NIST Interagency Report 8429

Summarizing Demographic Differentials

Patrick Grother

NIST

IFPC

November 17, 2022

Why?

Demographics *do* have an effect

What?

Quantifying the problem

Role?

As a target for designers to optimize

Validate how?

Exercise them (in FRVT…)

Standardize where?

ISO/IEC 19795-10 Demographics

For use in (for example)

ISO/IEC 9868 EU Regulation

Quoting Georgetown’s Report: “The Perpetual Line-up”

[Bias Exists]

- The most prominent study [Klare et al.] found that several leading algorithms performed worse on African Americans, women, and young adults than on Caucasians, men, and older people, respectively.[216]

[Consequence]

- If the suspect is African American rather than Caucasian, the system is more likely to erroneously fail to identify the right person, potentially causing innocent people to be bumped up the list—and possibly even investigated

[Awareness]

- “Q: Is the Booking Photo Comparison System biased against minorities[?]”
- “A: No… it does not see race, sex, orientation or age. The software is matching distance and patterns only, not skin color, age or sex of an individual.”- Frequently Asked Questions, Seattle Police Department

[No Bias Tests]

- There is no independent testing regime for racially biased error rates … two major face recognition companies admitted that they did not run these tests

[Priors]

- Racial bias intrinsic to an algorithm maybe compounded by outside factors. African Americans are disproportionately likely to come into contact with—and be arrested by—law enforcement.[218]

Clare Garvie, Alvaro M. Bedoya, Jonathan Frankle
The Perpetual Line-up Unregulated Police Face Recognition In America
Georgetown Law Center on Privacy and Technology
October 18, 2016 https://www.perpetuallineup.org/

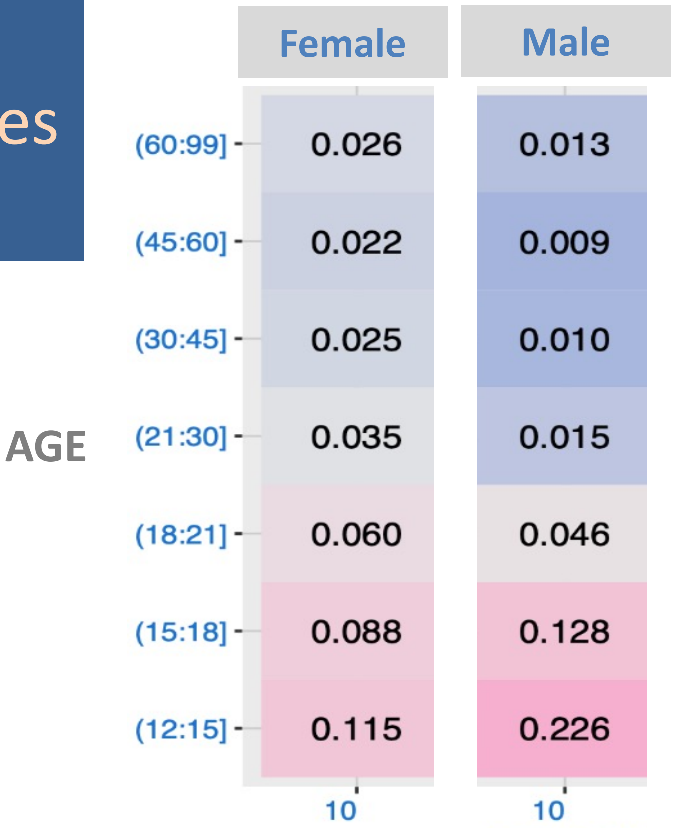
NIST Interagency Report 8280 - Dec 2019

- Distinguish between False Negatives and False Positives
- Distinguish between 1:1 and 1:N
- Consequences of differentials are application dependent
- Effects are algorithm dependent ⟶ know your algorithm ⟶ know your system
- Mitigation guidance

- ΔFNMR small cooperative images
- ΔFMR massive even in cooperative images
- Higher FNMR and FMR in women
- Higher FMR in East Asia and Africa
- Some Chinese algorithm give higher FMR in Europe
- Some 1:N algorithms effect low ΔFPIR

Effect of age:

False Negative Identification Rates aka “Miss Rates”

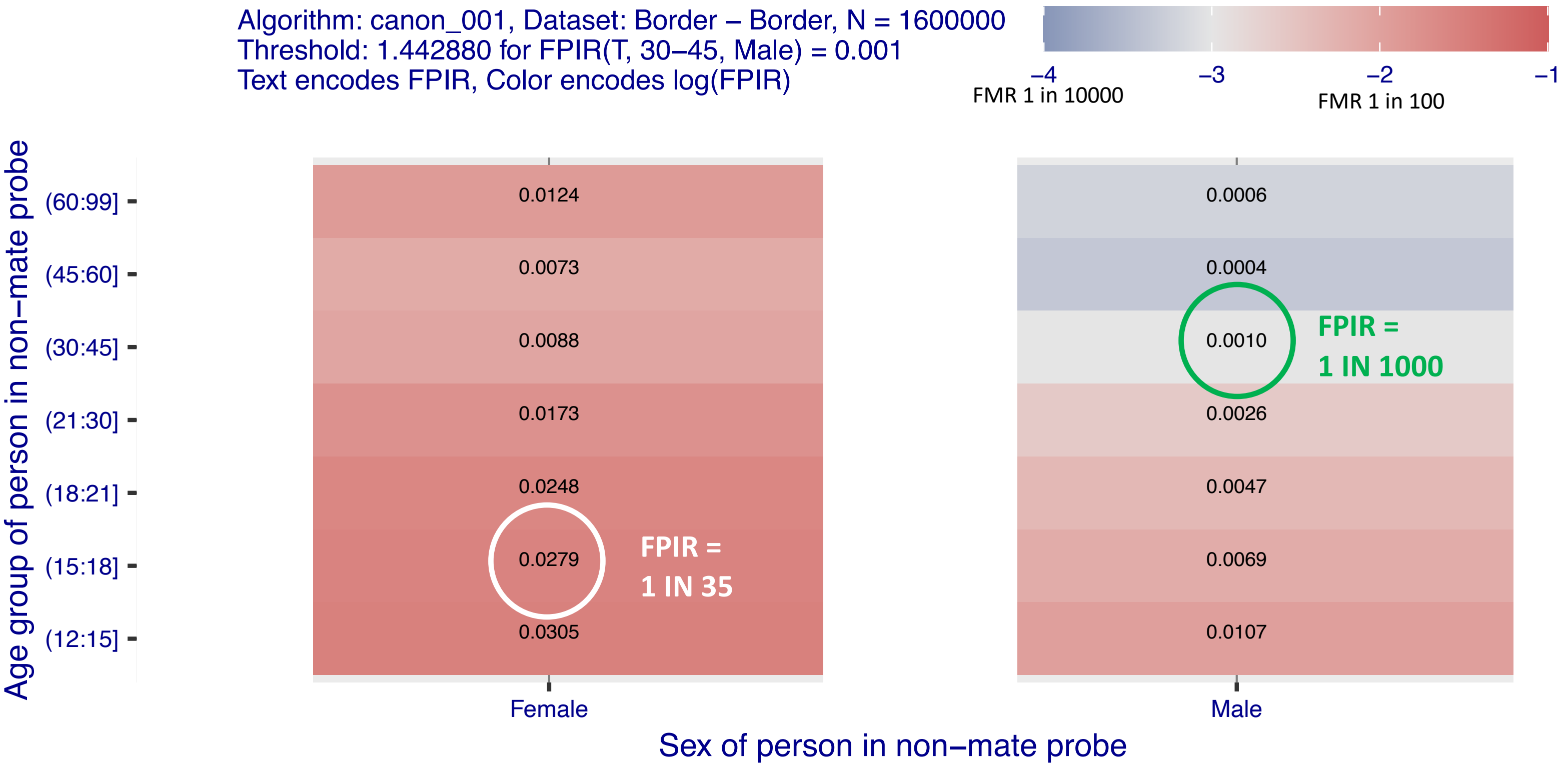


AGE	Female	Male
(60:99]	0.026	0.013
(45:60]	0.022	0.009
(30:45]	0.025	0.010
(21:30]	0.035	0.015
(18:21]	0.060	0.046
(15:18]	0.088	0.128
(12:15]	0.115	0.226

Algorithm: Canon-001 (2021-10-27)
Images: Airport immigration photos

False Negative Identification “Miss” Rates. N = 1.6 million
The threshold set to limit false positive outcomes to 1 in 1000 searches (FPIR = 0.001) for men age 30-45.

