**Abstract**

This report includes extensive visualization of false match rates across demographics. In the context of one-to-one matching these are some security related implications for authentication and non-repudiation applications. If the same algorithm is used in a one-to-many search context, implementing 1:N search as N 1:1 comparisons, then the false match rate variations will impact false positive identification rates in search applications. This paper documents a model of how false match rate measurements for various demographics will affect one-to-many applications

# 1 Relating 1:1 results to 1:N applications

This report card includes tabulation of error rates, differentials and summaries for 1:1 face comparison algorithms. This will be important also to that subset of 1:N identification search algorithms that implement search by computing N 1:1 scores, sorting them, and then returning candidate hits if the scores exceed a recognition threshold.

**Note that a majority of 1:N algorithms operate in this way. A significant minority however do not[1], such that the binomial model of recognition given below does not apply. In such cases, demographic effects can only be measured empirically by running one-to-many trials - this was done in NIST Interagency Report 8280.**

Using the the N 1:1 comparison construct, the following extends the well known Binomial model of false postive identification rate in an N-person gallery, namely that a false positive occurs unless *all* comparisons are below threshold:

$$\text{FPIR}(\tau) = 1 - (1 - \text{FMR}(\tau))^N \tag{1}$$

which is approximately

$$\text{FPIR}(\tau) = N\,\text{FMR}(\tau) \tag{2}$$

at high thresholds for which FMR $\ll N^{-1}$.

The following adapts points made in a presentation by Sirotin et al. at the March 2021 EAB Demographics Conference and then openly published [2] and then re-iterated by others [1]. Others have previously considered heterogeneous false match rates in identification systems [4].

Given demographic groups $i$ and $j$ and estimates for false match rate, $\text{FMR}_{ij}(\tau)$, for comparison of samples from those groups, at threshold $\tau$, we estimate one-to-many false positive identification rate for group $i$ for a enrollment database comprised of $n_j$ samples from demographic groups $1 \le j \le J$

$$\text{FPIR}_i(\tau) = 1 - \prod_j (1 - \text{FMR}_{ij})^{n_j} \tag{3}$$

where the matrix $\text{FMR}_{ij}$ expresses cross-demographic false match rates for all combinations of age, sex, and country-of-birth. If all $\text{FMR}_{ij} \ll 1/n_j$ this simplifies to

$$\text{FPIR}_i(\tau) = \sum_j \text{FMR}_{ij}(\tau)\, n_j \tag{4}$$

which has the convenient matrix notation:

$$\text{FPIR}_i(\tau) = \textbf{FMR}(\tau)\ \mathbf{n} \tag{5}$$

where $\mathbf{n}$ is the database composition vector, whose $i$-th element is the integer count of people in demographic $i$. Note

[1] See Figure J in algorithm-specific report cards for a one-to-many algorithm that has FPIR(T) scaling linearly as predicted by binomial models, and an example of one that does not.

that the matrix notation is an elegant device made possible by the approximation used for eq. 2 but is not necessary: We could re-write with the full Binomial from eq. 1.

Further, if this database is later searched with $p_i$ probes from each demographic group $1 \leq i \leq I$ then the expected number of false positives (NFP) for that group is

$$\text{NFP}_i(\tau) = p_i \ \text{FPIR}_i(\tau) \tag{6}$$

and the total number would be

$$\text{NFP}(\tau) = \mathbf{p}^T \ \mathbf{FMR}(\tau) \ \mathbf{n} \tag{7}$$

where $\mathbf{p}$ is the probe search count vector. An overall FPIR is available from its definition as the number of false positives divided by the number of searches:

$$\text{FPIR}(\tau) = \frac{\text{NFP}(\tau)}{\sum_i p_i} \tag{8}$$

**Special cases:** Worth considering are two special forms for **FMR**. First is the case of broadly homogeneous [3] false match rates in which $\mathbf{FMR} = f\mathbf{1}\mathbf{1}^T$ (with $\mathbf{1}^T = (1, 1, \ldots 1)$) meaning that false match rates don't depend on these demographics at all. In that case the number of false positives is

