VALIDATION OF THE FACIAL PHOTOGRAPHIC METHOD IN FETAL ALCOHOL SPECTRUM DISORDER SCREENING AND DIAGNOSIS

Marina Avner, Paul Henning, Gideon Koren, Irena Nulman
Division of Clinical Pharmacology and Toxicology, Hospital for Sick Children, University of Toronto, Canada

ABSTRACT

Objective
A prospective study to validate the computer-assisted method of measuring palpebral fissure length and philtrum smoothness using digital patient photographs. These are key diagnostic facial features of Fetal Alcohol Syndrome.

Participants
Motherisk Program (including *Breaking the Cycle*), Hospital for Sick Children, Toronto - a clinical, research and teaching program dedicated to antenatal drug, chemical, and disease risk counseling. 40 children referred for FASD assessment, 21 under 4 years old, 19 were 4 years or older.

Methods/ Materials
Facial measurements were obtained directly from the patient by physicians and compared to those obtained by computer software measurement of photographs of the same patient.

Outcome Measures
Palpebral fissure length and philtrum smoothness.

Results
The photographic measurements showed shorter palpebral fissure length than the direct measurements when analyzing all children (25.4±2.3 vs.23.2±2.4mm; $p<0.0001$), and children under four (n=21, 24.7±2.4 vs. 21.6±1.6mm; $p<0.0001$). The difference for older children (n=19) did not reach statistical significance. The computer found four false positive cases and no false negative cases of clinically short palpebral fissure (sensitivity=100%, specificity=64%). Direct measurement scores for philtrum smoothness were different from the computer's measurements using the frontal view ($p=0.0012$) but not using the ¾ view.

Conclusion
The method of computer-assisted measurement tends to underestimate the true length and, hence, over-diagnose short palpebral fissure, especially in children under four years old. This method may serve as a useful fetal alcohol syndrome screening tool.

Key Words: *Fetal alcohol syndrome, palpebral fissure length, philtrum smoothness, screening tools*

Ethanol is the most widely used human teratogen, and up to 60% of women worldwide drink alcohol at some point during pregnancy.[1] The adverse effects of alcohol on the developing human represent a spectrum of structural, behavioural and neurocognitive disabilities collectively termed Fetal Alcohol Spectrum Disorders (FASD).[2] FASD is one of the leading known causes of mental retardation in the Western World[3] and nearly 1% of America's population is affected by some expression of it.[4] Abel's 1998 analysis of 29 prospective studies

collected from indigenous populations around the world estimated that Fetal Alcohol Syndrome (FAS), the most extreme presentation of FASD, occurs an average of 0.97 times for every 1000 live births.[1] Two sets of diagnostic criteria are commonly used for the evaluation FAS in children: the 1996 Institute of Medicine criteria[5] and the Washington University criteria.[6] Both retain the four key diagnostic features of FAS. They are growth deficiency, characteristic FAS facial phenotype, central nervous system damage/dysfunction, and alcohol exposure in utero. The facial morphological features include palpebral fissure length (PFL) (Figure 1) and philtrum smoothness with upper lip thinness (Figure 2). While it is very important to accurately measurethe facial morphological features, this task is challenging, especially in young children.

.

