

NISTIR 8280

**Ongoing Face Recognition
Vendor Test (FRVT)
Part 3: Demographic Effects**

**Annex 16 : Identification error characteristics by race and
sex**

This document is an annex of NIST Interagency Report 8280:
<https://doi.org/10.6028/NIST.IR.8280>

2019/12/19

NIST
**National Institute of
Standards and Technology**
U.S. Department of Commerce

1 Overview

This annex includes figures that shows rank-based and threshold-based error characteristics for mugshot images from three identification experiments. Each experiment was designed to isolate just one demographic factor from all demographic factors related to race, sex, and age. The experiments also remove any effects from imbalance, specially by composing galleries from two demographic groups in equal proportion. The experiments were designed to be readily interpretable. They were not designed to mimic any specific operational scenario.

The three experiments are:

- ▷ **Black-White Males** The left panels in each figure summarize one-to-many search results for galleries in which all subjects are males in the age range 20 to 40, and the only known covariate is race, here black vs. white. The metadata labels are described in Annex [Annex 1](#).
- ▷ **Black Male-Female** The center panels in each figure summarize one-to-many search results for galleries in which all subjects are black in the age range 20 to 40, and the only known covariate is sex.
- ▷ **White Male-Female** As previously, with white instead of black subjects.

In the above we refer the “known” covariates. The word here is used to express that a large number of (we assume) less influential factors could be at play: e.g. the height of the subject¹. In virtually all modeling efforts, e.g. building regression models, there are unknown independent variables. Full knowledge of *all* covariates would mean the model would have the complexity of the system itself.

2 Data

The mugshot dataset and its metadata is described in Annex [Annex 1](#).

The figures in this annex were generated from the three experiments detailed in Table 1.

3 Fixed Threshold

A false positive is declared if a non-mated search yields *any* candidates (one or more) with similarity at or above, a threshold. The error tradeoff characteristics sweep the threshold over at least 1280 representative threshold values, to effect FPIR statements over four decades from one in ten thousand, up to one. Doing this for the various demographic curves individually yields the colored traces in the second row of panels.

The error tradeoff characteristics also include three grey lines. These connect points of fixed threshold, and thereby expose difference in FNIR and FPIR between two demographic groups.

The threshold value could be any value germane to that comparison algorithm. Here we set it to achieve one of three FPIR values on, arbitrarily, white males. The three FPIR values are 0.003, 0.03, 0.3.

¹This may matter because pitch angle relative to the camera is known to affect most face recognition algorithms, and many cameras would not be adjusted for height for each subject. The mugshot photographic configuration includes a larger stand-off distance, around 2 meters, so pitch angle variation is expected to be small. In addition, subject are usually directed to look at the camera.

		ENROLLMENT			SEARCH			
TYPE SEE	POPULATION			MATE		NON-MATE		
SECTION ??	FILTER	N-SUBJECTS	N-IMAGES	N-SUBJECTS	N-IMAGES	N-SUBJECTS	N-IMAGES	
Demographics								
1	RECENT	MALE, AGE21-40, ΔT ≤ 5 YR, BLACK AND WHITE BALANCED	800 000 B + 800 000 W	800 000 B + 800 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W	100 000 B + 100 000 W
2	RECENT	WHITE, AGE21-40, ΔT ≤ 5 YR, MALE AND FEMALE BALANCED	800 000 F + 800 000 M	800 000 F + 800 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M
3	RECENT	BLACK, AGE21-40, ΔT ≤ 5 YR, MALE AND FEMALE BALANCED	500 000 F + 500 000 M	500 000 F + 500 000 M	97 000 F + 97 000 M	97 000 F + 97 000 M	100 000 F + 100 000 M	100 000 F + 100 000 M

Table 1: **Enrollment and search sets.** Each row summarizes one identification trial. Unless stated otherwise, all entries refer to mugshot images. The term “natural” means that subjects were selected without heed to demographics, i.e. in the distribution native to this dataset. The probe images were collected in a different calendar year to the enrollment image. Missing values in rows 2-12 are the same as in row 1.

4 Plot

Each page contains two figures corresponding to two algorithm. Each figure contains two rows. The first shows FNIR as a function of rank, a metric appropriate to investigate “lead generation” applications. The second rows shows FNIR as a function of FPIR, plotted parametrically on threshold, T. Higher thresholds and lower FPIR are at the left of each plot.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

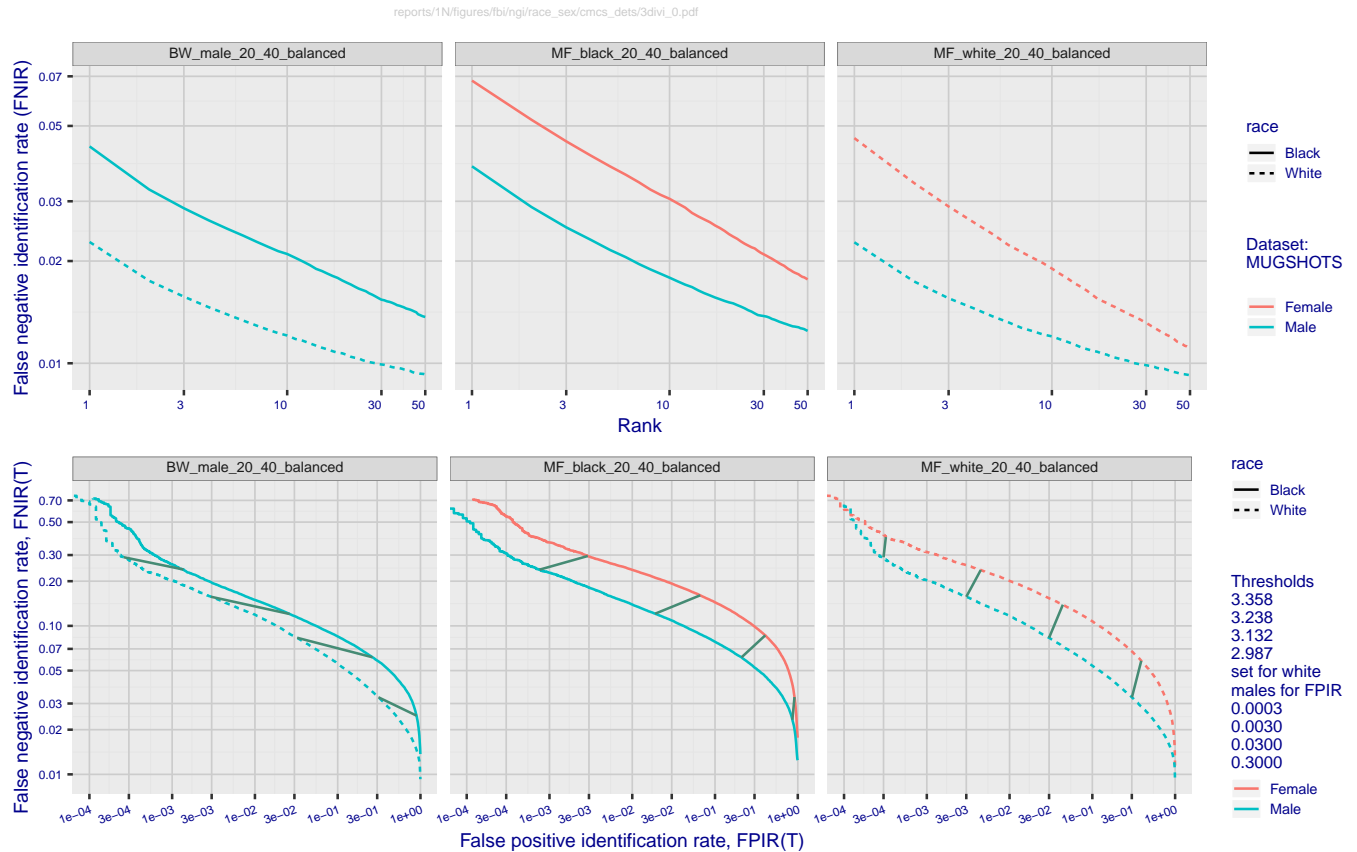


Figure 1: FNIR by sex and race for mugshot, 3divi-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

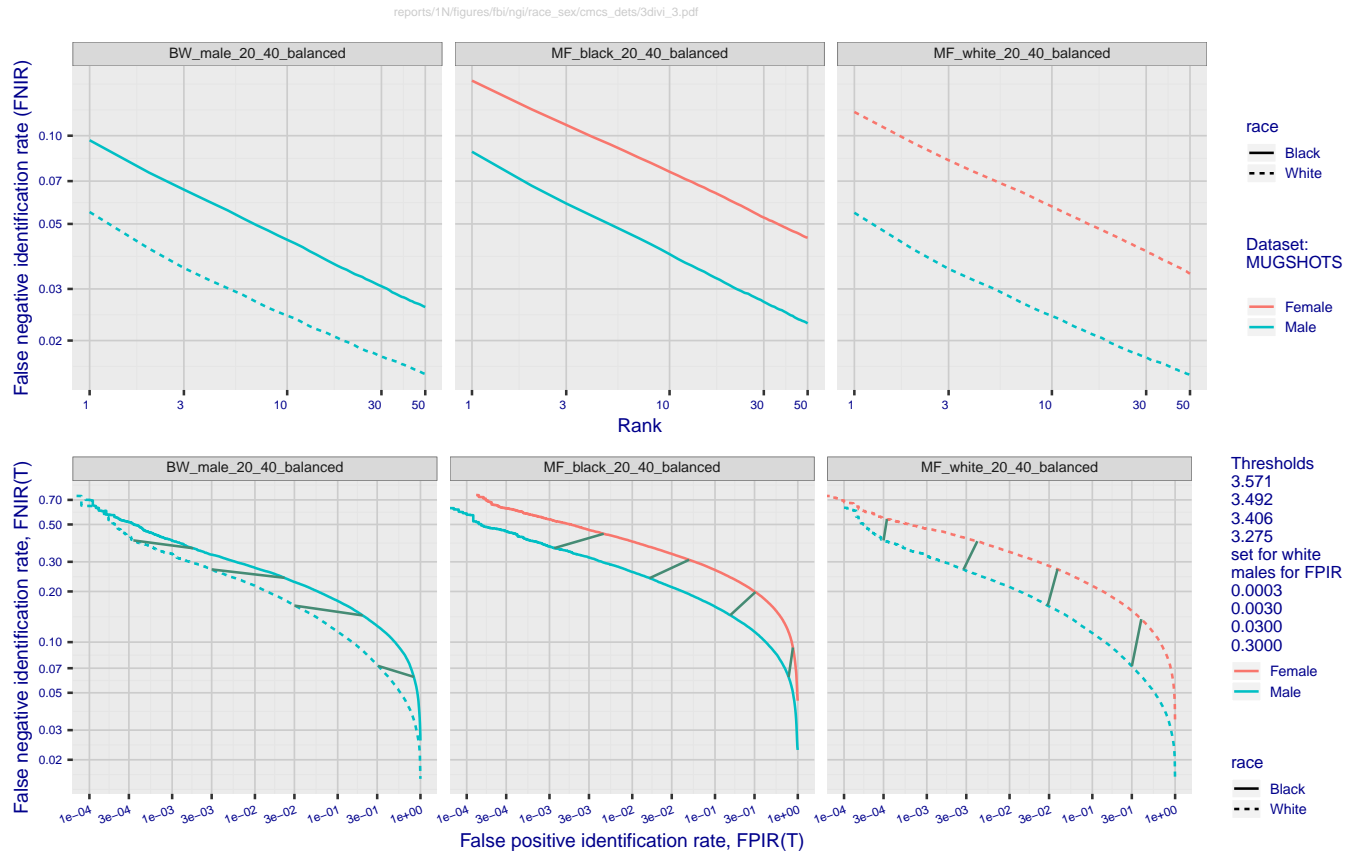


Figure 2: FNIR by sex and race for mugshot, 3divi-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

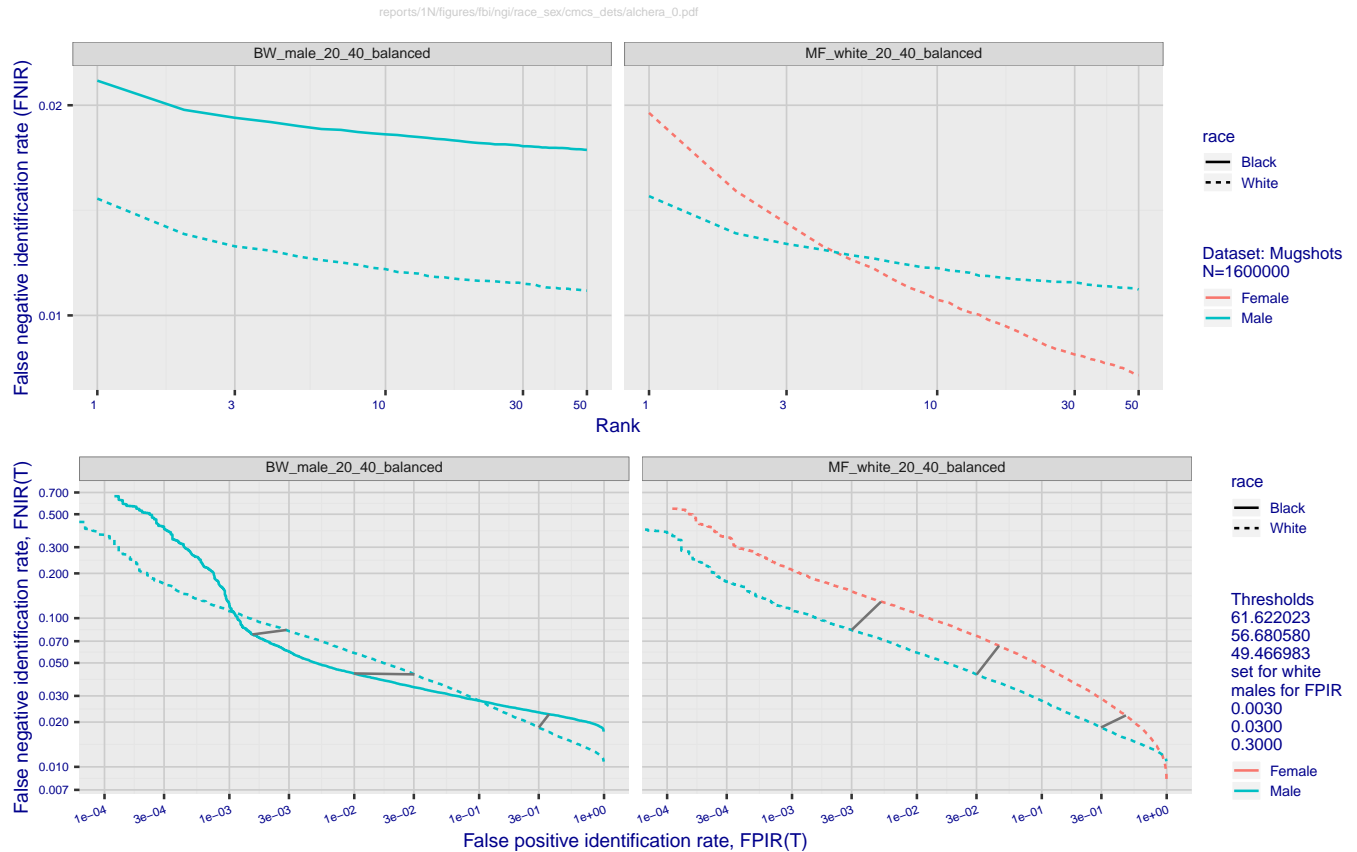


Figure 3: FNIR by sex and race for mugshot, alchera-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

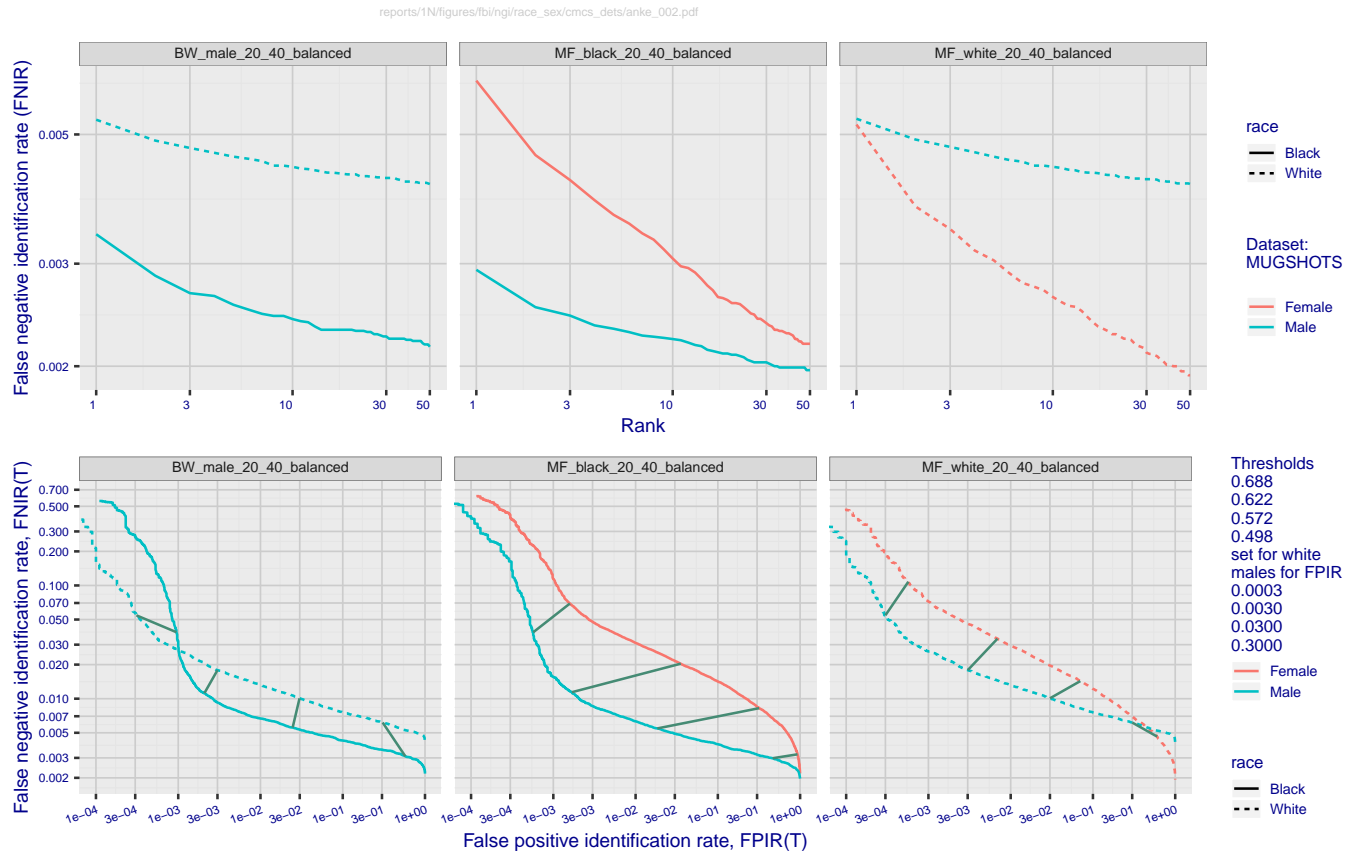


Figure 4: FNIR by sex and race for mugshot, anke-002. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

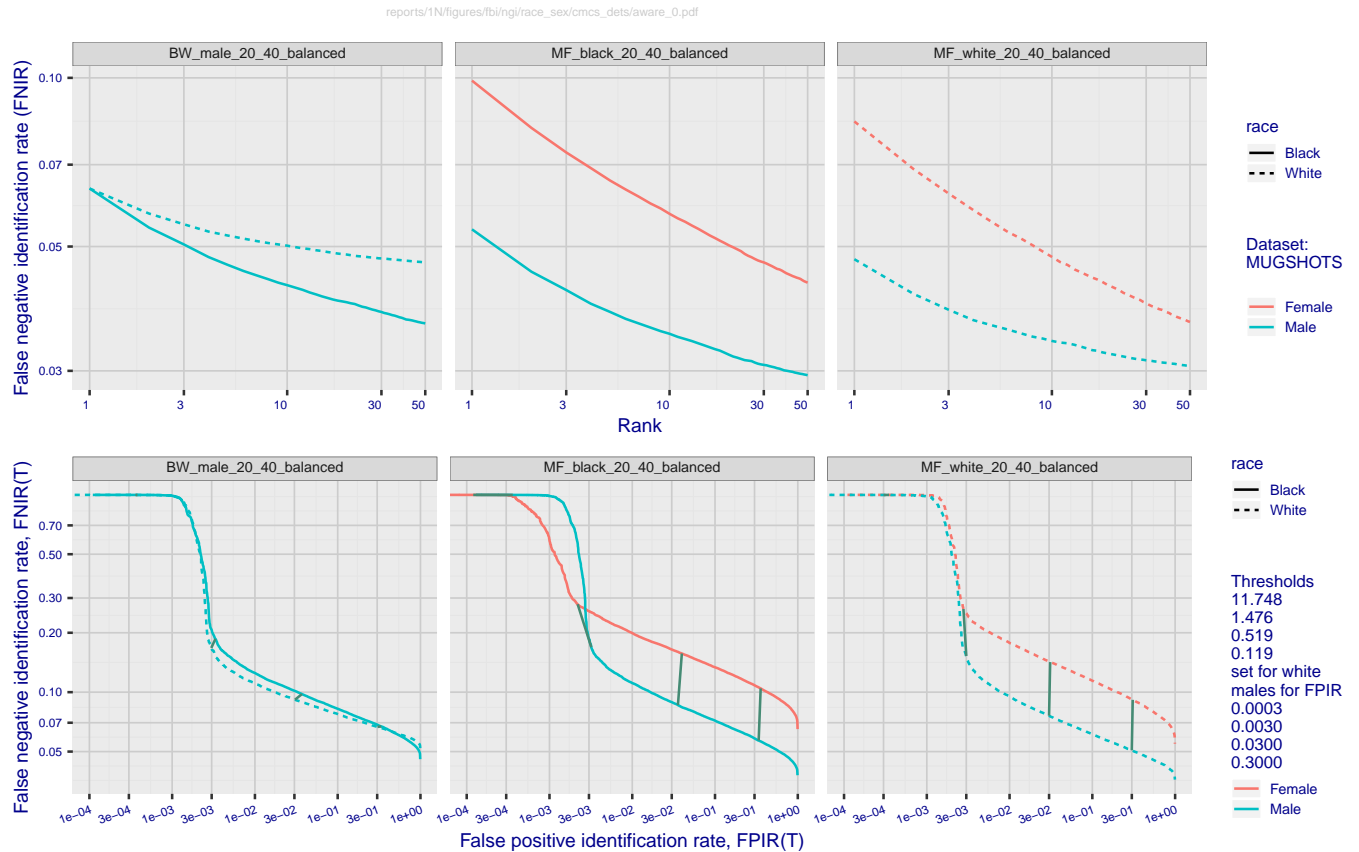


Figure 5: FNIR by sex and race for mugshot, aware-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

