

U.S. Department of Homeland Security

---

# SCIENCE AND TECHNOLOGY DIRECTORATE

2023 Remote Identity Validation Technology Demonstration (RIVTD) Webinar



Science and  
Technology

**Arun Vemury**  
Senior Advisor  
Biometric & Identity  
Technology Center

**John Howard**  
Lead Data Scientist  
The Maryland Test Facility

**Yevgeniy Sirotin**  
Technical Director  
The Maryland Test Facility

August 2024

# Agenda

---

- Introduction
- Remote Identity Validation Tech Overview
- Remote Identity Validation Tech Track 2: Match to Document Overview
  - Data Used
  - Match to Document System Properties
- Remote Identity Validation Tech Track 2” Results
  - Evaluation Criteria
  - Failure to Extract Rate
  - False Non-Match Rate
  - Demographic Analysis
  - False Match Rate
- Summary & Conclusions



# Biometric & Identity Technology Center

The Science & Technology Directorate (S&T) conducts foundational research to ensure advancements in science and technology are harnessed in the development of cutting-edge solutions to new and emerging operational challenges.

- ✓ Drive biometric and identity innovation at the Department of Homeland Security (DHS) through Research, Development, Test, and Evaluation (RDT&E) capabilities
- ✓ Facilitate and accelerate understanding of biometrics and identity technologies for new, DHS use cases
- ✓ Drive efficiencies by supporting cross-cutting methods, best practices and solutions across programs
- ✓ Deliver subject matter expertise across the DHS enterprise
- ✓ Engage industry and provide feedback
- ✓ Encourage innovation across industry and academia



# Remote Identity Validation Tech Demo

- Industry has developed new tools to authenticate documents and verify the identity of users remotely:
  - Remote Identity Validation (RIV)
- Difficult for industry to test the effectiveness and fairness of these systems:
  - Hard to obtain large samples of genuine documents and their owners
  - Testing for demographic differentials is costly
- S&T is studying the current performance of RIV to help industry to develop more secure, accurate and equitable technologies

# 2023 Remote Identity Validation Tech Demo

- S&T is evaluating component RIV technologies that are capable of:
  1. Assessing the validity of an identity document (U.S. driver's license);
  2. Matching a selfie to the photo on the identity document; or
  3. Assessing the “liveness” of the selfie.
- The demonstration has followed a phased approach, such that each step in the RIV process is demonstrated in a separate track.



# 2023 REMOTE IDENTITY VALIDATION TECH DEMO



Science and Technology

## Track 1: ID Validation

✓ ACCEPT ID

✗ REJECT ID



Dataset of over 1,000 genuine state ID card photos

Dataset of over 1,000 fraudulent ID card photos

## Track 2: Match to Document

✓ VERIFY IDENTITY

✗ FAIL TO MATCH



Dataset of selfie photos and genuine documents from over 1,000 people

Over 1,000 mated comparisons

Over 500,000 non-mated comparisons

## Track 3: Presentation Attack Detection

✓ ACCEPT SELFIE

✗ DETECT ATTACK



Tested with over 600 diverse bona fide users

Tested with over 1,200 presentation attacks



# Remote Identity Validation Tech

## Track 2: Match to Document Overview



# Selfie and Document Dataset Collection and Composition

- A total of 1,633 volunteers participated in Remote Identity Validation Tech (RIVTD) Track 2 over two data collections
  - Maryland Test Facility (MdTF), May 2023
  - Remote Collection, September 2023
- Each volunteer used each smartphone to provide:
  - one controlled and one uncontrolled selfie image
- Test team personnel used each smartphone to collect:
  - one controlled document image (only front of document used)
- Demographics:
  - Age (self-reported)
  - Gender (self-reported)
  - Race (self-reported)
  - Skin-Tone (measured)

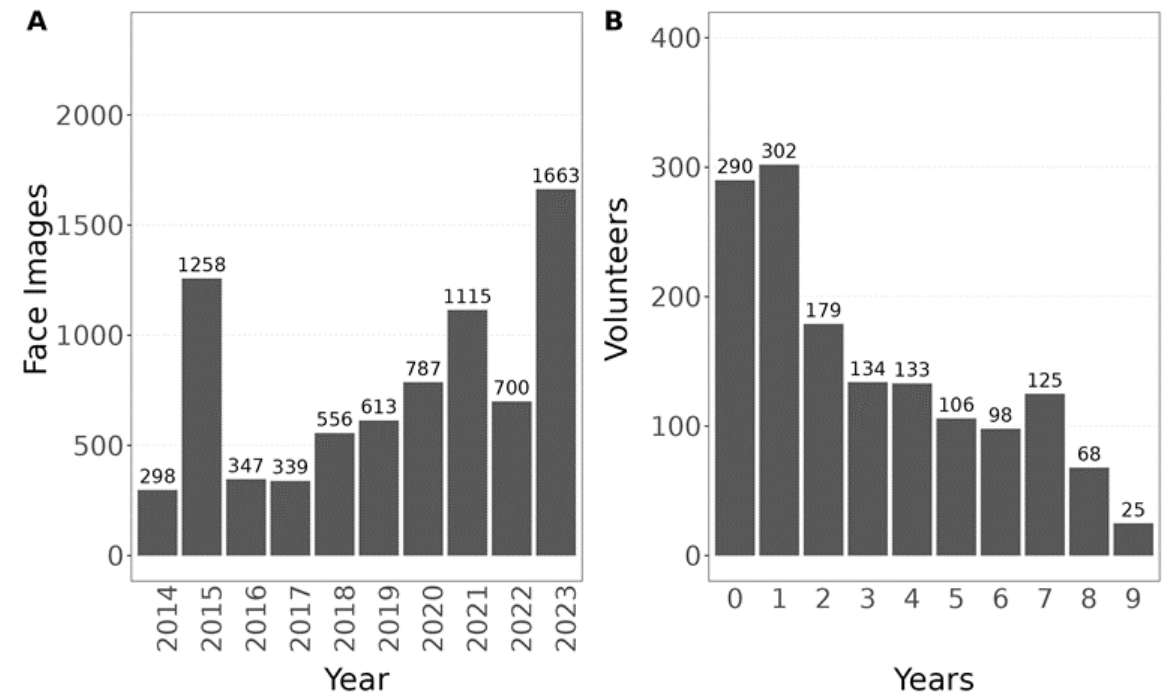
| Category  | Group    | n      |
|-----------|----------|--------|
| Gender    | Female   | 924    |
|           | Male     | 702    |
|           | Other    | 7      |
|           | Total    | 1,633  |
| Race      | Asian    | 355    |
|           | Black    | 289    |
|           | Hispanic | 467    |
|           | Other    | 94     |
|           | White    | 428    |
|           | Total    | 1,633  |
| Age Group | 18-30    | 296    |
|           | 31-45    | 525    |
|           | 46-60    | 432    |
|           | 61+      | 379    |
|           | Total    | 1,632* |

\*One volunteer did not report age information.



