On the Persistence of Fingerprints

Soweon Yoon and Anil K. Jain

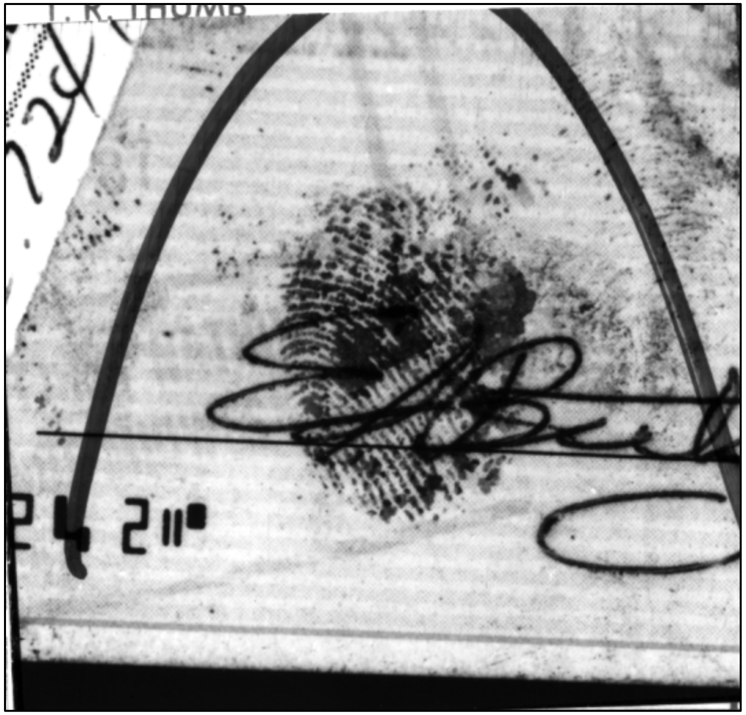
Michigan State University
http://biometrics.cse.msu.edu

IBPC 2014
April 3, 2014

Research supported by NSF CITeR

Fundamental Premise for Fingerprint Recognition

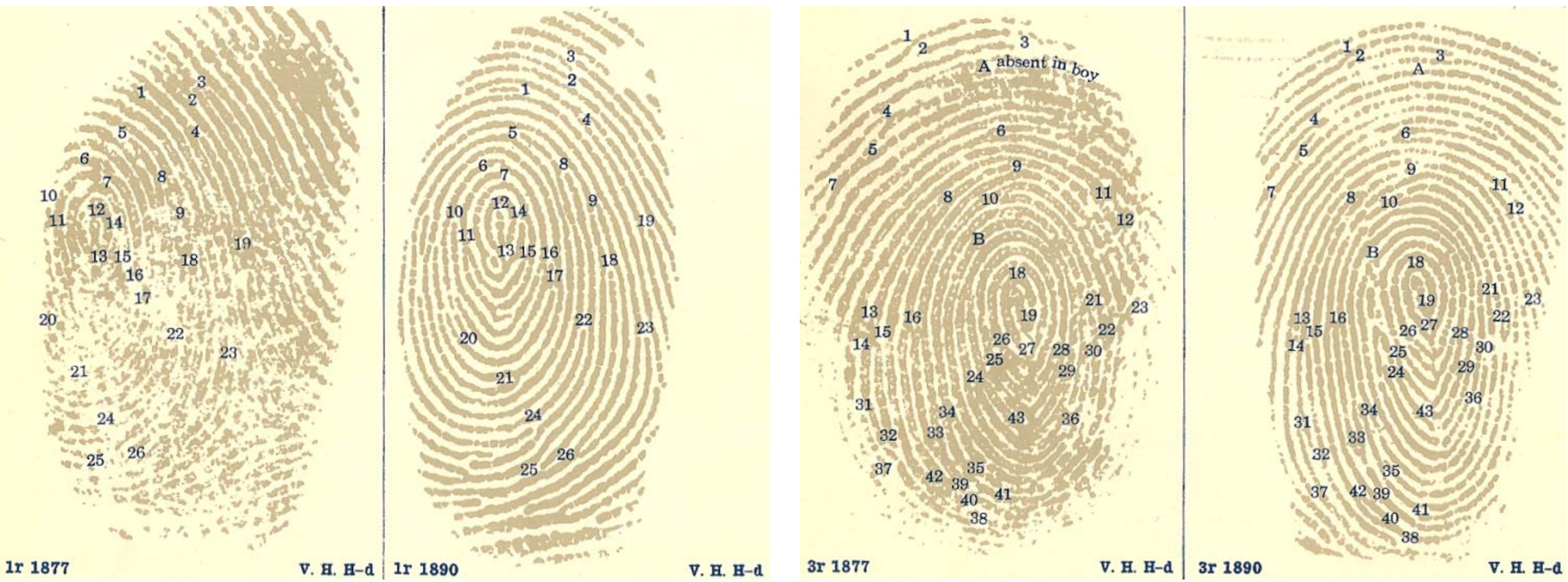
Do these two impressions come from the same finger?



