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.

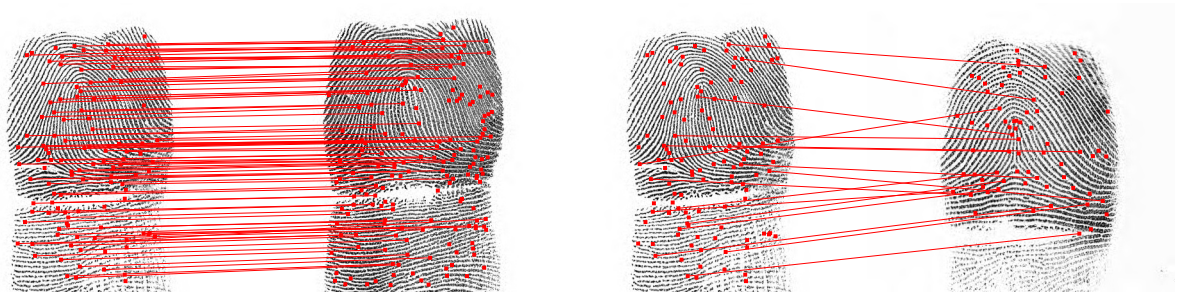
Thus, the modeling and analysis schemes used in this paper cannot be directly applied to latent fingerprints. Further, to the best of our knowledge, there is no latent fingerprint database available for longitudinal study, which contains multiple latent impressions from each finger over time.

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.

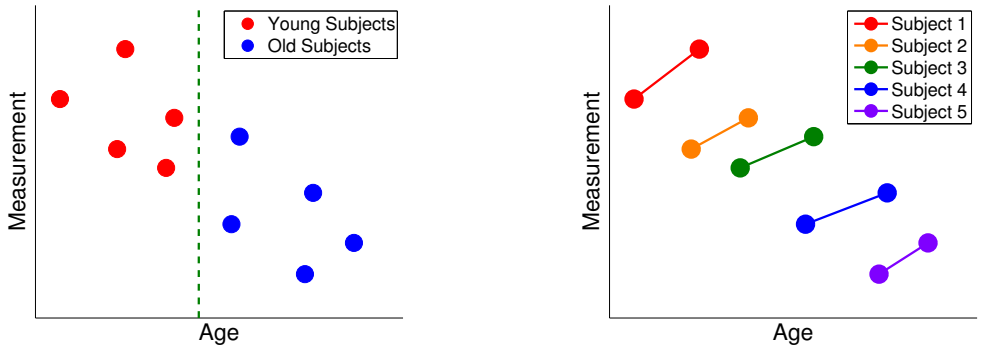
Compound symmetry The compound symmetry requires (i) homoscedasticity of variance: the variance of the measurements at a time instance across all subjects is the same as that of the measurements at another time instance, and (ii) constant covariance: the correlation between the measurements at the first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



(A) (B)

Figure S2: Cross-sectional analysis versus longitudinal analysis of balanced but time-unstructured longitudinal data (adapted from [S10]). For this dataset (two measurements for each of 5 subjects), (A) the cross-sectional analysis that discards subject labels on data makes an inference that *the measurement values tend to decrease with respect to subject's age*, while (B) the longitudinal analysis interprets the data as *the measurement values tend to increase with respect to subject's age*.

commercial matcher shows 1.9% false negative identification rate (FNIR) for single index fingers at a false positive identification rate (FPIR) of 0.1% when searching 100,000 fingerprints, and 0.1% FNIR for ten-finger fusion with rolled prints at the same FPIR when searching 5 million fingerprints. When latent fingerprints are analyzed, human experts are inevitably involved in the latent search procedure to compensate for the limitations of state-of-the-art AFIS in reliable feature extraction and identification [S4].

Improvements in fingerprint acquisition technology have led to the prevalent use of fingerprint recognition in various applications beyond law enforcement and forensics. Fingerprint impressions that were traditionally obtained by smearing fingers with ink and pressing them on paper are now acquired by optical sensors (e.g., at immigration counters in U.S. airports) and solid state sensors (e.g., in smart phones), and these digital images of fingerprints can be readily processed by AFIS.

S4 Persistence Study of Biometrics Traits

Early studies on persistence of fingerprints focused on demonstrating the invariance of ridge structure in fingerprints with respect to time. Herschel collected three fingerprints of his son when he was 7, 17, and 40 years old and verified that all ridge details in the three fingerprints did not change over time [10]. Galton collected 11 pairs of fingerprints from six different individuals at two different time instances [11]. The time interval between a pair of fingerprints in Galton's collection ranged from 11 years to 31 years. The six subjects in his study were selected from different age groups; the age of the subject at the second impression was as young as 15 years and as old as 79 years. Among the 389 minutiae pairs that were manually labeled by Galton, only a single minutia was missing in a fingerprint pair (see Plate 13, Fig. 20 in [11]).

More recently, a number of published studies have claimed template aging—an increase in the error rate in biometric recognition with respect to the time gap between a query and a reference template[1] [14]—for major biometrics modalities, including fingerprint [15], iris [S5, S6], and face [S7]. The question that these studies posed is essentially the following: "does the stored biometric template remain adequate for person authentication over time or should the template be updated to account for possible changes in a person's biometric trait, imaging conditions, or subject's behavior?"

These prior studies [15], [S5, S6, S7] use a method of cross-sectional analysis by grouping the longitudinal data that is unbalanced and time unstructured according to time interval between two acquisitions of a biometric trait and comparing the groups. However, cross-sectional analysis for unbalanced and/or

[1] A biometric template is a compact representation of a subject's biometric data. A template then becomes the reference against which subsequent acquisitions of the subject are compared for authentication.

June 2001 July 2002 April 2003 September 2007 March 2008 October 2008

Figure S3: Six different impressions of the right index finger of a subject in the longitudinal fingerprint dataset used in this study

time-unstructured longitudinal dataset may lead to incorrect inferences from the dataset; Fig. S2 shows a hypothetical example which illustrates that when the dataset is unbalanced and/or time unstructured, cross-sectional analysis makes incorrect inference against the actual longitudinal behavior.

On the other hand, the longitudinal study on iris recognition reported in [S8, S9] uses a nonlinear mixed-effects model which is adequate to analyze the longitudinal data used in the study. They show the relationship between genuine iris match scores and covariates such as time elapsed after enrollment and the difference in iris dilation. However, this study has the following limitations: (i) the dataset used is truncated in the sense that the iris match scores from falsely rejected genuine comparisons were not included, and (ii) the data is collected based on token-less identification. The truncated portion of the data precludes the analysis to determine the tendency of false rejection with respect to time. Also, the occurrence of false acceptances is not clear without the ground truth of image pairings.

S5 Longitudinal Fingerprint Data

The observed responses for Case II (when all ten of a subject's fingers are used) are defined as follows.

- Normalized genuine match score ($Y_{i,jk}$) following a sum-rule fusion:

$$Y_{i,jk} = \frac{S_{i,jk} - \mu}{\sigma}, \text{ where } S_{i,jk} = \sum_{m=1}^{10} s_{i,jk}^{(m)}, \qquad \textbf{[S1]}$$

  $s_{i,jk}^{(m)}$ is the match score between the impressions from finger m in the j-th and k-th tenprint cards of subject i, and μ and σ are the mean and standard deviation of $\{S_{i,jk}\}$, respectively.

- Binary identification decision made on a pair of genuine tenprint cards with fusion score of $S_{i,jk}$ by applying a decision threshold ($\widetilde{Th}$):

$$Y_{i,jk}^{*} = \begin{cases} 1, & \text{if } S_{i,jk} > \widetilde{Th} \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{[S2]}$$

- Normalized impostor fusion score ($Y_{ij,k}$) between the k-th tenprint of subject i and the first tenprint of subject j

- Binary identification decision ($Y_{ij,k}^{*}$) made on an impostor pair of tenprints with fusion match score of $S_{ij,k}$ by applying the decision threshold $\widetilde{Th}$.

S6 Models Investigated in the Study

- Multilevel models for genuine match score analysis are shown in Table 1.

- Multilevel models for impostor match score analysis

  Model B'_T

$$\begin{aligned} &\text{Level-1 Model: } y_{ij,k} = \varphi_{0i} + \varphi_{1i}\triangle T_{i,1k} + \varepsilon_{ij,k}, \\ &\text{Level-2 Model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}. \end{aligned} \quad \textbf{[S3]}$$

  Model B'_A

$$\begin{aligned} &\text{Level-1 Model: } y_{ij,k} = \varphi_{0i} + \varphi_{1i}AGE_{i,k} + \varphi_{2i}AGE_{j,1} + \varepsilon_{ij,k}, \\ &\text{Level-2 Model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}, \varphi_{2i} = \beta_{20}. \end{aligned} \quad \textbf{[S4]}$$

- Multilevel models for binary decisions on genuine fingerprint pairs

  Model B^*_T

$$\begin{aligned} &\text{Level-1 Model: } g(\pi_{i,jk}) = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varepsilon_{i,jk}, \\ &\qquad\qquad\quad y^*_{i,jk} \sim Bin(1, \pi_{i,jk}), \\ &\text{Level-2 Model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}. \end{aligned} \quad \textbf{[S5]}$$

  Model D^*

$$\begin{aligned} &\text{Level-1 Model: } g(\pi_{i,jk}) = \varphi_{0i} + \varphi_{1i}\triangle T_{i,jk} + \varphi_{2i}AGE_{i,jk} + \varphi_{3i}Q_{i,jk} + \varepsilon_{i,jk}, \\ &\qquad\qquad\quad y^*_{i,jk} \sim Bin(1, \pi_{i,jk}), \\ &\text{Level-2 Model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}, \varphi_{2i} = \beta_{20} + b_{2i}, \varphi_{3i} = \beta_{30} + b_{3i}. \end{aligned} \quad \textbf{[S6]}$$

- Multilevel model for binary decisions on impostor fingerprint pairs

  Model $B^{*\prime}_T$

$$\begin{aligned} &\text{Level-1 Model: } g(\pi_{ij,k}) = \varphi_{0i} + \varphi_{1i}\triangle T_{i,1k} + \varepsilon_{ij,k}, \\ &\qquad\qquad\quad y^*_{ij,k} \sim Bin(1, \pi_{ij,k}), \\ &\text{Level-2 Model: } \varphi_{0i} = \beta_{00} + b_{0i}, \varphi_{1i} = \beta_{10} + b_{1i}. \end{aligned} \quad \textbf{[S7]}$$

S7 Validation of Normality Assumptions in Multilevel Model

The multilevel model assumes that the residuals ($\varepsilon_{i,jk}$) and random effects (b_{ri}) follow normal distributions. The inference made based on the model fitting is valid only if the underlying assumptions of the multilevel model are satisfied. The normal probability plot is a way to visually verify the normality of the data. If the normal probability plot is linear, one can ascertain that the data is from a normal distribution. Fig. S5 shows the normal probability plots of $\varepsilon_{i,jk}$, b_{0i}, and b_{1i} when model B_T is fit to the genuine match scores obtained from the two COTS matchers.

