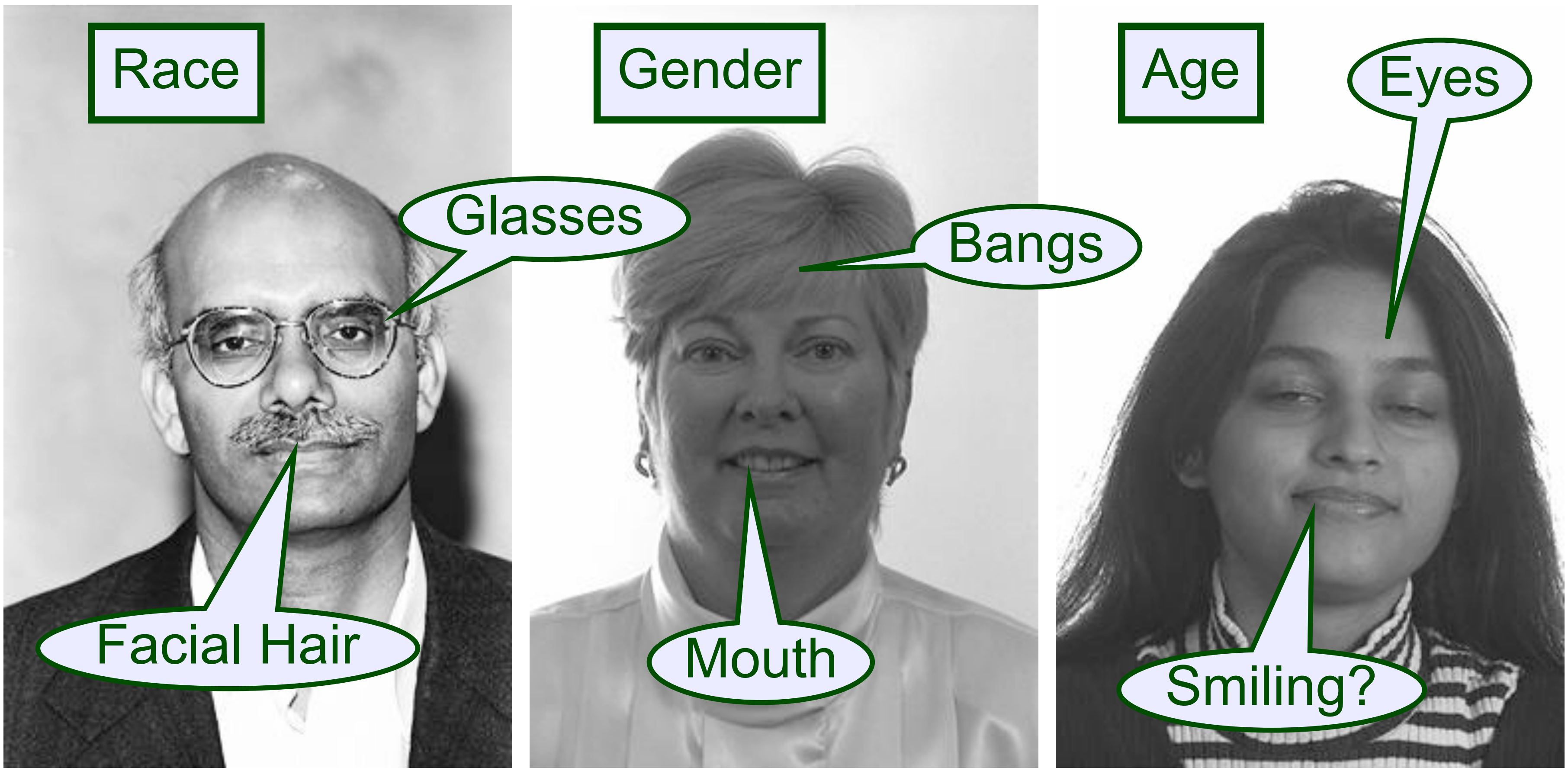
Linear and Generalized Linear Models for Analyzing Face Recognition Performance

J. Ross Beveridge
Colorado State University

Credit Where Credit is Due …

- Bruce Draper …….. CSU Computer Science
- Geof Givens ………. CSU Statistics
- Jonathon Phillips …. NIST
- Graduate Students
  – Wendy Yambor, Kai She, David Bolme, Kyungim Baek, Marcio Teixeira, David Bolme, Ben Randall, Trent Williams, Jilmil Saraf, Ward Fisher

What Factors (Covariates) ?



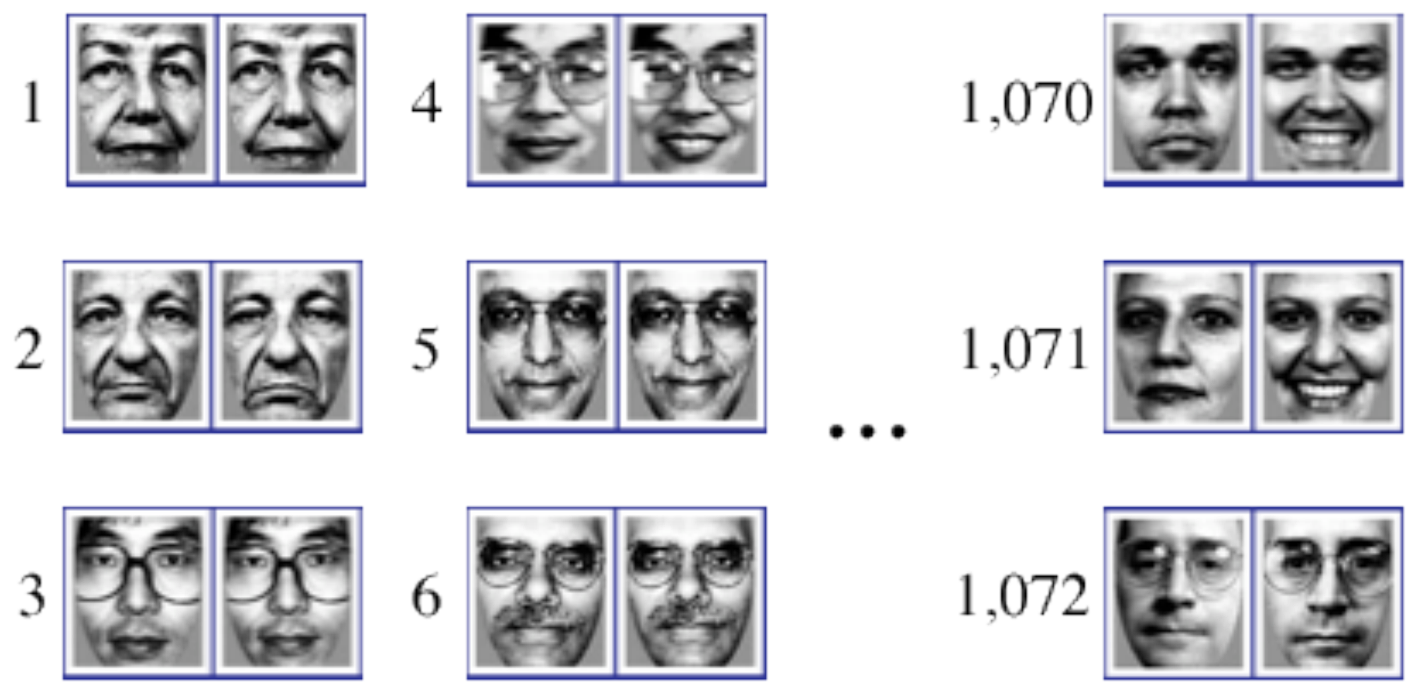
Subject Image Data

Yes, Yes, FER(R)ET Again …

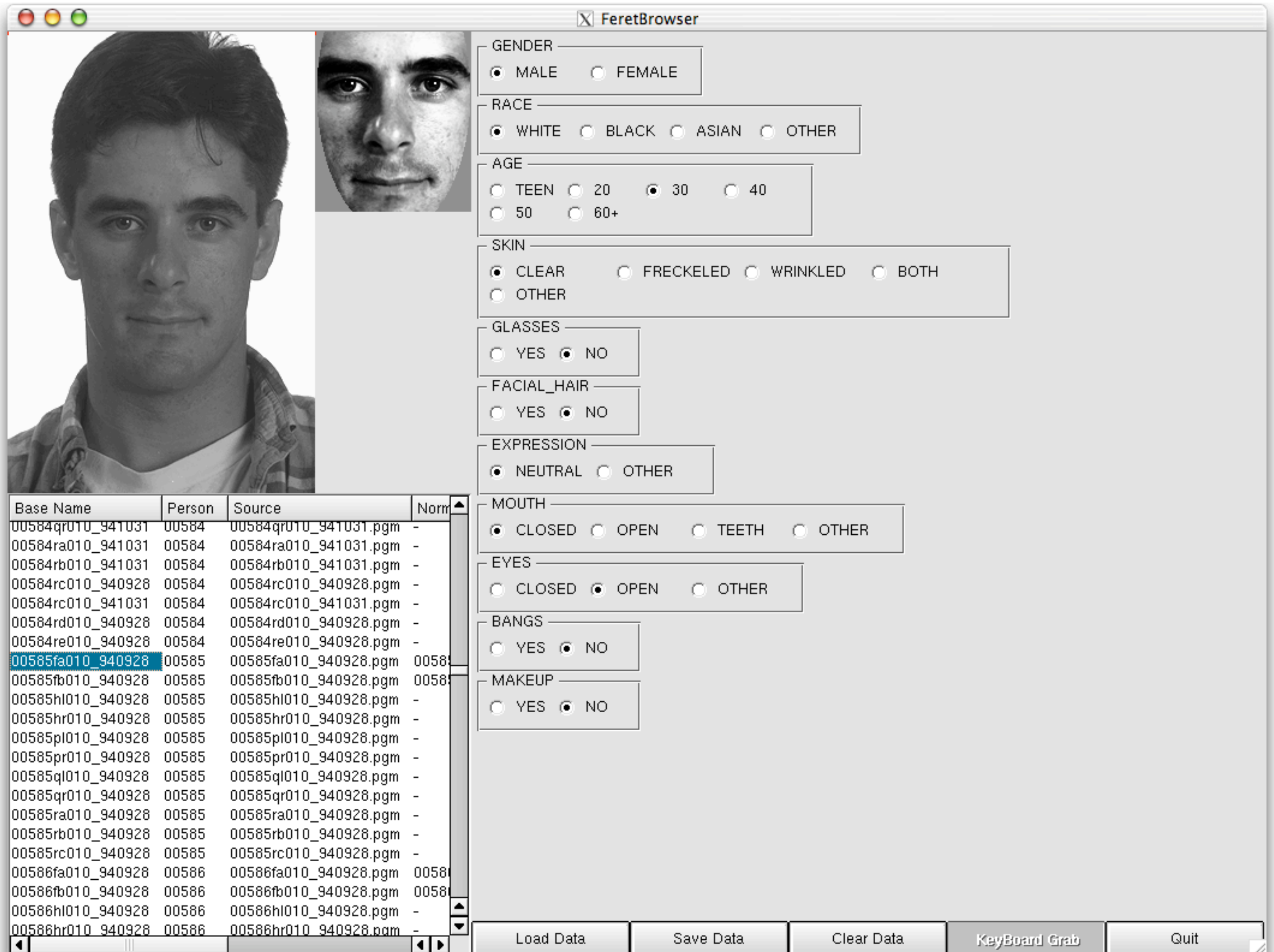
http://www.rollmop.org/ferrets/

Subject Image Data

- 1,072 Human Subjects from the FERET Data
- 2,144 FERET Images
- Exactly 2 images per subject, taken on same day