# Longitudinal Dataset Composition

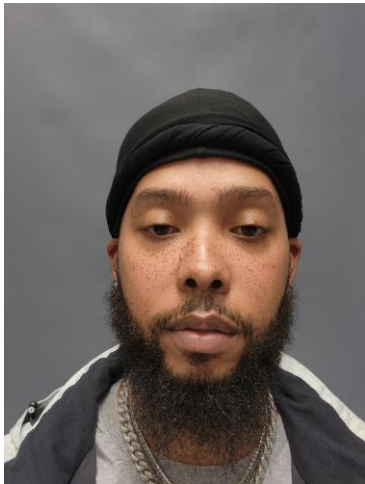
- An additional dataset, comprised of MdTF enrollment images over a 10-year span was used to calculate false match rate (FMR)
- Thousands of images, thousands of subjects, millions of non-mated comparisons evaluated
- FMR values assessed at thresholds provided with match to document systems



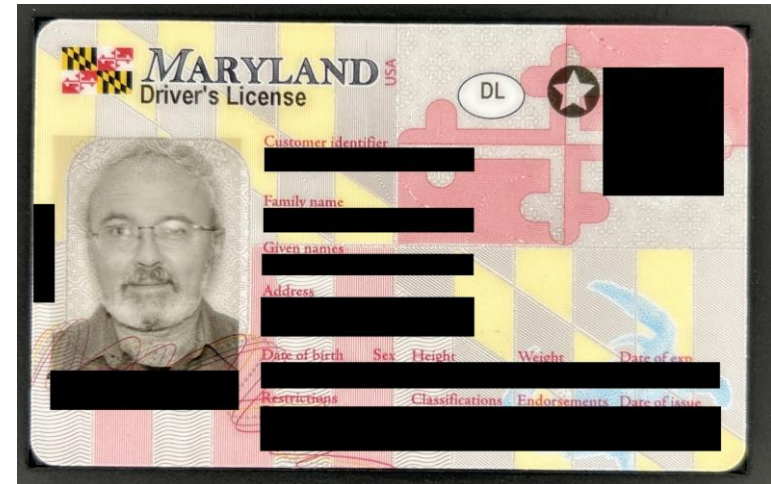
# Sample Images of Selfies and Identity Documents

- Sample images of selfies and U.S. person ID cards (e.g., driver's licenses) from test volunteers

- Selfies:



- U.S. Driver's License:

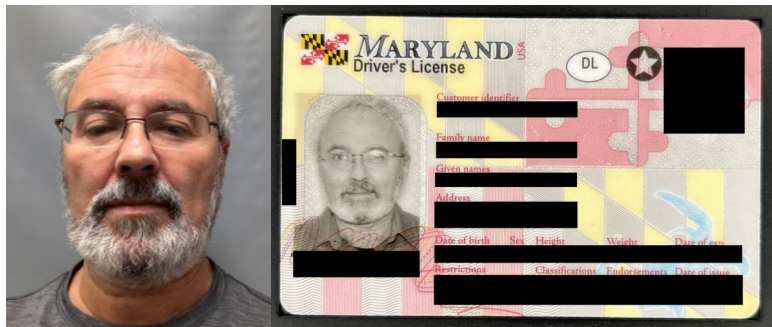


\*Volunteers shown consented to have images used in government presentations. ID documents redacted to protect privacy.

# Sample Images Across Devices

- Selfies and document images were acquired on each of three smartphones

Apple iPhone 14



Samsung Galaxy S22



Google Pixel 7



\*Volunteers shown consented to have images used in government presentations. ID documents redacted to protect privacy.

# System Requirements

- Implement the MdTF Match to ID Application Programming Interface (API)
- A single Linux-based docker container
  - HTTP server on port 8080
  - Less than 1.5 GB in size
- No outside functionality and no access to the internet
- Licensed to operate at MdTF

## The Maryland Test Facility Match-to-ID Interface 0.0.1

**OAS3**

This document specifies the API requirements for MdTF testing of algorithms that match facial biometric samples to identity document images ("match-to-id"). Match-to-ID testing at the MdTF is supported by the Department of Homeland Security, Science and Technology Directorate (DHS S&T) as part of the Remote Identity Validation Technology Demonstration (RIVTD). For more information please visit <https://mdtf.org> and <https://www.dhs.gov/science-and-technology/BI-TC>

[The MdTF - Website](#)  
[Send email to The MdTF](#)

Copyright (c) 2023, The Maryland Test Facility

### Biometric Operations ^

- POST** `/v1/create-template` Generate a template from the provided facial biometric sample or the identity document image. v
- POST** `/v1/compare-list` Compare a single template to a list of target templates. v