1:N False Positive Rates by Sex and Age

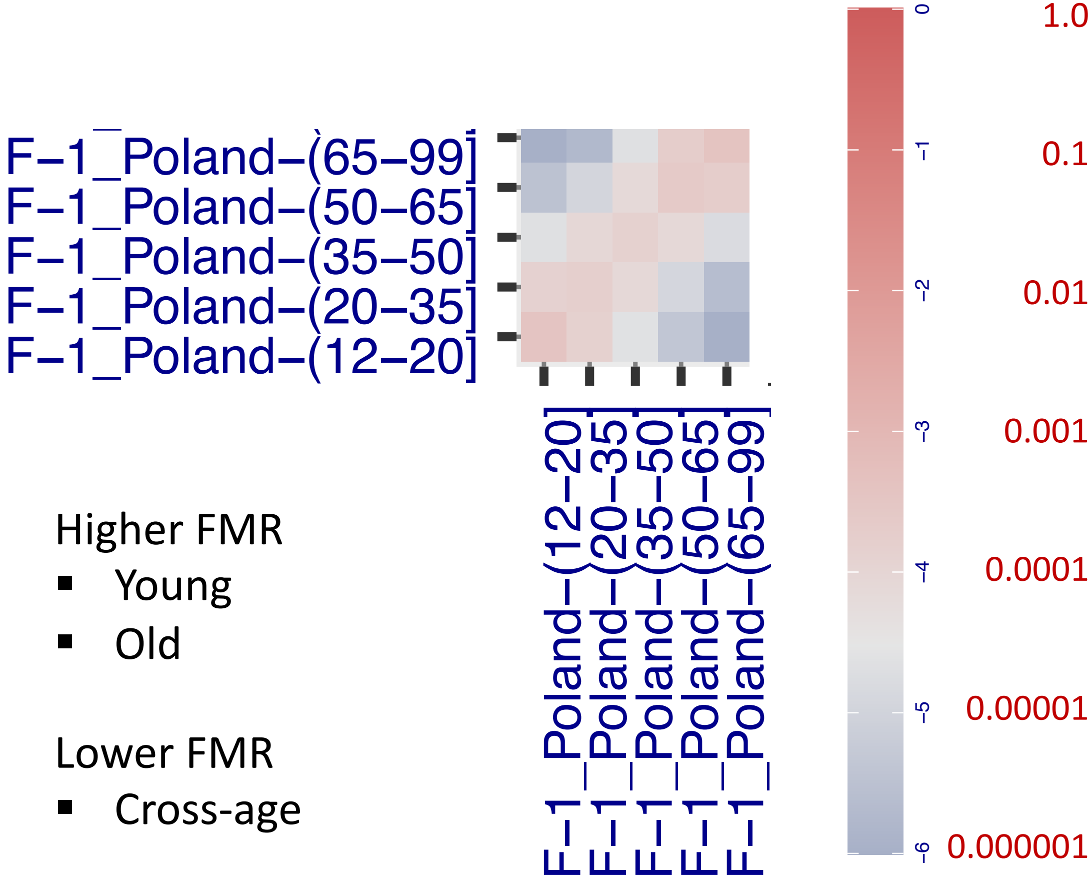


Algorithm: canon_001, Dataset: Border – Border, N = 1600000
Threshold: 1.442880 for FPIR(T, 30–45, Male) = 0.001
Text encodes FPIR, Color encodes log(FPIR)

FPIR = 1 IN 35

FPIR = 1 IN 1000

Cross age false match rates



Higher FMR

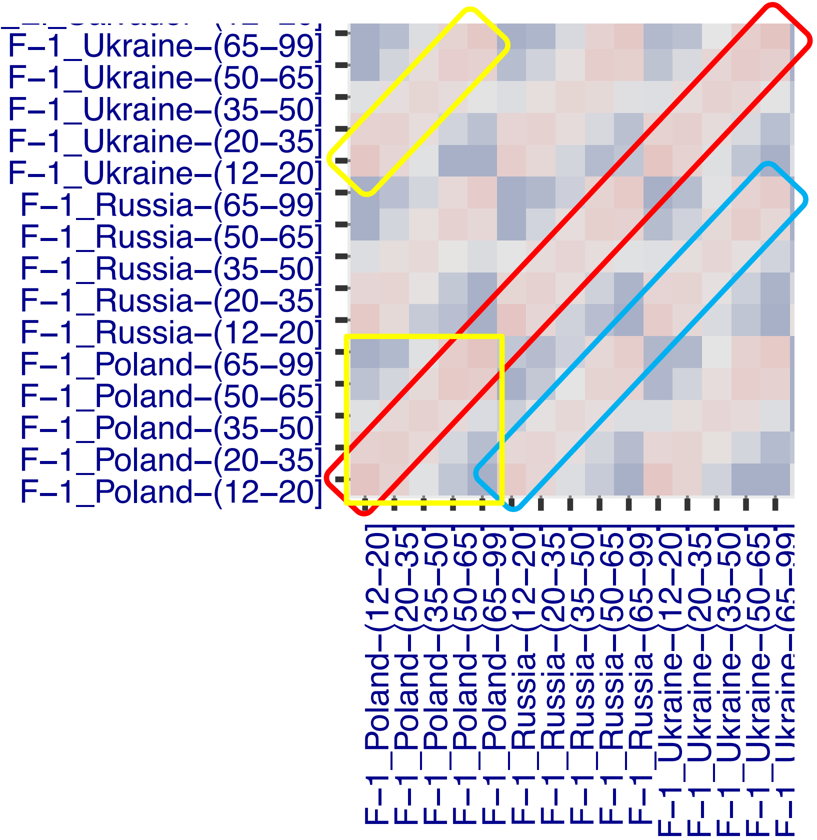
- Young
- Old

Lower FMR

- Cross-age

Algorithm: dahua_003 Threshold: 6430.000000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Cross country-of-birth and age false match rates

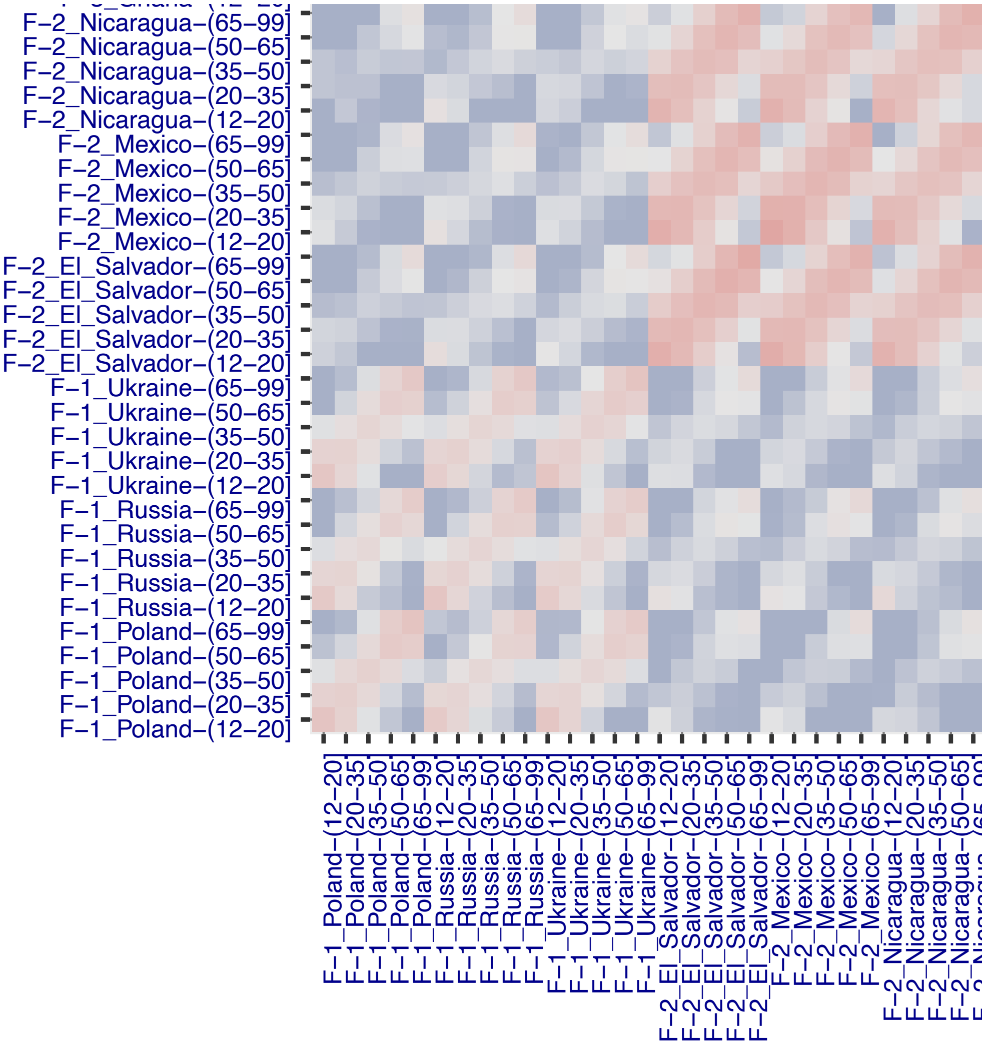


Adding women born in:

- Russia
- Ukraine

Algorithm: dahua_003 Threshold: 6430.000000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Cross country-of-birth and age false match rates

