

Short: Racial Disparities in Pulse Oximetry Cannot Be Fixed With Race-Based Correction

Neal Patwari

Washington Univ. in St. Louis

Missouri, USA

npatwari@wustl.edu

Di Huang

Washington Univ. in St. Louis

Missouri, USA

di.huang@wustl.edu

Francesca Bonetta-Misteli

Washington Univ. in St. Louis

Missouri, USA

fbonetta-misteli@wustl.edu

ABSTRACT

Pulse oximeters play a critical role in health monitoring. Pulse ox measurements have statistical bias that is a function of race, which results in higher rates of occult hypoxemia, i.e., missed detection of dangerously low oxygenation, in patients of color. This paper further characterizes the statistical distribution of pulse ox measurements, showing they also have a higher variance for patients racialized as Black, compared to those racialized as white. By analyzing the performance of hypoxemia detection as a detector, we show that no single race-based correction factor will provide equal performance by race. As a result, for racially equitable pulse oximetry, the pulse oximeter itself must be fixed, not just the hypoxemia thresholds.

ACM Reference Format:

Neal Patwari, Di Huang, and Francesca Bonetta-Misteli. 2023. Short: Racial Disparities in Pulse Oximetry Cannot Be Fixed With Race-Based Correction. In *ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE '23)*, June 21–23, 2023, Orlando, FL, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3580252.3586986>

1 INTRODUCTION

Pulse oximeters have been a key part of life-saving care during COVID-19. Those infected are advised to go to an emergency department only when their pulse ox readings fall below 90% [7]. Blood oxygenation percentage determines Medicare reimbursement, and eligibility for drug treatment therapies in most hospitals [4].

Unfortunately, pulse oximeters have a racial bias, which remains despite over 30 years of published reports [1, 4–6, 14, 16]. These reports show average pulse oximeter (SpO_2) measurements are biased higher for patients racialized as Black, particularly at low blood oxygenation levels. Since pulse oximeters are used to detect hypoxemia, i.e., critically low blood oxygenation, the higher positive bias in SpO_2 for Black patients results in more frequent occult (missed detection of) hypoxemia, as compared to white patients. The reported odds ratio for occult hypoxemia for patients racialized as Black, compared to white, is 2.2 [1], 1.4 [3] and “nearly three” [14].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHASE '23, June 21–23, 2023, Orlando, FL, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0102-3/23/06.
<https://doi.org/10.1145/3580252.3586986>

As a result, white patients are more able to obtain care than Black patients when faced with equally dangerous low blood oxygenation. For those hospitalized with COVID-19, “it is possible that unreliable measurements of the oxygen saturations have contributed to increased mortality reported in Black patients” [15]. By hiding a patient’s hypoxemia behind acceptable oxygenation readings, the pulse oximeter contributes to the racialized and gendered phenomenon that Sasha Ottey calls “health care gaslighting”, in which a patient in a medical crisis is told that they are not in danger.

The word “bias” has two meanings: *statistical bias*, and *differential assessment by group*. In the case of pulse oximetry, we observe both. Arterial oxygen saturation estimates from pulse oximeters themselves have a statistical bias, that is, the expected value of the pulse oximeter measurement is different from the arterial oxygenation saturation. This mean difference is the *statistical bias*. The racial discriminatory impact to patients stems, in part, from the racial differences in the statistical bias. That is, the bias is more severe in patients racialized as Black vs. white, which then results in disparate rates of care for hypoxemia.

If the racial difference in pulse oximetry was only in the *mean* of the measurements, one could design a race-based correction in which the statistical bias as a function of race would be subtracted out, as a function of the patient’s racial group. Indeed, this is suggested for patient care in [9] and [19]. There are many historical problems with race-based correction factors in health care, which have been used in the US to justify classifying Black patients as less in need of medical care [17]. Race-based correction is also problematic because people may be in more than one racial group; and due to the social construction of race, skin color is not the same as race.

Beyond these very real problems with race-based correction in general, in this paper we are the first, to our knowledge, to show that *no race-based correction factor will lead to equal performance in hypoxemia detection between patients racialized as Black and white*. In addition, we address the racial differences in the shape of the distribution of pulse ox measurements, in particular the variance, of pulse oximeter measurements across five racial groups. Clinicians have asked for this additional information [9].

