American Journal of Epidemiology **Submitted Manuscript**

Editor's note: *An invited commentary on this article is forthcoming*

Title: Racial Disparities in Pulse Oximeter Device Inaccuracy and Estimated Clinical Impact on COVID-19 Treatment Course

Authors: Sylvia E.K. Sudat, Paul Wesson, Kim F. Rhoads, Stephanie Brown, Noha Aboelata, Alice R. Pressman, Aravind Mani, Kristen M.J. Azar

Correspondence Address: Correspondence to Dr. Kristen M.J. Azar, Center for Health Systems Research, Sutter Health, 795 El Camino Real, Palo Alto, CA, USA 94301 (e-mail: azark@sutterhealth.org)

Affiliations: Institute for Advancing Health Equity, Sutter Health, Sacramento, California, United States (Sylvia E.K. Sudat, Stephanie Brown, Alice R. Pressman, Kristen M.J. Azar); Center for Health Systems Research, Sutter Health, Walnut Creek, California, United States (Sylvia E.K. Sudat); Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States (Paul Wesson, Alice R. Pressman, Kristen M.J. Azar);

School of Medicine, University of California San Francisco, San Francisco, California, United States (Kim F. Rhoads); Philip R. Lee Institute for Health Policy Studies, University of California San Francisco, San Francisco, California, United States (Kim F. Rhoads); Alta Bates Summit Medical Center, Oakland, California, United States (Stephanie Brown); Berkeley Emergency Medical Group, San Ramon, California, United States (Stephanie Brown); Roots Community Health Center, Oakland, California, United States (Noha Aboelata); Stanford University School of Medicine, Stanford, California, United States (Noha Aboelata); California Pacific Medical Center, San Francisco, California, United States (Aravind Mani); Pacific Inpatient Medical Group, San Francisco, California, United States (Aravind Mani); Center for Health Systems Research, Sutter Health, Palo Alto, California, United States (Kristen M.J. Azar).

Funding: We have no funding to report.

Data Availability Statement: Please reach out to the corresponding author with data availability questions.

Thanks: We thank Zijun Shen for his assistance with data extraction.

Conference presentation: This analysis has not been presented at any conferences.

Preprint Information: This article was not previously published as a preprint.

Disclaimer: The views expressed in this article are those of the authors and may not represent those of their affiliated institutions.

Conflict of Interest: The authors report no conflicts of interest.

Running Head: Impact of Pulse Oximeter Device Inaccuracy

Key words: health equity; real-world data; electronic health record; blood oxygenation; coronavirus; arterial blood gas

Abbreviations: ABG, arterial blood gas; CDC, Centers for Disease Control and Prevention; CI, confidence interval; ED, emergency department; EHR, electronic health record; FDA, Food and Drug Administration; ICU, intensive care unit; NH, non-Hispanic; NHB, NH Black/African American; NHW, NH white; OMHHE, Office of Minority Health and Health Equity; SH, Sutter Health.

ABSTRACT

Arterial blood oxygen saturation measured by pulse oximetry (SpO2) may be differentially less accurate for people with darker skin pigmentation, which could potentially affect COVID-19 treatment course. We analyzed pulse oximeter accuracy and association with COVID-19 treatment outcomes using electronic health record (EHR) data from Sutter Health, a large, mixed-payer, integrated healthcare delivery system in northern California, United States (US). We analyzed two cohorts: (1) 43,753 concurrent arterial blood gas (ABG) oxygen saturation (SaO2)/SpO2 measurement pairs taken January 2020-February 2021 for Non-Hispanic white (NHW) or Non-Hispanic Black/African American (NHB) adults, and (2) 8,735 adults who went to the emergency department (ED) with COVID-19 July 2020-February 2021. Pulse oximetry systematically overestimated blood oxygenation by 1% more in NHB individuals than in NHW individuals. For people with COVID-19, this was associated with lower admission probability (-3.1 percentage-points), dexamethasone treatment (-3.1 percentage-points), and supplemental oxygen treatment (-4.5 percentage-points), as well as increased time-to-treatment: +37.2 minutes before dexamethasone initiation and +278.5 minutes before initiation of supplemental oxygen. These results call for additional investigation of pulse oximeters, and suggest that current guidelines for development, testing, and calibration of these devices should be revisited, investigated, and revised.

INTRODUCTION

In February 2021, both the United States (US) Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDC) issued public statements regarding possible suboptimal accuracy of pulse oximeters in individuals with darker skin pigmentation.(1, 2) The CDC noted the potential for under-detection of hypoxemia in darker-skinned people with COVID-19, citing data from several studies.(3-9) Despite this caution, an oxygen saturation measured by pulse oximetry of less than 94% remains a key determinant of severe disease.(10, 11) Pulse oximetry therefore influences treatment decisions for people with COVID-19, such as whether to initiate hospital-based treatments (e.g. supplemental oxygen or intravenous dexamethasone).(11, 12)

Pulse oximeters are external monitoring devices that clip onto a part of the body, typically a fingertip or earlobe, and emit a light that passes through the tissue and blood that is then detected by a sensor.(13) The amount of light not absorbed is used to estimate the blood's oxygen saturation.(13) Pulse oximeter accuracy can be affected by multiple factors such as nail polish, skin features (thickness, pigmentation, temperature), and tobacco use.(1) Several studies have also shown pulse oximeters to be less accurate at lower oxygen saturations among severely ill patients.(14, 15)