### Algorithm Information ^

- GET** `/v1/info` Returns basic information for the algorithm. v

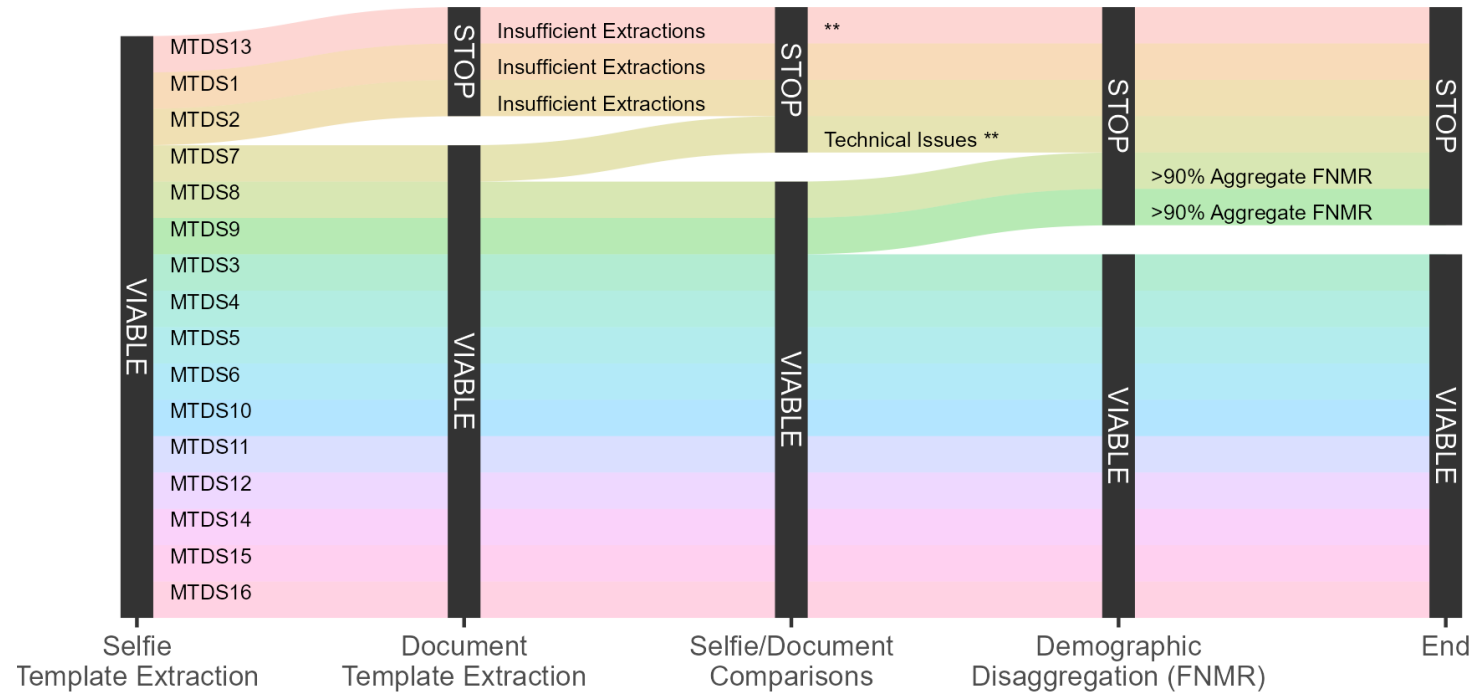
# Application and Selection Process

- All RIVTD Track 2 applications were evaluated by a panel of experts from DHS, DoD and NIST.
- 18 selfie Match to Document Systems (MTDSs) applied to participate
  - 16 were accepted
  - 10 were viable for all assessed metrics
  - Representative of industry state of the art
- Each system was given a unique alias (MTDS1, MTDS2, etc.)

# Remote Identity Validation Tech

## Track 2: Results

# Evaluation Criteria



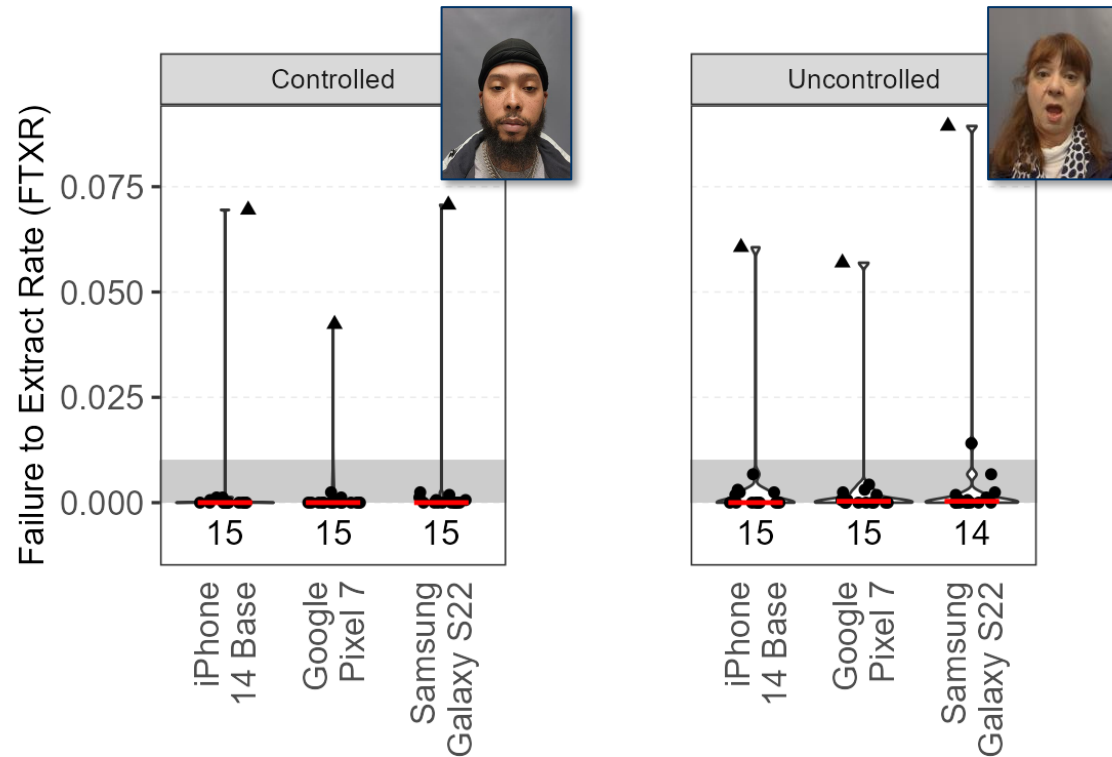
\*\* Longitudinal comparison issues as well

Nearly 38% of systems did not complete the demonstration

- All 16 MTDSs were able to extract templates from selfies
- 13 MTDSs were able to extract templates from >50% of documents
- 12 were able to successfully compare selfie and document templates
- 10 had <90% False Non-Match Rate



# Selfie Failure to Extract Rate



Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower Selfie FTXR.

