PNAS RESEARCH ARTICLE | ECONOMIC SCIENCES

Clinician behavior when skin tone affects test results

Marcella Alsan[a,b], Liran Einav[a,b], Amy Finkelstein[b,c,1], and Jonathan Zhang[b,d,e]

Contributed by Amy N. Finkelstein; received October 25, 2025; accepted February 20, 2026; reviewed by Amitabh Chandra and Hannes Schwandt

We compare care for Black vs. White patients following pulse oximetry, a widely used device for measuring blood oxygen levels which overestimates oxygen saturation in darker-skinned patients. Black patients are therefore medically more appropriate for follow-up care than White patients with the same pulse oximetry reading. Yet, using data from the Veterans Health Administration on 3.5 million emergency department visits between 2014 and 2018, we find that Black patients systematically receive lower rates of follow-up care than White patients with the same reading. Our findings illustrate how bias in a medical screening device can be propagated and potentially amplified in downstream care.

pulse oximetry | racial bias | health care

Significance

There is growing attention and concern about racial biases in some of the decision tools that clinicians use to decide on medical interventions, but considerably less evidence on the downstream consequences of these biases for patient treatment. To shed light on this, we compare medical care for Black and White patients when doctors use a medical technology (pulse oximetry) that overstates oxygen saturation among darker-skinned patients. Despite Black patients therefore being more medically appropriate for follow-up care than White patients with the same pulse oximetry reading, we find that Black patients systematically receive lower rates of follow-up care than White patients with the same reading.

There is growing attention and concern about the potential for "algorithmic bias," in which an ostensibly objective decision tool ends up making systemic errors that adversely impact a particular group. Within medicine, a particular focus has been on racial biases in the algorithms clinicians use to decide on medical interventions. Racial biases in clinical algorithms have been documented across many areas of medicine, including cardiology, nephrology, obstetrics, neonatal care, pulse oximetry, lung-function testing, and complex care management (e.g., refs. 1–4). In all these instances, the algorithms are biased toward underdiagnosing potential health problems in Black patients. For example, the long-standing clinical algorithm used to gauge kidney function from an easily obtained measure of serum creatinine level (the estimated Glomerular Filtration Rate, or eGFR) includes an unwarranted "race correction" which causes the algorithm to overstate the kidney function of Black patients (2).

There is considerably less evidence, however, on the downstream consequences of these algorithmic biases for patient treatment. The concern is that, by underestimating the severity of illness in Black patients, these clinical decision tools may result in Black patients receiving less of the recommended follow-up care than White patients of the same underlying health status. Of course, if health-care providers are rational and informed about the algorithmic bias, and there are no other race-specific frictions—such as differential trust, communication barriers, or implicit biases—in the provider–patient interaction, providers might (partly or fully) debias the output from the clinical algorithm. In this scenario, there would be higher rates of follow up care for Black patients with the same algorithmic reading as White patients, and potentially no differential treatment by patient race across patients of the same underlying health status. Whether and to what extent Black patients receive higher rates of follow-up care from biased devices and clinical decision tools is thus an empirical question.

We examine this question empirically in the specific context of pulse oximetry. The pulse oximeter is a widely used clinical screening and monitoring device, typically slipped onto the finger, that painlessly and quickly produces a measure of blood oxygenation. Recommended follow-up care after a reading of low blood oxygenation from the pulse oximeter includes an arterial blood gas (ABG) test to determine blood oxygenation rates directly, as well as timely administration of supplemental oxygen.

As we discuss in more detail below, both theoretical physics and existing clinical evidence indicate that the pulse oximeter overstates (biases) the blood oxygen content for darker-skinned individuals relative to lighter skinned individuals. This arises from the combination of the spectroscopic phenomenon the pulse oximeter exploits and its having been calibrated and validated on light-skinned patients. Absent any other race-specific frictions, if clinicians were unaware of or did not do anything to debias the pulse oximetry readings, Black and White patients with the same pulse oximeter reading would receive the same rates of recommended follow-up care. If, on the other hand, providers partially or fully debias the pulse oximetry readings, we should observe, on average, *higher* rates of subsequent interventions for Black patients than White patients with the same reading.

Author affiliations: [a]Department of Economics, Stanford University, Stanford, CA 94305; [b]National Bureau of Economic Research, Cambridge, MA 02138; [c]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139; [d]Sanford School of Public Policy, Duke University, Durham, NC 27708; and [e]United States Department of Veterans Affairs Palo Alto Health Care System, Menlo Park, CA 94025

Author contributions: M.A., L.E., A.F., and J.Z. designed research; performed research; analyzed data; and wrote the paper.

Reviewers: A.C., Harvard University; and H.S., Northwestern University.

Competing interest statement: J.Z. has received compensation from the US Department of Veterans Affairs as a graduate student, federal contractor, as well as via an intergovernmental personnel act.

[1]To whom correspondence may be addressed. Email: afink@mit.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2530585123/-/DCSupplemental.

Published May 8, 2026.

Using data on 3.5 million Emergency Department (ED) visits to 39 facilities of the Veterans Health Administration (VA) from 2014 to 2018, we therefore compare rates of follow-up care for Black and White patients with the same pulse oximeter reading. The VA's detailed electronic medical records and its high share of Black patients (approximately one third in our sample) make it an ideal empirical setting for this exercise.

