Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19

Ashraf Fawzy, MD, MPH; Tianshi David Wu, MD, MHS; Kunbo Wang, MS; Matthew L. Robinson, MD; Jad Farha, MD; Amanda Bradke, MD, MA; Sherita H. Golden, MD, MHS; Yanxun Xu, PhD; Brian T. Garibaldi, MD, MEHP

IMPORTANCE Pulse oximetry guides triage and therapy decisions for COVID-19. Whether reported racial inaccuracies in oxygen saturation measured by pulse oximetry are present in patients with COVID-19 and associated with treatment decisions is unknown.

OBJECTIVE To determine whether there is differential inaccuracy of pulse oximetry by race or ethnicity among patients with COVID-19 and estimate the association of such inaccuracies with time to recognition of eligibility for oxygen threshold–specific COVID-19 therapies.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study of clinical data from 5 referral centers and community hospitals in the Johns Hopkins Health System included patients with COVID-19 who self-identified as Asian, Black, Hispanic, or White.

EXPOSURES Concurrent measurements (within 10 minutes) of oxygen saturation levels in arterial blood (SaO₂) and by pulse oximetry (SpO₂).

MAIN OUTCOMES AND MEASURES For patients with concurrent SpO₂ and SaO₂ measurements, the proportion with occult hypoxemia (SaO₂<88% with concurrent SpO₂ of 92%-96%) was compared by race and ethnicity, and a covariate-adjusted linear mixed-effects model was produced to estimate the association of race and ethnicity with SpO₂ and SaO₂ difference. This model was applied to identify a separate sample of patients with predicted SaO₂ levels of 94% or less before an SpO₂ level of 94% or less or oxygen treatment initiation. Cox proportional hazards models were used to estimate differences by race and ethnicity in time to recognition of eligibility for guideline-recommended COVID-19 therapies, defined as an SpO₂ level of 94% or less or oxygen treatment initiation. The median delay among individuals who ultimately had recognition of eligibility was then compared.

RESULTS Of 7126 patients with COVID-19, 1216 patients (63 Asian [5.2%], 478 Black [39.3%], 215 Hispanic [17.7%], and 460 White [37.8%] individuals; 507 women [41.7%]) had 32 282 concurrently measured SpO₂ and SaO₂. Occult hypoxemia occurred in 19 Asian (30.2%), 136 Black (28.5%), and 64 non-Black Hispanic (29.8%) patients compared with 79 White patients (17.2%). Compared with White patients, SpO₂ overestimated SaO₂ by an average of 1.7% among Asian (95% CI, 0.5%-3.0%), 1.2% among Black (95% CI, 0.6%-1.9%), and 1.1% among non-Black Hispanic patients (95% CI, 0.3%-1.9%). Separately, among 1903 patients with predicted SaO₂ levels of 94% or less before an SpO₂ level of 94% or less or oxygen treatment initiation, compared with White patients, Black patients had a 29% lower hazard (hazard ratio, 0.71; 95% CI, 0.63-0.80), and non-Black Hispanic patients had a 23% lower hazard (hazard ratio, 0.77; 95% CI, 0.66-0.89) of treatment eligibility recognition. A total of 451 patients (23.7%) never had their treatment eligibility recognized, most of whom (247 [54.8%]) were Black. Among the remaining 1452 (76.3%) who had eventual recognition of treatment eligibility, Black patients had a median delay of 1.0 hour (95% CI, 0.23-1.9 hours; P = .01) longer than White patients. There was no significant median difference in delay between individuals of other racial and ethnic minority groups and White patients.

CONCLUSIONS AND RELEVANCE The results of this cohort study suggest that racial and ethnic biases in pulse oximetry accuracy were associated with greater occult hypoxemia in Asian, Black, and non-Black Hispanic patients with COVID-19, which was associated with significantly delayed or unrecognized eligibility for COVID-19 therapies among Black and Hispanic patients. This disparity may contribute to worse outcomes among Black and Hispanic patients with COVID-19.

Published online May 31, 2022.