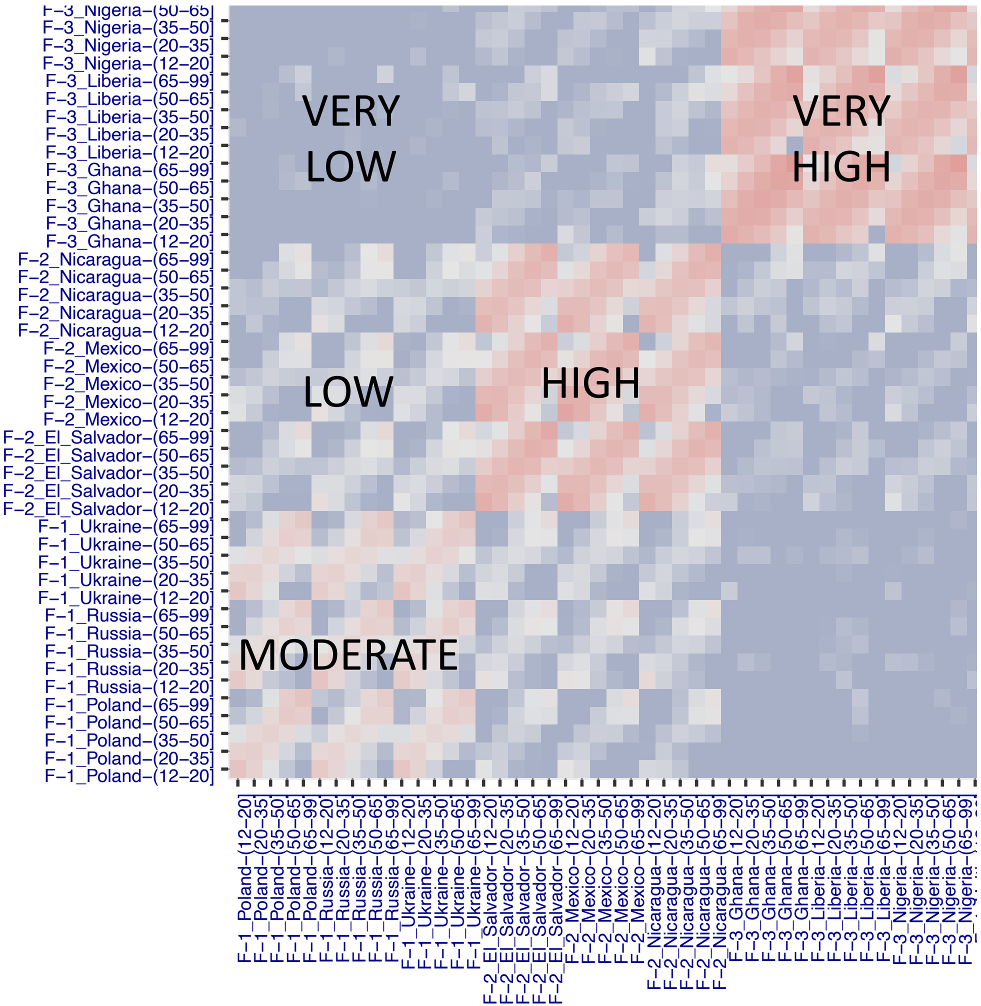


Adding people born in:

- El Salvador
- Mexico
- Nicaragua

Algorithm: dahua_003 Threshold: 6430.000000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Cross country-of-birth and age false match rates



Adding women born in:

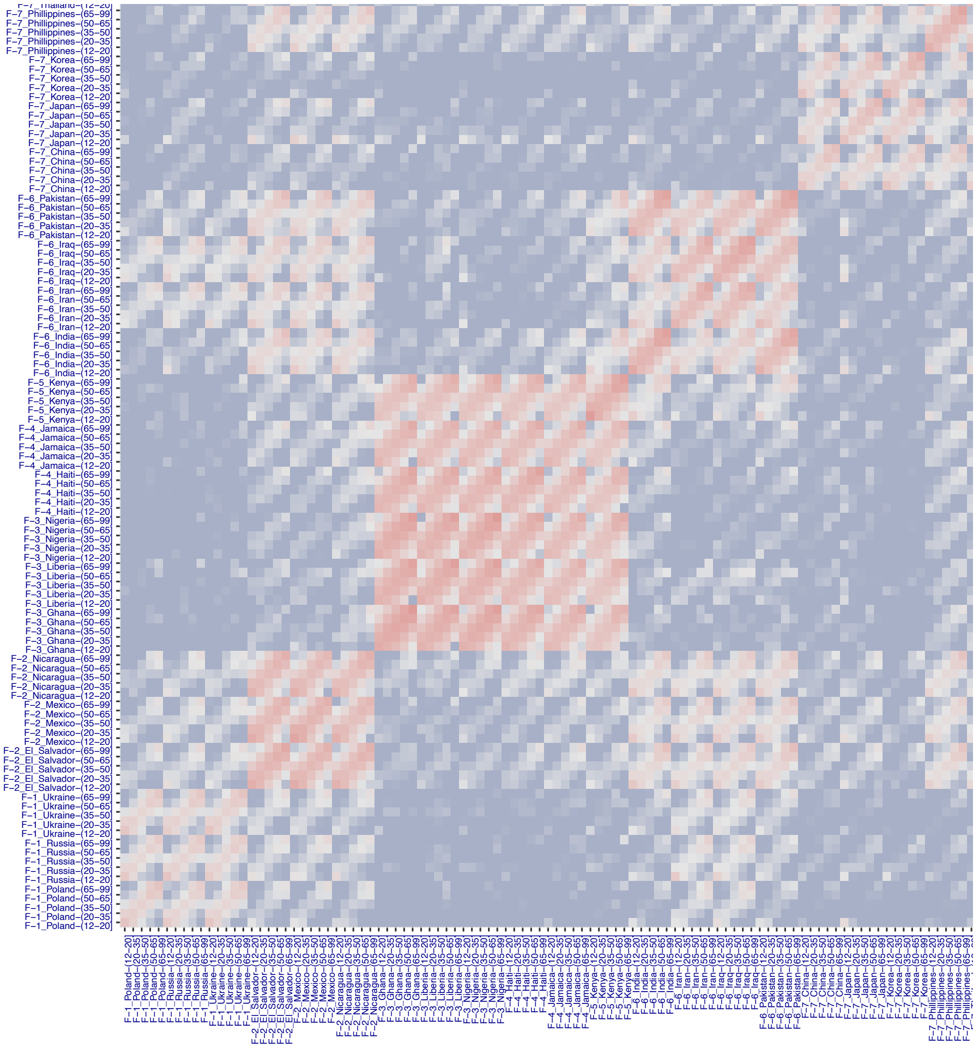
- Nigeria
- Liberia
- Ghana

Algorithm: dahua_003 Threshold: 6430.000000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Cross country-of-birth and age false match rates

Adding women in

- 22 countries
- 7 regions
  - E. Europe
  - C. America
  - W. Africa
  - Caribbean
  - E. Africa
  - S. Asia
  - E. Asia

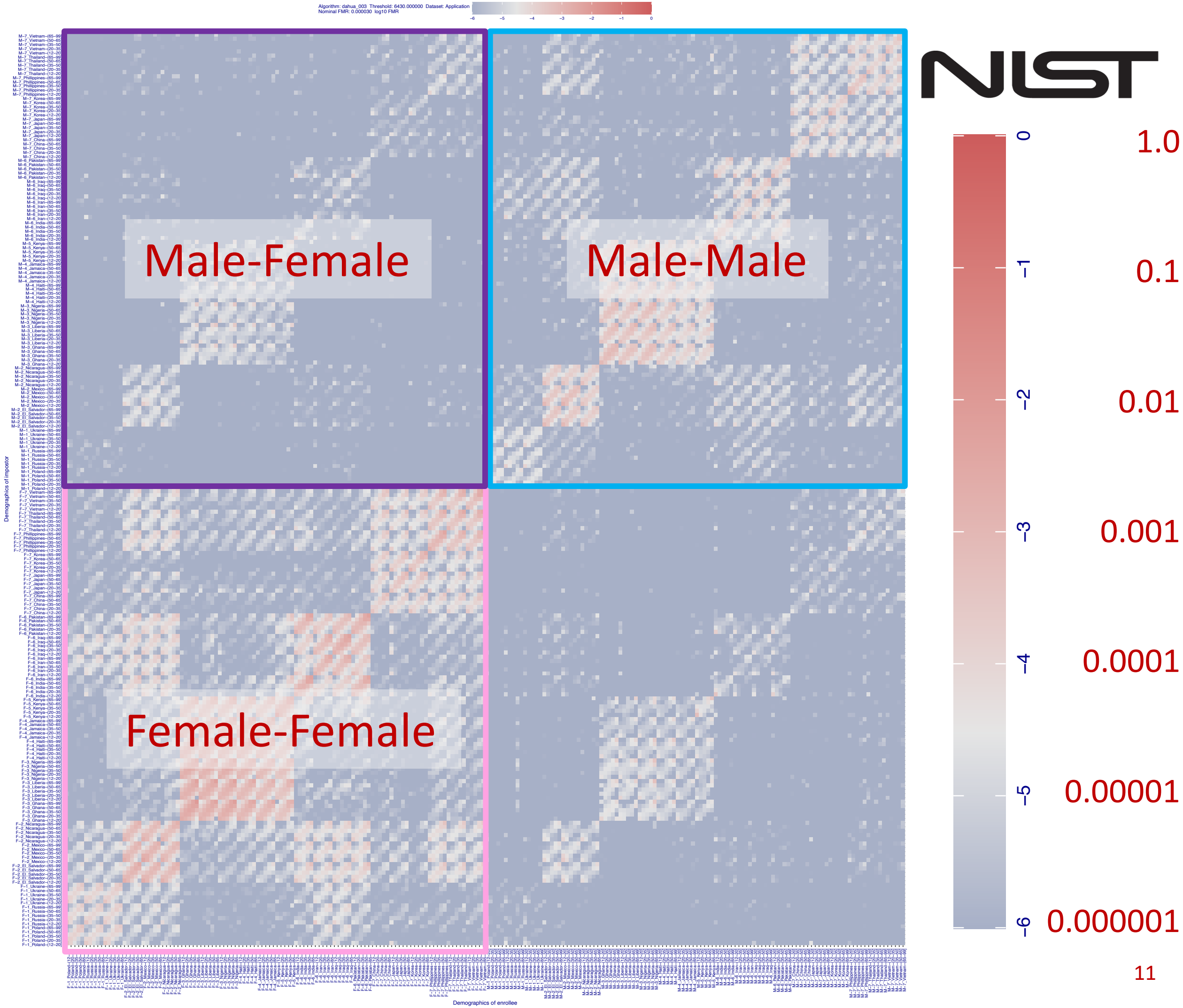


Algorithm: dahua_003 Threshold: 6430.000000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Cross country-of-birth, age, and sex false match rates

Adding:

- Males

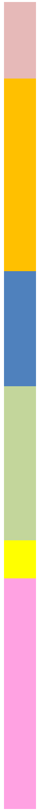


How false positives affect 1:N applications

Gallery composition:

1. Six demographic groups
2. Equally balanced – all have same number of people

How false positives affect 1:N applications



Gallery composition:

1. Six demographic groups
2. Now imbalanced - the usual case
3. Number of people in group i is n_i

4. Gallery size is N = $\sum n_i$

Private 1:N Watchlist: Copenhagen Brondby FC

Brondby fans scuffle with police during a match between the Copenhagen and Brondby soccer teams at Copenhagen's Telia Parken stadium in 2017.