Studies typically measure pulse oximeter accuracy by comparing pulse oximetry to concurrent oxygen saturations measured by arterial blood gas (ABG), which is considered the gold standard for measuring arterial blood oxygen saturation.(16) An ABG requires a sample of arterial blood, usually taken from the radial artery. This is more invasive and uncomfortable than pulse oximetry and cannot be performed as frequently. The ABG evaluates the partial pressures of gases (oxygen and carbon dioxide, among others) and blood pH (acid-base content) in the sample, more accurately assessing oxygenation.(17) Historical studies have reported mixed results when comparing the accuracy of pulse oximeters against that of ABG. Some studies have found that pulse oximeters overestimate oxygenation, while other studies suggest the opposite.(4, 7, 14, 18-20) Overestimation of blood oxygenation by pulse oximetry in darker-skinned (as compared to lighter-skinned) individuals has been documented in the medical literature for more

than 30 years, however the clinical importance of this systematic measurement inaccuracy has not been clearly established.(5, 6, 8, 21)

A 2020 observational study of two large hospital cohorts found that people with COVID-19 who identified as Black were nearly three times more likely than those who identified as white to have occult hypoxemia not detected by pulse oximetry.(9) The well-documented differences in pulse oximeter accuracy between darker- and lighter-skinned people could be a driver of these results. Given the central role of oxygen saturation in management of COVID-19 patients, even small systematic inaccuracies have the potential to impact access to treatment and patient outcomes.(13, 22) For example, one study of COVID-19 patients in New York City, US estimated that each 1% decrease in pre-hospital oxygen saturation (measured by pulse oximetry) was associated with 7% higher odds of death.(23) Systematic overestimation of true blood oxygenation by pulse oximeter within Black patients with COVID-19 – even by a modest amount – can therefore lead to underestimation of disease severity, potentially causing delays in care and contributing to health dispartities. This is supported by a 2022 study that showed a link between predicted pulse oximeter-induced underestimation of blood oxygenation and delayed identification of treatment eligibility for supplemental oxygen in COVID-19 patients.(24) The estimated associated delay in treatment (in minutes or hours) that could be directly attributable to differential pulse oximeter measurement accuracy, however, or its impact on other aspects of the COVID-19 treatment course, is not yet known.

To address this gap, we first investigated pulse oximeter/ABG discrepancies within non-Hispanic Black/African American (NHB) and non-Hispanic white (NHW) patients of a large, integrated healthcare system in northern California, USA. After establishing the existence of differential pulse oximeter accuracy, we then analyzed possible impacts of these differences on COVID-19-related treatment and outcomes. We hypothesized that differential overestimation of oxygen saturation by pulse oximeter could negatively impact timely access to treatment for NHB individuals with COVID-19.

METHODS

Study setting

We used electronic health record (EHR) data from Sutter Health (SH), a large, mixed-payer, integrated healthcare delivery system in northern California, US, described elsewhere.(25) SH delivers comprehensive medical services in 100+ ambulatory clinics and 24 acute-care hospitals, caring for approximately 3.5 million people each year. The SH Institutional Review Board approved this study.

Analytical populations and data extraction

We extracted blood oxygen saturation measurements by ABG and by pulse oximetry – hereafter referred to as SaO2 and SpO2, respectively – between January 2020 and February 2021 at any SH hospital for adults who self-identified as either NHW or NHB (Cohort 1). Individuals who self-identified as multi-race or Hispanic were excluded. We used racial/ethnic categories consistent with the US Census categories and the US Office of Budget and Management, and use the designation "Hispanic" instead of the commonly-used term "Latinx" for consistency with US Census categories (and with SH self-reported data collection).(26) We limited our sample to NHB and NHW patients for maximal potential contrast, and also for comparability to other recent studies.(9, 27) We paired each SaO2 measurement with the nearest recorded SpO2 for the same person, truncated at +/-10 minutes from the earlier of the ABG specimen taken time or result time. We defined hypoxemia as an SaO2<90%.(28)

Our second study population (Cohort 2) comprised all adults who visited the emergency department (ED) with COVID-19 between July 2020 and February 2021 and self-identified as NHW or NHB (as above). COVID-19 visits were defined by ICD-10 hospital diagnoses (any position) of either U07.1 (COVID-19) or J12.82 (COVID-19 pneumonia), and only the first qualifying visit for each individual was included. July 2020 was chosen because the guidelines for dexamethasone treatment based on oxygen saturation were released in late June of 2020.(29) We excluded hospital visits for which we had no documented SpO2.

We compiled patient-level sociodemographic and clinical characteristics, including SpO2 at ED presentation, age, sex, homelessness status, smoking status (ever smoked), and the Charlson comorbidity index (CCI).(30, 31) We separately considered diagnosed asthma, cancer, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), cardiovascular disease, type 2 diabetes, depression, hypertension, liver disease, obesity, and renal disease. We examined the following outcomes related to the COVID-19 treatment course: (1) measurement

of oxygen saturation by SaO2, (2) time spent in the ED, (3) hospital admission, (4) dexamethasone administration and timing, (5) oxygen supplementation and timing, and (6) return to the hospital after discharge home. Detailed information on data extraction and preparation can be found in Web Appendix 1 and Web Table 1.

Statistical analysis

Pulse oximetry measurement error analysis

We compared paired SpO2/SaO2 measurements between the NHB and NHW subgroups (Cohort 1) using pairwise statistical tests, specifically the Wilcoxon rank-sum test for quasi-continuous variables, and the chi-squared test for binary variables. The Wilcoxon rank-sum test was chosen due to the non-normality of the data. The unit of analysis was the individual oxygen saturation measurement pair. We also plotted median SpO2 for each value of SaO2 to graphically explore the agreement between SaO2 and SpO2 within subgroups. To evaluate the influence of measurement timing differences, we conducted a sensitivity analysis restricted to SpO2/SaO2 measurement pairs with a time difference of zero.