FIG. 1 Measuring Facial Features

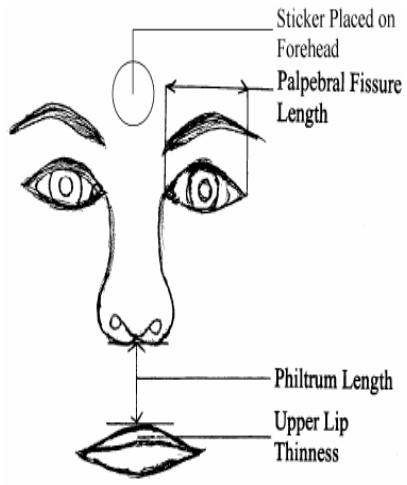
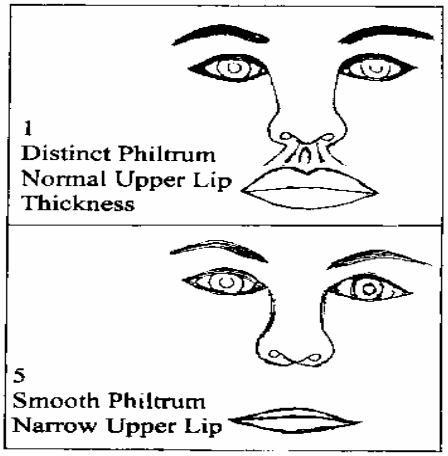
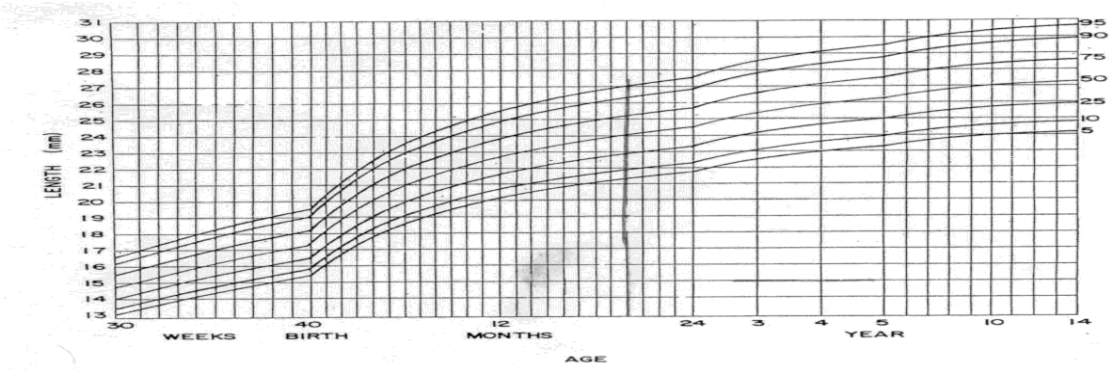


FIG. 2 Lip-philtrum Formations



Example of a normal lip-philtrum formation[1] and a lip-philtrum formation strongly affected by FAS[5]. Rankings were developed by Astley and Clarren (2001).

FIG. 3 Relationship of Palpebral Fissure Length to Age in American White Children



Direct measurement of the palpebral fissure is dependent on the examiner's skills and the child's cooperation, which is difficult to achieve when he/she is young. In 1996, Astley and Clarren offered a photographic approach for more efficient and accurate screening of facial morphology.[7]

To assess the photographs, the Facial Analysis Software was developed. Over 60 clinical teams were trained across the USA and Canada in its use. This very promising method has made telediagnosis possible, allowing FAS and FASD to be identified in remote populations where no appropriate medical expertise exists. Presently, no study has contrasted the use of the photographic method with direct measurement of facial morphology. Nor has any study validated the use of the photographic method in young children. It is critically important to validate the photographic method because changes as small as one or two millimeters in measurement of the eye can lead to over or under diagnosis of FASD (Figure 3). Therefore, in this study we aimed to validate the computer-assisted method of photographic measurement, by comparing its results to those obtained by a trained physician performing direct measurements of the same child.

METHODS

Study Population

Included in this study was a prospectively collected cohort of 40 children, subdivided into two age groups (under four years of age, four years or older). All the children in the study were referred for an FASD assessment where facial measurements were taken (regardless of diagnosis). Each child's facial features were measured both by hand and by using the photographic method. In this way, each child served as its own control. Parents gave consent to have their child's picture taken as part of the clinical recording of the cases. Children with structural facial dysmorphology precluding accurate measurements were excluded.

Setting

Recruitment took place at a FAS diagnostic clinic at the Hospital for Sick Children, and at a pediatric clinic at *Breaking the Cycle* - an early identification and prevention program designed to serve drug- and alcohol-addicted pregnant women and families with young children. The study was approved by the Hospital for Sick Children Research Ethics Board.