- **Uniqueness**: Ridge patterns on different fingers are distinctive
- **Persistence**: Friction ridge patterns do not change over time

Persistence of Fingerprints

- Traditional perspective: Persistence of fingerprint ridge structure
- Galton compared 11 pairs of fingerprints from six different individuals; only 1 out of 389 minutiae was found to be missing



F. Galton, Finger Prints, Macmillan, 1892

Uniqueness and Persistence

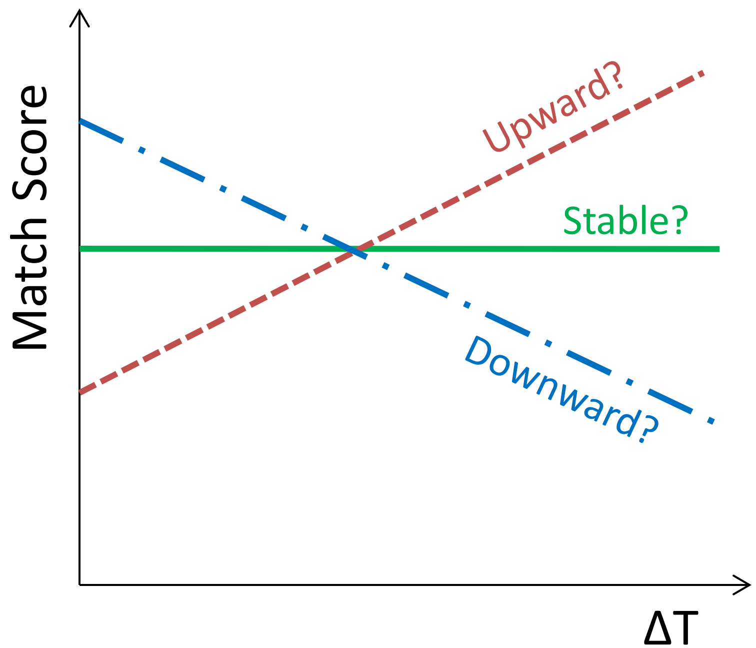
*"**Uniqueness and persistence** are necessary conditions for friction ridge identification to be feasible, but those conditions **do not imply that anyone can reliably discern whether or not two friction ridge impressions were made by the same person**."*

National Research Council, "Strengthening Forensic Science in the United States: A Path Forward", 2009