© 2022 American Medical Association. All rights reserved.
The modern pulse oximeter provides a noninvasive estimate of arterial blood oxygen saturation levels based on the relative absorbance of 2 wavelengths of light and the pulsatile flow of arterial blood (SpO2). Since its development in the 1970s, pulse oximetry has been integrated into the routine monitoring of hospitalized patients, particularly those with respiratory illnesses. Inaccurate estimation of true arterial blood oxygen saturation levels (SaO2) by pulse oximetry can occur for various physiologic, pathologic, technical or iatrogenic reasons. However, several studies have reported a systematic overestimation of SpO2 compared with SaO2 among individuals with skin of darker pigmentation compared with individuals with lighter pigmentation. Recently, Sjoding et al reported 3 times the frequency of occult hypoxemia among individuals self-identified as Black compared with White race, while Wong et al reported significantly greater risk of occult hypoxemia for Asian and Black patients and relatively higher in-hospital mortality for Black patients with occult hypoxemia compared with White patients.

Pulse oximetry has played a prominent role in guiding triage and therapy throughout the COVID-19 pandemic, during which recommendations for hospitalization and therapy have been commonly based on SpO2 thresholds. For example, while the US Food and Drug Administration has authorized the use of remdesivir in treating hospitalized patients with COVID-19, most institutions use a threshold of an SpO2 of 94% or less or the use of supplemental oxygen to determine eligibility for treatment with remdesivir. Similarly, dexamethasone is frequently used for patients who require supplemental oxygen. Overestimation of SpO2 may be associated with premature deescalation of therapies or hospital discharge, or could be associated with the delay or withholding of therapies that shorten the disease course, slow progression, or reduce mortality. SpO2 is also frequently used as a variable in clinical risk predictors, either alone or as part of the ratio of SpO2 to the fraction of inspired oxygen. An overestimation of oxygen saturation levels could be associated with an underappreciation of clinical risk as presented by these calculators.

To our knowledge, the performance of pulse oximetry in patients with COVID-19 and its potential association with clinical decision-making remains unexplored and can help identify factors that may explain the disproportionate COVID-19 mortality experienced by patients of certain racial and ethnic minority groups. The objective of this study was to investigate the frequency and magnitude of pulse oximeter bias in estimation of SaO2 associated with race and ethnicity within a diverse cohort of patients with COVID-19 who were treated in a large multihospital health system. Furthermore, this study aimed to determine the potential association of such biases with recognition of eligibility for guideline-recommended therapies. As a respiratory illness that affects many individuals with disproportionate morbidity and mortality among racial and ethnic minority populations, the COVID-19 pandemic provides a unique opportunity to investigate the effect of race and ethnicity-based inaccuracy of pulse oximetry.

Key Points

**Question** Are there systematic racial and ethnic biases in pulse oximetry among patients with COVID-19, and is there an association between such biases and unrecognized or delayed recognition of eligibility for oxygen threshold-specific therapy?

**Findings** In this retrospective cohort study of 7126 patients with COVID-19, an analysis of 1216 patients with oxygen saturation levels that were concurrently measured by pulse oximetry and arterial blood gas demonstrated that pulse oximetry overestimated arterial oxygen saturation among Asian, Black, and Hispanic patients compared with White patients. Separately, among 6673 patients with pulse oximetry measurements and available covariate data, predicted overestimation of arterial oxygen saturation levels by pulse oximetry among 1903 patients was associated with a systematic failure to identify Black and Hispanic patients who were qualified to receive COVID-19 therapy and a statistically significant delay in recognizing the guideline-recommended threshold for initiation of therapy.

**Meaning** The study results suggest that overestimation of arterial oxygen saturation levels by pulse oximetry occurs in patients of racial and ethnic minority groups with COVID-19 and contributes to unrecognized or delayed recognition of eligibility to receive COVID-19 therapies.

**Methods**

The study was conducted using the JH-CROWN registry, which includes data collected during routine clinical care at 5 hospitals comprising 2513 beds in the Johns Hopkins Health System (Johns Hopkins Hospital, Baltimore, Maryland; Bayview Medical Center, Baltimore, Maryland; Howard County General Hospital, Columbia, Maryland; Suburban Hospital, Bethesda, Maryland; and Sibley Hospital, Washington, DC). The institutional review board of each hospital approved this study as minimal risk and waived informed consent requirements. COVID-19 was defined as the detection of SARS-CoV-2 from any nucleic acid test combined with an International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) code indicating the presence of symptomatic disease. Patients who were evaluated in the emergency department or hospitalized between March 4, 2020, and November 12, 2021, were eligible for inclusion. Some patients were included in prior studies.