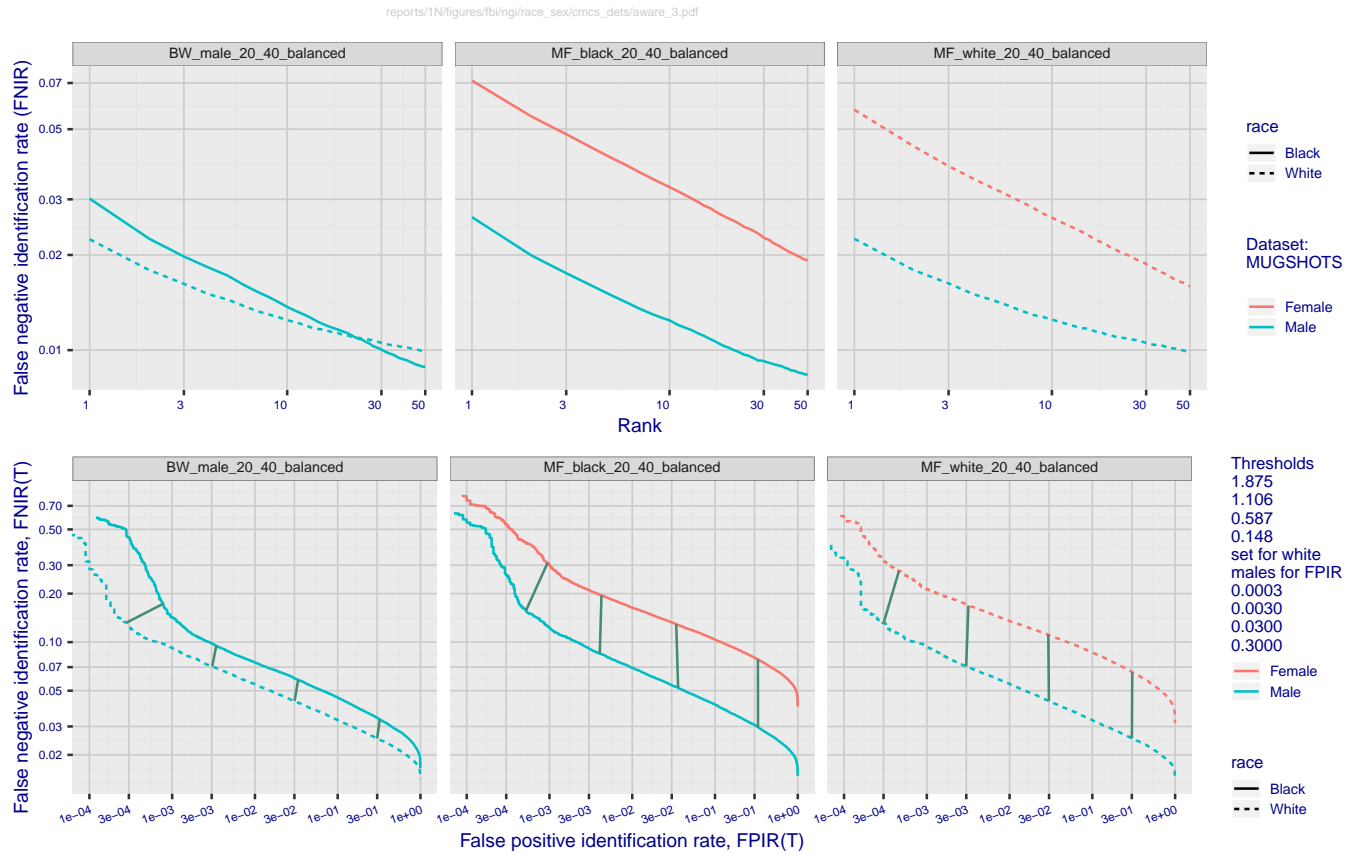


Figure 6: FNIR by sex and race for mugshot, aware-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

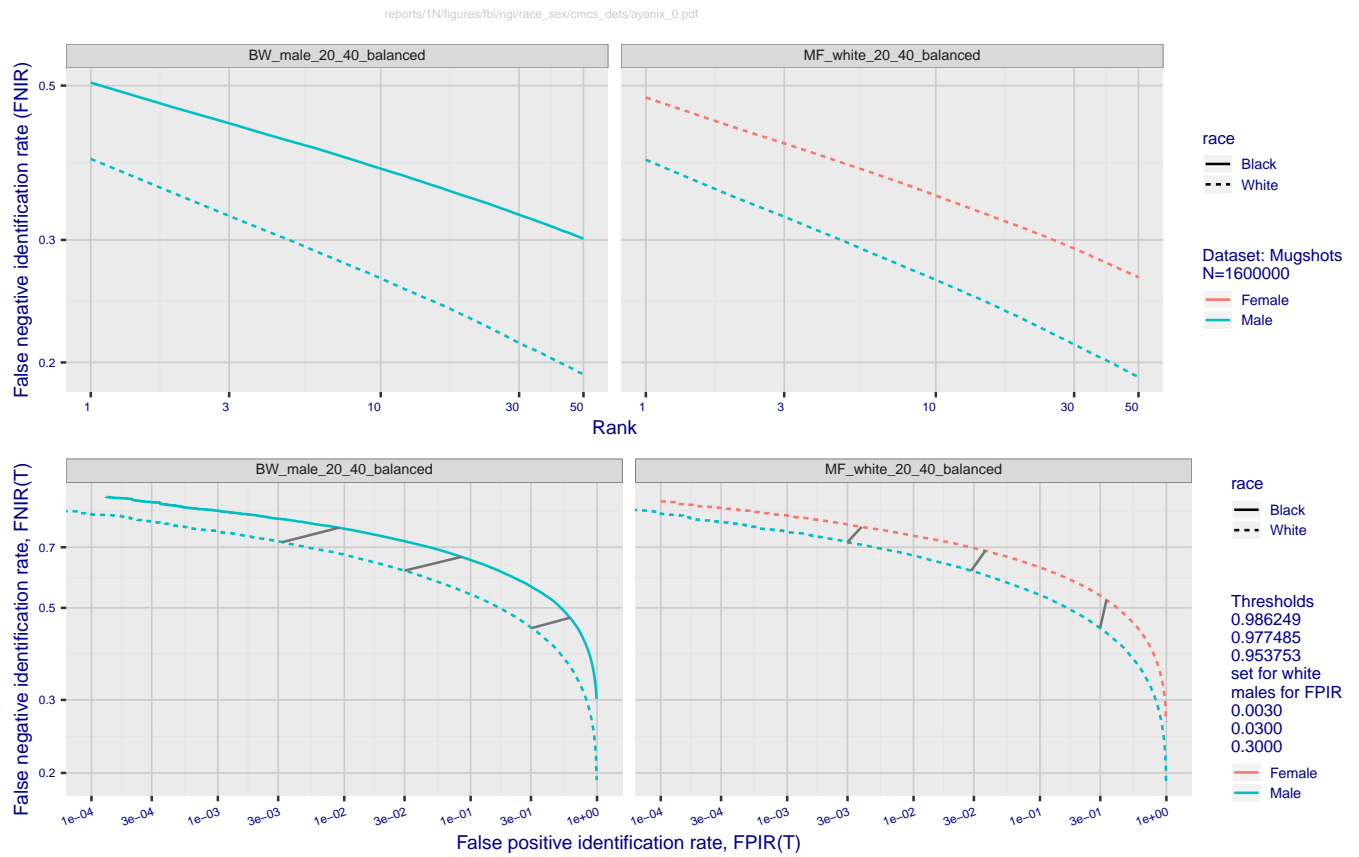


Figure 7: FNIR by sex and race for mugshot, ayonix-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

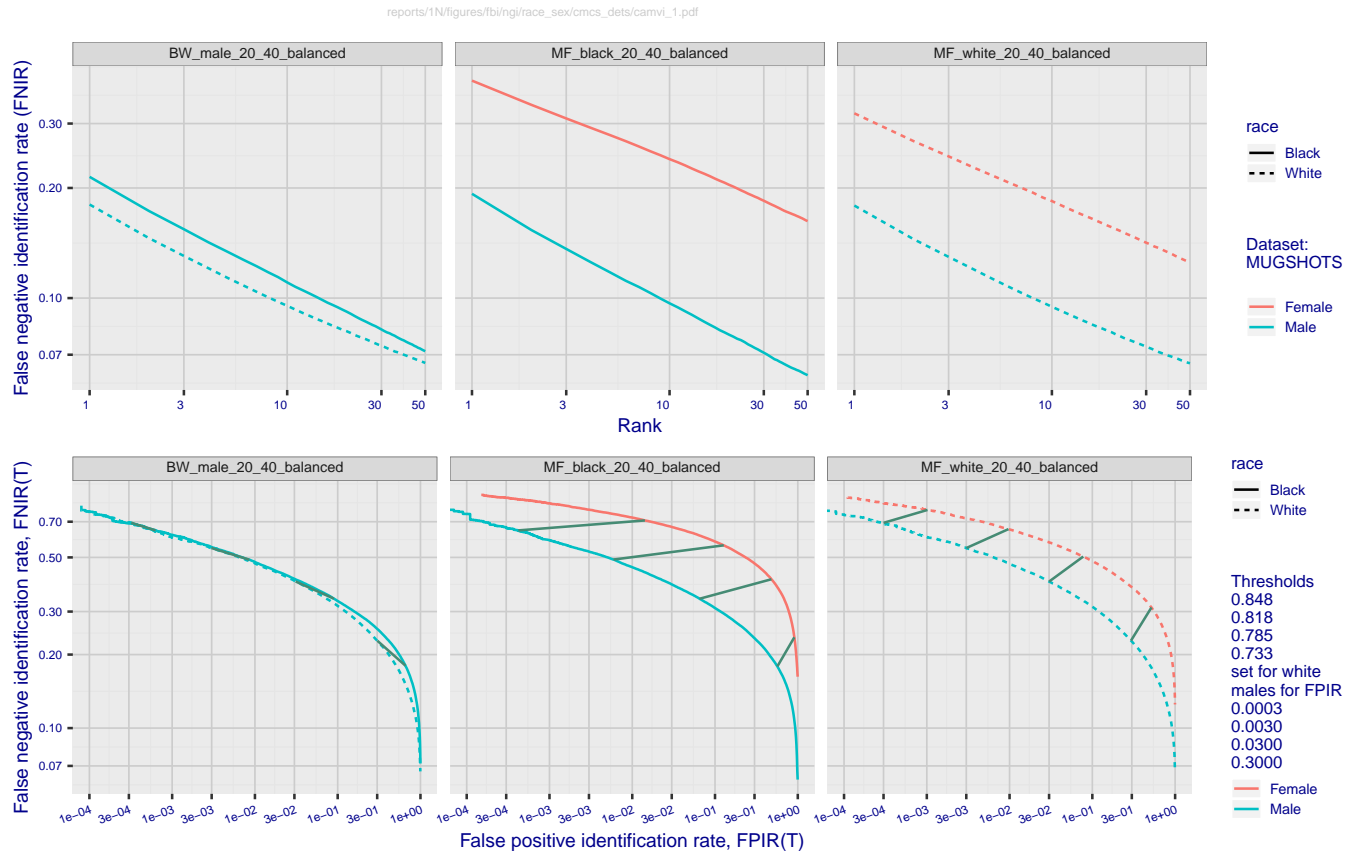


Figure 8: FNIR by sex and race for mugshot, camvi-1. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

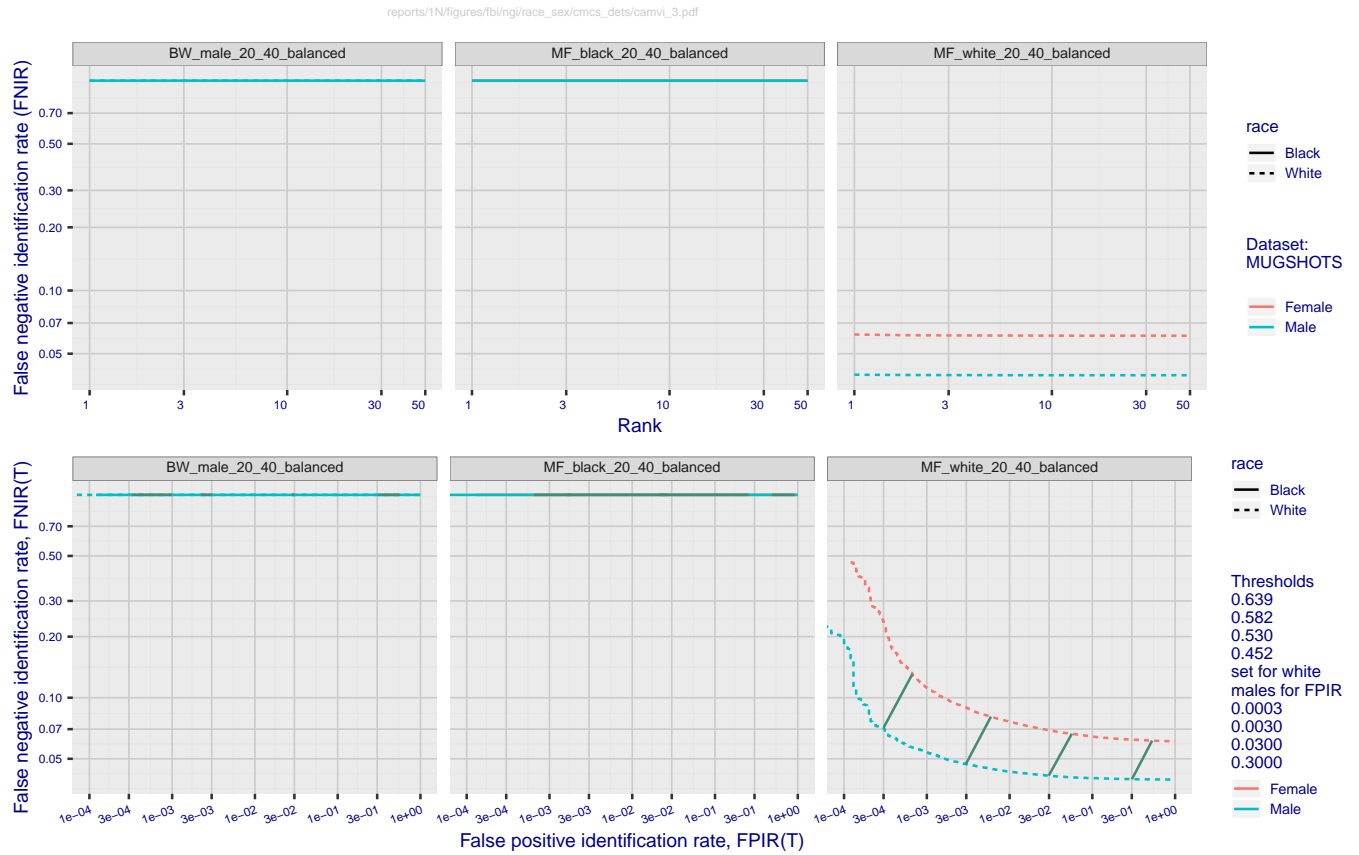


Figure 9: FNIR by sex and race for mugshot, camvi-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

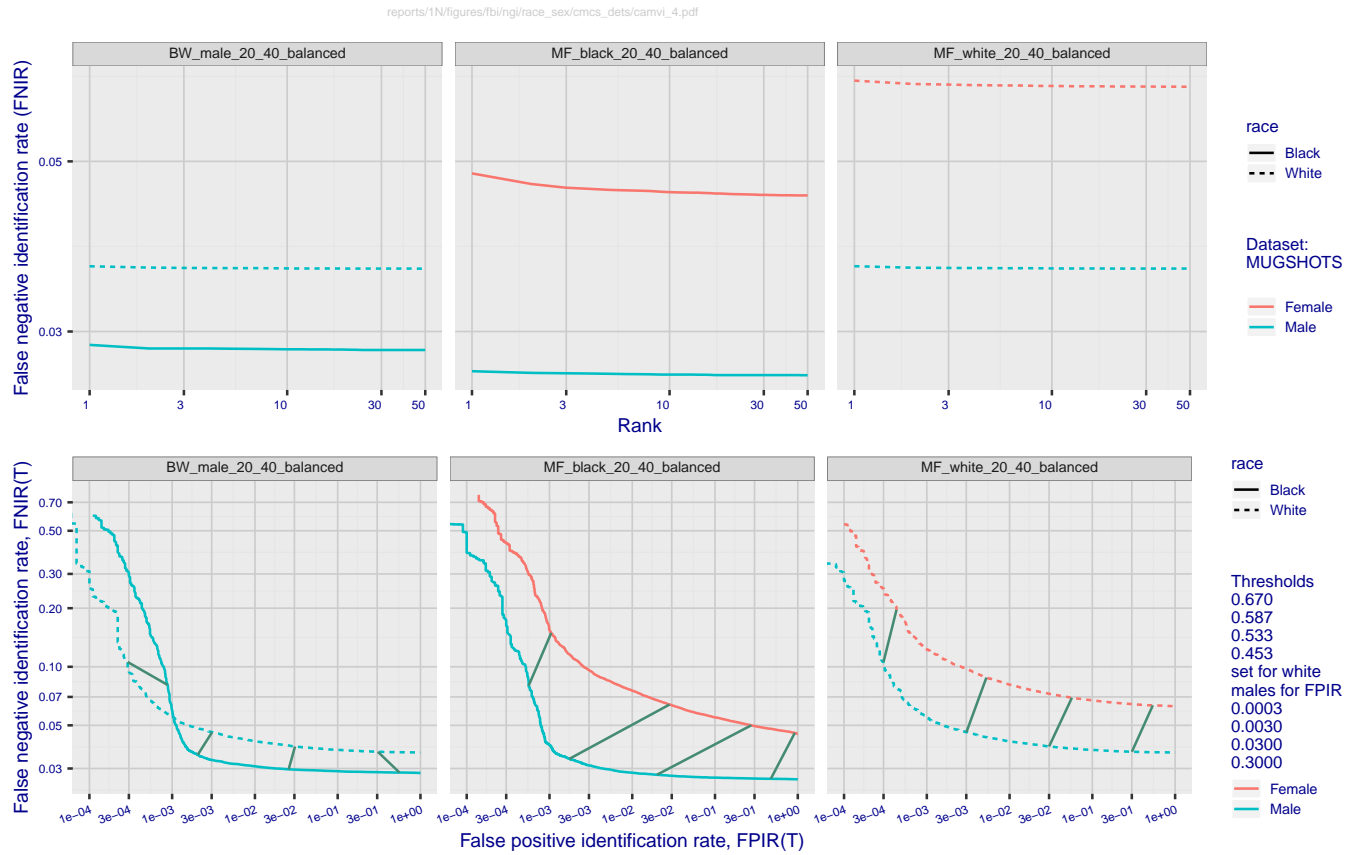


Figure 10: FNIR by sex and race for mugshot, camvi-4. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

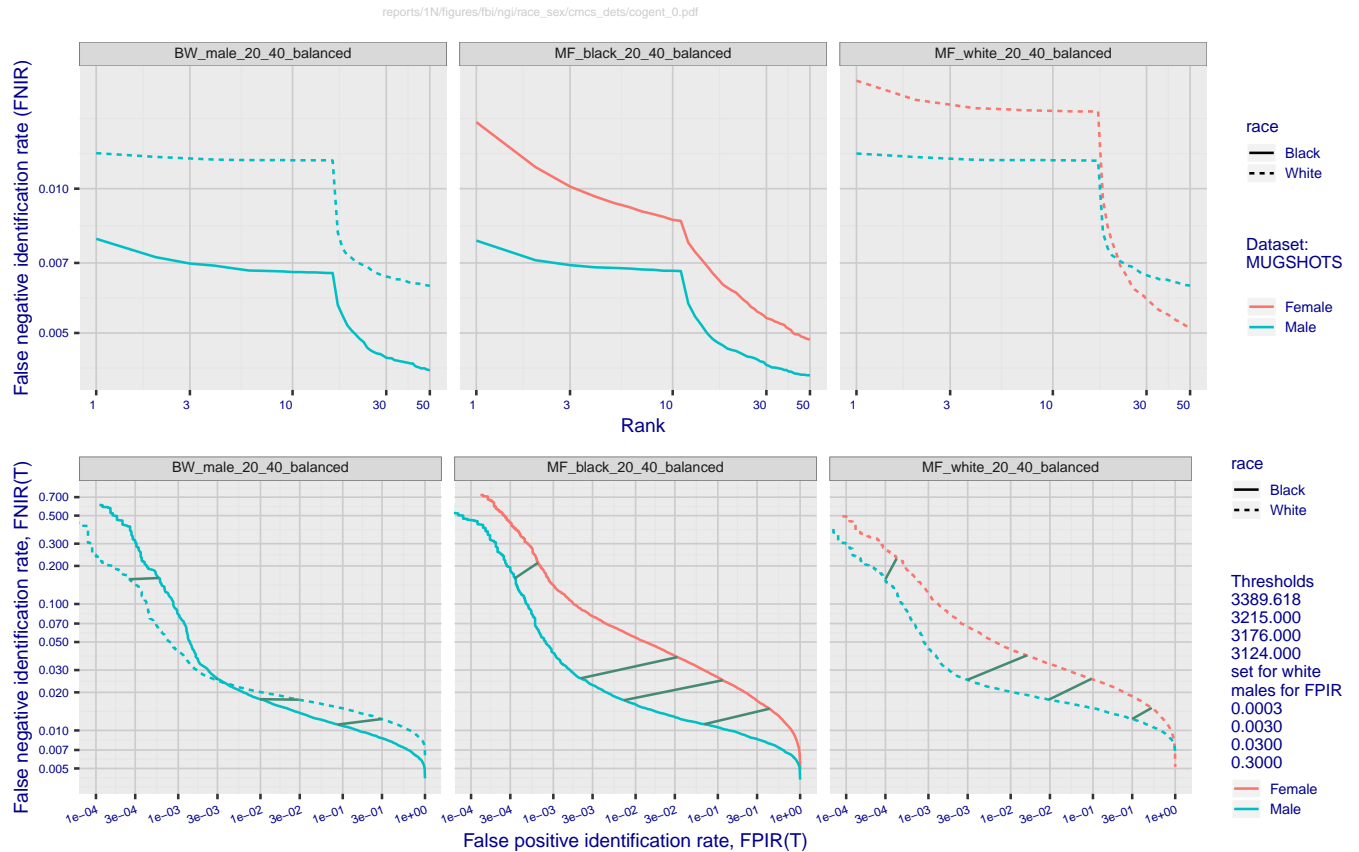


Figure 11: FNIR by sex and race for mugshot, cogent-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

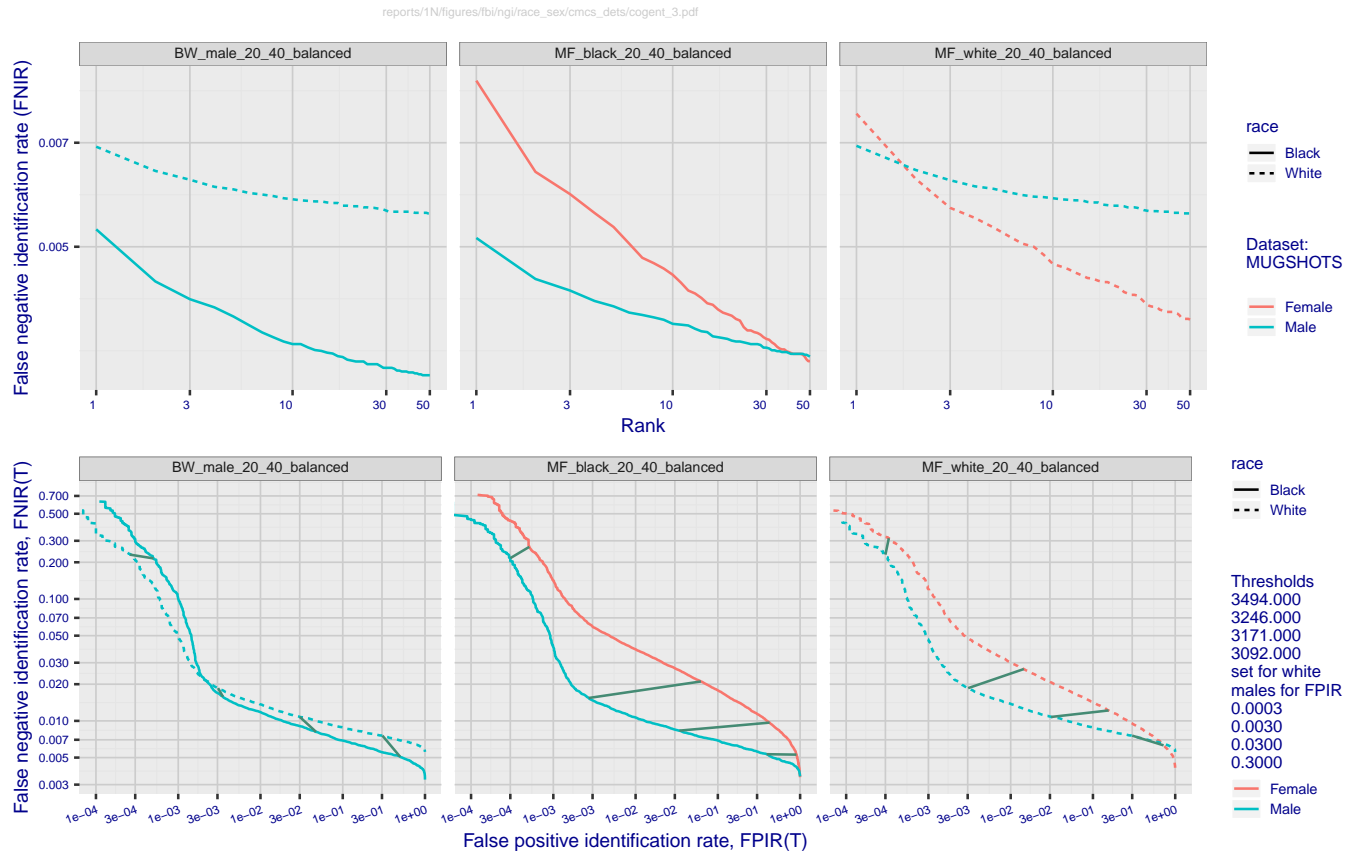


Figure 12: FNIR by sex and race for mugshot, cogent-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

