Longitudinal study of fingerprint recognition

Soweon Yoon[a,b] and Anil K. Jain[a,1]

[a]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824; and [b]Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899

Edited by Bin Yu, University of California, Berkeley, CA, and approved May 14, 2015 (received for review June 4, 2014)

Human identification by fingerprints is based on the fundamental premise that ridge patterns from distinct fingers are different (uniqueness) and a fingerprint pattern does not change over time (persistence). Although the uniqueness of fingerprints has been investigated by developing statistical models to estimate the probability of error in comparing two random samples of fingerprints, the persistence of fingerprints has remained a general belief based on only a few case studies. In this study, fingerprint match (similarity) scores are analyzed by multilevel statistical models with covariates such as time interval between two fingerprints in comparison, subject's age, and fingerprint image quality. Longitudinal fingerprint records of 15,597 subjects are sampled from an operational fingerprint database such that each individual has at least five 10-print records over a minimum time span of 5 y. In regard to the persistence of fingerprints, the longitudinal analysis on a single (right index) finger demonstrates that (*i*) genuine match scores tend to significantly decrease when time interval between two fingerprints in comparison increases, whereas the change in impostor match scores is negligible; and (*ii*) fingerprint recognition accuracy at operational settings, nevertheless, tends to be stable as the time interval increases up to 12 y, the maximum time span in the dataset. However, the uncertainty of temporal stability of fingerprint recognition accuracy becomes substantially large if either of the two fingerprints being compared is of poor quality. The conclusions drawn from 10-finger fusion analysis coincide with the conclusions from single-finger analysis.

biometrics | fingerprint recognition | persistence of fingerprints | longitudinal data analysis | multilevel statistical model

Friction ridge skin on fingers and palms has been purportedly known to be a physical characteristic of an individual that does not change over time (i.e., persistence or permanence of friction ridge pattern) and can be used as a person's "seal" or "signature" (i.e., uniqueness or individuality of ridge pattern). Starting with the first known case where the latent fingerprints found at a crime scene in Argentina in 1893 were officially accepted as evidence to convict a suspect (1), friction ridge analysis has become one of the most crucial methods in crime scene investigations worldwide. The decision made in *Frye v. United States* in 1923 (2) was widely cited as the basis for the admissibility of forensic evidence, including friction ridge pattern; *Frye* standard states that a scientific principle or discovery that has gained a general acceptance in the relevant field is admissible in the courts.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, in 1993 (3), however, the general acceptance test of *Frye* was superseded by the Federal Rules of Evidence. The *Daubert* ruling established a guideline for admitting forensic evidence, which consists of the following factors: (*i*) empirical testing, (*ii*) peer review and publication, (*iii*) known or potential error rate, (*iv*) standards controlling the operation, and (*v*) the *Frye* standard of general acceptance. The *Daubert* standard provoked challenges to admissibility of friction ridge evidence in the courts. Although all of about 40 such challenges resulted in a decision that friction ridge analysis is acceptable as forensic evidence, the *Daubert* case highlighted a lack of scientific basis of persistence and uniqueness of fingerprints and standards that can be universally referred to in friction ridge analysis.

Along with the development of standards and guidelines for friction ridge analysis (4) and retraining of latent examiners (5) as a result of the *Daubert* ruling, a body of research to demonstrate uniqueness and persistence of friction ridge patterns has emerged. Although the uniqueness of fingerprints has been studied by (*i*) estimating the probability of a random correspondence (i.e., two different fingerprints selected at random will be sufficiently similar to be claimed as a mate) (6–8) or (*ii*) measuring the evidential value[§] of latent fingerprint comparisons (9) (terminologies indicated with § are defined in *SI Appendix*, section S2), the persistence of fingerprints has been generally accepted based on anecdotal evidence, including case studies conducted by Herschel (10) and Galton (11) (*SI Appendix*, section S4), and the anatomical structure of friction ridge skin—the ridge pattern formed in the inner (dermal) layer during gestation remains unchanged with the protection of the outer (epidermal) layer (12).

The persistence of fingerprints typically refers to the invariance of friction ridge pattern itself. However, the pertinent question of interest is whether the fingerprint recognition methodology (*SI Appendix*, section S3) maintains high recognition accuracy as the time interval between two fingerprints being compared increases. The 2009 National Research Council report *Strengthening Forensic Science in the United States: A Path Forward* (13) pointed out, "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." Fingerprint recognition exhibits two types of comparison errors: (*i*) false rejection: two impressions of the same finger (a genuine fingerprint pair) are declared as a nonmatch due to large "intrafinger" variability, and (*ii*) false acceptance: impressions from two distinct fingers (an impostor fingerprint pair) are declared as a match due to large "interfinger" similarity. The

Significance

Fingerprint recognition, which is considered to be a reliable means for human identification, has been used in many applications ranging from law enforcement and forensics to unlocking mobile phones. Despite its successful deployment, the fundamental premise of fingerprint-based identification—persistence and uniqueness of fingerprints—has not yet been well studied, resulting in challenges to the admissibility of friction ridge evidence in courts of law. This study investigates the tendency of fingerprint similarity scores and recognition accuracy with respect to covariates characterizing properties of fingerprint impressions and demographics of subjects, including time interval between two fingerprints being compared in regard to the persistence of fingerprints. A multilevel statistical analysis is conducted with a longitudinal dataset of fingerprint records from 15,597 subjects.

Author contributions: S.Y. and A.K.J. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The dataset used in this paper can be made available to interested researchers for longitudinal study under appropriate nondisclosure agreement dictated by the Michigan State Police.

[1]To whom correspondence should be addressed. Email: jain@cse.msu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410272112/-/DCSupplemental.

intrafinger variability is observed due to changes in intrinsic skin condition (e.g., finger skin dryness, cuts, and abrasions) and extrinsic acquisition process (e.g., finger pressure and placement), and sensing technology (e.g., interoperability among various fingerprint sensors). The interfinger similarity is observed when the ridge patterns from two distinct fingers coincide partially.

The biometric recognition literature has reported a phenomenon called template aging, which refers to an increase in the error rate in biometric recognition with respect to the time gap between a query and a template (or reference) (14). A study comparing groups of fingerprint pairs with respect to time gap reported that the fingerprint comparisons with less than a 5-y time gap show lower error rate than comparisons with a larger time gap (15). However, cross-sectional analysis used in ref. 15 is valid only if the longitudinal data[§] are balanced[§] and time structured[§]; this condition is not typically satisfied in most biometrics data, including the dataset used in ref. 15. Longitudinal studies on fingerprint, face, and iris biometrics published in the literature are summarized in *SI Appendix*, section S4.

Study Objectives and Caveats

A longitudinal dataset of fingerprints from 15,597 subjects apprehended by Michigan State Police (MSP) is analyzed, which consists of five or more fingerprint records over a time span varying from 5 to 12 y for each subject. Multilevel statistical models with covariates characterizing properties of fingerprint impressions and demographics of subjects are designed to analyze the longitudinal dataset, which is unbalanced and time unstructured. Specifically, our study aims to address the following issues:

- Trend of fingerprint match scores of genuine and impostor pairs with respect to the following covariates: time interval between fingerprints in comparison, subject's demographic factors (age, sex, and race), and fingerprint image quality;
- Assessment and comparison of the multilevel models with various combinations of the covariates;
- Correlations and interactions among covariates;
- Temporal trend of fingerprint recognition accuracy in terms of probabilities of true acceptance and false acceptance;
- Trend of fingerprint match scores and recognition accuracy when all 10 of a subject's fingers are used for recognition, a prevailing practice in law enforcement.

The results and conclusions made in this paper should be interpreted with the following caveats:

i) Any variability appearing in fingerprint images collected over time may be induced by either physical changes in friction ridge structure or changes in imaging condition and subject's behavior during fingerprint acquisition. No distinction between the two sources of variability is made in the analysis presented in this paper.

ii) The inferences and conclusions presented in this paper are drawn from 10-print analysis and do not suggest the same conclusions for latent fingerprint (or finger mark) analysis.

The detailed explanations about the caveats can be found in *SI Appendix*, section S1.

Longitudinal Fingerprint Dataset

A longitudinal dataset of fingerprints was collected from the records of repeat offenders apprehended by the MSP. *SI Appendix*, Fig. S3 shows an example of six fingerprint impressions of the right index finger of a subject in the dataset acquired between June 2001 and October 2008. A total of 15,597 subjects were randomly extracted from the MSP fingerprint archive, such that each subject has at least five acquisitions from all 10 fingers on a formatted fingerprint card (called 10-print card) over a minimum of 5-y time span. The 10-print cards of a subject are ordered according to the time sequence; a set of 10-print cards of subject i ($i = 1,2, \ldots, N$; N is the total number of subjects in the dataset) is labeled as follows: $\mathcal{F}_i = \{F_{i,1}, F_{i,2}, \ldots, F_{i,n_i}\}$, such that $T_{i,1} < T_{i,2} < \cdots < T_{i,n_i}$, where $F_{i,j}$ is the jth 10-print card of subject i, $T_{i,j}$ is the time stamp of $F_{i,j}$, and n_i denotes the number of 10-print cards belonging to subject i.

A summary of the dataset is below, and the data statistics are shown in *SI Appendix*, Fig. S4.

- Each of the 15,597 subjects has at least five 10-print cards, providing 122,685 10-print cards in total. The average number of 10-prints per subject in the dataset is 8, and the maximum is 26 cards.
- The 10-print impressions of a subject have a minimum of 5-y time span (the time difference between the first and the last fingerprint acquisitions of a subject); that is, $\triangle T_{i,1n_i} \geq 5$ y for $i = 1,2, \ldots, N$, where $\triangle T_{i,1n_i} = T_{i,n_i} - T_{i,1}$. The average time span is 9 y, and the maximum time span in the dataset is 12 y.
- Any two consecutive 10-print impressions of a subject are obtained with at least a 2-mo time gap; $T_{i,j+1} - T_{i,j} \geq 2$ mo for $j = 1,2, \ldots, n_i - 1$ for subject i.
- Along with 10-print images, the following demographic information is also available for each subject: sex (male or female), race (white/Hispanic, black, American Indian/Eskimo, or Asian/Pacific Islander), and age at the time of 10-print acquisition (the youngest subject's age at the time of the first impression is 8 y; the oldest subject's age at the time of the last impression is 78 y).

Two commercial off-the-shelf (COTS) fingerprint matchers (denoted as COTS-1 and COTS-2) are used to compute match scores. For subject i with n_i fingerprint impressions, we conduct all pairwise comparisons; that is, $\binom{n_i}{2}$ genuine match scores are generated from each matcher. This is because law enforcement agencies often store all of the 10-print records for every subject and compare a query fingerprint to all records in the database. Note that the pairwise comparisons of the fingerprint records of a subject result in correlations among the genuine match scores of the subject. On the other hand, the impostor scores are obtained by comparing subject i's n_i fingerprint impressions to the first impressions from all other subjects. For each finger position, 481,181 genuine match scores and 1,913,395,260 impostor match scores are obtained by each of the COTS matchers.

Multilevel Statistical Model

For balanced and time-structured longitudinal dataset, cross-sectional analysis can be readily applied by grouping the data according to cohort (for example, short-term and long-term fingerprint comparison groups) under the assumption of compound symmetry.[§] In reality, however, it is not feasible to collect longitudinal fingerprint data by following an identical measurement schedule over a large number of subjects in the sample satisfying the compound symmetry. Multilevel statistical model (16, 17) is one of several statistical models that can handle the unbalanced and/or time-unstructured longitudinal data.

As a fingerprint comparison essentially involves two fingerprint impressions to generate a single match score, a simple linear two-level model with a single covariate for continuous match scores can be represented by the following:

Level 1 model (intrasubject variability):

$$y_{ijk} = \varphi_{0i} + \varphi_{1i} x_{ijk} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \qquad [1]$$

Level 2 model (intersubject variability):

$$\begin{aligned} \varphi_{0i} &= \beta_{00} + b_{0i}, \\ \varphi_{1i} &= \beta_{10} + b_{1i}, \end{aligned} \quad \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right). \qquad [2]$$

The level 1 model in Eq. **1** is regressed to the repeated measurements taken from each subject, and accounts for the

intrasubject variability. In case of genuine fingerprint comparisons, the variables and parameters in the level 1 model are defined as follows: y_{ijk} is the subject i's observed response of match score when the jth and kth fingerprints are compared, x_{ijk} is the explanatory variable (or covariate), φ_{0i} and φ_{1i} are the true parameters representing the intercept and slope of the linear model for subject i, and ε_{ijk} is the error in the observed response y_{ijk} from the model fit. The error is assumed to be normally distributed with a zero mean and a variance of σ_ε^2.

In the level 2 model in Eq. **2**, where the population-mean tendency and deviations of subjects from the mean trend are modeled to account for the intersubject variability, the true parameters for subject i (φ_{0i} and φ_{1i}) can be modeled by a mixture of fixed and random effects: fixed-effects parameters β_{00} and β_{10} represent the grand means of intercept and slope across all N subjects in the data, and random-effects parameters b_{0i} and b_{1i} represent the deviations of subject i's intercept and slope from β_{00} and β_{10}. The random effects are assumed to follow a normal distribution.

To determine whether two fingerprint impressions are from the same finger, a binary decision for a fingerprint pair is made by applying a predetermined decision threshold to the match score. If the match score of a fingerprint pair is greater than the threshold, the two fingerprints are determined to be a genuine match; otherwise, they are determined to be an impostor match. If a fingerprint pair determined to be a genuine match is indeed from the same finger, the binary decision is a true acceptance. If a genuine-match decision is made on two fingerprints that are in fact from different fingers, the decision is a false acceptance. In multilevel modeling, a binary response is viewed as a Bernoulli trial with the probability of true (or false) acceptance π_{ijk}, and the expected π_{ijk} is modeled after being transformed by a logit link function.

$$\begin{aligned} &\text{Level 1 model: } g(\pi_{ijk}) = \varphi_{0i} + \varphi_{1i} x_{ijk} + \varepsilon_{ijk}, \\ &\qquad\qquad\quad y^*_{ijk} \sim Bin(1, \pi_{ijk}), \\ &\text{Level 2 model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}, \end{aligned} \quad [3]$$

where $g(\cdot)$ is a logit link function: $g(\pi_{ijk}) = \log(\pi_{ijk}/(1-\pi_{ijk}))$.

The maximum likelihood (ML) and generalized least-squares (GLS) estimations are widely used to estimate parameters in a multilevel model (17). Under the assumption that the residuals are normally distributed, the ML estimates of the parameters are typically obtained by iterative GLS (16).