**Exposures and Outcomes**

Patients self-reported their race and ethnicity during hospital registration, which was then coded into prespecified categories (eTable 1 in the Supplement). Individuals who self-identified their race as Asian, Black or African American, or White and their ethnicity as Hispanic or non-Hispanic were included in this analysis. To most accurately reflect skin tone and because the health outcomes of Black Hispanic patients most closely align with non-Hispanic Black patients, individuals who reported Black race and Hispanic ethnicity were categorized as Black. Consequently, race was categorized as Asian, Black or African American.
(which comprised Hispanic and non-Hispanic ethnicity), non-Black Hispanic, or non-Hispanic White. The Spo2 readings and SaO2 values from arterial blood gases were collected from the electronic medical record. All arterial blood gases were analyzed using the ABL825, ABL827, or ABL90 blood gas analyzer (Radiometer America), which utilize cooximetry to determine oxygen saturation values. The primary outcome of interest was the difference between Spo2 and SaO2 for readings that occurred within 10 minutes of one another. Occult hypoxemia was defined as an $\text{SaO}\text{2}$ of less than 88% in the setting of an $\text{SpO}\text{2}$ of 92% to 96%.5 The secondary outcome of interest was the relative difference in time to recognition of eligibility for treatment with oxygen threshold–specific COVID-19 therapies between patients of racial and ethnic minority groups and White patients with COVID-19. Based on the joint guidelines of the US Centers for Disease Control and Prevention and Infectious Diseases Society of America, we set the threshold of interest as an oxygen saturation of 94% or less or use of supplemental oxygen, reflecting eligibility for treatment with dexamethasone or remdesivir.21,22

**Statistical Analysis**

For each SaO2 value, we first identified a concurrent Spo2 value, which was defined as an $\text{SpO}\text{2}$ measured within 10 minutes before or after the corresponding SaO2. If there were multiple values within the period, the closest one was selected. For the paired Spo2 and SaO2 values, the mean and standard error of the paired differences were calculated within each racial and ethnic category. The occurrence of occult hypoxemia (defined as SaO2 <88% with concurrent Spo2 measurement of 92%-96%) was presented as the proportion of individuals with at least 1 instance of occult hypoxemia during their hospital encounter and the proportion of measurements consistent with occult hypoxemia by racial and ethnic category.

The association between racial and ethnic category and the difference in paired Spo2 and SaO2 measurements (response variable) was examined using an unadjusted linear mixed-effects model that accounted for repeated measurements, with statistical significance defined as $P < .02$ based on Bonferroni correction. The linear mixed-effects model was then adjusted for covariates that captured disease severity or an underlying comorbidity or had a known or theoretical association with pulse oximetry accuracy, including demographic characteristics along with time-varying clinical and laboratory variables.1 All covariates and their interactions with race and ethnicity were included as fixed effects, and time-varying vital and laboratory measurements were included as random effects (complete model details are described in eMethods 1 in the Supplement). Only complete cases (ie, records with no missing data in covariates) were included in the adjusted linear mixed-effects model. R-squared (the proportion of variance explained by the model’s independent variables) was calculated to measure the goodness of fit of the model. The mean absolute prediction error from 10-fold cross validation was presented to show model prediction accuracy.

The fitted linear mixed-effects model was subsequently used to explore the association of differential pulse oximetry accuracy by race and ethnicity with recognition of eligibility for treatment with oxygen threshold–based COVID-19 pharmacologic therapy. Among all patients with COVID-19 who were admitted between March 4, 2020, and November 12, 2021, who were not already receiving supplemental oxygen or mechanical ventilation, we estimated the predicted SaO2 corresponding to each Spo2 measurement using the fitted linear mixed-effects model and nearest covariates. Individuals with missing covariates, those who did not have a predicted SaO2 of 94% or less, or those who initiated treatment with oxygen or had a measured Spo2 of 94% or less before a predicted SaO2 of 94% or less were excluded from this analysis.