Since pulse oximeter-based monitoring of blood oxygenation is a critical part of “smart” COVID-19 health care, both at home and in hospitals, these two contributions help to motivate and inform the effort to design racially equitable pulse oximeters. We present in Section 4 a discussion of the impact of our analysis on future pulse ox measurement and its part in smart and connected health.

2 DATA

We use the eICU Collaborative Research Database (eICU-CRD) in our retrospective analysis [10]. It is the larger of the two data sets used in the influential Sjoding et al. study [14] and is publicly available. In this section, we describe the database and how we extract pulse oximeter and arterial blood gas measurements.

The eICU-CRD has anonymized and timestamped records from over 139k unique patients during their critical care in an ICU between 2014 and 2015, from 208 hospitals, organized into several tables [10]. We extract 335k measurements of SaO₂ from 73k unique patients from the lab table, measured via arterial blood gas (ABG) test. We extract 140M measurements of SpO₂ from 196k unique patients from the vitalPeriodic (5-min median from the bedside pulse oximeter), physicalExam (only the ‘current value’ record), and nurseCharting (entered by a nurse from a bedside reading) tables. Each patient has exactly one race or ethnicity entry in the ‘patient’ table, with options are limited to: white, African American or Black, Hispanic, Asian, Native American, or “other/unknown”, pulled from the patient’s medical record [10]. We extract each arterial blood gas measurement (SaO₂), and then find the pulse oximeter value (SpO₂) for the same patient that is measured closest in time. If an SpO₂ was measured within $T = 10$ minutes of their SaO₂ measurement, and the patient’s race is not “other/unknown”, we store the (SaO₂, SpO₂) pair for use in our analysis. In total we extract 218k pairs of (SaO₂, SpO₂). Our choice of $T = 10$ matches that used in retrospective studies [4, 14, 16]. Further, we do not see significant differences in our subsequent results when we use $T = 5$ or $T = 20$ minutes.

Pairs from white patients are 80.8% of the dataset, as described in Table 1, which is skewed more white than the US population as a whole. At lower SpO₂ values, and for less well-represented racial categories, we have less data with which to characterize pulse ox performance. For example, for SpO₂ values < 87%, we do not have more than 10 data pairs from Native American patients, and data for Asian patients is often fewer than 10 pairs per SpO₂ value. Thus, in this paper, when displaying results for all five available race/ethnicity categories, we limit the analysis to SpO₂ ≥ 87%.

3 METHODS AND RESULTS

3.1 Racial Difference in Statistical Bias

Race/ Ethnicity	Data %	Bias		$P [\text{SpO}_2 - \text{SaO}_2 > 10]$
		Median	Average	
White	80.8%	1.00%	1.58%	0.0486
Black	11.3%	1.70%	2.60% ***	0.0813 ***
Asian	1.9%	1.30%	2.47% ***	0.0537
Hispanic	5.2%	1.20%	1.47%	0.0384
Native Am.	0.8%	1.40%	1.81%	0.0407
All Data	100%	1.00%	1.71%	0.0518

Table 1: Median & average pulse ox bias, & probability of large (> 10) error, vs. assigned patient race/ethnicity (with percentage in data set). * $p < 0.001$, from Welch’s t-test used to test difference of mean compared to white patients.**

From the (SpO₂, SaO₂) pairs, we first validate prior work that identified a higher statistical bias in SpO₂ values for Black patients as compared to white patients. We use “SpO₂ error” to refer to SpO₂ – SaO₂. The statistical bias is the expected value of the SpO₂ error. We use the average error over all data pairs and display this average vs. racial category in Table 1. To provide a robust estimate less influenced by very low and high errors, we also show the median difference. The statistical bias is consistently positive for all groups (1.71%), which indicates pulse oximeters are reporting values higher on average than the corresponding SaO₂ value.

In our data, the bias is statistically significantly higher for Black patients (2.60%) and Asian patients (2.47%) vs. white patients (1.58%). We use a two-sided Welch’s t-test to compare the data from each group to the white group; this test of difference in mean does not assume that the variances are identical across groups. The test gives p -values less than 10^{-13} for data from Black and Asian patients, but $p > 0.05$ for patients racialized as Hispanic or Native American.