With multilevel modeling for longitudinal data analysis, we investigate the following observed responses:

- Case I: A single finger (right index finger of the subjects that is typically chosen as the primary finger in fingerprint recognition with one finger) is used for recognition.
  - Normalized genuine match score obtained by the following:

$$y_{i,jk} = \frac{s_{i,jk} - \mu}{\sigma}, \quad [4]$$

    where $s_{i,jk}$ is the genuine match score between the jth and kth fingerprint impressions of the right index finger of subject i, and μ and σ are the mean and SD of $\{s_{i,jk}\}$, respectively.
  - Normalized impostor match score ($y_{ij,k}$) between the kth fingerprint impression of the right index finger of subject i and the right index fingerprint in the first 10-print card of subject j for $j \neq i$ and $k = 1,2, \ldots, n_i$.
  - Binary identification decision made on a genuine pair with match score of $s_{i,jk}$ by applying a decision threshold (Th):

$$y^*_{i,jk} = \begin{cases} 1, & \text{if } s_{i,jk} > Th \\ 0, & \text{otherwise} \end{cases}. \quad [5]$$

  - Binary identification decision ($y^*_{ij,k}$) made on an impostor pair with match score of $s_{ij,k}$ by applying the decision threshold Th.
- Case II: All 10 fingers are used for recognition.
  Similar to Case I, normalized genuine and impostor fusion scores ($Y_{i,jk}$ and $Y_{ij,k}$) and binary identification decisions made on genuine and impostor pairs of subjects ($Y^*_{i,jk}$ and $Y^*_{ij,k}$) are obtained based on the match score fused by a sum rule (*SI Appendix*, section S5).

The following covariates are investigated in the multilevel models for genuine fingerprint pairs:

- $\triangle T_{i,jk}$: Time interval between the jth and kth fingerprint impressions of subject i; $\triangle T_{i,jk} = T_{i,k} - T_{i,j}$ for $j < k$;
- $AGE_{i,jk}$: Age of subject i when the latter of the jth and kth 10-print impressions was made; the subject i's age at $T_{i,k}$ for $j < k$;
- $Q_{i,jk}$: The value corresponding to the lower of the qualities of the jth and kth fingerprint impressions of subject i. In this study, the National Institute of Standards and Technology Fingerprint Image Quality (NFIQ) measure (18) is used, which assigns one of the five discrete values ranging from 1 (the highest quality) to 5 (the lowest quality), to define fingerprint image quality. According to the definition of NFIQ, $Q_{i,jk} = \max(Q_{i,j}, Q_{i,k})$, where $Q_{i,j}$ is the NFIQ value of the jth fingerprint impression of subject i;
- bM_i: A binary indicator of sex of subject i; 1 for male, and 0 for female;
- bW_i: A binary indicator of race of subject i; 1 for whites, and otherwise 0.

Multilevel models for impostor pairs are designed with the following covariates:

- $\triangle T_{i,1k}$: Time elapsed since the first 10-print of subject i was obtained; $\triangle T_{i,1k} = T_{i,k} - T_{i,1}$, $k = 1,2, \ldots, n_i$;
- $AGE_{i,k}$ and $AGE_{j,1}$: Age of subject i when the kth 10-print impression was made, and the age of impostor subject j at the time of the first 10-print acquisition, respectively.

The two-level linear models investigated in our study are listed in Table 1 and *SI Appendix*, section S6.

Results

Population-Mean Trend of Genuine Fingerprint Match Scores. Given that the normality assumptions of the residuals and random effects in the multilevel model fit to the data are violated (*SI Appendix*, section S7), the parameters in the multilevel models are estimated by a fully nonparametric bootstrap (16). We generate 1,000 bootstrap samples for genuine match score analysis, where each bootstrap sample is obtained by a cluster bootstrap—N subjects are resampled with replacement at level 1, and all of the level 2 data belonging to those subjects are included in the sample—to preserve the hierarchy in the longitudinal data.

SI Appendix, Tables S2 and S3 report the mean of the parameter estimates of the bootstrap samples and the percentile confidence intervals for the genuine match score models. The population-mean trends of models B_T, B_A, and B_Q based on the fixed-effects parameter estimates (β_{00} and β_{10}) show that the genuine match scores tend to decrease when $\triangle T_{i,jk}$, $AGE_{i,jk}$, and $Q_{i,jk}$ increase (Fig. 1 and *SI Appendix*, Fig. S6). The null hypothesis—$\beta_{10} = 0$ in models B_T, B_A, and B_Q (i.e., the slope of the linear model is zero)—is rejected for all three models at a significance level of 0.05 because the 95% confidence interval for β_{10} does not contain zero.

Models D and E incorporate all three covariates ($\triangle T_{i,jk}$, $AGE_{i,jk}$, and $Q_{i,jk}$) into the model; model E includes interaction terms ($\triangle T_{i,jk} Q_{i,jk}$ and $AGE_{i,jk} Q_{i,jk}$), whereas model D does not. The covariance matrix in model D shows that the correlations (i) between $\triangle T_{i,jk}$ and $Q_{i,jk}$ and (ii) between $AGE_{i,jk}$ and $Q_{i,jk}$ are very small. Also, the population-mean trends of models D and E and their 95% confidence intervals, illustrated in *SI Appendix*, Fig. S7, indicate that the impact of the interactions (i) between $\triangle T_{i,jk}$ and $Q_{i,jk}$ and (ii) between $AGE_{i,jk}$ and $Q_{i,jk}$ on genuine

Table 1. Multilevel models for genuine match score analysis

Model	Level 1 model	Level 2 model
Model A	$y_{i,jk} = \varphi_{0i} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i}$
Model B_T	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i},\ \varphi_{1i} = \beta_{10} + b_{1i}$
Model B_A	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}AGE_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i},\ \varphi_{1i} = \beta_{10} + b_{1i}$
Model B_Q	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}Q_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i},\ \varphi_{1i} = \beta_{10} + b_{1i}$
Model C_G	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + \beta_{01}bM_i + b_{0i},\ \varphi_{1i} = \beta_{10} + \beta_{11}bM_i + b_{1i}$
Model C_R	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + \beta_{01}bW_i + b_{0i},\ \varphi_{1i} = \beta_{10} + \beta_{11}bW_i + b_{1i}$
Model D	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varphi_{2i}AGE_{i,jk} + \varphi_{3i}Q_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i},\ \varphi_{1i} = \beta_{10} + b_{1i},\ \varphi_{2i} = \beta_{20} + b_{2i},\ \varphi_{3i} = \beta_{30} + b_{3i}$
Model E	$y_{i,jk} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varphi_{2i}AGE_{i,jk} + \varphi_{3i}Q_{i,jk} + \varphi_{4i}\triangle T_{i,jk}Q_{i,jk} + \varphi_{5i}AGE_{i,jk}Q_{i,jk} + \varepsilon_{i,jk}$	$\varphi_{0i} = \beta_{00} + b_{0i},\ \varphi_{1i} = \beta_{10} + b_{1i},\ \varphi_{2i} = \beta_{20} + b_{2i},\ \varphi_{3i} = \beta_{30} + b_{3i},\ \varphi_{4i} = \beta_{40},\ \varphi_{5i} = \beta_{50}$

match scores is not significant (see *SI Appendix*, section S8, for the detailed analysis with models D and E).

Outlying Subjects in Model B_T. The parameter estimates of $(\varphi_{0i}, \varphi_{1i})$ for each subject in model B_T are shown in *SI Appendix*, Fig. S10, in addition to the population-mean trend (β_{00}, β_{10}). Several outlying subjects whose genuine match score trend markedly deviates from the population-mean trend in terms of Mahalanobis distance are identified. *SI Appendix*, Figs. S11–S15 show the individual trends of the outlying subjects and their fingerprint impressions.

- Outlying subject 1 (*SI Appendix*, Fig. S11): The estimated intercept of this subject is very small. The subject consistently gives low genuine match scores because his fingerprints are severely scarred.
- Outlying subject 2 (*SI Appendix*, Fig. S12): The intercept of the fitted model for this subject is rather large, although the slope is negative. This subject consistently gives high genuine match scores because his fingerprint impressions are of good quality.
- Outlying subject 3 (*SI Appendix*, Fig. S13): This subject shows a very sharp decrease in genuine match scores as a function of time interval. In *SI Appendix*, Fig. S13*A*, the genuine match scores involving the first fingerprint impression are very low. This fingerprint impression is indeed an impostor fingerprint (*SI Appendix*, Fig. S13*B*); it is of tented arch type whereas the actual pattern of this finger is a right loop. After inspection of the subject's 10-print cards, it turned out that the subject's left and right hands were swapped at the time of the first 10-print card acquisition; the first impression came from the subject's left index finger, instead of right index finger. This shows that operational fingerprint data can be mislabeled.
- Outlying subject 4 (*SI Appendix*, Fig. S14): This subject also has a steep slope. It turned out that the fingerprint impressions of this subject were collected during his adolescence (starting at the age of 11 until the age of 21). This explains the sharp decrease in genuine match scores due to growth in finger size (19).
- Outlying subject 5 (*SI Appendix*, Fig. S15): A positive slope is observed for this subject because the comparisons involving a lower quality fingerprint were made over a shorter time interval than the comparisons with higher quality fingerprints. This example illustrates that the fingerprint image quality does not necessarily vary with respect to time elapsed.

Model Assessment and Comparison. Goodness-of-fit of a model evaluates how well the model fits the data. Furthermore, the impact of covariates on the observed responses can be assessed by comparing the goodness-of-fits of different models. The following three criteria are used to measure the goodness-of-fit: (*i*) deviance, (*ii*) Akaike information criterion (AIC), and (*iii*) Bayesian information criterion (BIC). The deviance measure is used to compare nested models, whereas AIC and BIC add a constant term to the deviance for the sake of comparing nonnested models (*SI Appendix*, section S9). The smaller the deviance (AIC or BIC), the better the model fit.

SI Appendix, Table S1 shows the goodness-of-fit measures of the multilevel models fit to genuine match scores obtained by the two COTS matchers. The model comparisons based on the goodness-of-fit lead to the following observations:

- A decrease in deviance is observed when models B_T, B_A, and B_Q are compared with model A (unconditional mean model or empty model). This means that each individual covariate used in model B ($\triangle T_{i,jk}$, $AGE_{i,jk}$, or $Q_{i,jk}$) can explain some of the variation in genuine match scores.
- Model B_Q provides a better fit to the data than models B_T and B_A. This implies that fingerprint quality explains better the variation in genuine match scores than time interval or subject's age.
- Sex and race are not important factors to explain the variation in genuine match scores because the deviance barely decreases from model B_T to models C_G or C_R.
- Models D and E show significantly smaller goodness-of-fit values than the other models. In other words, including all of the three covariates ($\triangle T_{i,jk}$, $AGE_{i,jk}$, and $Q_{i,jk}$) in the multilevel model better explains the trend in genuine match scores

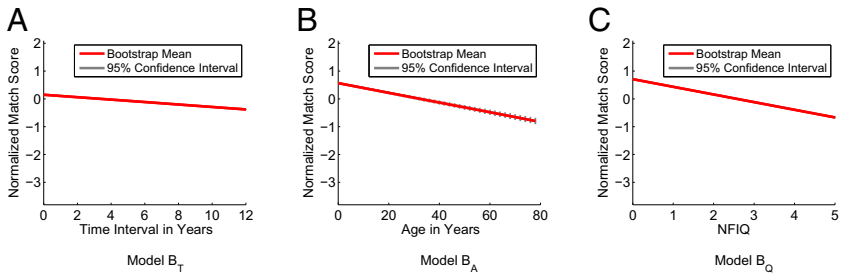


Fig. 1. Population-mean trends of genuine match scores obtained by COTS-1 matcher along with 95% confidence intervals with respect to (*A*) $\triangle T_{i,jk}$, (*B*) $AGE_{i,jk}$, and (*C*) $Q_{i,jk}$ when a single finger is used for recognition. The confidence intervals for models B_T and B_Q are too tight along the means to be visible.

compared with including only a single covariate. The additional interaction terms in model E slightly improve the model fit.

Population-Mean Trend of Impostor Fingerprint Match Scores. For impostor score analysis, 1,000 bootstrap samples are constructed to accurately capture the variability among distinct fingers, maintain the size of bootstrap sample feasible for model fitting, and avoid data dependency. The rth bootstrap sample contains a set of impostor scores for subject i $(i=1,2,\ldots,N)$, $\{s_{ij^*,k} | j^* \in J_i^{(r)}; k=1,2,\ldots,n_i\}$, where $J_i^{(r)}$ is a random set with indices of 10 impostor subjects for subject i. $J_i^{(r)}$ is sampled such that (i) $J_i^{(r_1)}$ and $J_i^{(r_2)}$, for $r_1 \neq r_2$, are independent, and (ii) $J_{j^*}^{(r)}$, where $j^* \in J_i^{(r)}$, does not include subject i.

The impact of time elapsed $(\triangle T_{i,1k})$ and subjects' ages ($AGE_{i,k}$ and $AGE_{j,1}$) on impostor match scores is evaluated in models B_T' and B_A', respectively. Note that model B_A' in *SI Appendix*, Eq. S4, has two covariates associated with the age of two subjects involved in an impostor comparison: age of the subject of interest for longitudinal analysis $(AGE_{i,k})$ and an impostor subject's age as a control $(AGE_{j,1})$.

Although the null hypothesis $\beta_{10}=0$ is rejected in both models B_T' and B_A' at a significance level of 0.05 (*SI Appendix*, Table S4 and Figs. S16–S18), (i) the decrease in the population-mean trend of impostor match scores is negligible for both COTS matchers as time elapsed since the first fingerprint acquisition increases up to 12 y, the maximum time span in the dataset, and (ii) subject's age appears to have marginal impact on impostor match scores in that the tendencies observed in two COTS matchers do not coincide (i.e., $\beta_{10}>0$ for COTS-1 matcher and $\beta_{10}<0$ for COTS-2 matcher), and the changes in impostor match score with the increase of subject's age up to 78 y are small. For a similar reason to (ii), the impostor subject's age has an insignificant impact on the impostor match scores.