While the residuals generally follow normal distributions, significant departures from normality are observed at the tails for the scores output by both the matchers. A possible cause of non-normality at the tails is that the scores from the COTS fingerprint matchers are typically censored, i.e., very low or high match scores are trimmed so that the output scores are in a finite range.

When the model assumptions are violated, the parameter estimates for fixed and random effects tend to be still reliable while the standard errors (consequently, confidence intervals) tend to be underestimated [S11]. In this case, bootstrapping is a useful way to estimate parameters and confidence intervals [S12].

Table S1: Goodness-of-fit of the models for genuine match scores shown in Table 1

Model	COTS-1			COTS-2		
	Deviance	AIC	BIC	Deviance	AIC	BIC
Model A	1,114,948	1,114,954	1,114,988	1,142,532	1,142,538	1,142,571
Model B_T	1,099,980	1,099,992	1,100,058	1,115,191	1,115,203	1,115,269
Model B_A	1,100,979	1,100,991	1,101,057	1,120,911	1,120,923	1,120,990
Model B_Q	1,028,899	1,028,911	1,028,978	1,060,037	1,060,049	1,060,115
Model C_G	1,099,969	1,099,985	1,100,074	1,115,117	1,115,133	1,115,222
Model C_R	1,099,817	1,099,833	1,099,921	1,114,378	1,114,394	1,114,483
Model D	1,003,908	1,003,938	1,004,105	1,019,412	1,019,442	1,019,608
Model E	1,003,839	1,003,873	1,004,062	1,018,986	1,019,020	1,019,209

S8 Genuine Match Score Analysis with Models D and E

In models D and E, the fixed-effects parameter estimates for $\triangle T_{i,jk}$ (β_{10}), $AGE_{i,jk}$ (β_{20}), and $Q_{i,jk}$ (β_{30}) remain negative, similar to models B_T, B_A, and B_Q. The correlations between any two covariates can be calculated from the estimated covariance matrices in models D and E. In particular, we are interested in (i) σ_{13} which gives the correlation between $\triangle T_{i,jk}$ and $Q_{i,jk}$ and (ii) σ_{23} which gives the correlation between $AGE_{i,jk}$ and $Q_{i,jk}$. Although the estimated values for σ_{13} and σ_{23} are negative, the correlations among the covariates are very small—in model D, the correlation coefficients for σ_{13} and σ_{23} based on COTS-1 match scores are -0.0324 and -0.0464; for COTS-2 matcher, they are -0.0174 and -0.1035. Moreover, σ_{13} in models D and E with COTS-2 matcher cannot be claimed to be significantly different from 0 since the null hypothesis $\sigma_{13} = 0$ is not rejected at a significance level of 0.05.

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 match scores is assessed by comparing the population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ at different values of $AGE_{i,jk}$ and $Q_{i,jk}$ in models D and E (see Fig. S7). As the 95% confidence intervals in models D and E are overlapped, it cannot be said that these interactions significantly affect the variation in genuine match scores with respect to $\triangle T_{i,jk}$.

The temporal trend of genuine match scores is analyzed by fixing one of the covariates $AGE_{i,jk}$ and $Q_{i,jk}$ in model E (see Figs. S8 and S9). In Fig. S8, the population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ for each subject's age group ($AGE_{i,jk}$ is (A) 20, (B) 40, (C) 60, and (D) 78) are shown. For the age group of 20, the population-mean trends of fingerprint comparisons grouped by the NFIQ ($Q_{i,jk}$ is 1, 3, or 5) are well separated at a significance level of 0.05. However, for subjects at older ages, the 95% confidence intervals of the fingerprint quality groups become overlapped. On the other hand, the population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ for each fingerprint quality group ($Q_{i,jk}$ is (A) 1, (B) 3, and (C) 5) are shown in Fig. S9. At any level of fingerprint quality, the impact of subject's age is not significant on genuine match scores since the 95% confidence intervals of all age groups are completely overlapped.

S9 Assessment of Goodness-of-Fit

The details of the goodness-of-fit measures used in this study are as follows.

- Deviance (D): Deviance can be used to compare the goodness-of-fit of nested models. The nested property is easily determined by checking if one model becomes equivalent to the other by setting the coefficients for some of the covariates to zero. For example, whereas models A and B_T are nested and models A and B_Q are nested, models B_T and B_Q are not nested. The deviance is defined as:

$$D = -2\log(L), \qquad \textbf{[S8]}$$

  where L is the maximum value of the likelihood function for the model.

- Akaike Information Criterion (AIC): AIC can be used for any model comparison task (models do not need to be nested). AIC is defined as:

$$AIC = 2k - 2\log(L), \qquad \textbf{[S9]}$$

where k is the number of parameters in the model, and L is the maximum value of the likelihood function for the model.

- Bayesian Information Criterion (BIC): Under the assumption that the data distribution is in the exponential family, BIC is defined as:

$$BIC = k\log(n) - 2\log(L), \qquad \textbf{[S10]}$$

where k is the number of parameters in the model, n is the number of data points, and L is the maximum value of the likelihood function for the model. BIC also can be used for comparisons of non-nested models.

References

S1. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) *Handbook of Fingerprint Recognition* (Second Edition), Springer-Verlag.

S2. Cole SA (2001) *Suspect Identities: A History of Fingerprinting and Criminal Identification*, Harvard University Press.

S3. Watson C, et. al. (2014) Fingerprint vendor technology evaluation. *National Institute of Standards and Technology Interagency Report 8034.*

S4. Indovina M, Dvornychenko V, Hicklin RA, Kiebuzinski GI (2012) Evaluation of latent fingerprint technologies: extended feature sets [evaluation #2]. *National Institute of Standards and Technology Interagency Report 7859.*

S5. Fenker SP, Bowyer KW (2011) Experimental evidence of a template aging effect in iris biometrics. *IEEE Workshop on Applications of Computer Vision*, pp. 232–239.

S6. Fenker SP, Bowyer KW (2012) Analysis of template aging in iris biometrics. *IEEE Computer Vision and Pattern Recognition Workshop on Biometrics*, pp. 45–51.

S7. Lanitis A (2010) A survey of the effects of aging on biometric identity verification. *International Journal of Biometrics*, 2(1):34–52.

S8. Grother P, Matey JR, Tabassi E, Quinn GW, Chumakov M (2013) IREX VI: temporal stability of iris recognition accuracy. *National Institute of Standards and Technology Interagency Report 7948.*

S9. Grother P, Matey JR, Quinn GW, Tabassi E (2014) Iris permanence: what we know, what we don't, and how to find out more. *Global Identity Summit.*

S10. Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data*, Oxford University Press.

S11. Maas CJM, Hox JJ (2004) The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis*, 46(3):427–440.

S12. Leeden R, Busing FMTA, Meijer E (1997) Bootstrap methods for two-level models. *Multilevel Conference.*

