dibbern, H. (1989) Alcohol embryopathy In *Atlas of Clinical Syndromes. A Visual Aid to Diagnosis*, 2nd edn, pp. 480–483, Mosby Year Book, St Louis, MO.

Wiig, E. H. and Secord, W. (1992) *Test of Word Knowledge*. The Psychological Corporation, San Antonio, TX.

Wilkinson, G. S. (1994) *The Wide Range Achievement Test 3*. Wide Range, Inc., Wilmington, DE.

Woodcock, R. W. and Johnson, M. B. (1990) *Woodcock–Johnson Psycho-Educational Battery — Revised: Tests of Achievement*. DLM Teaching Resources, Allen, TX.