Population-Mean Trend of Probability of True Acceptance. The population-mean trend of probability of true acceptance $(\pi_{i,jk})$ with respect to $\triangle T_{i,jk}$ in model B_T^* in *SI Appendix*, Eq. **S5** is investigated at various false acceptance rates (FARs) (0.01%, 0.00001%, and empirical 0%). The decision thresholds are determined based on the impostor score distribution obtained by making all pairwise comparisons among the first fingerprint impressions of N subjects. The threshold corresponding to FAR of p is determined by a maximum of match score x satisfying $cdf(x)<1-p$, where $cdf(x)$ is the cumulative distribution function of impostor match scores. The threshold corresponding to empirical 0% FAR is set to the maximum value in the impostor score set. The expected value of $\pi_{i,jk}$ at the operational FAR range between 0.01% and 0.00001% remains close to 1.0 for the both COTS matchers even though the time interval between two fingerprints in comparison increases up to 12 y (Fig. 2*A* and *SI Appendix*, Figs. S19 and S20 and Table S5). At empirical 0% FAR, an extreme FAR point, the expected $\pi_{i,jk}$ for COTS-1 matcher still remains close to 1.0, whereas COTS-2 matcher, which delivers inferior recognition accuracy compared with COTS-1 matcher, exhibits a considerable degradation in $\pi_{i,jk}$ with increase in $\triangle T_{i,jk}$.

To understand the joint impact of $\triangle T_{i,jk}$, $AGE_{i,jk}$, and $Q_{i,jk}$ on the probability of true acceptance, model D* in *SI Appendix*, Eq. **S6** is fit to the binary decisions on genuine fingerprint pairs when the threshold corresponding to 0.00001% FAR is applied to COTS-1 match scores. The population-mean trend of $\pi_{i,jk}$ with respect to $\triangle T_{i,jk}$ is shown in Fig. 3 and *SI Appendix*, Fig. S21, at various $AGE_{i,jk}$ and $Q_{i,jk}$ values. The bootstrap mean trend remains close to 1.0 across all age groups and fingerprint image quality levels. Interestingly, whereas 95% confidence intervals of $\pi_{i,jk}$ are consistent among different age groups at fixed $Q_{i,jk}$, 95% confidence interval of $\pi_{i,jk}$ increases substantially when the fingerprint quality decreases (recall, the higher the $Q_{i,jk}$, the lower the fingerprint quality) regardless of subject's age group. This indicates that the tendency of $\pi_{i,jk}$ with respect to $\triangle T_{i,jk}$ has a considerable uncertainty when either of two fingerprints in comparison is of poor quality. It conforms to the model assessment analysis that fingerprint image quality, rather than time interval between fingerprints in comparison or subject's age, explains well the variation in genuine match scores, and the large confidence intervals observed in models D and E fit to genuine match scores.

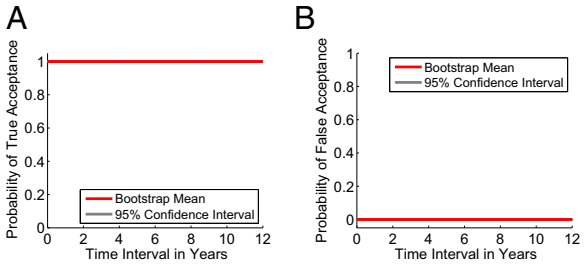


Fig. 2. Population-mean trend of fingerprint recognition accuracy along with 95% confidence interval with respect to $\triangle T_{i,jk}$. (*A*) Probability of true acceptance ($\pi_{i,jk}$ in model B_T^*) and (*B*) probability of false acceptance ($\pi_{ij,k}$ in model $B_T^{*\prime}$). Match scores are obtained by COTS-1 matcher when a single finger is used for recognition. The confidence intervals are too tight along the means to be visible in the plots.

Population-Mean Trend of Probability of False Acceptance. The population-mean trend of probability of false acceptance $(\pi_{ij,k})$ is also investigated at various FARs (0.01%, 0.1%, and 1%). Fig. 2*B* and *SI Appendix*, Figs. S22 and S23 and Table S5 show that the predicted value of $\pi_{ij,k}$ remains close to 0.0 regardless of $\triangle T_{i,1k}$ within 12 y. This implies that the binary decisions made on impostor fingerprint comparisons over time are not likely to be affected by the time elapsed.

Results When Using All 10 Fingers for Recognition. Ten-finger fusion results of genuine match scores (*SI Appendix*, Table S6 and Fig. S24) and impostor match scores (*SI Appendix*, Table S7 and Figs. S25–S27) conform to the results in the single-finger experiments. It is noteworthy that the probability of true acceptance remains 1.0 for both matchers when match scores from 10 fingers are fused, even at the extreme FAR point of 0% (*SI Appendix*, Table S8 and Figs. S28 and S29). The probability of false acceptance stays close to 0.0 with the increase in time elapsed (*SI Appendix*, Table S8 and Figs. S30 and S31).

Conclusions

Since ancient times, fingerprints have been accepted as persistent and unique to an individual. Early scientific studies on fingerprint recognition in the late 19th century claimed that there is no significant change in the friction ridge structure over time by examining small sets of genuine fingerprint pairs captured over a large time interval. Although fingerprint recognition is now prevalent in distinguishing a large number of individuals—for example, Federal Bureau of Investigation's Next Generation Identification searches a fingerprint database holding 106 million criminal and civil files (20)—acceptance of the persistence of fingerprints has been mostly based on anecdotal evidence.

To understand the behavior of fingerprint match score and recognition accuracy, multiple fingerprint records of 15,597 subjects apprehended by the MSP over a time span varying from 5 to 12 y were investigated. The genuine and impostor match scores obtained by two COTS fingerprint matchers were analyzed by linear multilevel statistical models with various covariates, including time interval between two fingerprints being compared, subject's demographic factors such as age, sex, and race, and fingerprint image quality. The longitudinal study of fingerprint recognition reported in this paper leads to the following conclusions:

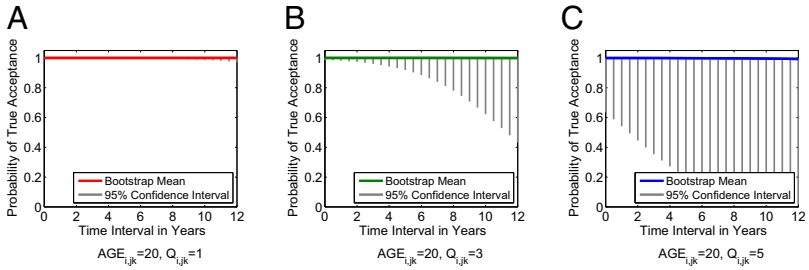


Fig. 3. Population-mean trend and 95% confidence interval of probability of true acceptance $(\pi_{i,jk})$ with respect to $\triangle T_{i,jk}$ when $AGE_{i,jk}$ is 20 and $Q_{i,jk}$ is (*A*) 1, (*B*) 3, and (*C*) 5 in model D*. The decision threshold is set to the value corresponding to FAR of 0.00001%. Match scores are obtained by COTS-1 matcher when a single finger is used for recognition.

- The hypothesis test for the slope of a linear model indicates that genuine match scores tend to decrease as the time interval between two fingerprints being compared increases. Furthermore, genuine match scores tend to decrease as the subject's age increases or when the fingerprint image quality decreases.
- Despite the downward trend in genuine match scores over time, the probability of true acceptance, at operational FAR settings, remains close to 1.0 (up to 12 y, the maximum time span in the dataset). However, if either of two fingerprints in comparison is of poor quality, the uncertainty in the expected probability of true acceptance becomes considerably large.
- The changes in impostor match scores with respect to time elapsed and subject's age are negligible. Hence, the probability of false acceptance remains close to 0.0 regardless of the time interval between two fingerprints.
- A comparison among the models with different covariates fit to the genuine match scores shows the following:
  - Time interval, subject's age, and fingerprint image quality can explain the variation in genuine match scores, whereas subject's sex and race have marginal impact.
  - Fingerprint image quality explains the variation in genuine match scores better than time interval and subject's age.
  - The correlations (*i*) between time interval and fingerprint image quality and (*ii*) between subject's age and fingerprint image quality are negligibly small in the population-mean trend analysis.
  - The impact of the interactions (*i*) between time interval and fingerprint image quality and (*ii*) between subject's age and fingerprint image quality on genuine match scores is not significant.
- Outlying subjects in the dataset who do not conform to the population-mean trend as determined by model fit are examined. These outlying subjects illustrate (*i*) a case where a COTS matcher consistently provides high genuine match scores due to high-quality fingerprints from the same finger, (*ii*) an example where the matcher generates low genuine match scores due to the scarring of a finger, (*iii*) a degradation in genuine match scores when a juvenile fingerprint is compared with the corresponding adult fingerprint, and (*iv*) presence of labeling errors in the operational fingerprint databases.
- The inferences from single (right index) finger analysis conform to the inferences from 10-finger score fusion analysis.
- The results from two different COTS fingerprint matchers used in the study coincide, except the tendency of impostor match scores with respect to subject's age and the temporal tendency of probability of true acceptance at empirical 0% FAR in the single-finger analysis.

Our future work will include the following: (*i*) given that we make all pairwise comparisons of the fingerprint impressions from each subject, the correlation among the genuine match scores of a subject needs to be reflected in the model; and (*ii*) nonlinear multilevel models will be investigated and compared with the linear models presented in this study.

ACKNOWLEDGMENTS. We thank Prof. Joseph Gardiner (Department of Epidemiology and Biostatistics, Michigan State University) and Karthik Nandakumar (IBM, Singapore) for helpful discussions and suggestions. We also acknowledge the help of Capt. Greg Michaud, Director of Forensic Science Division at the Michigan State Police, in providing us the fingerprint dataset used in this study. This research was supported by a grant from the National Science Foundation Center for Identification Technology Research.

1. Hawthorne MR (2009) *Fingerprints: Analysis and Understanding* (CRC, Boca Raton, FL).
2. Frye v. United States (1923) 54 App. D.C. 46, 293 F. 1013.
3. Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993) 509 U.S. 579, 113 S. Ct. 2786.
4. Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) (2013) *Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint)*. Available at www.swgfast.org/documents/examinations-conclusions/130427_Examinations-Conclusions_2.0.pdf. Accessed May 12, 2015.
5. Vanderkolk JR (2011) Examination process. *Fingerprint Sourcebook* (National Institute of Justice, Washington, DC), Chap 9.
6. Pankanti S, Prabhakar S, Jain AK (2002) On the individuality of fingerprints. *IEEE Trans Pattern Anal Mach Intell* 24(8):1010–1025.
7. Zhu Y, Dass SC, Jain AK (2007) Statistical models for assessing the individuality of fingerprints. *IEEE Trans Inform Forensics Secur* 2(3):391–401.
8. Chen Y, Jain AK (2009) Beyond minutiae: A fingerprint individuality model with pattern, ridge and pore features. *Third IAPR/IEEE International Conference on Biometrics*, Lecture Notes in Computer Science, eds Tistarelli M, Nixon MS (Springer, Berlin), Vol 5558, pp 523–533.
9. Neumann C, et al. (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *J Forensic Sci* 52(1):54–64.
10. Herschel WJ (1916) *The Origins of Finger-Printing* (Oxford Univ Press, Oxford).
11. Galton F (1892) *Finger Prints* (Macmillan, London).
12. Cummins H, Midlo M (1961) *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics* (Dover Publications, New York).
13. National Research Council (2009) *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, Washington, DC).
14. Mansfield AJ, Wayman JL (2002) *Best Practices in Testing and Reporting Performance* (Center for Mathematics and Scientific Computing, National Physical Laboratory, Teddington, UK).
15. Federal Office for Information Security (BSI) (2004) Evaluation of fingerprint recognition technologies–BioFinger. Available at https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Studies/BioFinger/BioFinger_pdf.pdf?__blob=publicationFile. Accessed May 12, 2015.
16. Goldstein H (2010) *Multilevel Statistical Models* (Wiley, Chichester, UK), 4th Ed.
17. Singer JD, Willett JB (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford Univ Press, Oxford).
18. Tabassi E, Wilson CL, Watson CI (2004) Fingerprint image quality. National Institute of Standards and Technology Interagency Report 7151 (National Institute of Standards and Technology, Gaithersburg, MD).
19. Gottschlich C, et al. (2011) Modeling the growth of fingerprints improves matching for adolescents. *IEEE Trans Inform Forensics Secur* 6(3):1165–1169.
20. Federal Bureau of Investigation (2014) Next Generation Identification (NGI). Available at www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi. Accessed May 12, 2015.

Longitudinal Study of Fingerprint Recognition
Supporting Information Appendix

Soweon Yoon*† and Anil K. Jain*
*Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824
†National Institute of Standards and Technology, Gaithersburg, MD 20899

S1 Caveats

The results and conclusions made in this paper should be interpreted with the following caveats:

(1) Any variability that appears in fingerprint images collected over time has two potential causes: (i) physical changes in inherent friction ridge structure of finger skin, and (ii) changes in imaging condition and subject's behavior during fingerprint acquisition (e.g., changes in sensing technology and methodology, finger skin condition, seasonal effect, subject's habituation to fingerprint capturing, etc.). It is not possible for us to distinguish between the two sources of variability due to the lack of information required for conducting such an analysis.

(2) The inferences and conclusions presented in this paper are drawn from tenprint analysis (more specifically, analysis based on rolled fingerprints) and do not suggest the same conclusions for latent fingerprint (or finger mark) analysis. This is because latent comparison is different from tenprint matching in the following aspects:

- As a latent fingerprint is lifted from the surface where the fingerprint impression was left unintentionally, the variability presented in the images is generally much greater than tenprint images which are typically captured in a controlled environment. A suitable metric for assessing the variability existing in the latent fingerprints is not yet available.
- In current practice, latent fingerprint identification involves both human examiners and Automated Fingerprint Identification Systems (AFIS) (i.e., a latent examiner compares the top-K candidates retrieved from a fingerprint database by AFIS) instead of making a binary decision based on match score.

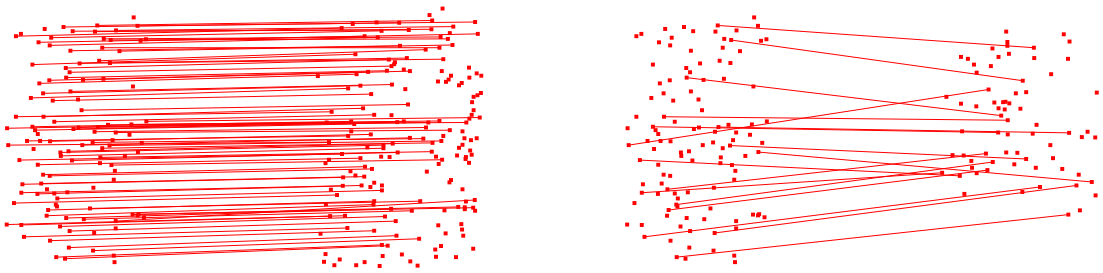
Thus, the modeling and analysis schemes used in this paper cannot be directly applied to latent fingerprints. Further, to the best of our knowledge, there is no latent fingerprint database available for longitudinal study, which contains multiple latent impressions from each finger over time.