Contrary to the hypothesis that we would find similar or greater rates of follow-up care for Black patients than White patients with the same pulse oximeter reading, we find that Black patients systematically receive lower rates of clinical follow-up care, at all pulse oximetry values we study. The differences in rates of follow-up care are statistically and substantively significant. For example, in the five hours after an initial pulse-oximetry reading, Black patients with a pulse oximetry measurement of 89% or lower (a measure well below the threshold at which standard guidelines would suggest follow up is expected) are about 6 percentage points (20%) less likely to receive an ABG test and about 4 percentage points (7%) less likely to receive supplemental oxygen relative to White patients with the same pulse oximetry reading. These findings are not affected by the inclusion of controls for demographic and (detailed) clinical characteristics.

It is important to emphasize that these differences in follow-up care occur in a context in which, due to the algorithmic bias, the appropriate rates of follow-up care (conditional on the same pulse oximetry reading) "should" have been higher for Black patients. Indeed, once we apply a correction for the bias based on estimates from the literature, the racial differences in follow-up care more than double. For example, we find that among patients whose skin-tone-corrected initial pulse oximetry measurement would be 89% or lower, Black patients are now about 12 percentage points less likely to receive an ABG test and about 19 percentage points less likely to receive supplemental oxygen. Our results suggest that differences in rates of follow-up care cannot be explained solely by bias in the underlying screening device but must also reflect decisions by the clinician. We also note that while the device bias is unambiguously the result of skin tone, the difference in clinician decisions conditional on the device reading may be the result of skin tone or of factors that may be correlated with skin tone, such as socioeconomic status.

Our paper relates to a large literature on racial bias in high stakes settings like health care and criminal justice. We add to scholarship that documents racial bias by individual decision makers, such as physicians, caseworkers, judges, and police (e.g., refs. 5–8), by examining potential biases introduced by the medical devices and decision support tools these individuals use. In this sense, our paper relates to the work of Bohren et al. (9) who note that total discrimination is the sum of "direct" and "systemic" discrimination. Our primary contribution is to use a well-established instance of racial bias in medical screening results to document how it is propagated through the care pathways.

Our study also relates to the large literature on racial and ethnic disparities in health care. This has been an important area of policy and academic attention since the landmark publication of Unequal Treatment by the Institute of Medicine in 2003 (10) and updated in 2024 (11). The causes of these gaps are multifaceted, and include social, economic, and institutional factors, in addition to medical ones. Within the study of medical factors, there are large literatures documenting both differences in access to care by race and ethnicity as well as the fact that patients of different race or ethnicity systematically receive differential medical treatment from the same health care provider. In particular, there is considerable evidence that Black and Hispanic patients receive less intensive treatment than otherwise similar White patients.* The disparate treatment has typically been interpreted as arising through individual behavior, such as incorrect clinical beliefs, communication challenges, or taste-based discrimination. Our study highlights the role of the development and commercialization of racially biased medical devices in affecting disparate treatment.

The rest of the paper proceeds as follows. Section 1 provides an overview of pulse oximetry, including its clinical uses, the effects of skin tone on its readings, and the history of the development and commercialization of racially biased pulse oximetry readers. Section 2 describes the VA data, and Section 3 presents our findings. The final section concludes.

1. Setting

1.1. Pulse Oximetry. Pulse oximetry is a noninvasive optical technique used ubiquitously in clinical medicine to gauge the oxygen saturation of blood (13–15). The technology was adopted widely in hospitals during the 1980s, and quickly became so widespread that it is often referred to as the "fifth vital sign," regularly collected along with temperature, blood pressure, pulse, and respiration rate to provide a rapid assessment of a patient's lung function and peripheral tissue oxygenation (16, 17). Today, it is used routinely in a variety of settings from acute care to the home, and for both diagnosis and treatment management.

Anyone who has ever been to a physician—and certainly to a hospital or ED—has likely had a pulse oximeter slipped onto their finger early in their visit. Its routine use is due to the fact that pulse oximetry is painless and quick: Typically, a probe is placed on a finger or ear and a reading instantaneously appears on a digital monitor. However, pulse oximetry provides only a noisy signal of the underlying blood oxygen saturation. By contrast, the gold standard alternative, an arterial blood gas (ABG), provides a direct measurement of oxygen saturation. Yet, an ABG requires puncturing an artery to sample arterial blood—an invasive and painful procedure—and subsequent laboratory processing introduces delays. Pulse oximetry is therefore used as a routine screen to quickly and painlessly gauge whether or not the patient may be hypoxemic.

Given the vital importance of oxygenation for bodily functions and the narrow window in which a patient oxygen's saturation level can become of serious clinical concern with potentially life-threatening consequences, the clinical guidelines are to take additional investigative steps if the pulse oximeter measure falls below a given threshold—typically 92% oxygenation or lower, but sometimes 95% or lower (18, 19). Standard protocol, if the pulse oximetry measures falls below the threshold, is to engage in follow-up care; depending on the circumstances, that could involve a repeat pulse oximetry measure, a more expensive and invasive—but also accurate—ABG, and/or administration of supplemental oxygen (20).

The clinical use of pulse oximetry to diagnose and treat acute illness is thus similar to that of many other noninvasive medical screens used in a broad range of settings, from pregnancy, to newborns, to cancer diagnosis and treatment. These screens allow clinicians to obtain a relatively inexpensive, risk-free, painless, yet noisy signal which can then guide decisions on whether to undergo more invasive, expensive, risky, and potentially painful procedures that can provide definitive diagnoses (21). Our focus here, however, is not on the lack of precision in the measure, but

*Currie, MacLeod, and Musen provide a recent review (12).

on the potential bias: Pulse oximetry readings, like many other optical sensors, perform with varied accuracy across different skin tones. As a result, this can lead to differential clinical decision-making and worse health outcomes among people with darker skin tones.