Lars Ronbog/FrontzoneSport via Getty Images

Num. enrolled ~ 50
Num. searches ~ 21000

Once the men's chant is over, the group moves toward the stadium's entrance, where the men — along with 21,000 other fans — are asked to remove masks, hats and glasses so a computer can scan their faces. The scans will be compared against a list of roughly 50 banned troublemakers and will be used to determine whether the spectators will be allowed in.

No one is stopped on this day. But since the system's launch in July, it has caught four people on the blacklist, who were then turned over to police.

https://www.npr.org/2019/10/21/770280447/a-soccer-team-in-denmark-is-using-facial-recognition-to-stop-unruly-fans?sc=tw&t=1572190088133

How false positives affect 1:N applications

Probe composition over some time period

1. Say a total of 21000 searches
2. Almost all non-mates
3. Again imbalanced
4. Number of people in group i is p_i

5. Potentially gallery and probe composition differ

How false positives affect 1:N applications

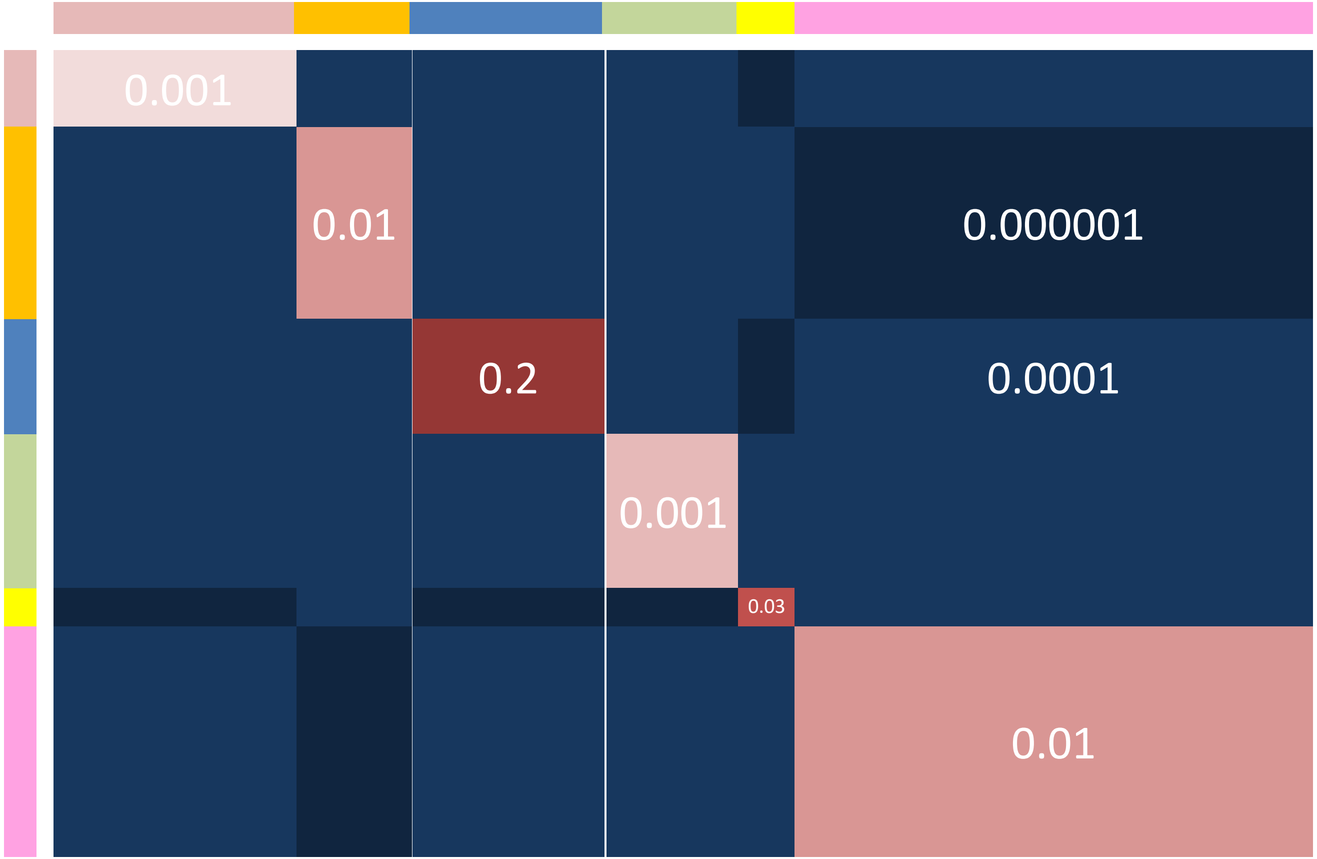
Number of expected false positives

- $NFP \quad = N \quad FMR(T) \quad P$

where

- N = gallery size
- P = number of non-mated searches
- FMR = monolithic 1:1 comparison false match rate
- assuming FMR doesn’t depend on demographics

How false positives affect 1:N applications



NFP = $\sum\sum p_j FMR_{ji} n_i$

In this toy case:

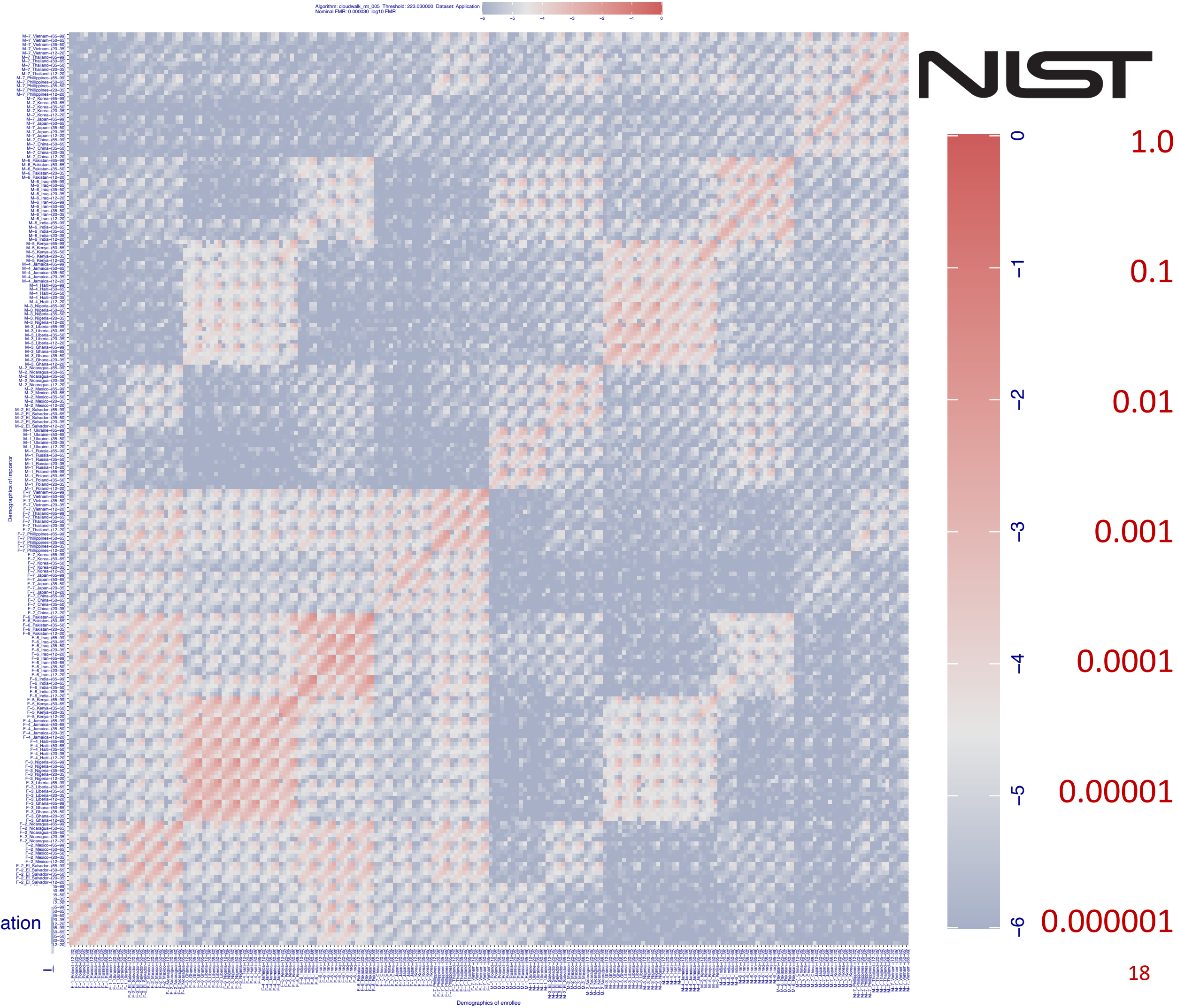
NFP is dominated by the high FMR group which has 20x higher FMR than any other.