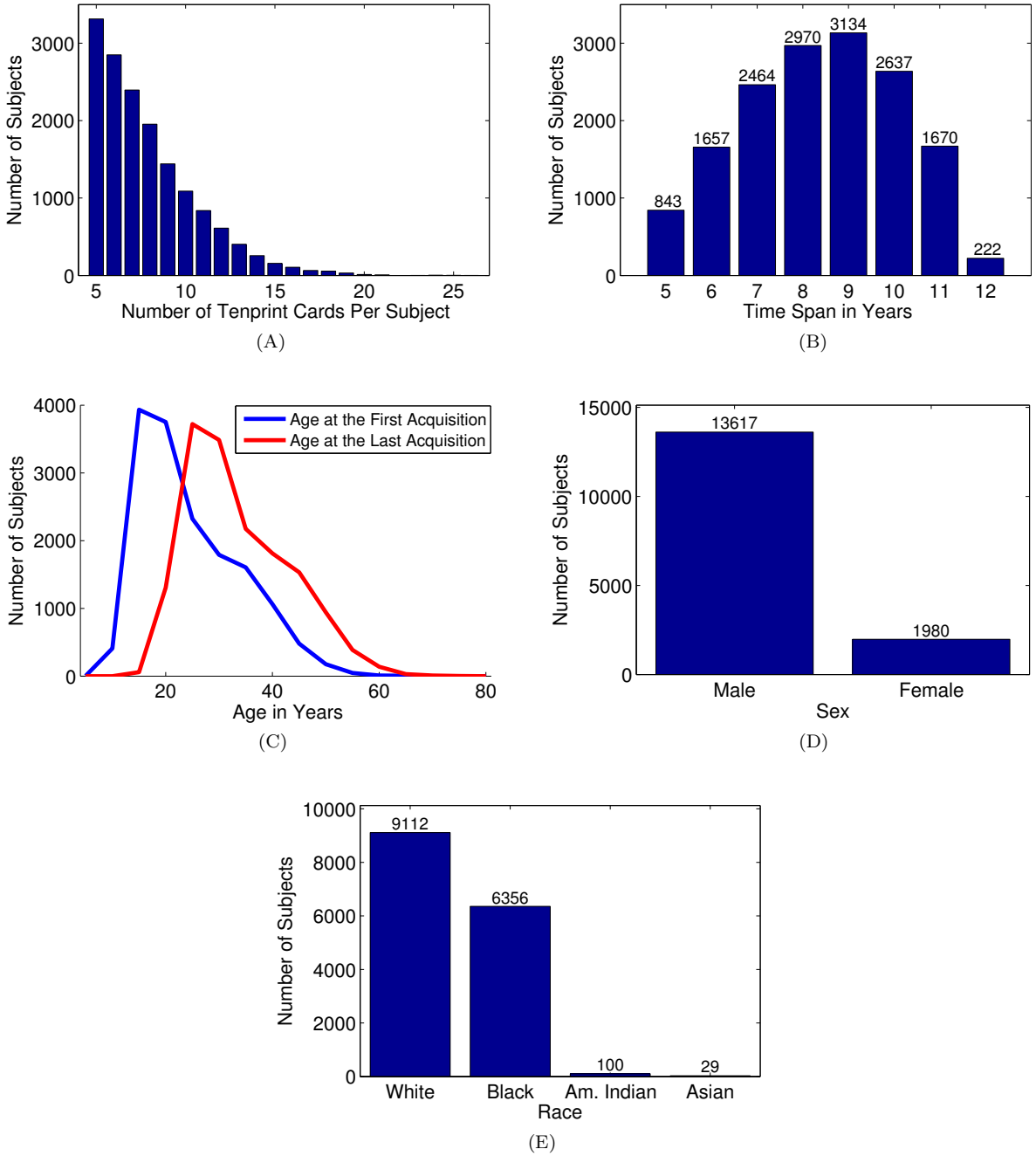
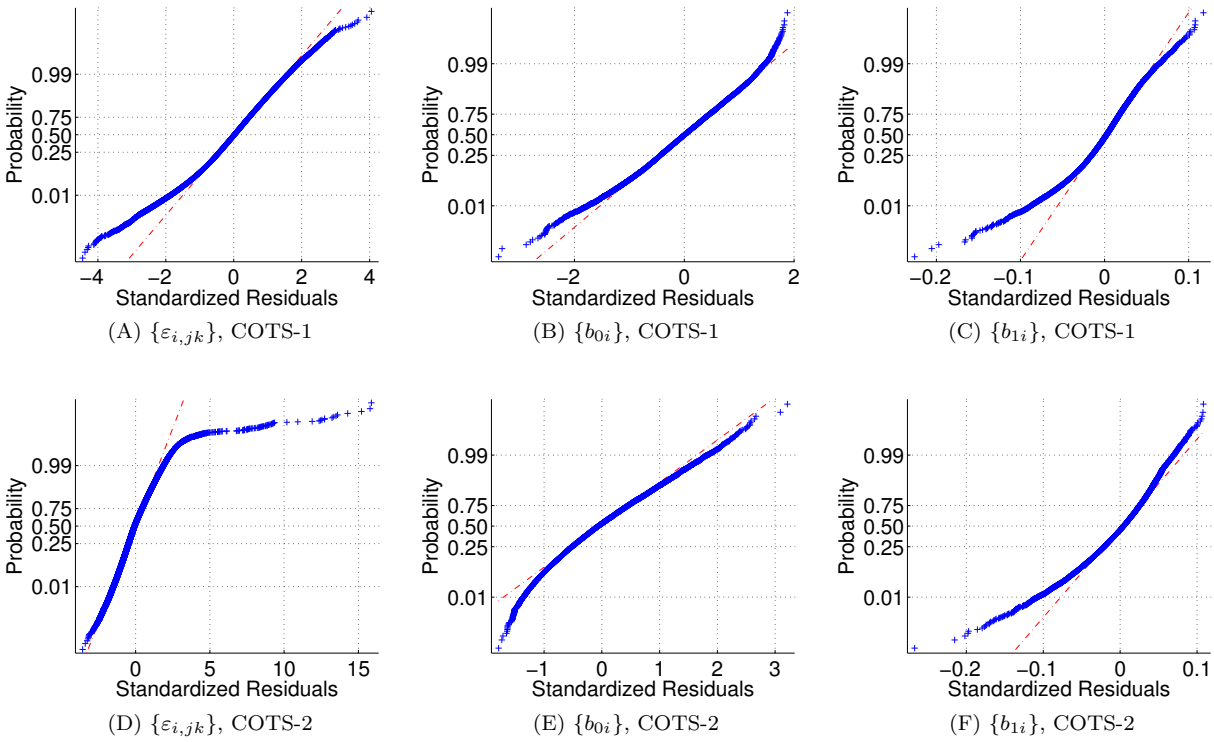


Figure S4: Statistics of the longitudinal fingerprint dataset used in this study. Histograms of (A) the number of tenprint cards per subject, (B) time span of data collection for a subject, (C) age at the first and last tenprint acquisitions, (D) sex, and (E) race.



(A) $\{\varepsilon_{i,jk}\}$, COTS-1 (B) $\{b_{0i}\}$, COTS-1 (C) $\{b_{1i}\}$, COTS-1

(D) $\{\varepsilon_{i,jk}\}$, COTS-2 (E) $\{b_{0i}\}$, COTS-2 (F) $\{b_{1i}\}$, COTS-2

Figure S5: Normal probability plots of (A) and (D) residuals at level 1 ($\varepsilon_{i,jk}$), (B) and (E) random-effects for intercept at level 2 (b_{0i}), and (C) and (F) random-effects for slope at level 2 (b_{1i}) of model $\mathrm{B_T}$ fit to the genuine match scores obtained from the two COTS matchers.

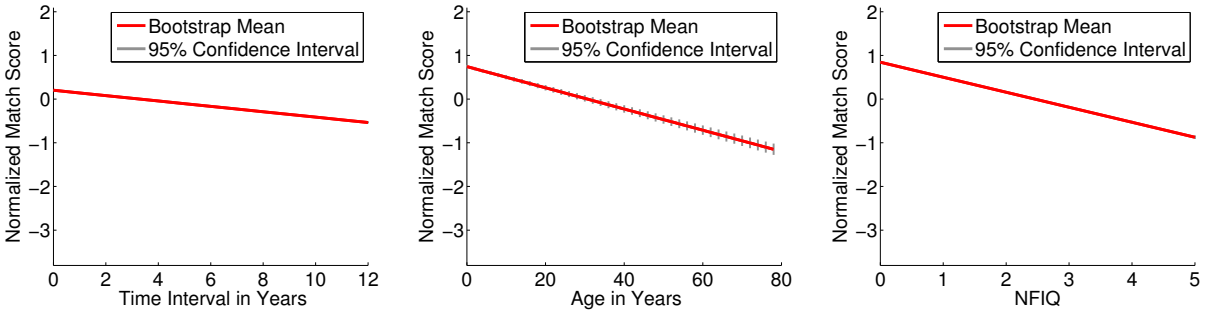
Table S2: Parameter estimates and 95% confidence intervals of genuine match scores obtained by COTS-1 matcher when a single finger (right index) is used for recognition

	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^2_ε	0.4980	0.4897	0.4210	0.3640	0.3639
	σ^2_0	0.5298	5.6056	0.9096	6.6323	6.6565
	σ^2_1	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^2_3				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

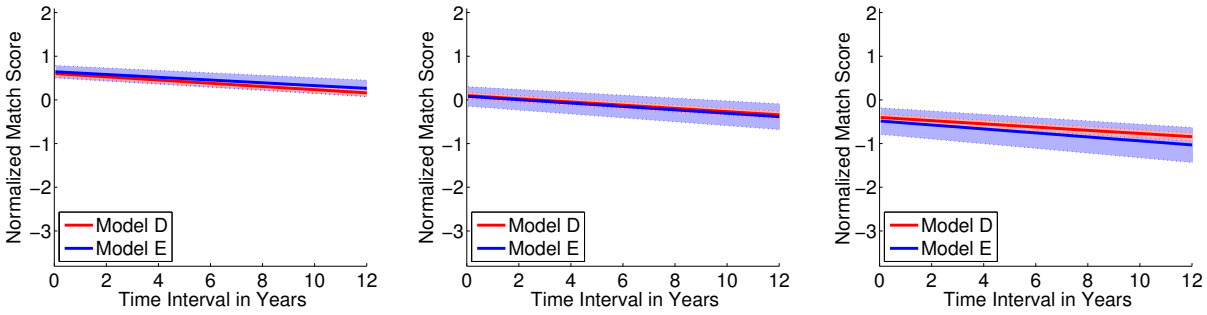
	Parameters	Model B_T	Model B_A	Model B_Q	Model D	Model E
Fixed Effects	β_{00}	0.2032	0.7447	0.8456	1.0353	1.4398
		(0.1939; 0.2127)	(0.7072; 0.7843)	(0.8316; 0.8595)	(0.9946; 1.0750)	(1.3706; 1.5102)
	β_{10}	-0.0616	-0.0243	-0.3439	-0.0533	-0.0654
		(-0.0625; -0.0606)	(-0.0254; -0.0231)	(-0.3489; -0.3385)	(-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^2_ε	0.5162	0.5070	0.4540	0.3755	0.3751
	σ^2_0	0.5744	7.5573	0.9105	7.8990	7.8357
	σ^2_1	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^2_3				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.

