Recognizing adult Aarskog-Scott syndrome carrier females based on craniofacial measurements using interaction variables and a surrogate covariance matrix

L. Wolf[1,2], L. Basel-Vanagaite[3,4,5], L. Karlinsky[1], M. Shohat[3,4,5], M. Porteous[6], R. Stevenson[7], J. Allanson[8]

1. FDNA Ltd., Herzlyia, Israel
2. The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
3. Schneider Children's Medical Center of Israel and Rabin Medical Center, Petah Tikva, Israel
4. Tel Aviv University, Tel Aviv, Israel
5. Felsenstein Medical Research Center, Petah Tikva, Israel
6. Centre for Molecular Medicine, University of Edinburgh, Edinburgh
7. Greenwood Genetic Center, Greenwood, SC USA
8. Dept. Genetics, Children's Hospital Eastern Ontario, Ottawa, ON, Canada.

Aarskog-Scott syndrome is an X-linked recessive syndrome caused by FDG1 mutations and characterized by dysmorphism, shawl scrotum, short stature and brachydactyly.

We demonstrate that carrier females are distinguishable based on craniofacial measurements. We evaluated 20 adult females, out of which 16 are obligatory carriers, 1 is a molecularly verified carrier, and 3 are verified non-carriers.

First, we compute the likelihood to be sampled from the matched control population for each of 21 craniofacial measurements. The combined likelihood scores demonstrate a moderate recognition rate with Area Under ROC Curve (auc) of 0.74. In order to improve the recognition rate, we consider the interaction variables that correspond to all pairs of measurements.

Since the individual measurements of the control population became unavailable once the statistics of each measurement were computed, we cannot directly estimate the distribution of the interaction variables. Instead, we employ a second data set of 21 adult female Navajos, for which five cranio-facial measurements that are common with the ones of the Aarskog-Scott syndrome carrier dataset are available per-person.

Employing only these five measurements and correcting for the correlations among the derived interaction variables, we are able to obtain an improved recognition rate (auc of 0.78). This is significantly higher than the recognition rate obtained from these five measurements without considering the interaction variables (auc of 0.65).

Therefore, our results highlight the utility of interaction variables in evaluation of facial features and support the usage of surrogate correlation matrices when such data are unavailable.

Recognizing adult Aarskog-Scott syndrome carrier females based on craniofacial measurements using interaction variables and a surrogate covariance matrix

Lior Wolf[1,2] Lina Basel-Vanagaite[1,3], Leonid Karlinsky[1], Mordechai Shohat[1,3], Mary Porteous[4], Roger Stevenson[5], Judith Allanson [6]

[1] FDNA Ltd., Herzlyia, Israel
[2] The School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel
[3] Raphael Recanati Genetics Institute, Rabin Medical Center, Israel.
[4] Centre for Molecular Medicine, University of Edinburgh, Edinburgh.
[5] Greenwood Genetic Center, Greenwood, SC USA.
[6] Dept. Genetics, Children's Hospital Eastern Ontario, Ottawa, ON, Canada.

Aarskog-Scott syndrome is an X-linked recessive syndrome caused by FDG1 mutations and characterized by dysmorphism, shawl scrotum, short stature and brachydactyly.

We demonstrate that carrier females are distinguishable based on craniofacial measurements. We evaluated 20 adult females, out of which 16 are obligatory carriers, 1 is a molecularly verified carrier, and 3 are verified non-carriers.

First, we compute the likelihood to be sampled from the matched control population for each of 21 craniofacial measurements. The combined likelihood scores demonstrate a moderate recognition rate with Area Under ROC Curve (auc) of 0.74.

In order to improve the recognition rate, we consider the interaction variables that correspond to all pairs of measurements. Since the individual measurements of the control population became unavailable once the statistics of each measurement were computed, we cannot directly estimate the distribution of the interaction variables. Instead, we employ a second data set of 21 adult female Navajos, for which five cranio-facial measurements that are common with the ones of the Aarskog-Scott syndrome carrier dataset are available per-person.

Employing only these five measurements and correcting for the correlations among the derived interaction variables, we are able to obtain an improved recognition rate (auc of 0.78). This is significantly higher than the recognition rate obtained from these five measurements without considering the interaction variables (auc of 0.65). Therefore, our results highlight the utility of interaction variables in evaluation of facial features and support the usage of surrogate correlation matrices when such data are unavailable.

