

Influence of Race on Training Data Quality for Artificial Intelligence (AI) Algorithms

Amitha Domalpally^{1,2}, Rick Volland², Robert Slater¹, Ellie Corkery², Pamela Vargo²,
Rebecca Kutzt², James Reimers², Roomasa Channa¹, Barbara Blodi²

¹ A-EYE Unit, ² WISCONSIN READING CENTER

Abs# 2999 – F0269

ARVO 2022

Background

- It is well known that human biases make their way into artificial intelligence(AI) models.
- Reading center grader evaluation of color photographs serves as the reference standard for prospectively validating AI algorithms for diabetic retinopathy (DR).¹
- Photographs of pigmented retinas can be dark, and graders can face challenges with DR assessment. This can have an impact on both training and validation of AI algorithms

Purpose

We explored the effect of **race-related retinal pigmentation on graders' confidence** in evaluation of DR

Methods

Three diverse datasets were identified to examine the association between image quality and race. DR severity distribution, grader's confidence score and grader reproducibility was compared across race in the three datasets.

Confidence Score		Description
High confidence	CS1	image quality is excellent and DR levels can be documented
Low confidence	CS2	image quality is borderline, but DR levels can be documented
Low confidence	CS3	image quality is poor and DR levels cannot be documented.

Dataset description:

- 1. Public data:** multiethnic data of selected high-quality images that is publicly available (<https://github.com/lgiancaUTH/HEI-MED>)
- 2. DM Clinical Trial :** single site prospective trial for diabetes obtaining fundus photos
- 3. DR Clinical Trial :** baseline images from 3 large multicenter clinical trials for DR with DRCR network