Direct Measurement of Facial Features

As part of the routine FAS assessment, the children's facial features were analyzed and ranked for the magnitude of expression of the FAS facial phenotype. This involved measurement of the PFL and evaluation of the philtrum-lip formation, by focusing on philtrum smoothness. The palpebral fissure length is defined as the distance between the inner and outer canthus of one eye.[8] The PFL was measured by trained physicians to the nearest millimeter with a clear plastic ruler held as close as possible to the eye without touching it or the eyelashes.[7] The children were instructed to open their eyes fully to allow for identification of the inner and outer canthus. The mean PFL for each patient was transformed into a standardized z-score using anthropometric charts[8] (such charts do not exist for the upper lip vermillion and, hence, it was not measured). The z-score is defined as the patient's PFL minus the mean PFL for the normal population divided by the standard deviation of the mean PFL for the normal population. Philtrum smoothness was scored using a five-point Likert scale using a pictorial lip-philtrum guide.[9]

Photographic Software Measurement of Facial Features

Three photographs (frontal, ¾, and lateral views) were taken of each child using a handheld digital camera with three mega pixel resolution. The pictures were taken with a 19.05 mm adhesive paper sticker (Figure 1) on the forehead for an internal measure of scale and children were instructed to relax with no smile and their eyes open wide, as described previously by Astley and Clarren.[9] The pictures were analyzed using the FAS Facial Photographic Analysis Software version 1.0.0 (FAS Diagnostic and Prevention Network, FAS DPN, University of Washington, Seattle, Washington, Copyright 2003) following the instructions given by the author. The right and left PFLs were measured by clicking the mouse on the inner and outer canthus and having the software calculate the distance between the landmarks in pixels (dots of light on the computer monitor). The length of the internal

measure of scale (paper sticker) was also translated into pixels. The real size of the PFL (mm) was computed from the PFL measured in pixels, the length of the paper sticker in pixels, and the real length of the paper sticker (mm) by the software. The z-score was also computed by the software using the same anthropometric chart[8] that was used to compute the z-score from the direct measurement of PFL in the clinic. Philtrum smoothness was ranked by the physician using the lip–philtrum guide.[9]

The guide ranks philtrum smoothness in 5 categories:
1. Deeply furrowed (Figure 2.1),
2. Somewhat furrowed,
3. Mid-range,
4. Somewhat smooth, and
5. Very smooth (Figure 2.5).

Categories four and five are diagnostic of alcohol toxicity. Philtrum smoothness was measured by holding the lip-philtrum guide next to the patient's face and assigning it the Likert rank of the photograph that best matches it.[9] The patient should be relaxed, but not smiling, and have his or her mouth closed. The philtrum smoothness of the same patient was assessed using the software by applying the lip-philtrum guide to a photograph of the patient.

Comparison of Direct and Photographic Software Measurements

Comparison of the PFL and the five-point Likert score for philtrum smoothness using direct and photographic software measurement was performed on all patients. In addition, a separate analysis of patients under four years of age, and patients four years of age or older was performed. Mean PFL measurements were compared by the two-tailed Student's t-test, while the five-point Likert scores for philtrum smoothness were compared by a two-tailed Mann-Whitney U test. In order to assess the clinical significance of our findings, the number of patients whose z-score was below the cut-off for FAS (z-score< -2) were counted and compared for each method of measurement. Sensitivity and specificity of the photographic measurements were subsequently defined in terms of how often they correctly detected a PFL that was below the cut-off.

RESULTS

Patient Characteristics

40 children, 20 from the Hospital for Sick Children (mostly four years of age or older) and 20 from Breaking the Cycle (mostly less than four years old) were recruited. Overall, the children were 40% female with a mean age of 4.8 ± 4.6 years (2 months-15 years). 21 of the patients were under four years of age, of which 43% were female and which had a mean age of 1.3 ± 1.0 years. 19 children were at least four years old, of which 37% were female and which had a mean age of 8.7 ± 3.8 years (Table 1).

Comparison of PFL Measurements

The photographic measurements yielded a statistically shorter PFL than the direct measures (Table 2). This result was apparent when analyzing all patients (25.4 ± 2.3 vs. 23.2 ± 2.4 mm; $p < 0.0001$) as well as with patients under four years of age (24.7 ± 2.4 vs. 21.6 ± 1.6 mm; $p< 0.0001$).The difference for patients four years and older did not reach statistical significance (26.1 ± 1.8 vs. 25.0 ± 1.8 mm; $p = 0.06$).