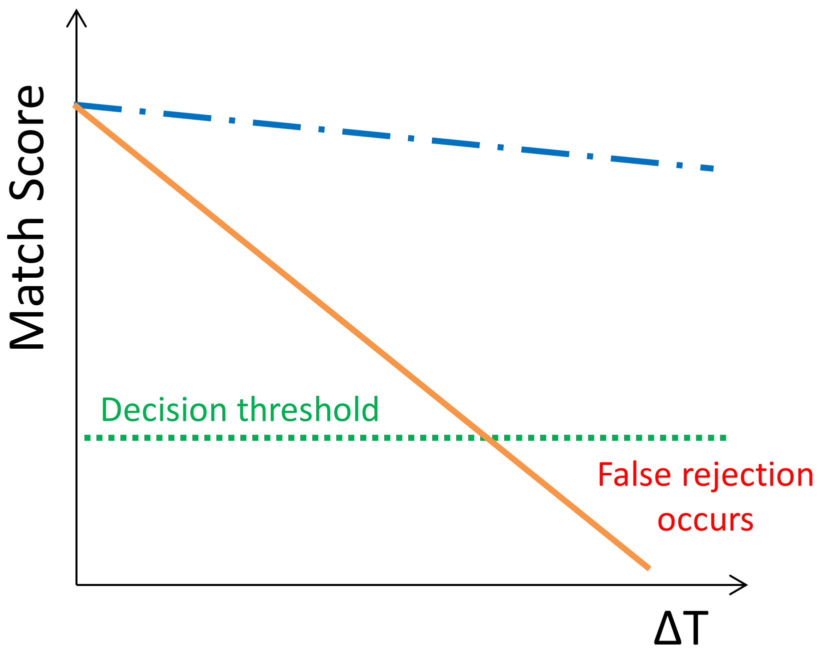
Problem Definition

Determine the persistence of fingerprints w.r.t. AFIS accuracy

Trend of genuine match scores



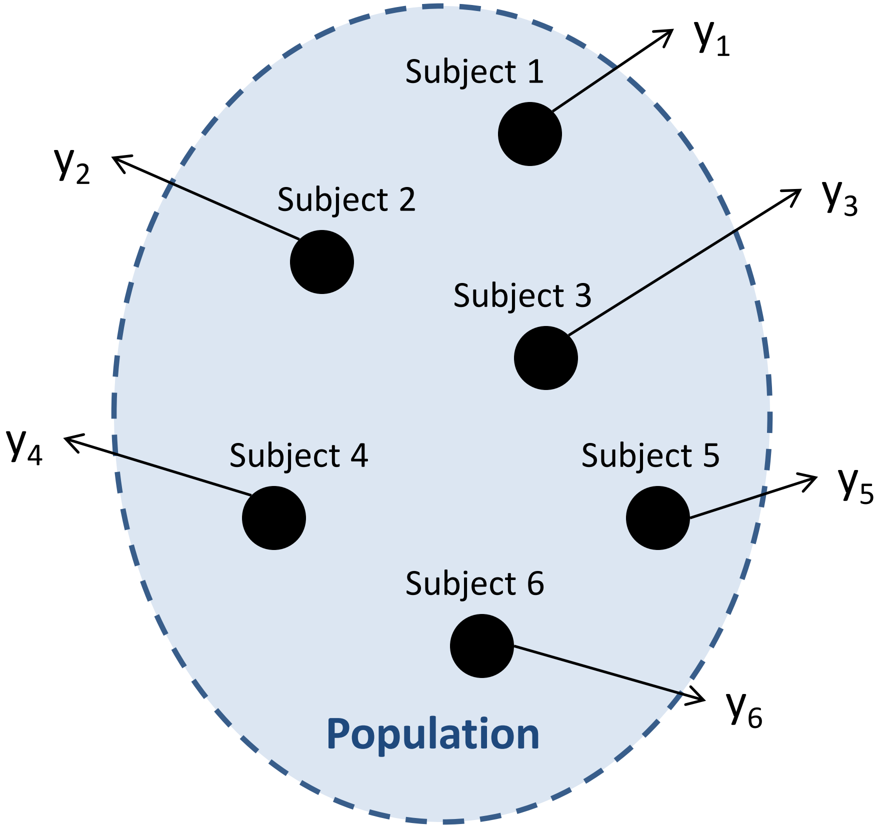
Trend of matching accuracy



Data Type: Longitudinal vs. Cross-Sectional

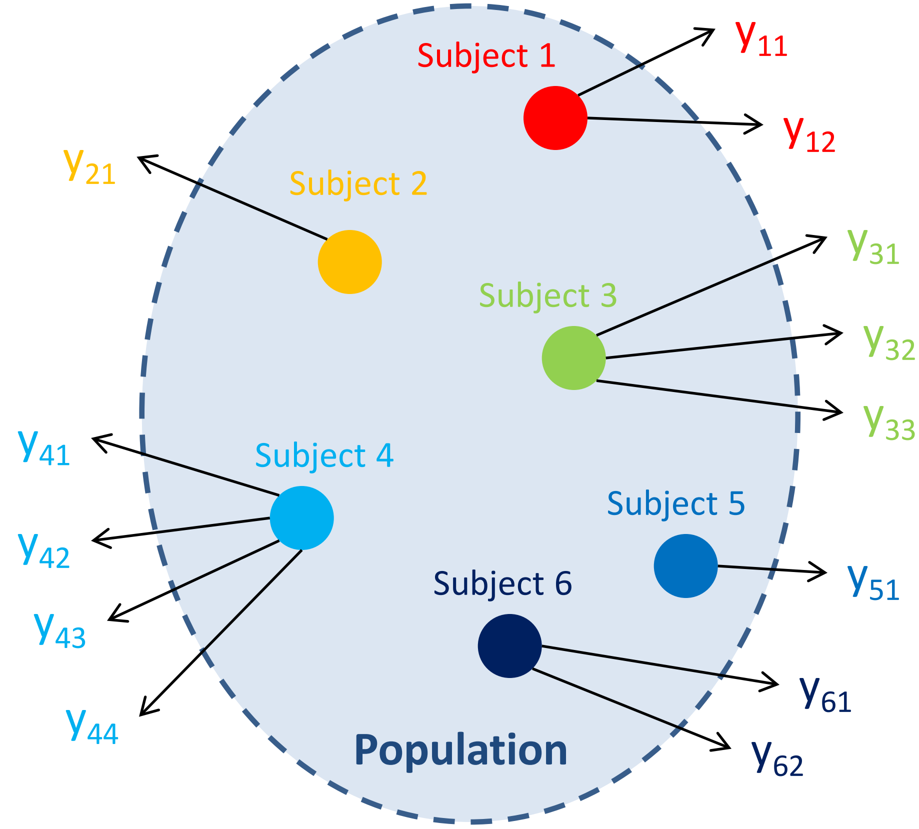
Cross-sectional data

A single measurement is made on each individual sampled from a population



Longitudinal data

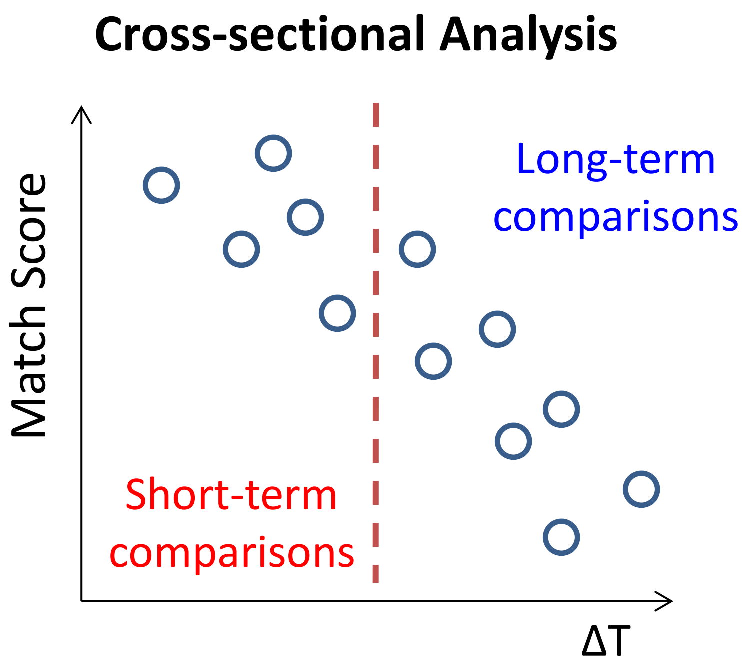
Repeated measurements on a collection of individuals sampled from a population



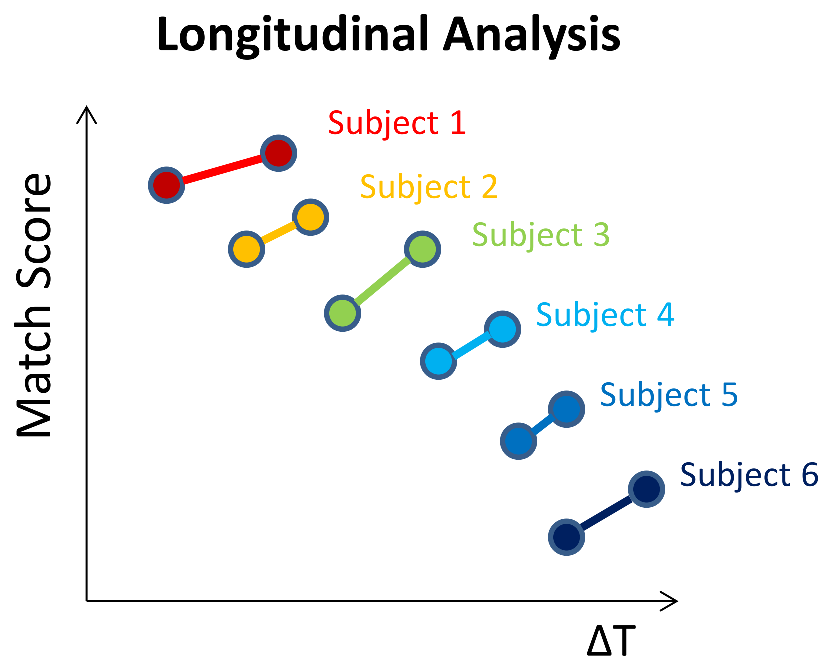
Longitudinal data are called

- Balanced data : Every subject has the same number of measurements
- Time-structured data: Repeated measurements follow an identical time schedule across individuals