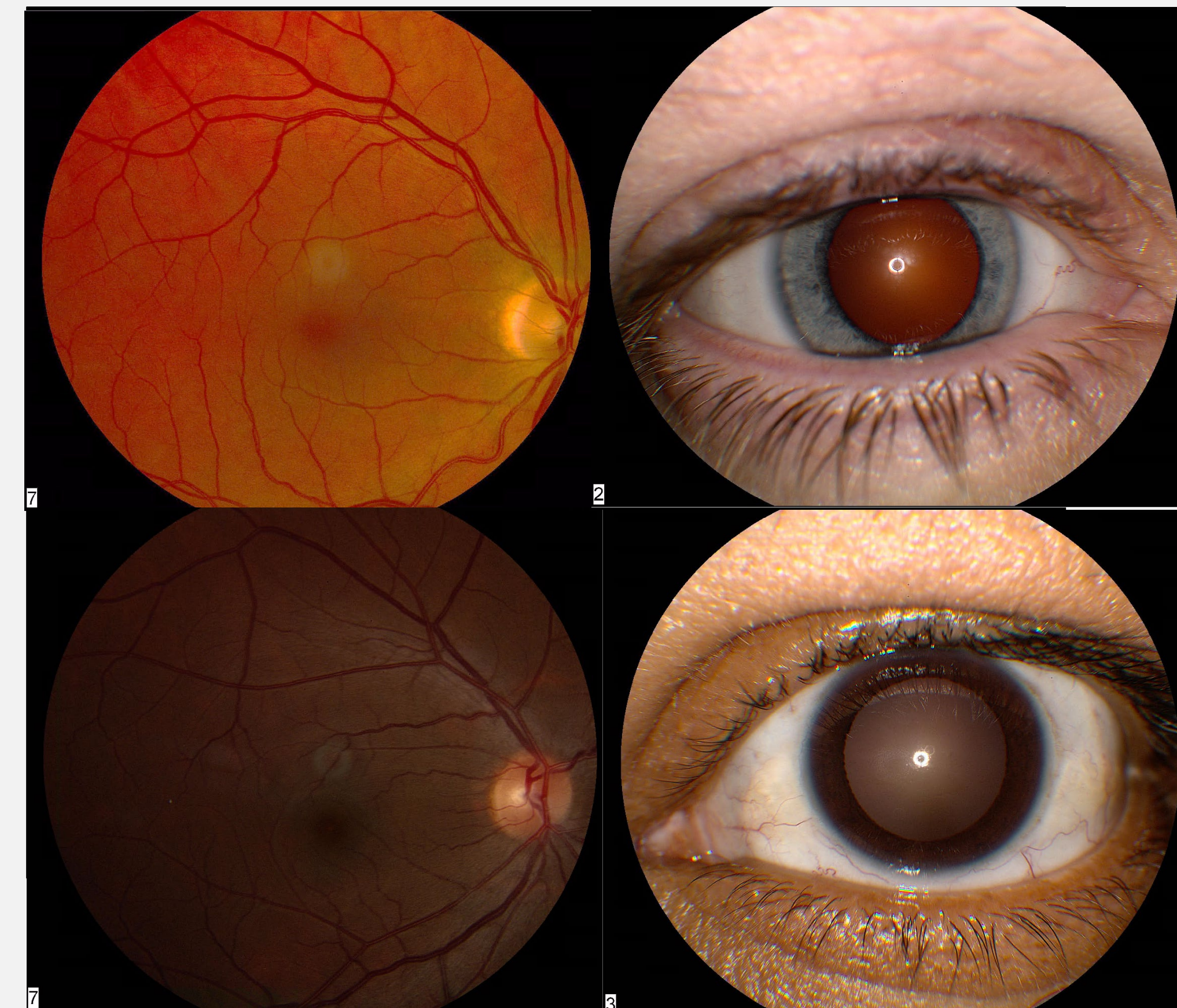


Figure 1: Fundus photograph and red reflex image of a White individual (above) and a Black individual (below). Image taken by same technician on same camera. Identification of microaneurysms in the image below can be challenging.