1.2. Skin-Tone Effects. The interpretation of a pulse-oximetry reading depends on skin tone—which has implications for people with different skin tones, which may be more prevalent in certain racial and ethnic groups—as well as on underlying disease (e.g. jaundice), and other factors such as nail polish and motion artifact. The role of skin tone arises from its impact on the rate of light absorption, and therefore affects the interpretation of many noninvasive medical screening devices that use optical sensing, such as temporal artery (forehead) thermometers (22) and electronic transcutaneous bilirubin (TcB) devices used to screen for neonatal jaundice (23), as well as wearable technologies such as fitbit tracking of heart rates (24, 25).

In the case of pulse oximetry, the oximeter gauges blood oxygenation rates by leveraging the fact that oxygenated and deoxygenated blood have different light absorption properties. The pulse oximeter uses two light emitting diodes (LEDs) which emit light of two different wavelengths: red and infrared. A sensor measures the relative amount of red vs. infrared light absorption and translates this into an estimate of blood oxygenation (O_2 saturation values) using an experimentally derived calibration curve (26, 27).

Skin tone affects that calibration curve, which is based on a ratio of ratios known as the R-value which measures the ratio of direct to alternating current in red vs. infrared light. Melanin affects the absorption of red light through skin tissue more so than the absorption of infrared light. The net effect is to skew the R-value in patients with darker skin (27). Because calibration curves were primarily calibrated using paired pulse oximeter and ABG readings on lighter-skinned individuals (26, 28), pulse oximetry readings therefore deliver an overestimate of the oxygenated hemoglobin for patients with darker skin tones.

Recently, Starnes et al. conducted a prospective study to carefully measure the extent of this miscalibration (29). They simultaneously took pulse-oximetry measures and ABG measures for a sample of patients undergoing cardiac catheterization. They report oxygen saturation bias—i.e., systematic measurement error, defined as oxygen saturation measured via pulse oximetry minus the corresponding ("true") ABG measure—across three skin-tone categories, measured using spectrometer, and for two different pulse oximeter devices (Nellcor and Masimo); we will use these for a bias-correction factor in some of our analyses below.[†]

1.3. History of Product Development and Commercialization. How did the differential light absorption properties of darker and lighter skin get baked into biased measurements in a near-universal screening technology? We briefly review the history of product development for pulse oximetry, emphasizing that the effect of skin tone on pulse oximetry measures was well documented when it was first developed in the 1970s, and that an alternative device that worked well on all skin types was deemed not commercially viable and not developed. This history provides a concrete example of how lack of representation in the development of medical technologies (31) as well as the market-based pull-factors that affect the development and commercialization of medical technologies (e.g., refs. 32 and 33) can affect clinical care.

The fundamental observation that oxygenated blood is more transparent to red light than deoxygenated blood has been known since the 1930s. But this insight was not sufficient for building a practical, continuous, noninvasive arterial oxygen monitoring device, since red light absorbance is also affected by the rate of blood flow. This issue was ultimately resolved by including a second spectra of light (infrared) where the absorption was more similar across the two types of blood and thus sensitive only to the amount of blood (34). Yet, by the mid-1970s accuracy issues remained a concern, including the fact that patient movement and skin tone could affect measurements, and that the pulsatile nature of arterial blood introduced noise into measurement (34, 35).

In the 1970s, engineers from Hewlett-Packard (HP) seeking to address these issues learned that by including additional wavelengths they could create a device that performed accurately on all skin tones, and that they could deal with sensitivity to patient movement by creating an ear cuff-based measurement device. In 1976, they published an article describing a model they had designed that used eight different wavelengths and that, when tested on a diverse set of individuals, achieved high accuracy "regardless of skin pigmentation" (34). However, their expensive and somewhat cumbersome over-the-ear design was not further commercialized (36).

Meanwhile, Takuo Aoyagli, an engineer in Japan, overcame the issue of noise introduced by the pulsatile nature of arterial blood by creating the "Ratio of Ratios" concept described above (37, 38). In the 1980s, the US-based companies Biox Technology and Nellcor began to commercialize products based on this "Ratio of Ratios" concept, but reverted to using two (red and infrared) light spectra rather than the eight used in the HP design (39). Although subsequent refinements were made to reduce measurement error created by patient movement, little attention was paid to skin tone in model design improvements (35), and it is unclear how salient the skin-tone bias was for medical care providers during our 2014–2018 study period.

1.4. Existing Literature. In the decades subsequent to the widespread diffusion of the modern pulse oximeter in the 1980s, the skin-tone bias in pulse oximetry received some, but relatively limited attention (e.g., refs. 40 and 41). All of this changed in 2020 with the arrival of the Covid-19 pandemic. Starting in 2020, a flurry of academic research papers (e.g., refs. 3 and 42–44), as well as articles in the broader media (e.g., refs. 28 and 36) documented and discussed the skin-tone bias in pulse oximetry.[‡] This upsurge of attention likely stemmed from both the increased prominence of pulse oximetry during the pandemic (28), the murder of George Floyd, and the increased attention to racial differences in health and health care treatment in the wake of strong racial differences in the pandemic's mortality impacts (e.g., refs. 47–49). In February 2021, the FDA issued a safety communication about pulse oximetry's accuracy in pigmented skin, among other situations (50). Prospective studies on the difference in accuracy across racial lines were launched and committees were convened by the FDA to discuss performance-

[†]A limitation is that the sample is comprised of mostly children and underrepresents Hispanic individuals. However, it is otherwise broadly comparable to that of the US population. A retrospective study of paired pulse oximetry and ABG measures among ICU patients (30) suggests a similar bias correction.