Collecting the Covariates



Our Subject Covariates

FERET Subject/Image Covariates				
Fixed Per Subject				
Age	*Young*	*Old*		
Gender	*Male*	*Female*		
Race	*White*	*Black*	*Asian*	*Other*
Skin	*Clear*	*Other*		
Fixed Per Image				
Bangs	*No*	*Yes*		
Expression	*Neutral*	*Other*		
Eyes	*Open*	*Other*		
Facial Hair	*No*	*Yes*		
Makeup	*No*	*Yes*		
Mouth	*Closed*	*Other*		
Glasses	No	Yes		

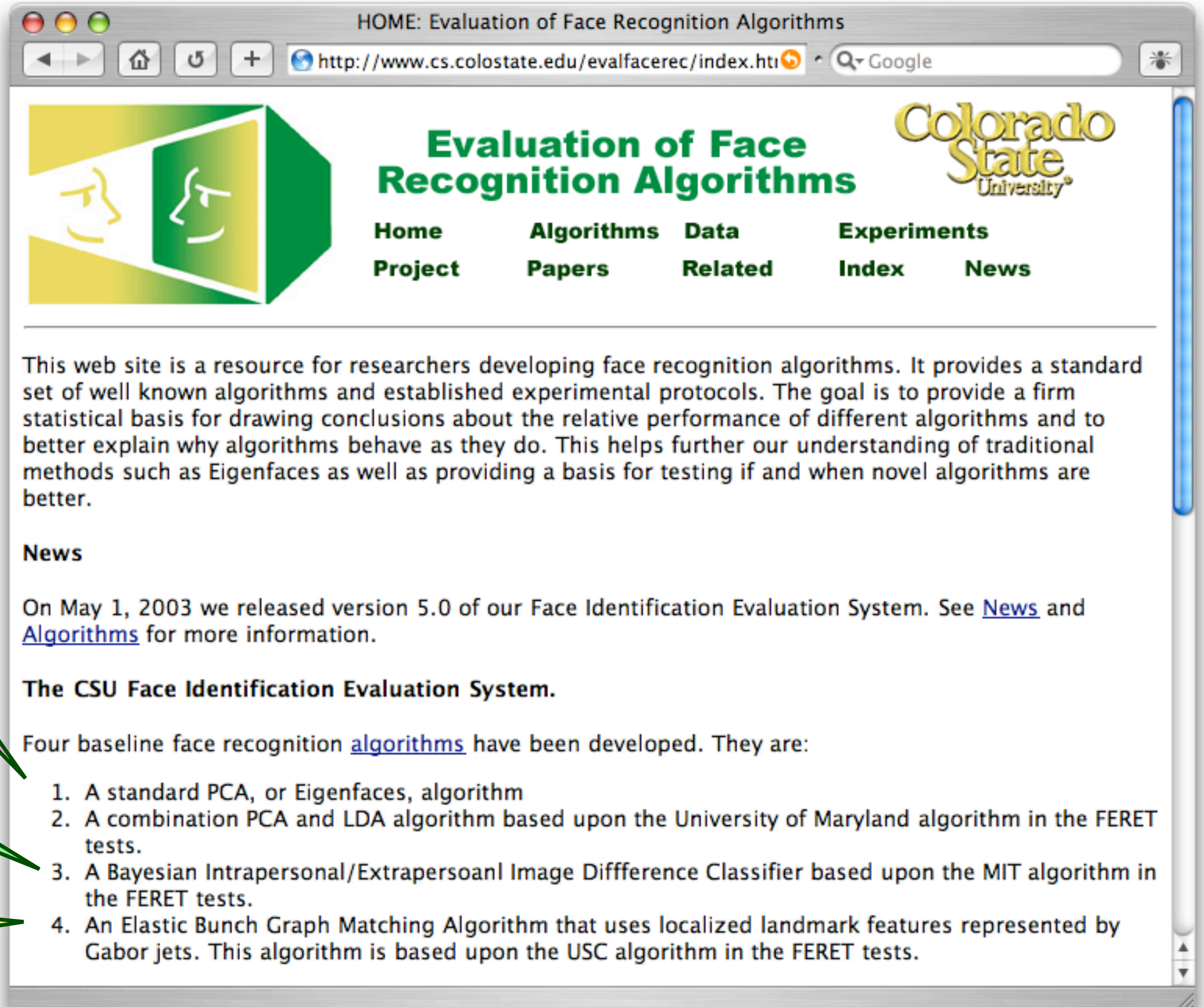
Standard Algorithms to Test

Three Algorithms

PCA

IIDC

EBGM


HOME: Evaluation of Face Recognition Algorithms

http://www.cs.colostate.edu/evalfacerec/index.htı

Evaluation of Face Recognition Algorithms

Home Algorithms Data Experiments

Project Papers Related Index News

This web site is a resource for researchers developing face recognition algorithms. It provides a standard set of well known algorithms and established experimental protocols. The goal is to provide a firm statistical basis for drawing conclusions about the relative performance of different algorithms and to better explain why algorithms behave as they do. This helps further our understanding of traditional methods such as Eigenfaces as well as providing a basis for testing if and when novel algorithms are better.

News

On May 1, 2003 we released version 5.0 of our Face Identification Evaluation System. See News and Algorithms for more information.

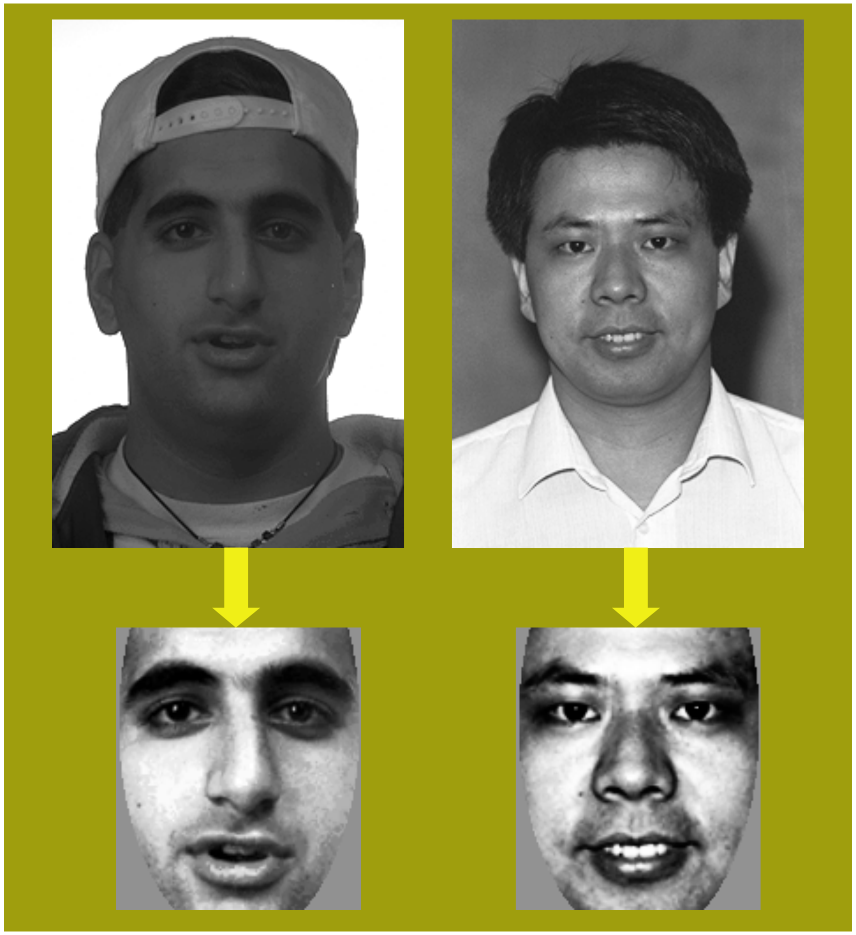
The CSU Face Identification Evaluation System.

Four baseline face recognition algorithms have been developed. They are:

1. A standard PCA, or Eigenfaces, algorithm
2. A combination PCA and LDA algorithm based upon the University of Maryland algorithm in the FERET tests.
3. A Bayesian Intrapersonal/Extrapersoanl Image Diffference Classifier based upon the MIT algorithm in the FERET tests.
4. An Elastic Bunch Graph Matching Algorithm that uses localized landmark features represented by Gabor jets. This algorithm is based upon the USC algorithm in the FERET tests.

http://www.cs.colostate.edu/evalfacerec/index.html

NIST FERET Image Preprocessing

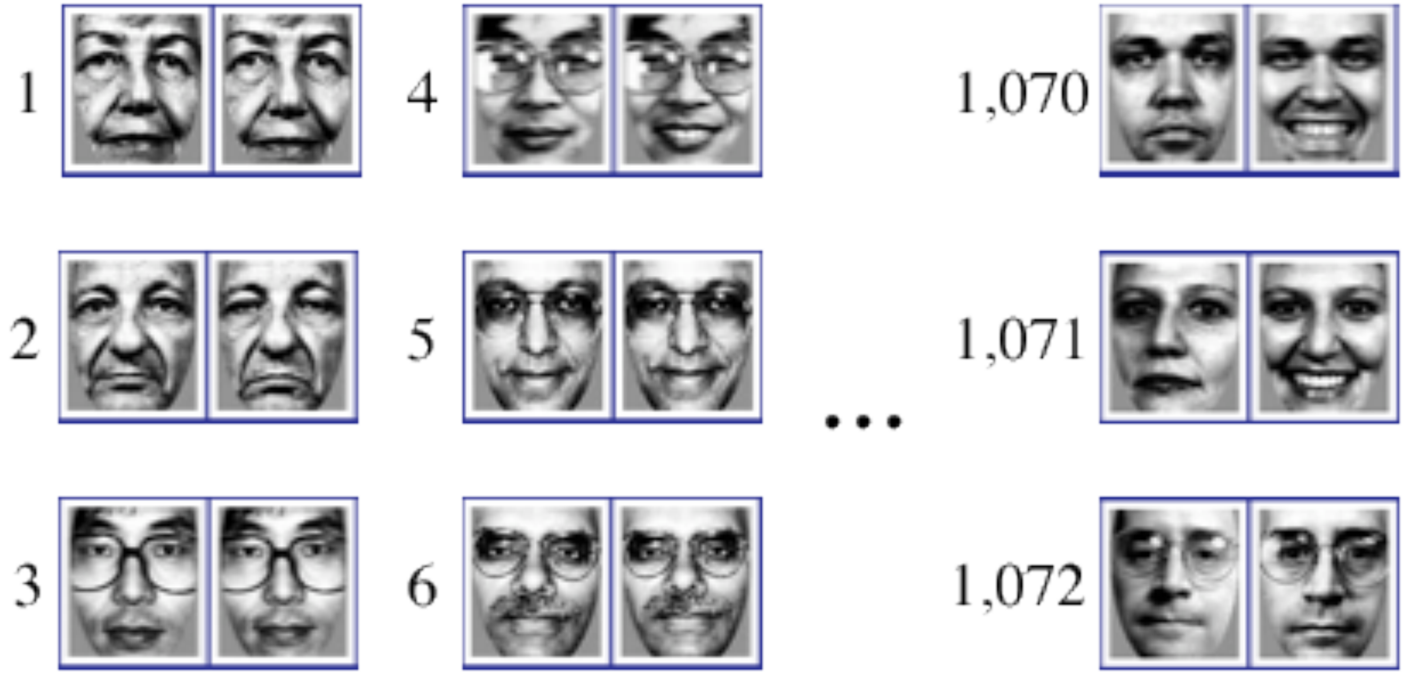


- Integer to float conversion
  - 256 gray levels to single-floats
- Geometric Normalization
  - Human chosen eye centers.
- Masking
  - Elliptical mask around face.
- Histogram Equalization
  - Equalize unmasked pixels
- Pixel normalization
  - Shift and scale pixel values so mean pixel value is zero and standard deviation over all pixels is one.