As observed across several papers in the literature [1, 5], the statistical bias is a function of blood oxygenation percentage. How do the racial differences in average error change as a function of the measured SpO₂ value? To show this, we divide the data into three ranges of SpO₂: Low [87 – 91]; Medium [92 – 96]; High [97 – 100]. Each range is inclusive, and the lower limit of 87 is because of the low count of data points for Asian and Native American patients below 87, as discussed. The 95% confidence interval for the statistical bias, i.e., the mean of SpO₂ – SaO₂, is plotted in Fig. 1 as a function of SpO₂ range and race. In general, for all groups, the statistical bias decreases as the measured SpO₂ decreases. However, *the racial disparity in the statistical bias increases as SpO₂ decreases*. At high SpO₂, the SpO₂ bias is 0.93 higher for Black vs. white patients. In comparison, at low and medium SpO₂, the SpO₂ bias is about 1.70 higher for Black vs. white patients.

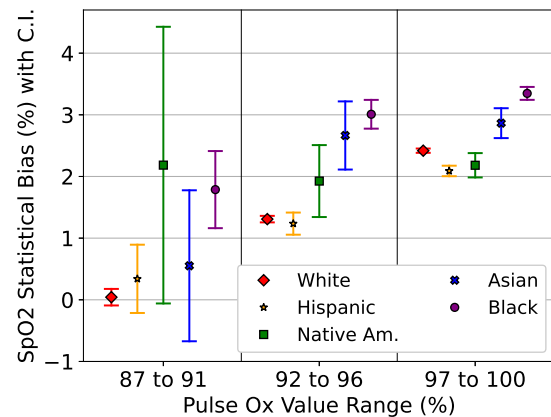


Figure 1: Statistical bias of SpO₂ values: 95% confidence interval on mean of SpO₂ – SaO₂, for 3 ranges of SpO₂ values, by race.

Low numbers of data points for Asian and Native American patients results in wider confidence intervals at all SpO₂ ranges, particularly in the lowest SpO₂ range. However, SpO₂ bias for data from Asian patients is generally higher than that from white patients and lower

than that from Black patients. The statistical bias for patients racialized as Hispanic and Native American is sometimes higher than and sometimes lower than that of patients racialized as white.

3.2 Racial Difference in Distribution Shape

In this subsection, we investigate the shape of the statistical distribution of the SpO_2 error and see that there are differences in how wide they are for different racial groups.

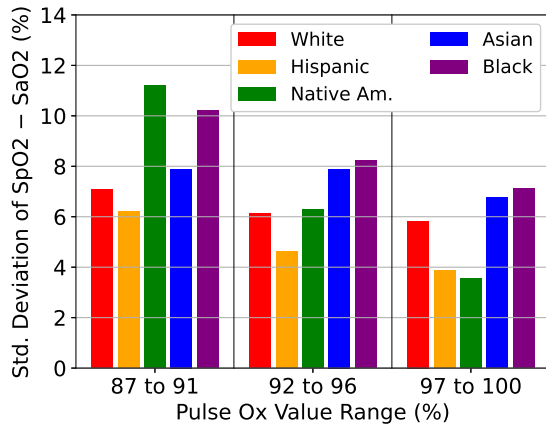


Figure 2: Estimated standard deviation of SpO_2 error in three ranges of SpO_2 values, by race.

We plot in Fig. 2 the standard deviation of the SpO_2 error, for three SpO_2 ranges and each group. We observe the standard deviation decreases as SpO_2 value increases. The standard deviation of SpO_2 error is consistently higher for Asian and Black patients vs. white patients. In fact, a one-sided F-test on the variance shows $p < 0.001$ for the variance of SpO_2 error from Asian patients or from Black patients, compared to white patients.

For more information about the shape of these distributions, we plot the probability mass functions (pmfs) for each range of SpO_2 and for data from patients racialized as Black and as white, the two racial groups with sufficient data to plot pmfs. With the exception of the high SpO_2 range, Fig. 3 shows SpO_2 error for white patients is narrower than for Black patients.