Longitudinal vs. Cross-Sectional Analysis



Match scores ***decrease*** w.r.t. ΔT



Match scores ***increase*** w.r.t. ΔT

- Longitudinal fingerprint data do not satisfy the properties of balanced & time structured required for cross-sectional analysis

P. J. Diggle, K-Y. Liang, and S. L. Zeger, Analysis of Longitudinal Data, Oxford Science Publications, 1994

Longitudinal Fingerprint Database

- Repeat offenders booked by the Michigan State Police
- 15,597 subjects with at least 5 tenprint cards, minimum time span of 5-years (max. time span is 12 years) and demographics (race, gender, age)
- All genuine pairwise comparisons by two COTS matchers
- Currently, only right index finger is used in the analysis

June 2001

July 2002

April 2003

Sept. 2007

March 2008

Oct. 2008

Approach

- Fit and evaluate a multilevel statistical model with time gap as covariate to genuine match scores
  - Null hypothesis: Slope of linear model is 0
- Compare time gap with other possible covariates (i.e., subject's age, fingerprint quality, race, and gender)
- Fit a multilevel model with time gap as covariate to binary match decisions

Multilevel Statistical Model

- Longitudinal data can be viewed as hierarchical data
  - *j*-th measurement (match score) for subject *i*
- A model in its simplest form

j-th measurement for subject i

Covariate (or predictor, explanatory variable)

Level-1 Model
(Within-person change)

$$y_{ij} = \varphi_{0i} + \varphi_{1i} x_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$$

Level-2 Model
(Between-person change)

$$\varphi_{0i} = \boxed{\beta_{00}} + \boxed{b_{0i}}$$
$$\varphi_{1i} = \boxed{\beta_{10}} + \boxed{b_{1i}}$$

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

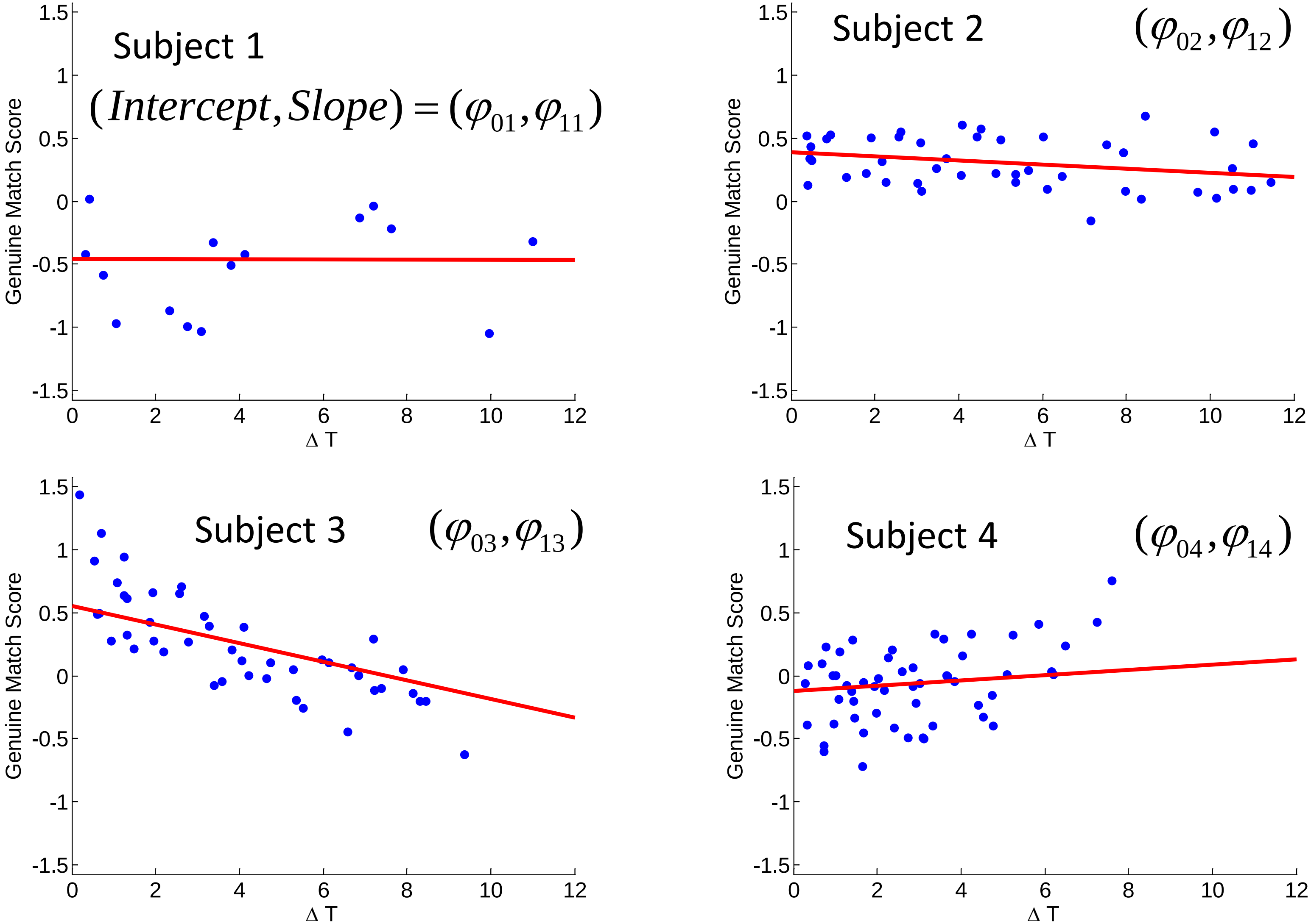
Fixed effects Random effects

Composite Model

$$y_{ij} = (\beta_{00} + b_{0i}) + (\beta_{10} + b_{1i}) x_{ij} + \varepsilon_{ij}$$