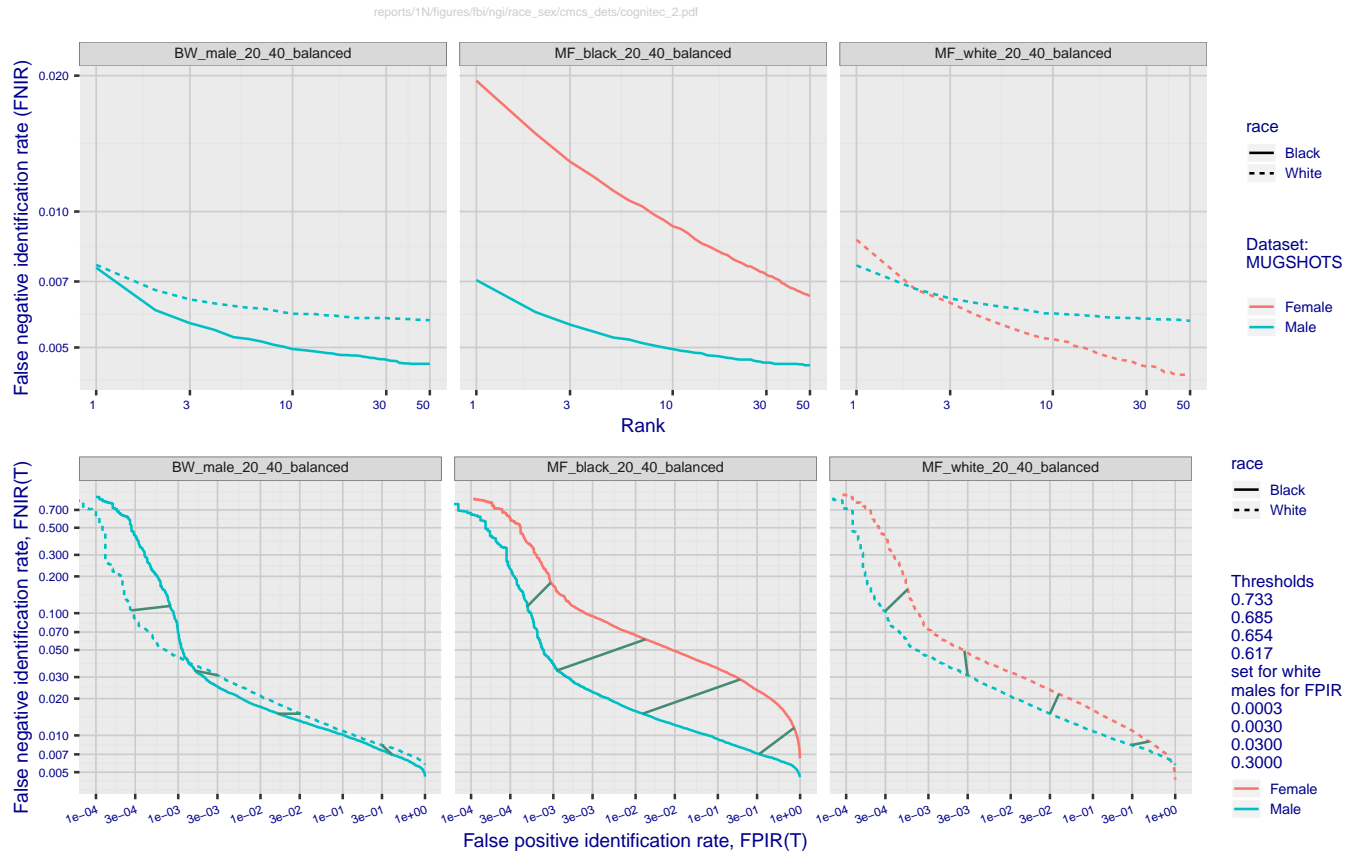


Figure 13: FNIR by sex and race for mugshot, cognitec-2. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

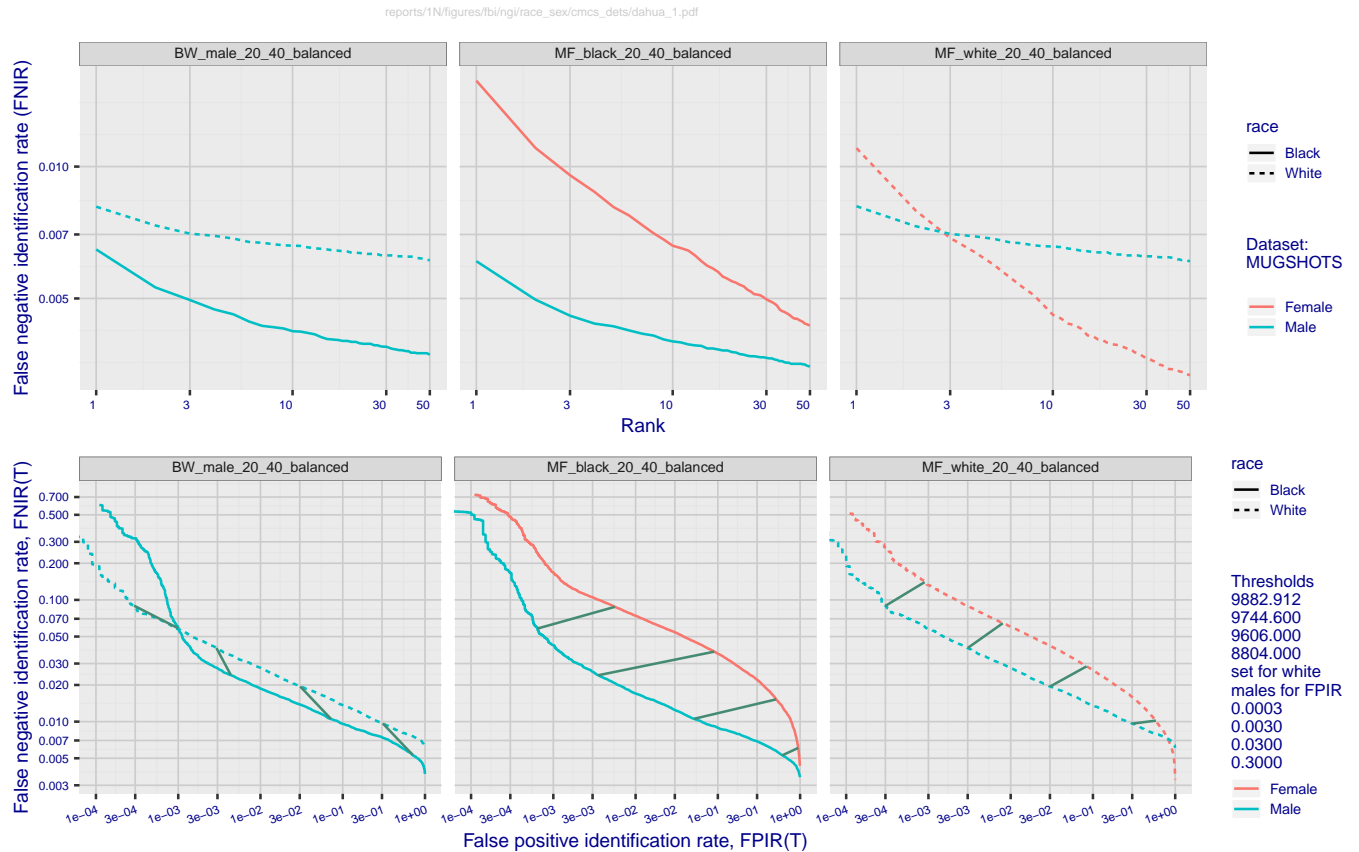


Figure 14: FNIR by sex and race for mugshot, dahua-1. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

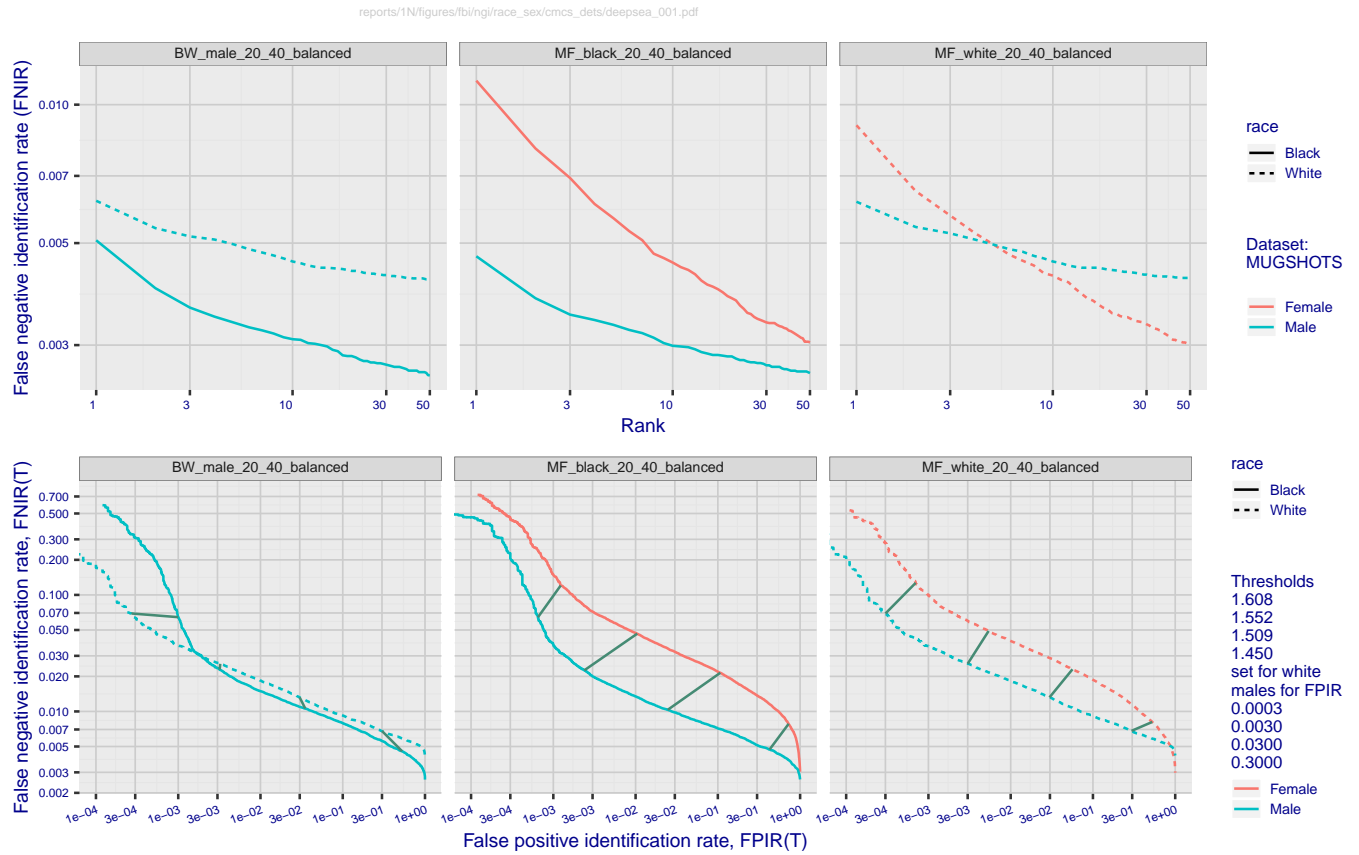


Figure 15: FNIR by sex and race for mugshot, deepsea-001. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

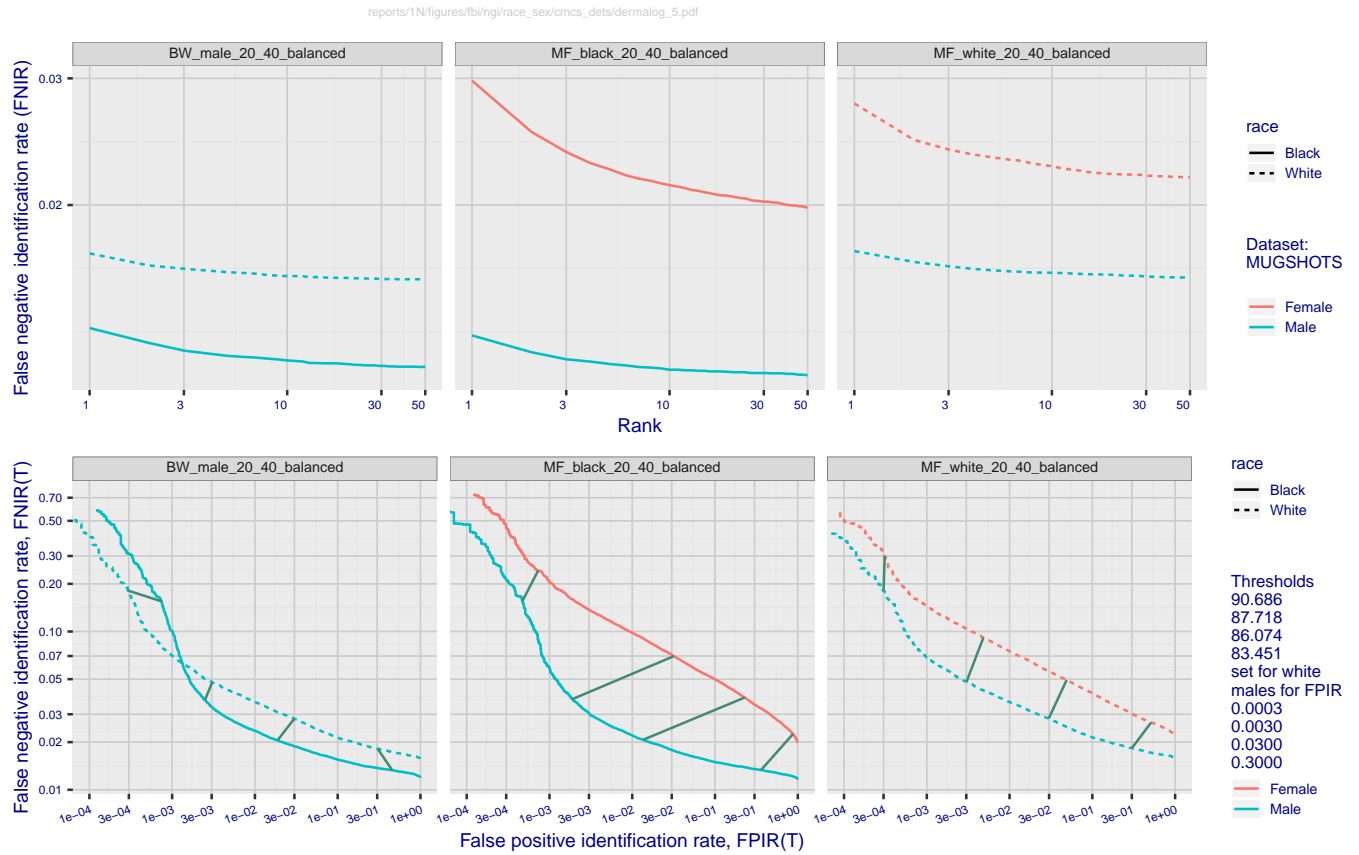


Figure 16: FNIR by sex and race for mugshot, dermalog-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

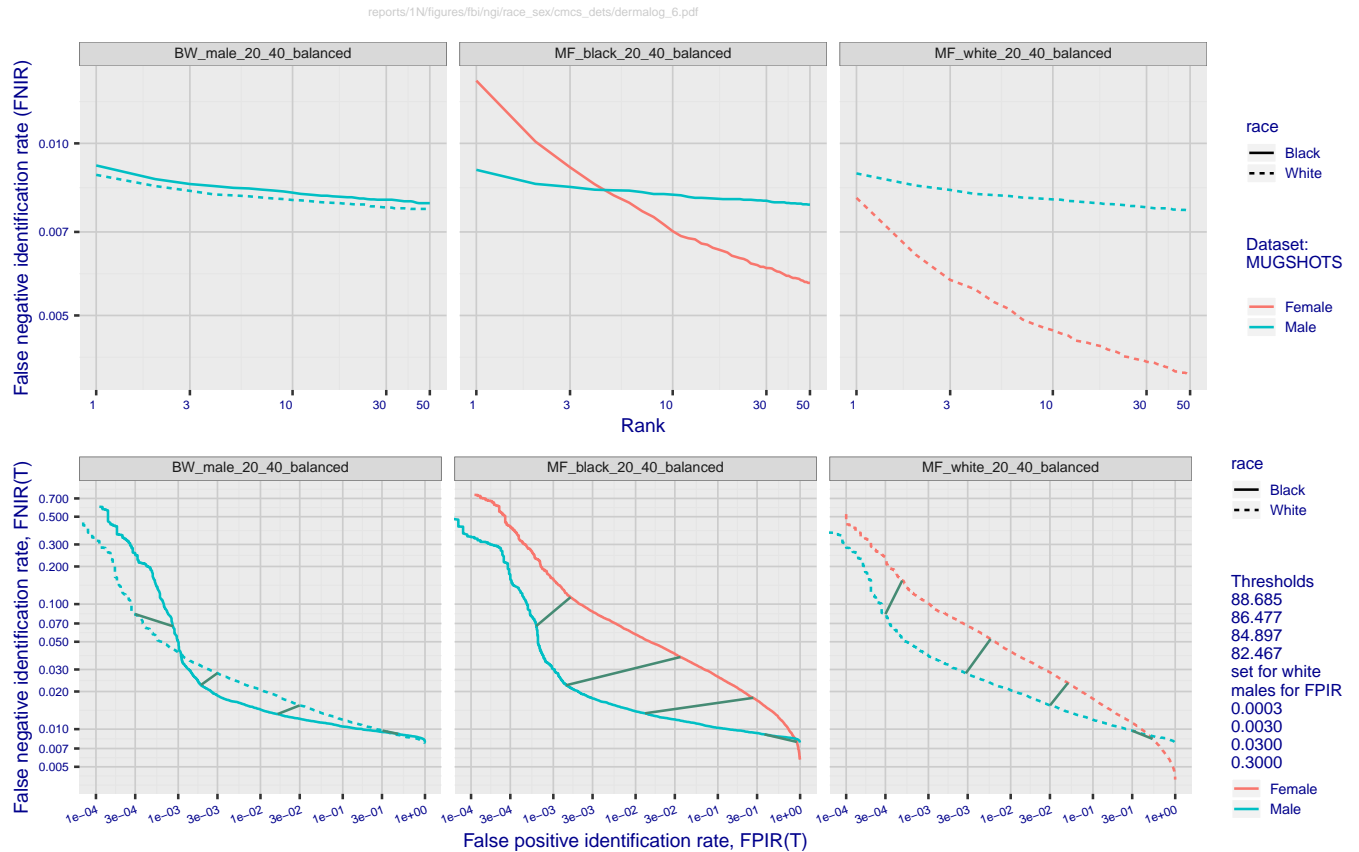


Figure 17: FNIR by sex and race for mugshot, dermalog-6. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

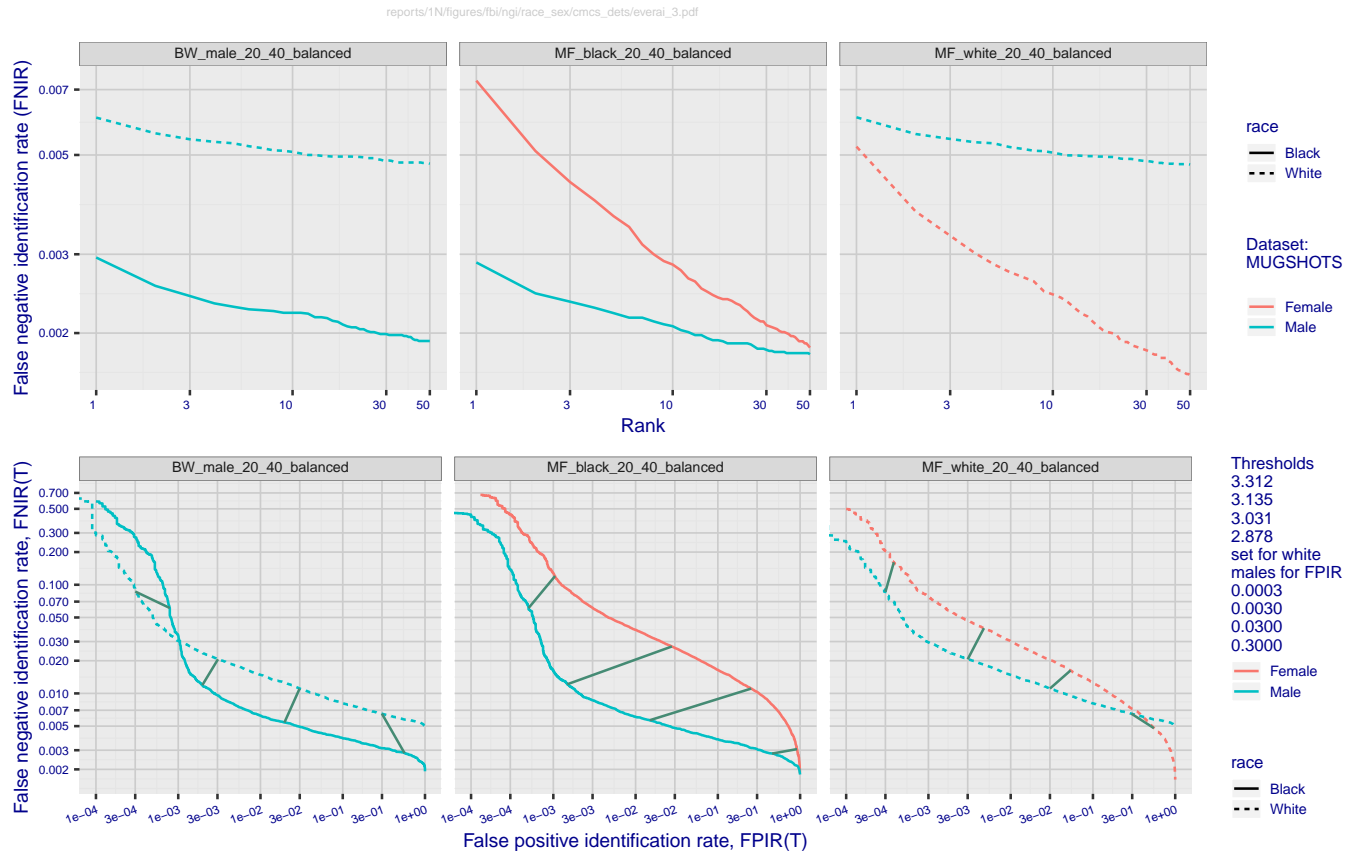


Figure 18: FNIR by sex and race for mugshot, everai-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