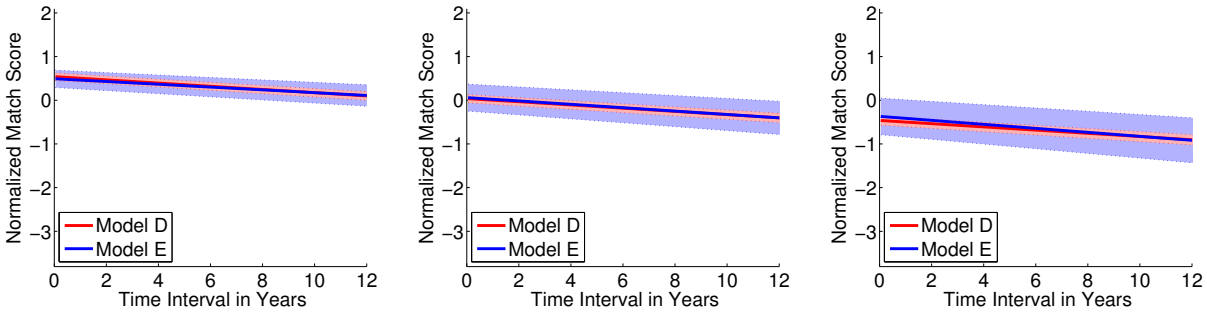


(A) Model B_T (B) Model B_A (C) Model B_Q

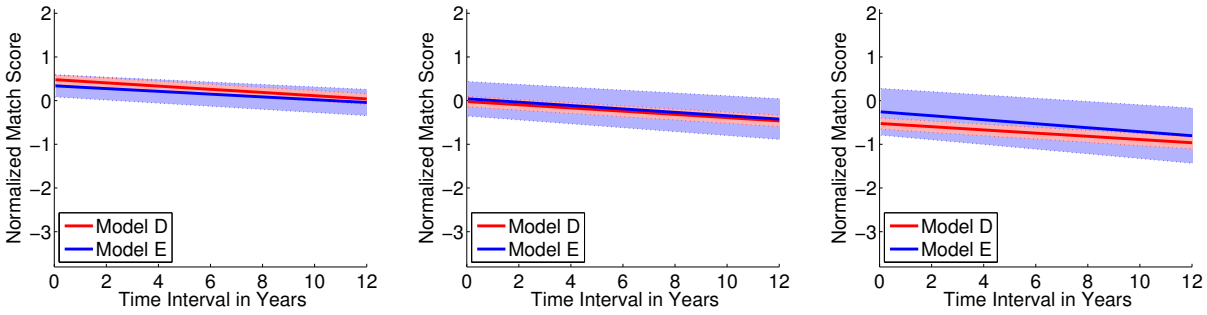
Figure S6: Population-mean trends of genuine match scores obtained by COTS-2 matcher and 95% confidence intervals with respect to (A) $\triangle T_{i,jk}$, (B) $AGE_{i,jk}$, and (C) $Q_{i,jk}$, when a single finger is used for recognition. The confidence intervals for models B_T and B_Q are too tight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



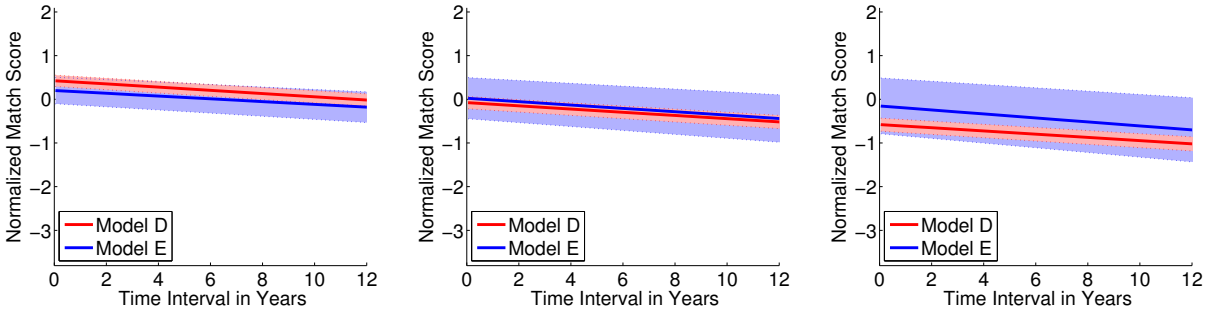
(A) $AGE_{i,jk} = 20$, $Q_{i,jk} = 1$ (B) $AGE_{i,jk} = 20$, $Q_{i,jk} = 3$ (C) $AGE_{i,jk} = 20$, $Q_{i,jk} = 5$



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

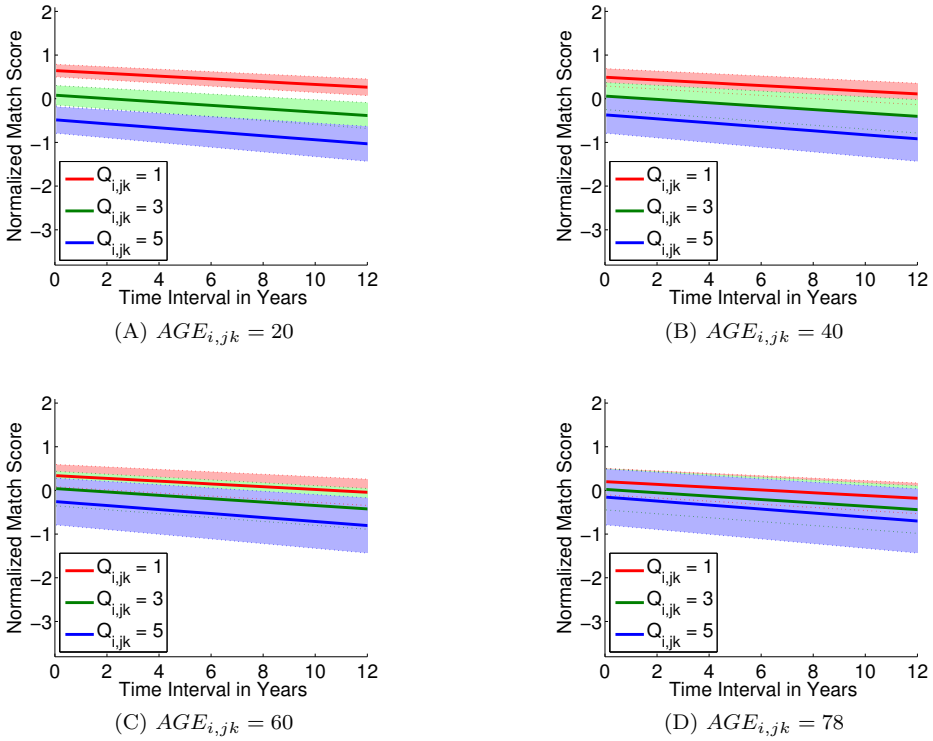


(G) $AGE_{i,jk} = 60$, $Q_{i,jk} = 1$ (H) $AGE_{i,jk} = 60$, $Q_{i,jk} = 3$ (I) $AGE_{i,jk} = 60$, $Q_{i,jk} = 5$



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

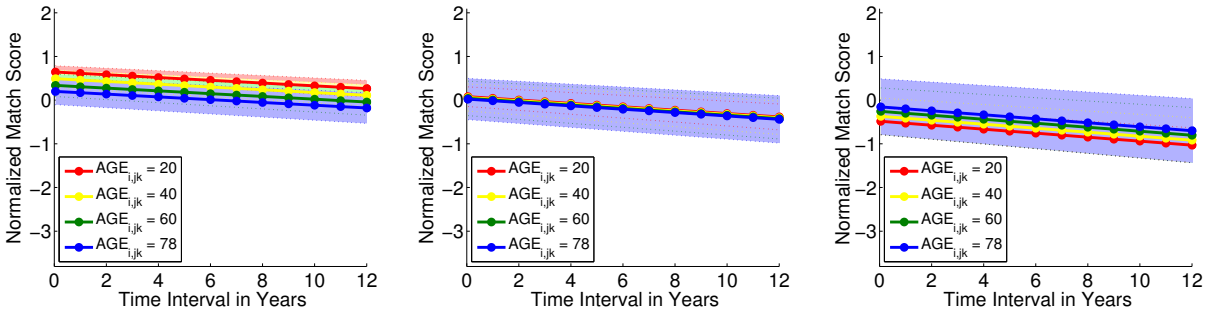
Figure S7: Comparison between models D and E. Population-mean trends of genuine match scores with respect to $\triangle T_{i,jk}$ are shown when $AGE_{i,jk}$ varies from 20 to 78 and $Q_{i,jk}$ varies from 1 to 5 in models D and E. Solid lines are the bootstrap means, and the shaded areas represent the 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) $AGE_{i,jk} = 40$

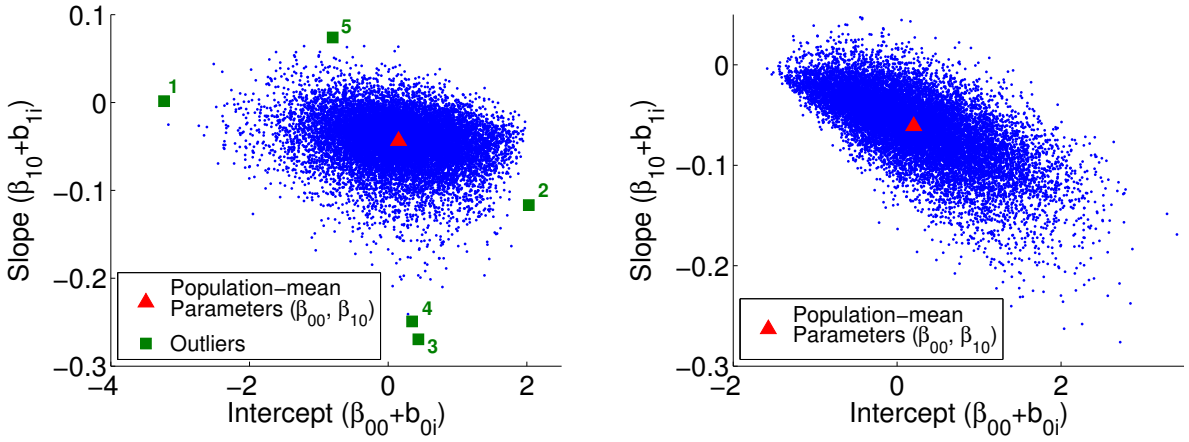
(C) $AGE_{i,jk} = 60$ (D) $AGE_{i,jk} = 78$

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.