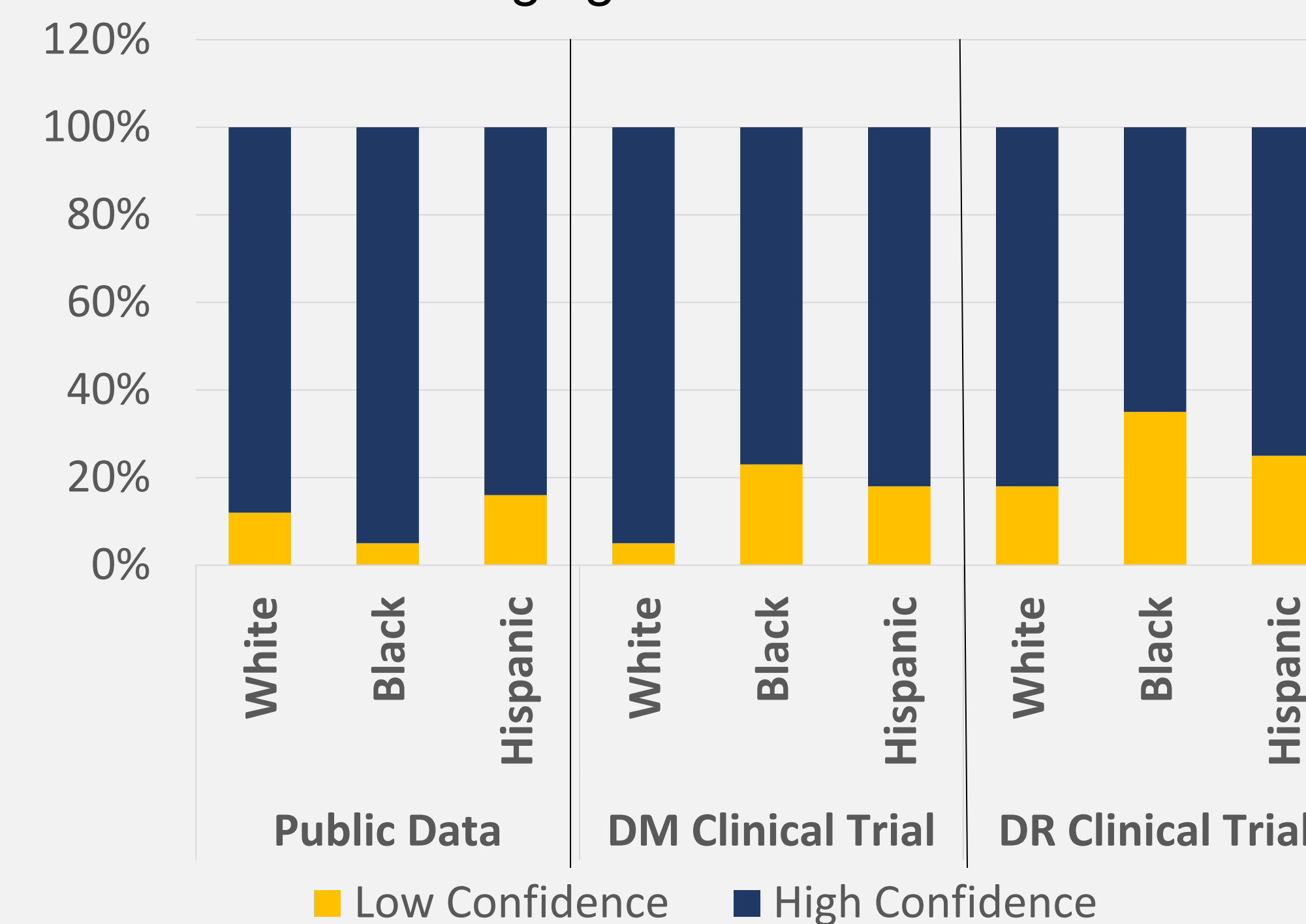


Figure 3: Grader Confidence Score and Race in three datasets

Kappa weighted	White	Black	Hispanic
Public data	0.77	0.86	0.70
DM clinical trial	0.91	NA	0.92
DR clinical trials	0.83	0.83	0.90

Table 2: Intergrader agreement and race

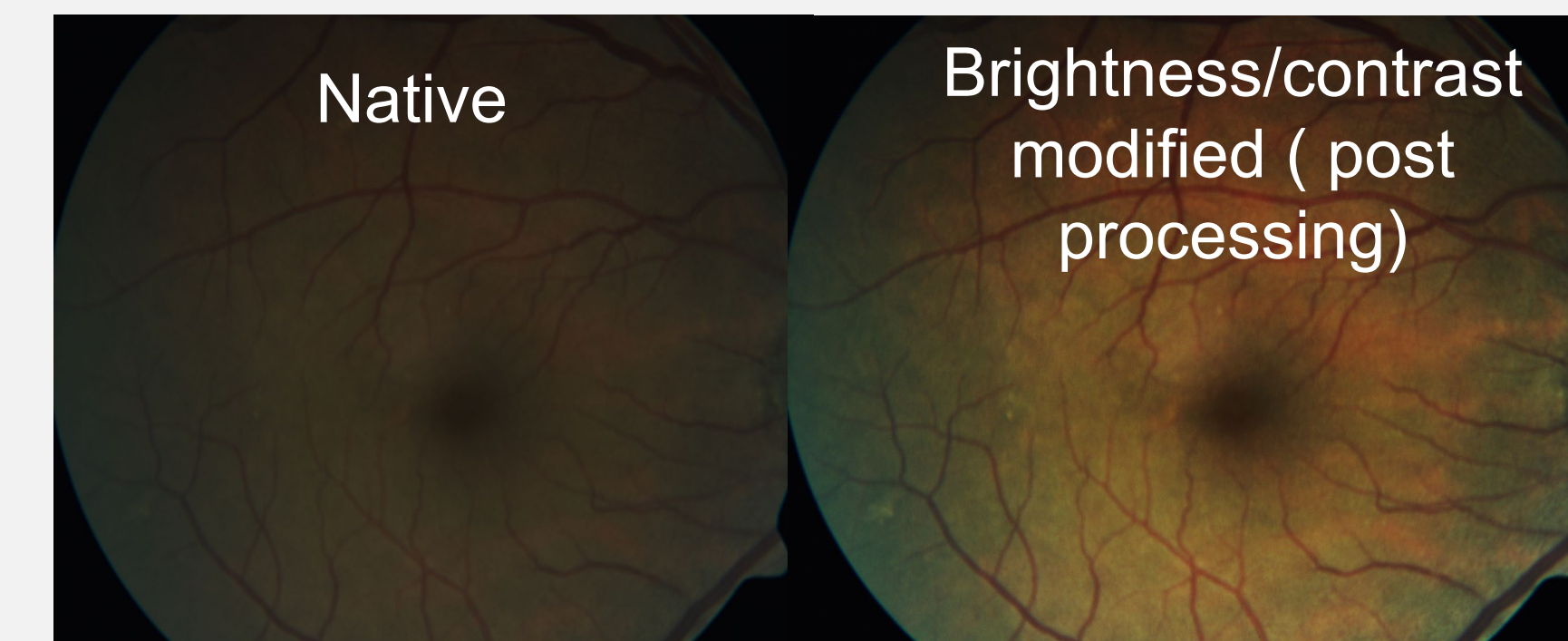


Figure 2: Changing brightness and contrast for dark images can cause artifacts in the image