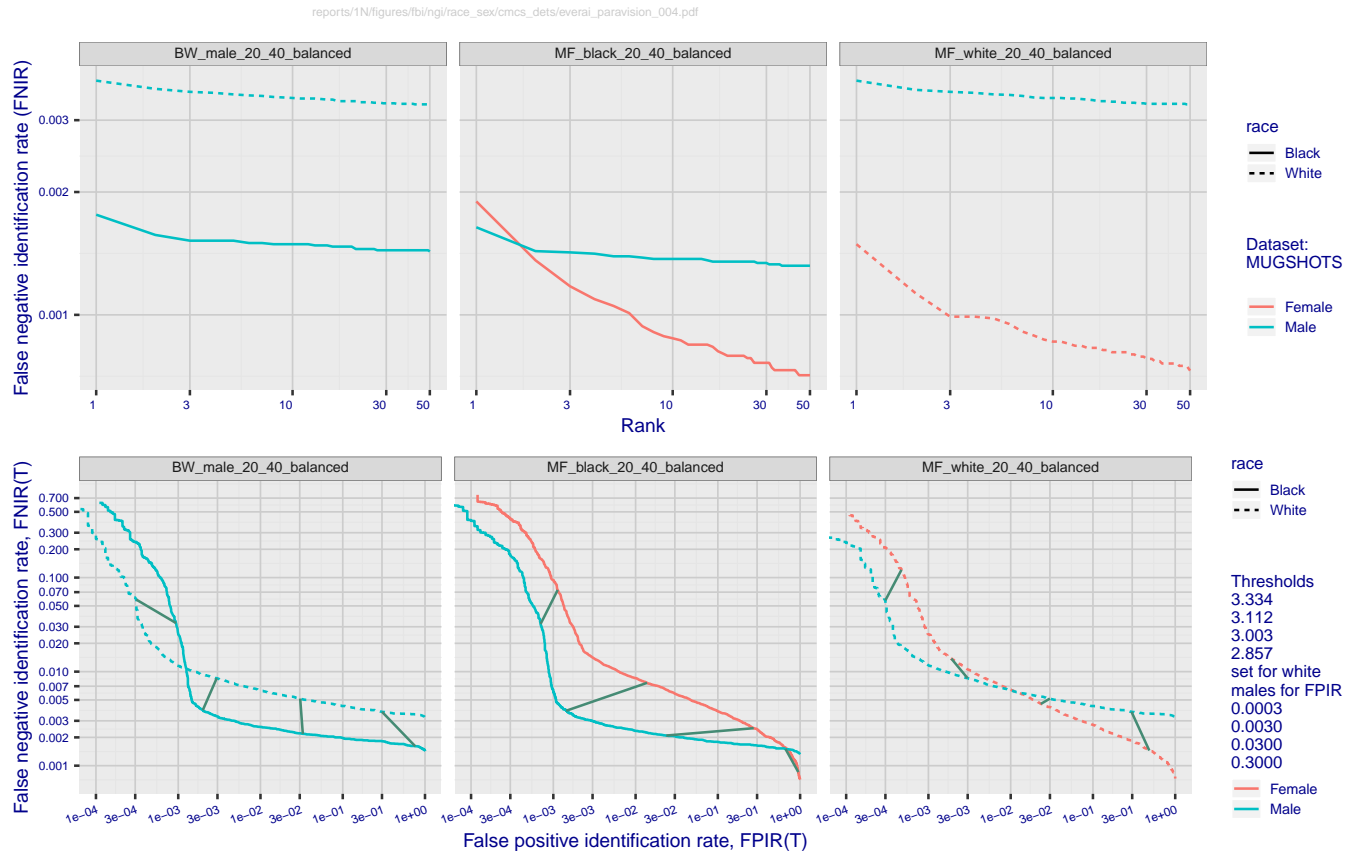


Figure 19: FNIR by sex and race for mugshot, everai-paravision-004. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

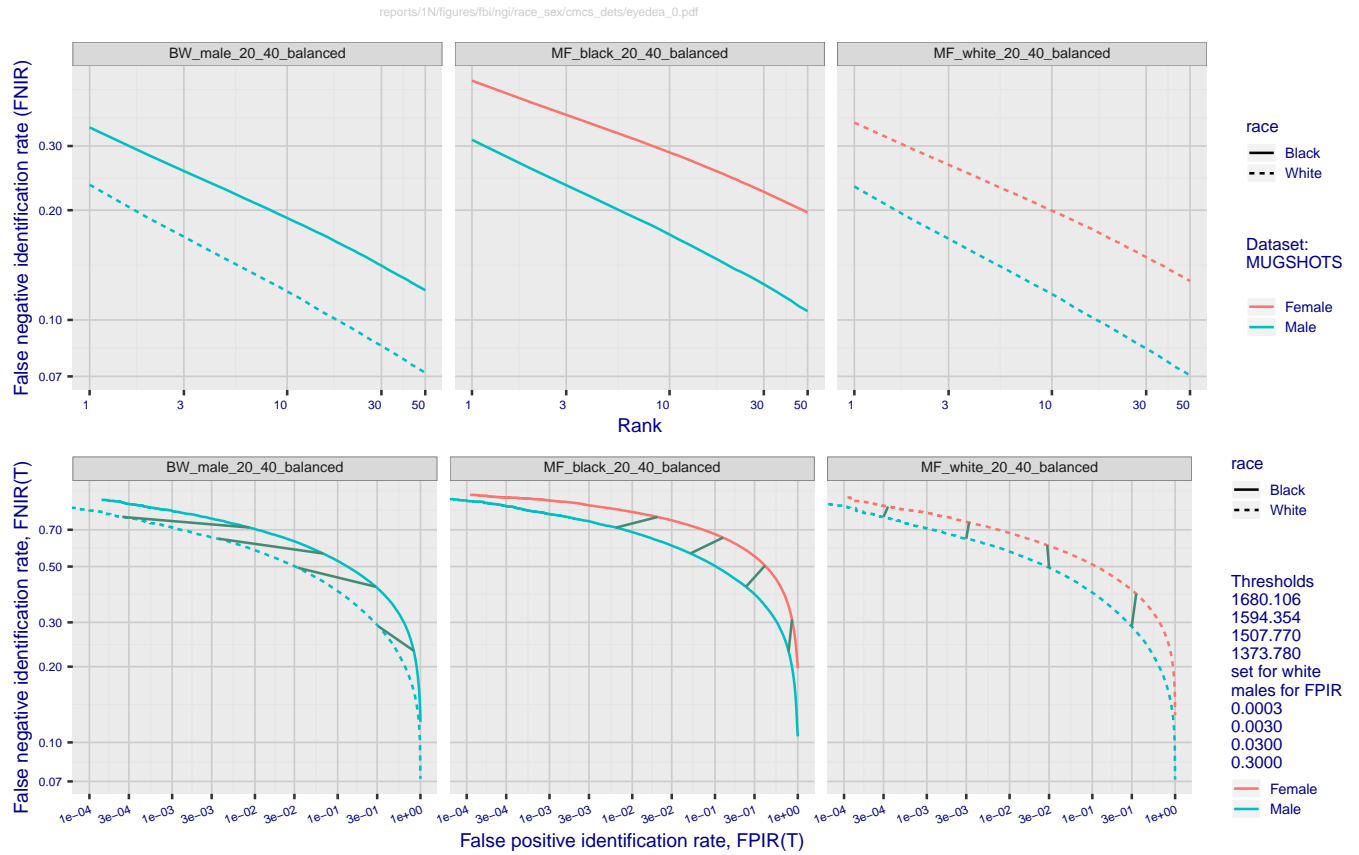


Figure 20: FNIR by sex and race for mugshot, eyede0-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

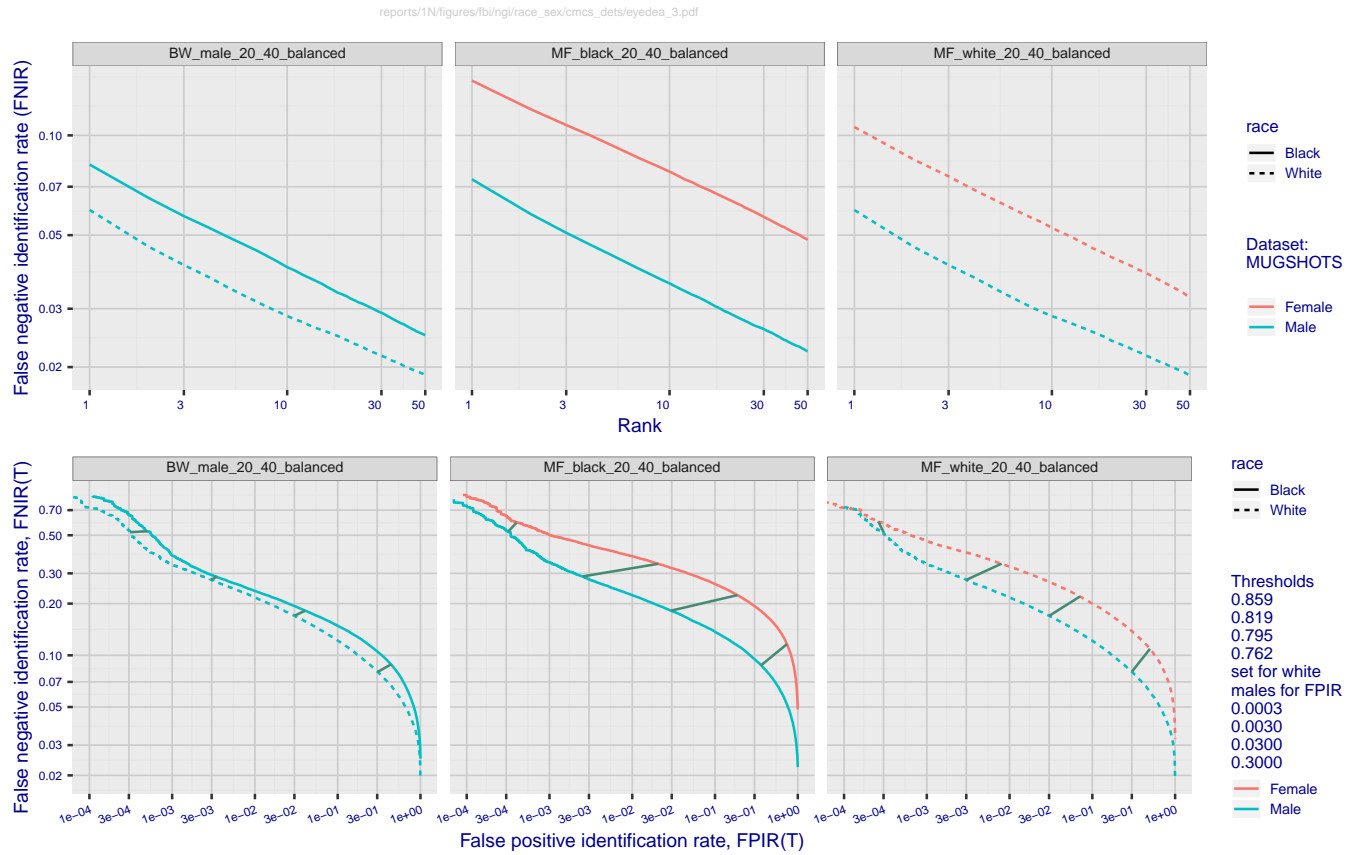


Figure 21: FNIR by sex and race for mugshot, eyede3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

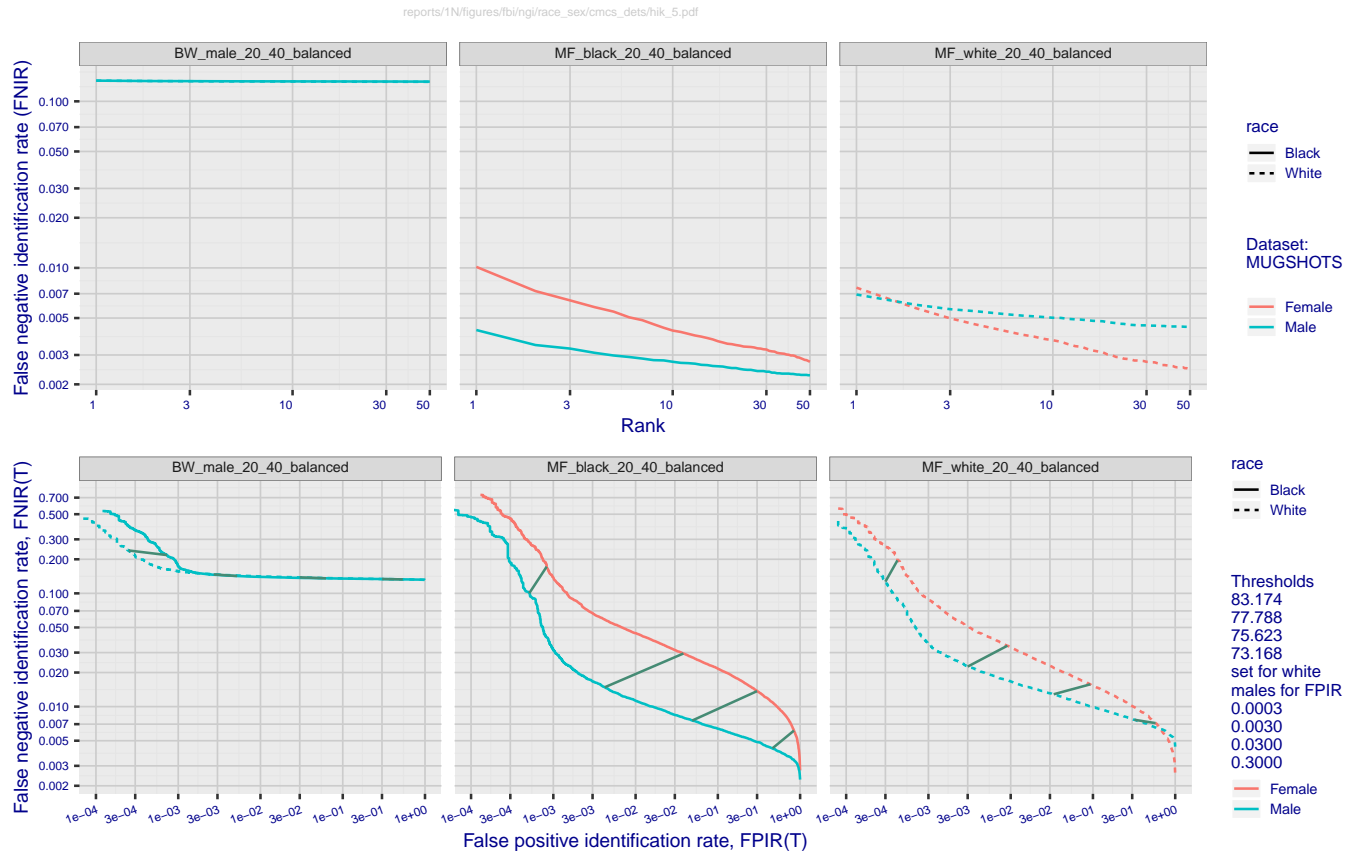


Figure 22: FNIR by sex and race for mugshot, hik-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

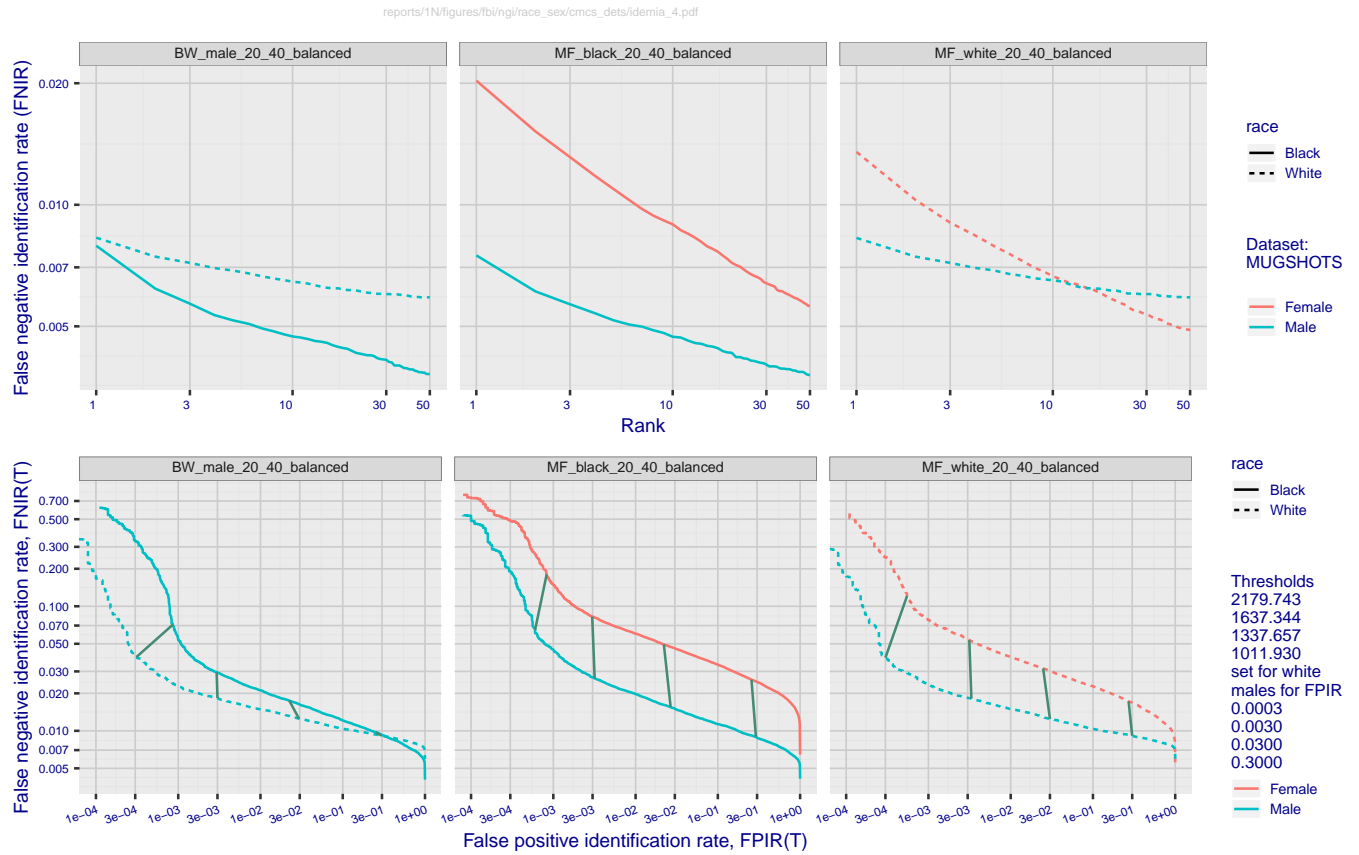


Figure 23: FNIR by sex and race for mugshot, idemia-4. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

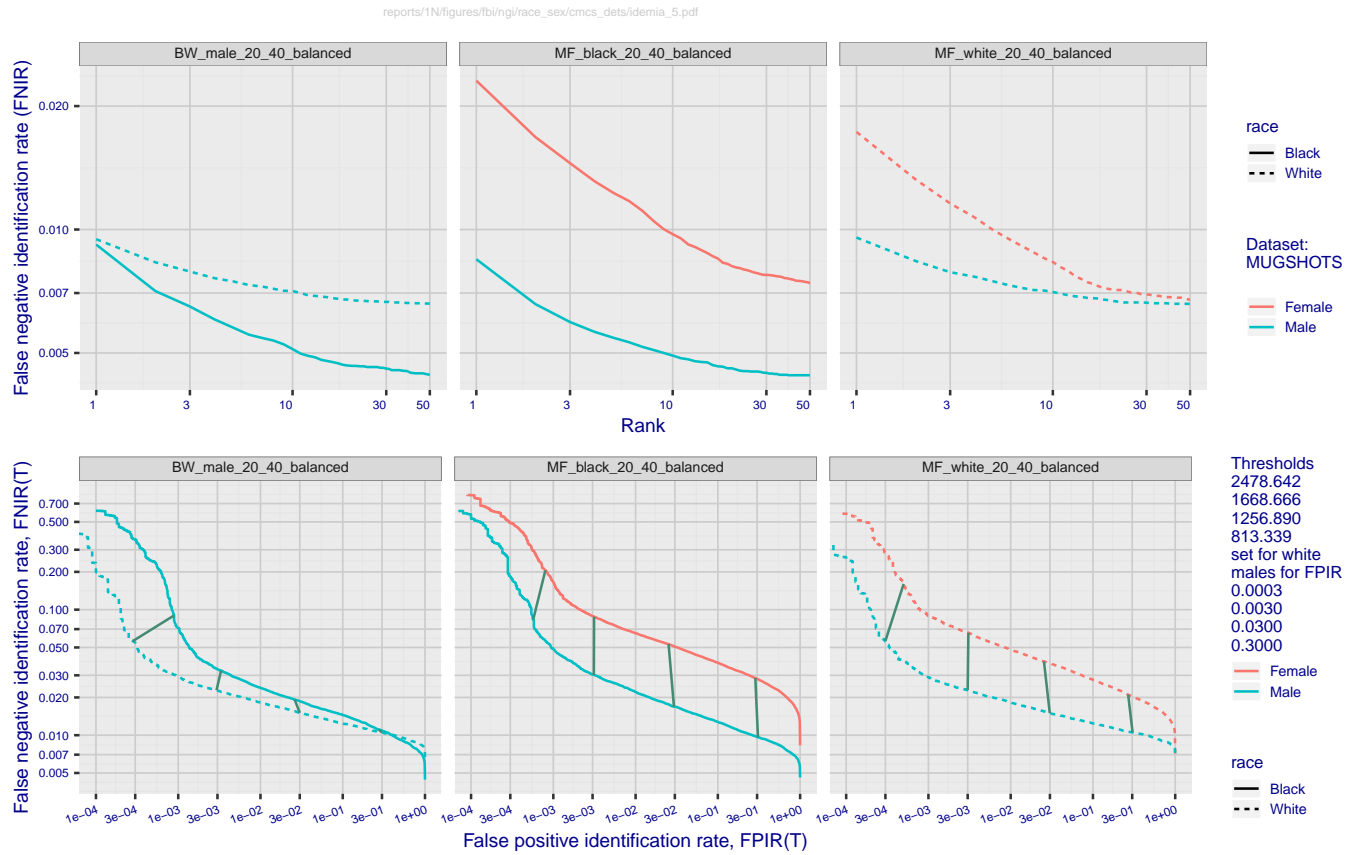


Figure 24: FNIR by sex and race for mugshot, idemia-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

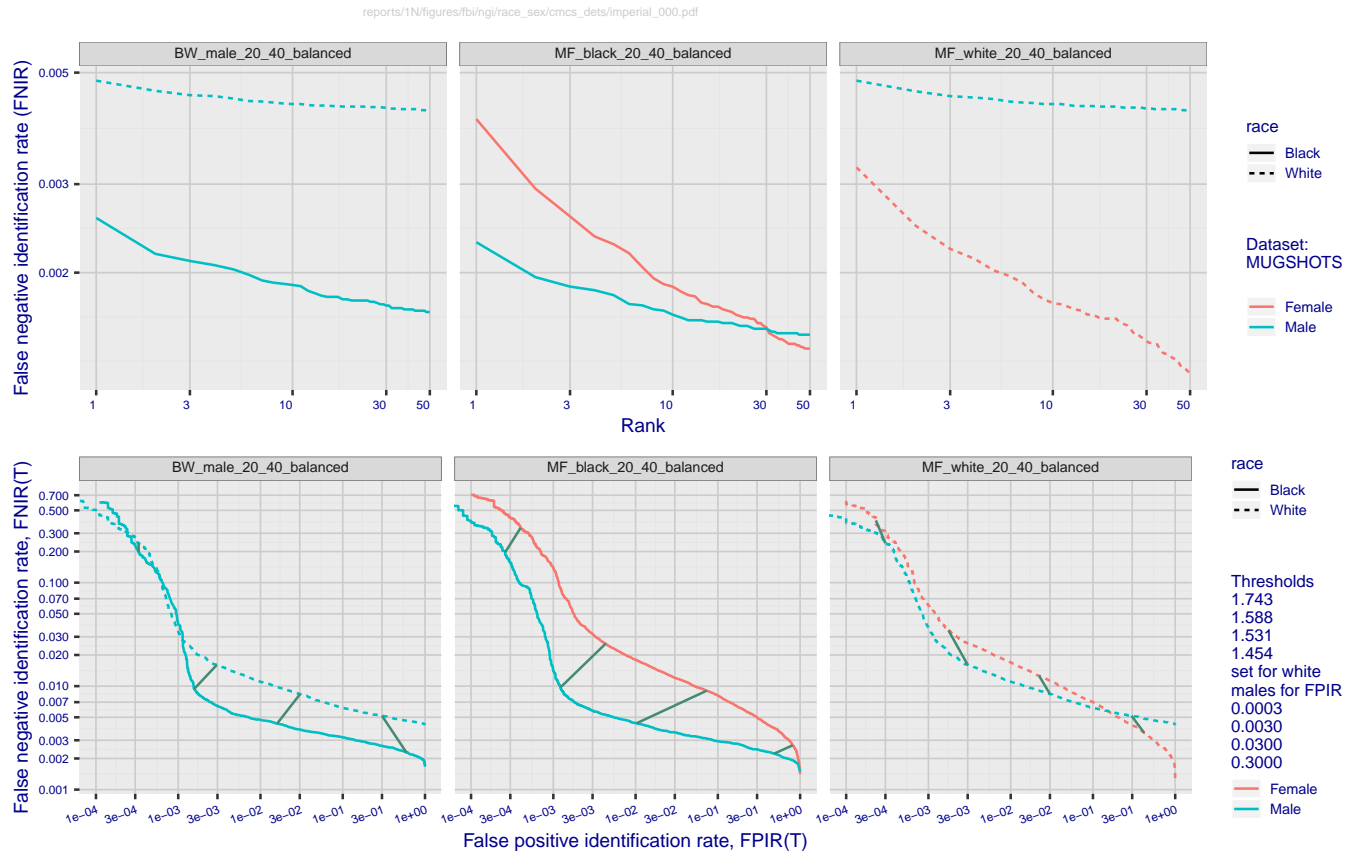


Figure 25: FNIR by sex and race for mugshot, imperial-000. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