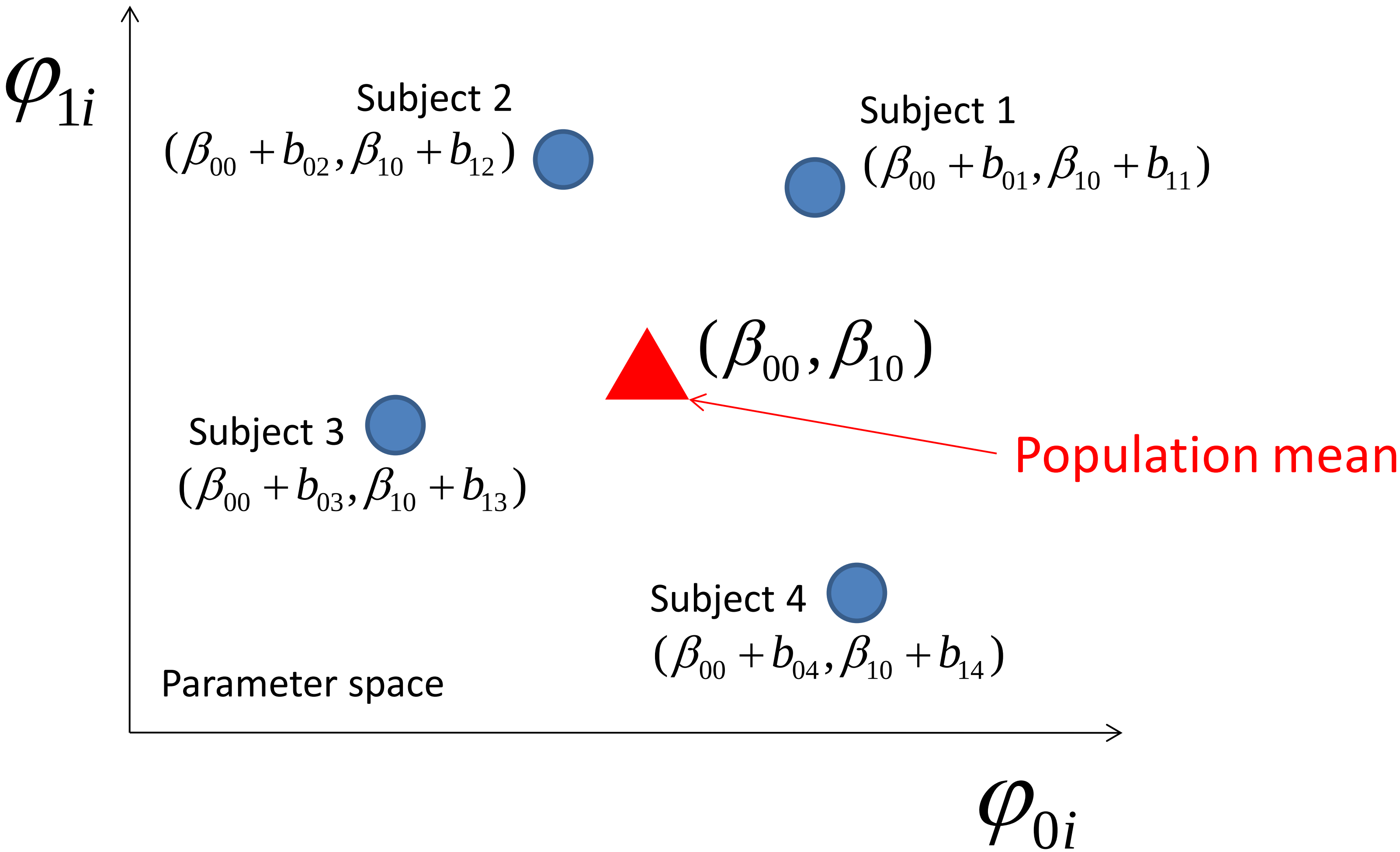
Level-1 Model

$$y_{ij} = \varphi_{0i} + \varphi_{1i} x_{ij} + \varepsilon_{ij}$$



Level-2 Model

$$\varphi_{0i} = \beta_{00} + b_{0i}$$
$$\varphi_{1i} = \beta_{10} + b_{1i}$$



Part I. Genuine Match Score Modeling

Level-1 | **Level-2**

Model A (Unconditional mean model)

