

EVALUATION OF FACIAL MORPHOLOGY: COMPARISON OF ABSOLUTE AND RELATIVE MEASUREMENTS, OBJECTIVE AND SUBJECTIVE APPROACHES

L. Wolf^{1,2}, L. Basel-Vanagaite^{1,2,3}, S. Cassidy⁴, T. Cole⁵, R. Newbury-Ecob⁶, A. Smith⁷, L. Biesecker⁷, K. Gripp⁸, C. Curry⁹, M. Innes¹⁰, J. Allanson¹¹.

1. FDNA Ltd., Herzlyia, Israel; 2. Tel Aviv University, Israel; 3. Schneider Children's Medical Center of Israel, Rabin Medical Ctr, Petah Tikva, Israel; 4. UC Irvine, USA; 5. West Midlands Regional Genetics Service, UK; 6. Clinical Genetics, Bristol, UK; 7. NHGRI, National Institutes of Health, USA; 8. Division of Medical Genetics, A. I. du Pont Hospital for Children, Wilmington, USA; 9. Genetic Medicine Central California, Fresno, USA; 10. Department of Medical Genetics, Alberta Children's Hospital and University of Calgary, Calgary, Canada; 11. Children's Hospital of Eastern Ontario and University of Ottawa, Canada



We examined the reliability of identifying facial features in a population of individuals with various syndromes (n=150), using (i) absolute anthropometric measurements; (ii) the ratio between each measurement and head circumference (OFC); or (iii) relative measurements automatically extracted by a computer software from patients' 2D facial photographs. The findings of each method are compared with the majority vote of a panel of 6 dysmorphology experts annotating the facial features in each photograph (average agreement rate: 82%).

Collecting the data

- Method:**
- Six expert dysmorphologists evaluated 150 images of individuals with various syndromes
 - The evaluation was based on the images alone and was done independently by each expert
 - The *established ground truth* is based on the most common vote among the expert panel
 - The anthropometric measurements were taken by one expert (JA) using calipers and tape measure
 - Image measurements were extracted automatically using the Facial Dysmorphology Novel Analysis (FDNA®) technology

- Sample results:**
- The table below demonstrates results obtained for three sample images.
 - +1 indicates the facial feature is present, -1 indicates it is absent.

		Broad Face	Face Long	Face Short	Telecanthus	Eyes, Widely Spaced	Eyes, Closely Spaced	Nose Long	Nose Short	Nose Wide	Nose Narrow	Mouth Wide	Mouth Narrow	Jaw Broad	Jaw Narrow
Image 1	Expert Majority	-1	+1	-1	-1	-1	-1	+1	-1	+1	-1	-1	-1	-1	-1
	Anthropometric Measurements	100%	83%	100%	100%	100%	83%	83%	100%	75%	100%	92%	83%	100%	67%
	Measurements Relative to OFC	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	Image Extracted Rel. Measurements	-1	+1	-1	-1	-1	-1	+1	-1	-1	-1	-1	-1	-1	-1
Image 8	Expert Majority	+1	-1	-1	+1	-1	-1	-1	+1	-1	-1	-1	+1	+1	-1
	Anthropometric Measurements	58%	100%	100%	58%	50%	100%	100%	100%	100%	100%	92%	75%	58%	75%
	Measurements Relative to OFC	-1	-1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	Image Extracted Rel. Measurements	-1	-1	-1	+1	-1	-1	-1	+1	+1	-1	-1	-1	+1	-1
Image 117	Expert Majority	-1	+1	-1	+1	-1	-1	-1	-1	-1	-1	+1	-1	-1	-1
	Anthropometric Measurements	83%	100%	100%	83%	83%	100%	83%	83%	100%	100%	58%	83%	100%	83%
	Measurements Relative to OFC	-1	+1	-1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
	Image Extracted Rel. Measurements	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