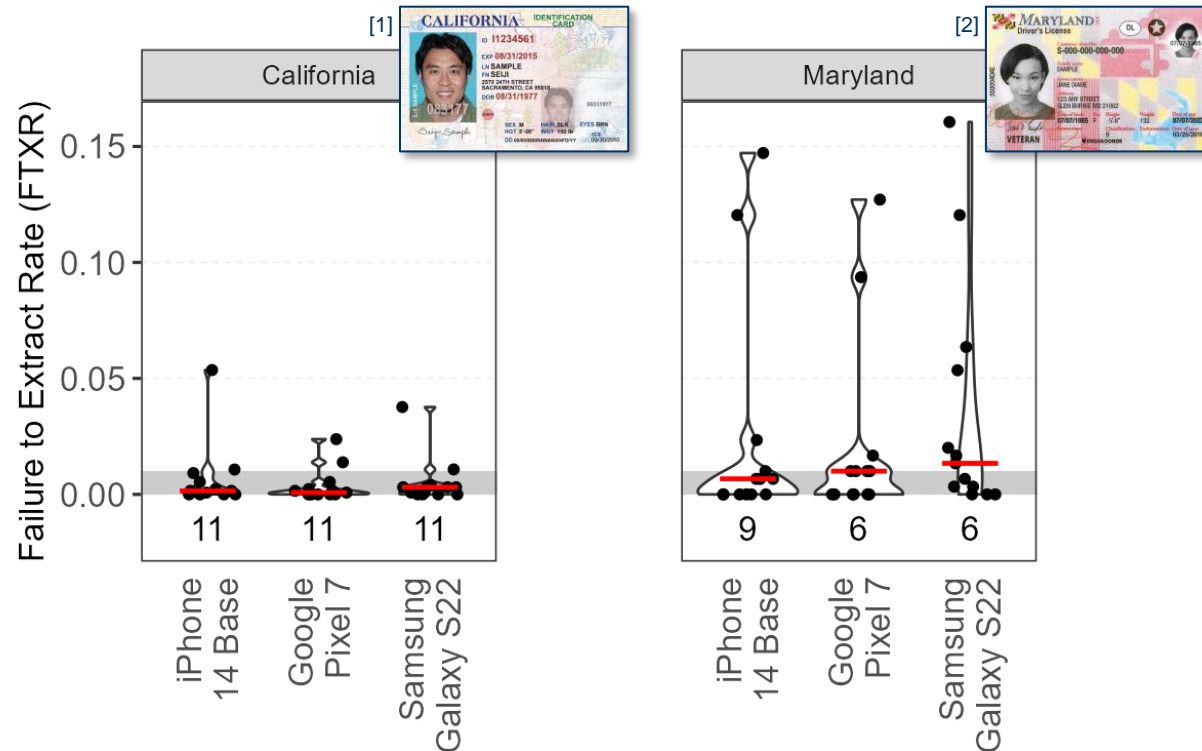
Red lines show median system combination Selfie FTXR.

Numbers indicate how many system combinations, out of 16, had 1% or lower Selfie FTXR.

- 14 of 16 systems had FTXR below 1% across devices and levels of selfie control
- Median Selfie FTXR was 0%
- Smartphone used to capture selfies had limited impact
- Larger variation in FTXR across systems observed for uncontrolled selfies
- Selfies did not pose a challenge for the majority of MTDSs

\*All volunteers shown here consented to have their images used in government presentations.

# Document Failure to Extract Rate



Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower Document FTXR.

Red lines show median system combination Document FTXR.

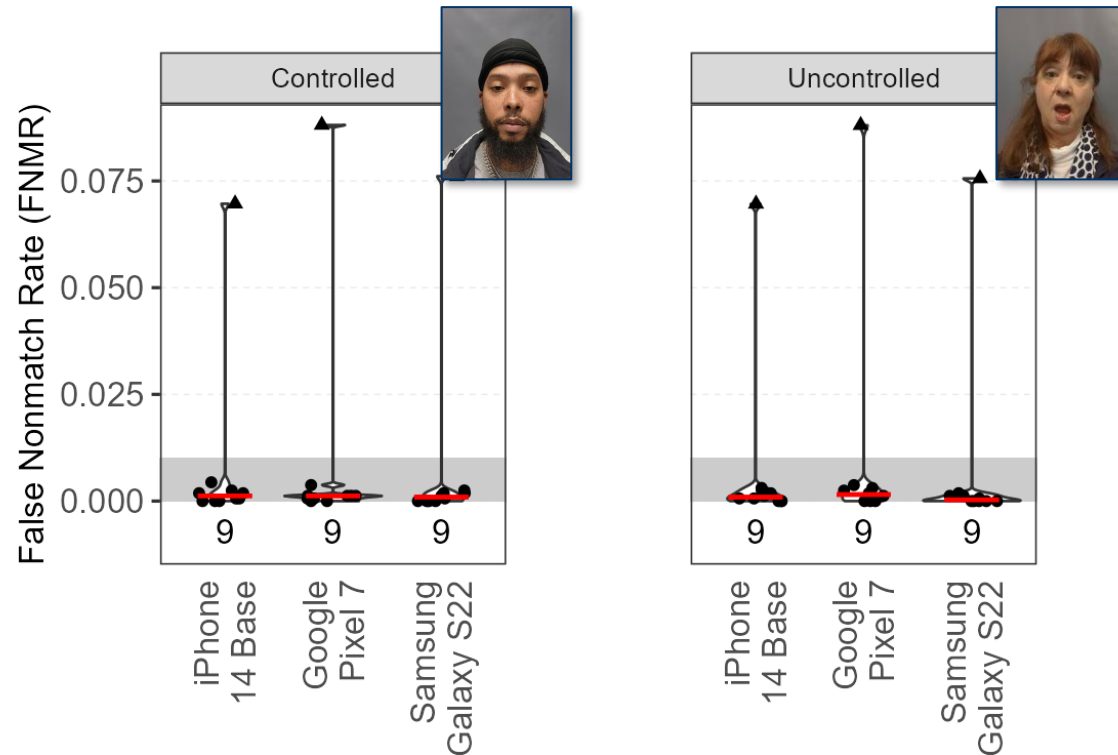
Numbers indicate how many system combinations, out of 13, had 1% or lower Document FTXR.

[1] [https://commons.wikimedia.org/wiki/File:Identity\\_card\\_of\\_the\\_State\\_of\\_California\\_sample\\_\(2010\).jpg](https://commons.wikimedia.org/wiki/File:Identity_card_of_the_State_of_California_sample_(2010).jpg)

[2] <https://www.delmarvanow.com/story/news/local/maryland/2016/05/09/mva-unveils-new-maryland-licenses-ids/84147078/>

- **Outliers with high error rates** - Three (3) systems (not shown) had Document FTXR >50% on one or more smartphones
- **Some systems had consistently low error rates** – Six of 16 systems had document FTXR below 1% across devices and states
- **Minor impacts of document state of issue and smartphone:**
  - Median Document FTXR for California IDs ranged from 0.08% (iPhone) to 0.31% (Samsung)
  - Median Document FTXR for Maryland IDs ranged from 0.67% (iPhone) to 1.34% (Samsung)

# False Non-Match Rate



Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower FNMR.