S2 Definitions

The terminologies used in this paper are defined as follows.

Evidential value Evidential value of a comparison between two fingerprints refers to the strength of the fingerprint comparison that can be used as evidence to claim whether or not they come from the same finger [9, 11].

Longitudinal data Longitudinal data refers to repeated measurements on a collection of individuals sampled from a population over time. This is in contrast to cross-sectional data where a single measurement is made on each individual [16].



(A) (B)

Figure S1: Fingerprint comparison using minutiae configuration. Minutiae correspondences are shown for (A) a genuine fingerprint pair and (B) an impostor fingerprint pair. The fingerprint match (similarity) scores obtained by the COTS-2 matcher are (A) 389 and (B) 11 (note that the match score corresponding to false acceptance rate (FAR) of 0.01% is 24).

Balanced and time-structured data A longitudinal dataset is characterized by (i) the number of measurements per individual and (ii) the time schedule used to make the measurements [17]. *Balanced* dataset means that every subject has the same number of measurements. *Time-structured* dataset consists of the repeated measurements following an identical time schedule across individuals. The sequence of measurements for each individual can be spaced either regularly or irregularly.

Compound symmetry The compound symmetry requires (i) homoscedasticity of variance: the variance of the measurements at a time instance across all subjects is the same as that of the measurements at another time instance, and (ii) constant covariance: the correlation between the measurements at the first and second time instances, for example, is the same as that between the measurements at the first and third time instances, and so on.

S3 Fingerprint Recognition

A fingerprint pattern consists of intervening ridge lines that are equidistantly spaced. Fingerprint features used for matching, both by forensic experts and machines (i.e., AFIS), are typically represented at three different levels: (i) level-1 features (orientation field and singular points) describe ridge flow and pattern type (e.g., arch, loop, and whorl), (ii) level-2 features (minutiae) represent ridge details such as ridge ending and bifurcation points, and (iii) level-3 features (pore, incipient ridges, etc.) represent the finest details in fingerprints [S1].

A comparison between two fingerprints is primarily based on the spatial configurations of minutiae in the corresponding impressions. If two fingerprint impressions show a high degree of agreement in minutiae configurations (resulting in high match score), the fingerprints are claimed to be a genuine pair, originating from the same finger (Fig. S1(A)). Otherwise, they are claimed to be an impostor pair (Fig. S1(B)). Note that these decisions may not be concurrent with ground truth; in such case, decision errors occur.

Starting around 1900, the Scotland Yard included fingerprints in anthropometric identification cards which recorded measurements of various physical attributes of criminals [S2]. Since then, the use of fingerprints has spread rapidly worldwide primarily for the purpose of tracking habitual criminals (repeat offenders) and identifying suspects based on latent fingerprints found at crime scenes. With a phenomenal and continual increase in the size of fingerprint databases held by various law enforcement agencies, fingerprint recognition technology has made great strides both in terms of matching accuracy and matching speed (throughput). The Federal Bureau of Investigation (FBI) alone currently holds over 106 million tenprint records of apprehended criminals and civilian government job applicants as of March 2015 [20]. The FBI's Next Generation Identification (NGI) responds to a tenprint record of arrests and prosecutions (RAP) sheet request in 1 minute and 11 seconds on average (97% of the requests are completed within 15 seconds) [20]. In the Fingerprint Vendor Technology Evaluation (FpVTE) reported in 2014 [S3], the best performing

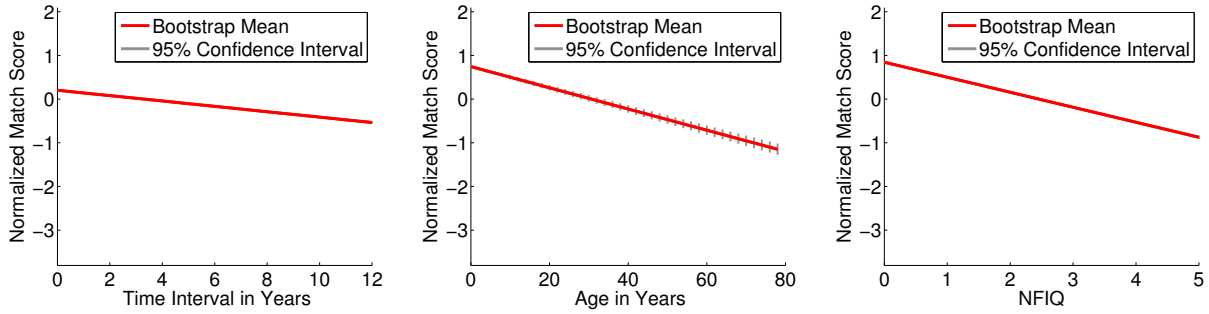
Table S2: Parameter estimates and 95% confidence intervals of genuine match scores obtained by COTS-1 matcher when a single finger (right index) is used for recognition

	Parameters	Model B_T	Model B_A	Model B_Q	Model D	Model E
Fixed Effects	β_{00}	0.1496	0.5682	0.7087	0.9137	1.1472
		(0.1406; 0.1590)	(0.5335; 0.6015)	(0.6954; 0.7221)	(0.8765; 0.94710)	(1.0827; 1.2100)
	β_{10}	-0.0440	-0.0175	-0.2750	-0.0368	-0.0283
		(-0.0450; -0.0430)	(-0.0185; -0.0164)	(-0.2798; -0.2702)	(-0.0378; -0.0358)	(-0.0313; -0.0255)
	β_{20}				-0.0030	-0.0110
					(-0.0042; -0.0019)	(-0.0130; -0.0090)
	β_{30}				-0.2509	-0.3486
					(-0.2558; -0.2463)	(-0.3739; -0.3241)
	β_{40}					-0.0035
						(-0.0045; -0.0024)
	β_{50}					0.0033
						(0.0026; 0.0041)
Variance Components	σ_{ε}^2	0.4980	0.4897	0.4210	0.3640	0.3639
	σ_0^2	0.5298	5.6056	0.9096	6.6323	6.6565
	σ_1^2	0.0034	0.0050	0.1163	0.0041	0.0041
	σ_{01}	-0.0134	-0.1576	-0.2543	0.0949	0.0950
	σ_2^2				0.0068	0.0068
	σ_{02}				-0.1941	-0.1945
	σ_{12}				-0.2092	-0.2207
	σ_3^2				0.1165	0.1181
	σ_{03}				-0.0036	-0.0036
	σ_{13}				-0.0007	-0.0011
	σ_{23}				-0.0013	-0.0010

Table S3: Parameter estimates and 95% confidence intervals of genuine match scores obtained by COTS-2 matcher when a single finger is used for recognition

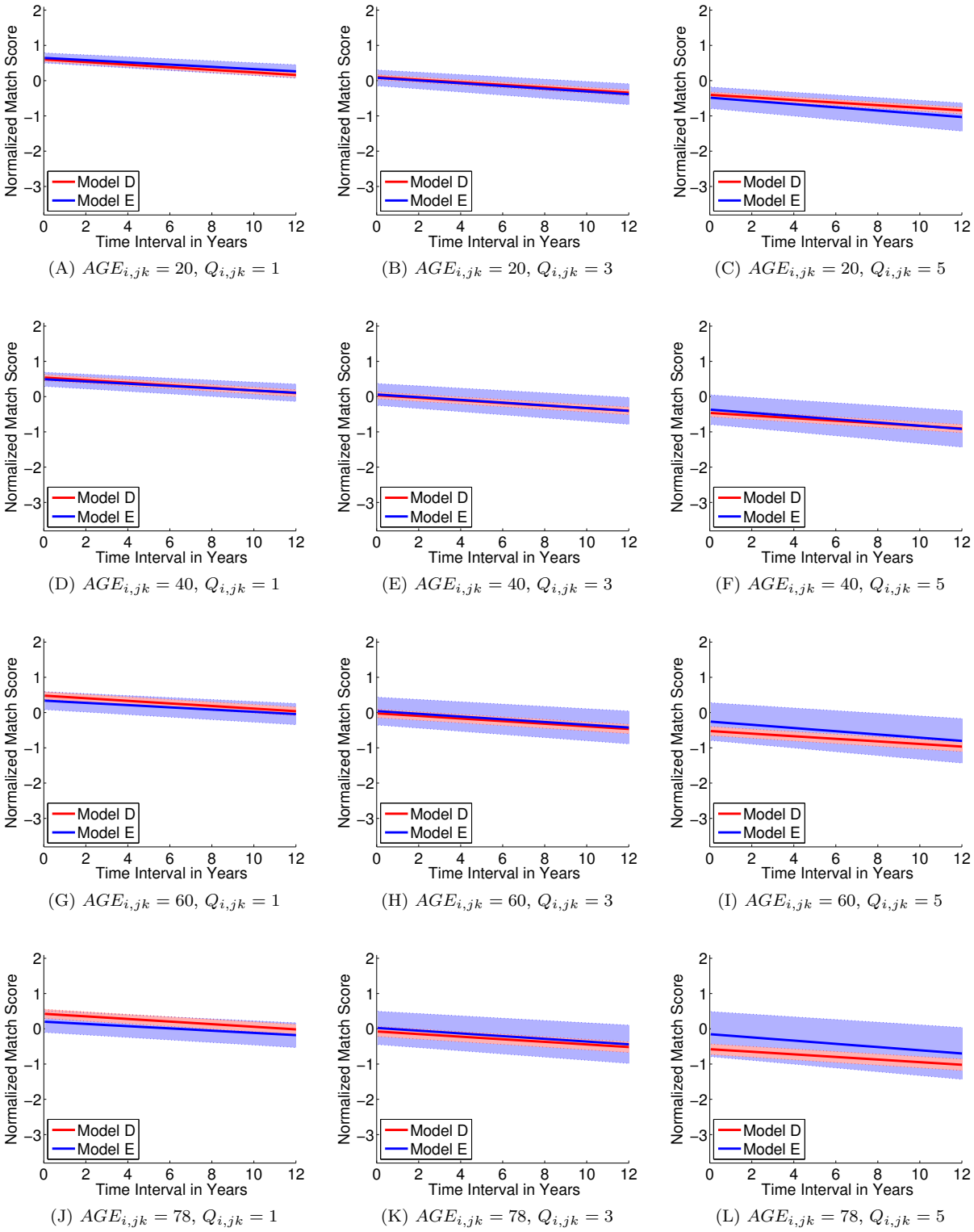
	Parameters	Model B_T	Model B_A	Model B_Q	Model D	Model E
Fixed Effects	β_{00}	0.2032	0.7447	0.8456	1.0353	1.4398
		(0.1939; 0.2127)	(0.7072; 0.7843)	(0.8316; 0.8595)	(0.9946; 1.0750)	(1.3706; 1.5102)
	β_{10}	-0.0616	-0.0243	-0.3439	-0.0533	-0.0654
		(-0.0625; -0.0606)	(-0.0254; -0.0231)	(-0.3489; -0.3385)	(-0.0543; -0.0522)	(-0.0678; -0.0629)
	β_{20}				-0.0024	-0.0130
					(-0.0036; -0.0011)	(-0.0152; -0.0107)
	β_{30}				-0.3064	-0.4693
					(-0.3112; -0.3015)	(-0.4925; -0.4466)
	β_{40}					0.0048
						(0.0039; 0.0057)
	β_{50}					0.0043
						(0.0036; 0.0050)
Variance Components	σ_ε^2	0.5162	0.5070	0.4540	0.3755	0.3751
	σ_0^2	0.5744	7.5573	0.9105	7.8990	7.8357
	σ_1^2	0.0039	0.0066	0.1027	0.0040	0.0040
	σ_{01}	-0.0277	-0.2136	-0.2473	0.0800	0.0825
	σ_2^2				0.0081	0.0081
	σ_{02}				-0.2335	-0.2314
	σ_{12}				-0.1466	-0.1646
	σ_3^2				0.1039	0.1077
	σ_{03}				-0.0033	-0.0033
	σ_{13}				-0.0004*	-0.0005*
	σ_{23}				-0.0030	-0.0027

* The hypothesis test indicates that the null hypothesis that the parameter is zero is not rejected at a significance level of 0.05.



(A) Model B_T (B) Model B_A (C) Model B_Q

Figure S6: Population-mean trends of genuine match scores obtained by COTS-2 matcher and 95% confidence intervals with respect to (A) $\triangle T_{i,jk}$, (B) $AGE_{i,jk}$, and (C) $Q_{i,jk}$, when a single finger is used for recognition. The confidence intervals for models B_T and B_Q are too tight along the means to be visible.



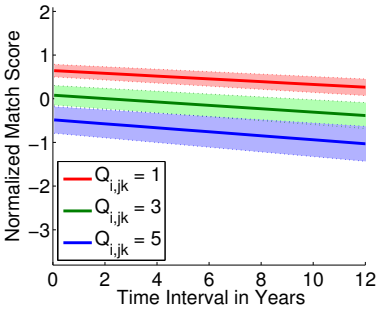
(A) $AGE_{i,jk} = 20$, $Q_{i,jk} = 1$ (B) $AGE_{i,jk} = 20$, $Q_{i,jk} = 3$ (C) $AGE_{i,jk} = 20$, $Q_{i,jk} = 5$

(D) $AGE_{i,jk} = 40$, $Q_{i,jk} = 1$ (E) $AGE_{i,jk} = 40$, $Q_{i,jk} = 3$ (F) $AGE_{i,jk} = 40$, $Q_{i,jk} = 5$

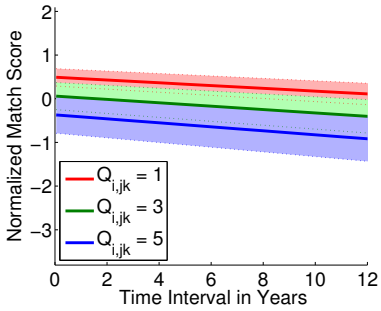
(G) $AGE_{i,jk} = 60$, $Q_{i,jk} = 1$ (H) $AGE_{i,jk} = 60$, $Q_{i,jk} = 3$ (I) $AGE_{i,jk} = 60$, $Q_{i,jk} = 5$

(J) $AGE_{i,jk} = 78$, $Q_{i,jk} = 1$ (K) $AGE_{i,jk} = 78$, $Q_{i,jk} = 3$ (L) $AGE_{i,jk} = 78$, $Q_{i,jk} = 5$

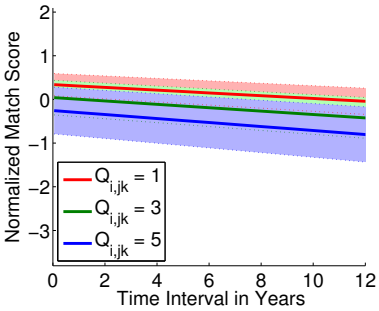
Figure S7: Comparison between models D and E. Population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ are shown when $AGE_{i,jk}$ varies from 20 to 78 and $Q_{i,jk}$ varies from 1 to 5 in models D and E. Solid lines are the bootstrap means, and the shaded areas represent the 95% confidence intervals. A single finger is used for recognition and match scores are obtained from COTS-1 matcher.