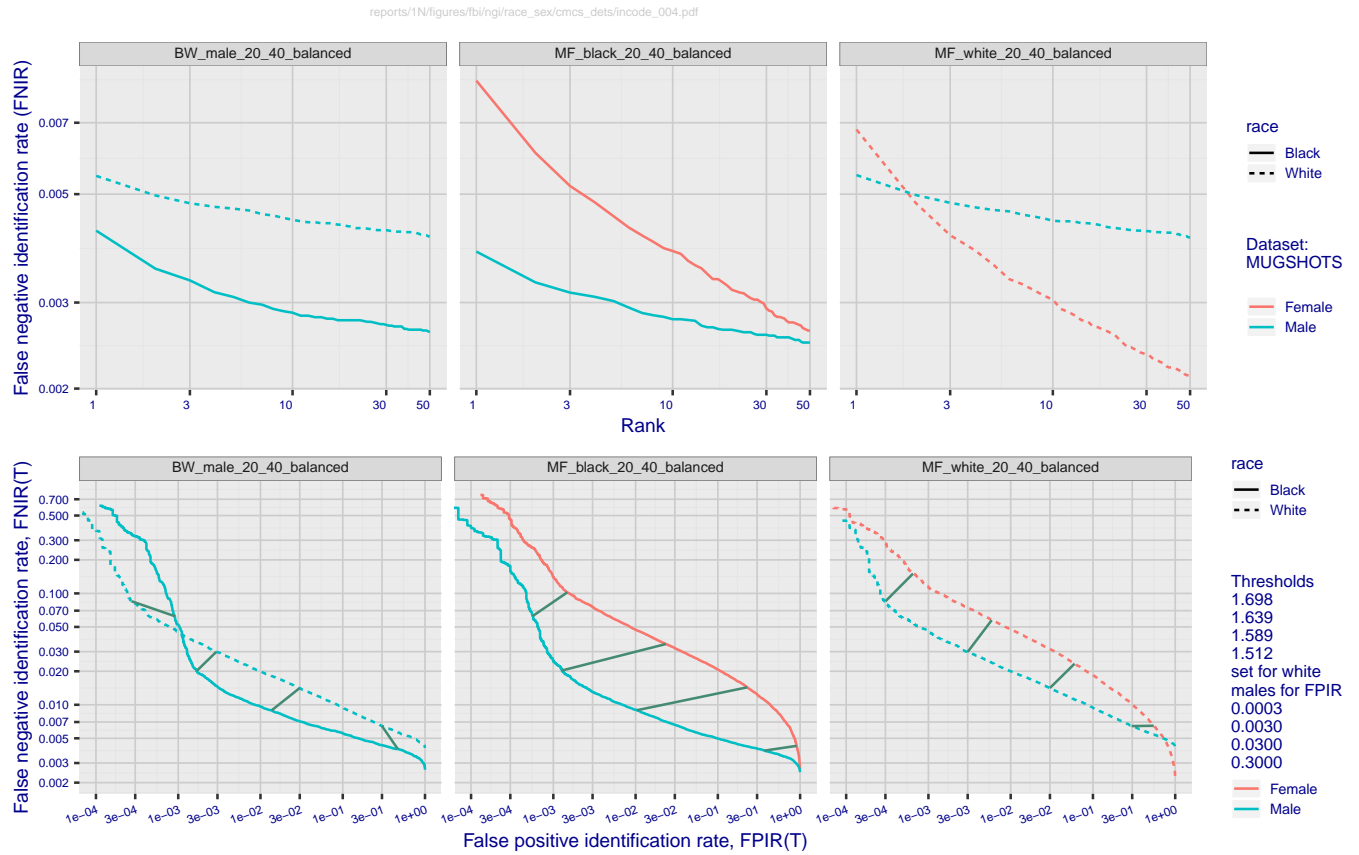


Figure 26: FNIR by sex and race for mugshot, incode-004. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

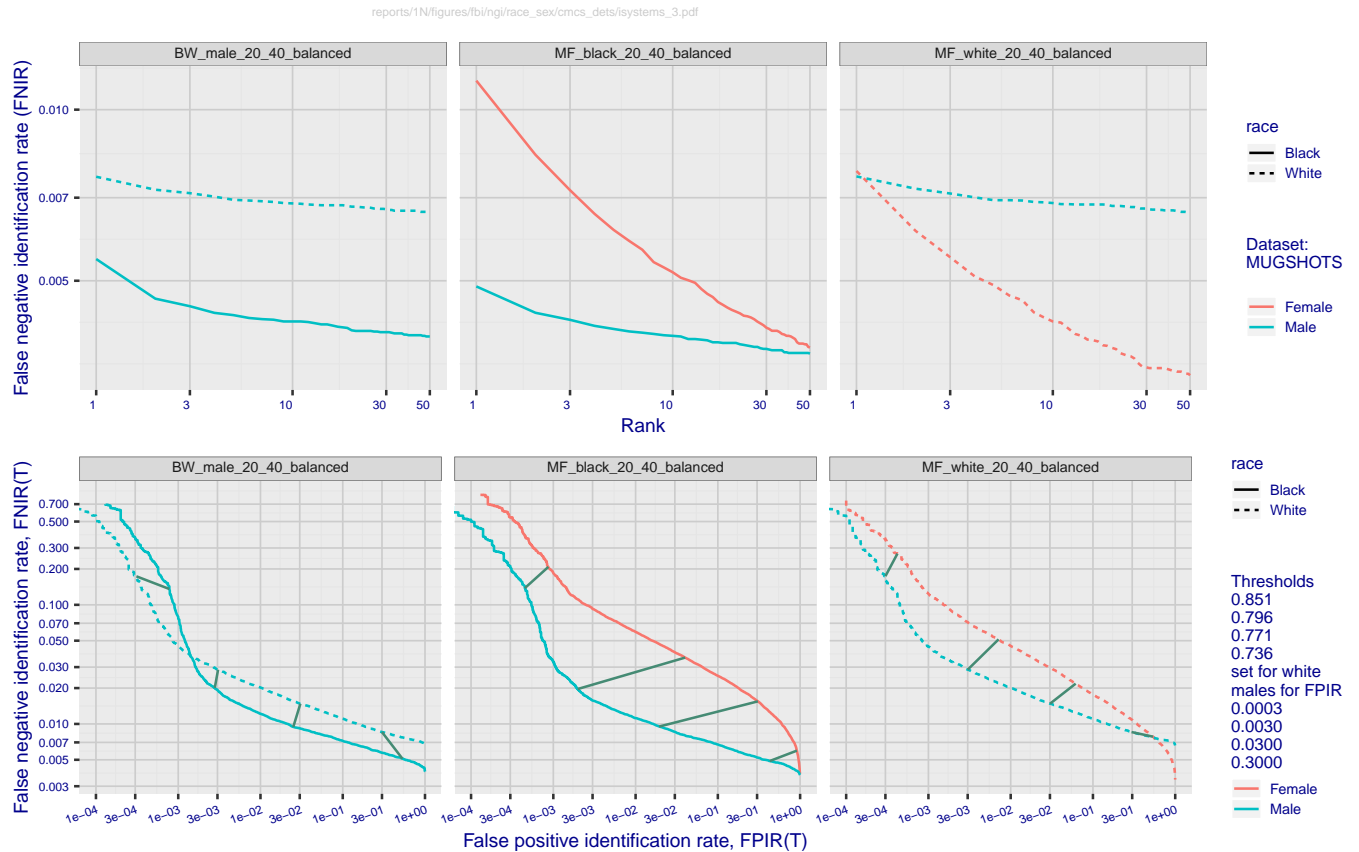


Figure 27: FNIR by sex and race for mugshot, isystems-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

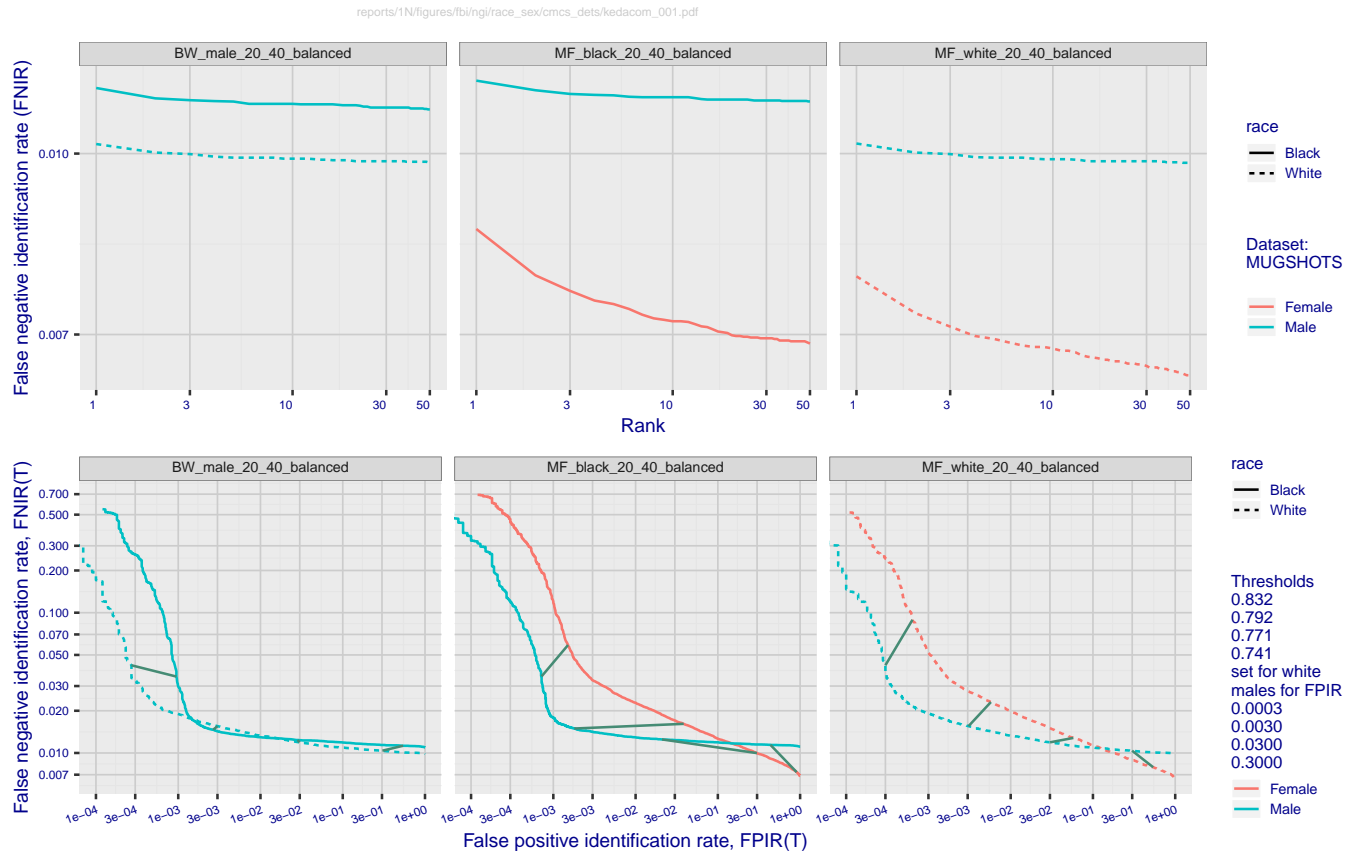


Figure 28: FNIR by sex and race for mugshot, kedacom-001. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

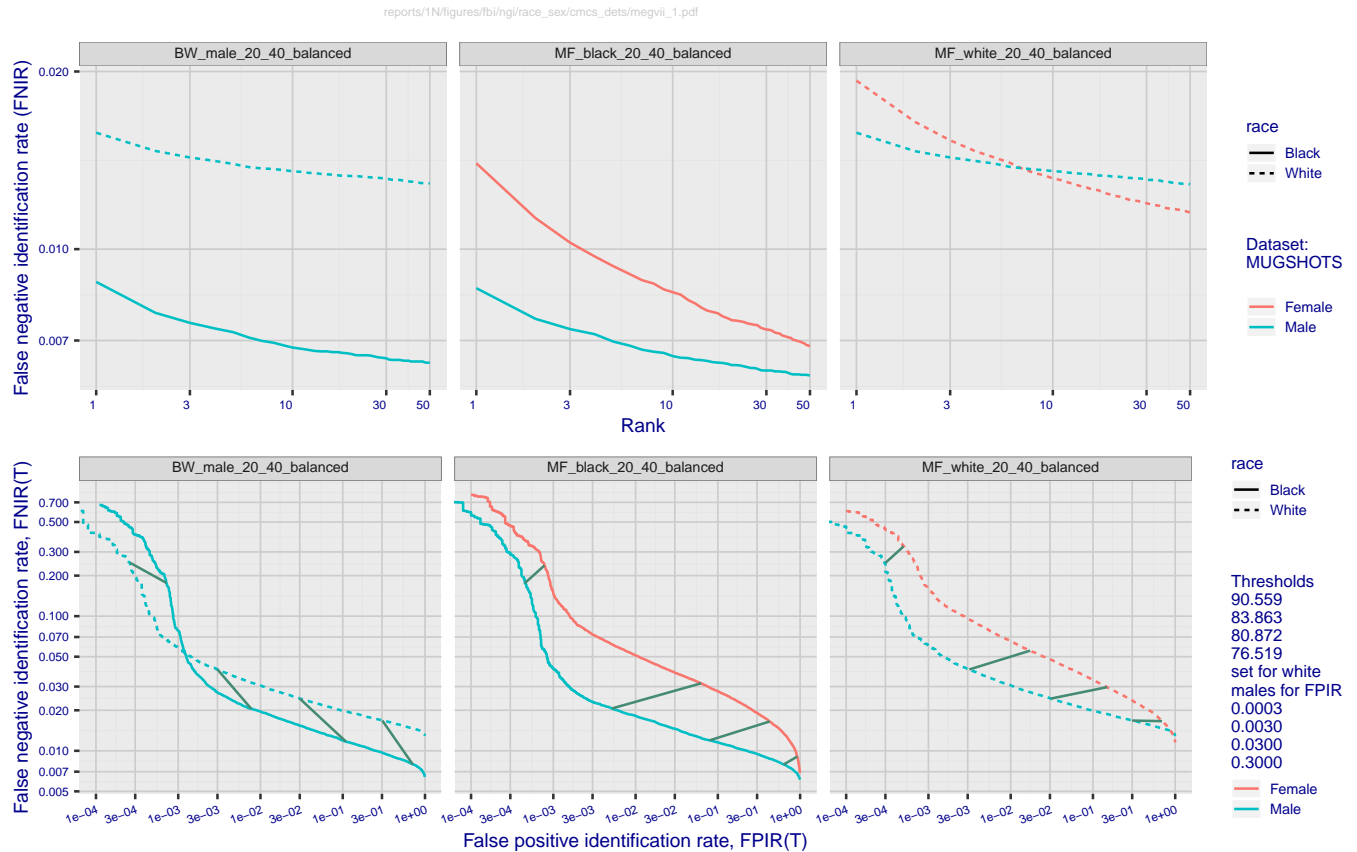


Figure 29: FNIR by sex and race for mugshot, megvii-1. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

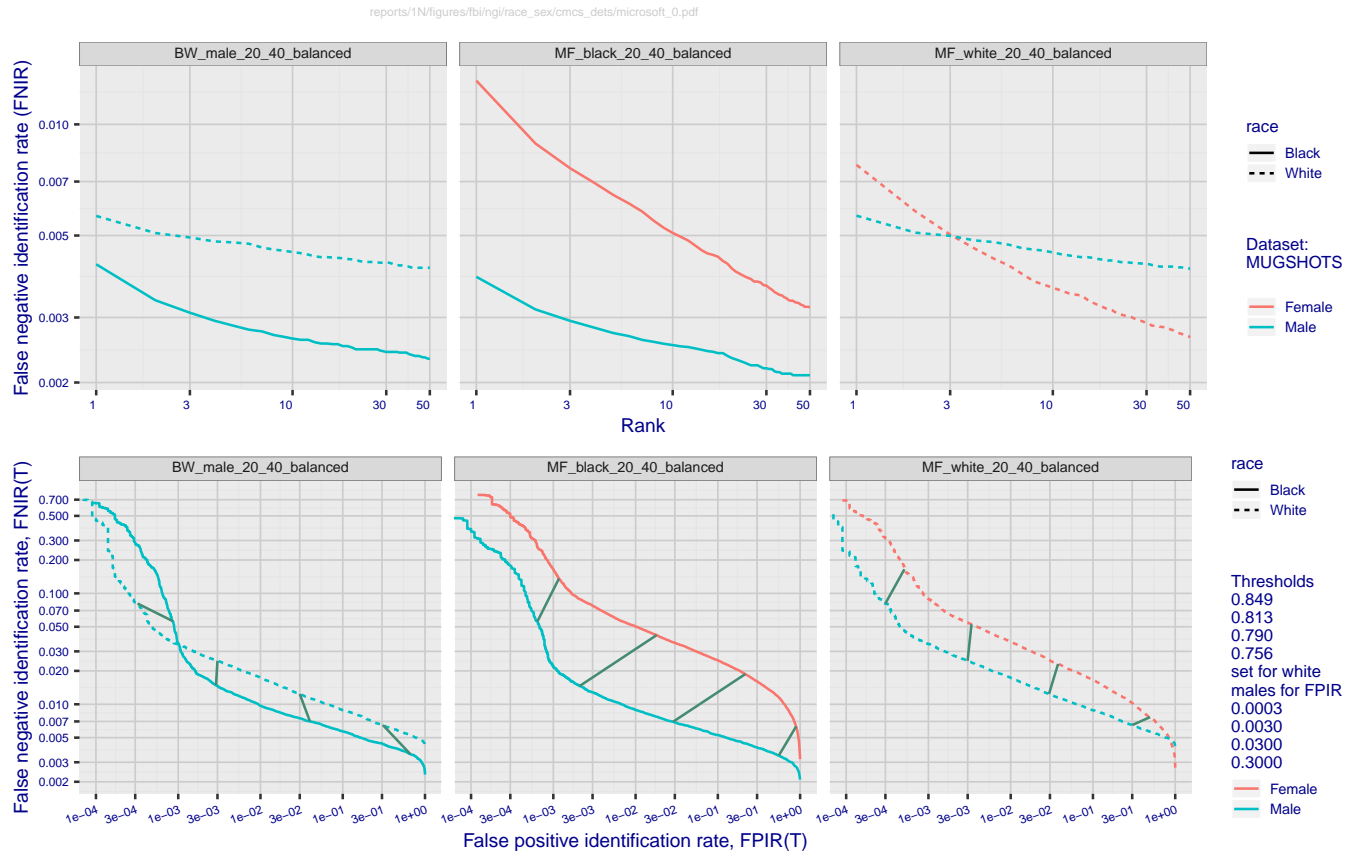


Figure 30: FNIR by sex and race for mugshot, microsoft-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

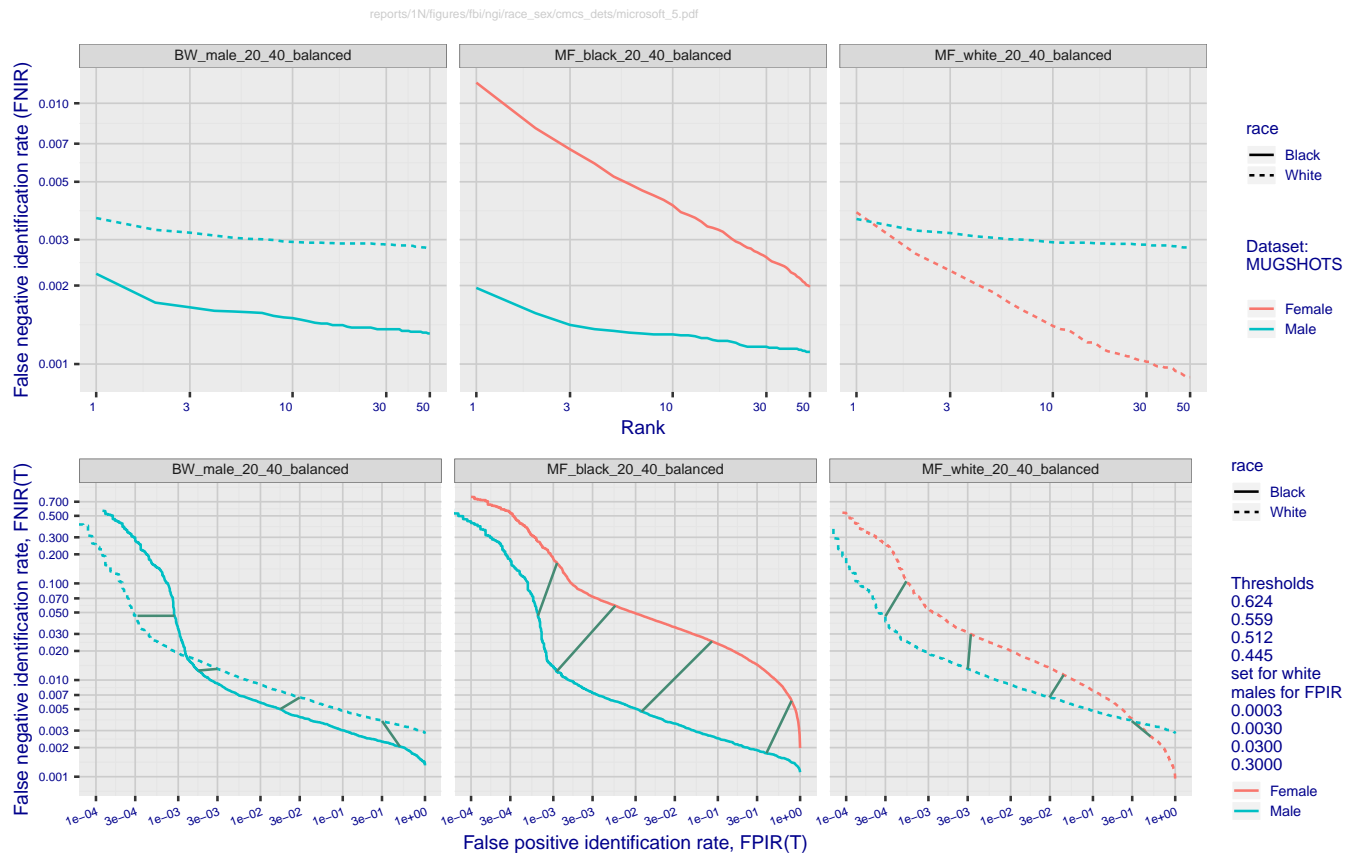


Figure 31: FNIR by sex and race for mugshot, microsoft-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

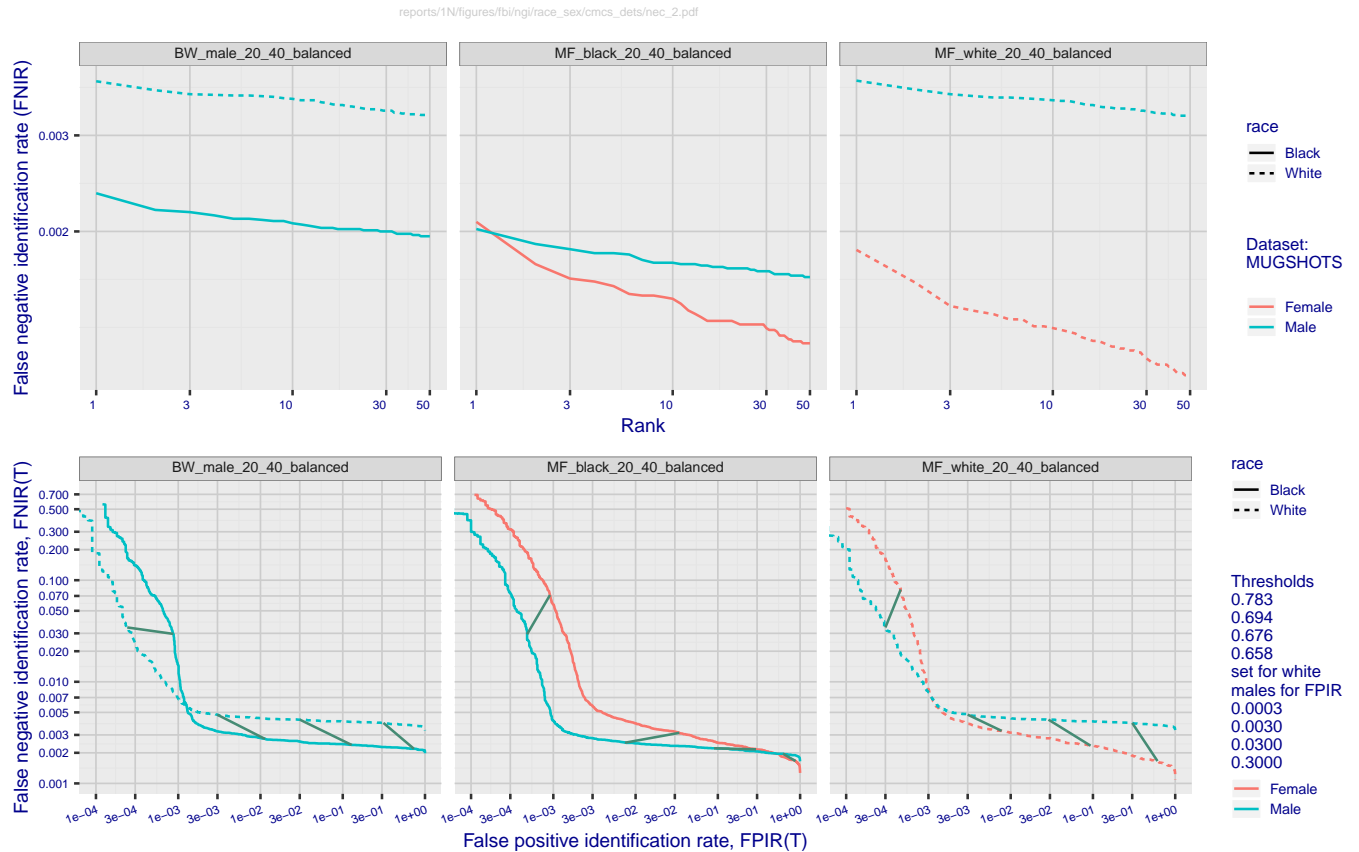


Figure 32: FNIR by sex and race for mugshot, nec-2. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