Proportions were presented by racial and ethnic category for patients with delayed recognition or unrecognized treatment eligibility, defined as those patients with a predicted SaO2 of 94% or less before a measured Spo2 of 94% or less or oxygen treatment initiation (delayed recognition) or those with a predicted SaO2 of 94% or less who did not initiate treatment with oxygen or have a recorded Spo2 of 94% or less at any time (unrecognized). The difference in time to recognition of treatment eligibility between patients of racial and ethnic minority groups and White patients was estimated using a Cox proportional hazards model. Patients were censored at 96 hours if they were discharged before experiencing an event or experienced the event after 96 hours (censoring 2.3% of events).

Among individuals with delayed recognition of treatment eligibility, we calculated the length of delayed recognition as the time between predicted SaO2 of 94% or less and measured Spo2 of 94% or less or oxygen treatment initiation and compared the distributions between patients of racial and ethnic minority groups and White patients using Wilcoxon rank sum tests. A sensitivity analysis excluding observations in which titration of oxygen may have occurred between Spo2 and SaO2 measurement was also conducted (eMethods 2 in the Supplement). All statistical analyses were conducted using R, version 4.0.2 (R Foundation for Statistical Computing), and the linear mixed-effects model was fitted using the lme4 package.23

**Results**

Of the 7448 individuals evaluated in the emergency department or hospitalized for COVID-19 during the period of interest, 1216 patients (16.3%) had a total of 32 282 concurrently measured Spo2 and SaO2 levels during their hospital encounter, with a median (IQR) of 10 (2-30) measurements per patient. A patient flow diagram is presented in Figure 1, and characteristics stratified by race and ethnicity are presented in the Table. Compared with White patients, Black and non-Black Hispanic patients were younger, had longer hospital stays, and more SaO2 measurements per person. In addition, non-Black Hispanic patients had a smaller proportion of women and fewer comorbidities. The median SaO2 for Asian and Black patients was consistently lower than the Spo2 value (ie, Spo2 overestimated the true oxygen saturation) for all Spo2 readings of 88% to 96%, while the median SaO2 was consistently higher than the Spo2 value (SpO2 underestimated the true oxygen saturation) for White patients with an Spo2 of 88% to 96%.
Figure 1. Patient Flow Diagram Detailing Patients Included in the Pulse Oximetry Inaccuracy Analysis and Analysis of Unrecognized or Delayed Treatment Eligibility

- **7448** With hospital encounter at JHHS for COVID-19
- **322** Reporting race and ethnicity other than Asian, Black, Hispanic, or White
- **7126** Patients
  - **141579** SpO2 measurements
  - **59974** SaO2 measurements
- **1216** Patients with 32,282 SpO2 measurements matched to SaO2 within 10 min
- **6673** Patients with 906,140 SpO2 measurements and complete covariates available for SaO2 estimation
  - **352** Asian
  - **2642** Black
  - **1170** Hispanic
  - **2509** White
- **603** Excluded
  - **591** No predicted SaO2 ≤94%
  - **12** Multiple admissions
- **4167** Excluded
  - **2714** SpO2 ≤94% predicted SaO2 <94%
  - **1453** Initiated oxygen before predicted SaO2 ≤94%
- **1004** Patients with 27,367 matched SpO2 and SaO2 measurements with complete covariates
- **1452** With predicted SaO2 ≤94% before SpO2 ≤94% or oxygen initiation
  - **54** Asian
  - **399** Black
  - **188** Hispanic
  - **363** White
  - **21** Asian
  - **681** Black
  - **323** Hispanic
  - **427** White
- **451** With predicted SaO2 ≤94% never had SpO2 ≤94% or oxygen therapy
  - **4** Asian
  - **247** Black
  - **122** Hispanic
  - **78** White

JHHS indicates Johns Hopkins Health System; SaO2, oxygen saturation levels in arterial blood; SpO2, pulsatile flow of arterial blood.
Using the unadjusted linear mixed-effects model that was limited to SpO2 values from 88% to 96%, compared with White patients, SpO2 overestimated SaO2 by 2.1% for Asian patients ($P < .001$), 1.4% for Black patients ($P < .001$), and 0.8% for non-Black Hispanic patients ($P < .001$). Results were attenuated but qualitatively similar when considering the entire range of SpO2 measurements (eTable 2 in the Supplement).