Refinement of NIST preprocessing used in FERET.

Training

- Best, but infeasible, solution
  - Disjoint images, same set of human subjects.
  - But, subject replicate images limited in FERET.
- Next best choice
  - Train on exactly those images used in the study.



Performance Variable?

- Recognition Rate?
  - Defined over a set of people, not per person.
- Similarity score?
  - Defined per person.
  - Linear models, ...
  - But, what does this tell us about actual performance?
- Probability of being recognized at Rank 1?
  - Defined per person.
  - Non-linear modeling problem.
- Probability of being correctly verified at given FAR?
  - Defined per person.
  - Non-linear modeling problem.

Statistical Modeling Overview

Sampled Normalized Similarity Scores

Covariates
- ***Algorithm***
- Age
- Race
- Gender
- Skin
- Glasses
- Facial Hair
- Makeup
- Bangs
- Expression
- Mouth
- Eyes

Linear Model (ANOVA)

Predict: similarity scores from covariate combinations

Statistical Modeling Overview

Sampled Normailzed Similarity Scores

Linear Model (ANOVA)

Predict: similarity scores from covariate combinations

Covariates

- ***Algorithm***
- Age
- Race
- Gender
- Skin
- Glasses
- Facial Hair
- Makeup
- Bangs
- Expression
- Mouth
- Eyes

Sampled Recognition Ranks

Generalized Linear Model

Predict: probability of correct recognition from covariate combinations

Linear Model - Similarity (Distance)

Y_i = Similiarity (Distance) metric for image pair *i*.

$\underline{X}_i$ = Algorithm & Human covariate factors for image pair *i*.

$\underline{\beta}$ = Parameters quantifying factor effects.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \varepsilon_i$$

with $\varepsilon_i \sim$ iid Normal$(0, \sigma^2)$

Generalized Linear Model
Pr(correct rank one recognition)

Y_i = Was the *i*th image pair matched at rank 1 ?
(i.e. $Y_i = 1$ if $R_i = 1$ and otherwise $Y_i = 0$)

$\underline{X}_i$ = Algorithm & Human covariate factors for image pair *i*.

$\underline{\beta}$ = Parameters quantifying factor effects.

$$g(\mu_{Y_i|\underline{X}_i}) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \varepsilon_i$$

$$Y_i \mid \underline{X}_i \sim f(\mu_{Y_i|\underline{X}_i}) \text{ independently}$$

Now: $g(z) = \log(z/(1-z))$, $f(\mu_{Y_i|\underline{X}_i}) = \text{Bernoulli}(\mu_{Y_i|\underline{X}_i})$

What Do Models Tell Us? PCA Algorithm Example.

Look at age holding all other covariates fixed.

Covariate	Base	Old
Age	Young	**Old**
Gender	Male	Male
Race	White	White
Skin	Clear	Clear
Bangs	No	No
Expression	Neutral	Neutral
Eyes	Open	Open
Facial Hair	No	No
Makeup	No	No
Mouth	Closed	Closed
Glasses	No	No

Similarity Scores - LM

- 13.0% Increase in similarity
- p-value < 0.0001
- Older is easier.

Pr(rank-one) - GLM

- Pr(crk=1) = 0.916 Base
- Pr(crk=1) = 0.951 Old
- p-value = 0.009
- Older is easier.

What Do Models Tell Us? PCA Algorithm Example.

Look at gender holding all other covariates fixed.

Covariate	Base	Old
Age	Young	Young
Gender	Male	**Female**
Race	White	White
Skin	Clear	Clear
Bangs	No	No
Expression	Neutral	Neutral
Eyes	Open	Open
Facial Hair	No	No
Makeup	No	No
Mouth	Closed	Closed
Glasses	No	No

Similarity Scores - LM

- 1.7% decrease in similarity
- p-value < 0.33
- Gender is not significant.

Pr(rank-one) - GLM

- Pr(crk=1) = 0.915 Base
- Pr(crk=1) = 0.884 Female
- p-value = 0.0925
- Gender is not significant

Model Validation & p-values

Table 1: ANOVA results for the linear model. 'B'='both images', 'O'='Other', 'Ch'='changes from one image to the other', and ':' indicates an interaction.

Predictor	Est.	S.E.	t	p
Intercept	-8.44	0.08	-107.76	< 0.0001
IIDC	5.48	0.11	49.46	< 0.0001
EBGM	3.54	0.11	31.98	< 0.0001
Old	-0.57	0.08	-7.09	< 0.0001
Female	0.18	0.09	2.14	0.0324
Afr.-American	-0.19	0.11	-1.76	0.0790
Asian	-0.64	0.10	-6.43	< 0.0001
O Race	-0.07	0.12	-0.59	0.5534
O Skin	-0.29	0.09	-3.08	0.0021
B Bangs	-0.82	0.08	-9.74	< 0.0001
Bangs Ch	-1.08	0.19	-5.63	< 0.0001
B O Expression	0.65	0.15	4.39	< 0.0001
Expression Ch	1.63	0.08	19.94	< 0.0001
B Eyes Not Open	-1.66	0.32	-5.22	< 0.0001
Eyes Ch	1.56	0.11	13.79	< 0.0001
B Facial Hair	0.25	0.10	2.40	0.0164
Facial Hair Ch	-0.75	0.32	-2.34	0.0191
B Glasses	-2.43	0.13	-18.14	< 0.0001
B Makeup	-0.23	0.11	-2.02	0.0439
Makeup Ch	0.32	0.26	1.23	0.2179
B O Mouth	[illegible]	[illegible]	[illegible]	[illegible]
Mouth Ch	[illegible]	[illegible]	[illegible]	[illegible]
IIDC : Old	0.37	0.11	3.22	0.0013
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 2: Summary of generalized linear model results.

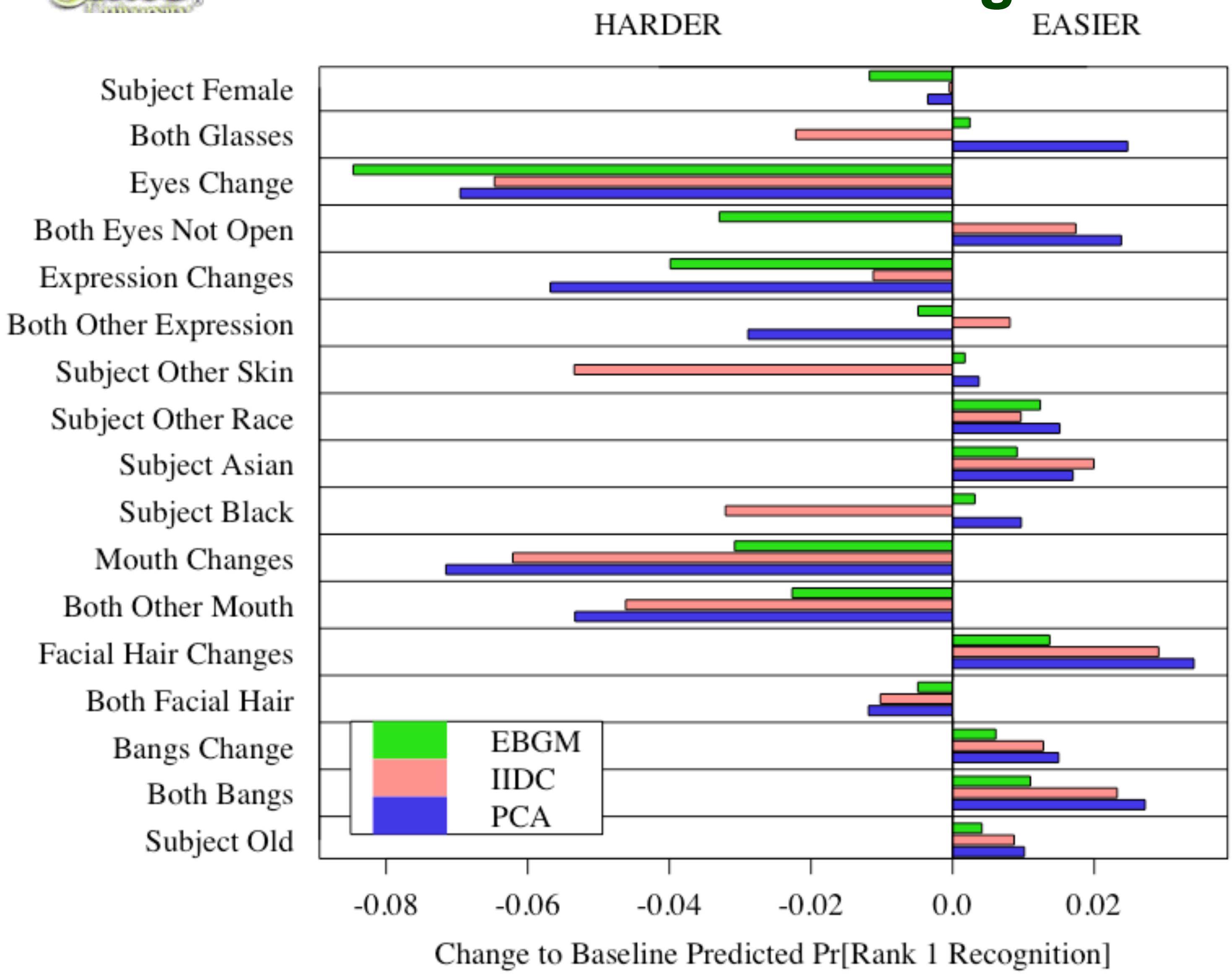
	df	ΔDeviance	p
Intercept	1	*Note 1*	
Algorithm	2	*Note 2*	
Age	1	5.73	0.0167
Bangs	2	63.99	< 0.0001
Facial Hair	2	11.12	0.0039
Mouth	2	76.50	< 0.0001
Race & Alg. : Race	9	46.48	< 0.0001
Skin & Alg. : Skin	3	24.00	< 0.0001
Expr. & Alg. : Expr.	6	54.64	< 0.0001
Eyes & Alg. : Eyes	6	131.87	< 0.0001
Glasses & Alg. : Glasses	3	8.15	0.0430
Gender & Alg. : Gender	3	9.55	0.0228

Note 1 The null model deviance is 4,266.9 on 6,425 df. The model using all terms given above has residual deviance of 3,676.9 on 6,386 df—highly significant.