TABLE 1 Patient Characteristics

	Mean Age ± SD (years)	Sex (% Female)
All Patients	4.8 ± 4.6	40
Patients Under 4	1.3 ± 1.0	43
Patients 4 and Over	8.7 ± 3.8	37

SD - Standard deviation

TABLE 2 Comparison of Palpebral Fissure Length (PFL) by Direct and Photographic Software Measurement

	Direct	Software	Significance*
All Patients - Mean PFL ± SD (mm)	25.4 ± 2.3	23.2 ± 2.4	p < 0.0001
Patients Under 4 - Mean PFL ± SD (mm)	24.7 ± 2.4	21.6 ± 1.6	p < 0.0001
Patients 4 and Over - Mean PFL ± SD (mm)	26.1 ± 1.8	25.0 ± 1.8	p = 0.0616

PFL - Palpebral Fissure Length SD – Standard Deviation* Using two-tailed student's t-test

TABLE 3 Comparison of 5-Point Likert Scale Scores for Philtrum by Direct and Photographic Software Measurement

	Direct Measurement: n (%)	Software Measurement (Frontal View): n (%)	Significance (Frontal)*	Software Measurement (3/4 View): n (%)	Significance (3/4)*
Deeply Furrowed	5 (15)	0 (0)	p = 0.0012	2 (6)	p = 0.091
Somewhat Furrowed	18 (55)	13 (33)		15 (44)	
Mid Range	10 (30)	22 (56)		16 (47)	
Somewhat Smooth	0 (0)	4 (10)		1 (3)	
Very Smooth	0 (0)	0 (0)		0 (0)	
Deeply Furrowed	3 (16)	0 (0)	p < 0.0001	1 (6)	p = 0.219
Somewhat Furrowed	9 (47)	5 (25)		6 (38)	
Mid Range	7 (37)	12 (60)		8 (50)	
Somewhat Smooth	0 (0)	3 (15)		1 (6)	
Very Smooth	0 (0)	0 (0)		0 (0)	
Deeply Furrowed	2 (14)	0 (0)	p = 0.037	1 (6)	p = 0.211
Somewhat Furrowed	9 (64)	8 (42)		9 (50)	
Mid Range	3 (21)	10 (53)		8 (44)	
Somewhat Smooth	0 (0)	1 (5)		0 (0)	
Very Smooth	0 (0)	0 (0)		0 (0)	

*Using Mann-Whitney U test

Comparison of Five-Point Likert Score for Philtrum Smoothness

The five-point Likert score for philtrum smoothness, determined by direct measurement, was compared with the photographic software measurement of philtrum smoothness using both the frontal view and the ¾ view (Table 3). The direct measurement values were significantly different from the photographic software measurement using the frontal view for all patients ($p = 0.0012$), patients under four ($p < 0.0001$), and patients four and older ($p = 0.037$). However, the direct measurement values were not statistically different from the photographic software measurement using the ¾ view for all patients ($p = 0.091$), patients under four ($p = 0.219$), and patients four and older ($p = 0.211$).

Clinical Significance

Since a z-score of < -2 would be indicative of short PFL it was deemed appropriate to determine which children would have screened positive for short PFL by direct measurement compared to photographic software measurement (Table 4). When comparing all children, five would be found to have a short PFL (z-score <-2) using photographic software but not by direct measurements of a physician. Of these patients, two were under four years old, and three were over four years old. The sensitivity of the photographic measurement was 100% (i.e. no true cases of z<-2 PFL were missed). The specificity of the photographic technique was 64% (9/14). The sensitivity and specificity refer to the likelihood of correctly measuring a short PFL and not necessarily correctly diagnosing FAS.

TABLE 4 Number of patients with short Palpebral Fissure Length as determined by z-score < -2

	Direct Measurement- # with z- score < - 2	Software Measurement - # with z-score < -2
All Patients	9	14
Patients Under 4	0	2
Patients 4 and Over	9	12

COMMENTS

Short palpebral fissure is the most diagnostic facial feature of FASD; hence its valid measurement is vital. The use of photography for this measure would allow screening and telediagnosis of thousands of potentially affected children who have no access to a physician trained in FASD diagnosis. However, before such a technique can be implemented, its accuracy must be established. Our study compared direct measures of PFL by FASD experts with those obtained by digital photography and the software developed by Astley and Clarren. We deliberately chose a range of children from toddler hood to school age so that we could explore the challenges encountered with young children who may not be cooperative. This is especially critical with measurement of PFL, where even a one or two millimeter difference may bring a child below two standard deviations of the mean.

Our study reveals that there is a significant trend toward shorter PFL with photography than with direct measurement overall. However, this occurs mostly among young children who are less likely to cooperate with the examiner. Since it is difficult to measure the angles of the eye in less cooperative children, this trend is logical.