Occult hypoxemia (SaO2 <88% with concurrent SpO2 measurement of 92%-96%) was identified in 3.7% of samples from Asian patients, 3.7% of samples from Black patients, 2.8% of samples from non-Black Hispanic patients, and 1.7% of samples from White patients. At the patient level, 19 Asian patients (30.2%), 136 Black patients (28.5%), 64 non-Black Hispanic patients (29.8%), and 79 White patients (17.2%) had occult hypoxemia at some point during the hospital encounter.

Results of the adjusted fully specified model are presented in eTable 3 and eFigure 1 in the Supplement. The interaction of race and ethnicity with SpO2 and oxygen device was statistically significant and included in the final parsimonious model (eTable 4 in the Supplement), which had an $R^2$ of 0.88 and mean (SD) absolute prediction error of 2.64 (0.11). In the adjusted parsimonious linear mixed-effects model, compared with White patients, SpO2 overestimated SaO2 by an average of 1.7% among Asian patients (95% CI, 0.5%-3.0%; $P = .01$), 1.2% among Black patients (95% CI, 0.6%-1.9%; $P < .001$), and 1.1% among non-Black Hispanic patients (95% CI, 0.3%-1.9%; $P = .01$; Figure 3). The absolute differences between SaO2 and SpO2 by race and ethnicity as stratified by oxygen device are presented in eFigure 2 in the Supplement.

Among 6673 patients with SpO2 measurements and available covariate data for estimation of predicted SaO2 using the
mixed-effects model, 1903 (28.5%) had a predicted \( \text{SaO}_2 \) of 94% or less before a measured \( \text{SpO}_2 \) of 94% or less or oxygen treatment initiation and were thus included in the analysis of unrecognized or delayed treatment eligibility (see Table 5 in the Supplement for baseline characteristics). Within each race and ethnicity, more Black and Hispanic patients unrecognized or delayed recognition of treatment eligibility compared with White patients (25 Asian [7.1%], 928 Black [35.1%], and 445 non-Black Hispanic patients [38.0%] vs 505 White patients [20.1%]). Compared with White patients, Black patients had a 29% lower hazard of eligibility recognition (hazard ratio [HR], 0.71; 95% CI, 0.63-0.80; \( P < .001 \)) and non-Black Hispanic patients had a 23% lower hazard of eligibility recognition (HR 0.77; 95% CI, 0.66-0.89; \( P < .001 \)), with no difference between Asian and White patients (HR, 0.97; 95% CI, 0.62-1.5; \( P = .90 \); Figure 4). Of these 1903 patients, 451 (23.7%) never had a recorded \( \text{SpO}_2 \) of 94% or less or received oxygen therapy and were considered to have unrecognized treatment eligibility. Most of these patients were Black (247 [54.8%]), followed by non-Black Hispanic (122 [27.1%]), White (78 [17.3%]), and Asian (4 [0.9%]). Among the remaining 1452 patients with eventual recognition of eligibility, the median (IQR) time until recognition was highest for Asian patients (7.7 [3.5-13.6] hours) and Black patients (7.0 [1.9-20.8] hours), followed by non-Black Hispanic patients (5.0 [1.2-15.8] hours) and White patients (5.3 [1.4-15.2] hours). Compared with White patients, Black patients had a significantly higher median difference in time to recognition of eligibility by 1 hour (95% CI, 0.23-1.9 hours; \( P = .01 \)). There was no statistically significant difference in treatment delay between Asian and Hispanic patients and White patients (eFigure 3 and eTable 6 in the Supplement). The sensitivity analysis refitting the linear mixed-effects model after excluding observations in which treatment with supplemental oxygen may have been titrated between \( \text{SpO}_2 \) and \( \text{SaO}_2 \) measurement produced similar results (eResults in the Supplement).