Analysis of impact on COVID-19 treatment outcomes

Following the potential outcomes framework from causal inference, we used G-computation to build two counterfactuals to assess the possible impacts of differential SpO2 measurement error on COVID-19-related outcomes for NHB patients.(32)

We first modeled the relationship between baseline characteristics and each COVID-19 treatment outcome. We used the subset of NHW patients rather than the entire study population to avoid identification of the NHB subgroup via some combination of predictor variables (e.g. due to differences in utilization patterns). This assumes that the associations between baseline characteristics and outcomes are similar across the entire study population. Logistic regression was used for binary outcomes and negative binomial regression for time-related outcomes (minutes in the ED, etc.). Negative binomial models were chosen over Poisson models both to allow zero values and due to mean/variance inequality.

To estimate the impact of SpO2 measurement error, we first used the counterfactual models described above to predict the expected marginal outcomes for NHB patients under the

assumption that the relationship between the predictors and outcomes does not differ between the NHB and NHW subgroups. Mean predicted (counterfactual) outcomes were then subtracted from the observed mean NHB patient outcomes to yield marginal estimates of the mean difference. We then shifted the observed SpO2 value by the NHW-NHB measurement difference observed in the pulse oximetry measurement error analysis described previously and used this updated SpO2 value to generate a second set of predictions from our counterfactual models. This second set of predictions estimated the expected (counterfactual C_1) outcomes again assuming both that the predictor/outcome relationship and the SpO2 measurement error does not differ between the NHB and NHW subgroups. This second set of predicted (counterfactual C_2) mean outcomes were again compared to observed mean outcomes from NHB patients. To isolate the estimated effect of differential SpO2 measurement error on outcomes for NHB patients, we computed the mean difference between the two predicted (counterfactual) outcomes for each individual (C_1-C_2). If differential SpO2 measurement error does not impact a given outcome, we would expect the difference between the two counterfactual outcomes to be zero. Additional methodological detail is provided in Web Appendix 2. Standard errors and 95% confidence intervals for the estimated mean differences in outcomes were computed using the nonparametric bootstrap with 2000 iterations. Analyses were conducted in R version 4.1.0 (R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria). We also summarized descriptive data for Cohort 2 and conducted unadjusted pairwise comparisons using chi-squared and Wilcoxon rank-sum tests.

RESULTS

Pulse oximetry measurement error analysis

We identified a total of 43,753 SaO2/SpO2 measurement pairs (8,626 NHB, 35,127 NHW) (Table 1). SpO2 values were higher for the NHB population on average (median 99, mean 96.94) than for the NHW population (median 98, mean 96.42) ($p<0.001$). SaO2 values were similar numerically between the two groups (median 96%, mean 94.49% NHB; median 96%, mean 94.88% NHW), however the distributions were found to be statistically different ($p=0.001$). SaO2 measurements were lower on average than concurrent SpO2 measurements (median difference -2%, mean difference -2.45% NHB; median difference -1%, mean difference

-1.53% NHW), and the measurement difference was 1% larger for the NHB population ($p<0.001$). Hypoxemia not detected by SpO2 within the NHB patient group was 5.50% as compared to 3.01% within the NHW patient group ($p<0.001$). NHB individuals were also more likely to have an SaO2 measurement below 94% paired with an SpO2 of 94% or above (14.71% vs. 9.52%, $p<0.001$). The sensitivity analysis agreed with the main analysis (see Web Table 2).

Figure 1A shows a consistently larger discrepancy between measurements for NHB patients, although the median SpO2 is higher than SaO2 for both groups across the range of SaO2. There is also an increase in SpO2 overestimation as SaO2 decreases. The distribution of SaO2-SpO2 for NHB individuals is shifted to the left of the NHW distribution and is slightly more dispersed (see Figure 1B).

Analysis of impact on COVID-19 treatment outcomes

A total of 8,735 encounters met our study criteria (7,036 NHW, 1,699 NHB). NHB individuals were younger on average (median 51 NHB vs. 60 NHW, $p<0.001$) and had higher SpO2 at presentation (median 98% NHB vs. 97% NHW, $p<0.001$). A higher proportion of NHB patients were experiencing homelessness (3.1% vs. 1.8%, $p=0.001$), and patient insurance types differed between the two groups (see Table 2). The comorbidity profiles at ED presentation also differed, with higher proportions of cancer, COPD, and depression in the NHW patient group; the NHB patient group had a higher percentage of asthma, congestive heart failure, obesity, and renal disease (Table 2). Unadjusted outcome comparisons are shown in Web Table 3.

Differences between observed and counterfactual (C_1) COVID-19 outcomes for NHB patients represent the estimated racial bias in outcomes for the NHB patient group (Table 3). NHB patients had a higher probability of being admitted (+2.4 percentage-points, 95% CI +0.2,+4.6), of returning to the hospital after discharge home from either the ED (+8.3 percentage-points, 95% CI +4.6,+11.9) or after inpatient admission (+8.7 percentage-points, 95% CI +0.7,+16.7), and of having SaO2 measured prior to the decision to admit or discharge (+12.9 percentage-points, 95% CI +4.1,+21.7). Observed mean time to admit decision was lower than its counterfactual estimate (-15.9 minutes, 95% CI -31.5,-0.3).

The second set of counterfactual outcomes (C_2) represent the expectation under the assumption that the SpO2 value for the NHB population is shifted up by 1%, thereby equalizing

the NHB and NHW measurement error (Table 4). In contrast to the results above, there was no difference (racial bias) in the probability of admission under this assumption, although the probability of treatment with supplemental oxygen was lower (-3.0 percentage-points, 95% CI -5.2,-0.8). Compared with C_1, the difference in probability of return to the hospital after discharge home was significant, but smaller for patients not admitted (+7.0 percentage-points, 95% CI +3.3,+10.7) and larger for those who were admitted (+9.7 percentage-points, 95% CI +1.7,+17.6).