$$y_{ij} = \varphi_{0i} + \varepsilon_{ij}$$

$$\varphi_{0i} = \beta_{00} + b_{0i}$$

Model B

$$y_{ij} = \varphi_{0i} + \varphi_{1i} x_{ij} + \varepsilon_{ij}$$

$x_{ij} = \Delta T_{ij}$ B_T: Time interval

$x_{ij} = AGE_{ij}$ B_A: Subject's age

$x_{ij} = Q_{ij}$ B_Q: Max. of NFIQ of fingerprints in comparison

$$\varphi_{0i} = \beta_{00} + b_{0i}$$

$$\varphi_{1i} = \beta_{10} + b_{1i}$$

Model C

$$y_{ij} = \varphi_{0i} + \varphi_{1i} \Delta T_{ij} + \varepsilon_{ij}$$

$$\varphi_{0i} = \beta_{00} + \beta_{01} C_i + b_{0i}$$

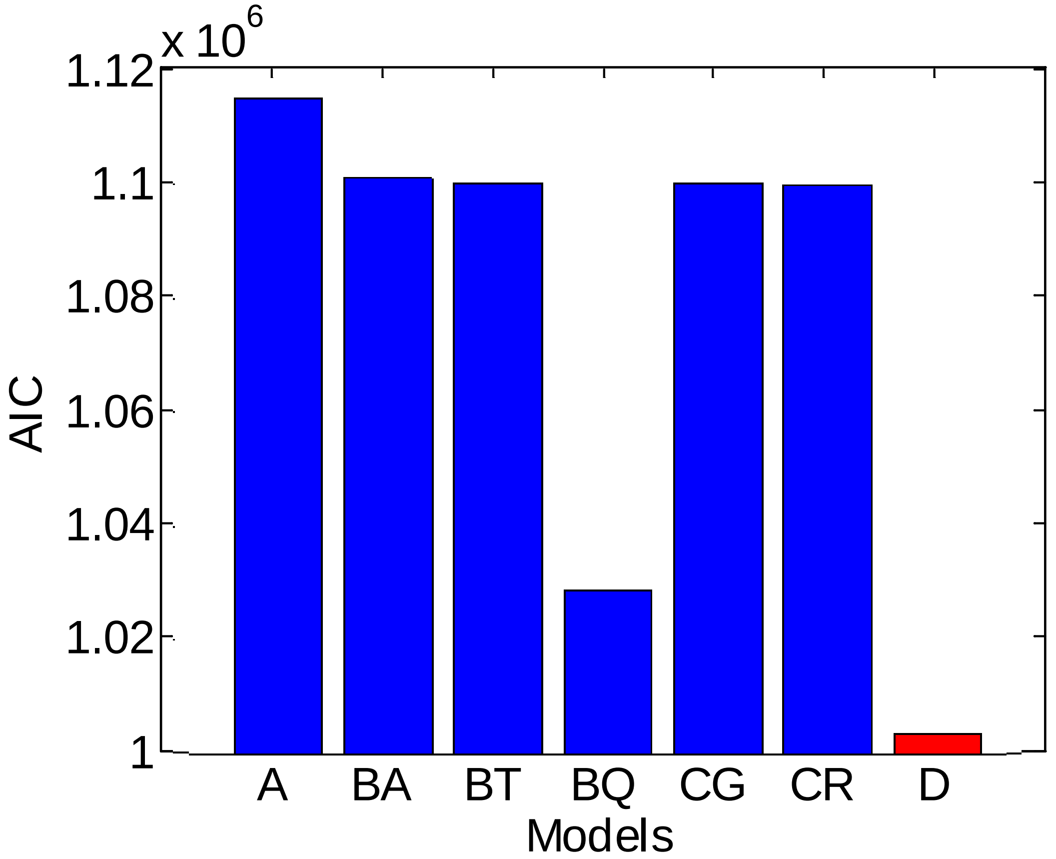
$$\varphi_{1i} = \beta_{10} + \beta_{11} C_i + b_{1i}$$

$C_i = bMale_i$ C_G: Gender

$C_i = bWhite_i$ C_R: Race

Model Comparisons

- Goodness-of-Fit
  - Smaller the value, better the model fit



- AIC (Akaike Information Criterion)
- Decrease in AIC observed for Models BT, BA, BQ vs. Model A
- ΔT, AGE & Q explain the variance in genuine match scores
- Q is the best covariate
- AIC barely decreases for Model BT vs. Models CG, CR
- Gender and race are not important covariates
- Model D with ΔT, AGE, and Q explains variance the best

Validation of Model Assumptions

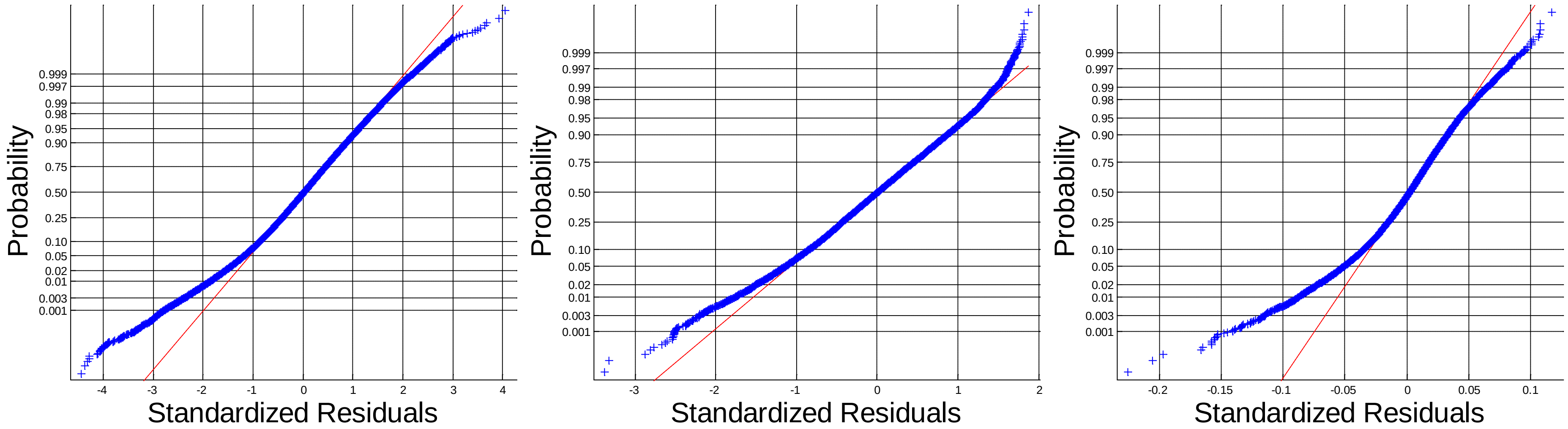
- Normal probability plots
  - If linear, the distribution is normal