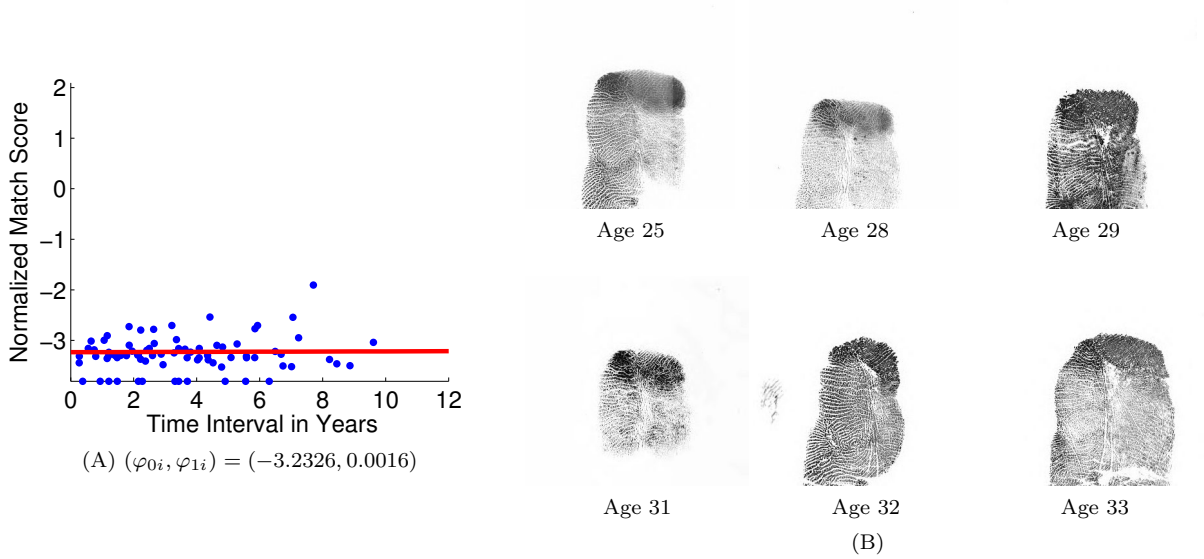
(A) $Q_{i,jk} = 1$ (B) $Q_{i,jk} = 3$ (C) $Q_{i,jk} = 5$

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.



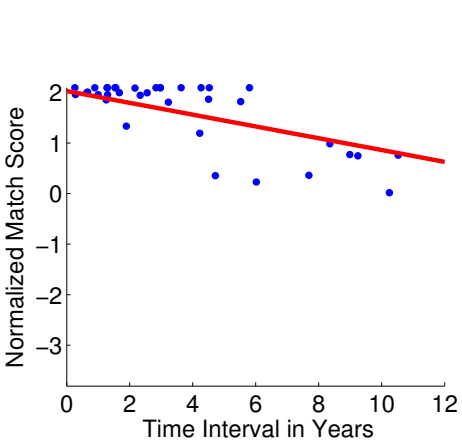
(A) COTS-1 (B) COTS-2

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).

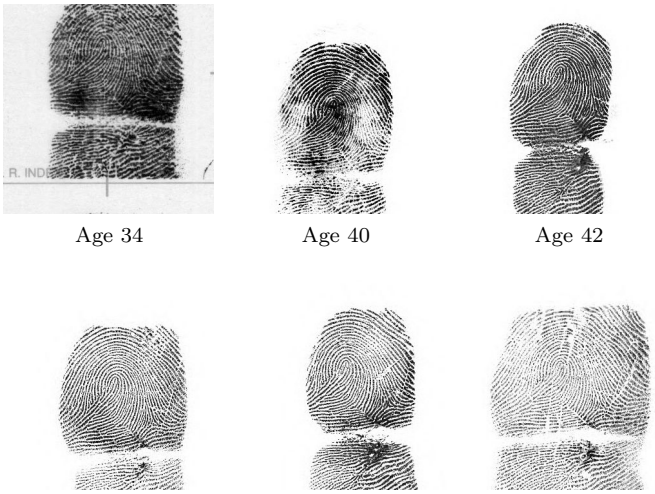


(A) $(\varphi_{0i}, \varphi_{1i}) = (-3.2326, 0.0016)$ (B)

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

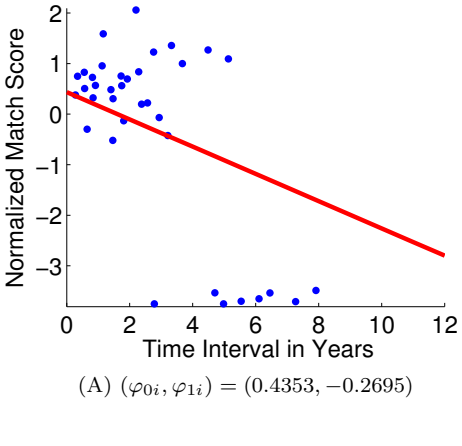


Age 34 Age 40 Age 42

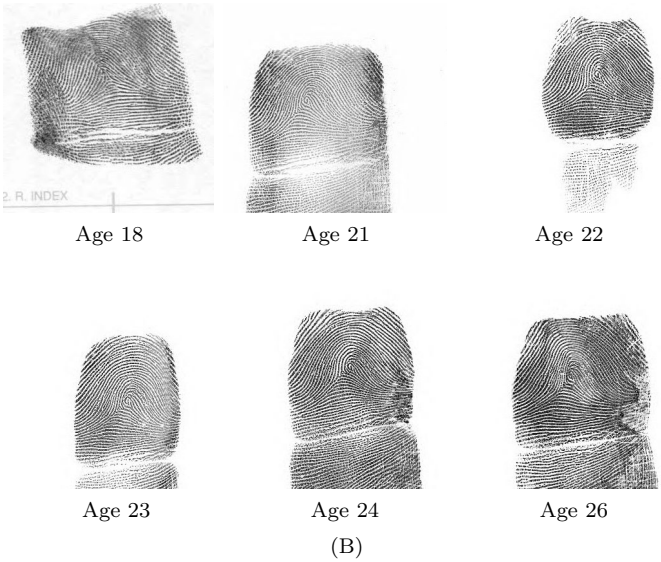
Age 43 Age 44 Age 45

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

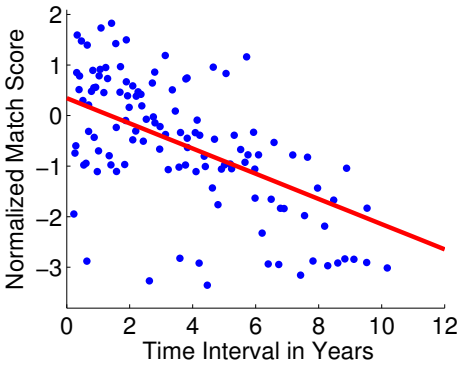


Age 18 Age 21 Age 22

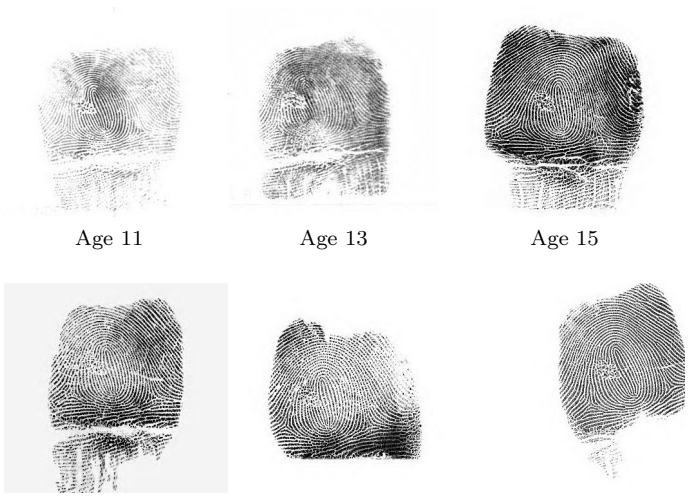
Age 23 Age 24 Age 26

(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.

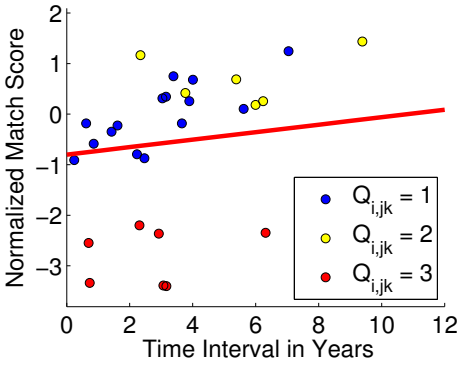


(A) $(\varphi_{0i}, \varphi_{1i}) = (0.3443, -0.2491)$

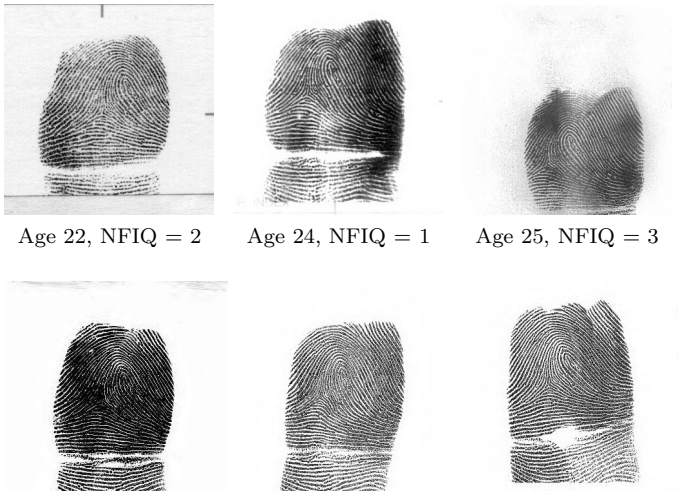


(B)

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



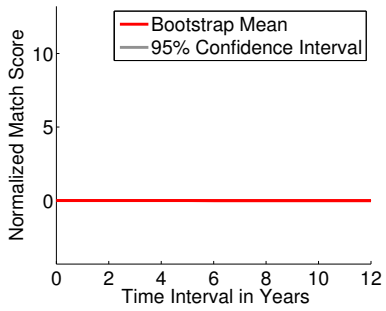
(B)

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.

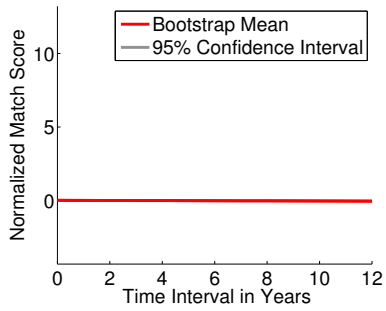
Table S4: Parameter estimates and 95% confidence intervals of impostor match scores obtained by two COTS matchers when a single finger is used for recognition

	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.