Comparing the two counterfactuals (C_1-C_2) allows us to estimate any differences in outcomes for NHB individuals that may be attributable to the differential SpO2 measurement error observed between NHB and NHW patients (Table 4). This differential measurement error was associated with a decreased probability of admission (-3.1 percentage-points, 95% CI -3.4,-2.8), dexamethasone treatment (-3.1 percentage-points, 95% CI -3.4,-2.7), treatment with supplemental oxygen (-4.5 percentage-points, 95% CI -4.9,-4.2), and return to the hospital after discharge from the ED (-1.2 percentage-points, 95% CI -1.9,-0.5). It was also associated with an increase in the probability of returning to the hospital after admission (+1.0 percentage-points, 95% CI +0.2,+1.7) and with an increase in time to treatment: 37.2 additional minutes before dexamethasone initiation (95% CI +20.1,+54.3) and 278.5 additional minutes before initiation of supplemental oxygen (95% CI +181.0,+376.0). Differences (<10 minutes) were also estimated in minutes to admit decision and spent in the ED (see Table 4).

DISCUSSION

In this observational study, we investigated whether or not pulse oximetry systematically underestimated oxygen saturation in patients who identify as NHB as compared to NHW counterparts. We also assessed whether or not differences in oxygen saturation measurement impacted admissions, care delivered, and return to the hospital post-discharge among patients with COVID-19. We found evidence of differential pulse oximeter measurement error in NHB individuals, resulting in a non-random overestimation of blood oxygenation as compared with NHW individuals. NHB individuals were also more likely to have hypoxemia not detected by pulse oximetry. For NHB patients presenting in the ED with COVID-19, we found overestimation of oxygen saturation was associated with underestimation of the need for

admission, and of treatment with dexamethasone and supplemental oxygen. Additionally, we observed associated delays in dexamethasone initiation and initiation of oxygen supplementation.

Our findings are in line with and strengthen the results of several recent studies examining SpO2 measurement error in the context of race and ethnicity in the intensive care unit (ICU).(9, 27, 33, 34) Our results also extend those of a 2022 study that established a relationship between underestimation of blood oxygenation by pulse oximeter and delays in identification of eligibility for supplemental oxygen.(24) Our results diverge from one 2022 study, in which investigators assessed pulse oximeter accuracy by race/ethnicity in patients with COVID-19 admitted to a single critical care unit in the United Kingdom (UK). Unlike our study, they reported no statistically significant difference Black patients and white patients in hypoxemia not detected by pulse oximetry. The study was limited by small sample size and a single institution, and the reported study results also do not support the conclusions of the article; the study states that the proportion of SaO2 measurements <90% with paired SpO2 measurements of 94% or above was 29.8% (71/238) for White patients and 71.1% (27/38) for Black patients. This is a more than twofold increase for Black patients, and a comparison by chi-squared test yields a p-value <0.001.

We also observed differential timing of the initial ABG for patients with SaO2 measurements taken, which was not examined by other studies. NHB patients had a much higher rate of SaO2 measurement in the ED, prior to the decision to admit or discharge home, a difference that persisted after accounting for baseline demographic and clinical characteristics. This may indicate that clinicians were less inclined to rely on the SpO2 measures alone given the full clinical pictures and other indicators of need for intervention upon presentation to the ED. A recent study looking at ICU patients found a higher likelihood of assessment by ABG among white patients (5.6%) as compared to other racial/ethnic groups (Asian 3.4%, Black 2.8%, Hispanic 1.9%), however this cohort was not focused on patients with COVID-19 and did not consider timing of the ABG measurement in relation to the decision to admit or discharge from the ED.(33)

Unlike many other studies of discrepancies in pulse oximeter accuracy between NHB and NHW patients, our study did not restrict to a specific patient subgroup (such as critical care or

ICU patients) and instead included all available measurements. This makes our results more generalizable, and suggests that the observed discrepancies hold true across a broader spectrum of hospitalized patients. Perhaps the most important contribution of the current study to the literature, however, is our novel analysis of how the discrepancy in oxygen saturation as measured by pulse oximeter is associated with differences in the delivery of evidence-based care for COVID-19. There is only one other study to-date that examines this relationship; that study examined delays in eligibility for oxygen supplementation only, and did not quantify downstream impacts on treatment course.(24) Our study therefore is an important contribution to this topic.

Missed hypoxemia by pulse oximeter in NHB patients at intake may have given a false impression to providers that NHB patients' presentations were less severe than the clinical reality. The much higher rate of ABG blood oxygen measurement for NHB patients prior to the decision to admit or discharge home could also indicate that clinicians were less inclined to rely on pulse oximetry alone to determine the need for intervention. Pulse oximeter-mediated underestimation of the severity of COVID-19 illness at ED presentation could have driven the delays in care and differences in the perceived need for therapies described in this study. These differences and delays in care have the potential to exacerbate pre-existing disparities in COVID-19 survival for NHB patients across the US.

Diagnostic tools and devices that differ in accuracy based on skin tone can be associated with differences in care, as we have shown in our study. This presents a serious threat to ongoing efforts to achieve health equity. Current FDA guidelines for pulse oximeters only require that they be tested for accuracy within at least 10 healthy subjects (producing at least 200 paired SpO2/SaO2 measurements), which must include at least 2 darkly-pigmented individuals (or 15% of the study group, whichever is larger).(35) Our results suggest that this may not be adequate to calibrate pulse oximetry machines to function adequately in NHB patients or other populations of patients with darkly pigmented skin. There is a growing recognition that one important driver of differential device inaccuracy may be due to lack of adequate representation by diverse race/ethnic groups in clinical trial and validation studies.(36-38) Recently, the US FDA established Office of Minority Health and Health Equity (OMHHE) launched an initiative to enhance equity in clinical trials by identifying barriers to clinical trial enrollment for

underrepresented populations (including, but not limited to, ethnicity, race, age, disability and geography).(39) To achieve equity in clinical device development and validation studies, novel and targeted interventions are needed to increase engagement for people from underrepresented groups in device trials and validation studies, develop metrics and standards to facilitate equitable inclusion and transparency.