Level-1

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Level-2

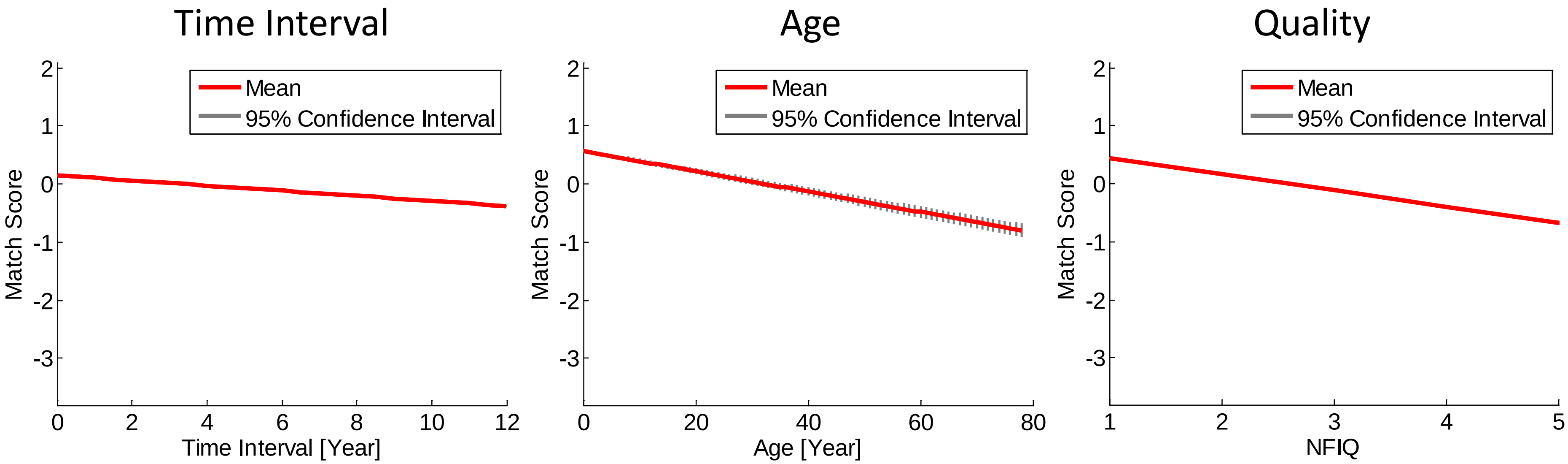
$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$



- Departures from normality are observed at tails

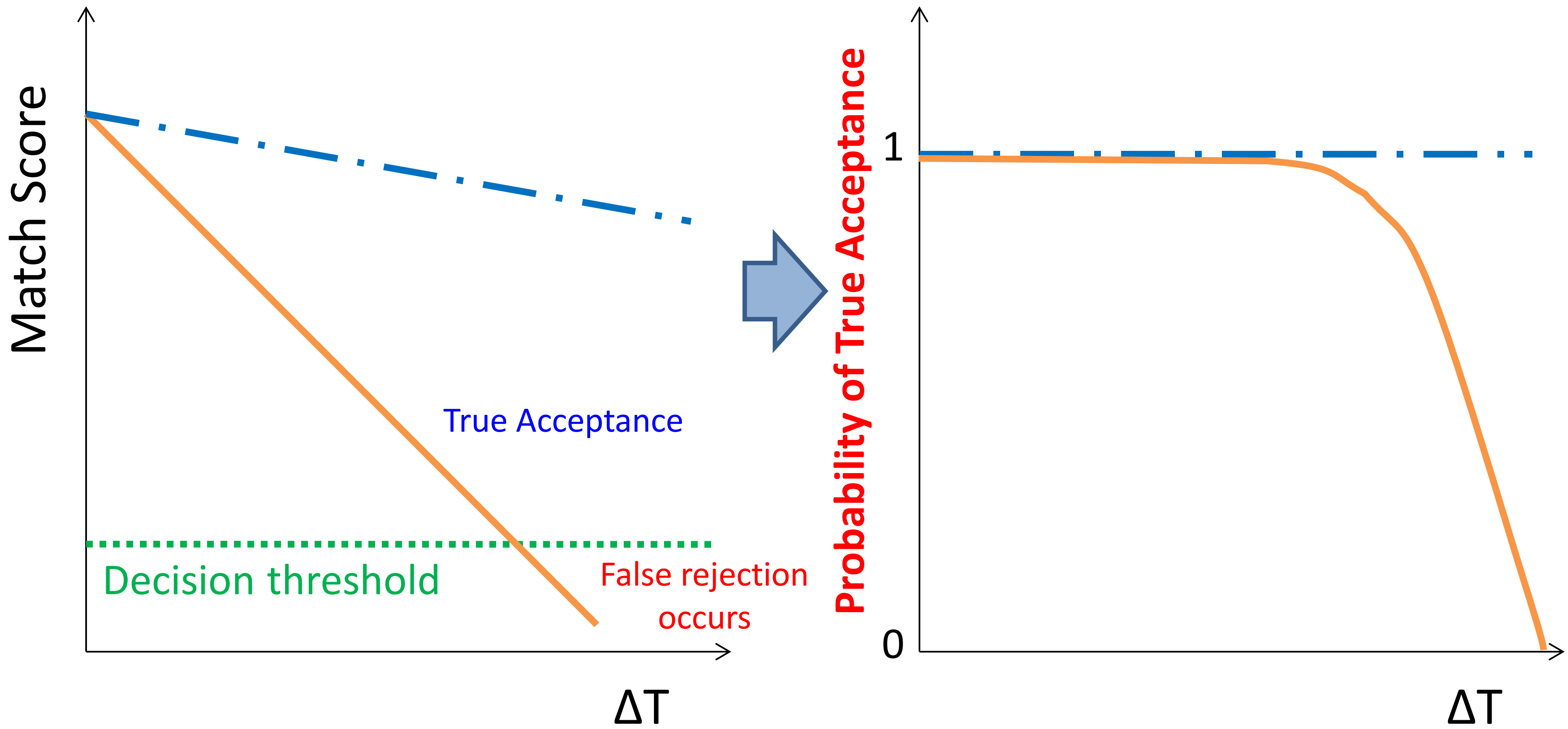
Parameter Estimates and Hypothesis Tests