Experts commonly label palpebral fissures that are two standard deviations below the mean as indicative of FASD. Since the photographic method tends to overestimate the number of short PFLs (sensitivity=100%, specificity=64%), it is likely to over-diagnose FAS. Therefore,

telediagnosis would be most useful if it were followed up by direct measurement (since the photographic method alone produced false positives but no false negatives). The unique configuration of the upper lip in humans is mainly determined by the existence of the philtrum. The philtrum consists of two structures, the lateral philtral ridges (two lateral ridges) and the midline philtral depression (the space between them). Research done on fetuses revealed that lateral philtral ridges start to develop at 14 weeks post conception. The orbicularis oris muscle (surrounding the mouth and controlling the lip movements) is detected at 11 weeks of pregnancy and, at 15-16 weeks, the muscle fibers from each lateral aspect of the lip cross towards the midline creating the two lateral philtral ridges and the midline philtral depression. Meanwhile, underneath the future philtral area, an area of loose connective tissue is found (11 weeks post conception) called the frenulum-associated connective tissue. This connective tissue seems to have a major role in the formation of the philtrum, perhaps by directing the orbicularis oris muscle fibers. Thus the formation of the philtrum is the result of an interaction between the connective tissues and the orbicularis oris muscle. Only the existence of both components will result in a normal philtrum-lip formation. A smooth upper lip is the result of the absence of philtral structures. Alcohol toxicity has been associated with the absence of frenulum-associated connective tissue, orbicularis oris muscle dysmorphology, and forebrain developmental abnormalities.[10]

The forebrain plays a major role in the development of the midline portion of the upper lip, primarily through the neural crest cell mesoderm, which forms the mid-face region. Sulik et al demonstrated that in mice, prenatal exposure to ethanol severely affects the development of the forebrain and the mid-face.[10] Deficiencies in neural plate development at early stages would lead, not only to abnormal brain morphogenesis, but also to deficient eye formation, since the eyes develop at later stages as evaginations from the neural epithelium.10 This shows that forebrain abnormalities caused by prenatal ethanol exposure may be the mechanism by which the philtrum fails to develop normally.

CONCLUSIONS

In summary, alcohol is a potent teratogen. Exposure to alcohol in early pregnancy may be associated with impaired brain formation and facial dysmorphology. Therefore, PFL and philtrum smoothness are important facial diagnostic criteria of alcohol toxicity in early pregnancy. The computer-assisted photographic method demonstrated 100% sensitivity and 64% specificity, with no false negatives recorded when looking for clinically short PFL. In toddlers younger than four years of age, the bias in measurement is larger; it tends to underestimate PFL. This method may be used as an efficient screening tool, and for telediagnosis of FASD patients in remote areas, where the number of local experts in the field is limited.

Funding Acknowledgements
Supported by grants from CIHR.

Corresponding Author:
irena.nulman@sickkids.ca

REFERENCES

1. Abel EL. Fetal Alcohol Syndrome. Plenum Press, New York, 1998.
2. Barr HM, Streissguth AP. Identifying maternal self-reported alcohol use associated with fetal alcohol spectrum disorders. Alcohol Clin Exp Res 2001;25:283-287.
3. Spohr HL, Willms J, Steinhausen HC. Prenatal alcohol exposure and long-term developmental consequences. Lancet 1993;341:907-10.
4. Sampson PD, et al. Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. Teratology 1997;56:317-26.
5. Stratton KR, Howe CJ, Battaglia FC. Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment. National Academy Press, Washington, 1996.
6. Astley SJ, Clarren SK. Diagnosing the full spectrum of fetal alcohol-exposed individuals:

introducing the 4-digit diagnostic code. Alcohol Alcohol 2000;35:400-410.
7. Astley SJ, Clarren SK. A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome. J Pediatr 1996;129:33-41.
8. Hall JG, Froster-Iskenius UG, Allanson JE. Handbook of Normal Physical Measurements. Oxford University Press, Oxford, 1989.
9. Astley SJ, Clarren SK. Measuring the facial phenotype of individuals with prenatal alcohol exposure: correlations with brain dysfunction. Alcohol Alcohol 2001;36:147-59.
10. Sulik KK, Johnston MC, Webb MA. Fetal alcohol syndrome: embryogenesis in a mouse model. Science 1981;214:936-8.