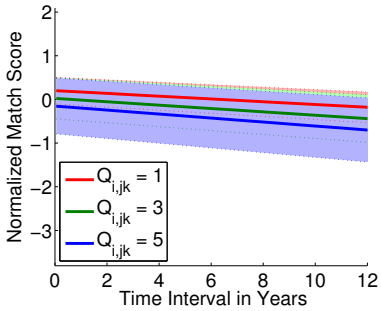
(A) $AGE_{i,jk} = 20$



(B) $AGE_{i,jk} = 40$

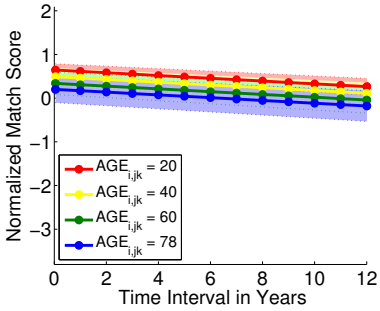


(C) $AGE_{i,jk} = 60$

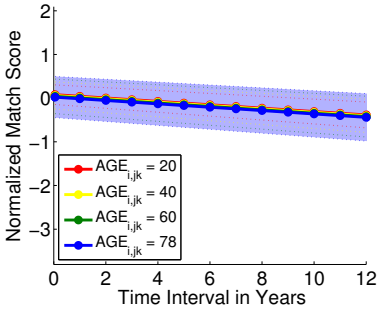


(D) $AGE_{i,jk} = 78$

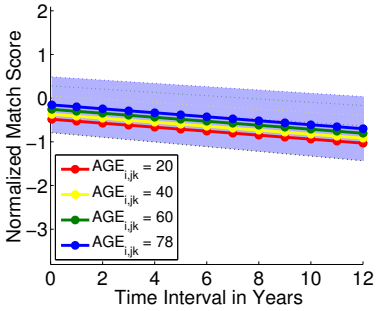
Figure S8: Population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ when $AGE_{i,jk}$ is fixed and $Q_{i,jk}$ varies from 1 to 5 in model E. Solid lines are the bootstrap means, and the shaded areas represent the 95% confidence intervals. A single finger is used for recognition and match scores are obtained from COTS-1 matcher.



(A) $Q_{i,jk} = 1$

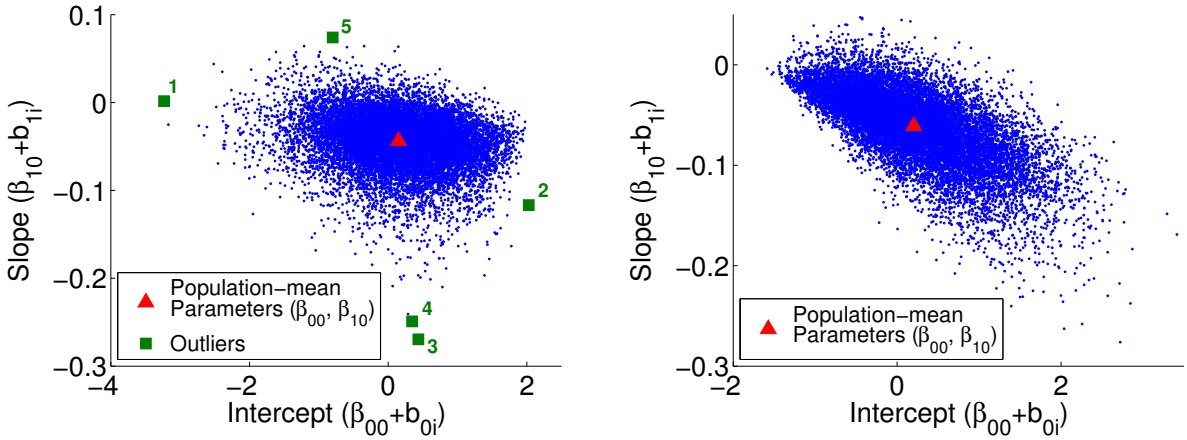


(B) $Q_{i,jk} = 3$



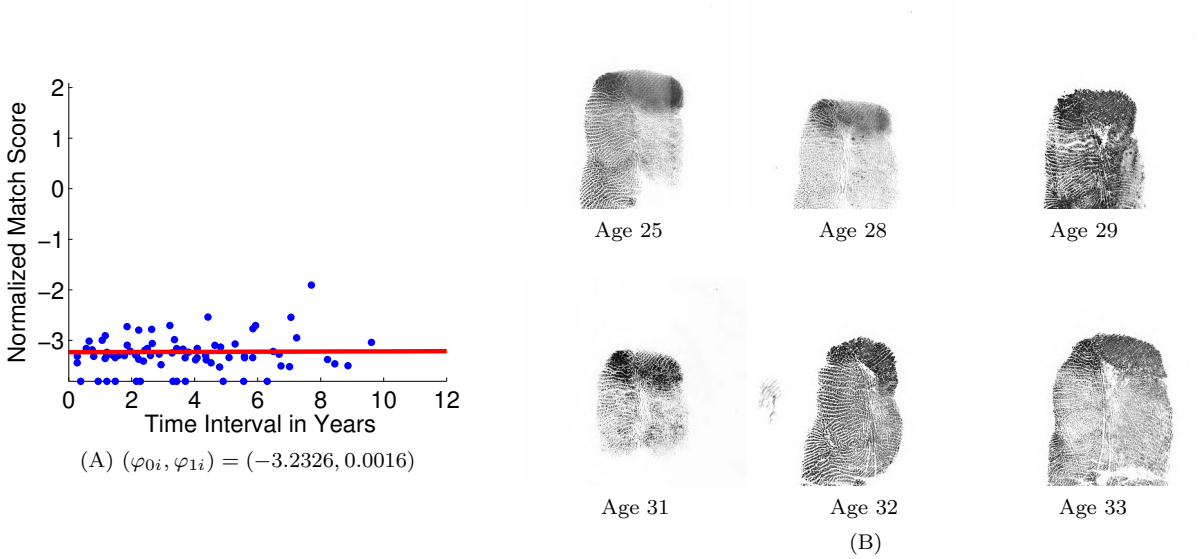
(C) $Q_{i,jk} = 5$

Figure S9: Population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ when $Q_{i,jk}$ is fixed and $AGE_{i,jk}$ varies from 20 to 78 in model E. Solid lines are the bootstrap means, and the shaded areas represent the 95% confidence intervals. A single finger is used for recognition and match scores are obtained from COTS-1 matcher.



(A) COTS-1 (B) COTS-2

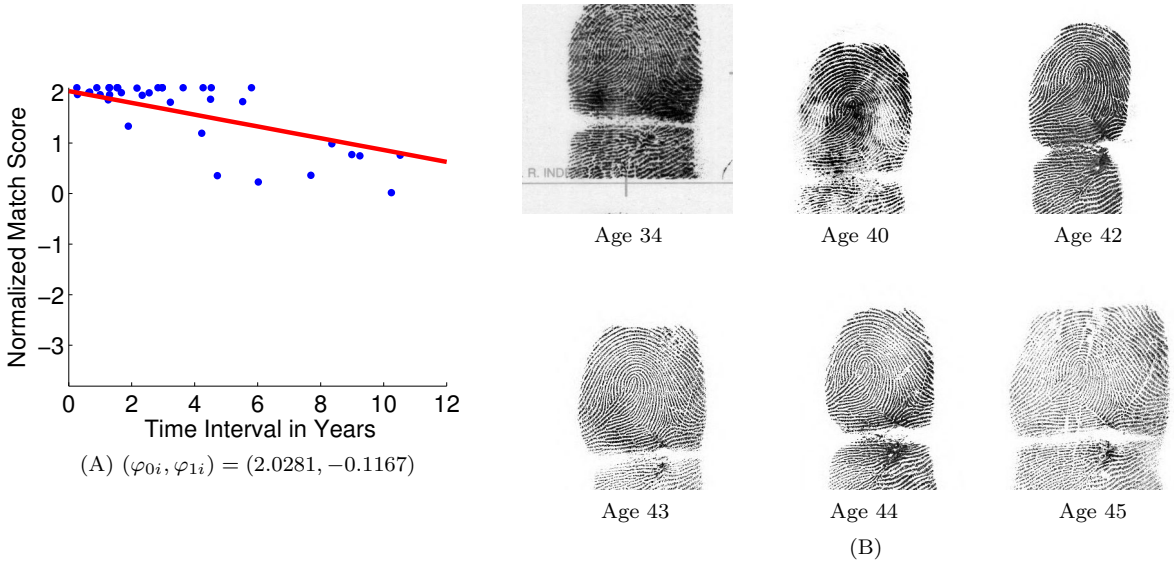
Figure S10: Parameter estimates of model B_T with genuine match scores provided by two COTS matchers. The estimates for the population-mean parameters (β_{00}, β_{10}) and the parameters for each subject $(\varphi_{0i}, \varphi_{1i}) = (\beta_{00} + b_{0i}, \beta_{10} + b_{1i})$ are represented as red triangles and blue dots, respectively. The parameters associated with five outlying subjects are marked as green squares in (A).



(A) $(\varphi_{0i}, \varphi_{1i}) = (-3.2326, 0.0016)$

(B)

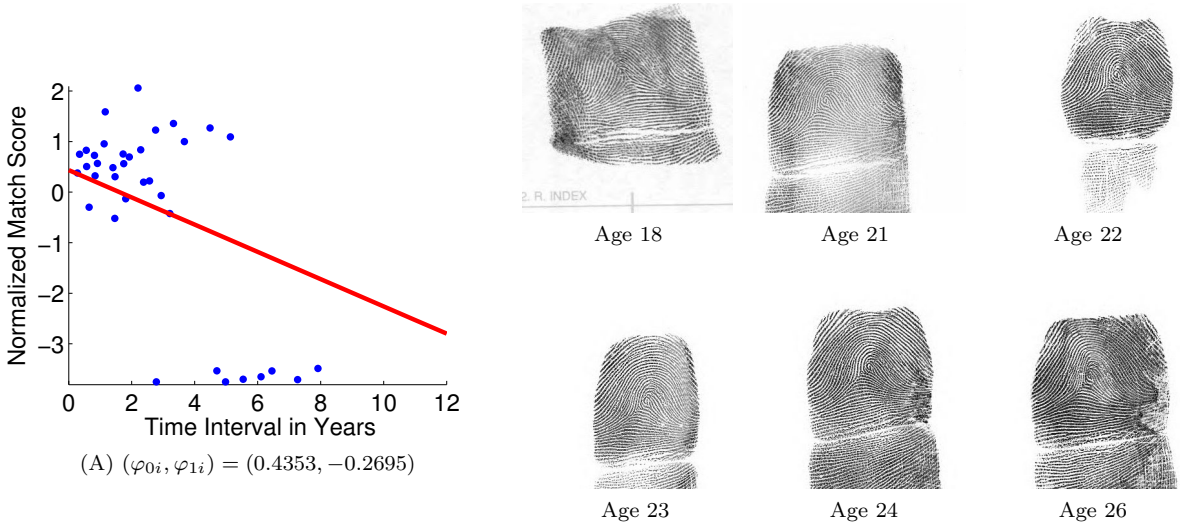
Figure S11: A subject whose intercept in model B_T is very small due to the severe alteration (i.e., scarring) of the fingerprint pattern (outlying subject 1). (A) The observed responses and fitting result of the subject, and (B) fingerprint impressions of the subject at different ages.



(A) $(\varphi_{0i}, \varphi_{1i}) = (2.0281, -0.1167)$

(B)

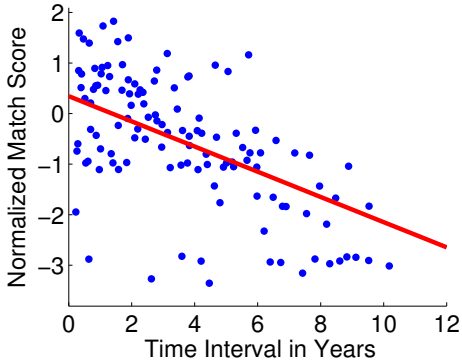
Figure S12: A subject with high quality ridge pattern resulting in the large intercept in model B_T (outlying subject 2). (A) The observed responses and fitting result of the subject, and (B) fingerprint impressions of the subject at different ages.



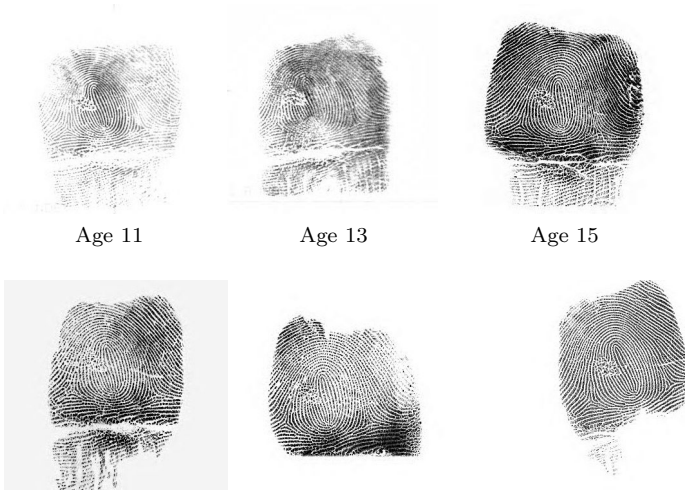
(A) $(\varphi_{0i}, \varphi_{1i}) = (0.4353, -0.2695)$

(B)

Figure S13: A subject with steep negative slope (outlying subject 3) resulting from a mislabeled fingerprint (the first impression). (A) The observed responses and fitting result of the subject, and (B) fingerprint impressions of the subject at different ages.