Red lines show median system combination FNMR.

Numbers indicate how many system combinations, out of 13, had 1% or lower FNMR.

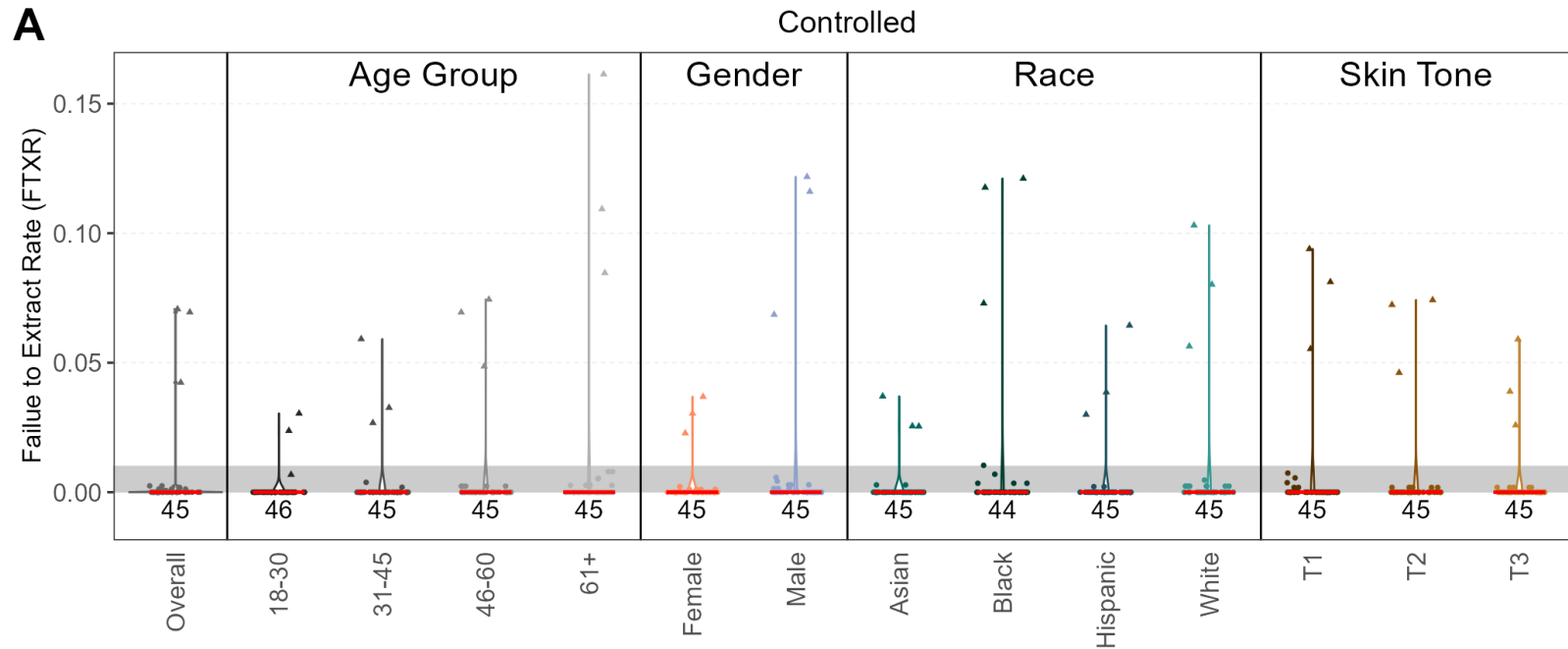
\*All volunteers shown here consented to have their images used in government presentations.



Illustrative only, not an actual identity document.

- **Outliers with high error rates** - 2 systems (not shown) had FNMR  $>90\%$  on one or more smartphones
- **Generally, high performance**
  - Document state of issue and smartphone type did not affect MTDS performance
  - Median FNMR was uniformly below 0.2%

# Demographics: Controlled Selfie Failure to Extract Rate



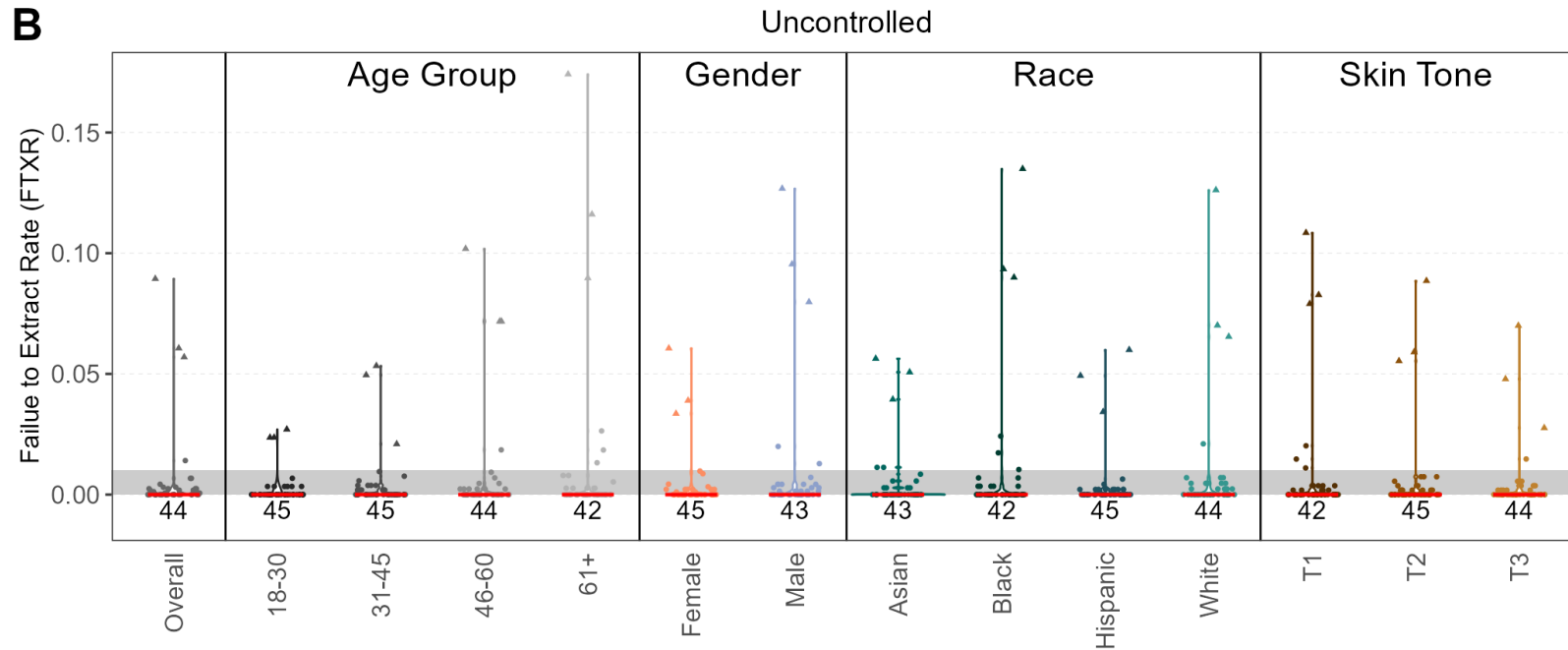
Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower controlled selfie FTXR.