Discussion

In this retrospective analysis of a large clinical data set of patients with COVID-19 who were admitted to a regional health care system, we identified persistent overestimation of arterial oxygen saturation among Asian, Black, and Hispanic individuals. This was associated with systematic misclassification of patient oxygenation status based on racial or ethnic

Figure 2. Difference Between Arterial Blood Oxygen Saturation and Noninvasive Oxygen Saturation Levels

Figure 3. Relative Mean Differences With 95% CIs of \( \text{SaO}_2 \)-\( \text{SpO}_2 \) for Patients of Racial and Ethnic Minority Groups Based on the Adjusted Parsimonious Linear Mixed-Effects Model

<table>
<thead>
<tr>
<th>Race and ethnicity</th>
<th>Patients, No.</th>
<th>Observations, No.</th>
<th>Mean difference (95% CI)</th>
<th>( \text{SpO}_2 ) underestimates</th>
<th>( \text{SpO}_2 ) overestimates</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>54</td>
<td>1696</td>
<td>-1.73 (-2.98 to -0.48)</td>
<td>0.007</td>
<td>&lt;.001</td>
<td>0.006</td>
</tr>
<tr>
<td>Black</td>
<td>399</td>
<td>10 517</td>
<td>-1.23 (-1.87 to -0.59)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>388</td>
<td>6693</td>
<td>-1.13 (-1.93 to -0.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>363</td>
<td>8461</td>
<td>0 [Reference]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Median, IQR, and range of difference between arterial blood oxygen saturation levels (\( \text{SaO}_2 \)) and noninvasive oxygen saturation levels measured by pulse oximetry (\( \text{SpO}_2 \)) by \( \text{SpO}_2 \) level among 950 patients (15,984 measurements) with an \( \text{SpO}_2 \) between 88% and 96%.
self-identification. When comparing pulse oximetry measurements with arterial oxygen saturation, approximately one-third of patients from each racial or ethnic minority group had at least 1 unidentified episode of hypoxia compared with fewer than one-fifth of White patients. In addition, we found a systematic failure to identify Black and Hispanic patients who were likely qualified to receive COVID-19 therapy and a statistically significant delay in recognizing the guideline-recommended threshold for initiation of therapy among Black patients compared with White patients.

A recent investigation also reported overestimation of arterial oxygen saturation levels by pulse oximetry among Black individuals, with a 2.7-fold greater incidence of occult hypoxemia in Black patients compared with White patients, which was corroborated in a pre–COVID-19 pandemic sample of patients who were about to undergo extracorporeal membrane oxygenation. By limiting the patient sample to individuals who had a common pulmonary diagnosis (COVID-19), the present study minimizes the patient heterogeneity in the prior investigations while accounting for several factors that may be associated with pulse oximetry inaccuracy. This study also extends these findings by demonstrating similar overestimation of oxygen saturation levels, occult hypoxemia, and associated disparities in recognition of treatment eligibility among non-Black Hispanic patients.

The finding that occult hypoxemia affected more patients from all racial and ethnic minority groups compared with White patients is at odds with the finding by Wong et al6 that occult hypoxemia was similar between Asian and White patients. This discrepancy may result from Wong et al6 limiting their investigation to the first arterial blood gas measurement or heterogeneity of the Asian populations between the studies. Two recent studies from the United Kingdom of individuals with COVID-19 reported conflicting results and were limited by overall small sample sizes comprising predominantly White patients with few Black patients (<10%) and no Hispanic patients.25,26

By leveraging a large data set of paired SaO2-SpO2 measurements to predict the time at which therapies would become indicated, we found that most patients who had an unrecognized indication for COVID-19 therapy were Black, along with statistically significant differences in time to recognition of treatment eligibility for Black and Hispanic patients compared with White patients. While long-term clinical data may inform whether the delay in recognition of treatment eligibility had clinical consequences, these data are not yet available for this cohort. However, recognizing that delays in care may exist because of inaccurate pulse oximetry measurements informs studies that have investigated COVID-19 outcomes among patients of racial and ethnic minority groups.6,22,23 Because occult hypoxia is associated with higher in-hospital mortality, particularly among racial and ethnic minority groups,6,6 it is possible that occult hypoxia in COVID-19 contributes to racial and ethnic disparities in COVID-19 outcomes.31-33