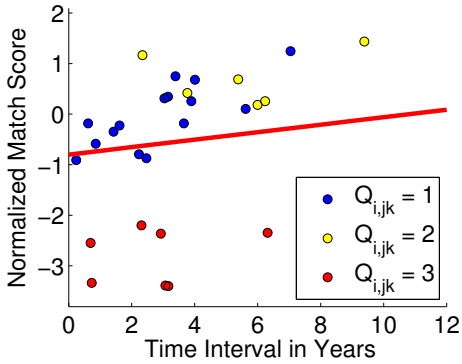


(A) $(\varphi_{0i}, \varphi_{1i}) = (0.3443, -0.2491)$

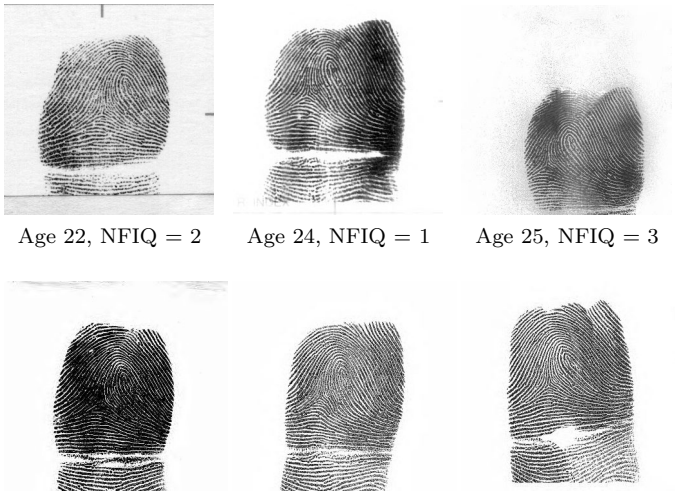


(B)

Figure S14: A subject with steep negative slope due to fingerprint impressions made during his adolescence (outlying subject 4). (A) The observed responses and fitting result of the subject, and (B) fingerprint impressions of the subject at different ages.



(A) $(\varphi_{0i}, \varphi_{1i}) = (-0.8006, 0.0740)$



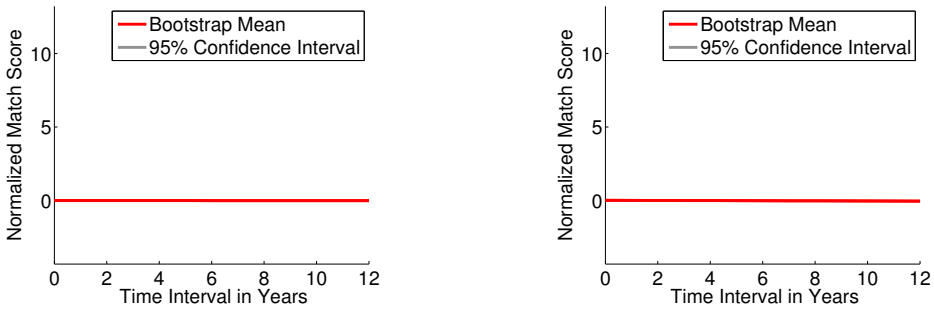
(B)

Figure S15: A subject with positive slope (outlying subject 5) where the comparisons involving a lower quality fingerprint (at age 25) have short time intervals. (A) The observed responses and fitting result of the subject, and (B) fingerprint impressions of the subject at different ages.

Table S4: Parameter estimates and 95% confidence intervals of impostor match scores obtained by two COTS matchers when a single finger is used for recognition

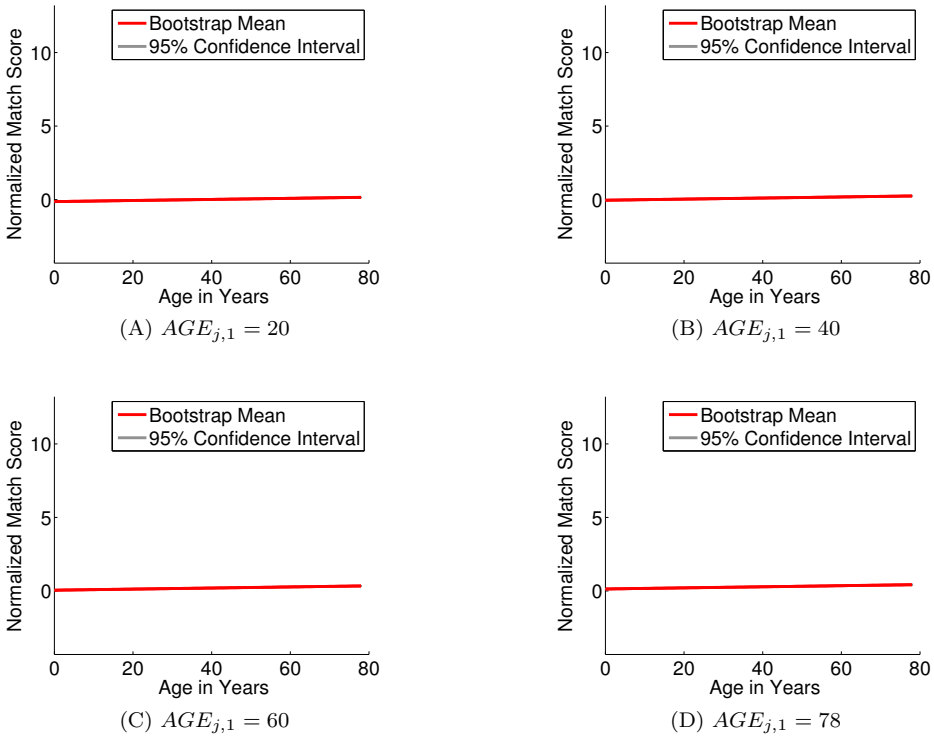
	Parameters	COTS-1		COTS-2	
		Model B'_T	Model B'_A	Model B'_T	Model B'_A
Fixed Effects	β_{00}	0.0017*	-0.2271	0.0158	0.0578
		(-0.0005; 0.0039)	(-0.2407; -0.2139)	(0.0135; 0.0180)	(0.0462; 0.0696)
	β_{10}	-0.0006	0.0037	-0.0042	-0.0008
		(-0.0010; -0.0001)	(0.0034; 0.0040)	(-0.0047; -0.0037)	(-0.0011; -0.0005)
	β_{20}		0.0047		-0.0014
			(0.0043; 0.0050)		(-0.0017; -0.0011)
Variance Components	σ^2_ε	0.8760	0.8744	0.9454	0.9477
	σ^2_0	0.1322	0.1841	0.0590	0.0647
	σ^2_1	1.0065e-05	1.5156e-05	0.0002	1.5917e-06
	σ_{01}	-0.0009	-0.0014	-0.0013	-0.0003

* The hypothesis test indicates that the null hypothesis that the parameter is zero is not rejected at a significance level of 0.05.



(A) COTS-1 (B) COTS-2

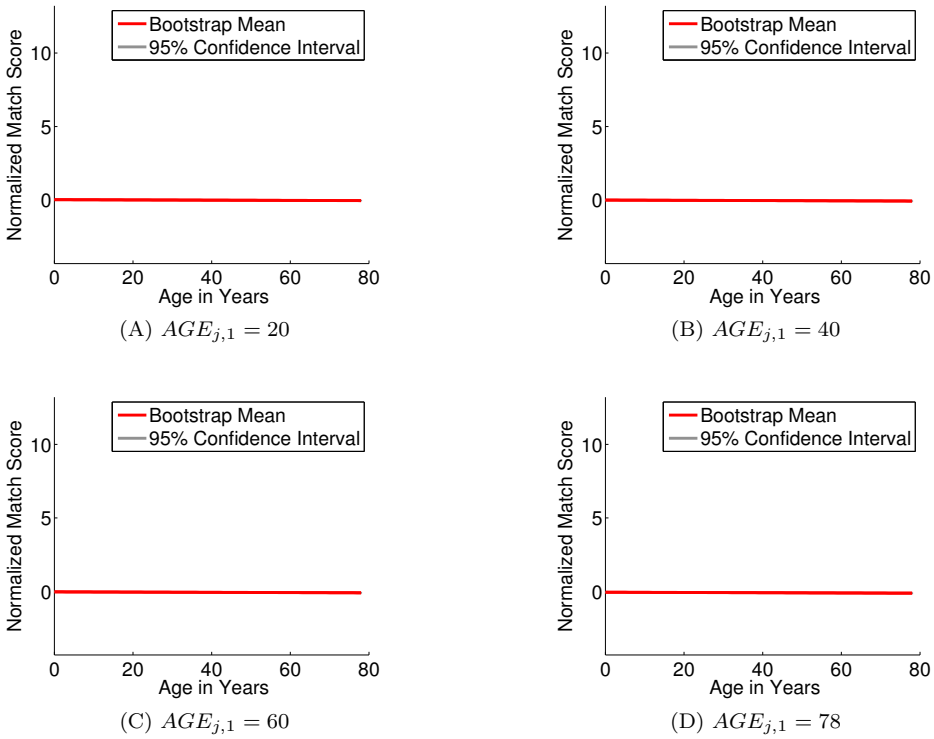
Figure S16: Population-mean trends and 95% confidence intervals of impostor match scores obtained by two COTS matchers with respect to $\triangle T_{i,1k}$ (model B'_{T}), when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.



(A) $AGE_{j,1} = 20$ (B) $AGE_{j,1} = 40$

(C) $AGE_{j,1} = 60$ (D) $AGE_{j,1} = 78$

Figure S17: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-1 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.



(A) $AGE_{j,1} = 20$ (B) $AGE_{j,1} = 40$

(C) $AGE_{j,1} = 60$ (D) $AGE_{j,1} = 78$

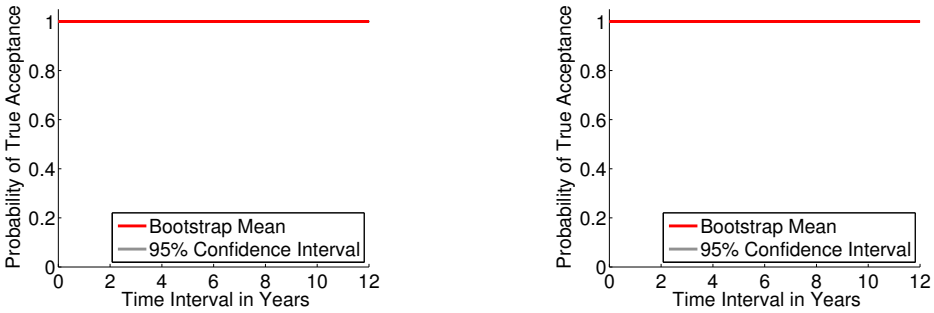
Figure S18: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-2 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.

Table S5: Expected probability of true acceptance ($\pi_{i,jk}$) and probability of false acceptance ($\pi_{ij,k}$) and their 95% confidence intervals, when a single finger is used for recognition

		COTS-1			COTS-2		
	Time Interval in Years	0	6	12	0	6	12
$\pi_{i,jk}$	FAR = 0.01%	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
		(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(0.9999; 1.0000)	(0.9998; 0.9999)
	FAR = 0.00001%	1.0000	1.0000	0.9999	0.9988	0.9976	0.9954
		(1.0000; 1.0000)	(0.9999; 1.0000)	(0.9999; 0.9999)	(0.9986; 0.9990)	(0.9969; 0.9982)	(0.9933; 0.9968)
	FAR = 0%	0.9999	0.9999	0.9998	0.9791	0.9536	0.9004
		(0.9999; 0.9999)	(0.9998; 0.9999)	(0.9997; 0.9999)	(0.9778; 0.9803)	(0.9488; 0.9579)	(0.8863; 0.9124)
$\pi_{ij,k}$	FAR = 0.01%	0.0264e-05	0.0215e-05	0.0175e-05	0.0146e-05	0.0005e-05	0.0000e-05
		(0.0144; 0.1012)	(0.0053; 0.1402)	(0.0020; 0.1942)	(0.0053; 7.3642)	(0.0000; 5.3964)	(0.0000; 3.9544)
	FAR = 0.1%	2.2000e-05	1.8930e-05	1.6288e-05	0.1555e-05	0.0266e-05	0.0045e-05
		(1.1828; 6.1468)	(0.4414; 6.7883)	(0.1648; 7.4968)	(0.0961; 0.8369)	(0.0000; 0.4699)	(0.0000; 0.2638)
	FAR = 1%	0.0062	0.0057	0.0053	0.0056	0.0040	0.0028
		(0.0058; 0.0065)	(0.0051; 0.0064)	(0.0044; 0.0063)	(0.0053; 0.0060)	(0.0035; 0.0046)	(0.0023; 0.0035)

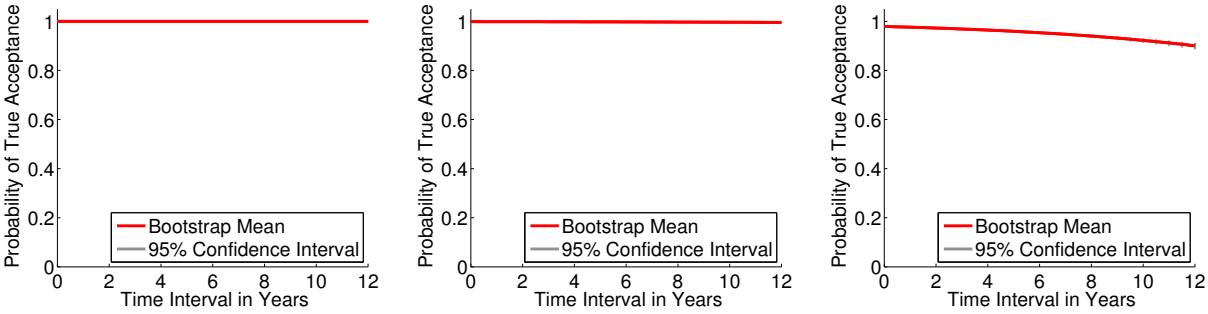
* Numbers in the tables are rounded off to four decimal points unless written in exponential notation. For some very small numbers, round-off to nine decimal points is used.

** If the expected probability is in exponential notation, its confidence interval is also in the same exponential notation.