[‡]Indeed, the leading internal medicine textbook, *Harrison's Principles of Internal Medicine*, updated its guidance on respiratory monitoring in its 22nd Edition to note: "Studies during the COVID pandemic noted that pulse oximetry overestimates oxygen saturation in patients with darker skin, thus making correlation with arterial PaO_2 more imperative." This statement was absent from earlier editions, including the 19th Edition (2015), despite otherwise nearly identical text in the relevant section (45, 46).

related benchmarks for oxygen measurement devices moving forward, culminating in a January 2025 draft guidance from the FDA for future blood oximetry device testing (51).

Many of the academic papers documenting bias, including those described above, follow a similar format: They examine pairs of measures of oxygen saturation by pulse oximetry and ABG among patients who received not only the routine pulse oximetry but also the follow-up ABG test, and compare how ABG levels (which provide a definitive measure of blood oxygenation) vary across patient race, conditional on the corresponding pulse oximetry reading. These papers consistently find evidence that, conditional on the pulse oximetry measure, Black patients or darker-skinned patients who receive a follow-up ABG test have lower rates of arterial oxygen saturation (as measured by the ABG test) than White patients or lighter-skinned patients. Martin et al. provide a systematic review of almost four dozen such studies documenting bias in paired measures of oxygen saturation from pulse oximetry and ABG from 1970 through 2023, more than half of which were published in 2020 or later (52). A challenge, however, with all of these studies is that they condition on having a follow-up ABG test, something that is not a common occurrence and could introduce selection, especially if the rate of follow-up testing, conditional on the pulse-oximetry reading, differs by race.

Compared to the voluminous number of papers documenting bias in paired measures of oxygen saturation from pulse oximetry and ABG, there is relatively less work on the consequences of this bias for follow-up care or downstream health outcomes. One study that did look at downstream outcomes documented that, during the Covid-19 pandemic, pulse oximetry bias was associated with delays in receiving treatment for Covid-19 among Black patients (42); this study attracted considerable attention (e.g. refs. 53 and 54). Other studies have documented lower rates of administration of supplemental oxygen and higher in-hospital mortality among Black ICU and surgical patients that is attributed to bias in pulse oximetry readings (e.g., refs. 55 and 56). However, like the large literature examining bias by comparing paired measures from pulse oximetry and ABG tests, all of these papers on downstream care and outcomes also restricted the analysis to the potentially selected sample of patients who received an ABG test following their pulse oximetry reading. We complement these existing studies by conditioning only on a pulse oximetry reading so that we can examine the differences in follow-up care among Black and White patients with the same pulse oximetry reading.[§]

2. Data

Our data come from electronic health records at the Veterans Health Administration (VA)—the largest integrated healthcare system in the United States, treating over 6 million veterans in 2018 (58). Its electronic medical records and high share of Black patients make it a particularly useful setting for our analyses. Prior work at the VA has documented that among VA inpatients who receive a follow-up ABG test, Black patients have lower rates of blood oxygenation (based on ABG tests) than White patients with the same initial pulse oximetry reading (43).

Our setting is patients in the Emergency Department (ED) where, as noted earlier, a pulse oximeter is routinely applied to all patients as part of collecting vital signs at intake. We begin with the universe of all ED visits in the VA over the five-year period from 2014 to 2018. This initial sample contains a total of 10.3 million ED visits to 105 VA facilities.

We make several sample restrictions to arrive at our primary analysis sample. First, given our focus on racial differences in measurement, we focus on Black non-Hispanic and White non-Hispanic patients only [because the correction factor we employ underrepresents Hispanic patients (29)]. These account for 8.7 million (84%) of the original sample. Second, although vital signs—including pulse oximetry—are almost always measured during ED visits at the VA, they are not always stored (within the electronic health records) in structured fields (59). Structured pulse oximetry measurements are available in only about 70% of all ED visits during this period; in the remaining cases these data are often stored as free text from clinical notes or device outputs, making it difficult to consistently extract. To ensure consistent measurement, we restrict our analysis to the 39 facilities where at least 90% of ED visits include a (structuredly stored) pulse oximetry reading, yielding 3,689,401 ED visits. In this sample, 96% of ED visits are associated with a pulse-oximetry reading, with similar rates of measurement of other vital signs (such as blood pressure). Finally, we further restrict attention to the 3,513,937 (95%) visits associated with a pulse oximetry of 80% or higher due to the relatively small number of patients with lower pulse oximetry rates and the possibility that these lower rates may reflect data errors. These 3.5 million visits form our baseline sample. 1,147,893 (33%) of these visits are associated with a (non-Hispanic) Black patient, while the remaining 2,366,044 visits involve a (non-Hispanic) White patient.[¶]

For each ED visit, we observe the patient's self-reported demographics, including age, sex, race, and ethnicity; we do not observe the skin tone of patients, and instead use Black race to proxy for darker skin tones. We also observe the time the patient arrived at the ED, the time stamped procedure codes for any subsequent tests or treatment they receive in the ED or after being admitted to the hospital, and the results of any of these subsequent tests. Most importantly for our context, we observe vital signs (including pulse oximetry readings), diagnostic laboratory results of any ABG tests, and administered procedures including supplemental oxygen; all of these are associated with time stamps, so that we know the sequencing of the different procedures and measurements. Finally, we construct indicators for preexisting Elixhauser comorbidities (60) based on medical diagnoses (in outpatient and inpatient settings) in the year prior to the ED visit; there are 30 such conditions including, for example, hypertension, diabetes, and depression.