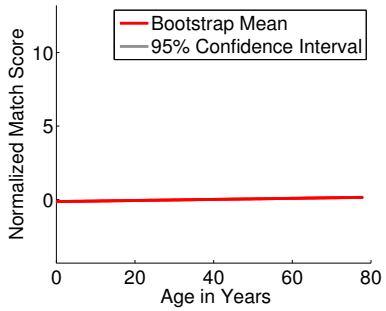


(A) COTS-1

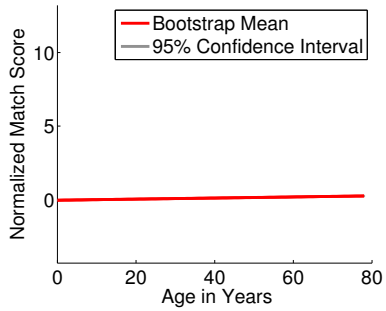


(B) COTS-2

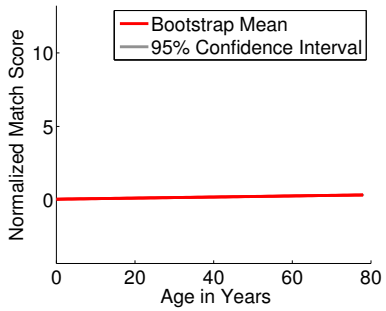
Figure S16: Population-mean trends and 95% confidence intervals of impostor match scores obtained by two COTS matchers with respect to $\triangle T_{i,1k}$ (model B'_{T}), when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.



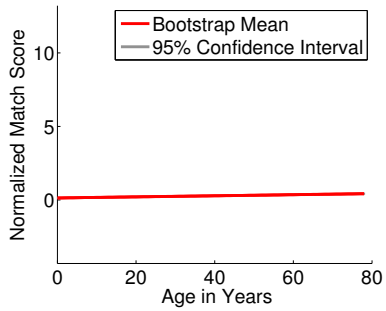
(A) $AGE_{j,1} = 20$



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

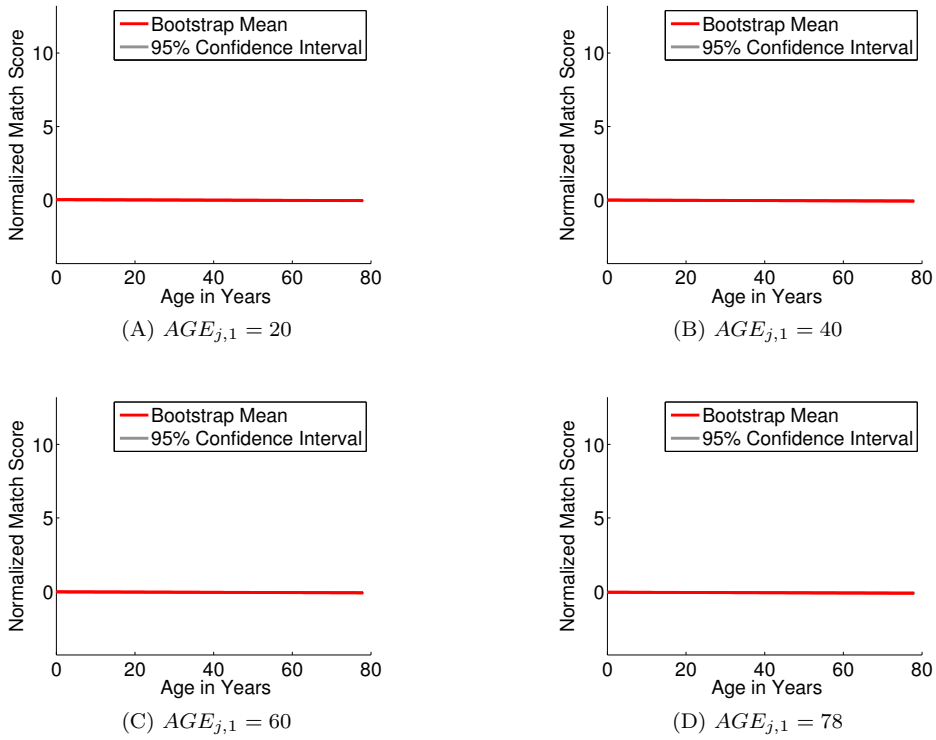


(C) $AGE_{j,1} = 60$



(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) $AGE_{j,1} = 78$

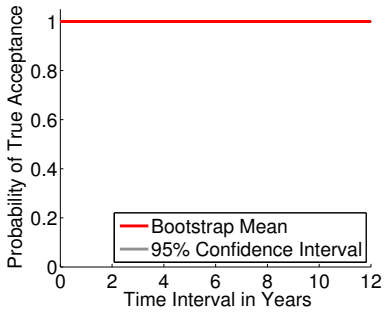
Figure S18: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-2 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when a single finger is used for recognition. The confidence intervals are too small to be visible in the plots.

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

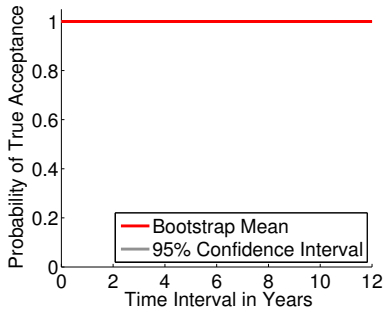
		COTS-1			COTS-2		
	Time Interval in Years	0	6	12	0	6	12
$\pi_{i,jk}$	FAR = 0.01%	1.0000 (1.0000; 1.0000)	1.0000 (1.0000; 1.0000)	1.0000 (1.0000; 1.0000)	1.0000 (1.0000; 1.0000)	0.9999 (0.9999; 1.0000)	0.9999 (0.9998; 0.9999)
	FAR = 0.00001%	1.0000 (1.0000; 1.0000)	1.0000 (0.9999; 1.0000)	0.9999 (0.9999;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

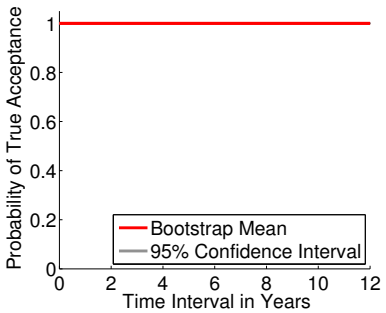


(A) FAR = 0.00001%

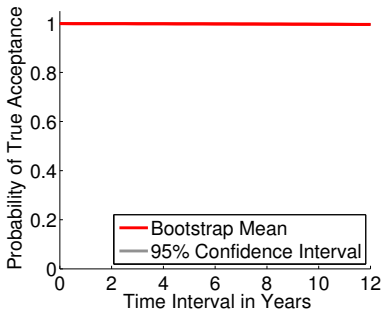


(B) FAR = 0%

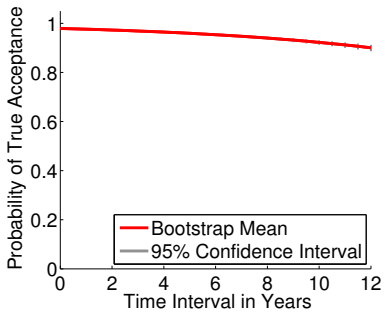
Figure S19: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{i,jk}$) with respect to $\triangle T_{i,jk}$ (A) when the decision threshold corresponding to FAR of 0.00001% is used, and (B) when the maximum impostor score is used as the decision threshold (corresponding 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%

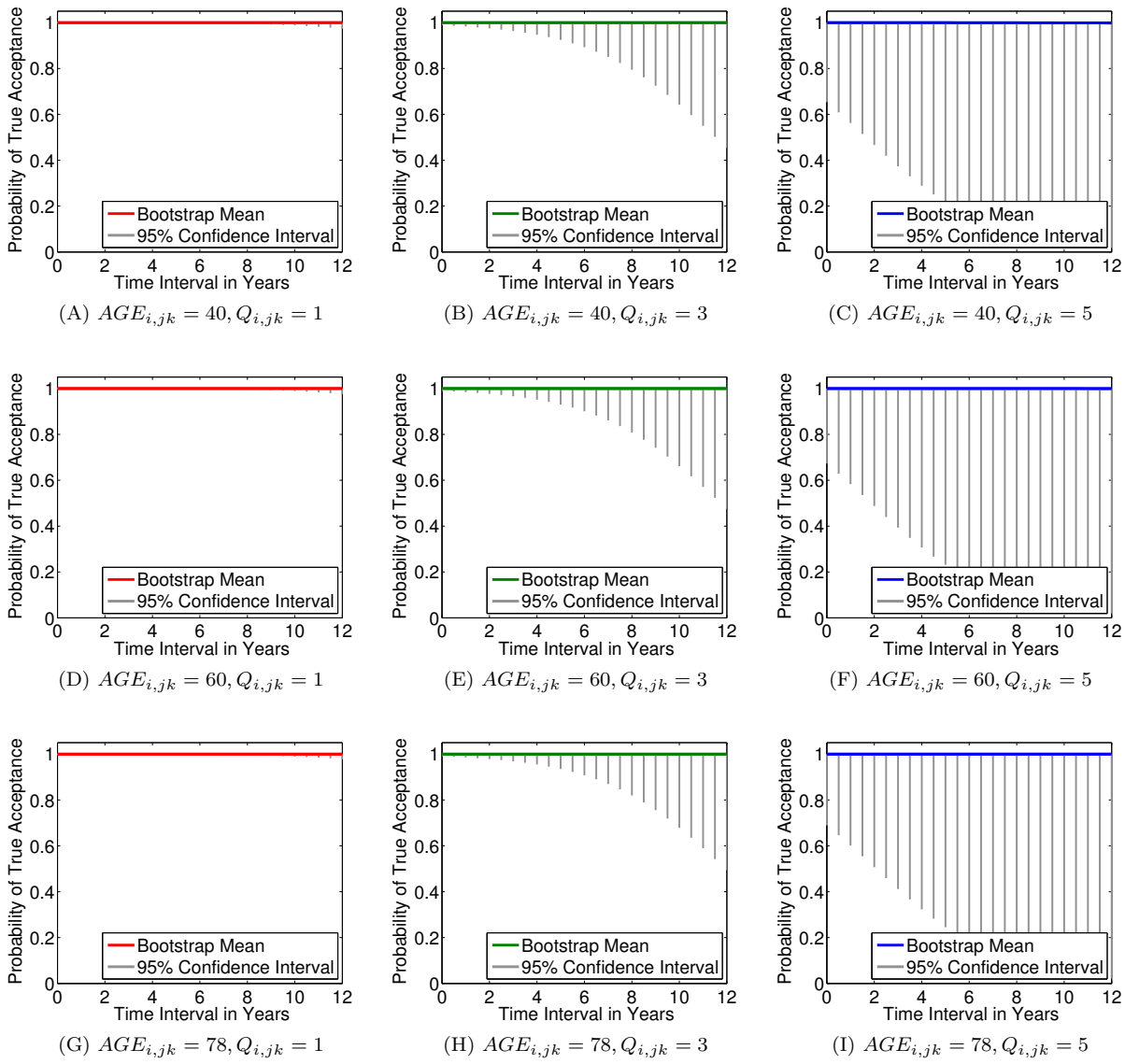


(B) FAR = 0.00001%



(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).