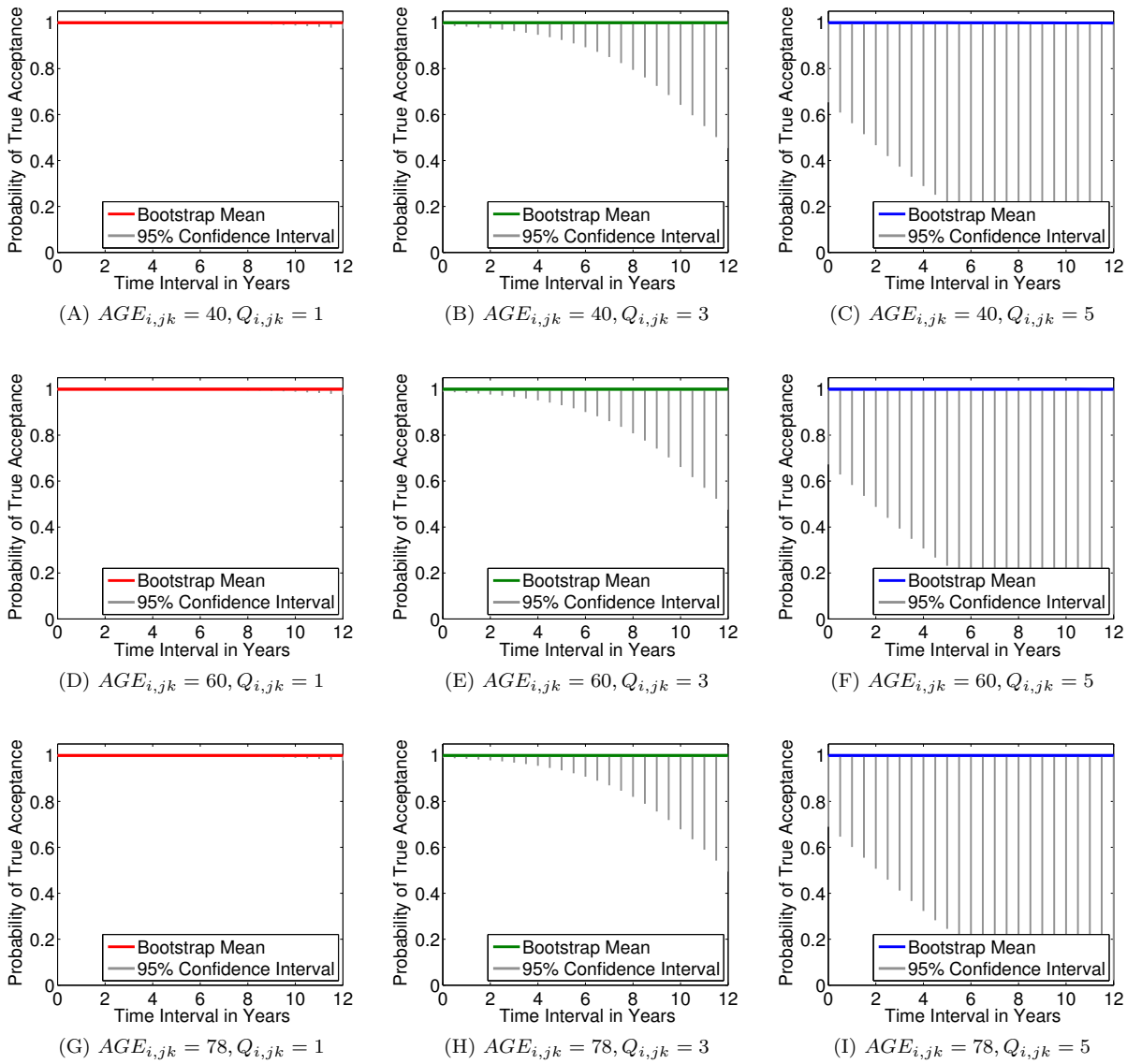
(A) FAR = 0.00001% (B) FAR = 0%

Figure S19: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ (A) when the decision threshold corresponding to FAR of 0.00001% is used, and (B) when the maximum impostor score is used as the decision threshold (corresponding to empirical 0% FAR). Match scores are obtained by COTS-1 matcher when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.



(A) FAR = 0.01% (B) FAR = 0.00001% (C) FAR = 0%

Figure S20: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ (A) when the decision threshold corresponding to FAR of 0.01% is used, (B) when the decision threshold corresponding to FAR of 0.00001% is used, and (C) when the maximum impostor score is used as the decision threshold (corresponding to empirical 0% FAR). Match scores are obtained by COTS-2 matcher when a single finger is used for recognition. The confidence intervals are too small to be visible in (A) and (B).

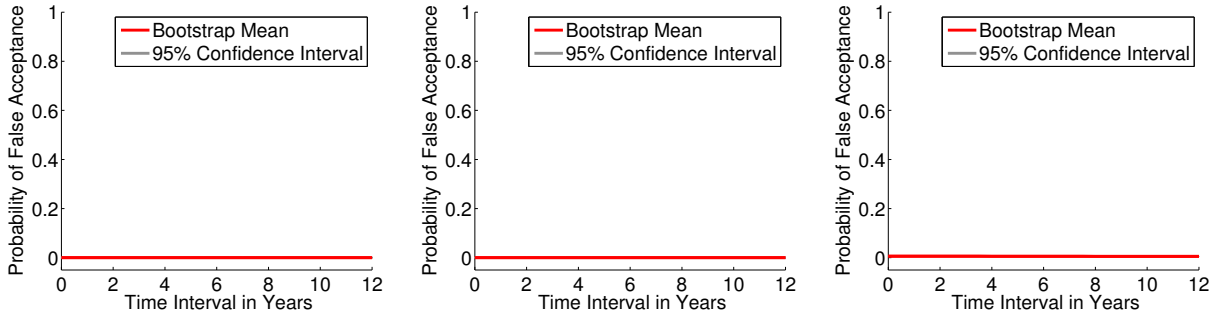


(A) $AGE_{i,jk} = 40, Q_{i,jk} = 1$ (B) $AGE_{i,jk} = 40, Q_{i,jk} = 3$ (C) $AGE_{i,jk} = 40, Q_{i,jk} = 5$

(D) $AGE_{i,jk} = 60, Q_{i,jk} = 1$ (E) $AGE_{i,jk} = 60, Q_{i,jk} = 3$ (F) $AGE_{i,jk} = 60, Q_{i,jk} = 5$

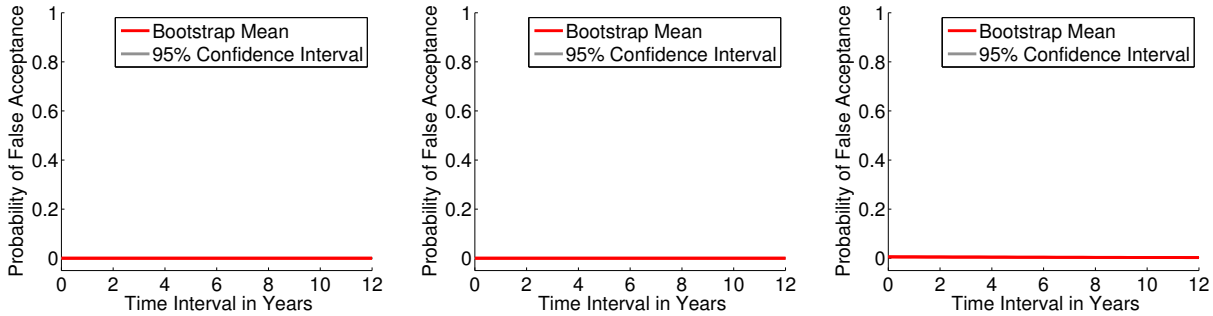
(G) $AGE_{i,jk} = 78, Q_{i,jk} = 1$ (H) $AGE_{i,jk} = 78, Q_{i,jk} = 3$ (I) $AGE_{i,jk} = 78, Q_{i,jk} = 5$

Figure S21: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ when $AGE_{i,jk}$ varies from 40 to 78 and $Q_{i,jk}$ varies from 1 to 5 in model D* for binary decisions on genuine fingerprint pairs. The decision threshold is set to the value corresponding to FAR of 0.00001%. Match scores are obtained by COTS-1 matcher when a single finger is used for recognition.



(A) FAR = 0.01% (B) FAR = 0.1% (C) FAR = 1%

Figure S22: Population-mean trend and 95% confidence interval of probability of false acceptance ($\pi_{ij,k}$) with respect to $\triangle T_{i,1k}$ when the decision threshold corresponding to (A) FAR of 0.01%, (B) FAR of 0.1%, and (C) FAR of 1% is used. Match scores are obtained by COTS-1 matcher when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.

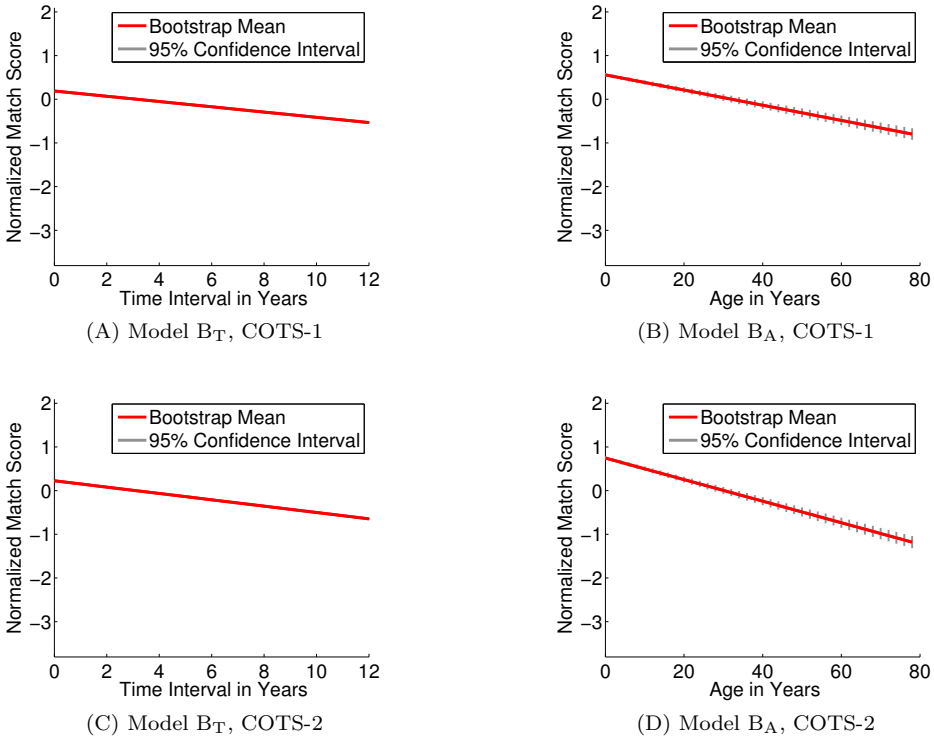


(A) FAR = 0.01% (B) FAR = 0.1% (C) FAR = 1%

Figure S23: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{ij,k}$) with respect to $\triangle T_{i,1k}$ when the decision threshold corresponding to (A) FAR of 0.01%, (B) FAR of 0.1%, and (C) FAR of 1% is used. Match scores are obtained by COTS-2 matcher when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.

Table S6: Parameter estimates and 95% confidence intervals when the genuine match scores from ten fingers obtained by each of two COTS matchers are fused by a sum rule

	Parameters	COTS-1		COTS-2	
		Model B_T	Model B_A	Model B_T	Model B_A
Fixed Effects	β_{00}	0.1896	0.5588	0.2258	0.7461
		(0.1800; 0.1995)	(0.5167; 0.5991)	(0.2159; 0.2360)	(0.7040; 0.7887)
	β_{10}	-0.0603	-0.0174	-0.0726	-0.0247
		(-0.0612; -0.0594)	(-0.0185; -0.0162)	(-0.0736; -0.0717)	(-0.0259; -0.0234)
Variance Components	σ^2_ε	0.4306	0.4266	0.4375	0.4325
	σ^2_0	0.5986	7.3744	0.6599	9.1532
	σ^2_1	0.0037	0.0064	0.0040	0.0078
	σ_{01}	-0.0144	-0.2053	-0.0304	-0.2553



(A) Model B_T, COTS-1

(B) Model B_A, COTS-1

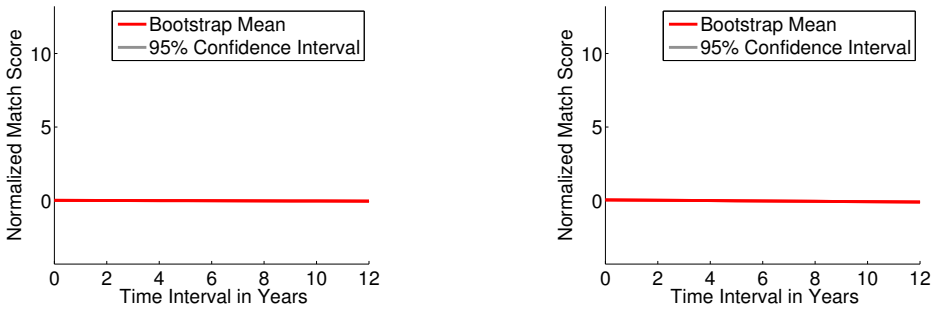
(C) Model B_T, COTS-2

(D) Model B_A, COTS-2

Figure S24: Population-mean trends of genuine match scores obtained by two COTS matchers and 95% confidence intervals with respect to (A) and (C) $\triangle T_{i,jk}$ and (B) and (D) $AGE_{i,jk}$, when the scores from ten fingers are fused. The confidence intervals for model B_T are too small to be visible in the plots.