2.1. Summary Statistics. Consistent with pulse oximetry being routinely used, there are no differences by race in the probability of receiving an initial pulse oximetry reading in our data (*SI Appendix*, Table S1). This is true both across all 105 VA EDs where—due to the idiosyncrasies of the electronic medical record storage system described above—we observe pulse oximetry readings for only 70% of patients [columns (1) and (2)] and in our primary analysis sample of the 39 EDs, where we observe pulse oximetry readings for 96% of patients [columns (3) and (4)].

Table 1 presents summary statistics for our primary analysis sample, separately for Black and White patients. Not surprisingly, there are noticeable differences in the characteristics of Black

[§]Matos et al. find lower rates of ABG testing for ICU patients who are Asian, Black, or Hispanic than for White patients, but do not compare ABG rates across patients with the same pulse oximeter reading (57).

[¶]For some of our analyses we also use the full sample of visits from all 105 facilities, after applying the same restrictions above regarding self-reported race and a pulse oximetry measure of 80% or higher. This expanded sample includes 6.1 million ED visits, of which 31% are associated with Black patients.

Table 1. Summary statistics

	Black (1)	White (2)	Difference (3)
Female	0.14	0.09	0.05
Age	57.6	61.0	−3.3
Age: 18 to 34	0.08	0.10	−0.02
Age: 35 to 49	0.17	0.13	0.04
Age: 50 to 64	0.45	0.29	0.16
Age: 65 to 79	0.25	0.37	−0.13
Age: 80+	0.06	0.11	−0.05
Number of elixhauser comorbidities	3.19	3.33	−0.14
Comorbidity: Hypertension	0.62	0.58	0.04
Comorbidity: Diabetes	0.33	0.31	0.02
Comorbidity: Depression	0.28	0.29	−0.01
Comorbidity: Chronic pulmonary disease	0.19	0.27	−0.08
Comorbidity: Cardiac arrhythmia	0.13	0.21	−0.08
N (number of visits)	1,147,893	2,366,044	
N (number of patients)	273,599	674,954	

Notes: This table presents summary statistics at the ED visit level. Comorbidities are diagnoses in the year prior to the ED visit. There are a maximum of 30 Elixhauser comorbidities; we report the average number, as well as rates for the top 5 most common ones in our sample. Column 1 and 2 shows the average for Black non-Hispanic and White non-Hispanic patients. Column 3 shows the Black–White difference (Black - White). All differences are statistically significant at the 0.001 level (with and without adjustment for observable differences by race in sex and age (or just sex or just age when looking at age and sex respectively).

and White veterans who visit the ED at the VA. Compared to the average ED visit by a White patient, the average ED visit by a Black patient involves a patient who is more likely to be female (14% compared to 9%), and is, on average, 3 y younger (average age of 57.6 among Black patients relative to 61 for Whites). Black patients also have slightly fewer diagnosed comorbidities (0.14 fewer; 3.19 compared to 3.33), and this remains true even conditional on sex and age (0.09 fewer). However, we show below that our main results are not affected by controlling for these demographic and health differences between Black and White patients, which is consistent with the clinical guidelines of treating the pulse oximetry reading as a "sufficient statistic" for follow-up oxygenation tests or treatments.

Fig. 1 presents the distribution of the initial pulse oximetry measurement, separately by race, in our baseline sample. Recall that clinical guidelines typically suggest that a pulse oximetry measurement of 92% or lower warrants additional tests and/or treatment. As one would expect, the vast majority of pulse oximetry readings are "normal,"—with readings of 95% or higher—but there is a left tail of lower rates that would potentially trigger follow-up care. Pulse oximetry measurements for Black patients are slightly higher on average (97.5%, compared to 96.7% for White patients), which may reflect underlying health differences, skin-tone effects on measurement, or both. In all of our subsequent analyses we will condition on the initial pulse oximetry measure.

3. Results

3.1. Racial Differences in Follow-Up Care. Fig. 2 presents our main findings: the propensity of subsequent follow-up care, by race, conditional on the initial pulse oximetry measurement. Panel (*A*) plots the fraction of patients to receive a follow-up ABG test within 300 min of the first pulse oximetry reading across race, and Panel (*C*) plots the fraction of patients to receive supplemental oxygen, again within 300 min of the first pulse oximetry reading. In both cases Black patients are less likely to receive follow-up care conditional on their initial pulse oximetry measurement.[#]

These differences are statistically significant—Panels (*B*) and (*D*) plots the racial differences and 95% CI constructed using heteroskedastic robust SEs—and quantitatively not trivial. For example, conditional of an initial pulse oximetry measurement of 89% or lower, Black patients are about 6 percentage points (20%) less likely to receive an ABG test; 31.9% of White patients receive an ABG test compared to only 26.4% of Black patients. Black patients are also about 4 percentage points (7%) less likely to receive supplemental oxygen (51.4% do, compared to 55.8% for White patients).[‖] Moreover, these results are unaffected by controlling for observable differences across Black and White patients (*SI Appendix*, Fig. S2), or measuring follow-up care through 150 min rather than 300 (*SI Appendix*, Fig. S3), and are similar to what we find if we use ED visits in all VA facilities (*SI Appendix*, Fig. S4).