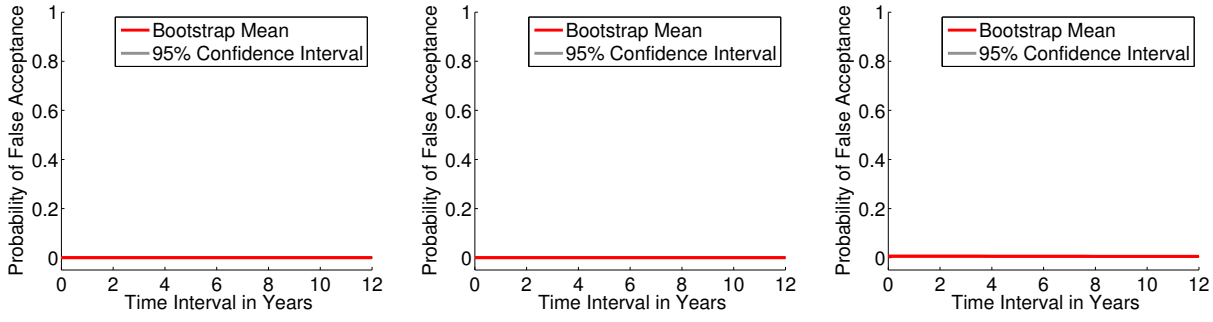


(A) $AGE_{i,jk} = 40, Q_{i,jk} = 1$ (B) $AGE_{i,jk} = 40, Q_{i,jk} = 3$ (C) $AGE_{i,jk} = 40, Q_{i,jk} = 5$

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

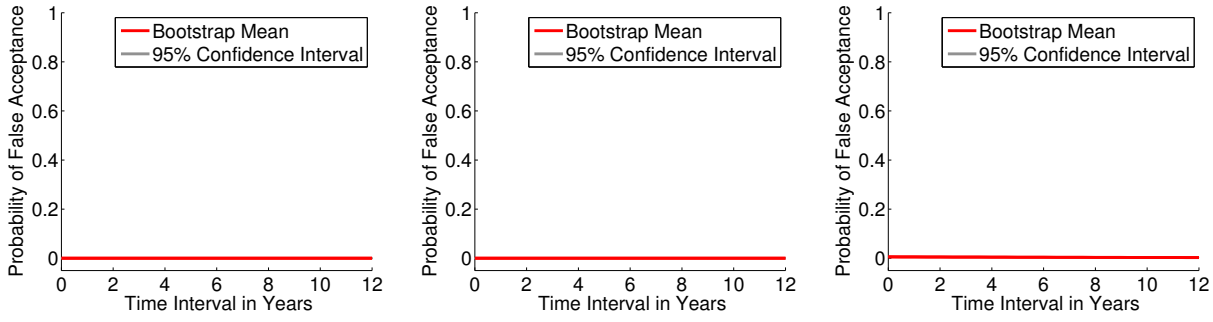
(G) $AGE_{i,jk} = 78, Q_{i,jk} = 1$ (H) $AGE_{i,jk} = 78, Q_{i,jk} = 3$ (I) $AGE_{i,jk} = 78, Q_{i,jk} = 5$

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.

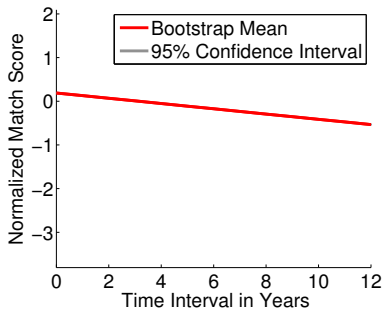


(A) FAR = 0.01% (B) FAR = 0.1% (C) FAR = 1%

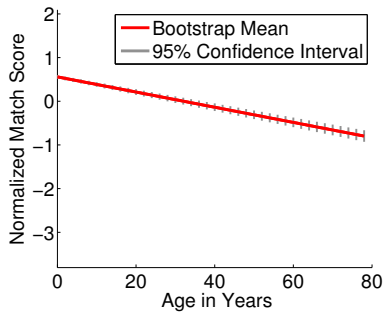
Figure S23: Population-mean trend and 95% confidence interval of probability of true acceptance ($\pi_{ij,k}$) with respect to $\triangle T_{i,1k}$ when the decision threshold corresponding to (A) FAR of 0.01%, (B) FAR of 0.1%, and (C) FAR of 1% is used. Match scores are obtained by COTS-2 matcher when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

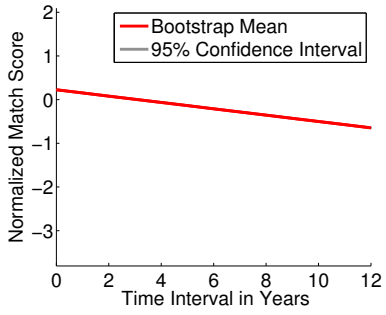
	Parameters	COTS-1		COTS-2	
		Model B_T	Model B_A	Model B_T	Model B_A
Fixed Effects	β_{00}	0.1896 (0.1800; 0.1995)	0.5588 (0.5167; 0.5991)	0.2258 (0.2159; 0.2360)	0.7461 (0.7040; 0.7887)
	β_{10}	-0.0603 (-0.0612; -0.0594)	-0.0174 (-0.0185; -0.0162)	-0.0726 (-0.0736; -0.0717)	-0.0247 (-0.0259; -0.0234)
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



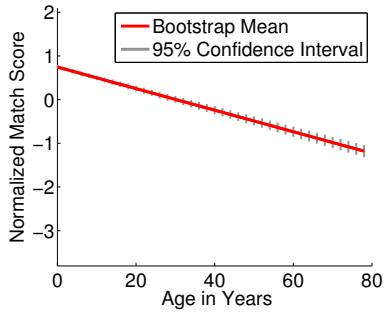
(A) Model B_T, COTS-1



(B) Model B_A, COTS-1



(C) Model B_T, COTS-2

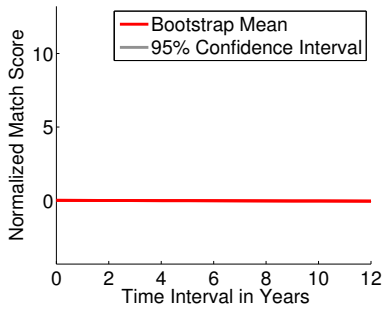


(D) Model B_A, COTS-2

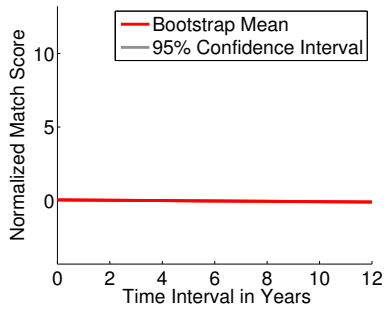
Figure S24: Population-mean trends of genuine match scores obtained by two COTS matchers and 95% confidence intervals with respect to (A) and (C) $\triangle T_{i,jk}$ and (B) and (D) $AGE_{i,jk}$, when the scores from ten fingers are fused. The confidence intervals for model B_T are too small to be visible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

		COTS-1		COTS-2	
	Parameters	Model B'_T	Model B'_A	Model B'_T	Model B'_A
Fixed Effects	β_{00}	0.0155 (0.0135; 0.0172)	-0.5357 (-0.5495; -0.5226)	0.0474 (0.0453; 0.0494)	0.2029 (0.1886; 0.2176)
	β_{10}	-0.0044 (-0.0047; -0.0040)	0.0089 (0.0086; 0.0093)	-0.0122 (-0.0126; -0.0118)	-0.0031 (-0.0034; -0.0027)
	β_{20}		0.0107 (0.0103; 0.0111)		-0.0047 (-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

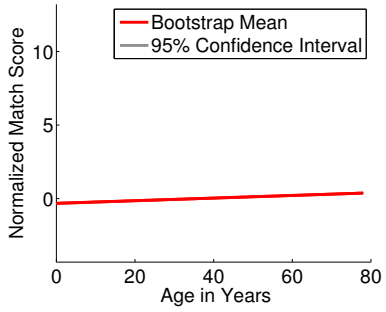


(A) COTS-1

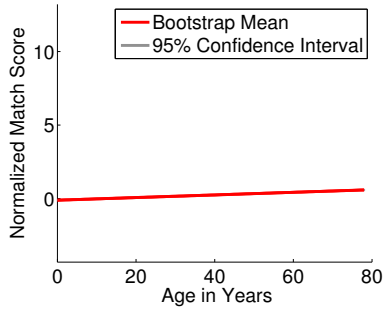


(B) COTS-2

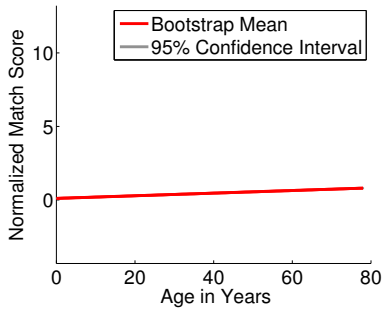
Figure S25: Population-mean trends and 95% confidence intervals of impostor match scores obtained by two COTS matchers with respect to $\triangle T_{i,1k}$ (model B'_{T}), when the scores from ten fingers are fused. The confidence intervals are too small to be visible in the plots.