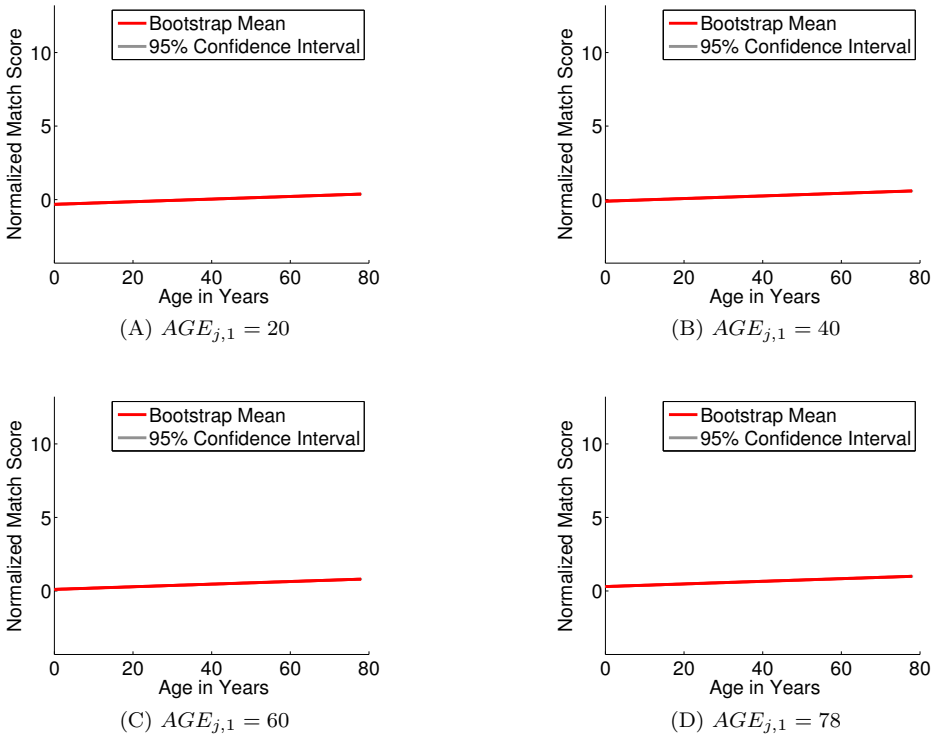
Table S7: Parameter estimates and 95% confidence intervals when the impostor match scores from ten fingers obtained by each of two COTS matchers are fused by a sum rule

	Parameters	COTS-1		COTS-2	
		Model B'_T	Model B'_A	Model B'_T	Model B'_A
Fixed Effects	β_{00}	0.0155	-0.5357	0.0474	0.2029
		(0.0135; 0.0172)	(-0.5495; -0.5226)	(0.0453; 0.0494)	(0.1886; 0.2176)
	β_{10}	-0.0044	0.0089	-0.0122	-0.0031
		(-0.0047; -0.0040)	(0.0086; 0.0093)	(-0.0126; -0.0118)	(-0.0034; -0.0027)
	β_{20}		0.0107		-0.0047
			(0.0103; 0.0111)		(-0.0050; -0.0043)
Variance Components	σ^2_ε	0.7498	0.7453	0.8293	0.8384
	σ^2_0	0.2697	0.3213	0.1829	0.2146
	σ^2_1	0.0005	9.9563e-06	0.0010	1.3781e-05
	σ_{01}	-0.0037	-0.0017	-0.0048	-0.0012



(A) COTS-1 (B) COTS-2

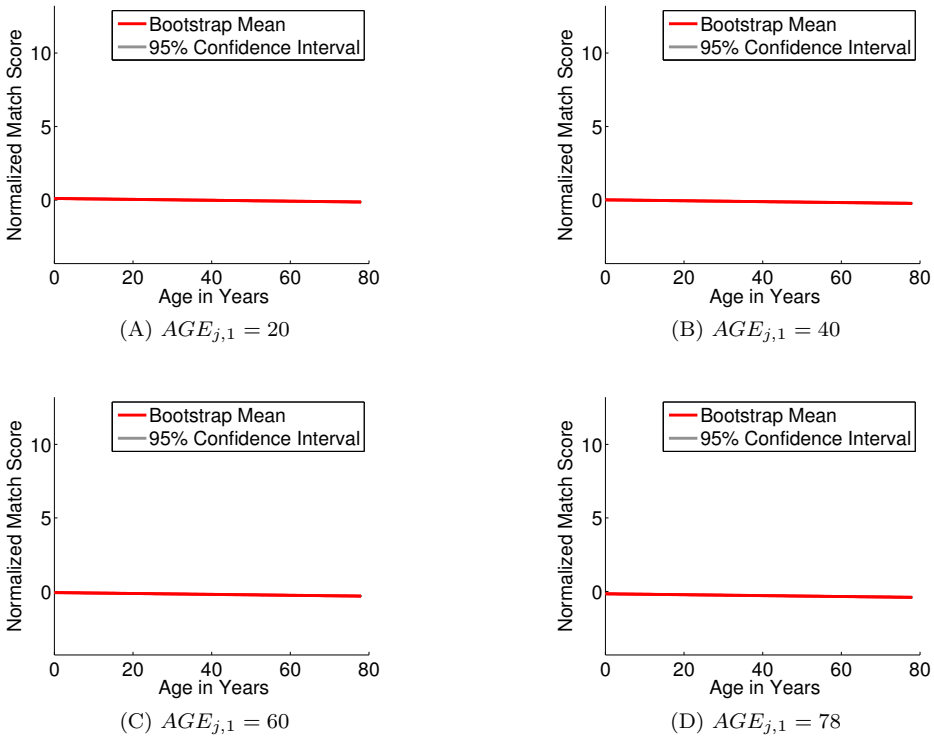
Figure S25: Population-mean trends and 95% confidence intervals of impostor match scores obtained by two COTS matchers with respect to $\triangle T_{i,1k}$ (model B'_{T}), when the scores from ten fingers are fused. The confidence intervals are too small to be visible in the plots.



(A) $AGE_{j,1} = 20$ (B) $AGE_{j,1} = 40$

(C) $AGE_{j,1} = 60$ (D) $AGE_{j,1} = 78$

Figure S26: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-1 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when the scores from ten fingers are fused. The confidence intervals are too small to be visible in the plots.



(A) $AGE_{j,1} = 20$ (B) $AGE_{j,1} = 40$

(C) $AGE_{j,1} = 60$ (D) $AGE_{j,1} = 78$

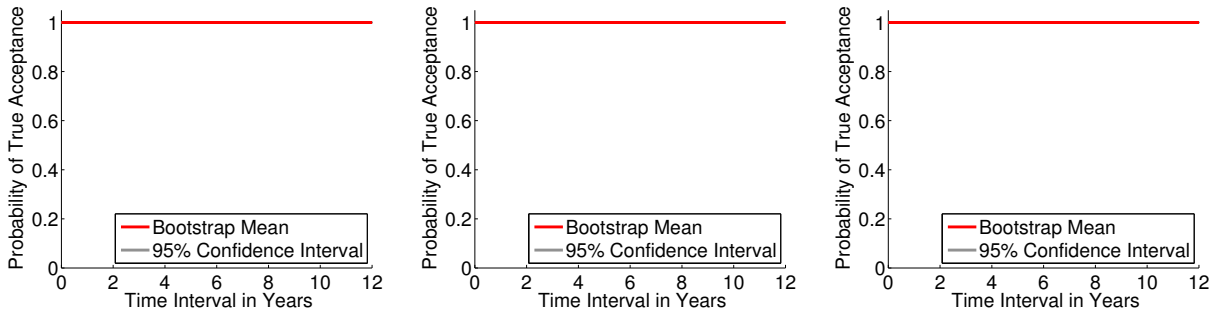
Figure S27: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-2 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when the scores from ten fingers are fused. The confidence intervals are too small to be visible in the plots.

Table S8: Expected probability of true acceptance ($\pi_{i,jk}$) and probability of false acceptance ($\pi_{ij,k}$) and their 95% confidence intervals, when the match scores from all ten fingers are fused by a sum rule

		COTS-1			COTS-2		
	Time Interval in Years	0	6	12	0	6	12
$\pi_{i,jk}$	FAR = 0.01%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)
	FAR = 0.001%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)
	FAR = 0%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)	(1.0000; 1.0000)
$\pi_{ij,k}$	FAR = 0.01%	0.0279e-05	0.0186e-05	0.0124e-05	0.0253e-05	0.0002e-05	0.0000e-05
		(0.0148; 0.1055)	(0.0016; 0.1333)	(0.0002; 0.1684)	(0.0060; 10.1003)	(0.0000; 3.6459)	(0.0000; 1.3160)
	FAR = 0.1%	2.2612e-05	1.7002e-05	1.2784e-05	0.5792e-05	0.1964e-05	0.0666e-05
		(1.3119; 4.5754)	(0.4309; 4.5704)	(0.1415; 4.5654)	(0.4099; 1.2990)	(0.0231; 0.7103)	(0.0013; 0.3884)
	FAR = 1%	0.0047	0.0037	0.0029	0.0037	0.0014	0.0006
		(0.0044; 0.0051)	(0.0032; 0.0043)	(0.0023; 0.0036)	(0.0030; 0.0042)	(0.0010; 0.0025)	(0.0003; 0.0015)

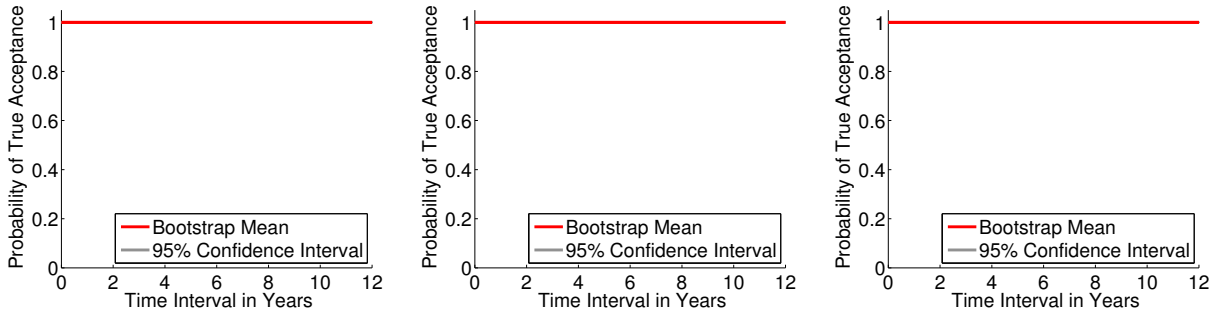
* Numbers in the tables are rounded off to four decimal points unless written in exponential notation. For some very small numbers, round-off to nine decimal points is used.

** If the expected probability is in exponential notation, its confidence interval is also in the same exponential notation.



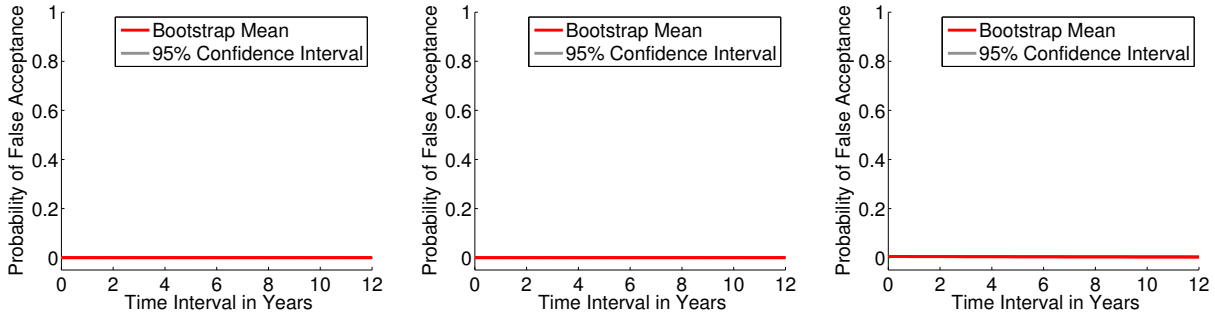
(A) FAR = 0.01% (B) FAR = 0.001% (C) FAR = 0%

Figure S28: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ (A) when the decision threshold corresponding to FAR of 0.01% is used, (B) when the decision threshold corresponding to FAR of 0.001% is used, and (C) when the maximum impostor score is used as the decision threshold (corresponding to empirical 0% FAR). Match scores are obtained by COTS-1 matcher from all ten fingers and fused by a sum rule. The confidence intervals are too small to be visible in the plots.



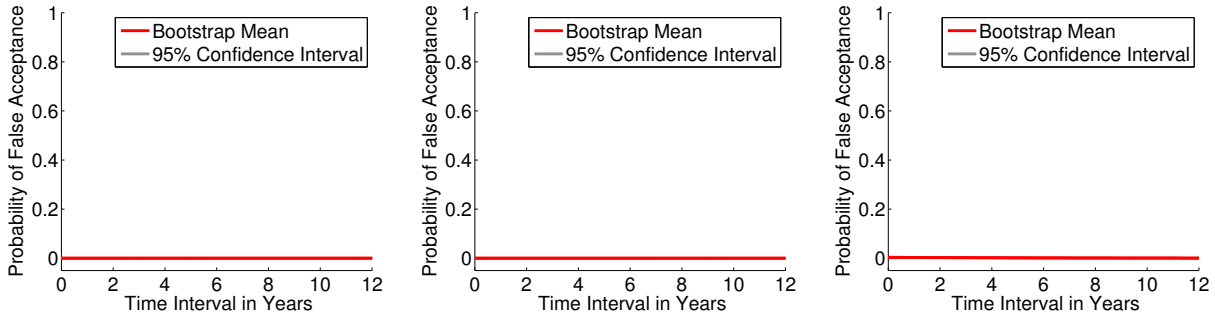
(A) FAR = 0.01% (B) FAR = 0.001% (C) FAR = 0%

Figure S29: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ (A) when the decision threshold corresponding to FAR of 0.01% is used, (B) when the decision threshold corresponding to FAR of 0.001% is used, and (C) when the maximum impostor score is used as the decision threshold (corresponding to empirical 0% FAR). Match scores are obtained by COTS-2 matcher from all ten fingers and fused by a sum rule. The confidence intervals are too small to be visible in the plots.



(A) FAR = 0.01% (B) FAR = 0.1% (C) FAR = 1%

Figure S30: Population-mean trend and 95% confidence interval of probability of false acceptance ($\pi_{ij,k}$) with respect to $\triangle T_{i,1k}$ when the decision threshold corresponding to (A) FAR of 0.01%, (B) FAR of 0.1%, and (C) FAR of 1% is used. Match scores are obtained by COTS-1 matcher from all ten fingers and fused by a sum rule. The confidence intervals are too small to be visible in the plots.



(A) FAR = 0.01% (B) FAR = 0.1% (C) FAR = 0.01%

Figure S31: Population-mean trend and 95% confidence interval of probability of false acceptance ($\pi_{ij,k}$) with respect to $\triangle T_{i,1k}$ when the decision threshold corresponding to (A) FAR of 0.01%, (B) FAR of 0.1%, and (C) FAR of 1% is used. Match scores are obtained by COTS-2 matcher from all ten fingers and fused by a sum rule. The confidence intervals are too small to be visible in the plots.