This study has several limitations. First, we relied upon self-reported race/ethnicity as a proxy for skin pigmentation. Standardized and accurate assessment of skin tone (e.g. via the Fitzpatrick scale) is not consistently done in health systems for all patients.(40) Our approach will likely produce conservative estimates given that there will be a spectrum of actual skin tone in our study population. Given this, we cannot establish definitively from this study that differential device accuracy based on skin pigmentation is the sole driver of the observed differences between NHB and NHW patient measurements. Second, our study is reliant upon the accuracy and completeness of the EHR data upon which it is based. There could have been inaccuracies in the documented SaO2 and SpO2 measurement times. It is also possible that comorbidity information could have been missing due on differential healthcare utilization; this could bias our estimates if the missingness were associated with race/ethnicity. To assess for non-random distribution of missing data, we compared the rates of available diagnosis data between the NHW and NHB COVID-19 patient groups, and found no evidence of non-random missingness. Third, as this is not a population-based study, we cannot extrapolate larger impacts of differential pulse oximeter accuracy, such as underestimation of COVID-19 disease severity amongst non-hospitalized patients, or associated delays in hospitalization. We hope that our work will help inspire future studies that will delve further into these and other important questions, such as how differential device error could impact treatment course for other conditions.

The novel coronavirus SARS-CoV-2 and the resulting COVID-19 pandemic created a uniquely urgent need for treatment guidelines concurrent with the discovery of new knowledge about the virus and disease. Our study suggests that the CDC's guidelines for hospital admission and subsequent care for inpatient COVID-19 care are based on measures obtained from a clinical tool which may be insufficiently sensitive to detect hypoxia in NHB patients. These results call for additional investigation of pulse oximeters by the FDA, and suggest that current guidelines

for development, testing, and calibration of these devices should be revisited, investigated, and revised. The number of dark-skinned individuals needed to increase accuracy should be clearly defined and new guidelines disseminated. Until recommendations for systemic change occur, physicians and care providers at the frontlines of COVID-19 triage must remain vigilant. Our results suggest that NHB patients with COVID-19 may benefit from liberal use of confirmatory ABG, especially when other clinical factors are inconsistent with pulse oximetry measurements. There are also broader implications beyond COVID-19, as differential pulse oximeter accuracy has the potential to exacerbate disparities for any condition that relies upon blood oxygenation to inform clinical decision-making.

REFERENCES

1. @US_FDA. Pulse Oximeter Accuracy and Limitations: FDA Safety Communication | FDA: @US_FDA; 2021 [Available from: https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication. Accessed 9/9/2022.
2. @CDCgov. Management of Patients with Confirmed 2019-nCoV | CDC: @CDCgov; 2022. Available from: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html. Accessed 9/9/2022.
3. Luks AM, Swenson ER. Pulse Oximetry for Monitoring Patients with COVID-19 at Home. Potential Pitfalls and Practical Guidance. Ann Am Thorac Soc. 2020;17(9):1040-6.
4. Philip KEJ, Bennett B, Fuller S, Lonergan B, McFadyen C, Burns J, et al. Working accuracy of pulse oximetry in COVID-19 patients stepping down from intensive care: a clinical evaluation. BMJ Open Respir Res. 2020;7(1).
5. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg. 2007;105(6 Suppl):S18-23, tables of contents.
6. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology. 2005;102(4):715-9.
7. Striebel HW, Steinhoff U, Krause H, Kretz FJ. [The reliability of pulse oximetry monitoring of arterial oxygen saturation in centrally intubated and hypothermic patients]. Anasth Intensivther Notfallmed. 1988;23(4):200-4.
8. Cahan C, Decker MJ, Hoekje PL, Strohl KP. Agreement between Noninvasive Oximetric Values for Oxygen Saturation. Chest. 1990;97(4):814-9.
9. Sjoding M, Dickson R, Iwashyna T, Gay S, Valley T. Racial Bias in Pulse Oximetry Measurement. The New England journal of medicine. 2020;383(25).
10. COVID-19 Treatment Guidelines Panel. *Coronavirus Disease 2019 (COVID-19) Treatment Guidelines [5-12-2020]*. National Institutes of Health; 2020 [Available from: https://files.covid19treatmentguidelines.nih.gov/guidelines/archive/covid19treatmentguidelines-05-12-2020.pdf.