$$\text{NFP}(\tau) = f(\tau) \sum_i n_i \ \sum_i p_i \tag{9}$$

and the false positive identification rate is

$$\text{FPIR}(\tau) = f(\tau) \sum_i n_i \quad = N \, \text{FMR}(\tau) \tag{10}$$

which is equation 2. This is widely considered to hold for the features extracted from fingerprint and iris characteristics, and yields the situation where demographic false positive counts are driven simply by representation of the groups in the enrollee population, with $f(\tau)$ being a pan-demographic FMR scalar value.

A second case is of narrow homogeneity, $\mathbf{FMR} = f\text{I}$, meaning that false matches only occur within-demographic and all groups have the same rate, $f$.

$$\text{NFP}(\tau) = f(\tau)\mathbf{p}^T\mathbf{In} = f(\tau)\mathbf{p}^T\mathbf{n} = f(\tau)\sum_i n_i p_i \tag{11}$$

$$\text{FPIR}(\tau) = f(\tau)\frac{\sum_i n_i p_i}{\sum_i p_i} \tag{12}$$

This means that false positive outcomes depend now on the demographic structure of the searches, in addition to the enrollments. This point was made by Howard et al. [2].

For a given $f$, equation 9 gives a higher value than 11 but a biometric modality or algorithm that offered broad homogeneity could be configured with a different threshold $\tau$ to give lower $f$.

In summary, the expected number of false positives for a demographic will depend on

- ▷ **Gallery presence:** How commonly members of the particular demographic are present in the gallery.
- ▷ **False match rates within demographic:** The $\text{FMR}_{ii}$ values govern how often individuals false match against people with the same demographics.
- ▷ **False match rates against other demographics:** As is evident in the heatmaps, false matches with other demo-

graphic groups are not insignificant, and must be accounted for. The full matrix shows, for example, significant male-female false match rates in the young, $(12 - 20]$.

- ▷ **Search volumes:** Once an FR system is deployed, the frequency with which individuals from a particular group are searched will increase the number of false positives for that group. This is separate to their presence in the enrollment database and their propensity to match within and across demographic groups.

**Important:** An important subset of 1:N search algorithms do not implement search as N 1:1 comparisons, and the Binomial formulation above does not apply. In particular, as noted in NIST Interagency Report 8280: FRVT Part 3: Demographics, some algorithms, specifically stabilize the right tail of the impostor distribution so that gallery size does not affect FPIR (FPIR is constant vs. linear in $N$) and they thereby reduce demographic variations in FPIR. This caveat is not present in the cited publications.

## References

[1] Pawel Drozdowski, Christian Rathgeb, and Christoph Busch. The watchlist imbalance effect in biometric face identification: Comparing theoretical estimates and empiric measurements. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3750–3758, 2021.

[2] John J. Howard, Yevgeniy B. Sirotin, Jerry L. Tipton, and Arun R. Vemury. Quantifying the extent to which race and gender features determine identity in commercial face recognition algorithms. Technical paper series, DHS Science and Technology Directorate, May 2021. https://www.dhs.gov/sites/default/files/publications/21_0922_st_quantifying-commercial-face-recognition-gender-and-race_updated.pdf.

[3] John J. Howard, Yevgeniy B. Sirotin, and Arun R. Vemury. The effect of broad and specific demographic homogeneity on the imposter distributions and false match rates in face recognition algorithm performance. In *Proc. 10th International Conference on Biometric Theory, Applications, and Systems*. IEEE, 2019. https://mdtf.org/publications/broad-and-specific-homogeneity.pdf.

[4] J. Sherrah. False alarm rate: a critical performance measure for face recognition. In *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition.*, pages 189–194, 2004.