Red lines show median system combination performance for controlled selfie FTXR.

Numbers indicate how many system combinations, out of 48, had 1% or lower controlled selfie FTXR. Skin tone is coded as T1: darkest skin tone tertile, T2: medium skin tone tertile, T3: lightest skin tone tertile.

- 16 MTDSs were assessed for FTXR, making 48 MTDS-smartphone combinations
- **Error rates were generally low across demographics for majority of MTDS in controlled Selfie FTXR**
- 44 of 48 MTDS-smartphone combinations met the 1% controlled Selfie FTXR benchmark for each demographic group
- Median error rates were 0%
- Though errors were low, demographic trends were observed for some system combinations

# Demographics: Uncontrolled Selfie Failure to Extract Rate



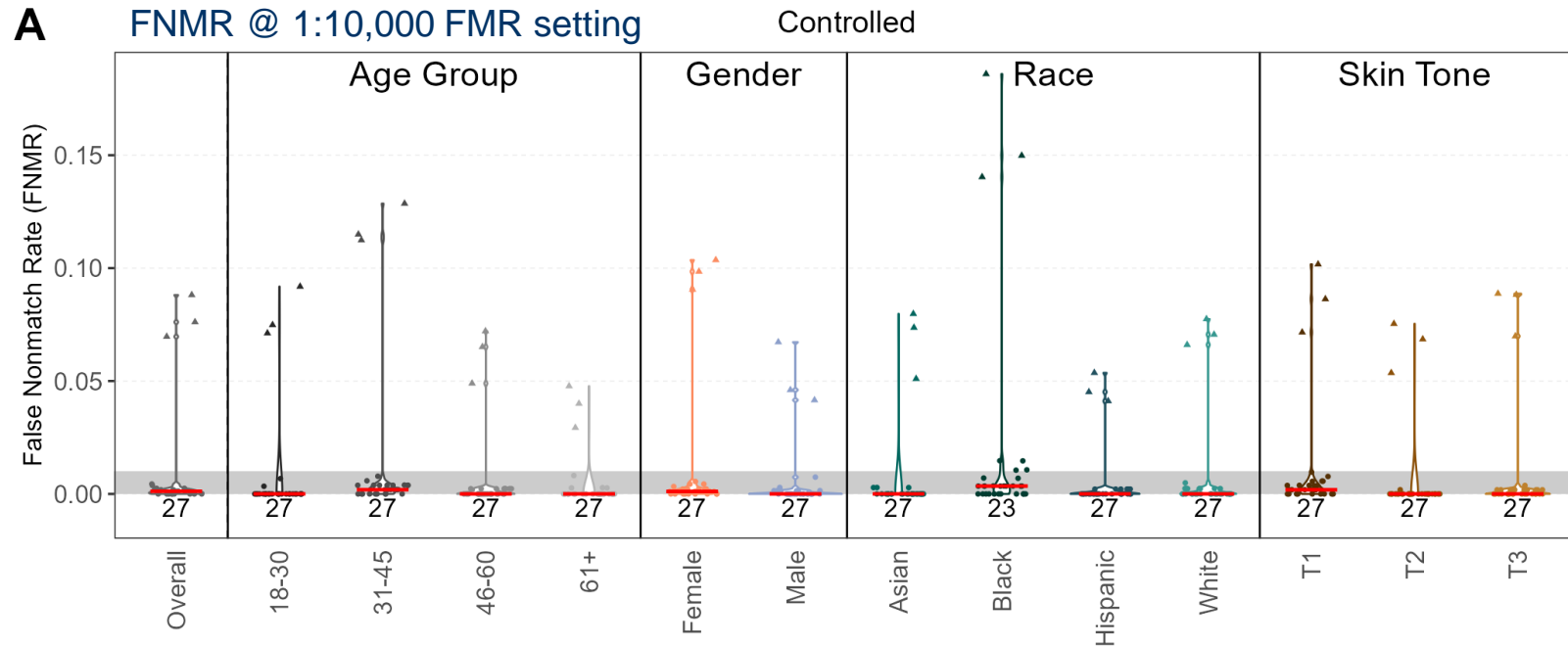
Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower Uncontrolled Selfie FTXR.

Red lines show median system combination performance for Uncontrolled Selfie FTXR.

Numbers indicate how many system combinations, out of 48, had 1% or lower Uncontrolled Selfie FTXR. Skin tone is coded as T1: darkest skin tone tertile, T2: medium skin tone tertile, T3: lightest skin tone tertile.