11. COVID-19 Treatment Guidelines Panel. *Coronavirus Disease 2019 (COVID-19) Treatment Guidelines [1-14-2021]*. National Institutes of Health; 2021 [Available from: https://files.covid19treatmentguidelines.nih.gov/guidelines/archive/covid19treatmentguidelines-01-14-2021.pdf.
12. Therapeutic Management of Hospitalized Adults With COVID-19: National Institutes of Health; [updated 12/16/2021. Available from: https://files.covid19treatmentguidelines.nih.gov/guidelines/archive/covid19treatmentguidelines-12-16-2021.pdf. Accessed 9/9/2022.
13. Moran-Thomas A. How a Popular Medical Device Encodes Racial Bias 2021 [Available from: http://bostonreview.net/science-nature-race/amy-moran-thomas-how-popular-medical-device-encodes-racial-bias. Accessed 9/9/2022.
14. Singh AK, Sahi MS, Mahawar B, Rajpurohit S. Comparative Evaluation of Accuracy of Pulse Oximeters and Factors Affecting Their Performance in a Tertiary Intensive Care Unit. J Clin Diagn Res. 2017;11(6):OC05-OC8.
15. Wilson BJ, Cowan HJ, Lord JA, Zuege DJ, Zygun DA. The accuracy of pulse oximetry in emergency department patients with severe sepsis and septic shock: a retrospective cohort study. BMC Emerg Med. 2010;10:9.
16. Davis MD, Walsh BK, Sittig SE, Restrepo RD. AARC clinical practice guideline: blood gas analysis and hemoximetry: 2013. Respir Care. 2013;58(10):1694-703.
17. Castro D, Patil S, Keenaghan M. Arterial Blood Gas. 2021 7/29/2021. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536919/. Accessed 9/9/2022.
18. Pilcher J, Ploen L, McKinstry S, Bardsley G, Chien J, Howard L, et al. A multicentre prospective observational study comparing arterial blood gas values to those obtained by pulse oximeters used in adult patients attending Australian and New Zealand hospitals. BMC Pulm Med. 2020;20(1):7.
19. Seguin P, Le Rouzo A, Tanguy M, Guillou YM, Feuillu A, Malledant Y. Evidence for the need of bedside accuracy of pulse oximetry in an intensive care unit. Crit Care Med. 2000;28(3):703-6.
20. Taylor MB, Whitwam JG. The accuracy of pulse oximeters. A comparative clinical evaluation of five pulse oximeters. Anaesthesia. 1988;43(3):229-32.
21. Bothma PA, Joynt GM, Lipman J, Hon H, Mathala B, Scribante J, et al. Accuracy of pulse oximetry in pigmented patients. S Afr Med J. 1996;86(5 Suppl):594-6.
22. Hidalgo DC, Olusanya O, Harlan E. Critical care trainees call for pulse oximetry reform. Lancet Respir Med. 2021;9(4):e37.
23. Lancet EA, Gonzalez D, Alexandrou NA, Zabar B, Lai PH, Hall CB, et al. Prehospital hypoxemia, measured by pulse oximetry, predicts hospital outcomes during the New York City COVID-19 pandemic. J Am Coll Emerg Physicians Open. 2021;2(2):e12407.
24. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19. JAMA Internal Medicine. 2022.
25. Azar KMJ, Shen Z, Romanelli RJ, Lockhart SH, Smits K, Robinson S, et al. Disparities In Outcomes Among COVID-19 Patients In A Large Health Care System In California. Health Aff (Millwood). 2020:101377hlthaff202000598.
26. Office of Budget and Management EOotP. Revisions to the standards for the classification of federal data on race and ethnicity. Federal Register 62 No. 210. p. 58782–90.

27. Wiles MD, El-Nayal A, Elton G, Malaj M, Winterbottom J, Gillies C, et al. The effect of patient ethnicity on the accuracy of peripheral pulse oximetry in patients with COVID-19 pneumonitis: a single-centre, retrospective analysis. Anaesthesia. 2022;77(2):143-52.
28. Hypoxemia (low blood oxygen): Mayo Clinic; [Available from: https://www.mayoclinic.org/symptoms/hypoxemia/basics/definition/sym-20050930. Accessed 9/9/2022.
29. COVID-19 Treatment Guidelines Panel. *The National Institutes of Health COVID-19 Treatment Guidelines Panel Provides Recommendations for Dexamethasone in Patients with COVID-19 2021*. National Institutes of Health; 2020 [Available from: https://files.covid19treatmentguidelines.nih.gov/guidelines/archive/recommendations-for-dexametha-06-25-2020.pdf. Accessed 9/9/2022.
30. Charlson ME, Pompei P, Ales KL, MacKenzie RC. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of Chronic Diseases. 1987;40(5):373-83.
31. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. American Journal of Epidemiology. 2011;173(6):676-82.
32. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol. 2011;173(7):731-8.
33. Wong AI, Charpignon M, Kim H, Josef C, de Hond AAH, Fojas JJ, et al. Analysis of Discrepancies Between Pulse Oximetry and Arterial Oxygen Saturation Measurements by Race and Ethnicity and Association With Organ Dysfunction and Mortality. JAMA Netw Open. 2021;4(11):e2131674.
34. Valbuena VSM, Barbaro RP, Claar D, Valley TS, Dickson RP, Gay SE, et al. Racial Bias in Pulse Oximetry Measurement Among Patients About to Undergo Extracorporeal Membrane Oxygenation in 2019-2020: A Retrospective Cohort Study. Chest. 2021.
35. Center for Devices and Radiological Health. *Pulse Oximeters - Premarket Notification Submissions [510(k)s]: Guidance for Industry and Food and Drug Administration Staff*. Food and Drug Administration; 2013. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug. Accessed 9/9/2022.
36. Administration USFaD. Clinical Trial Diversity [updated 5/13/2022. Available from: https://www.fda.gov/consumers/minority-health-and-health-equity/clinical-trial-diversity. Accessed 9/9/2022.
37. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. Health & social care in the community. 2004;12(5):382-8.
38. Ma MA, Gutiérrez DE, Frausto JM, Al-Delaimy WK, editors. Minority representation in clinical trials in the United States: trends over the past 25 years. Mayo Clinic Proceedings; 2021: Elsevier.
39. Administration USFaD. Enhance EQUITY in Clinical Trials [updated 6/2/2021. Available from: https://www.fda.gov/consumers/enhance-equity-initiative/enhance-equity-clinical-trials. Accessed 9/9/2022.
40. Gupta V, Sharma VK. Skin typing: Fitzpatrick grading and others. Clinics in Dermatology. 2019;37(5):430-6.