In the highest SpO_2 group, the high statistical bias may artificially make the variance lower for all racial groups, since the maximum oxygenation is 100%. This can be seen in the bottom subplot of Fig. 3, corresponding to the high SpO_2 range, in which the negative tail of the pmf is compressed compared to the heavy positive tail. Since the SpO_2 error is defined as $\text{SpO}_2 - \text{SaO}_2$, the SpO_2 is in the range 97-100, and the SaO_2 can be at most 100, there is much more opportunity for the error to be positive rather than negative.

In low and medium SpO_2 ranges, the pmf plots in Fig. 3 show the significantly heavier tails of the SpO_2 error for patients racialized as Black vs. white. For Black patients, very large errors of 10 or more have relatively high probabilities, as compared to white patients. We compute and show the probabilities of errors larger than 10, for each racial category in Table 1. Patients racialized as Black have

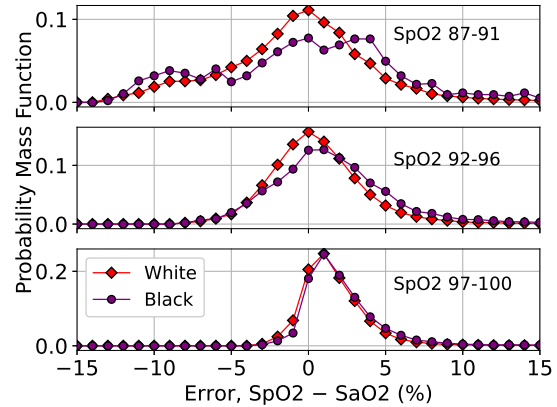


Figure 3: Probability mass function (pmf) of error, $\text{SpO}_2 - \text{SaO}_2$, in each SpO_2 range, for patients classified as white and Black.

a 67% higher chance of a large SpO_2 error compared to patients racialized as white. Using a one-sided test of equal proportions, via a normal approximation since the number of data points is large [11], we see that this difference in large error probability is statistically significant for Black compared to white patients with $p < 0.001$. The data from patients racialized as Asian has $p = 0.068$.

3.3 Hypoxemia Detection Performance

How do these distributions of SpO_2 impact detection of hypoxemia? Hypoxemia is defined as having an arterial oxygenation saturation less than 88% [14]. We use SaO_2 as a gold standard for arterial oxygenation saturation measurement, as before, in our analysis of the errors of SpO_2 . Thus we label any data pair as true hypoxemia when it has an $\text{SaO}_2 < 88\%$. We study a detector based solely on the single SpO_2 measurement taken within 10 minutes of the SaO_2 measurement, and our question is: what is the performance of a hypoxemia detector that uses one SpO_2 measurement as its input? We consider a null hypothesis H_0 : “does not have hypoxemia” and an alternate hypothesis H_1 : “has hypoxemia”; and a detector that uses a single threshold γ : it decides H_1 if the SpO_2 is less than γ , and decides H_0 if the SpO_2 is higher than γ .

The detector threshold γ (which does not need to be 88%) becomes the parameter that can be tuned to trade off performance between the two types of detection errors:

- (1) Type I error or “false alarm”: the patient does not have hypoxemia, but $\text{SpO}_2 < \gamma$ and thus the detector decides H_1 .
- (2) Type II error or “missed detection”: the patient *has* hypoxemia, but $\text{SpO}_2 > \gamma$ and the detector decides H_0 — also called “occult hypoxemia”.

Possible detector performance points are plotted for a wide range of thresholds γ , and for data from patients racialized as Black and white, in Fig. 4(a). Each performance point is labelled with the γ threshold value used to obtain it, and since SpO_2 values are all integers, we avoid confusion by setting γ values to be halfway between integers, i.e., ending in “.5”. The performance gap between Black and white patients is shown by connecting the points with a blue line.