- 16 MTDSs were assessed for FTXR, making 48 MTDS-smartphone combinations
- **Selfie control impacted performance** - 41 of 48 system combinations met the 1% uncontrolled Selfie FTXR benchmark for each demographic group
- Median error rates were 0%
- Fewer system combinations met the 1% benchmark based on some demographic variables:
  - Gender (Male)
  - Age (61+)
  - Skin Tone (T1: darkest)

# Demographics: Controlled False Non-Match Rate



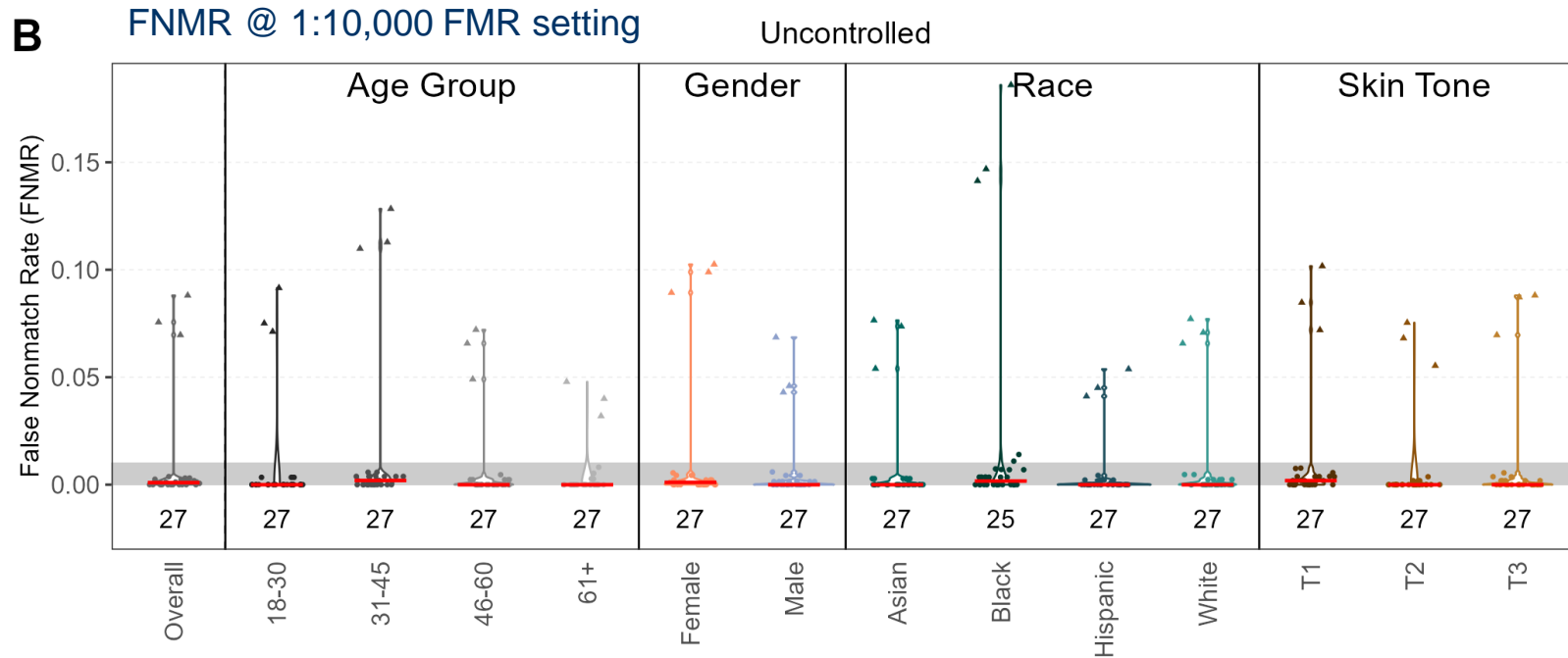
- 10 MTDSs were assessed for FNMR, making 30 MTDS-smartphone combinations
- 23 of 30 system combinations met the 1% FNMR benchmark for all demographic groups
- Median FNMR ranged from 0% (multiple groups) to 0.35% for volunteers that self-reported as Black

Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower Controlled Selfie FNMR.

Red lines show median system combination Controlled Selfie FNMR.

Numbers indicate how many system combinations, out of 30, had 1% or lower Controlled Selfie FNMR. Skin tone is coded as T1: darkest skin tone tertile, T2: medium skin tone tertile, T3: lightest skin tone tertile.

# Demographics: Uncontrolled False Non-Match Rate



- 10 MTDSs were assessed for FNMR, making 30 MTDS-smartphone combinations
- 25 of 30 system combinations met the 1% FNMR benchmark for all demographic groups
- Median FNMR ranged from 0% (multiple groups) to 0.19% for volunteers in the 31-45 age group

Points correspond to performance of combinations of smartphone and MTDS. Gray shaded area indicates 1% or lower Uncontrolled Selfie FNMR.

Red lines show median system combination Uncontrolled Selfie FNMR.

Numbers indicate how many system combinations, out of 30, had 1% or lower Uncontrolled Selfie FNMR. Skin tone is coded as T1: darkest skin tone tertile, T2: medium skin tone tertile, T3: lightest skin tone tertile.



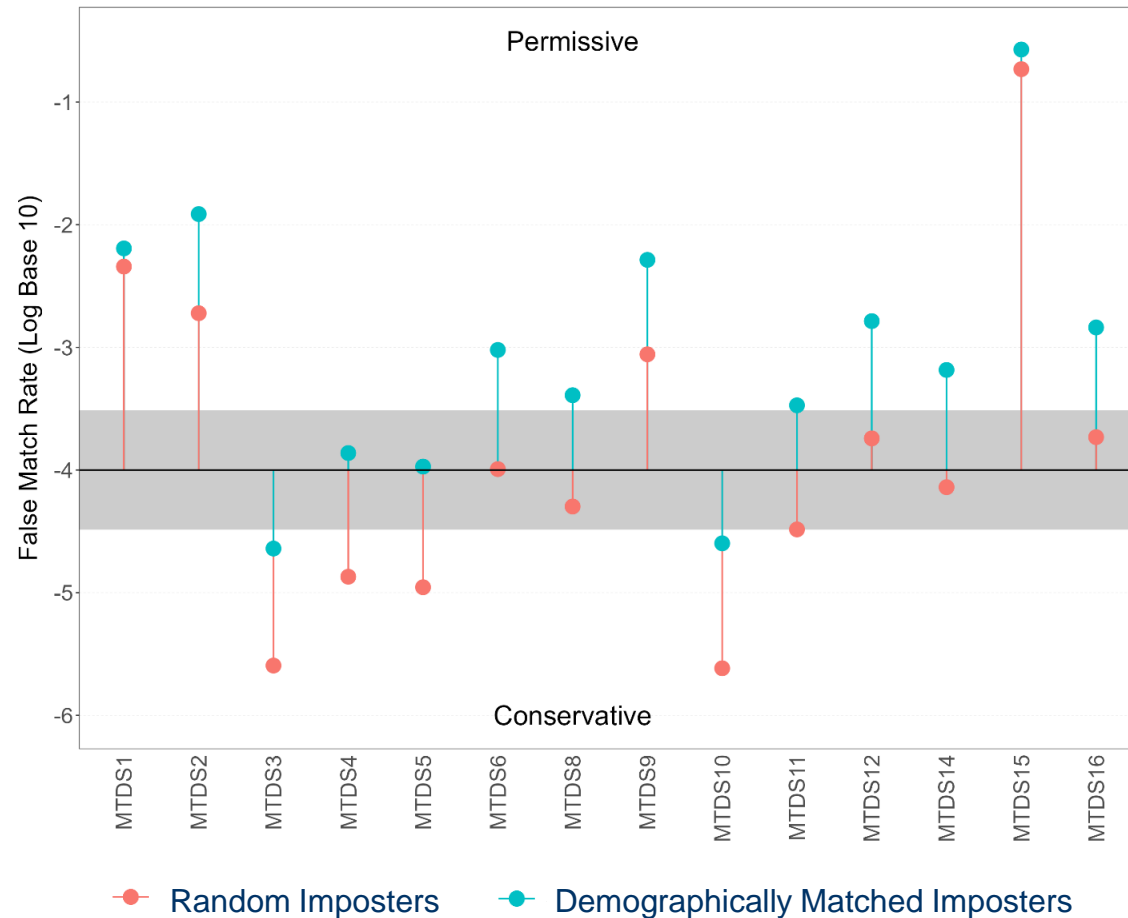
# Threshold Assessment: False Match Rate