Table 1. Comparison of paired SaO2 and SpO2 measurements taken between January 2020 and February 2021 for NH Black/African American and NH white individuals

Race/ethnicity	Paired measurements[a]	Unique encounters	Unique patients	SpO2		SaO2		Measurement difference		% Hypoxemia not detected by SpO2[b]	% SaO2 below 94, not detected by SpO2[c]
				Median	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)		
NH Black/African American	8,626	2,616	2,126	99	96.94 (5.55)	96	94.49 (6.42)	-2	-2.45 (6.91)	5.5	14.71
NH White	35,127	10,514	9,113	98	96.42 (4.76)	96	94.88 (5.77)	-1	-1.53 (5.52)	3.01	9.52
p-value for difference[d]				<0.001		0.001		<0.001		<0.001	<0.001

NH = Non-Hispanic; SaO2 = % blood oxygen saturation via arterial blood gas; SpO2 = % blood oxygen saturation by pulse oximetry
[a]Each SaO2 measurement was paired nearest recorded SpO2 for the same person, truncated at +/- 10 minutes from the SaO2 specimen taken time or the SaO2 result time, whichever was earlier.
[b]SaO2 below 90% paired with a concurrent SpO2 measurement of 94% or above.
[c]SaO2 below 94% paired with a concurrent SpO2 measurement of 94% or above.
[d]Wilcoxon rank-sum test for quasi-continuous variables, chi-squared test for categorical variables.

Table 2. Baseline characteristics of COVID-19 encounters originating in the ED between July 2020 and February 2021 for individuals identifying as NH Black/African American or NH White

Variable Name	NH Black/African American, % (N=1,699)	NH White, % (N=7,036)	p-value (difference)[a]
Age (mean)[b]	51.3	58.8	<0.001
Age (median)[c]	51	60	
Female	52.2	52.3	0.919
Initial SpO2 (mean)	96.7	95.9	<0.001
Initial SpO2 (median)	98	97	
Homeless	3.1	1.8	0.001
Insurance[d]			
Medicaid	41.8	22.3	<0.001
Medicare	30.6	45.0	<0.001
Uninsured or unknown	4.9	3.0	<0.001
Charity care	2.9	1.6	<0.001
Commercial or other insurance	24.7	31.0	<0.001
Charlson comorbidity index (mean)	1.3	1.2	0.472
Comorbidities[e]			
Asthma	8.7	5.6	<0.001
Cancer	3.3	5.3	<0.001
Congestive heart failure	14.4	11.4	0.001
COPD	5.4	6.9	0.020
Cardiovascular disease	8.8	7.9	0.188
Type 2 Diabetes	8.7	8.7	0.908
Depression	1.6	3.6	<0.001
Hypertension	19.7	21.0	0.225
Liver disease	2.7	2.2	0.228
Obesity	7.9	6.3	0.017
Renal disease	19.2	14.9	<0.001
Smoker (ever)	37.6	39.2	0.211

COPD = chronic obstructive pulmonary disease; ED = emergency department; NH = Non-Hispanic; SpO2 = % blood oxygen saturation by pulse oximetry

[a]Wilcoxon rank-sum test for quasi-continuous variables, chi-squared test for percentages.

[b]Values are expressed as mean age in years

[c]Values are expressed as median age in years

[d]Will not total to 100%, as the same encounter may have more than 1 payer (e.g. Medicare/Medicaid).

[e]Comorbidities assessed as of ED presentation, and do not include conditions diagnosed during the hospital visit.

Table 3. Differences in observed COVID-19 treatment outcomes for NH Black/African American patients compared to counterfactual outcomes under the assumption of no racial bias (N=1,699).

Variable Name	Observed (O) (Mean)	Counterfactual (C_1) (Mean)	Difference (O-C_1)	95% CI
Admitted[a]	34.5	32.1	+2.4	(0.2, +4.6)
Treated with dexamethasone[a]	22.5	21.2	+1.3	(-0.9, +3.5)
Treated with supplemental oxygen[a]	29.1	27.6	+1.5	(-0.6, +3.6)
Return to the hospital after discharge home (ED visits)[a,b]	37.0	28.7	+8.3	(+4.6, +11.9)
Return to the hospital after discharge home (admissions)[a,c]	43.2	34.5	+8.7	(+0.7, +16.7)
SaO2 measured before admit/discharge decision[a,d]	62.7	49.8	+12.9	(+4.1, +21.7)
Minutes to admit decision[e]	176.6	192.5	-15.9	(-31.5, -0.3)
Minutes in the ED (admitted)[f]	541.2	502.0	+39.1	(-5.0, +83.2)
Minutes in the ED (discharged home from ED)[b]	250.6	227.3	+23.4	(-10.5, +57.2)
Minutes to dexamethasone[g]	1069.8	961.4	+108.4	(-328.0, +544.8)
Minutes to supplemental oxygen[h]	718.5	1,030.1	-311.7	(-715.2, +91.9)

CI = Confirdence interval; ED = Emergency department; NH = Non-Hispanic; SaO2 = % blood oxygen saturation via arterial blood gas; SpO2 = % blood oxygen saturation by pulse oximetry

[a]Values are expressed as mean predicted probability, shown in percentage units (%);

[b]Patients discharged home after ED visit only, N=997.

[c]Patients discharged home after inpatient admission, N=236.

[d]Patients with SaO2 measured, N=217.

[e]Patients with admit decision, N=641.

[f]Patients admitted, N=586.

[g]Patients given dexamethasone, N=383.

[h]Patients given supplemental oxygen, N=484.

Table 4. Differences in observed COVID-19 treatment outcomes for NH Black/African American patients after correction for differential SpO2 measurement (N=1,699).