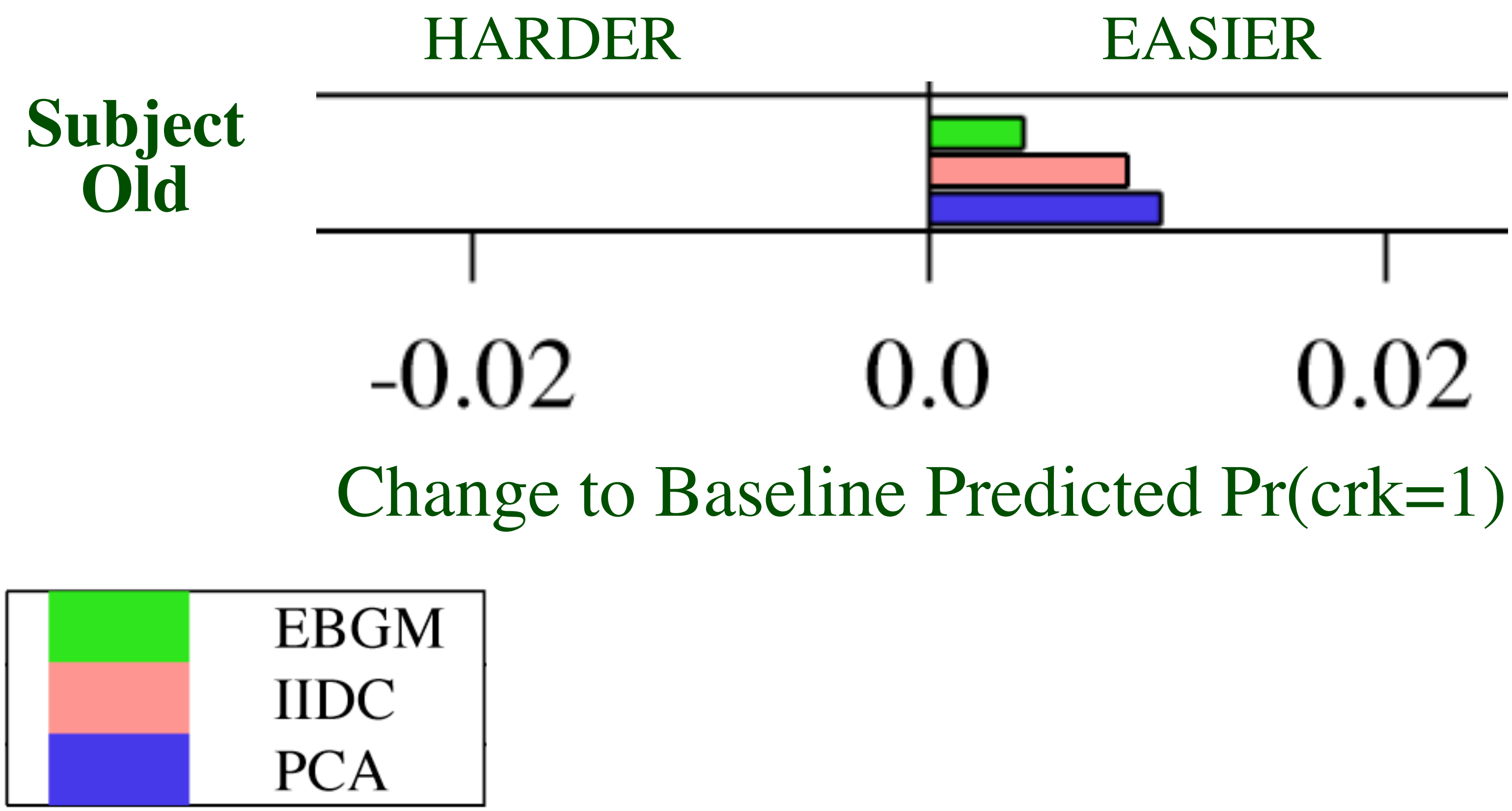
Note 2 The factor indicating algorithm has many significant interactions in this model and is highly significant. In a table organized to show subject covariate effects, an analogous test for algorithm would be distracting.

- Don't try to read this…
- Standards for evaluating and reporting results important.