See NISTIR 8429 Annex B

Cross country-of-birth, age, and sex false match rates

Adding:

- Males



Algorithm: cloudwalk_mt_005 Threshold: 223.030000 Dataset: Application
Nominal FMR: 0.000030 log10 FMR

Prior publication demographic consequences of FMR differentials on one-to-many search

Quantifying the Extent to Which Race and Gender Features Determine Identity in Commercial Face Recognition Algorithms

John J. Howard, Yevgeniy B. Sirotin, Jerry L. Tipton, A. Vemury

Published 15 October 2020

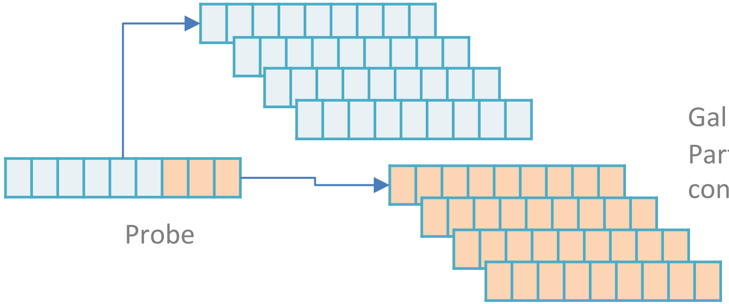
https://mdtf.org/publications/TPS-Features.pdf

Also see the older literature on (binomial) models of 1:N accuracy with heterogeneous error rates.

BUT … (And this is why NIST evaluates 1:N separately)

Caveat esp. for reviewers coming back to this deck later

- Many developers implement 1:N search as N 1:1 comparisons
- **But some do not**
  - **The enrollment database is not just N separate templates**
  - **It could be a tree, or a dictionary, or some exotic data structure**
- Some developers field both types of algorithms



Probe

Gallery,
Partitioned,
conditionally searched

- This has beneficial consequences for:
  - False positive rates
  - How false positive rates grow when N goes up
  - Demographic dependencies
  - Speed
- This has complexity
  - Deleting somebody from a database may not be a simple operation

Yu A. Malkov, D. A. Yashunin, **Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs.** IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 42 No. 4 April 2020 pp. 824–836 https://doi.org/10.1109/TPAMI.2018.2889473

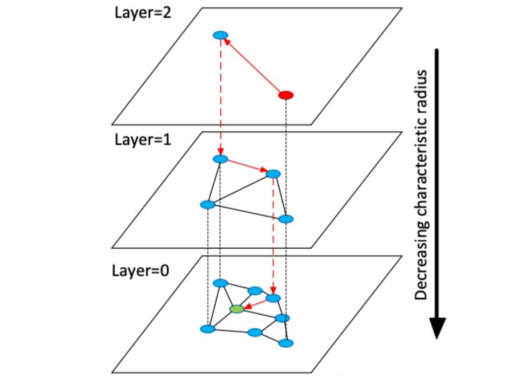
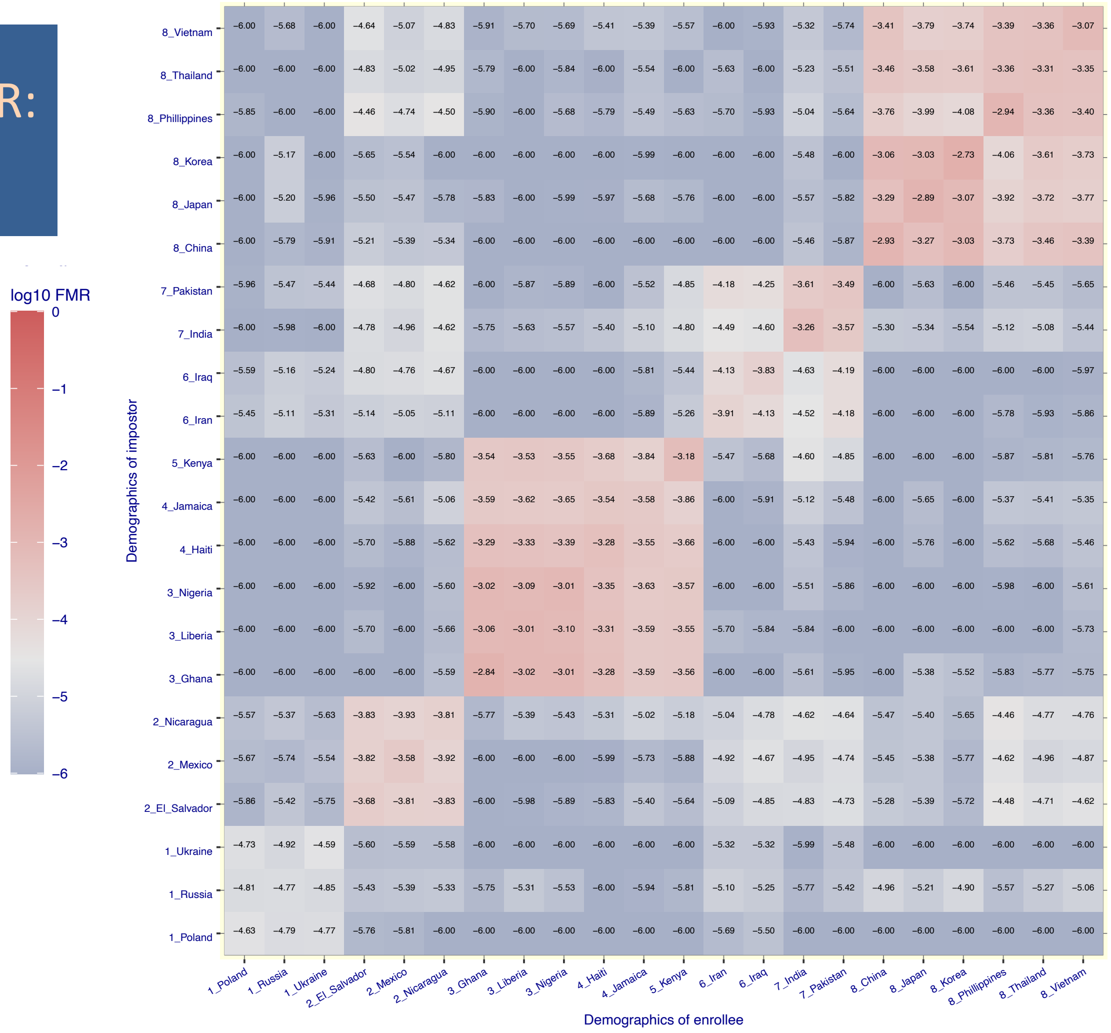


Fig. 1. Illustration of the Hierarchical NSW idea. The search starts from an element from the top layer (shown red). Red arrows show direction of the greedy algorithm from the entry point to the query (shown green).

Cross-country FMR: NTechLab (Russia)