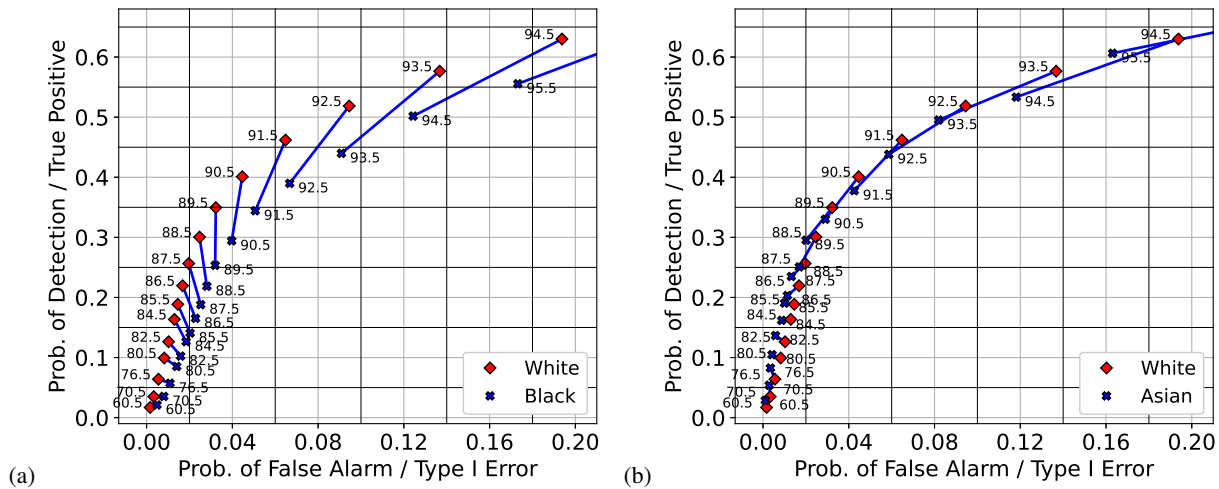


Figure 4: Probability of correct detection vs. probability of false alarm, for patients racialized as (a) Black vs. white, and (b) Asian vs. white, vs. threshold γ . Ideal performance is at the top left corner. Hypoxemia detection performance for same γ is connected (—). No single threshold results in equal performance. No race-conscious thresholds can achieve equal performance between Black and white patients, but they can achieve similar results between Asian and white patients.

Once can observe that no single threshold results in equal performance for Black and white patients. For example, $\gamma = 88.5$ results in a probability of detection of 30% and probability of false alarm of 2.5% for white patients; while the same probabilities are 22% and 2.8% for Black patients. We emphasize the point that *no single threshold can achieve the same performance point (probabilities of detection and false alarm) for both Black and white patients.*

Further, *no “race-based” correction will be able to equalize detection performance.* Medical algorithms with different settings as a function of race have typically led to better outcomes for white patients as compared to patients of color [17]. But even if we attempt to detect hypoxemia with different race-dependant thresholds with the goal of reducing disparities, we cannot achieve equal performance points. For example, we might use $\gamma = 88.5$ (as above) for patients racialized as white and $\gamma = 90.5$ for patients racialized as Black. The $\gamma = 90.5$ allows the probability of detection to be increased to 29.4% for Black patients, which would be nearly the same probability of detection as for white patients using $\gamma = 88.5$. However, the probability of false alarm would still be disparate: 2.5% and 4.0% for patients racialized as white and Black, respectively.

As a comparison, we look at the same performance points for patients racialized as Asian in Fig. 4(b). Similar to the above case, no single threshold γ produces identical detection performance for patients racialized as Asian vs. white. However, the points are nearly on the same curve. If we did consider a race-based correction factor, for example, subtracting 1.0 from the SpO₂ measurements for patients racialized as Asian, or equivalently increasing γ by 1.0 for Asian patients, we could make a race-based correction factor for γ for Asian patients that allows the detector to approximately match the hypoxemia detection performance for patients racialized as Asian and white. In Fig. 4(b), one can see that for γ in the 87.5 to 92.5 range for white patients, and γ in the 88.5 to 93.5 range for Asian

patients, performance points are close and approximately on the same curve. Further study is required for conclusive results due to having 20 times less data for Asian compared to white patients. Note that such a correction would not fix the discrimination against patients racialized as Black.

4 DISCUSSION AND FUTURE RESEARCH

If the statistical bias we validate in Section 3 were the only racial disparity in SpO₂ measurements, one could in theory have a race-based correction using different thresholds for each racial/ethnic category to achieve identical detection performance. But this is not the only disparity between patients racialized as Black and white. The significantly higher standard deviation and heavier tails of the SpO₂ distribution show that SpO₂ has larger errors for patients racialized as Black compared to white. Essentially, SpO₂ for Black patients is *both* less accurate and less precise than for white patients. Correcting for a known statistical bias, as suggested in [9, 19], is insufficient to achieve equal performance for patients racialized as Black vs. white. The literature about pulse oximeter racial disparities primarily addresses the mean, but fixing both the mean and the variance of the SpO₂ errors is required for equitable monitoring.