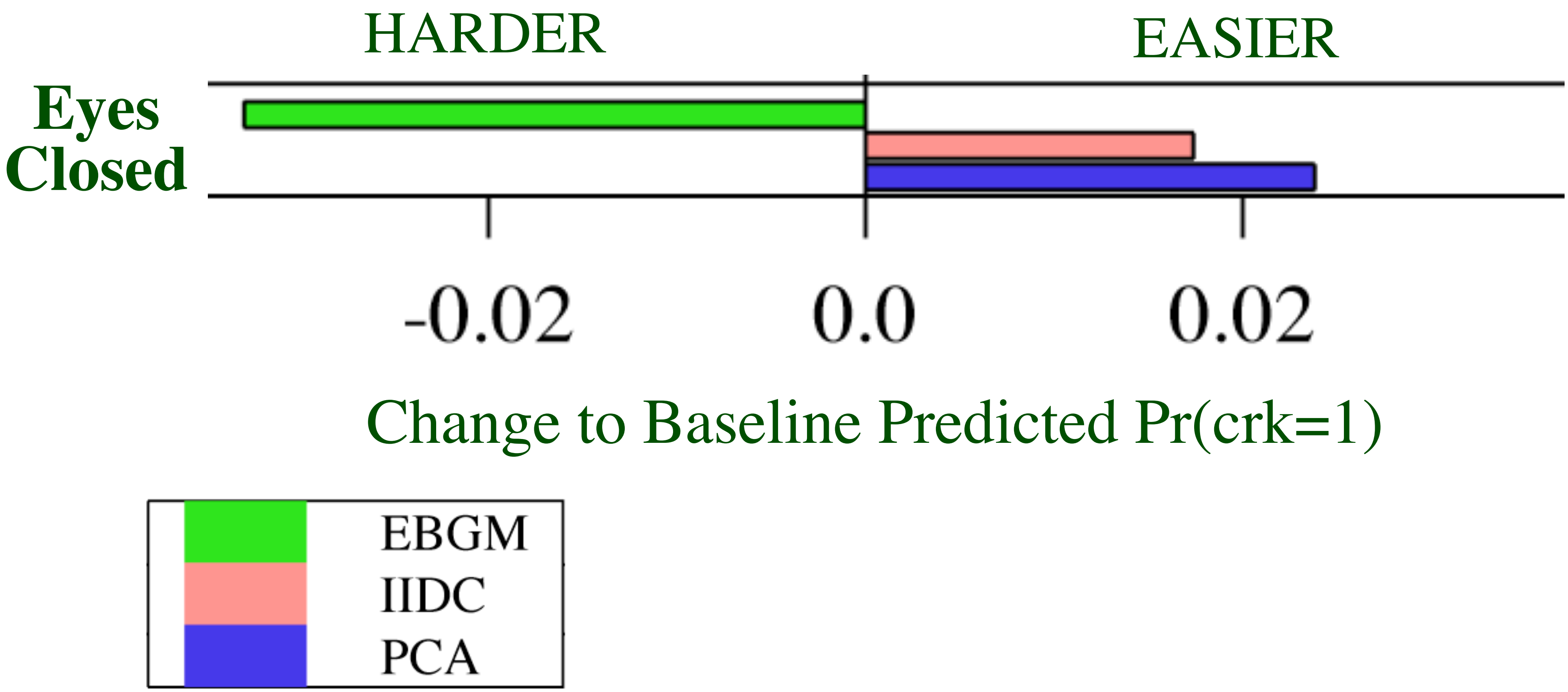
GLM with Three Algorithms



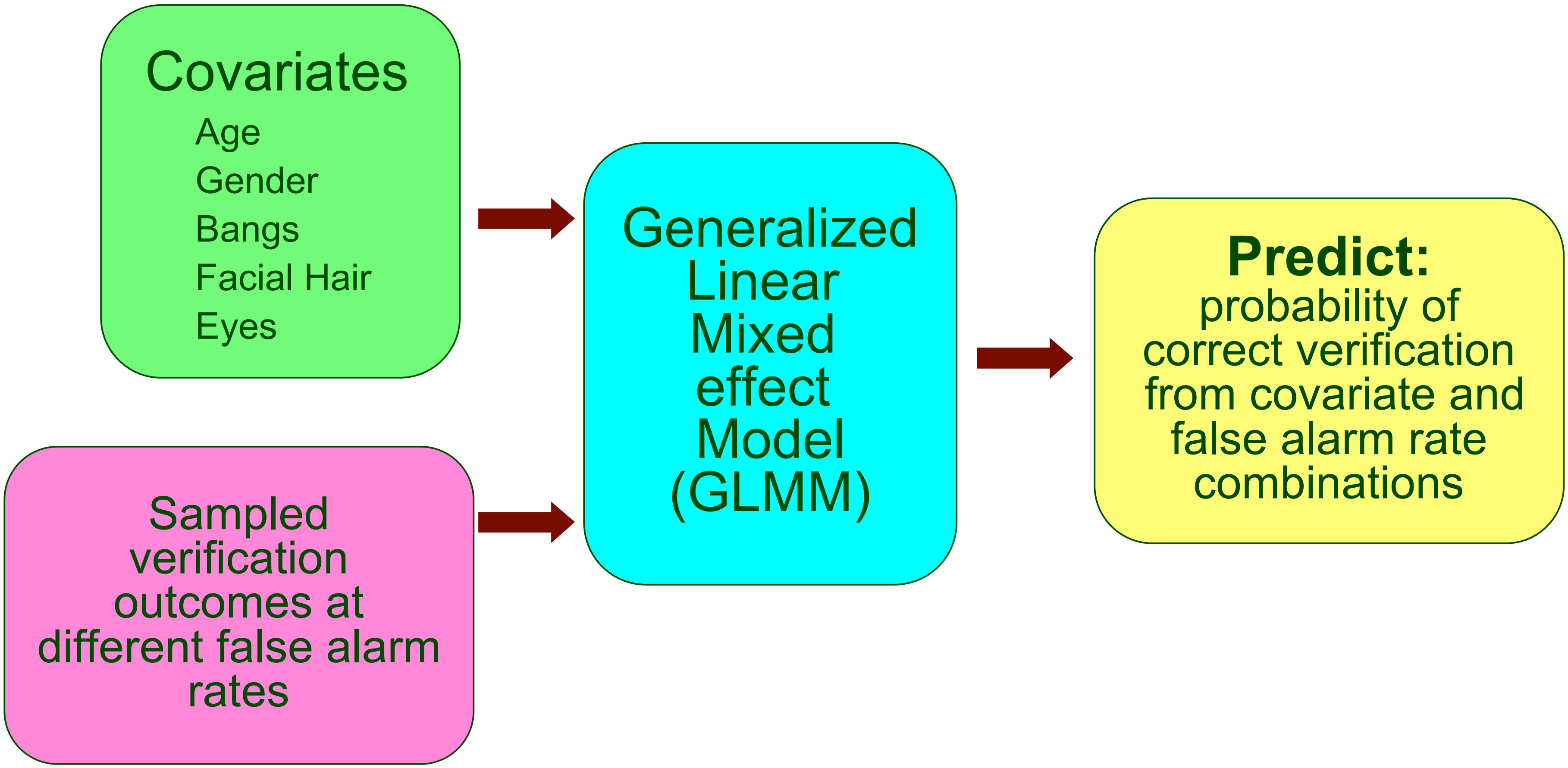
Age: Young vs. Old



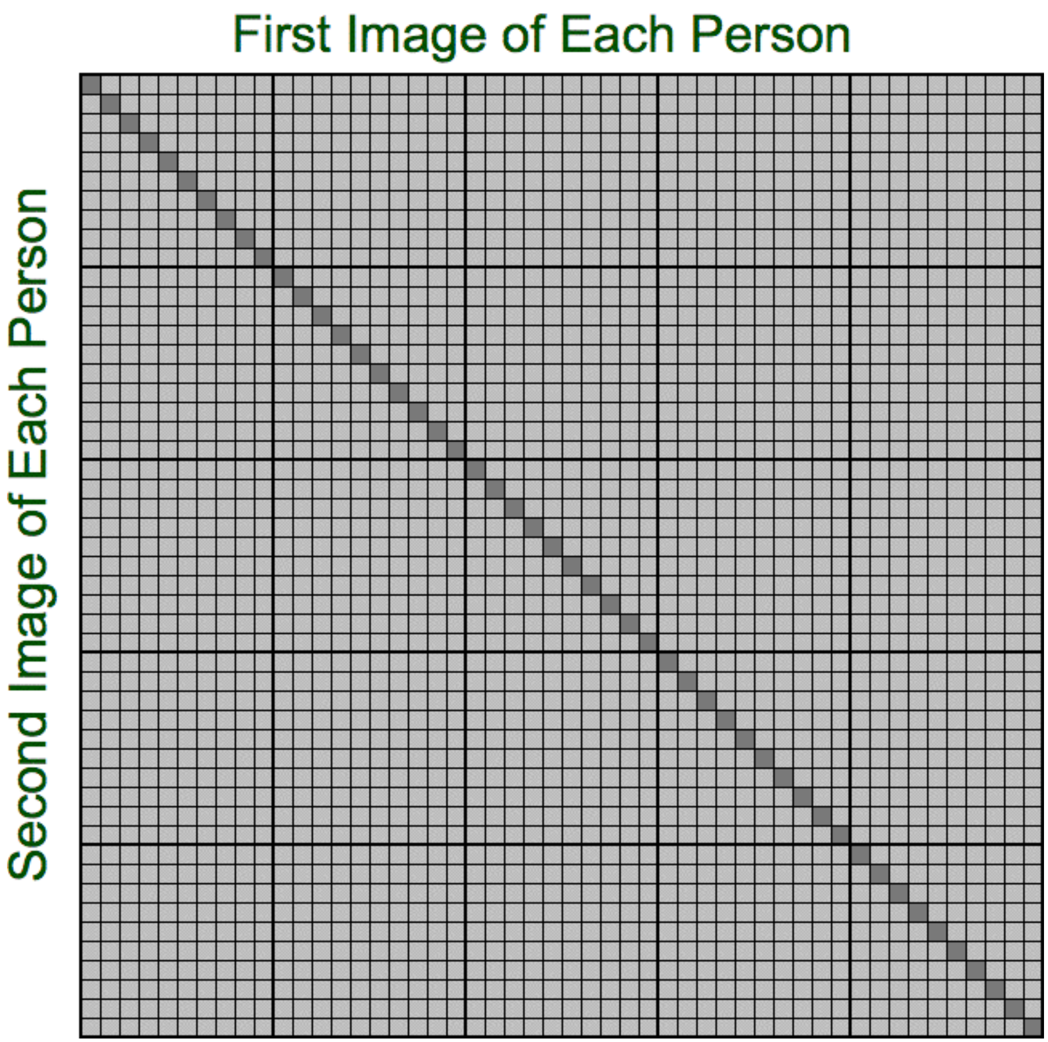
Eyes: Open vs. Closed



Verification Performance



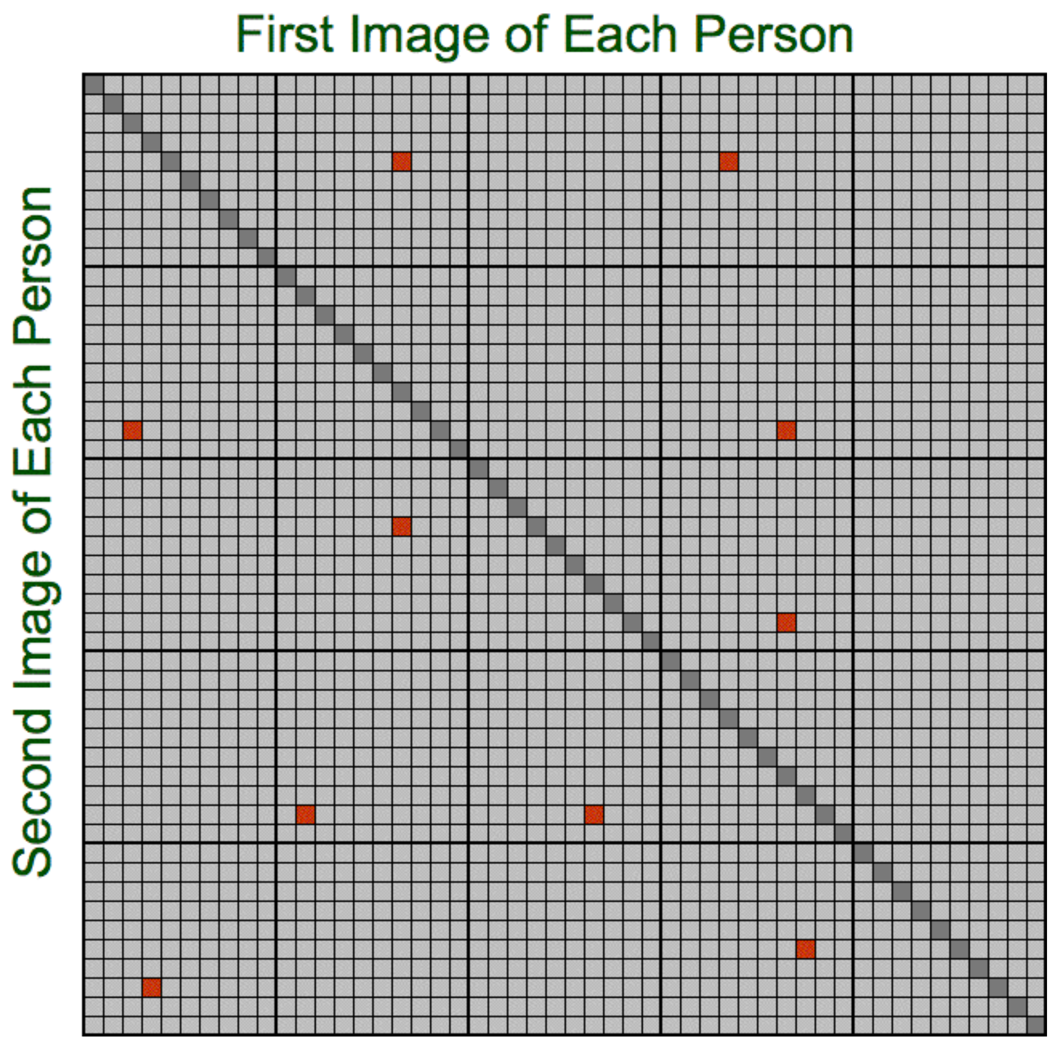
Verification Outcomes at Fixed False Alarm Rate α



Two Images per Subject Example

50 x 50 Similarity Matrix

Verification Outcomes at Fixed False Alarm Rate α



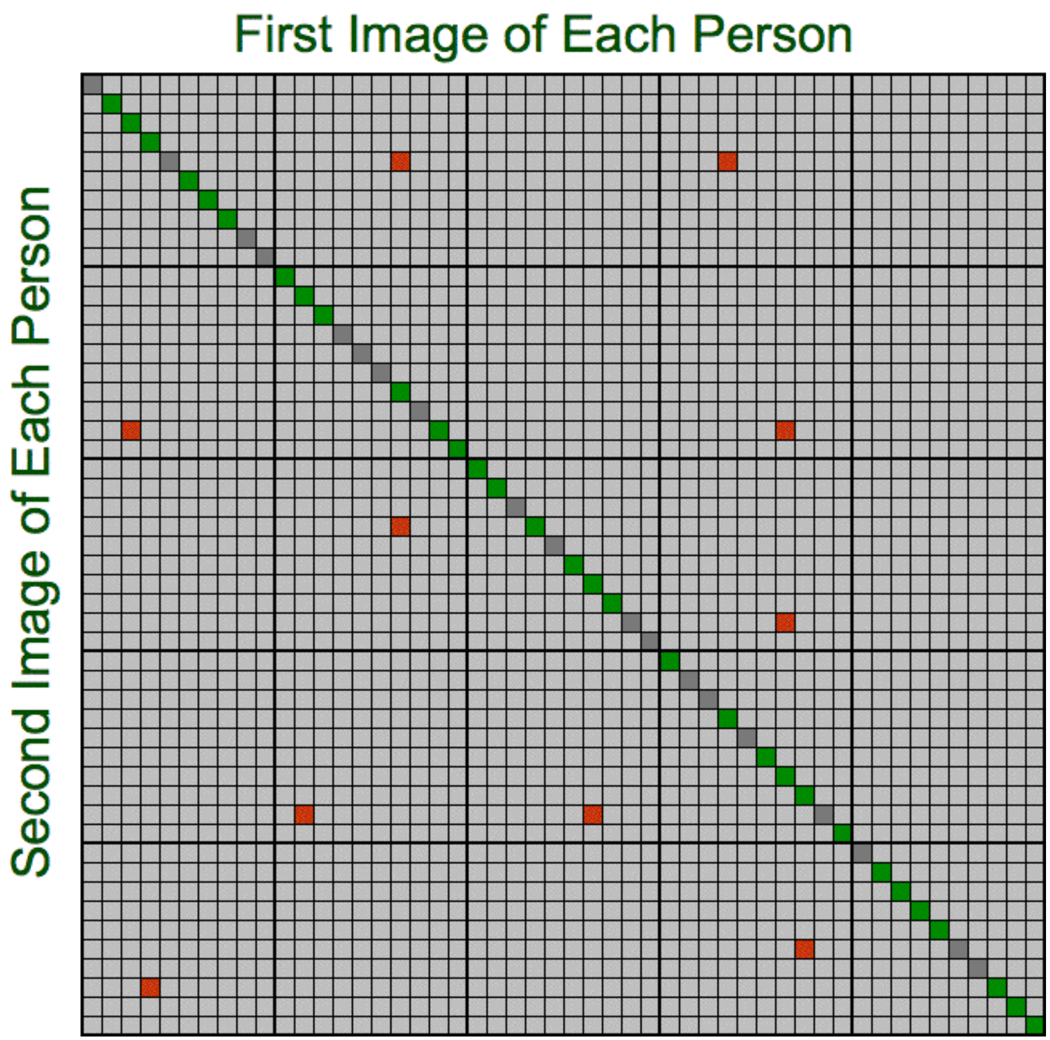
Two Images per Subject Example

50 x 50 Similarity Matrix

1) **Set FAR α,**

   e.g. $\alpha = 1/250$

Verification Outcomes at Fixed False Alarm Rate α



Two Images per Subject

Example

50 x 50 Similarity Matrix

1) **Set FAR α,**

   e.g. α = 1/250

2) **Indicate people correctly verified at threshold corresponding to α**

Verification Indicator Variable and FAR settings

- Our study - 1,072 x 1,072 similarity matrix.
  - 1,072 match scores,
  - 1,148,112 non-match scores.