Variable Name	Observed (O) (Mean)	Counterfactual (C_1) (Mean)	Counterfactual (C_2) (corrected SpO2) (Mean)	Difference (O-C_2) (corrected SpO2)	95% CI	Difference in counterfactuals (C_1-C_2)	95% CI
Admitted[a]	34.5	32.1	35.2	-0.7	(-2.9, +1.6)	-3.1	(-3.4, -2.8)
Treated with dexamethasone[a]	22.5	21.2	24.3	-1.8	(-4.0, +0.5)	-3.1	(-3.4, -2.7)
Treated with supplemental oxygen[a]	29.1	27.6	32.1	-3.0	(-5.2, -0.8)	-4.5	(-4.9, -4.2)
Return to the hospital after discharge home (ED visits)[ab]	37.0	28.7	30.0	+7.0	(+3.3, +10.7)	-1.2	(-1.9, -0.5)
Return to the hospital after discharge home (admissions)[ac]	43.2	34.5	33.5	+9.7	(+1.7, +17.6)	+1.0	(+0.2, +1.7)
Minutes to admit decision[d]	176.6	192.5	186.5	-9.9	(-24.2, +4.4)	+6.0	(+3.4, +8.6)
Minutes in the ED (admitted)[e]	541.2	502.0	498.5	+42.7	(-1.2, +86.5)	+3.6	(+1.3, +5.8)
Minutes in the ED[b] (discharged home from ED)	250.6	227.3	236.4	+14.3	(-20.6, +49.1)	-9.1	(-12.2, -6.0)
Minutes to dexamethasone[f]	1069.8	961.4	924.2	+145.6	(-287.1, +578.3)	+37.2	(+20.1, +54.3)
Minutes to supplemental oxygen[g]	718.5	1030.1	751.7	-33.2	(-370.3, +304.0)	+278.5	(+181.0, +376.0)

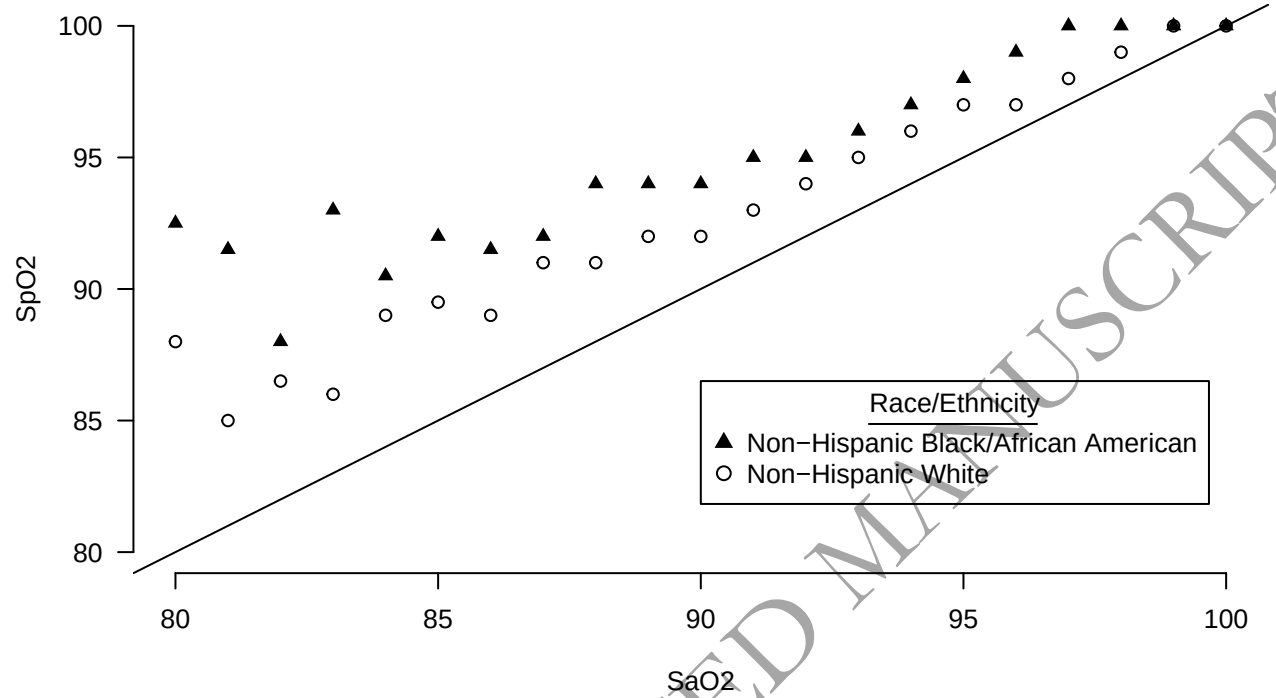
CI = Confidence interval; ED = Emergency department; NH = Non-Hispanic; SaO2 = % blood oxygen saturation via arterial blood gas; SpO2 = % blood oxygen saturation by pulse oximetry

[a]Values are expressed as mean predicted probability, shown in percentage units (%).

[b]Patients discharged home after ED visit only, N=997.
[c]Patients discharged home after inpatient admission, N=236.
[d]Patients with admit decision, N=641.
[e]Patients admitted, N=586.
[f]Patients given dexamethasone, N=383.
[g]Patients given supplemental oxygen, N=484.

Figure 1. Comparisons between paired arterial blood gas (SaO2) and pulse oximetry (SpO2) blood oxygenation measurements taken between January 2020 and February 2021 by race/ethnicity (N=8,626 non-Hispanic Black, N=35,127 non-Hispanic white). SaO2 and SpO2 are expressed in percentage units.

A)



B)