Regarding the racial disparity in variance, consider the impact of melanin on pulse ox measurements. In the visible light range, melanin is the main absorbent chromophore in human skin, while hemoglobin dominates absorbance in the blood. The absorption of each is represented by their *extinction spectra*, e.g., as in Figures 1.3 & 1.4 of [18]. Typical pulse oximeters use two LEDs, one in the red visible light range (~650 nm) and one in the infrared range (~900 nm). The degree to which light is attenuated is proportional to the concentration of melanin in the skin at the recording site, and red light is particularly attenuated by the presence of melanin.

One way in which the variance of SpO₂ is impacted is due to this signal attenuation. Because of the attenuation of light due to melanin, the signal to noise ratio (SNR) of the pulse oximeter light measurements is reduced, particularly at the red wavelength. As the oxygenation estimate is a function of light measurements at these two wavelengths, the lower SNR of the measurements will lead to higher variance estimates for those with darker skin.

Another way in which the variance of SpO₂ is lower among patients racialized as white may be due to the particular way in which race has been socially constructed in the US over its 400 year history. During slavery, the construction of race “allowed white men to profit from their sexual assaults on Black women” by enslaving people if they were descendant from a person and their enslaver [13]. After the end of slavery, laws enforced white supremacy with the *one-drop rule*, that any person with any degree of Black ancestry would be considered Black [2]. Both ensured that people with a wider variety of melanin levels are racialized as Black vs. white. If the statistical bias in SpO₂ is a function of melanin level, then SpO₂ values for those racialized as Black will include a wide variety of SpO₂ biases; grouped together this will make in the SpO₂ distribution among Black patients wider, i.e., with a higher variance.

It is essential that future large pulse oximetry studies record skin color in addition to race, as suggested in [8]. The lack of skin color in existing large datasets is a major limitation for research. Confounding factors, such as organ dysfunction score (e.g., SOFA), should be controlled in future analysis. For equity, SpO₂ must be unbiased despite any SOFA differences. Further, occult hypoxemia leads to further organ dysfunction [19], complicating analysis. Future studies must also include larger numbers of patients of color to quantify the implications of skin color on pulse oximetry. The high proportion of white patients in the eICU-CRD makes it difficult to analyze errors in Asian and Native American patients. We note that pulse oximeters are currently approved by the US FDA even when their subject studies include very few people of color. The FDA 510(k) process recommends, but does not require, that the testing subject population includes, “at least 2 darkly pigmented subjects or 15%” of the total pool [8]. If instead the FDA process had a requirement to over-represent of people of color in the subject population, a pulse ox study could definitively evaluate bias and potential mediation mechanisms prior to public use.

5 CONCLUSION

For pulse oximetry’s use in smart health care, it is clear, as Inioluwa Deborah Raji says, “the data that we handle are human fates, not footnotes” [12]. If racial disparities had been addressed with new pulse oximeter designs starting when they were first published over three decades years ago [15], we might not have had pulse oximetry contributing to the racial disparities in treatment and outcomes during the SARS-CoV-2 pandemic. In this paper, we use the eICU-CRD [10] data previously reported on by Sjoding et al. [14]. We contribute to the discussion with an in-depth analysis of the statistical distributions of SpO₂ errors as a function of race. We show that, in addition to a statistical bias that differs by race, that the SpO₂ error distribution is wider for patients racialized as Black vs. white.

We further show that a hypoxemia detector cannot achieve identical detection performance between white and Black patients. In particular, race-based correction cannot achieve equal performance between white and Black patients. We provide specific guidance for the direction of future research and testing and argue that smart and equitable health care requires fixing the racial disparities, both in mean and in variance, of the pulse oximeter.