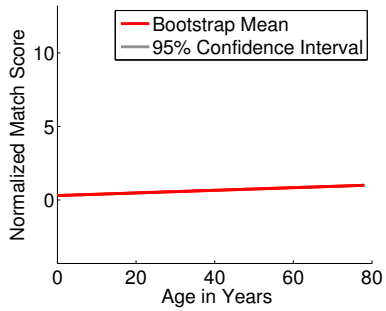
(A) $AGE_{j,1} = 20$



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

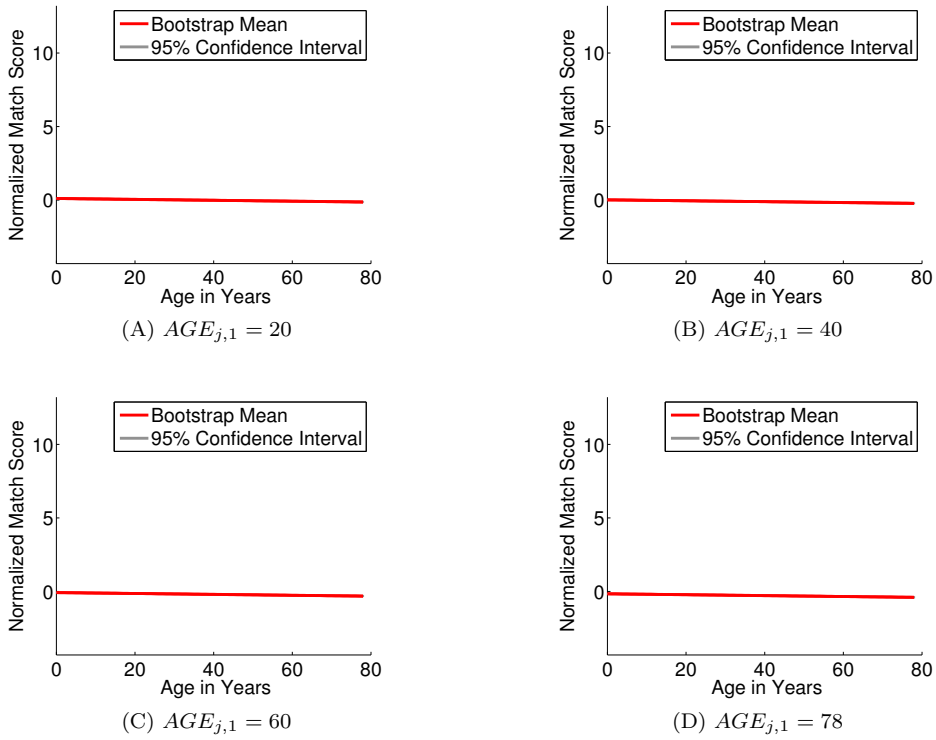


(C) $AGE_{j,1} = 60$



(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) $AGE_{j,1} = 78$

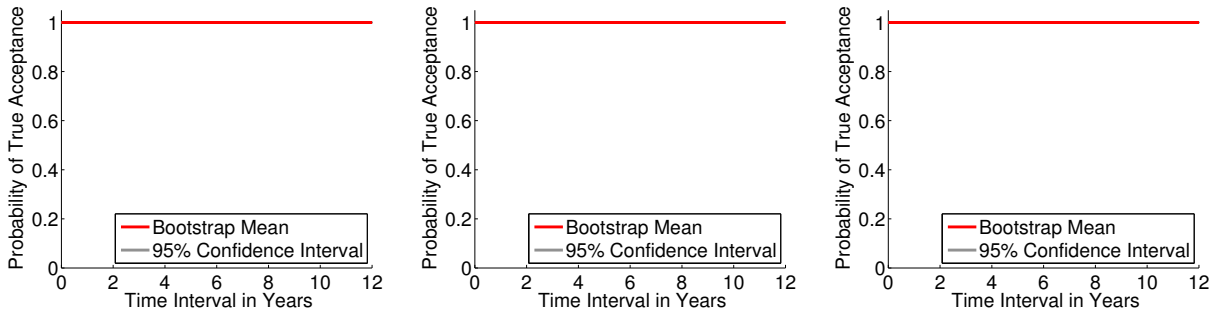
Figure S27: Population-mean trends and 95% confidence intervals of impostor match scores obtained by COTS-2 matcher with respect to $AGE_{i,k}$ (model B'_{A}), when the scores from ten fingers are fused. The confidence intervals are too small to be visible in the plots.

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 (0.0148; 0.1055)	0.0186e-05 (0.0016; 0.1333)	0.0124e-05 (0.0002; 0.1684)	0.0253e-05 (0.0060; 10.1003)	0.0002e-05 (0.0000; 3.6459)	0.0000e-05 (0.0000; 1.3160)
	FAR = 0.1%	2.2612e-05 (1.3119; 4.5754)	1.7002e-05 (0.4309; 4.5704)	1.2784e-05 (0.1415; 4.5654)	0.5792e-05 (0.4099; 1.2990)	0.1964e-05 (0.0231; 0.7103)	0.0666e-05 (0.0013; 0.3884)
	FAR = 1%	0.0047 (0.0044; 0.0051)	0.0037 (0.0032; 0.0043)	0.0029 (0.0023; 0.0036)	0.0037 (0.0030; 0.0042)	0.0014 (0.0010; 0.0025)	0.0006 (0.0003; 0.0015)

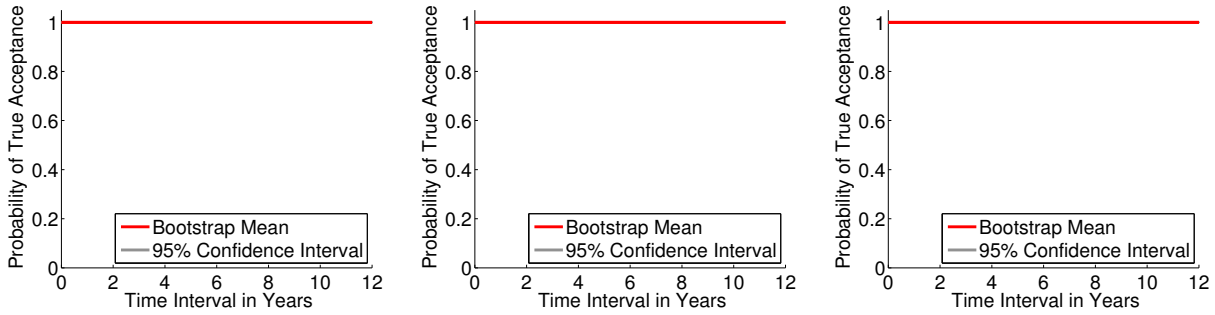
* Numbers in the tables are rounded off to four decimal points unless written in exponential notation. For some very small numbers, round-off to nine decimal points is used.

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



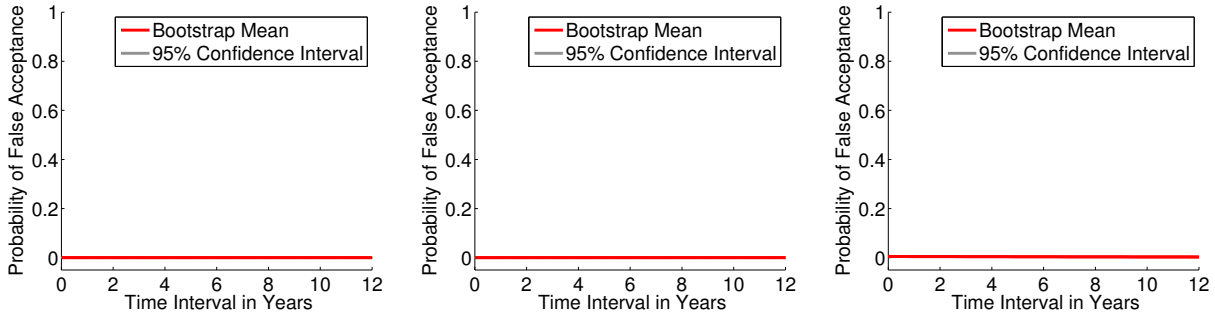
(A) FAR = 0.01% (B) FAR = 0.001% (C) FAR = 0%

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.



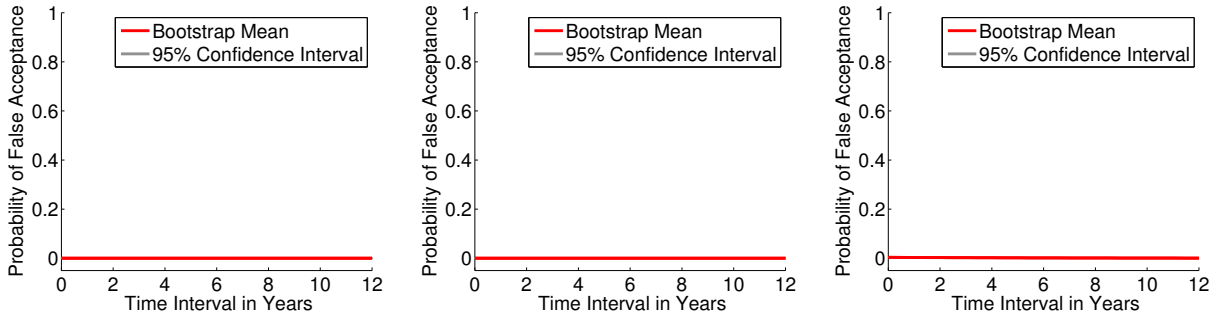
(A) FAR = 0.01% (B) FAR = 0.001% (C) FAR = 0%

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.