1. FMR EU ~ 1:33000
2. FMR Nigeria 1:1000
3. FMR Korea 1:500
4. Relevance to 1:N



	1_Poland	1_Russia	1_Ukraine	2_El_Salvador	2_Mexico	2_Nicaragua	3_Ghana	3_Liberia	3_Nigeria	4_Haiti	4_Jamaica	5_Kenya	6_Iran	6_Iraq	7_India	7_Pakistan	8_China	8_Japan	8_Korea	8_Phillippines	8_Thailand	8_Vietnam
8_Vietnam	–6.00	–5.68	–6.00	–4.64	–5.07	–4.83	–5.91	–5.70	–5.69	–5.41	–5.39	–5.57	–6.00	–5.93	–5.32	–5.74	–3.41	–3.79	–3.74	–3.39	–3.36	–3.07
8_Thailand	–6.00	–6.00	–6.00	–4.83	–5.02	–4.95	–5.79	–6.00	–5.84	–6.00	–5.54	–6.00	–5.63	–6.00	–5.23	–5.51	–3.46	–3.58	–3.61	–3.36	–3.31	–3.35
8_Phillippines	–5.85	–6.00	–6.00	–4.46	–4.74	–4.50	–5.90	–6.00	–5.68	–5.79	–5.49	–5.63	–5.70	–5.93	–5.04	–5.64	–3.76	–3.99	–4.08	–2.94	–3.36	–3.40
8_Korea	–6.00	–5.17	–6.00	–5.65	–5.54	–6.00	–6.00	–6.00	–6.00	–6.00	–5.99	–6.00	–6.00	–6.00	–5.48	–6.00	–3.06	–3.03	–2.73	–4.06	–3.61	–3.73
8_Japan	–6.00	–5.20	–5.96	–5.50	–5.47	–5.78	–5.83	–6.00	–5.99	–5.97	–5.68	–5.76	–6.00	–6.00	–5.57	–5.82	–3.29	–2.89	–3.07	–3.92	–3.72	–3.77
8_China	–6.00	–5.79	–5.91	–5.21	–5.39	–5.34	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–5.46	–5.87	–2.93	–3.27	–3.03	–3.73	–3.46	–3.39
7_Pakistan	–5.96	–5.47	–5.44	–4.68	–4.80	–4.62	–6.00	–5.87	–5.89	–6.00	–5.52	–4.85	–4.18	–4.25	–3.61	–3.49	–6.00	–5.63	–6.00	–5.46	–5.45	–5.65
7_India	–6.00	–5.98	–6.00	–4.78	–4.96	–4.62	–5.75	–5.63	–5.57	–5.40	–5.10	–4.80	–4.49	–4.60	–3.26	–3.57	–5.30	–5.34	–5.54	–5.12	–5.08	–5.44
6_Iraq	–5.59	–5.16	–5.24	–4.80	–4.76	–4.67	–6.00	–6.00	–6.00	–6.00	–5.81	–5.44	–4.13	–3.83	–4.63	–4.19	–6.00	–6.00	–6.00	–6.00	–6.00	–5.97
6_Iran	–5.45	–5.11	–5.31	–5.14	–5.05	–5.11	–6.00	–6.00	–6.00	–6.00	–5.89	–5.26	–3.91	–4.13	–4.52	–4.18	–6.00	–6.00	–6.00	–5.78	–5.93	–5.86
5_Kenya	–6.00	–6.00	–6.00	–5.63	–6.00	–5.80	–3.54	–3.53	–3.55	–3.68	–3.84	–3.18	–5.47	–5.68	–4.60	–4.85	–6.00	–6.00	–6.00	–5.87	–5.81	–5.76
4_Jamaica	–6.00	–6.00	–6.00	–5.42	–5.61	–5.06	–3.59	–3.62	–3.65	–3.54	–3.58	–3.86	–6.00	–5.91	–5.12	–5.48	–6.00	–5.65	–6.00	–5.37	–5.41	–5.35
4_Haiti	–6.00	–6.00	–6.00	–5.70	–5.88	–5.62	–3.29	–3.33	–3.39	–3.28	–3.55	–3.66	–6.00	–6.00	–5.43	–5.94	–6.00	–5.76	–6.00	–5.62	–5.68	–5.46
3_Nigeria	–6.00	–6.00	–6.00	–5.92	–6.00	–5.60	–3.02	–3.09	–3.01	–3.35	–3.63	–3.57	–6.00	–6.00	–5.51	–5.86	–6.00	–6.00	–6.00	–5.98	–6.00	–5.61
3_Liberia	–6.00	–6.00	–6.00	–5.70	–6.00	–5.66	–3.06	–3.01	–3.10	–3.31	–3.59	–3.55	–5.70	–5.84	–5.84	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–5.73
3_Ghana	–6.00	–6.00	–6.00	–6.00	–6.00	–5.59	–2.84	–3.02	–3.01	–3.28	–3.59	–3.56	–6.00	–6.00	–5.61	–5.95	–6.00	–5.38	–5.52	–5.83	–5.77	–5.75
2_Nicaragua	–5.57	–5.37	–5.63	–3.83	–3.93	–3.81	–5.77	–5.39	–5.43	–5.31	–5.02	–5.18	–5.04	–4.78	–4.62	–4.64	–5.47	–5.40	–5.65	–4.46	–4.77	–4.76
2_Mexico	–5.67	–5.74	–5.54	–3.82	–3.58	–3.92	–6.00	–6.00	–6.00	–5.99	–5.73	–5.88	–4.92	–4.67	–4.95	–4.74	–5.45	–5.38	–5.77	–4.62	–4.96	–4.87
2_El_Salvador	–5.86	–5.42	–5.75	–3.68	–3.81	–3.83	–6.00	–5.98	–5.89	–5.83	–5.40	–5.64	–5.09	–4.85	–4.83	–4.73	–5.28	–5.39	–5.72	–4.48	–4.71	–4.62
1_Ukraine	–4.73	–4.92	–4.59	–5.60	–5.59	–5.58	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–5.32	–5.32	–5.99	–5.48	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00
1_Russia	–4.81	–4.77	–4.85	–5.43	–5.39	–5.33	–5.75	–5.31	–5.53	–6.00	–5.94	–5.81	–5.10	–5.25	–5.77	–5.42	–4.96	–5.21	–4.90	–5.57	–5.27	–5.06
1_Poland	–4.63	–4.79	–4.77	–5.76	–5.81	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–5.69	–5.50	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00	–6.00

Asia

S. Asia

MidEast

E. Africa

Caribbean

W. Africa

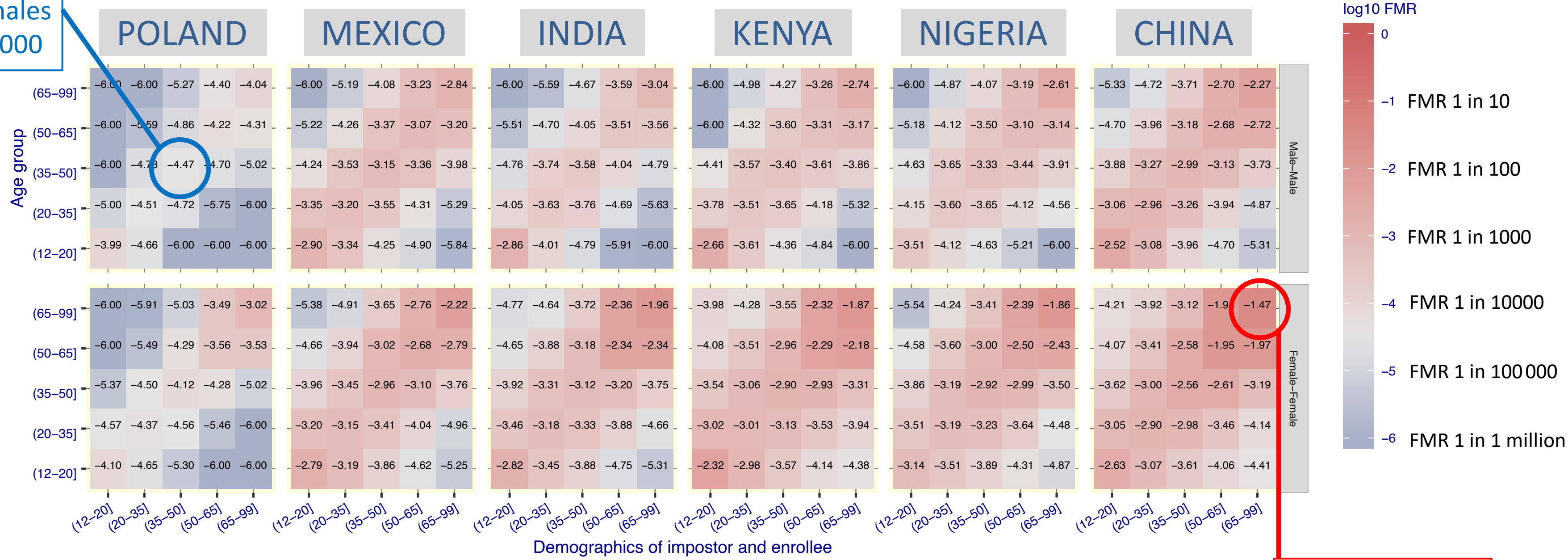
C. America

E. Europe

Magnitude matters: Age x Age for six countries

Algorithm:
imperial_002
Threshold:
1.381120
Dataset: Application
Nominal FMR: 0.000030

FMR on
Polish males
1 in 30 000



POLAND	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−6.00	−6.00	−5.27	−4.40	−4.04
	(50–65]	−6.00	−5.59	−4.86	−4.22	−4.31
	(35–50]	−6.00	−4.73	−4.47	−4.70	−5.02
	(20–35]	−5.00	−4.51	−4.72	−5.75	−6.00
	(12–20]	−3.99	−4.66	−6.00	−6.00	−6.00
Female–Female	(65–99]	−6.00	−5.91	−5.03	−3.49	−3.02
	(50–65]	−6.00	−5.49	−4.29	−3.56	−3.53
	(35–50]	−5.37	−4.50	−4.12	−4.28	−5.02
	(20–35]	−4.57	−4.37	−4.56	−5.46	−6.00
	(12–20]	−4.10	−4.65	−5.30	−6.00	−6.00

MEXICO	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−6.00	−5.19	−4.08	−3.23	−2.84
	(50–65]	−5.22	−4.26	−3.37	−3.07	−3.20
	(35–50]	−4.24	−3.53	−3.15	−3.36	−3.98
	(20–35]	−3.35	−3.20	−3.55	−4.31	−5.29
	(12–20]	−2.90	−3.34	−4.25	−4.90	−5.84
Female–Female	(65–99]	−5.38	−4.91	−3.65	−2.76	−2.22
	(50–65]	−4.66	−3.94	−3.02	−2.68	−2.79
	(35–50]	−3.96	−3.45	−2.96	−3.10	−3.76
	(20–35]	−3.20	−3.15	−3.41	−4.04	−4.96
	(12–20]	−2.79	−3.19	−3.86	−4.62	−5.25

INDIA	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−6.00	−5.59	−4.67	−3.59	−3.04
	(50–65]	−5.51	−4.70	−4.05	−3.51	−3.56
	(35–50]	−4.76	−3.74	−3.58	−4.04	−4.79
	(20–35]	−4.05	−3.63	−3.76	−4.69	−5.63
	(12–20]	−2.86	−4.01	−4.79	−5.91	−6.00
Female–Female	(65–99]	−4.77	−4.64	−3.72	−2.36	−1.96
	(50–65]	−4.65	−3.88	−3.18	−2.34	−2.34
	(35–50]	−3.92	−3.31	−3.12	−3.20	−3.75
	(20–35]	−3.46	−3.18	−3.33	−3.88	−4.66
	(12–20]	−2.82	−3.45	−3.88	−4.75	−5.31