	White	Black	Hispanic
Public data (328)	84 (26%)	206 (63%)	38 (12%)
DM Clinical Trial (366)	74 (20%)	22(6%)	270 (74%)
DR Clinical Trial Data (1611)	1080 (67%)	302(19%)	229 (14%)

Table 1: Distribution of race in the three datasets

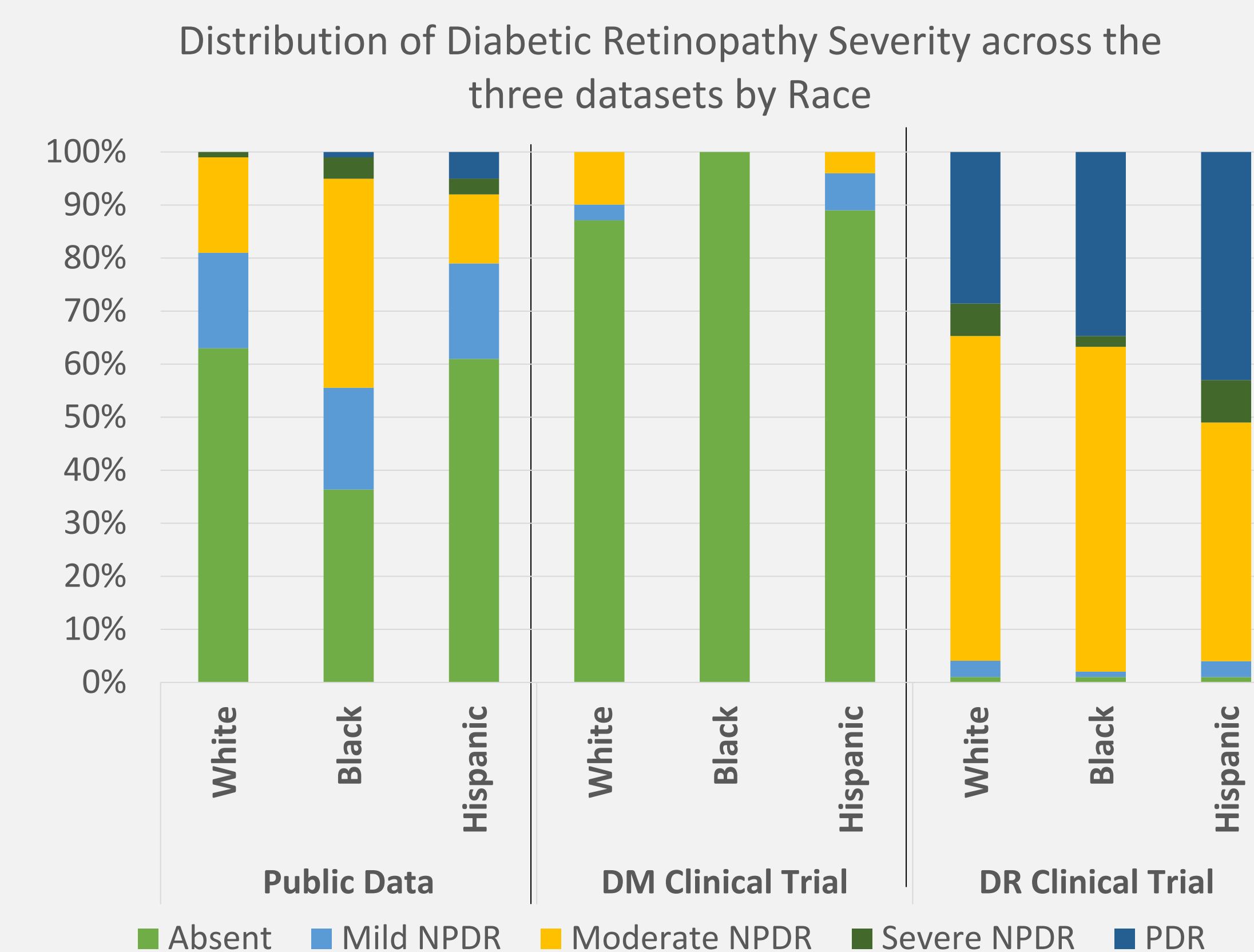


Figure 4: Distribution of DR severity scores. The Public dataset is selected for stratified representation of DR scores, DM clinical trial is skewed towards absent and mild spectrum, DR clinical trial data is skewed towards moderate DR and more.