- Bootstrap to obtain parameter estimates and confidence interval
  - Resample N (= 15,597) subjects with replacement; 1,000 bootstrap samples
- **H_0: $\beta_{10} = 0$ (slope of linear model is 0)**
  - H_0 is rejected at 0.05 level for Model B_T, B_A, and B_Q



- Genuine match scores decrease w.r.t. time interval, subject's age, and NFIQ

Part II. Matching Accuracy Modeling



Multilevel Model for Binary Responses

(Generalized Linear Mixed-effects Model)

Level-1

$$y_{ij}^{*} = \begin{cases} 1, & y_{ij} > Th \\ 0, & \text{otherwise} \end{cases}$$

$$\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^{2})$$

$$y^{*}_{ij} \sim Bin(1, \pi_{ij})$$

$$g(\pi_{ij}) = \varphi_{0i} + \varphi_{1i} x_{ij} + \varepsilon_{ij}$$

$g(\cdot)$ is a link function;
for binary responses,
g(·) is a logit function

Level-2

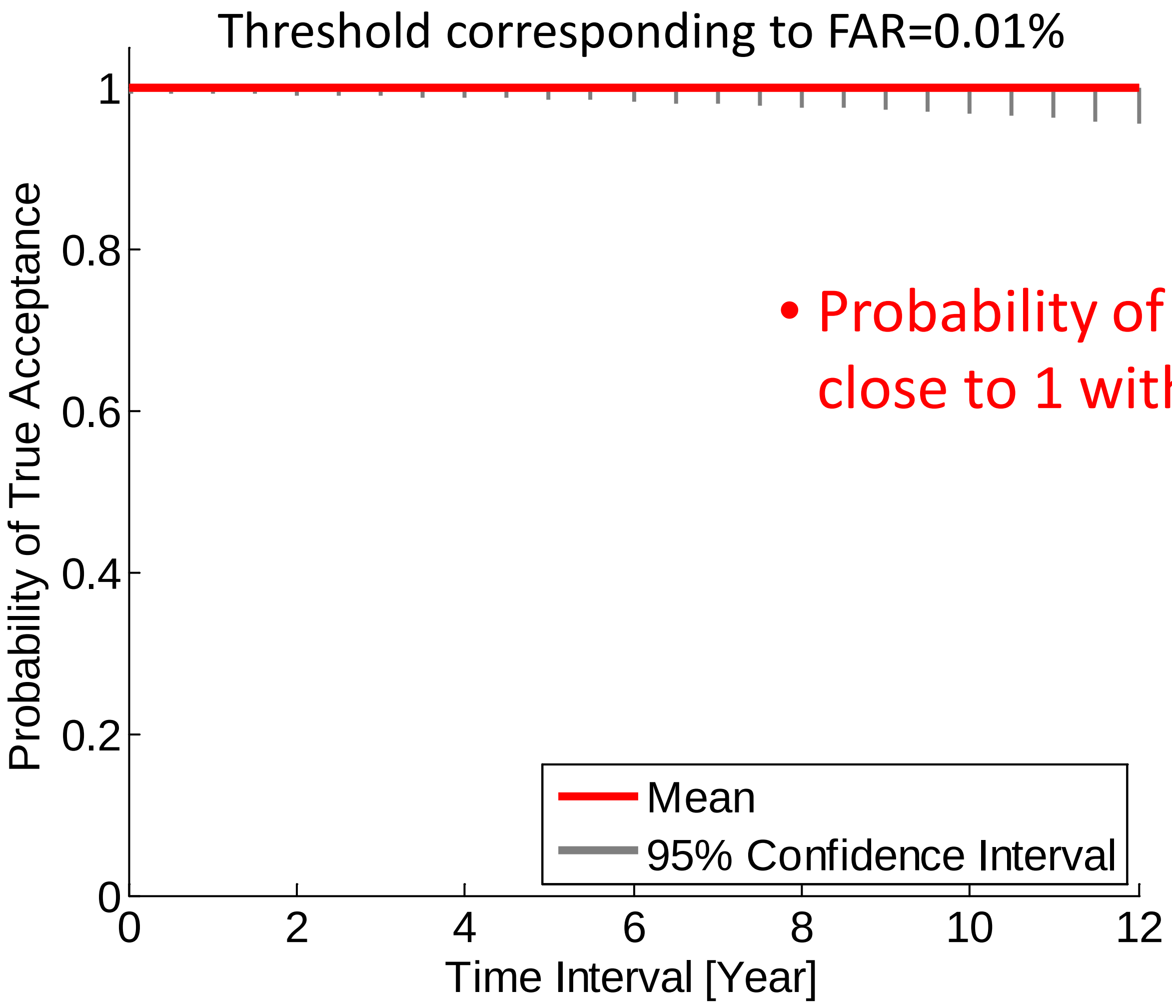
$$\varphi_{0i} = \beta_{00} + b_{0i}$$

$$\varphi_{1i} = \beta_{10} + b_{1i}$$

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

Matching Accuracy over Time

- 400 bootstrap samples



- Probability of true acceptance remains close to 1 within 12-year time interval

Summary and Conclusions

- Statistical analysis with multilevel models for longitudinal fingerprint data (15,597 subjects with 12-year time span)
- Based on the results of hypothesis test and bootstrap confidence interval, we can make following inferences
  - **Genuine match score tends to decrease over time**
  - **Matching accuracy tends to remain stable over time with high confidence**
- Future work
  - Analyze longitudinal data with longer time span
  - Explore nonlinear models and interaction terms

Thank you.