KENYA	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−6.00	−4.98	−4.27	−3.26	−2.74
	(50–65]	−6.00	−4.32	−3.60	−3.31	−3.17
	(35–50]	−4.41	−3.57	−3.40	−3.61	−3.86
	(20–35]	−3.78	−3.51	−3.65	−4.18	−5.32
	(12–20]	−2.66	−3.61	−4.36	−4.84	−6.00
Female–Female	(65–99]	−3.98	−4.28	−3.55	−2.32	−1.87
	(50–65]	−4.08	−3.51	−2.96	−2.29	−2.18
	(35–50]	−3.54	−3.06	−2.90	−2.93	−3.31
	(20–35]	−3.02	−3.01	−3.13	−3.53	−3.94
	(12–20]	−2.32	−2.98	−3.57	−4.14	−4.38

NIGERIA	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−6.00	−4.87	−4.07	−3.19	−2.61
	(50–65]	−5.18	−4.12	−3.50	−3.10	−3.14
	(35–50]	−4.63	−3.65	−3.33	−3.44	−3.91
	(20–35]	−4.15	−3.60	−3.65	−4.12	−4.56
	(12–20]	−3.51	−4.12	−4.63	−5.21	−6.00
Female–Female	(65–99]	−5.54	−4.24	−3.41	−2.39	−1.86
	(50–65]	−4.58	−3.60	−3.00	−2.50	−2.43
	(35–50]	−3.86	−3.19	−2.92	−2.99	−3.50
	(20–35]	−3.51	−3.19	−3.23	−3.64	−4.48
	(12–20]	−3.14	−3.51	−3.89	−4.31	−4.87

CHINA	Age group	(12–20]	(20–35]	(35–50]	(50–65]	(65–99]
Male–Male	(65–99]	−5.33	−4.72	−3.71	−2.70	−2.27
	(50–65]	−4.70	−3.96	−3.18	−2.68	−2.72
	(35–50]	−3.88	−3.27	−2.99	−3.13	−3.73
	(20–35]	−3.06	−2.96	−3.26	−3.94	−4.87
	(12–20]	−2.52	−3.08	−3.96	−4.70	−5.31
Female–Female	(65–99]	−4.21	−3.92	−3.12	−1.94	−1.47
	(50–65]	−4.07	−3.41	−2.58	−1.95	−1.97
	(35–50]	−3.62	−3.00	−2.56	−2.61	−3.19
	(20–35]	−3.05	−2.90	−2.98	−3.46	−4.14
	(12–20]	−2.63	−3.07	−3.61	−4.06	−4.41

FMR ~ 1 in 30
on Chinese
women > 65

Summary Stat. #1:	Maximum / Minimum	~ 1000
Summary Stat. #2:	Maximum / Geometric Mean:	~ 43

Candidate Measures

IDIAP

$$\mathrm{A}(\tau) = \max_{d_i} \mathrm{FMR}_{d_i}(\tau) - \min_{d_j} \mathrm{FMR}_{d_j}(\tau)$$

NIST

$$\mathrm{A}(\tau) = \frac{\max_{d_i} \mathrm{FMR}_{d_i}(\tau)}{\min_{d_j} \mathrm{FMR}_{d_j}(\tau)} \qquad \forall d_i, d_j \in \mathcal{D}$$

AWS

$$\mathrm{A}(\tau) = \sum_{d \in \mathcal{D}} \left| \log_{10} \frac{\mathrm{FMR}_d(\tau)}{\mathrm{FMR}^{\dagger}(\tau)} \right|$$

IDEMIA

$$\mathrm{A}(\tau) = \frac{\max_{d_i} \mathrm{FMR}_{d_i}(\tau)}{\mathrm{FMR}^{\dagger}}$$

MDTF (GINI)

$$\mathrm{A}(\tau) = \frac{\sum_i \sum_j \left|\mathrm{FMR}_{d_i}(\tau) - \mathrm{FMR}_{d_j}(\tau)\right|}{2n^2\, \mathrm{FMR}^{\diamond}} \frac{n}{n-1}$$

Two demographic summary measures

$$\mathrm{A}(\tau) = \frac{\max_{d_i} \mathrm{FMR}_{d_i}(\tau)}{\mathrm{FMR}^{\dagger}}$$

Worst-case error rate over all demographic groups divided by the geometric mean

$$\mathrm{A}(\tau) = \frac{\sum_i \sum_j \left|\mathrm{FMR}_{d_i}(\tau) - \mathrm{FMR}_{d_j}(\tau)\right|}{2n^2\,\mathrm{FMR}^{\diamond}} \, \frac{n}{n-1}$$

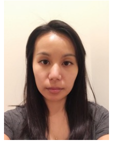
Mean absolute error rate difference over all demographic groups, divided by the arithmetic mean

$$x^{\dagger} = \left(\prod_i x_i\right)^{1/n}$$

$$x^{\diamond} = n^{-1} \sum_i x_i$$

False Match Rates have bigger demographic variations

Visa



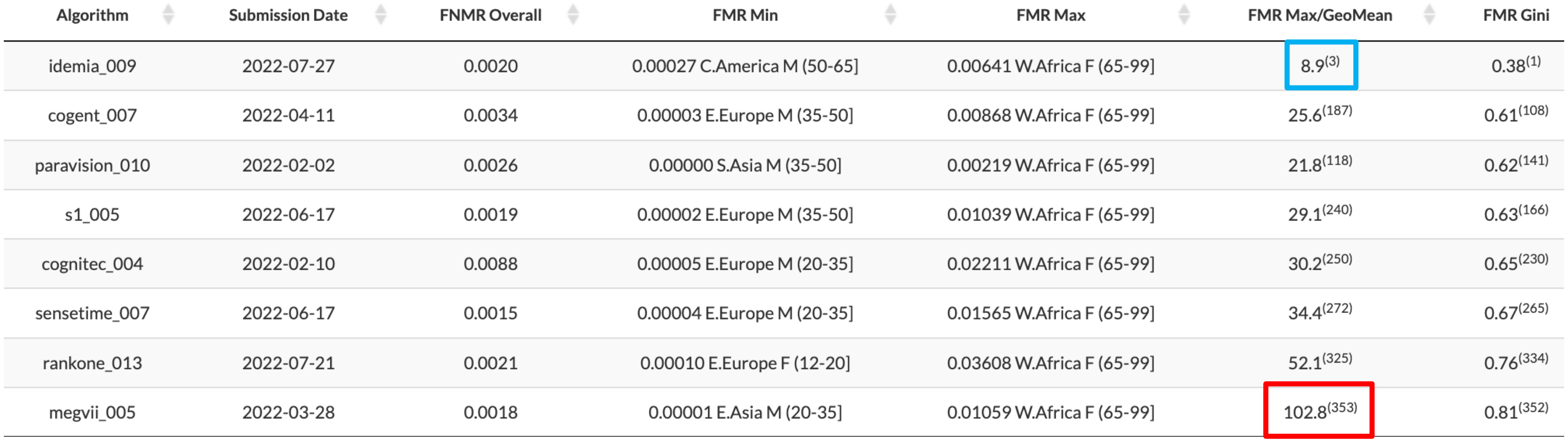
Border

A: Lowest false match rates often in E. European men

B: Highest false match rates in older W. African women

C: False match rates 10-100 times higher

D: Economists standard measure “Gini” is much higher



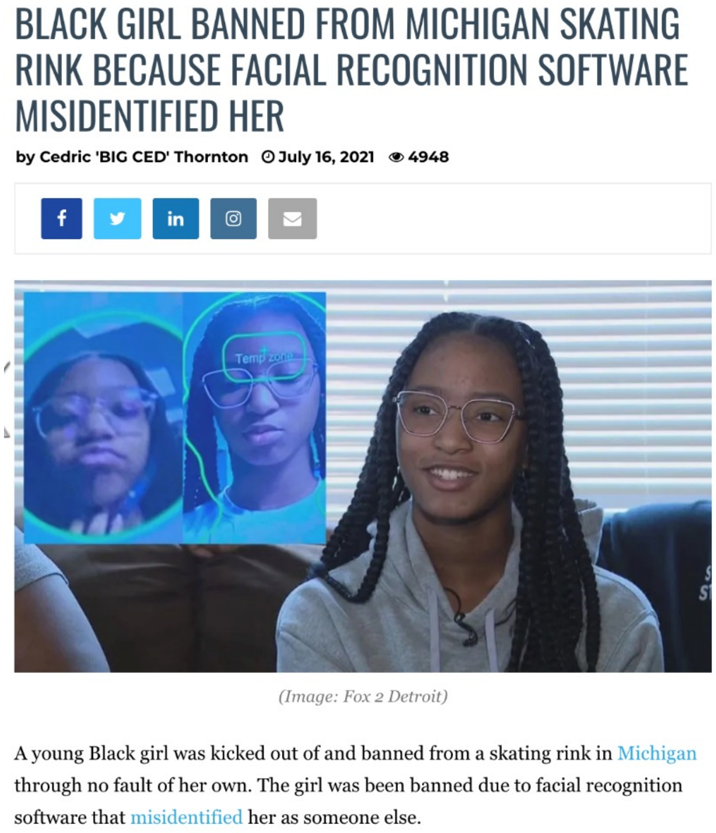
Algorithm	Submission Date	FNMR Overall	FMR Min	FMR Max	FMR Max/GeoMean	FMR Gini
idemia_009	2022-07-27	0.0020	0.00027 C.America M (50-65]	0.00641 W.Africa F (65-99]	8.9$^{(3)}$	0.38$^{(1)}$
cogent_007	2022-04-11	0.0034	0.00003 E.Europe M (35-50]	0.00868 W.Africa F (65-99]	25.6$^{(187)}$	0.61$^{(108)}$
paravision_010	2022-02-02	0.0026	0.00000 S.Asia M (35-50]	0.00219 W.Africa F (65-99]	21.8$^{(118)}$	0.62$^{(141)}$
s1_005	2022-06-17	0.0019	0.00002 E.Europe M (35-50]	0.01039 W.Africa F (65-99]	29.1$^{(240)}$	0.63$^{(166)}$
cognitec_004	2022-02-10	0.0088	0.00005 E.Europe M (20-35]	0.02211 W.Africa F (65-99]	30.2$^{(250)}$	0.65$^{(230)}$
sensetime_007	2022-06-17	0.0015	0.00004 E.Europe M (20-35]	0.01565 W.Africa F (65-99]	34.4$^{(272)}$	0.67$^{(265)}$
rankone_013	2022-07-21	0.0021	0.00010 E.Europe F (12-20]	0.03608 W.Africa F (65-99]	52.1$^{(325)}$	0.76$^{(334)}$
megvii_005	2022-03-28	0.0018	0.00001 E.Asia M (20-35]	0.01059 W.Africa F (65-99]	102.8$^{(353)}$	0.81$^{(352)}$