While pulse oximetry has become a fundamental tool in diagnosis, triage, and management decisions in the acute care setting, the device’s lack of accuracy in certain populations has not been adequately investigated or addressed, although it has been recognized for several decades34 and was highlighted in a 2020 safety communication by the US Food and Drug Administration.35 With the integration of economical light-emitting diodes and semiconductors, pulse oximeters have also migrated out of the acute care setting and are now available and affordable to the average consumer for home use. The expanded use of a differentially inaccurate device potentially exacerbates racial and ethnic health disparities. For example, a Black individual who tests positive for COVID-19 in the outpatient setting may be advised against or decide to delay seeking care based on false reassurance from normal pulse oximetry readings. Such a scenario was reflected in the overrepresentation of Black patients who had unrecognized eligibility for treatment with remdesivir and dexamethasone. The differential inaccuracy of pulse oximetry in racial and ethnic minority groups may be similarly associated with treatment and triage decisions for other respiratory illnesses, such as pneumonia and acute respiratory distress syndrome. Although increased awareness of the limitations of pulse oximetry may mitigate some of the adverse effects, innovative approaches, such as the integration of additional wavelengths or calibration based on skin pigmentation, are needed. Unlike the use of race and ethnicity in the calculation of glomerular filtration rate or interpretation of pulmonary function testing, which have come under recent scrutiny, the race and ethnicity-based discrepancy of pulse oximetry exposes a fundamental flaw in the acquisition rather than interpretation of data, although all the aforementioned biases are associated with systematic underdiagnosis of disease or withholding of therapies for racial and ethnic minority groups.

Although the principal objective of the study was to characterize relative biases in pulse oximetry, the absolute biases identified among all racial and ethnic minority groups are notable. While some variability may be explained by the allowance of 10 minutes between paired SpO2 and SaO2 measure-
ments, such variability would not explain the systematic error among White patients that is separate from the pulse oximeter’s imprecision. This imprecision was more pronounced among racial and ethnic minority groups, as reflected by the standard errors in eTable 2 in the Supplement, which corroborates previous findings and may exacerbate the biases identified in this investigation.44 This study’s results suggest that for critical treatment decisions that rely on arterial oxygen saturation, use of pulse oximetry may be inadequate and produce opportunities for undertreatment and overtreatment irrespective of the patient’s race or ethnicity. This mirrors inaccuracies in capillary blood glucose levels reported among patients with critical illness.45

Strengths and Limitations
The strengths of this study include a large diverse patient population spanning multiple hospitals that allows statistical inferences to be made for several racial or ethnic minority groups; many paired measurements, which decreases susceptibility to bias; a curated data set with prior validation of important covariates; and the relative uniformity of underlying disease state. However, there are several limitations. Self-reported race and ethnicity is used in this study as a surrogate marker for skin tone, which is not routinely collected in the clinical setting. Thus, we were unable to measure or account for heterogeneity in skin tones within each racial or ethnic group. Rapid fluctuations in clinical condition among patients with COVID-19 may introduce downward bias in SpO2-SaO2 calculations if there is a preponderance of patients who quickly deteriorate after their SpO2 measurement is logged. However, such biases should not be different between racial and ethnic minority groups. Because the analysis estimating racial and ethnic difference in SpO2-SaO2 required a blood gas measurement, results may not be generalizable to healthy individuals or those with less acute illness. Furthermore, use of an oxygen saturation threshold to represent recognition of drug eligibility is an imperfect estimate of the actual delay in treatment that is attributable to pulse oximetry, as clinical attention, time to ordering, medication delivery, and drug availability are all contributory factors. However, this would tend to conservatively bias estimates. Finally, treatment delay estimates depend on the frequency with which SpO2 was measured or recorded, which may potentially differentially underestimate delay in recognition of treatment eligibility based on the patient’s level of care.

Conclusions
In this cohort study using a large multihospital COVID-19 clinical data set, we found statistically significant and persistent overestimation of arterial oxygen saturation by pulse oximetry among Asian, Black, and Hispanic patients compared with non-Hispanic White patients. Black and Hispanic patients were more likely to experience unrecognized and delayed recognition of eligibility to receive COVID-19 therapy. Differential inaccuracies in pulse oximetry should be examined as a potential explanation for disparities in COVID-19 outcomes and may have implications for the monitoring and treatment of other respiratory illnesses.