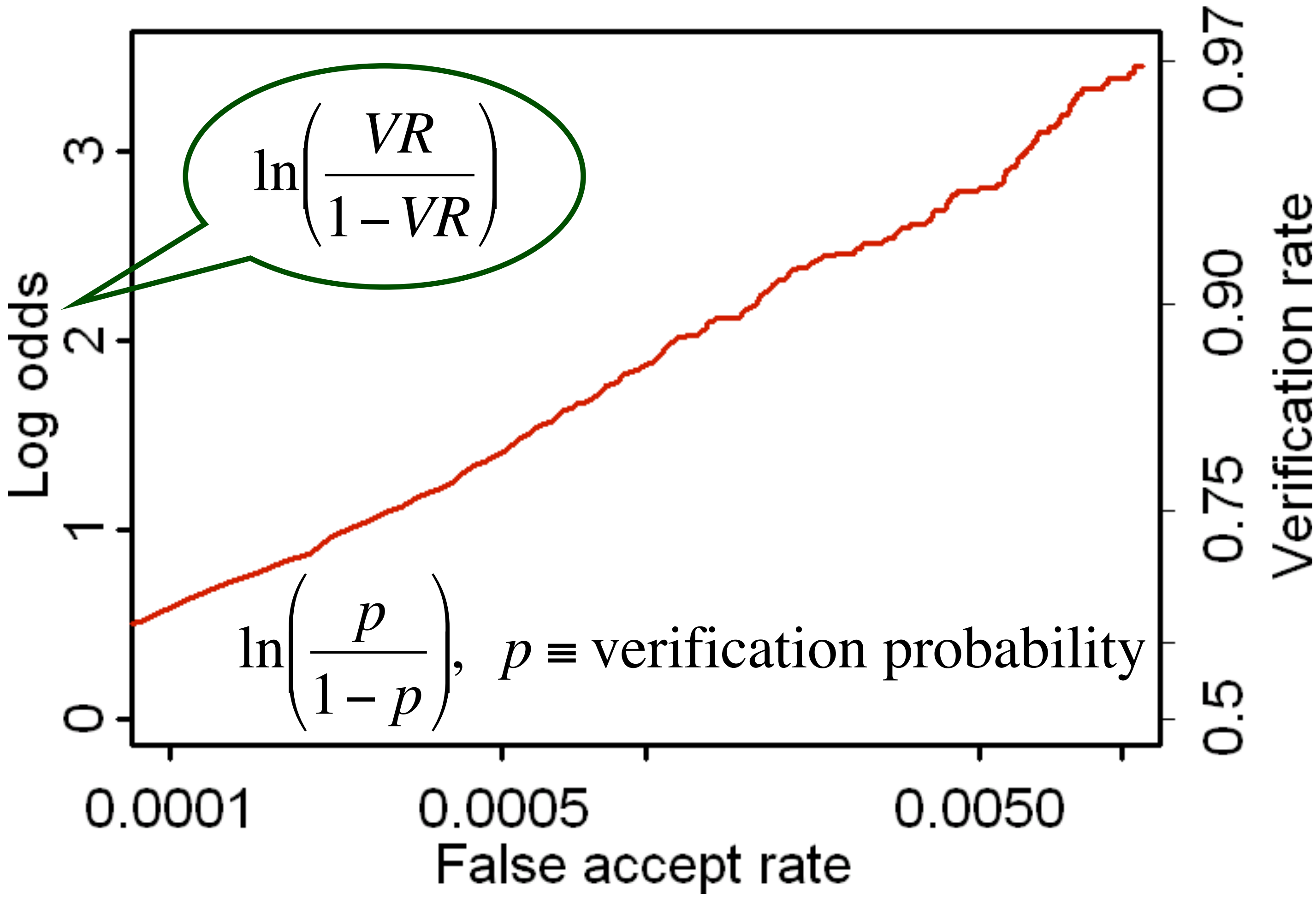
Indicator Variable *Y* for each subject for each FAR setting:

1 verified
0 otherwise

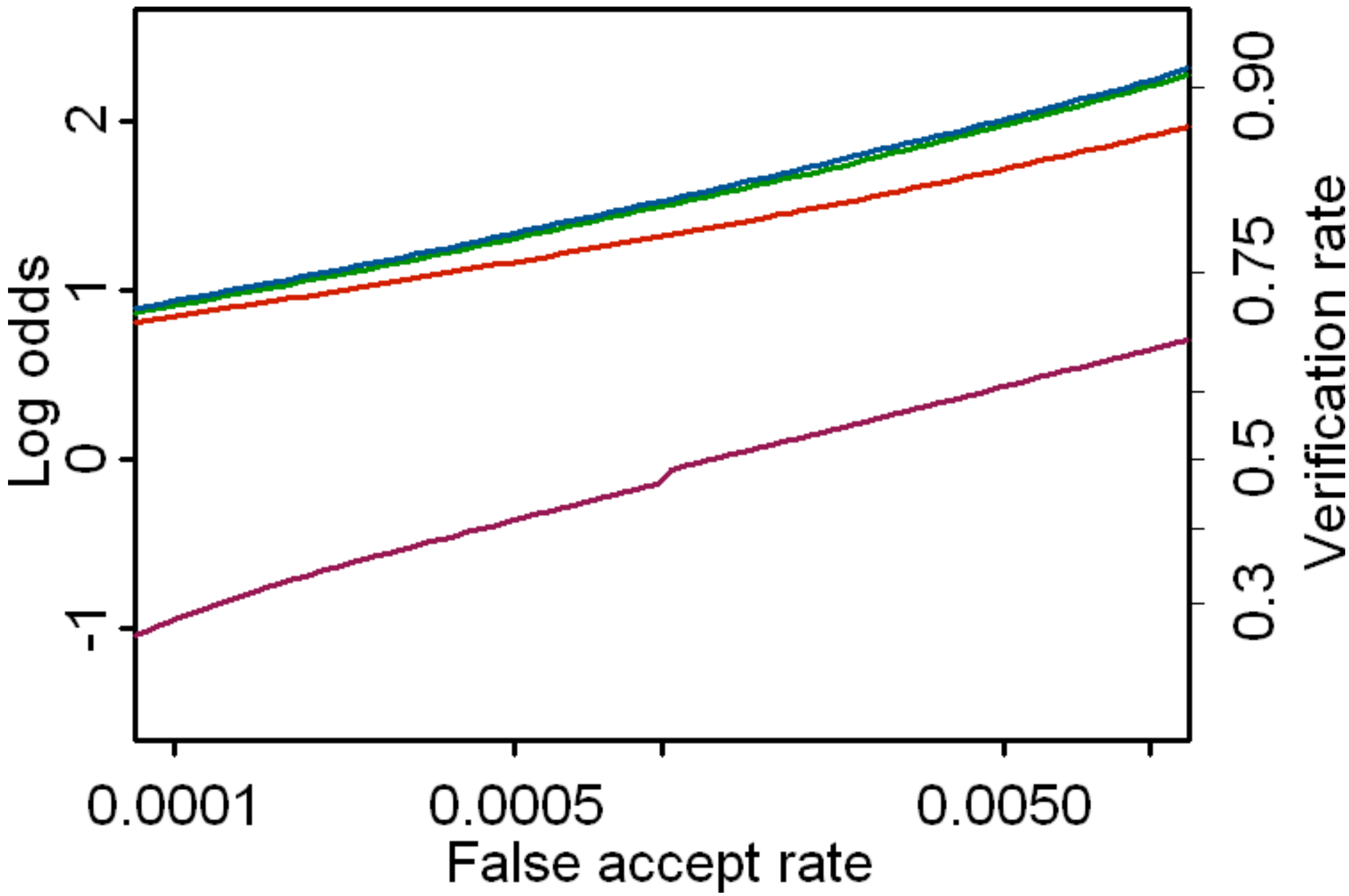
7 settings total.

Setting	FAR (α)	Rate per 10,000
1	1/10,000	1
2	1/5,000	2
3	1,2,500	4
4	1/1,000	10
5	1/500	20
6	1/250	40
7	1/100	100

Linearity of Log Odds against Log FAR - FERET+PCA



Linearity of Log Odds against Log FAR - FRVT



Generalized Linear Mixed Model (GLMM)

Analysis is: ***Mixed Effects Logistic Regression with Repeated Measures on People.***

- Let A and B be 2 factors that might influence algorithm performance. For example, age and gender.
  - Example factor settings A=a and B=b.
- Let j index the FAR setting, α_j
- Y_{pabj} is
  - 1 if Person p is verified correctly,
  - 0 otherwise.
- Y_{pabj} depends on:
  - person p,
  - factors A and B, and
  - false alarm rate α_j.

GLMM Model Continued …

Y_{pabj} is Bernoulli R.V. with success probability p_{pabj}

$$\log\left(\frac{p_{pabj}}{1-p_{pabj}}\right) = \mu + A_a + B_b + \gamma_j \log(\alpha_j) + A_a \gamma_{aj} \log(\alpha_j) + \pi_p$$

μ	=	grand mean
A_a	=	effect of setting a of factor A
B_b	=	effect of setting b of factor B
$\gamma_j \log(\alpha_j)$	=	log linear effect of α_j
$\gamma_{aj} A_a \log(\alpha_j)$	=	interaction effect
π_p	=	subject id. random effect (next page)

Subject Variation - The Mixed in Generalized Linear **Mixed** effect Model

$$\left[\pi_1,\ ..\ ,\ \pi_{1,072}\right]^T \sim \text{ Multivariate Normal where}$$

$$\mathrm{E}\left(\pi_p\right) = 0,\ \ \mathrm{Var}\ \ \pi_p \ = \ \sigma_\pi^2,$$

$$\mathrm{Cor}\left(y_{pab\alpha},\ y_{p'a'b'\alpha'}\right) = \begin{cases} \phi & \text{if} \quad p = p' \\ 0 & \text{if} \quad p \neq p' \end{cases}$$

This means:

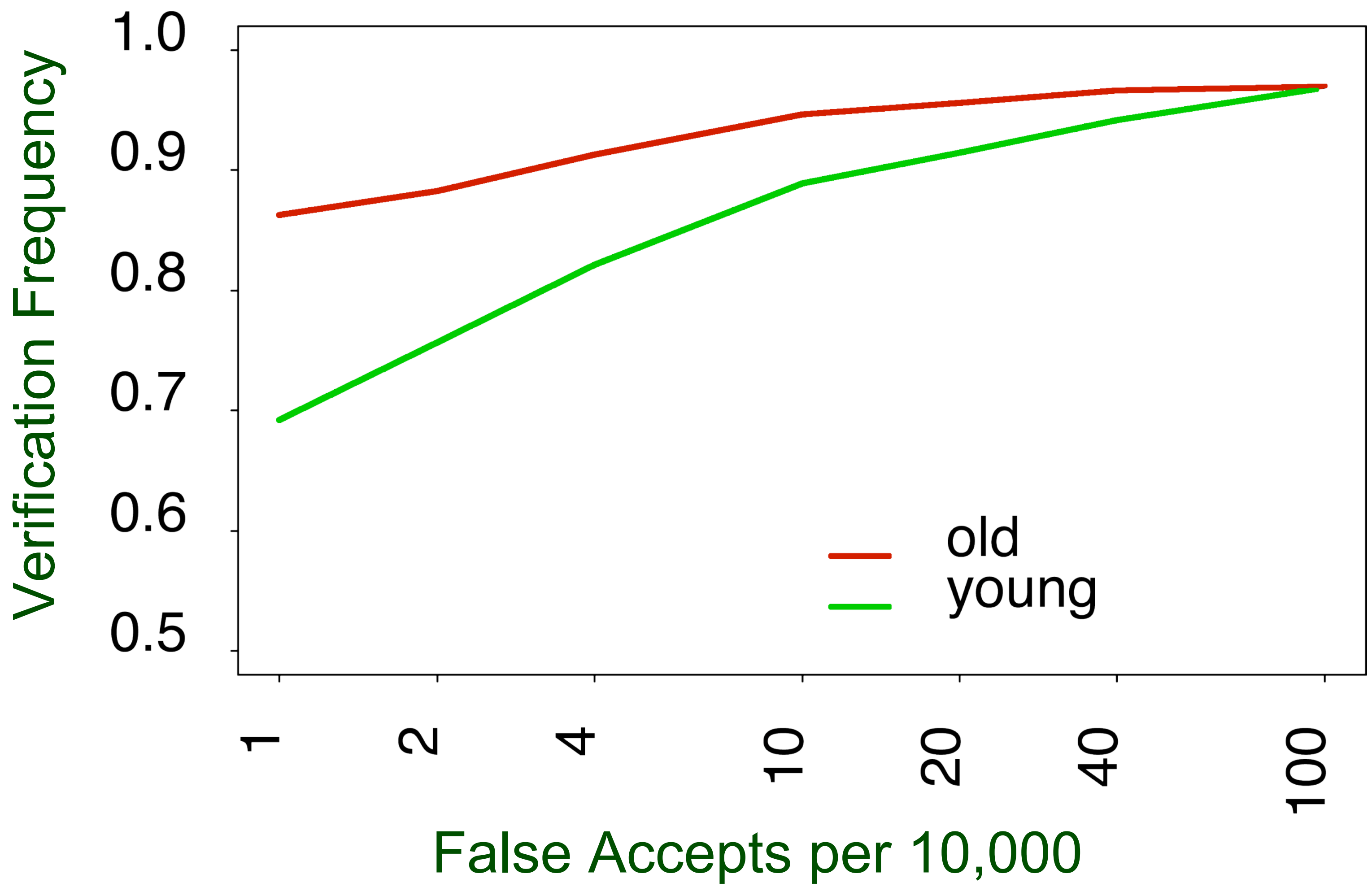
The outcomes, i. e. verification success/failure, are uncorrelated when testing different people but correlated when testing the same person under different configurations.

Random Effects are Important GLMM vs. GLM

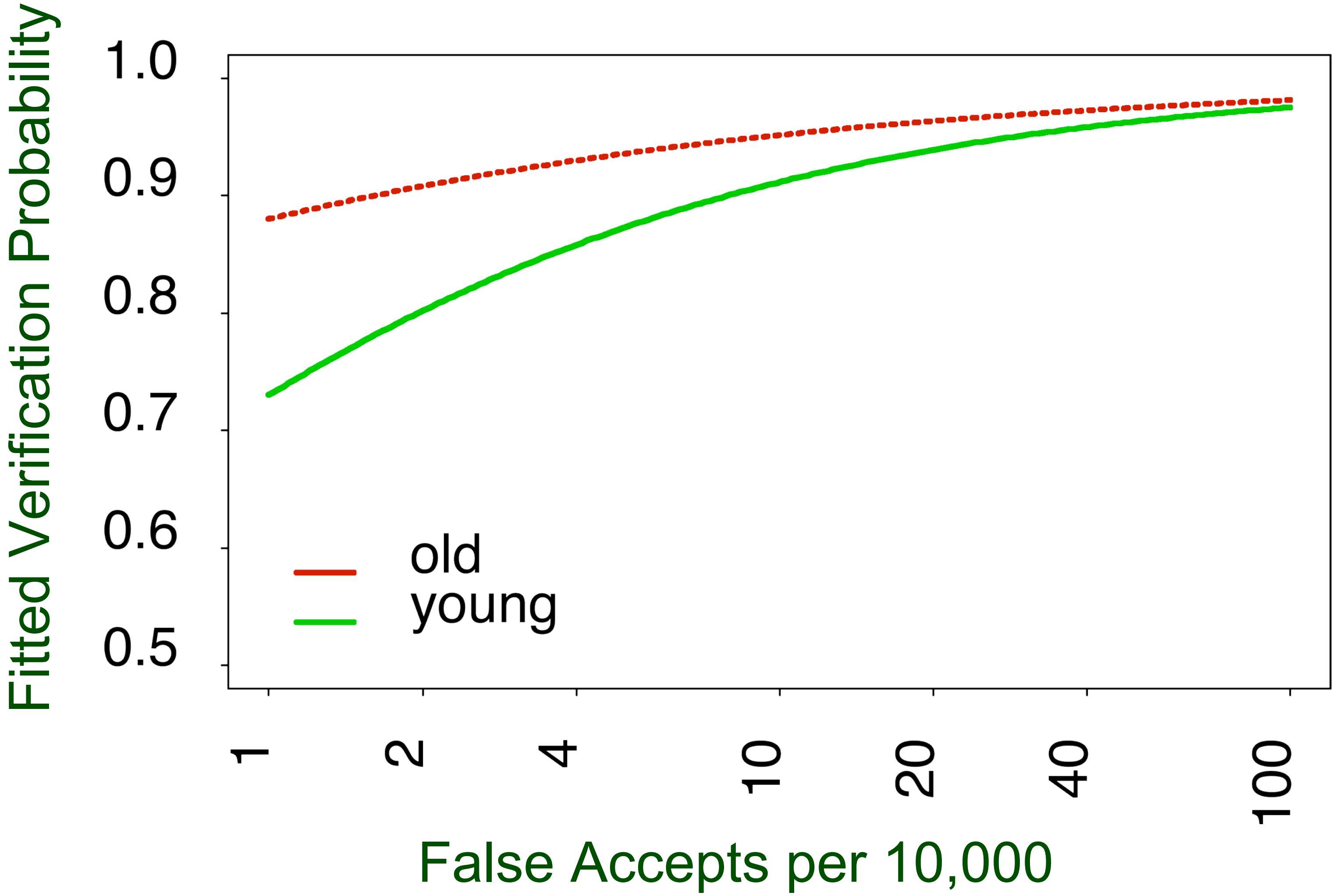
- Some people are harder to recognize then others.
- But, we don't care who specifically is hard or easy.

Removing the "noise" of random effects helps reveal other significant effects of interest.