Source: https://pages.nist.gov/frvt/html/frvt11.html

Conclusions:

1. False negative rates vary greatly across demographic groups (age, gender, region-of-birth)
2. Some developers have improved

Demographics: A False Positive Anecdote



BLACK GIRL BANNED FROM MICHIGAN SKATING RINK BECAUSE FACIAL RECOGNITION SOFTWARE MISIDENTIFIED HER

by Cedric 'BIG CED' Thornton July 16, 2021 4948

(Image: Fox 2 Detroit)

A young Black girl was kicked out of and banned from a skating rink in Michigan through no fault of her own. The girl was been banned due to facial recognition software that misidentified her as someone else.

https://www.zdnet.com/article/backlash-to-retail-use-of-facial-recognition-grows-after-michigan-teen-kicked-out-of-skating-rink-after-false-match/

THANKS PATRICK.GROTHER@NIST.GOV FRVT@NIST.GOV

NIST INTERAGENCY REPORT 8429
SUMMARIZING DEMOGRAPHIC
DIFFERENTIALS

ISO/IEC 9868 WD7
PASSIVELY CAPTURED SUBJECTS

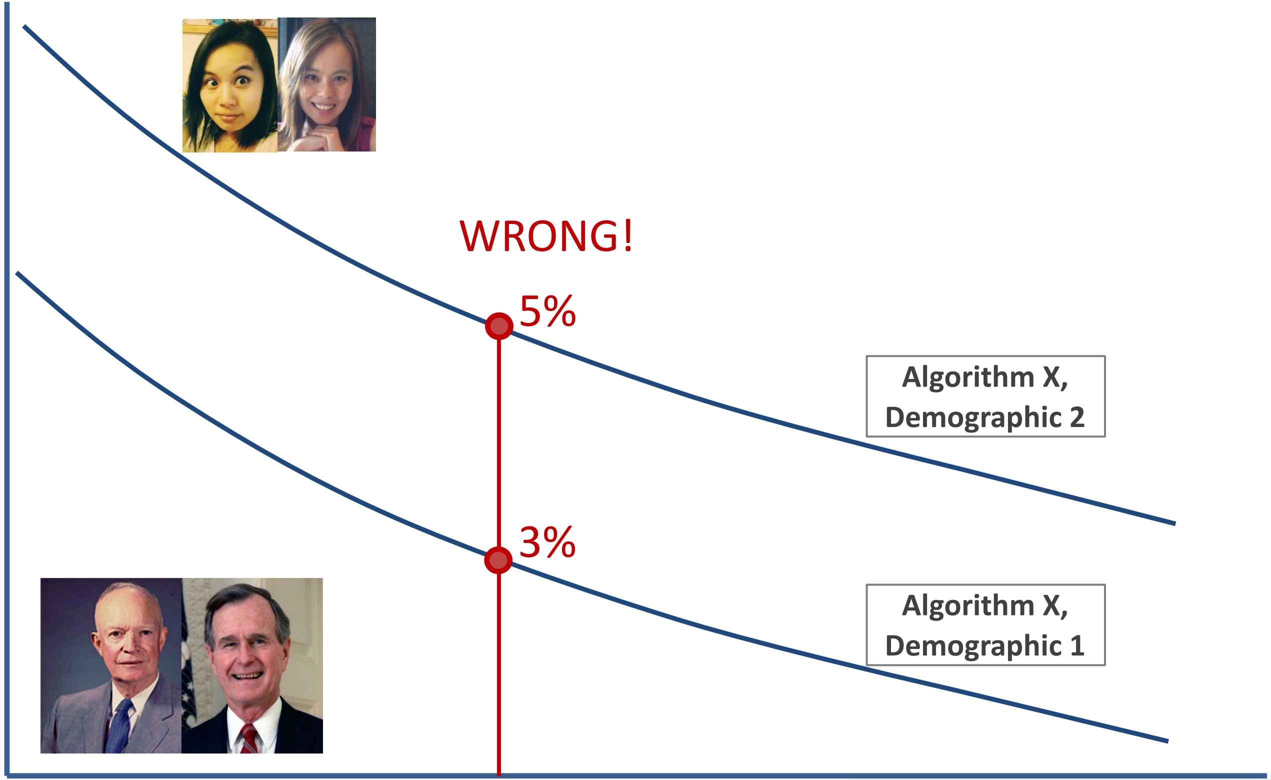
ISO/IEC 19795-10 WD4
QUANTIFYING DEMOGRAPHIC EFFECTS

Methodology: Error tradeoff characteristics for two demographics

FNMR
False non-match rate

Proportion of genuine comparisons producing score below threshold, T.

See ISO/IEC 19795-1



Low FMR values achieved with higher, i.e. more stringent, thresholds.

$\mathrm{FMR}_{\mathrm{POLICY}}$

Log-scale is often required because low FMR values are operationally relevant.

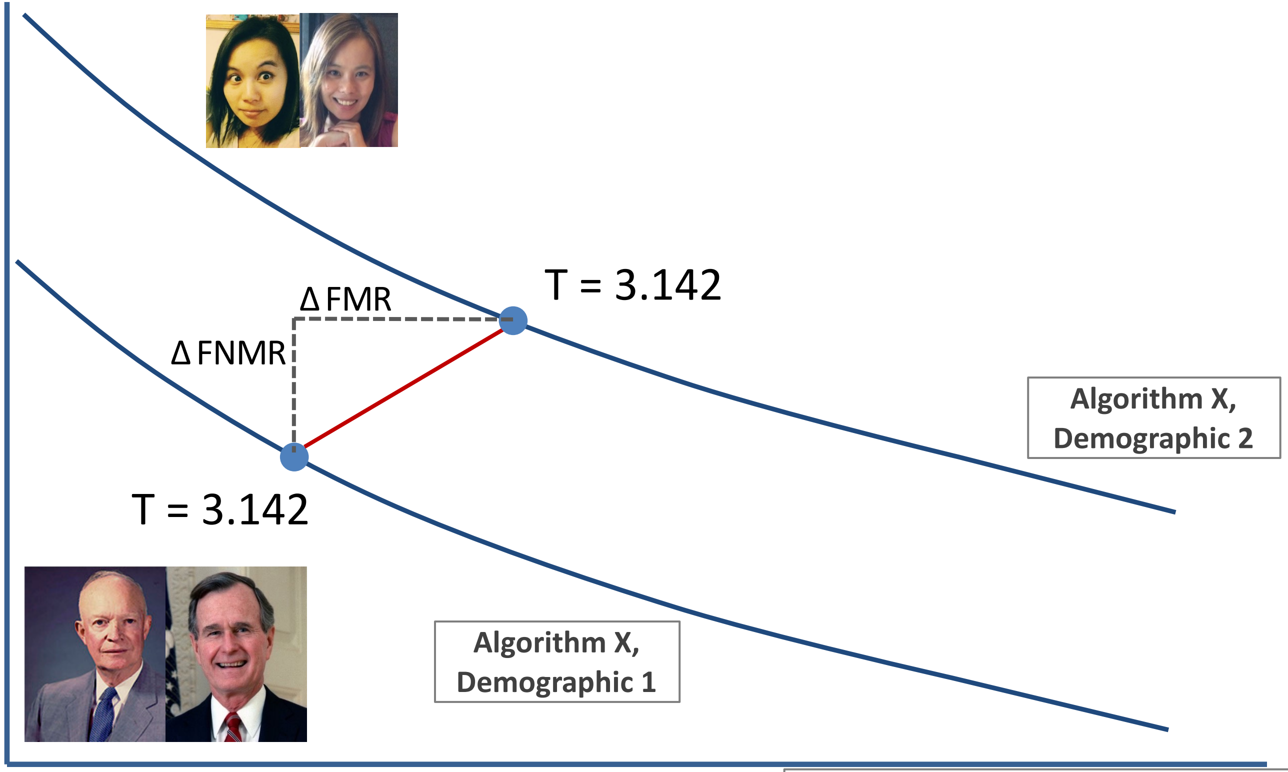
FMR False match rate
Proportion of impostor comparisons searches yielding any candidates at or above threshold, T.

Algorithms are configured with a fixed threshold

FNMR
False non-match rate

Proportion of genuine comparisons producing score below threshold, T.

See ISO/IEC 19795-1



Low FMR values achieved with higher, i.e. more stringent, thresholds.

Log-scale is often required because low FMR values are operationally relevant.

FMR False match rate
Proportion of impostor comparisons searches yielding any candidates at or above threshold, T.