Results

- Image quality: Grader confidence score was significantly lower in Black and Hispanic population compared to White in both clinical trials. Results were reversed in public dataset due to selective high-quality data.
- Grader reproducibility: there was no effect of race on intergrader agreement
- Distribution of DR in the DM clinical trial was skewed with no DR in Black population.

Conclusions

- Public datasets can be selective for image quality and can bias the training and validation of AI algorithms
- Grader confidence for evaluating DR features is lower in pigmented retinas across all levels of DR
- In the DM clinical trial group, it is possible that DR was under called in Black population due to difficulty in detecting early features
- Photographer education in image capture and grader training in racially diverse datasets is needed.
- Race intelligent cameras can potentially help resolve these issues

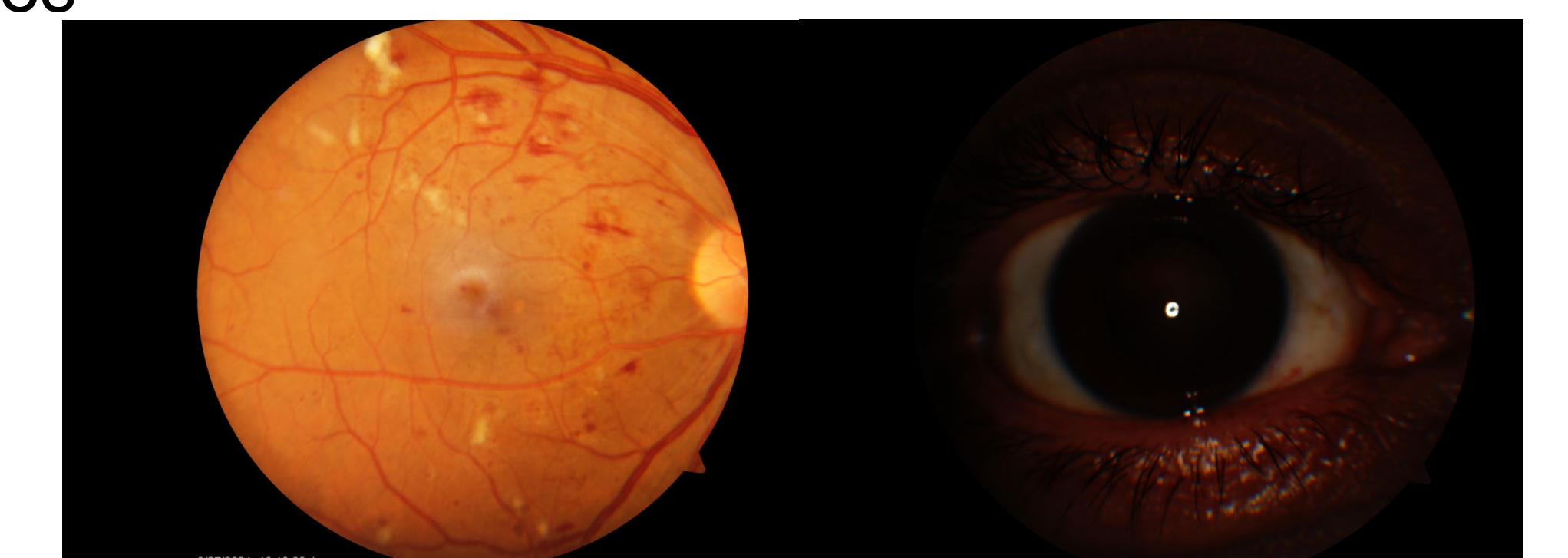


Figure 4: Fundus photo from a Black individual showing improved quality by increasing flash intensity.

Acknowledgements:

Public Data: Dr. Luca Giancardo, UTH, Houston
DM clinical trial: Dr. Ralph DeFronzo, Dr.Curtis Triplitt,, UTHSC, San Antonio
DRCRNet : The source of the data is the DRCR Retina Network (Unique Federal Award Identification Number (FAIN) UG1EY014231),but the analyses, content and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by DRCR.net.

References

1. Abramoff MD, Cunningham B, Patel B, et al. Foundational Considerations for Artificial Intelligence Using Ophthalmic Images. Ophthalmology 2021.

Commercial interest disclosures: NONE for all authors