While we have focused on the two main types of follow-up testing and care, another plausible response to a low pulse oximetry measure is to conduct an additional pulse oximetry reading. Indeed, the top two panels of *SI Appendix*, Fig. S5 show that additional pulse oximetry readings are common for all patients (about 30% of patients with an initial pulse oximetry reading of 95% or higher receive an additional pulse oximetry reading in the next 300 min) and more common for patients whose first reading is low. It also shows that for low levels of initial pulse oximetry, Black patients are more likely to receive a subsequent (biased) pulse oximetry reading. This suggests that subsequent pulse oximetry is being used as a substitute for administering an ABG to Black patients. Consistent with partial, but incomplete substitution, panels (*C*) and (*D*) of *SI Appendix*, Fig. S5 show that the probability of any follow up (ABG, supplemental oxygen, or additional pulse oximetry) is

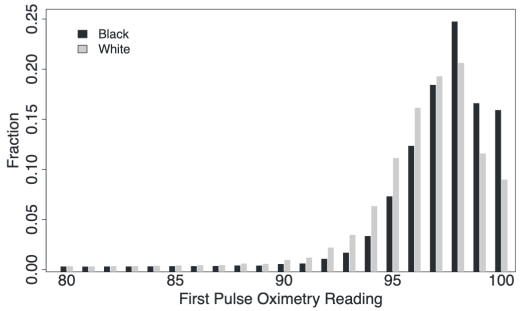

Fig. 1. Distribution of first pulse oximetry measurements, by race. *Notes:* This figure plots the distribution of the first pulse oximetry reading taken in the emergency department for Black non-Hispanic (black bars) and White non-Hispanic patients (gray bars) in our primary analysis, which consists of 1,147,893 ED visits and 2,366,044 ED visits with pulse oximetry readings for Black and White patients, respectively. The average pulse oximetry for Black (White) patients is 97.5 (96.7), SD for Black (White) patients is 2.0 (2.4), and the share of pulse oximetry between 80 to 89, 90 to 94, and 95 to 100 for Black (White) patients are 0.006 (0.013), 0.06 (0.13), 0.94 (0.86).

[#] And, naturally, also unconditionally, given that—as we saw in Fig. 1—Black patients are associated with higher initial measurement.

[‖] In addition, consistent with prior work both at the VA (43) and in many other settings (52), conditional on receiving an ABG test, the tests reveal that Black patients have much lower rates of oxygen saturation than White patients with the same initial pulse oximetry reading. *SI Appendix*, Fig. S1 shows these results, which are typical of the studies documenting lower underlying blood oxygenation for Black patients than for White patients with the same pulse oximetry reading, conditional on receiving an ABG test.

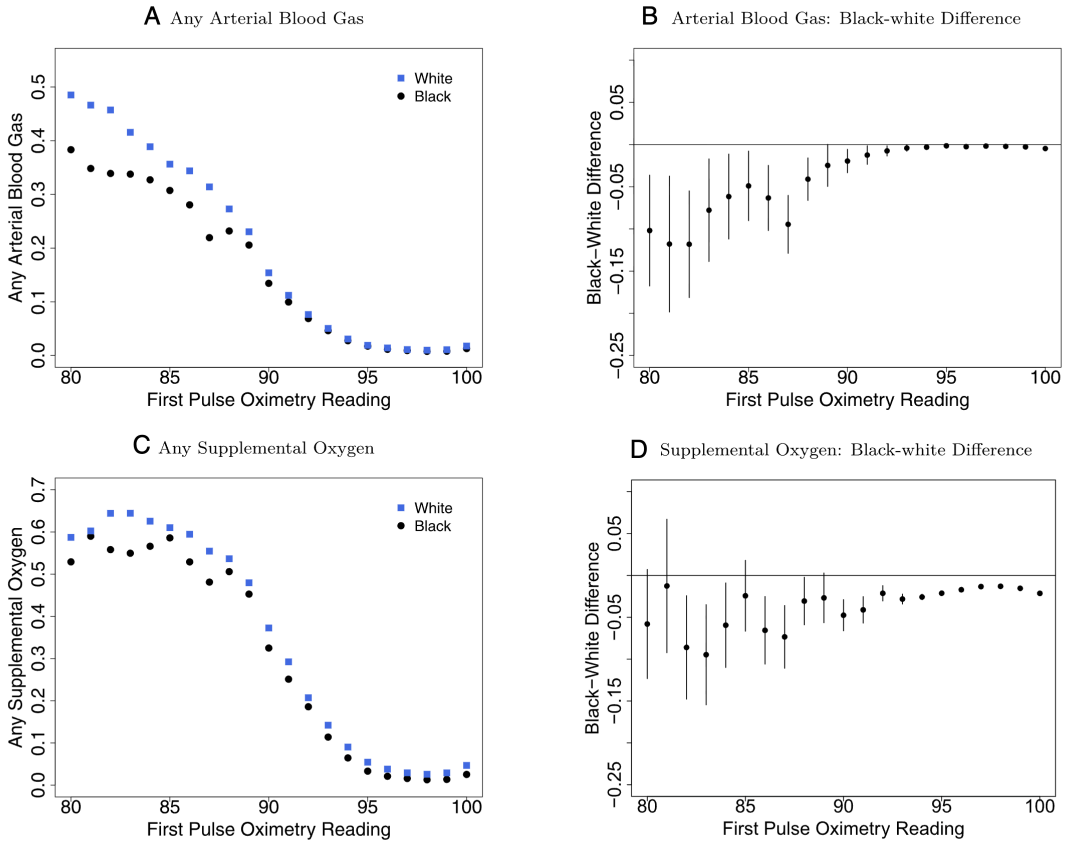

Fig. 2. Racial differences in follow-up care. Notes: This figure plots two clinical decisions by first pulse oximetry reading in the ED and by race: any ABG (*Top* row) and any supplemental oxygen (*Bottom* row), both within 300 min of the first pulse oximetry measurement. Panel (*A*) shows the fraction of patients who receive an arterial blood gas within 300 min by first pulse oximetry reading in the ED. Blue squares and black circles correspond to White and Black patients, respectively. Panel (*B*) plots the average Black–White difference in receiving an arterial blood gas within 300 min and the 95% CI (calculated using heteroskedastic robust SEs). Panels (*C*) and (*D*) is the same but with any supplemental oxygen as the outcome. The sample consists of 1,147,893 ED visits and 2,366,044 ED visits with pulse oximetry readings for Black and White patients, respectively.

more similar for Black and White patients at low levels of initial pulse oximetry than the probability of ABG or of oxygen, although Black patients still receive less follow up care of any kind.