- MTDS providers were asked to include threshold settings to achieve specific levels of FMR performance:
  - 1:1,000; **1:10,000**; 1:100,000; 1:1,000,000
  - NIST 800-63b will require FMR = 1:10,000 for authentication
- The 1:10,000 FMR settings were independently confirmed using the longitudinal dataset of face images with ground-truth demographics (Gender, Race, Age)
- FMR settings were confirmed based on:
  - Random imposters
  - Demographically matched imposters<sup>[1,2]</sup>
    - i.e., comparisons between people of the same gender, same race, and similar age

[1] Howard, John J., Yevgeniy B. Sirotnin, 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." *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2019.

[2] Grother, P. , Ngan, M. and Hanaoka, K., "Face Recognition Vendor Test Part 3: Demographic Effects, NIST Interagency/Internal Report (NISTIR)," National Institute of Standards and Technology, Gaithersburg, MD, [online], 2019, <https://doi.org/10.6028/NIST.IR.8280> (Accessed July 18, 2024)

# Threshold Assessment: False Match Rate



- FMR could be assessed for 14 MTDSs
- FMR measured at the provided 1:10,000 FMR setting
  - Permissive: FMR 3-fold larger than expected
  - Conservative: FMR 3-fold smaller than expected
- Random imposters
  - 5 of 14 MTDSs behaved as expected
  - 4 of 14 MTDSs had permissive threshold
  - 5 of 14 MTDSs had conservative threshold
- Demographically matched imposters:
  - 2 of 14 MTDSs behaved as expected
  - 10 of 14 MTDSs had permissive threshold
  - 2 of 14 MTDSs had conservative threshold

# Remote Identity Validation Tech

## Summary & Conclusions

# Results Summary

| MTDS              | 1     | 2    | 3   | 4   | 5   | 6   | 7   | 8                | 9    | 10   | 11  | 12  | 13    | 14   | 15               | 16  |
|-------------------|-------|------|-----|-----|-----|-----|-----|------------------|------|------|-----|-----|-------|------|------------------|-----|
| Selfie FTXR       | 0%    | <1%  | <1% | <1% | 0%  | 0%  | <2% | <9%              | 0%   | <1%  | 0%  | <1% | <1%   | 0%   | 0%               | <1% |
| Document FTXR     | <100% | <80% | <3% | <1% | <1% | <1% | 0%  | <13%             | 0%   | <17% | <6% | <2% | ≤100% | <13% | 0%               | <2% |
| FNMR              | NA    | NA   | <1% | <1% | 0%  | <1% | NA  | <100%            | <95% | 0%   | <1% | <9% | NA    | <1%  | <1%              | <1% |
| Threshold Setting | P     | P    | C   | C   | C   | E   | NA  | E <sup>[1]</sup> | P    | C    | C   | E   | NA    | E    | P <sup>[2]</sup> | E   |

<sup>[1]</sup> MTDS6 had a longitudinal FTXR of 56%.

<sup>[2]</sup> MTDS15 provided thresholds leading to FMR values (1:10,000) several orders of magnitude larger than expected.

|      |                |    |                             |   |              |   |
|------|----------------|----|-----------------------------|---|--------------|---|
| Key: | Met Benchmark: | X  | Did Not Meet Benchmark:     | X | Large Error: | X |
|      | Not Assessed:  | NA | FMR Outside Expected Range: | X |              |   |

- Benchmark performance was set at <1% error rate
- Threshold assessment outcomes were expected [E], permissive [P], or conservative [C]
- Performance issues centered around document processing and threshold setting
- MTDS6 met the benchmarks across all performance measures
- MTDS4 and MTDS5 met the benchmarks for FTXR and FNMR but supplied conservative thresholds
- MTDS15 met the benchmarks for FTXR and FNMR but supplied highly permissive thresholds

# Conclusions

- Face verification can perform well as part of the RIV process, but many systems still encountered performance and technical issues.
- When it works, face recognition performance in the RIV use case is largely robust against smart phone type and selfie control and across demographics.
- Some systems encounter errors processing faces from documents.
  - Notable issues were observed with rotated face images.
  - Document related processing issues were the largest identified source of error.
- Setting thresholds based on FMR targets can be challenging.
  - FMR estimates may be dataset dependent.
  - Demographically matched imposters yield higher FMR.

# Questions & Answers

- Contact information
  - [peoplescreening@hq.dhs.gov](mailto:peoplescreening@hq.dhs.gov)
  - [rivtd@mdtf.org](mailto:rivtd@mdtf.org)
- Visit our websites for additional information
  - To see additional work DHS S&T supports, visit [www.dhs.gov/science-and-technology](http://www.dhs.gov/science-and-technology).
  - For information about this and other DHS S&T technology evaluations, visit <https://mdtf.org>.