REFERENCES

- [1] Erica Andrist, Mark Nuppau, Ryan P. Barbaro, Thomas S. Valley, and Michael W. Sjoding. 2022. Association of Race With Pulse Oximetry Accuracy in Hospitalized Children. *JAMA Network Open* 5, 3 (03 2022). <https://doi.org/10.1001/jamanetworkopen.2022.4584>
- [2] Daniel Brook. 2019. *The accident of color: A story of race in reconstruction*. WW Norton & Company.
- [3] Garrett W. Burnett, Blaine Stannard, David B. Wax, Hung-Mo Lin, Chantal Pyram-Vincent, Samuel DeMaria Jr., and Matthew A. Levin. 2022. Self-reported Race/Ethnicity and Intraoperative Occult Hypoxemia: A Retrospective Cohort Study. *Anesthesiology* 136, 5 (May 2022), 688–696.
- [4] Ashraf Fawzy, Tianshi David Wu, Kunbo Wang, Matthew L. Robinson, Jad Farha, Amanda Bradke, Sherita H. Golden, Yanxun Xu, and Brian T. Garibaldi. 2022. Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19. *JAMA Internal Medicine* 182, 7 (2022), 730–738.
- [5] John R. Feiner, John W. Severinghaus, and Philip E. Bickler. 2007. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesthesia & Analgesia* 105, 6 (2007), S18–S23.
- [6] Amal Jubran and Martin J. Tobin. 1990. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. *Chest* 97, 6 (1990), 1420–1425.
- [7] Amy Moran-Thomas. 2020. How a popular medical device encodes racial bias. *Boston Review* 8, 5 (2020).
- [8] Olunmi E. Okunlola, Michael S. Lipnick, Paul B. Batchelder, Michael Bernstein, John R. Feiner, and Philip E. Bickler. 2022. Pulse oximeter performance, racial inequity, and the work ahead. *Respiratory Care* 67, 2 (2022), 252–257.
- [9] Keir E.J. Philip, Robert Tidswell, and Charles McFadyen. 2021. Racial bias in pulse oximetry: more statistical detail may help tackle the problem. *British Medical Journal* 372 (2021).
- [10] Tom J. Pollard, Alistair E.W. Johnson, Jesse D. Raffa, Leo A. Celi, Roger G. Mark, and Omar Badawi. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data* 5, 1 (2018), 1–13.
- [11] Jack Prins, Don McCormack, Di Michelson, and Karen Horrell. 2012. *NIST/SEMATECH e-Handbook of Statistical Methods*. Chapter 7, Product and Process Comparisons. <https://doi.org/10.18434/M32189>
- [12] Inioluwa Deborah Raji. 2020. Handle with Care: Lessons for Data Science from Black Female Scholars. *Patterns* 1, 8 (2020), 100150.
- [13] Dorothy Roberts. 2021. *Race. In The 1619 Project*, Nikole Hannah-Jones, Caitlin Roper, Ilena Silverman, and Jake Silverstein (Eds.). One World, New York, Chapter 2, 44–61.
- [14] Michael W. Sjoding, Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. 2020. Racial bias in pulse oximetry measurement. *New England Journal of Medicine* 383, 25 (2020), 2477–2478.
- [15] Martin J. Tobin and Amal Jubran. 2022. Pulse oximetry, racial bias and statistical bias. *Annals of Intensive Care* 12, 1 (Jan 2022), 1–2.
- [16] Valeria SM Valbuena, Sarah Seelye, Michael W. Sjoding, Thomas S. Valley, Robert P. Dickson, Steven E. Gay, Dru Claar, Hallie C. Prescott, and Theodore J. Iwashyna. 2022. Racial bias and reproducibility in pulse oximetry among medical and surgical inpatients in general care in the Veterans Health Administration 2013-19: multicenter, retrospective cohort study. *bmj* 378 (2022).
- [17] Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones. 2020. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine* 383, 9 (2020), 874–882.
- [18] Lihong V. Wang and Hsin-i Wu. 2012. *Biomedical optics: principles and imaging*. John Wiley & Sons.
- [19] An-Kwok Ian Wong, Marie Charpignon, Han Kim, Christopher Josef, Anne AH de Hond, Jhalique Jane Fojas, Azade Tabaei, Xiaoli Liu, Eduardo Mireles-Cabodevila, Leandro Carvalho, et al. 2021. Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality. *JAMA Network Open* 4, 11 (2021), e2131674–e2131674.