3.2. Measurement-Corrected Results. Because of the skin-tone differences in the expected blood oxygen rates for Black and White patients with the same pulse oximetry measurement, comparing rates of follow-up care conditional on the pulse oximetry measurement will understate differences in rates of follow-up care conditional on "true" underlying blood oxygen rates.

We therefore reproduce the analysis of Fig. 2 after attempting to "correct" for the pulse oximetry skin-tone effect so that the corrected pulse-oximetry measurements are more comparable across patients of different race. Specifically, we draw on the estimates from Starnes et al. (29) which, as discussed in Section 1.2, compare pulse oximetry measures of blood oxygenation to ABG measures taken at the same time. Starnes et al. report the average bias (i.e., difference between these two measures) for three different skin tone categories; they do this for two different, common pulse oximetry devices (the Nellcor and the Masimo device). We calculate the average bias across these two devices for the darkest skin tone category (ITA 5 or 6) and the lightest category (ITA 1 or 2), and assign these to Black and White patients, respectively. We then adjust each observed pulse oximetry value by subtracting the relevant bias according to patient race. This serves as our corrected pulse oximetry measurement.

The resulting analysis is shown in Fig. 3. Conditioning on the "true" blood oxygen level (that is, the skin-tone-corrected initial pulse oximetry measurement) rather than the "raw" pulse oximetry measure greatly amplifies the lower propensity of Black patients to receive follow-up care. For example, conditional on an initial corrected estimate of blood oxygen measurement of 89% or lower, Black patients' propensity to receive ABG is now about 12 percentage points lower than White patients (13.9% compared to 26.3%) and their propensity to receive supplemental oxygen