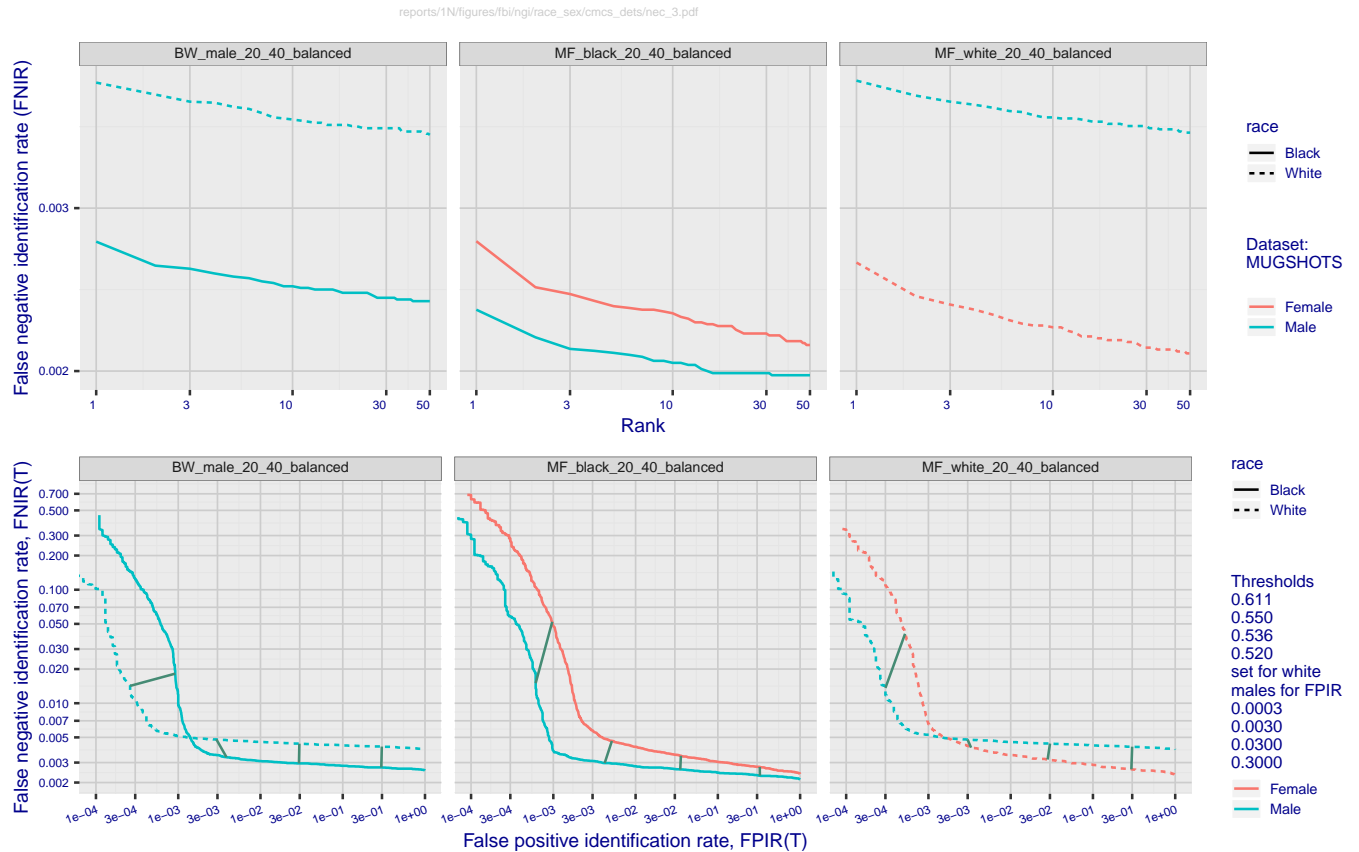


Figure 33: FNIR by sex and race for mugshot, nec-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

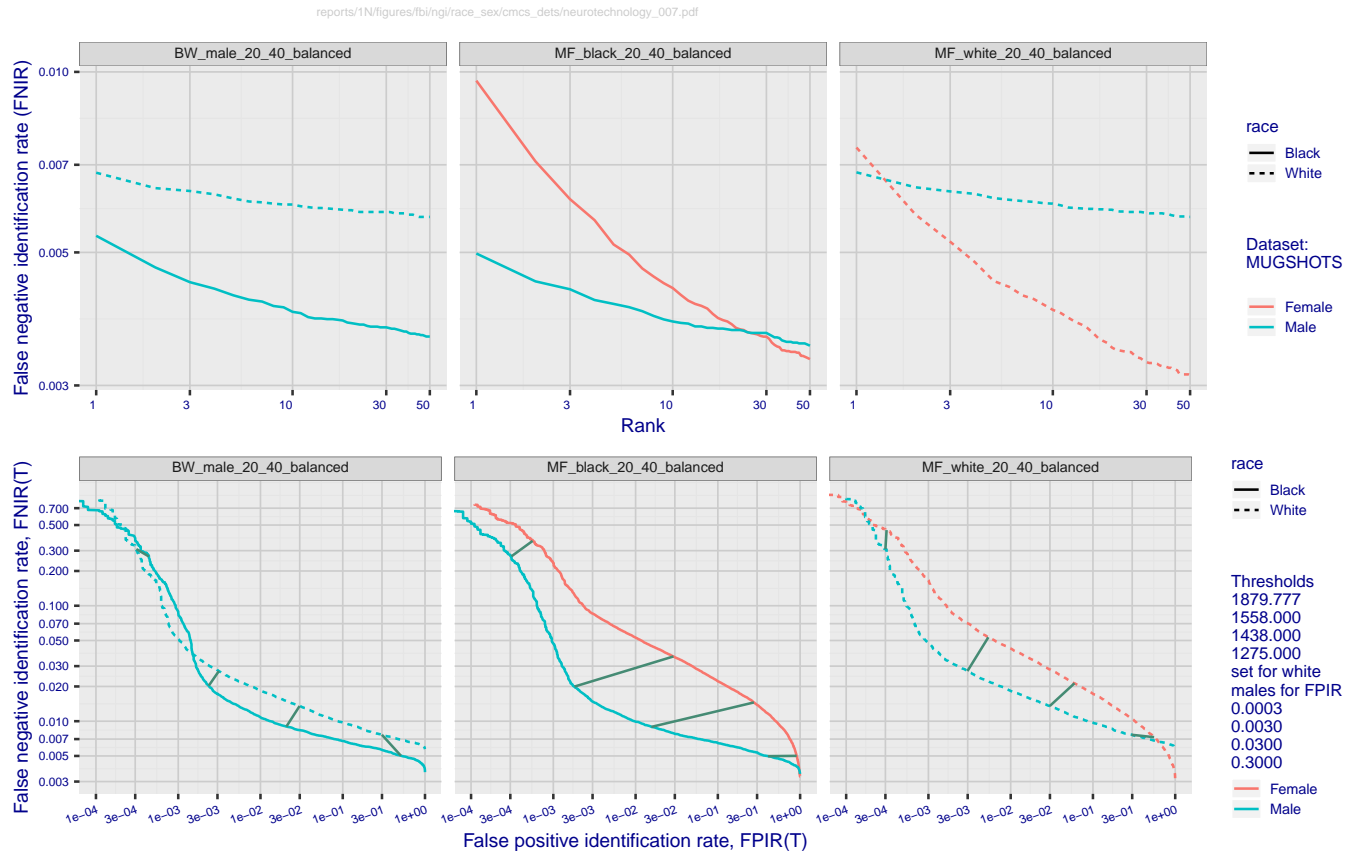


Figure 34: FNIR by sex and race for mugshot, neurotechnology-007. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

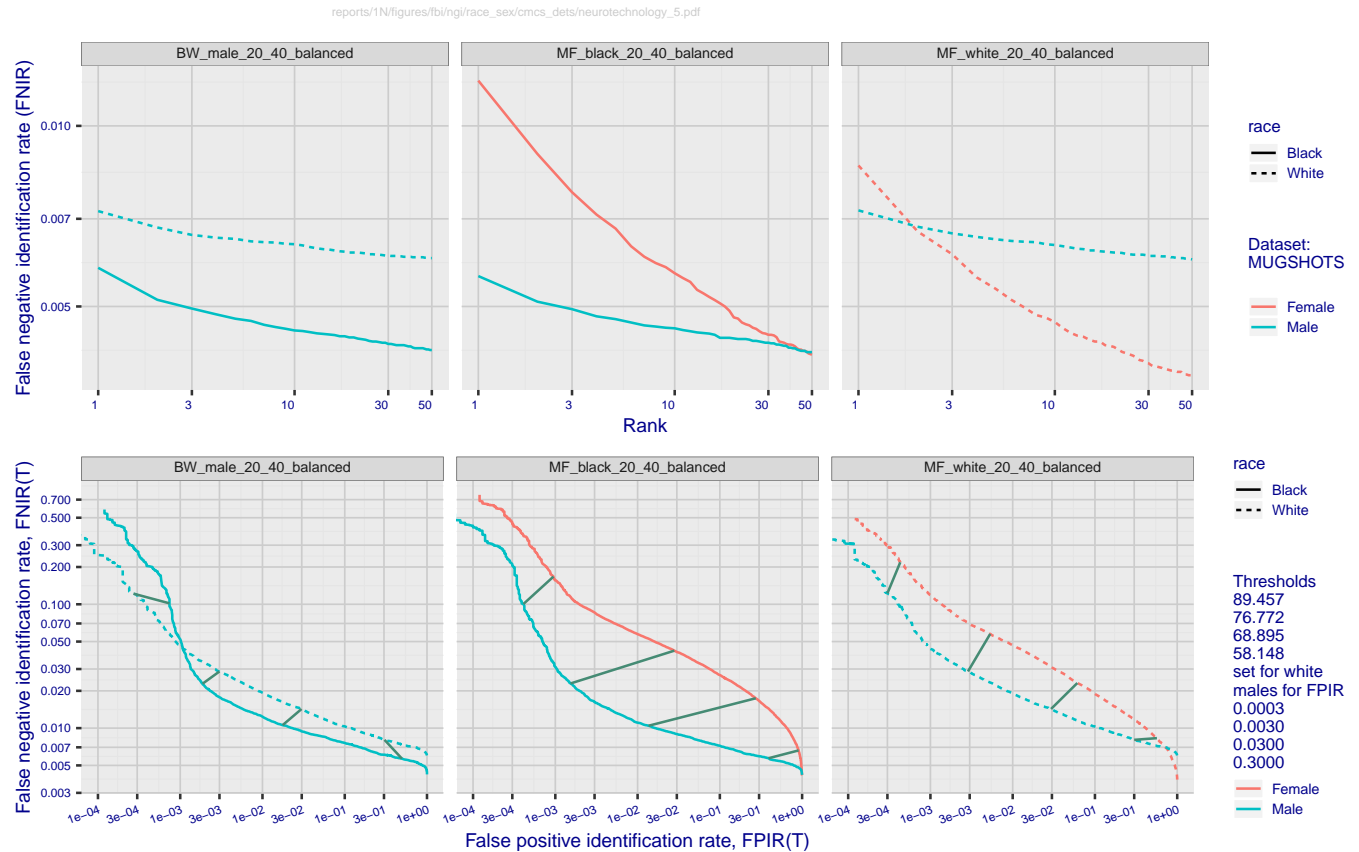


Figure 35: FNIR by sex and race for mugshot, neurotechnology-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

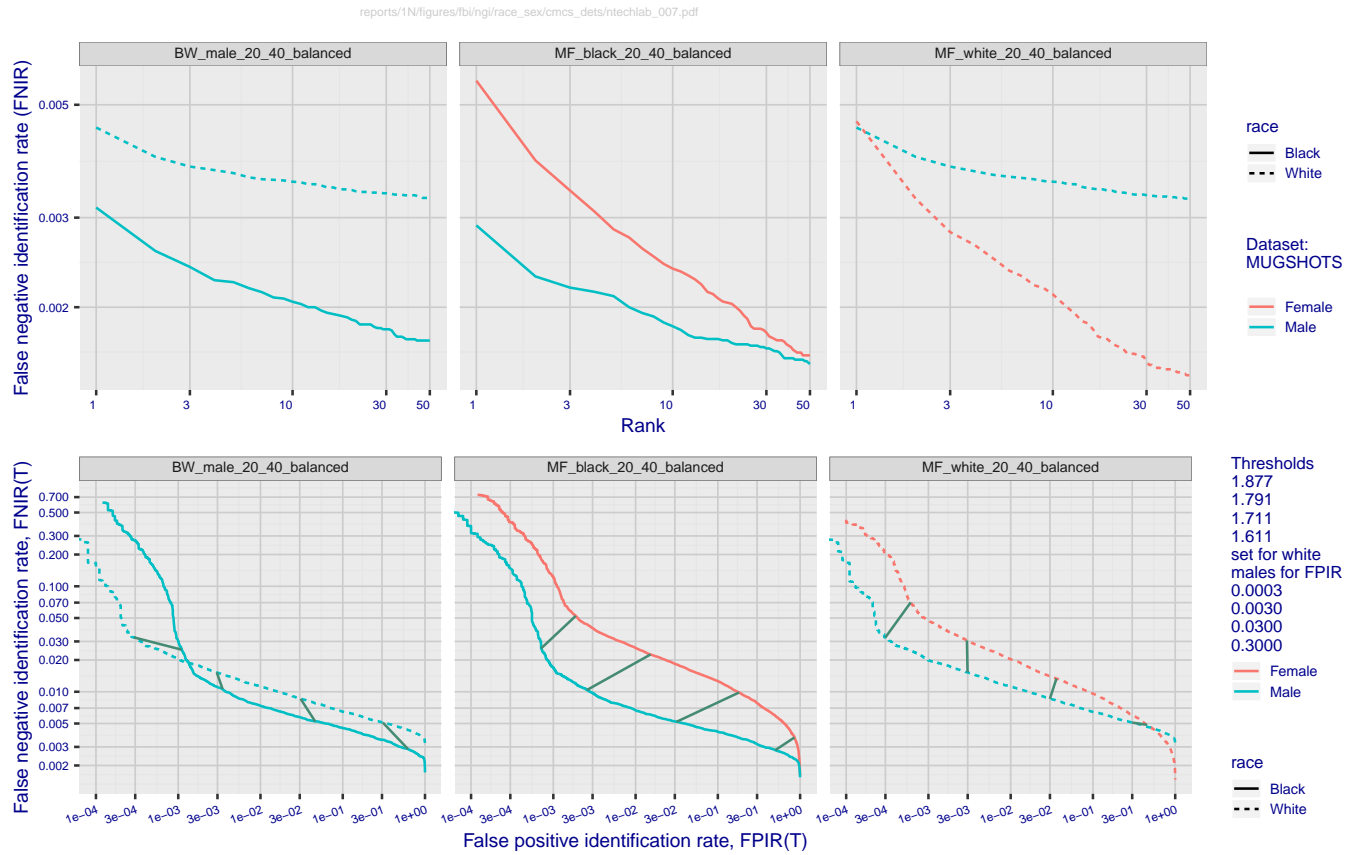


Figure 36: FNIR by sex and race for mugshot, ntechlab-007. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

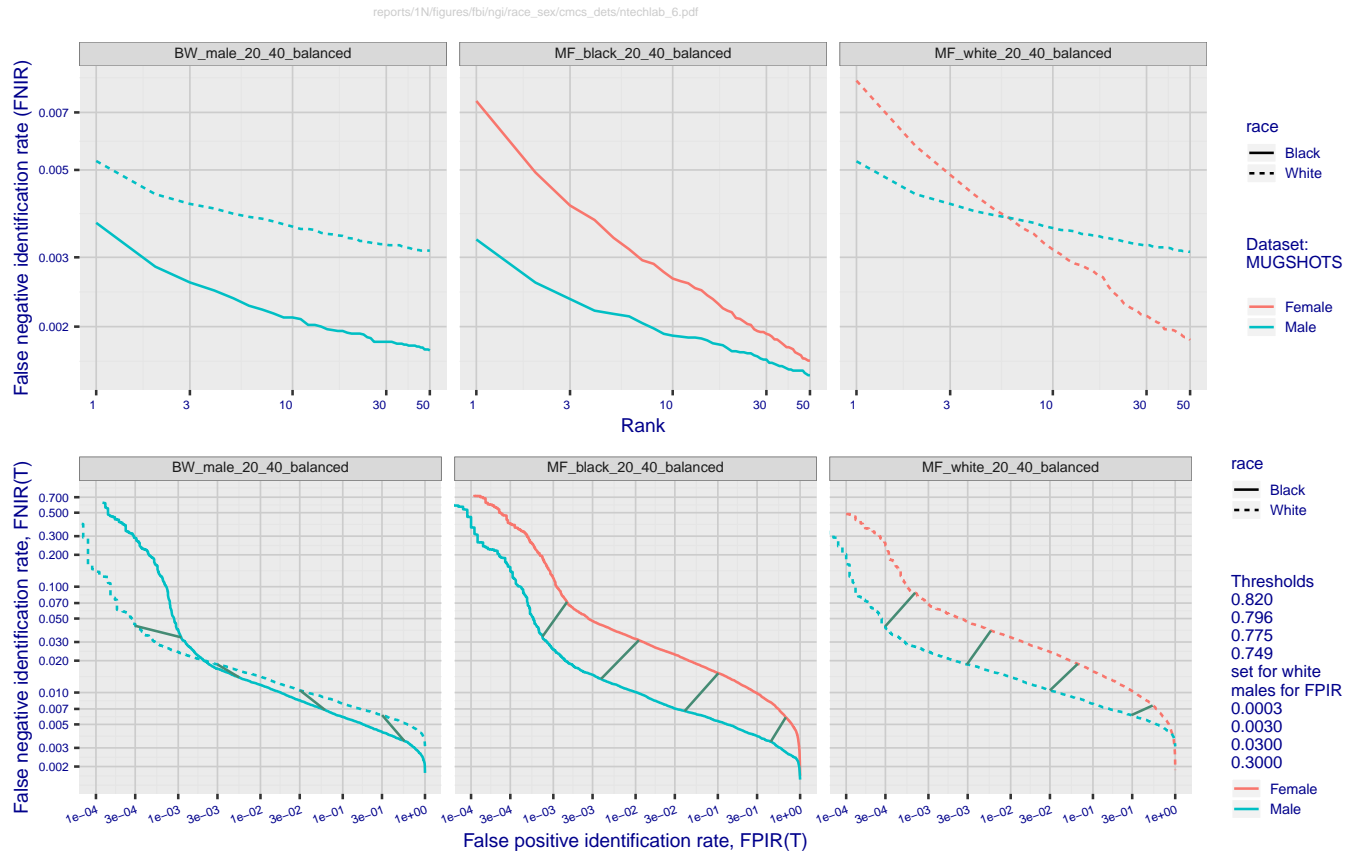


Figure 37: FNIR by sex and race for mugshot, ntechlab-6. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

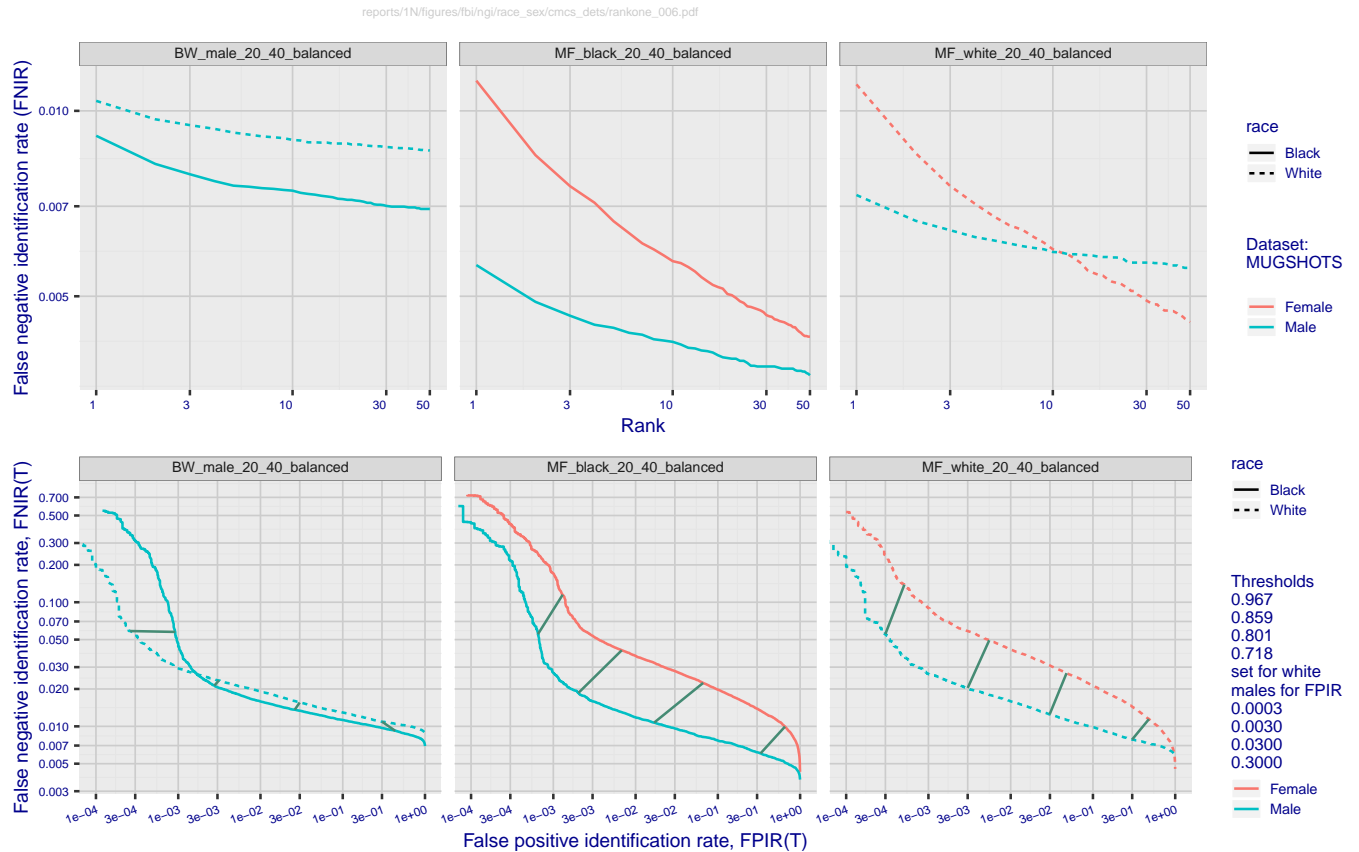


Figure 38: FNIR by sex and race for mugshot, rankone-006. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