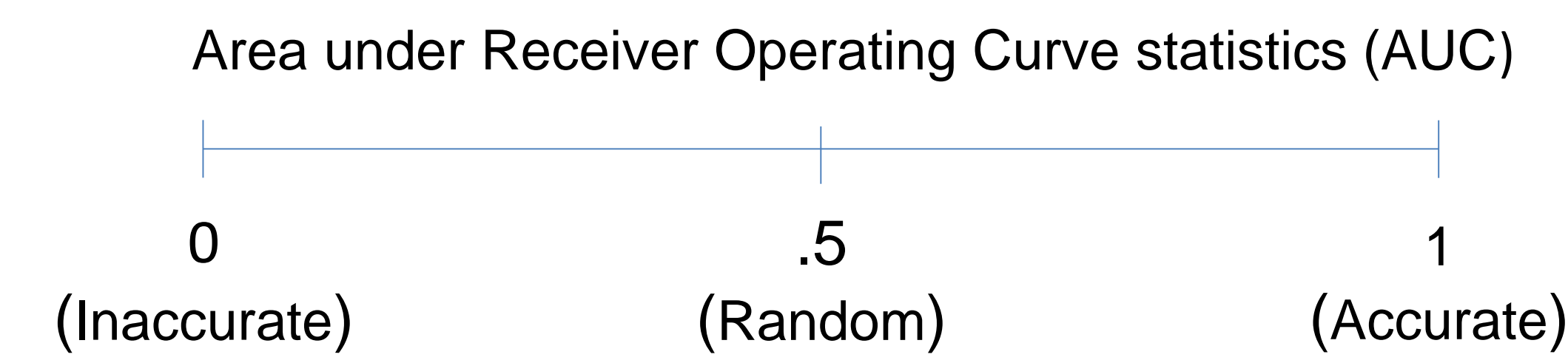
Comparison of Feature Detection Methods

Method:

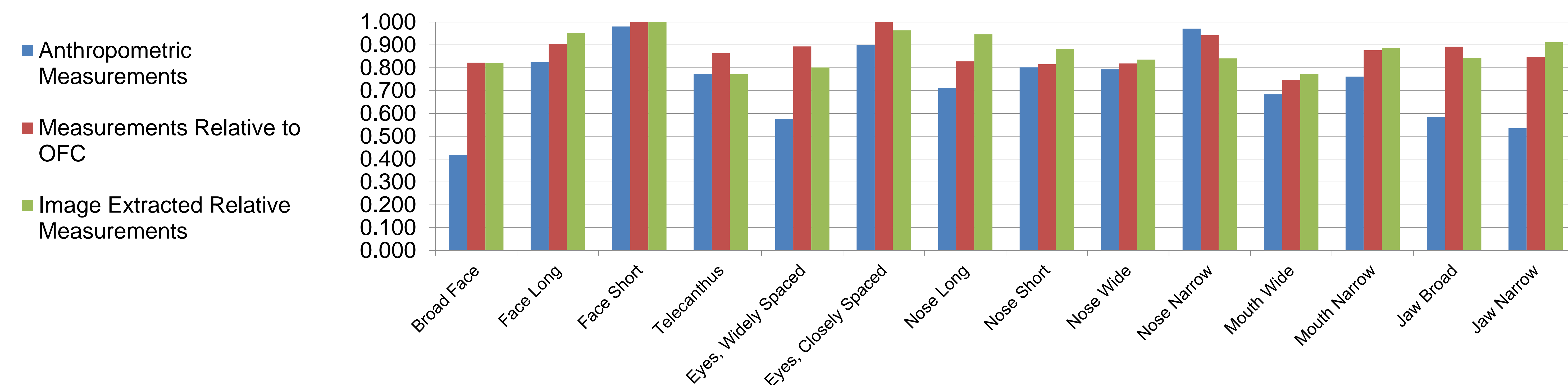
- The three types of analysis are being compared to the expert panel
- Results for individual facial features are reported in Area under Receiver Operating Curve statistics (AUC). The highest AUC for each feature is bolded
- The final "Combined Annotation" column measures mean accuracy for the entire set of features in each image

Results:

- Using the experienced dysmorphologist panel majority vote as *grounded truth*, of the three methodologies, automatically extracted image-based measurements provided the most accurate assessment of facial features
- When comparing the experienced eye to absolute measurements, subjective assessments of face width, widely spaced eyes, and jaw width were the most inaccurate