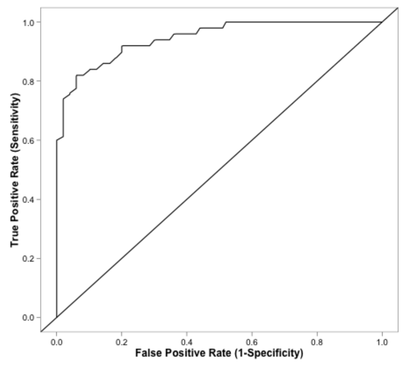
Aarskog-Scott syndrome female carrier

Evaluating discriminative power

Each variable is normalized with accordance to the matched population mean and standard deviation. The value of the variable after the normalization represents the probability of obtaining such a value by chance in the general population.

To evaluate the discriminative power of a single variable the area under the ROC curve is used.

When combining multiple variables, a product of the probabilities is used. This is sometimes called the "Naïve Bayes" method.



Interaction variables

A long ear on an unusually long head is less likely to be clinically dysmorphic than a long ear on a normally sized head. Interaction variables combine pairs of features to allow the computer to model such relations.

First, each variable is normalized with accordance to the mean and standard deviation of that measurement. Then, products of normalized variables are considered.

Surrogate covariance matrix

In order to combine the interaction variables, there is a need to normalize each such variable. The product of two normalized variables is *not* normalized, and its statistics depend on the covariance of these two variables.

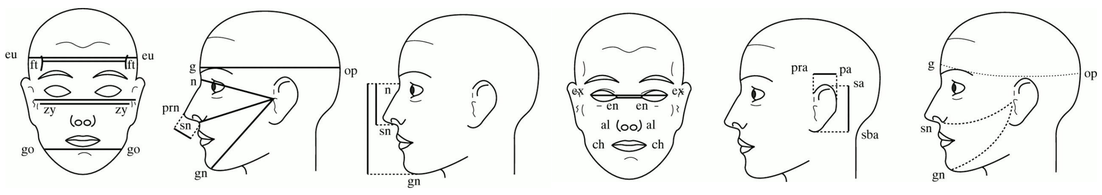
Here, since the required information was not retained, data from another survey (the Navajo data set) was used.

Description of measurements – Aarskog carrier survey

	Dysmorphic features	points		Dysmorphic features	points
1	Head Width	eu.eu	12	Upper Facial Height	n.sn
2	Skull Base Width	t.t	13	Nasal Width	al.al
3	Minimal Frontal	ft.ft	14	Mouth Width	ch.ch
4	Upper Facial Width	zy.zy	15	Inner Canthal Distance	en.en
5	Lower Facial Width	go.go	16	Outer Canthal Distance	ex.ex
6	Head Length	g.op	17	Ear Width	pra.pa
7	Upper Facial Depth	t.n	18	Ear Length	sa.sba
8	Mid Facial Depth	t.sn	19	Maxillary Arc	t.sn.t
9	Lower Facial Depth	t.gn	20	Mandibular Arc	t.gn.t
10	Nasal Protusion	sn.prn	21	Head Circumference	Ofc
11	Total Facial Height	n.gn			

Description of measurements – Navajo survey

	Dysmorphic features	Abbrv.	points
1	Head circumference	HC	Ofc
2	Outer canthal distance	OCD	ex.ex
3	Inner canthal distance	ICD	en.en
4	Interpupillary distance	IPD	
5	Ear length	EAR.L	sa.sba
6	Philtrum length	PH	sn.ls
7	Mouth width	MO	ch.ch



Results

	Experiment	AUC (measure of success)
1	21 Carrier features	74%
2	5 Common features	65%
3	Interactions of 5 Common features	78%

Practical implications for clinical and research use

- It is likely that Aarskog-Scott carriers are identifiable based on dysmorphology
- Interaction variables might provide better statistical diagnostics
- Imaging can facilitate the computation of a multitude of interactions
- Individual data (including facial images) should be retained and not just per-measurement statistics
- When such statistics are unavailable, covariance from another population can be used