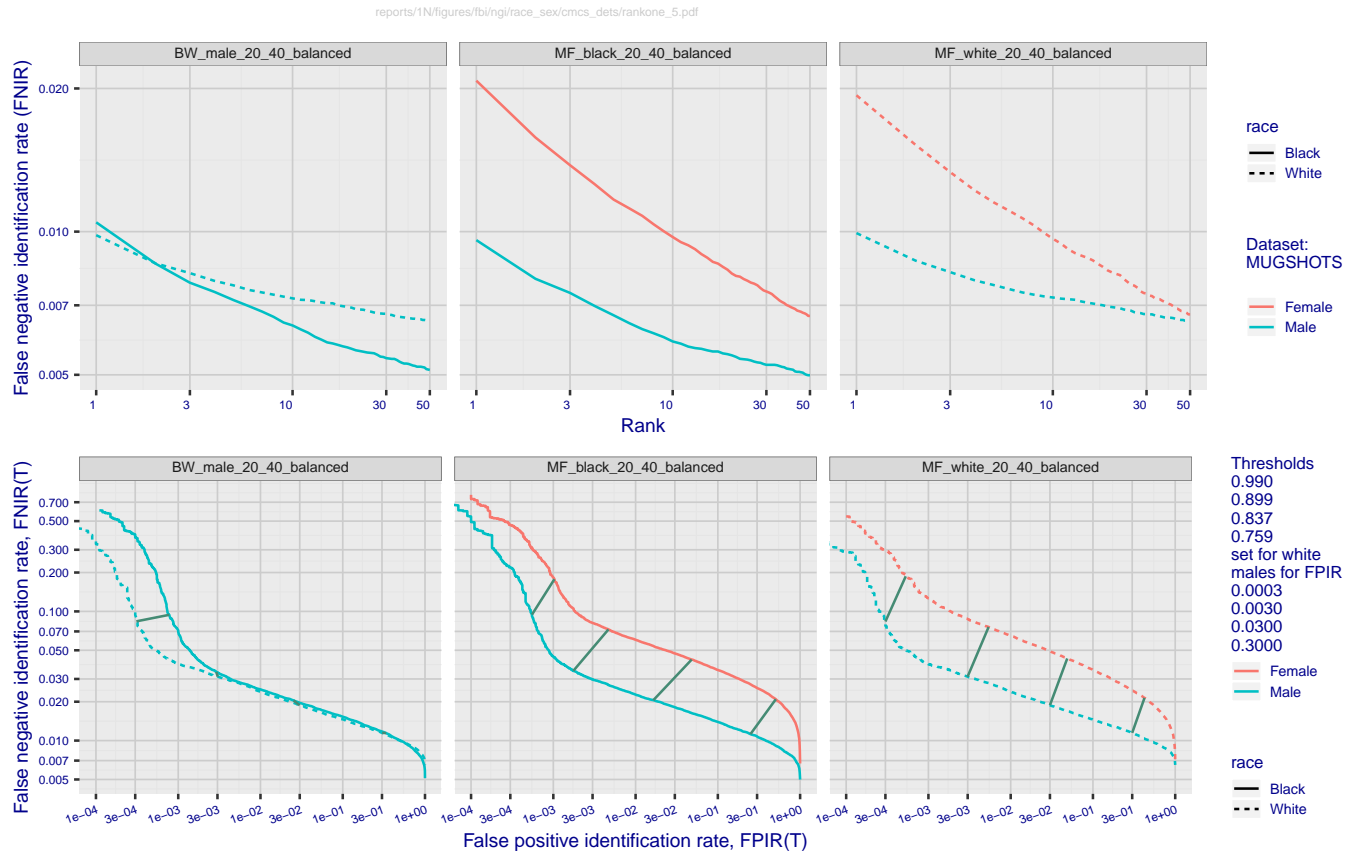


Figure 39: FNIR by sex and race for mugshot, rankone-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

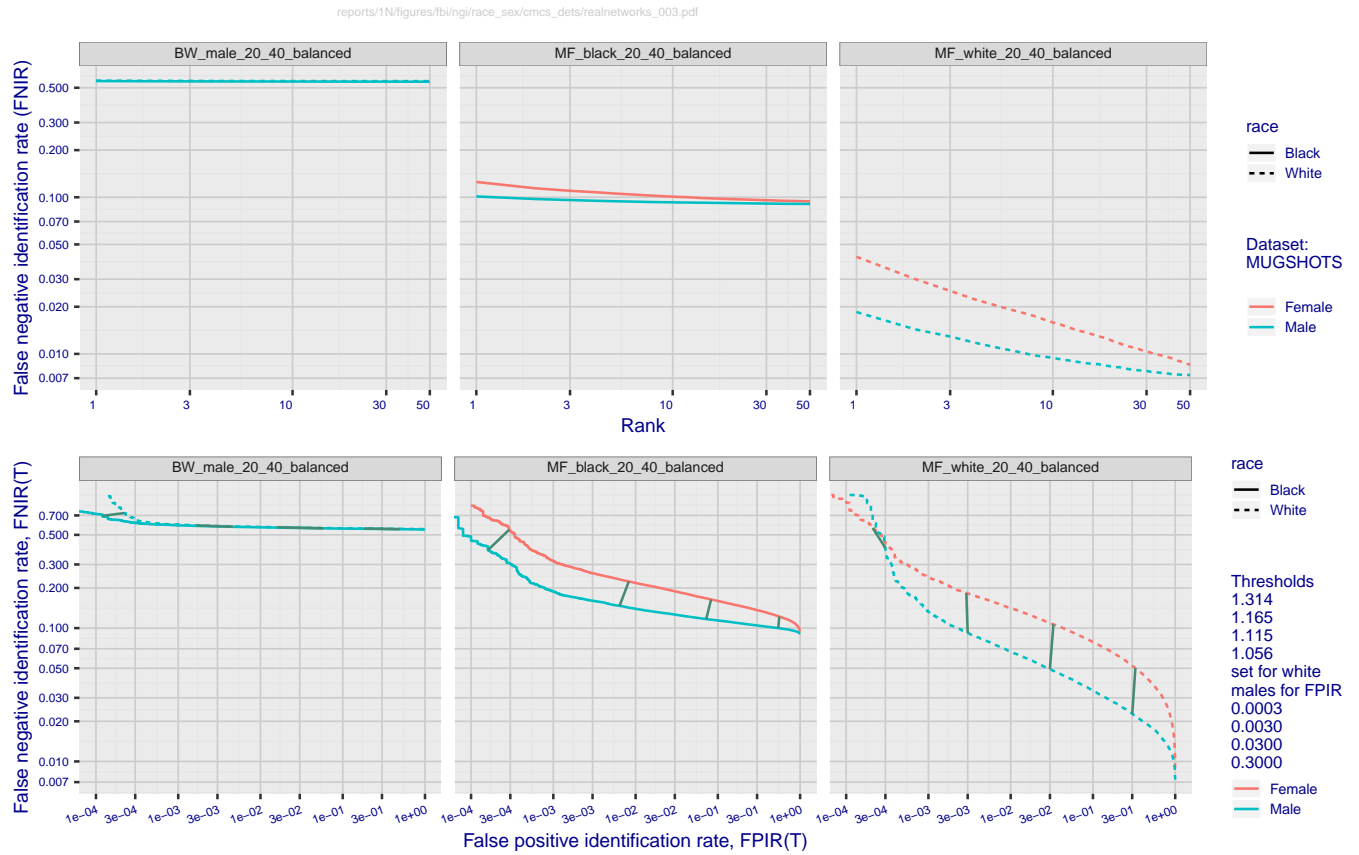


Figure 40: FNIR by sex and race for mugshot, realnetworks-003. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

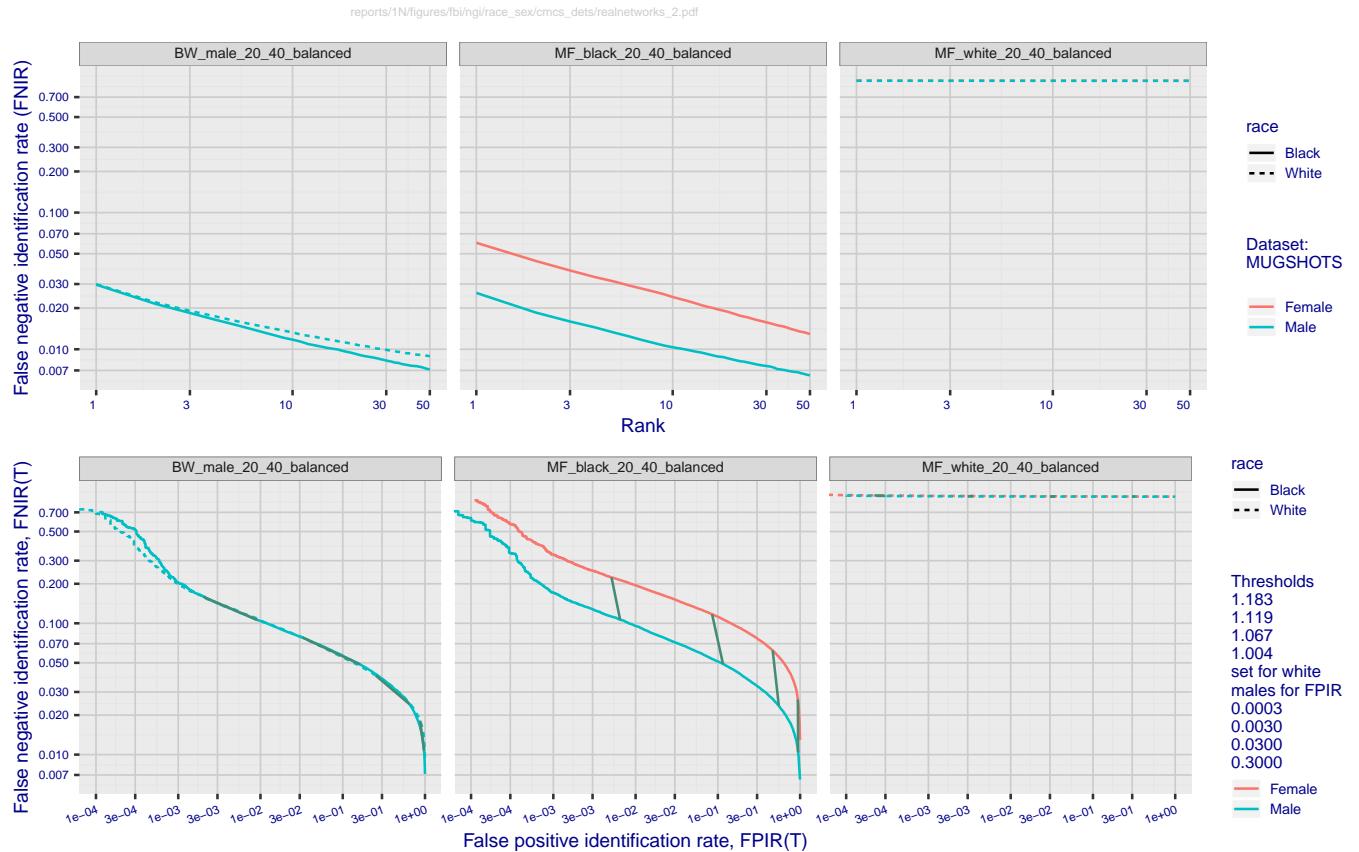


Figure 41: FNIR by sex and race for mugshot, realnetworks-2. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8280

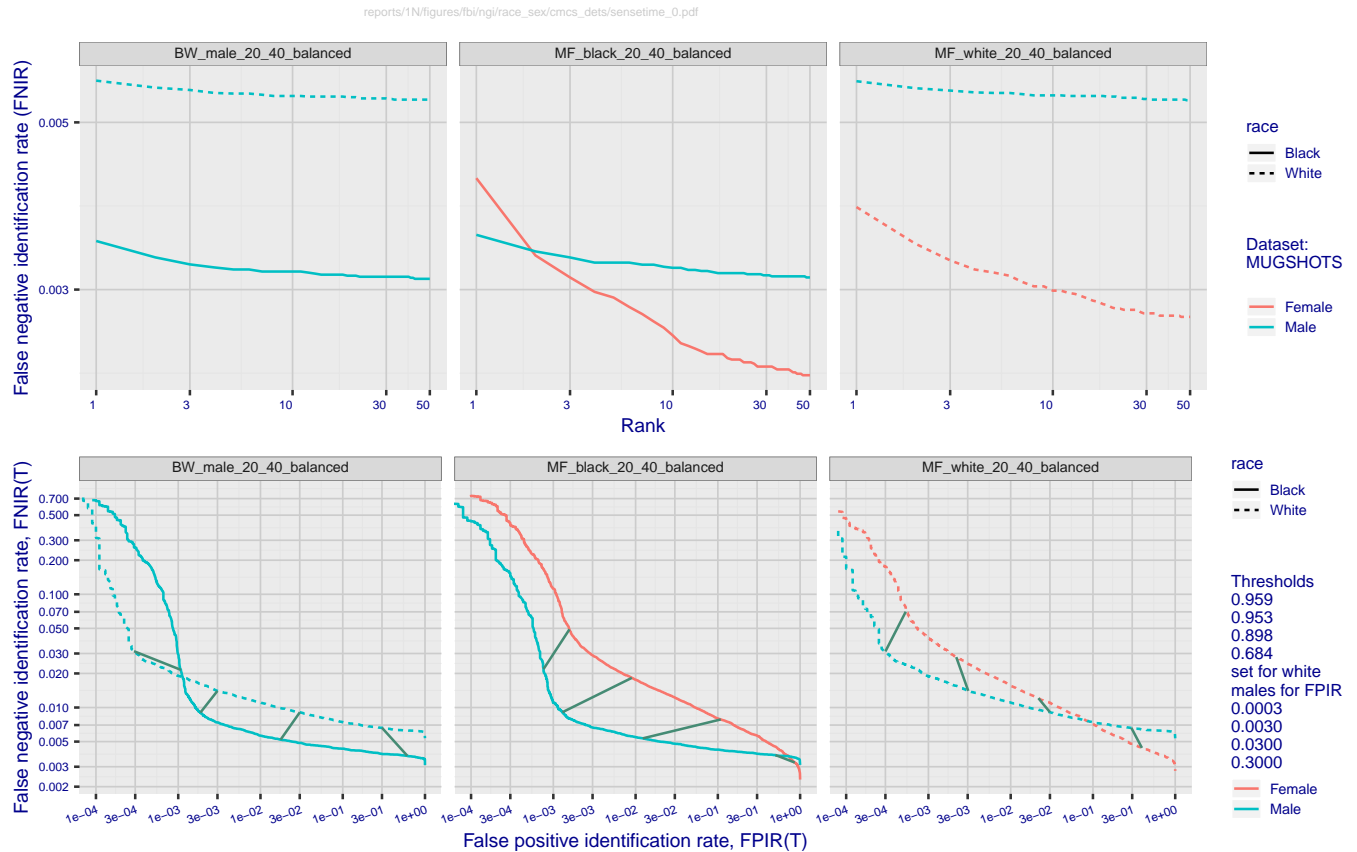


Figure 42: FNIR by sex and race for mugshot, sensetime-0. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

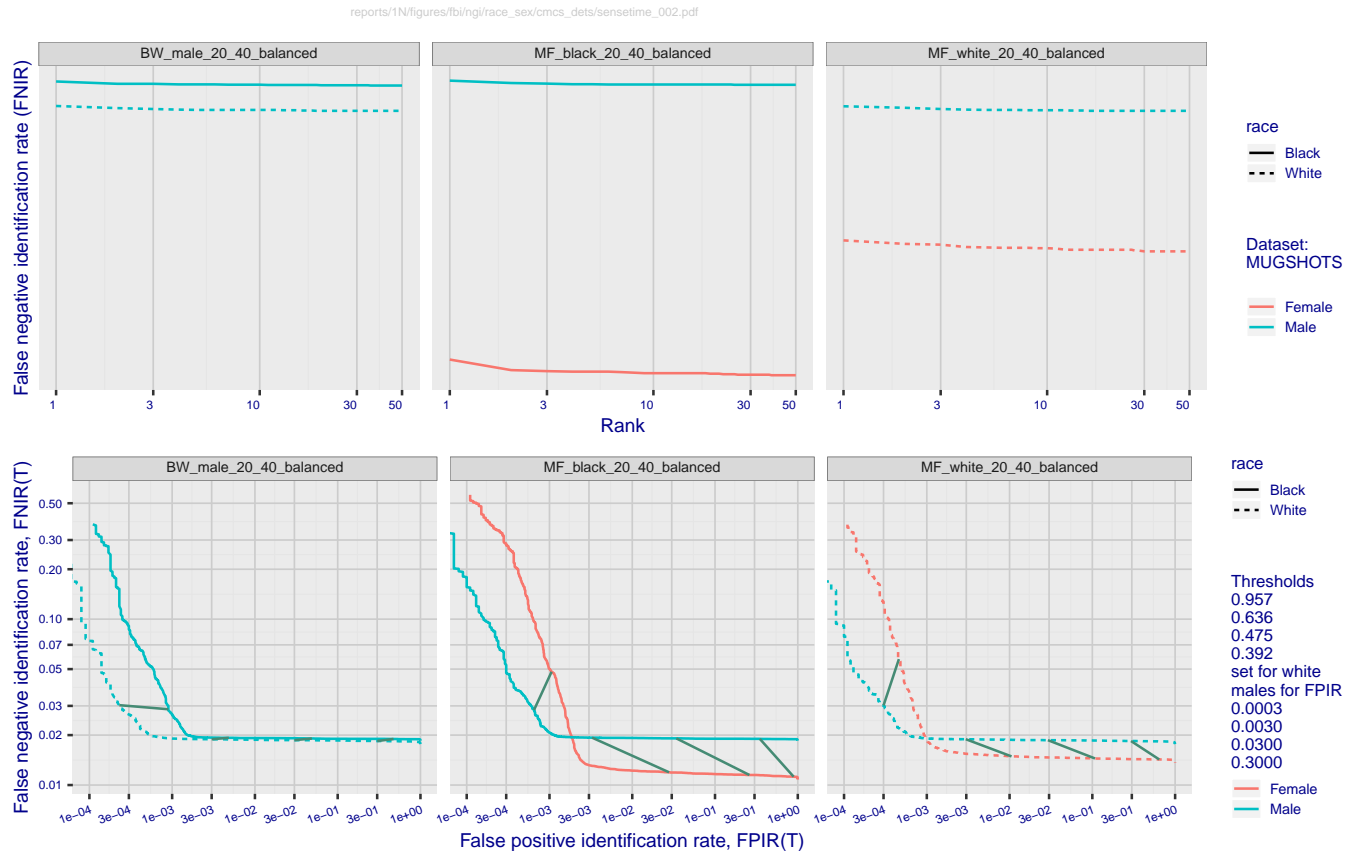


Figure 43: FNIR by sex and race for mugshot, sensetime-002. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

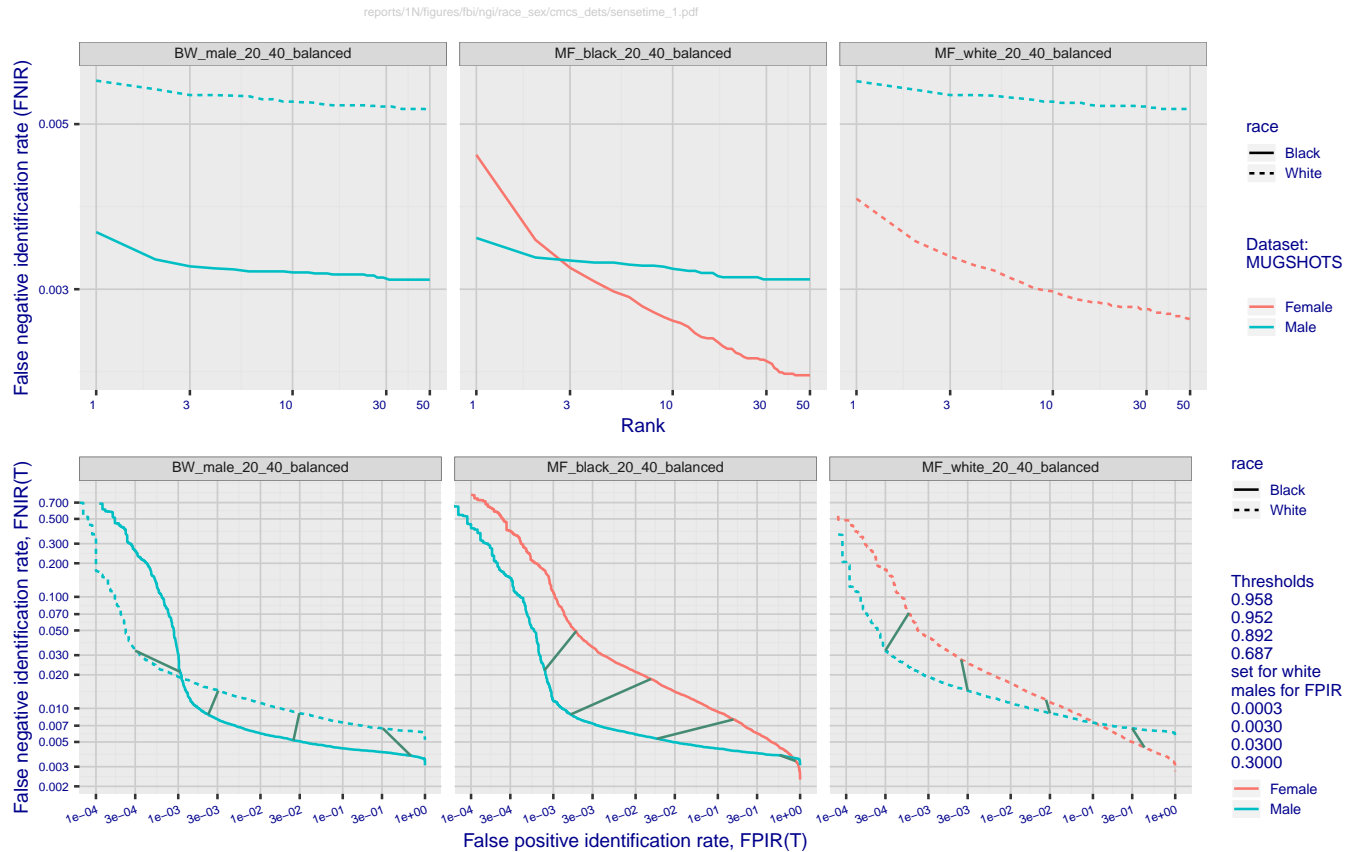


Figure 44: FNIR by sex and race for mugshot, sensetime-1. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

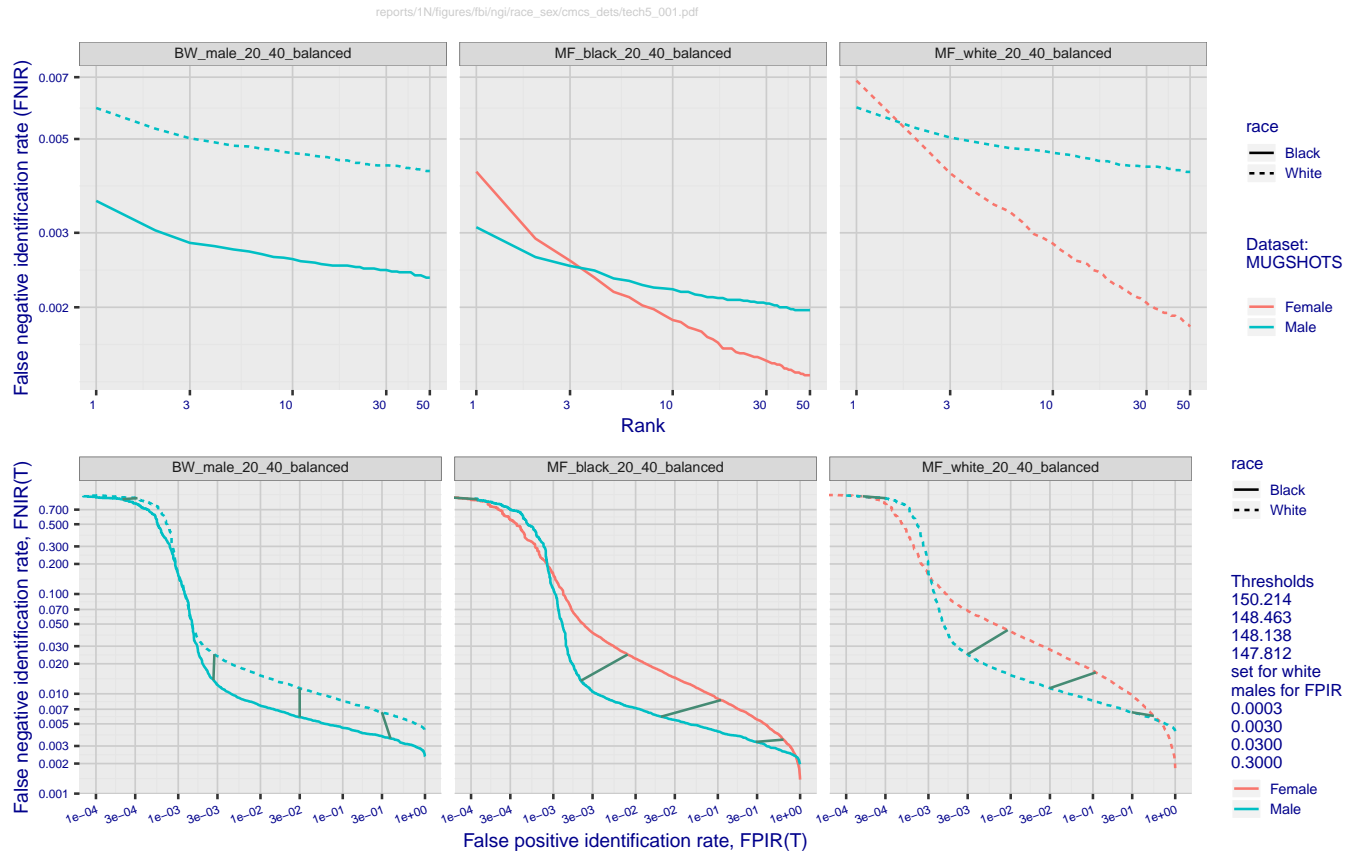


Figure 45: FNIR by sex and race for mugshot, tech5-001. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