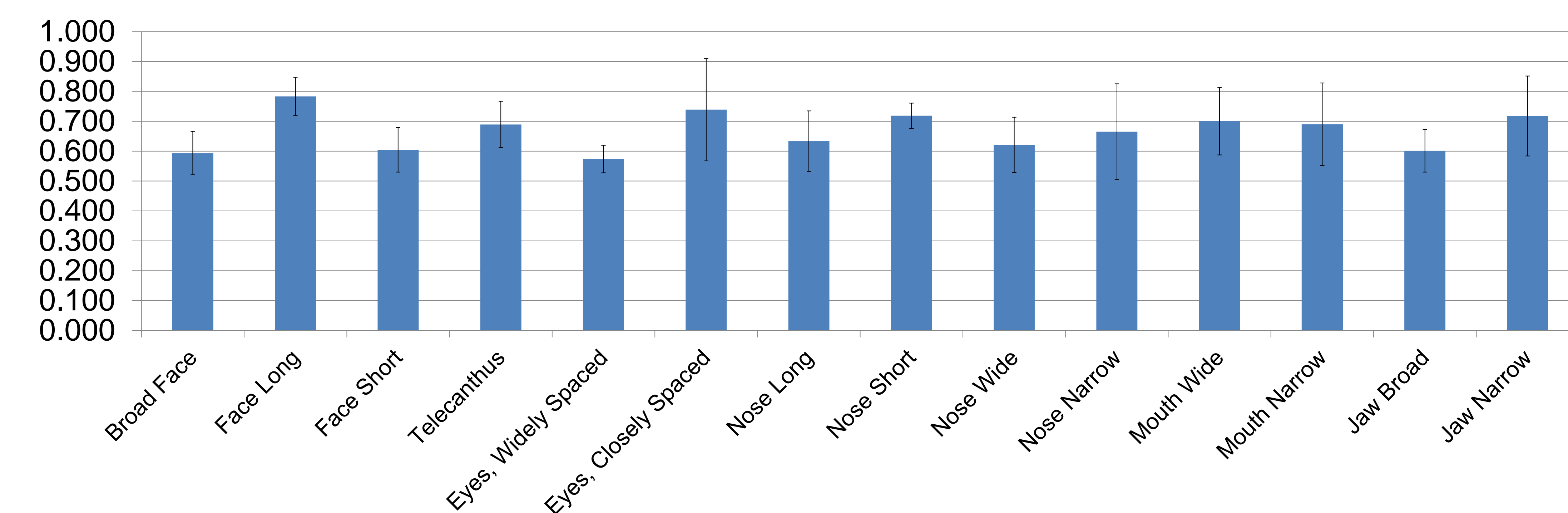
	Broad Face	Face Long	Face Short	Telecanthus	Eyes, Widely Spaced	Eyes, Closely Spaced	Nose Long	Nose Short	Nose Wide	Nose Narrow	Mouth Wide	Mouth Narrow	Jaw Broad	Jaw Narrow	Mean AUC	Combined Annotation
Anthropometric Measurements	0.419	0.825	0.980	0.773	0.577	0.900	0.711	0.802	0.793	0.971	0.684	0.761	0.585	0.535	0.737	0.883
Measurements Relative to OFC	0.660	0.617	1.000	0.791	0.749	0.993	0.763	0.749	0.819	0.771	0.714	0.768	0.718	0.801	0.779	0.882
Image Extracted Relative Measurements	0.821	0.952	1.000	0.772	0.801	0.964	0.947	0.883	0.835	0.841	0.773	0.888	0.844	0.912	0.874	0.911



Estimating Expert Variability in Feature Annotation

Method: We performed 6 experiments in a *leave-one-out* manner. Each time one expert is tested and the other five are used to create *ground truth*. AUC is used to measure the accuracy of the solo expert compared to the consensus of the five. The table shows AUC mean and SD for the six measured accuracies per facial feature.

Results: There is a significant amount of inter-expert variability, most notably for broad face, short face, widely spaced eyes and broad jaw.



Conclusions:

- When comparing the experienced eye to absolute measurements, subjective assessments of face width, widely spaced eyes, and jaw width are the most inaccurate
- Neither the expert majority vote nor the image measurement is consistently concordant with actual measurements
- Relative measurements, extracted automatically from facial photographs by FDNA® image analysis, identify most unusual features as reliably as a panel of experienced dysmorphologists observing the same images
- Since most individuals evaluated in the clinic would not be measured using calipers, and since expert panels are not accessible outside research settings, these results might suggest a clinical use for automatic facial analysis
- The current lack of norms for relative measurements hinders such clinical usage