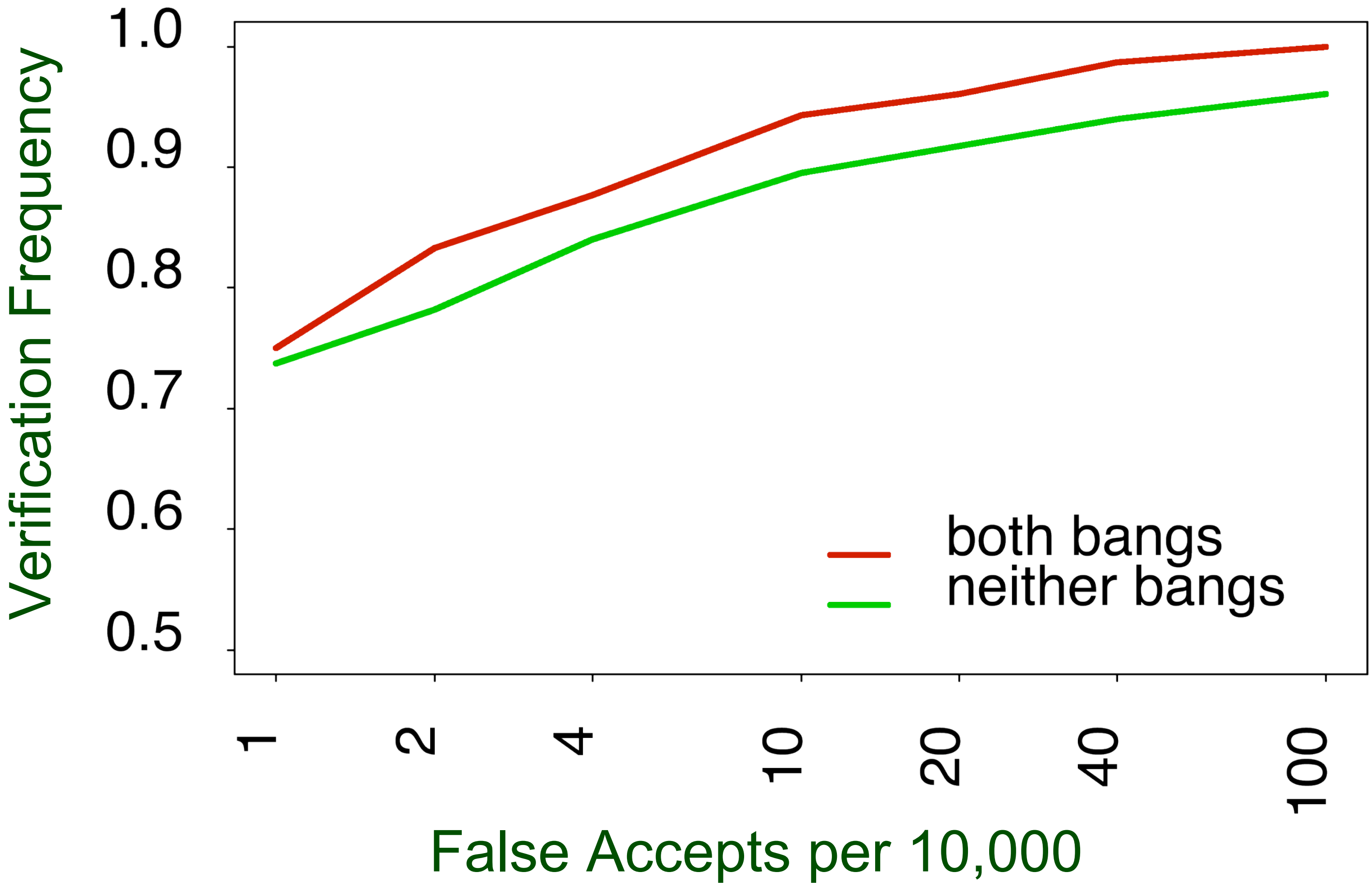
Marginal Verification Rates - Age



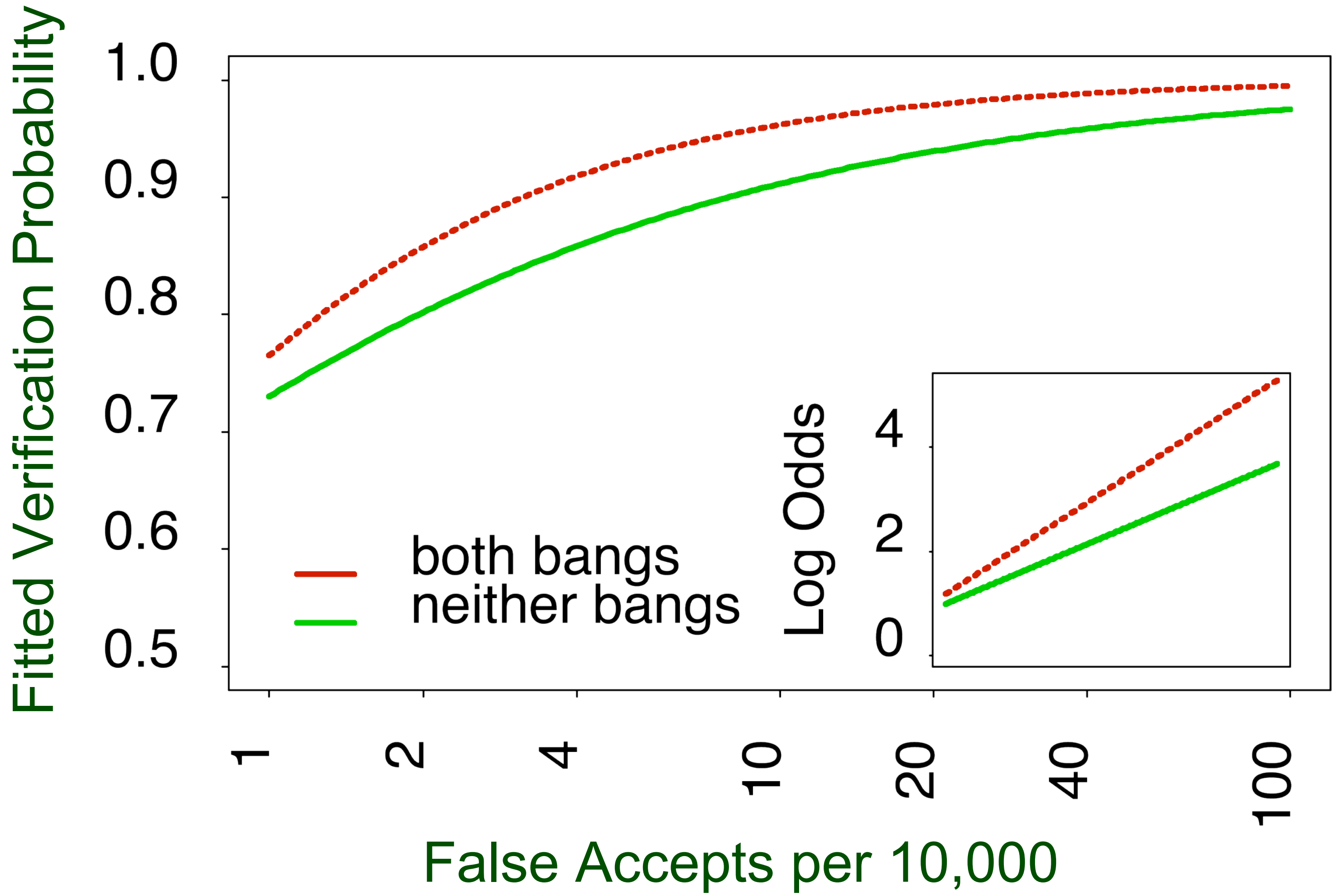
Results of the Model - Age



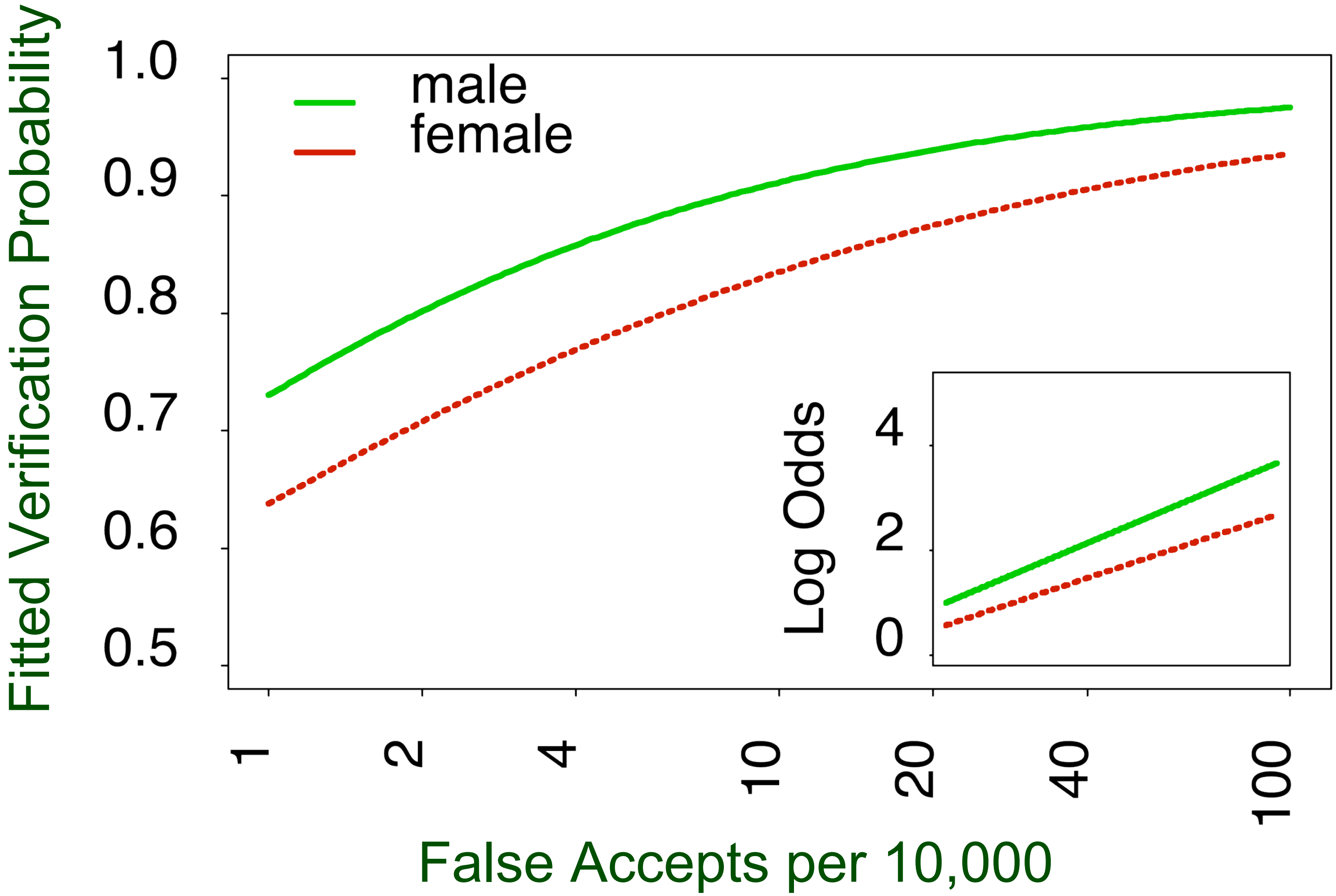
Marginal Verification Rates - Bangs



Results of the Model - Bangs



Results of the Model - Gender



Step Back: Why use Linear Models and Generalized Linear Models

F_1

F_2

F_3

$\vdots$

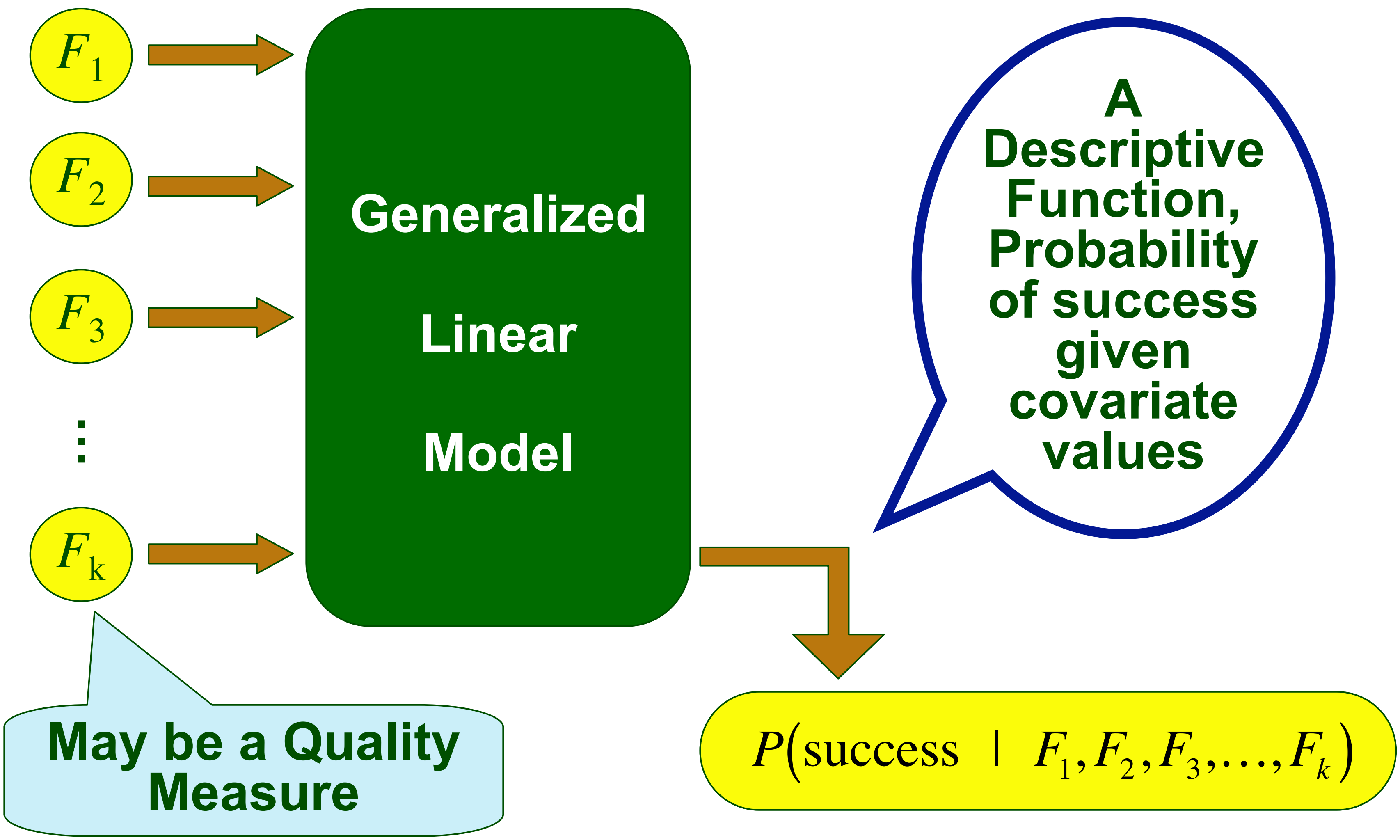
F_k

Start with a set of factors - covariates

These may be …

- ***Properties of the subject: age, etc.***
- ***Properties of the scene: lighting, etc.***
- ***Properties of the image:***
  - ***Focus***
  - ***Resolution***
  - ***Contrast***
  - ***…***

Step Back: Why use Linear Models and Generalized Linear Models



Thank You

LM with Three Algorithms