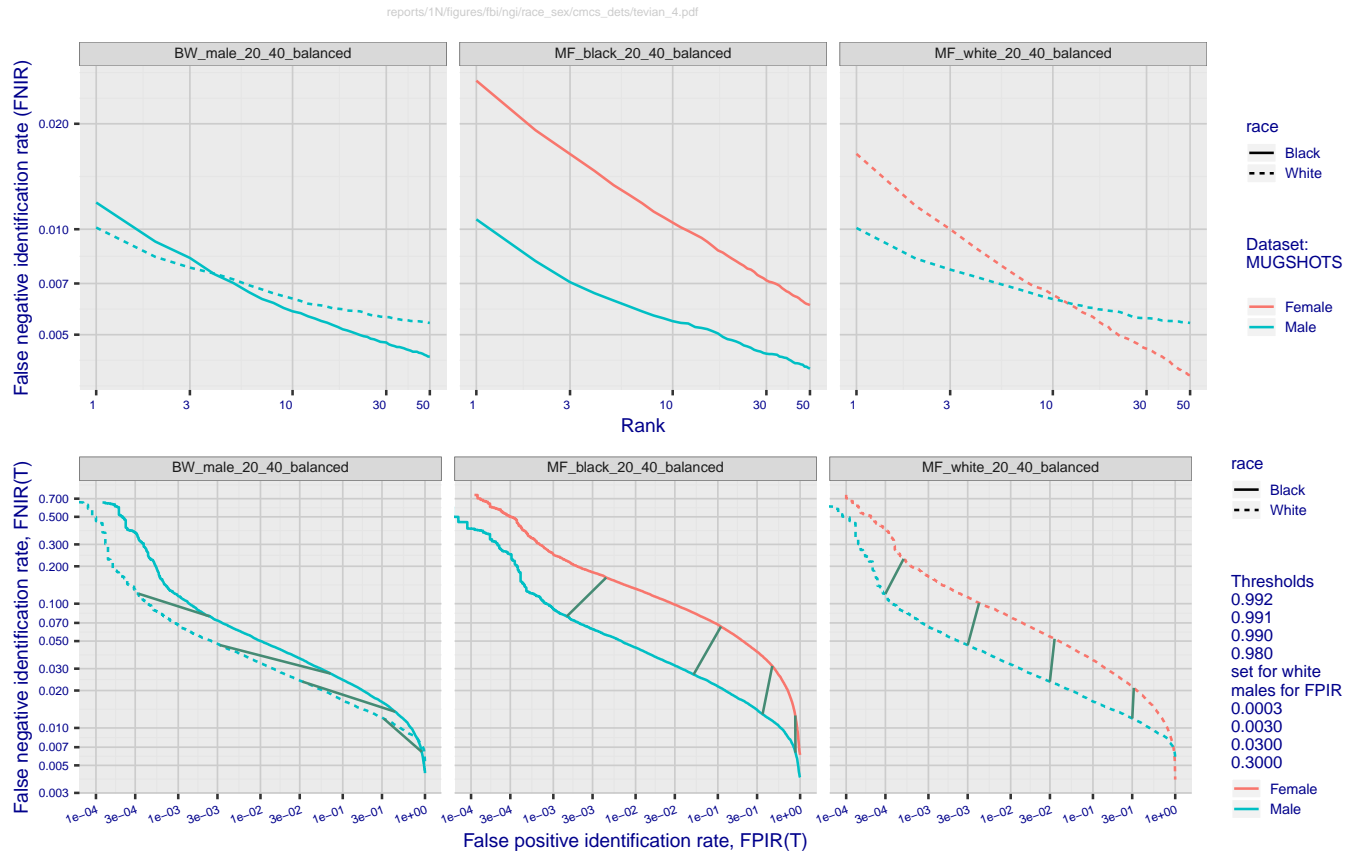


Figure 46: FNIR by sex and race for mugshot, tevia-4. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

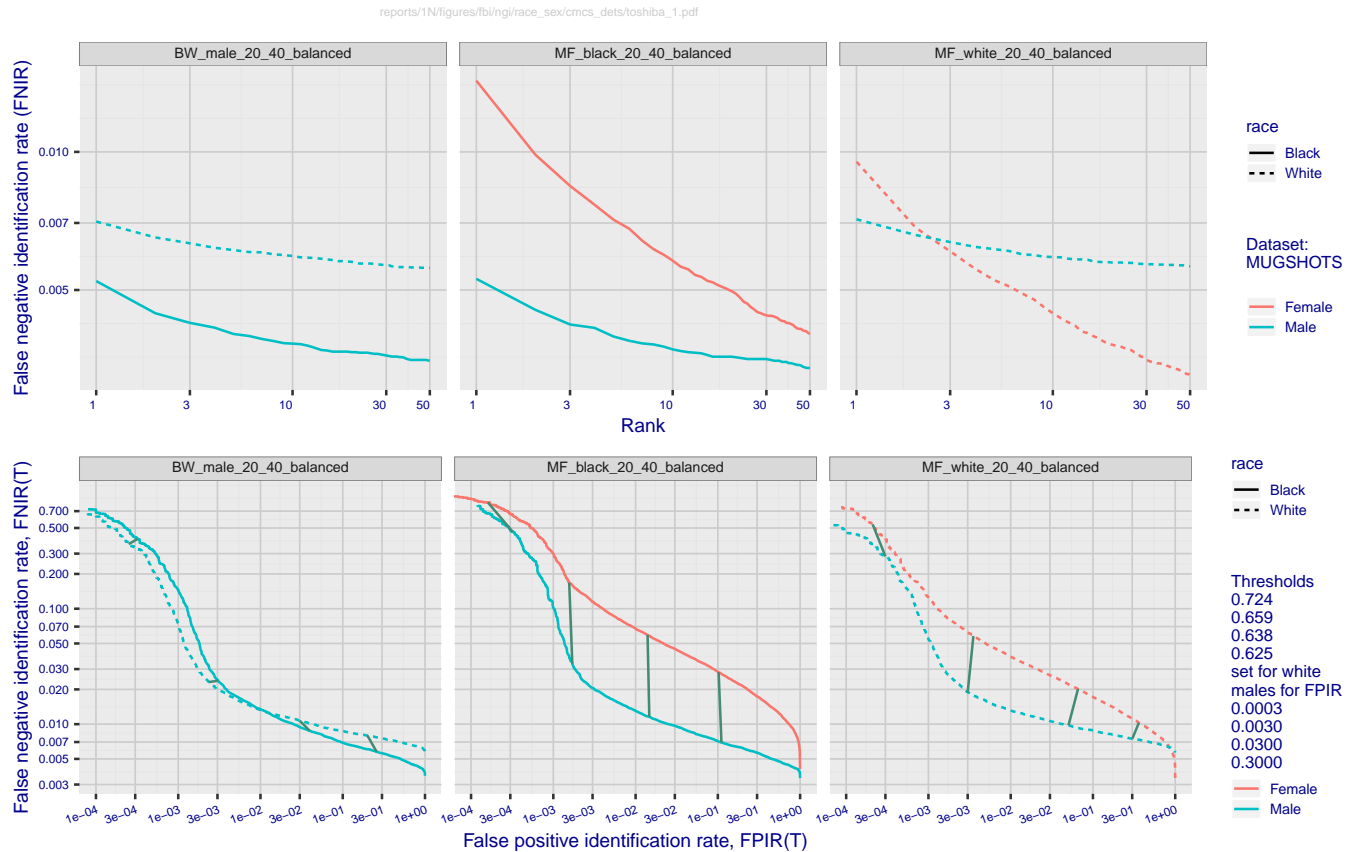


Figure 47: FNIR by sex and race for mugshot, toshiba-1. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

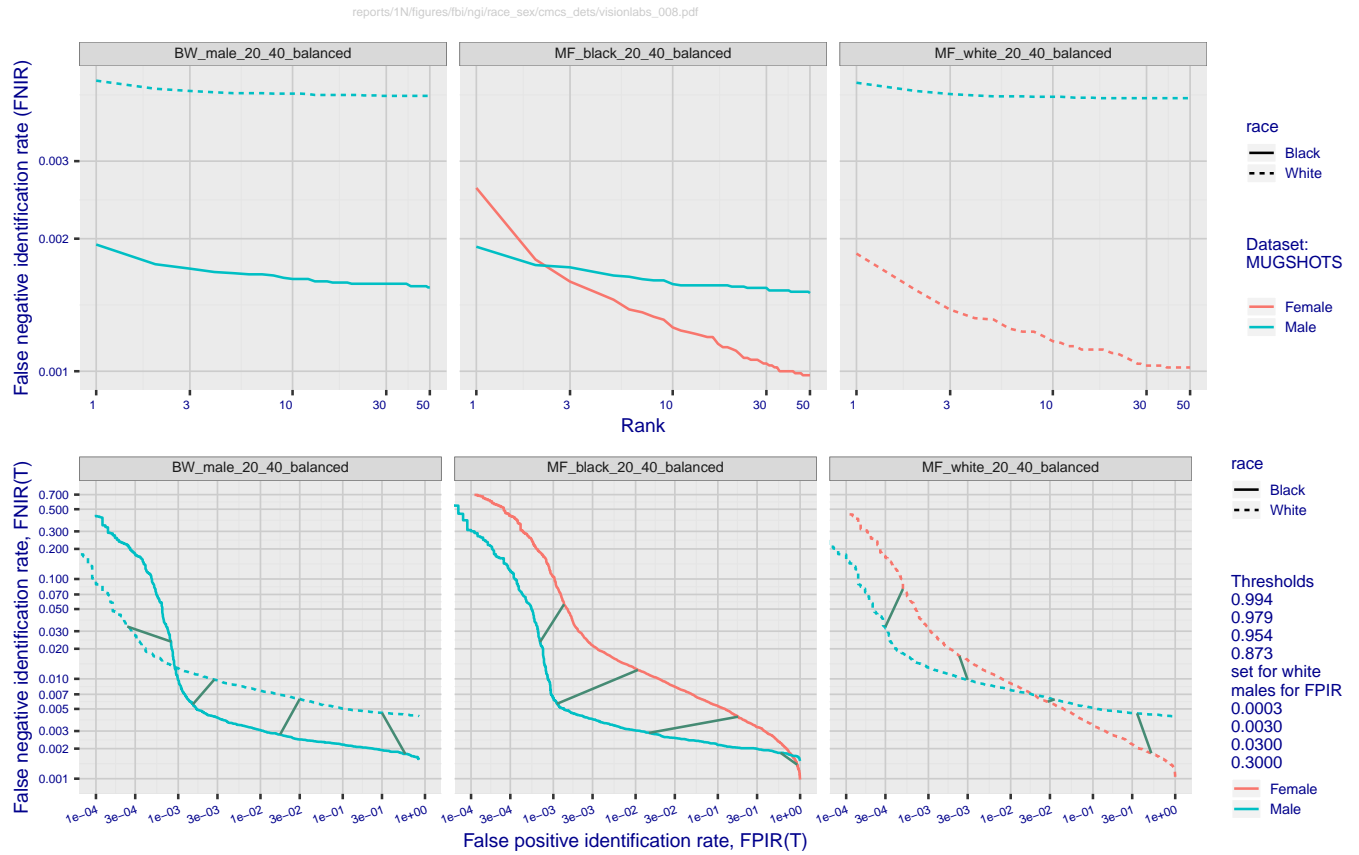


Figure 48: FNIR by sex and race for mugshot, visionlabs-008. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

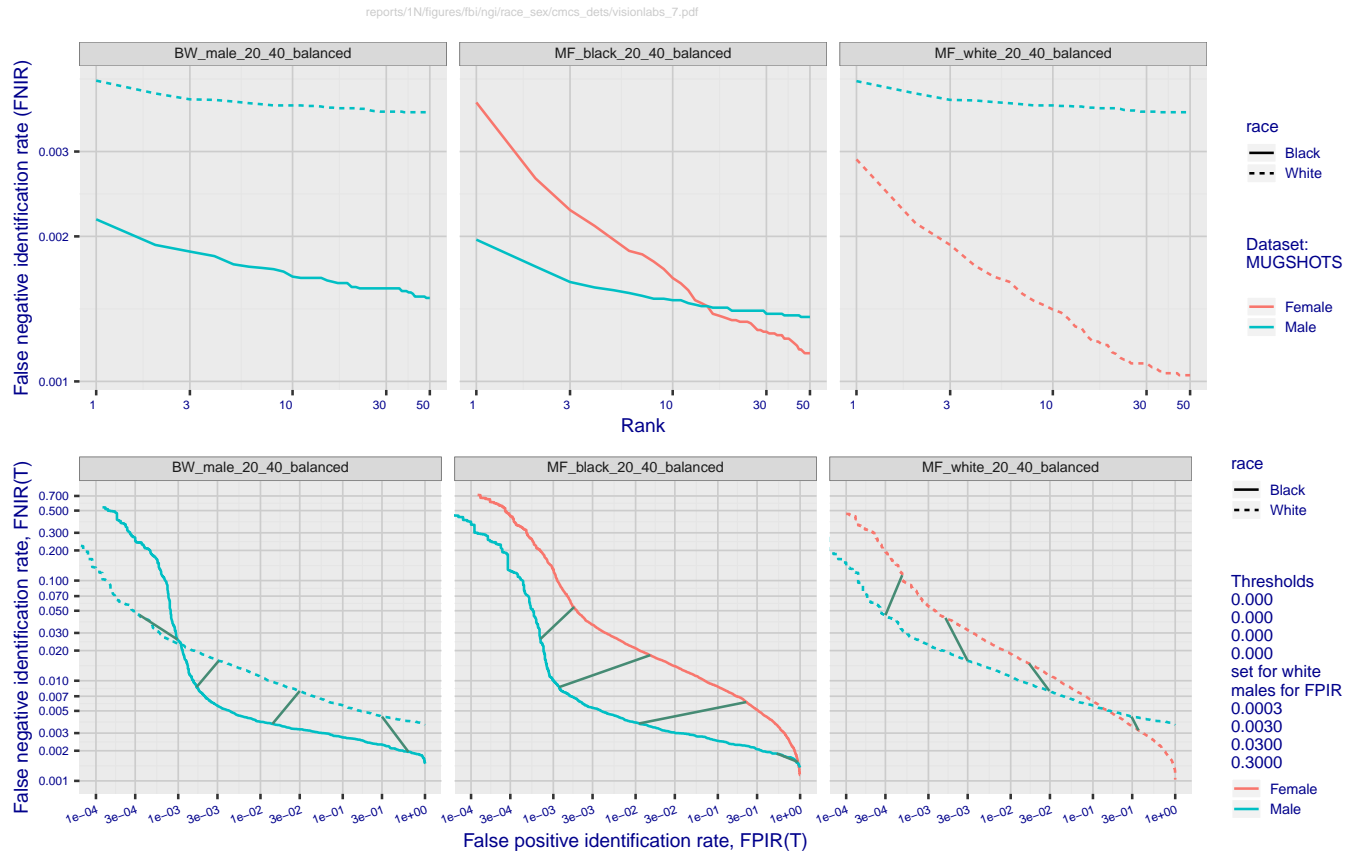


Figure 49: FNIR by sex and race for mugshot, visionlabs-7. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

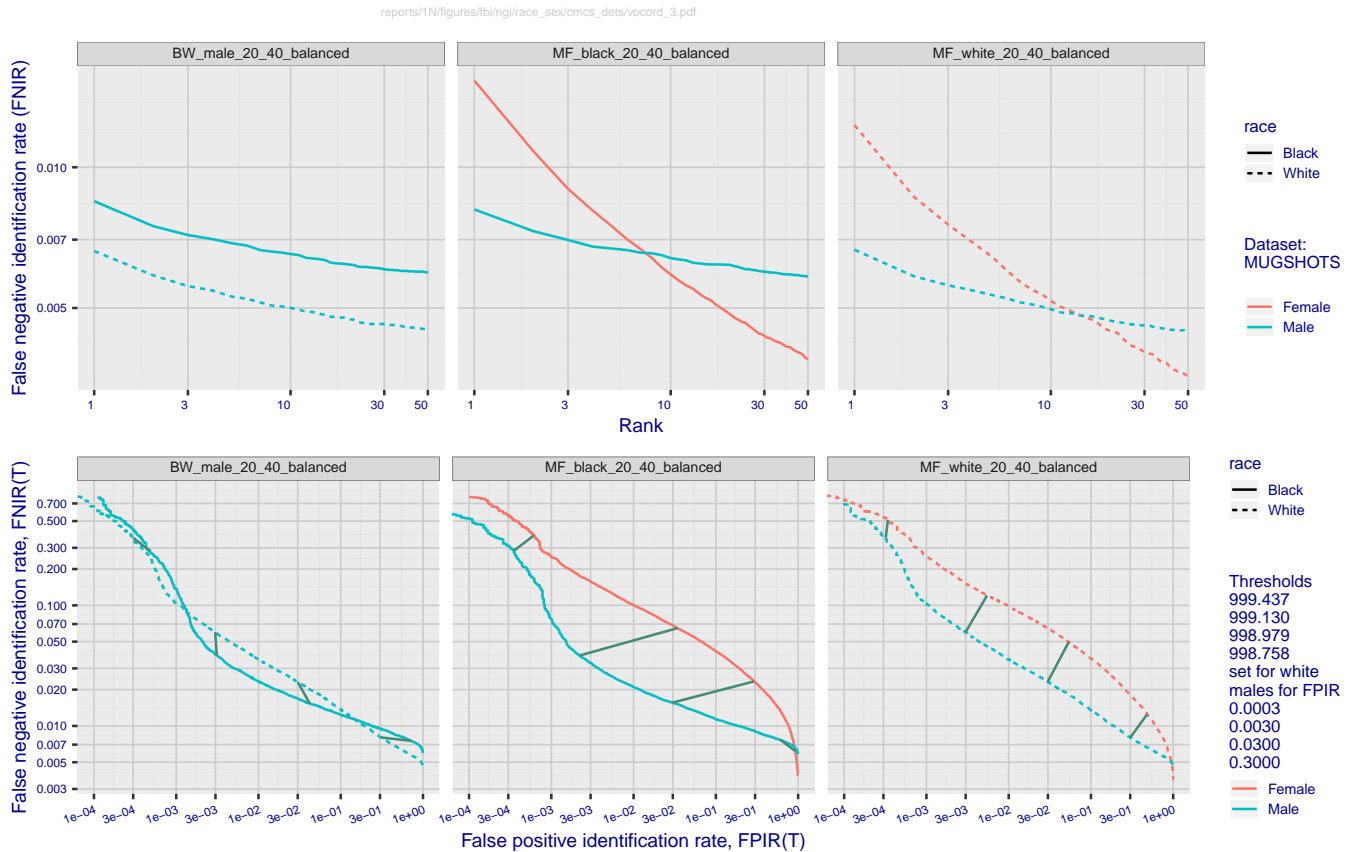


Figure 50: FNIR by sex and race for mugshot, vocord-3. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>

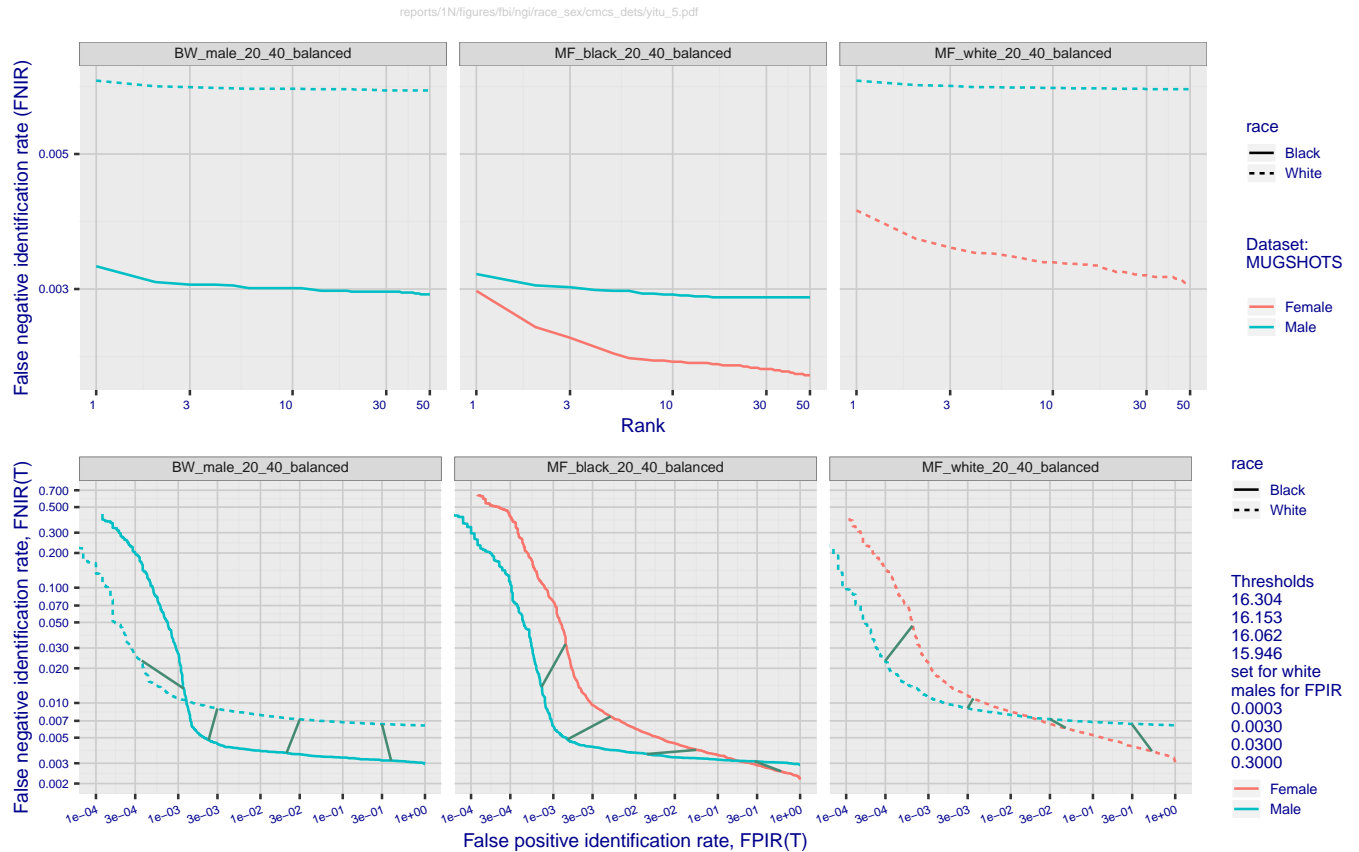


Figure 51: FNIR by sex and race for mugshot, yitu-5. The three columns of panels represent three one-to-many search trials with, respectively, balanced galleries of black and white males, black males and females, and white males and females. The upper row shows false negative identification rate vs. rank for investigative search applications with zero threshold. The lower row shows false negative vs. false positive identification rates for higher threshold applications where only strong hits should be returned, either because that causes an acceptable decision or because labor is available to review the candidates returned. The grey lines join points of equal threshold. The three thresholds are chosen to give FPIR of 0.0003, 0.003, 0.03, 0.3 respectively one baseline demographic, here white males.

This publication is available free of charge from: <https://doi.org/10.6028/NIST.IR.8280>