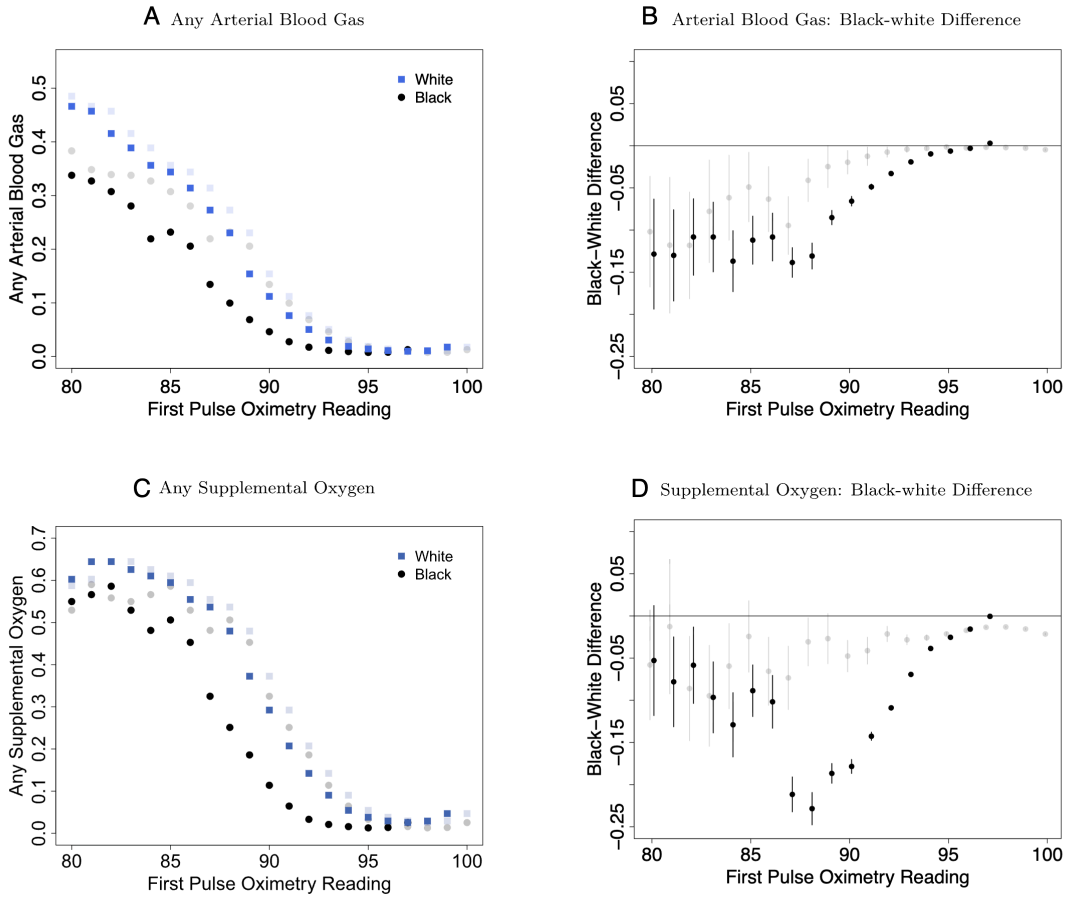

Fig. 3. This figure plots two clinical decisions by "corrected" first pulse oximetry reading in the ED and by race: any ABG (Top row) and any supplemental oxygen (Bottom row), both within 300 minutes of the first pulse oximetry measurement. Panel (*A*) shows the fraction of patients who receive an arterial blood gas within 300 minutes by "corrected" first pulse oximetry reading in the ED. Panel (*B*) plots the average Black-white difference in receiving an arterial blood gas within 300 minutes and the 95% CI (calculated using heteroskedastic robust SEs). Panels (*C*) and (*D*) are the same but with any supplemental oxygen as the outcome. The "corrected" pulse oximetry readings use the Starnes et al. correction (29). Specifically, using their Table 1 results, rounded to the nearest integer, we "correct" pulse oximetry by subtracting 1 from white patients and 3 from Black patients. Across all panels, the lighter transparent points are noncorrected and identical to Fig. 2. The darker points are corrected. Blue squares and black circles correspond to white and Black patients, respectively. The 95% CIs are calculated using heteroskedastic robust SEs. The sample consists of 1,147,893 ED visits and 2,366,044 ED visits with pulse oximetry readings for Black and white patients, respectively.

is now about 19 percentage points lower (31.0% compared to 49.5%).

4. Conclusion

In this paper, we investigated how racial bias embedded in technology, specifically in pulse oximetry, contributes to disparities in follow-up care between Black and White patients. A widely used, noninvasive tool to measure blood oxygenation, pulse oximeters overestimate blood oxygenation for darker-skinned individuals relative to lighter-skinned ones due to the combination of differences in physical light absorption properties by skin tone and device calibration based primarily on lighter-skinned individuals.

Using data from 3.5 million ED visits in the VA, we find that Black patients with the same pulse oximetry readings as White patients receive significantly less follow-up care, including an intervention listed by the World Health Organization as being essential: supplemental oxygen (61). These gaps persist even after controlling for observable characteristics. Our findings highlight how systemic bias in product development and commercialization can have downstream consequences for clinical care, and underscore the importance of representation in all steps of the scientific process. Further research would be useful to study how these downstream consequences for clinical care affect morbidity and mortality.

This study has several limitations. First, skin tone is not directly observed; instead, race is used as a proxy, which may also capture other patient characteristics—such as underlying health

conditions or socioeconomic status—that could independently influence care. Second, we do not observe patients' true underlying health status, limiting our ability to fully adjust for clinical severity. Third, our measurement correction is derived from a sample composed primarily of children and relies on a discretized skin tone scale. Fourth, it is unclear to what extent clinicians were aware of skin-tone–related bias in pulse oximetry during the study period. Finally, observed differences in care reflect a combination of device measurement bias and clinician responses to those measurements, which we are unable to separately identify.

Data, Materials, and Software Availability. The data used in this study were obtained from the U.S. Department of Veterans Affairs (VA) Corporate Data Warehouse (62) and contain sensitive patient-level information. These data are not publicly available due to legal and ethical restrictions under U.S. federal law and VA policy. Access to VA data may be granted to qualified researchers with an approved VA appointment and an approved research protocol or operations project. The analytic code used in this study is available from the corresponding author upon request.

ACKNOWLEDGMENTS. We are grateful to Vanessa Jia, Bhayva Pandey, Maya Rosen, and Sam Wolf for research assistance. We thank two referees, David Au, Justin List, Amy Moran-Thomas, Ernest Moy, and Thomas Valley for invaluable discussions and comments. This project was conducted as a quality assurance/clinical operations activity within the Veterans Health Administration and was determined by Office of Health Equity not to constitute human subjects research. Accordingly, the project did not undergo institutional review board review and approval. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

1. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
2. D. A. Vyas, L. G. Eisenstein, D. S. Jones, Hidden in plain sight–reconsidering the use of race correction in clinical algorithms. *New Engl. J. Med.* **383**, 874–882 (2020).
3. M. W. Sjoding, R. P. Dickson, T. J. Iwashyna, S. E. Gay, T. S. Valley, Racial bias in pulse oximetry measurement. *New Engl. J. Med.* **383**, 2477–2478 (2020).
4. J. A. Diao *et al.*, Implications of race adjustment in lung-function equations. *New Engl. J. Med.* **390**, 2083 (2024).
5. C. Tuttle, Racial disparities in federal sentencing: Evidence from drug mandatory minimums. SSRN 3080463 (2019). https://dx.doi.org/10.2139/ssrn.3080463. Accessed 27 May 2025.
6. D. Arnold, W. Dobbie, P. Hull, Measuring racial discrimination in bail decisions. *Am. Econ. Rev.* **112**, 2992–3038 (2022).
7. E. J. Baron, J. J. J. Doyle, N. Emanuel, P. Hull, J. P. Ryan, "Unwarranted disparity in high-stakes decisions: Race measurement and policy responses" NBER Working Paper No. 33104 (2024).
8. A. Corredor-Waldron, J. Currie, M. Schnell, "Drivers of racial differences in c-sections, (National Bureau of Economic Research)" NBER Working Paper No. 32891 (2024).
9. A. J. Bohren, P. Hull, A. Imas, "Systemic discrimination: Theory and measurement" NBER Working Paper No. 29820 (2022).
10. B. D. Smedley *et al.*, "Racial and ethnic disparities in health care: An ethical analysis of when and how they matter" in *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, B. D. Smedley, A. Y. Stith, A. R. Nelson, Eds. (National Academies Press US, Washington, DC, 2003).
11. E National Academies of Sciences, Medicine, *Ending Unequal Treatment: Strategies to Achieve Equitable Health Care and Optimal Health for All*, G. C. Benjamin, J. E. DeVoe, F. K. Amankwah, S. J. Nass, Eds. (The National Academies Press, Washington, DC, 2024).
12. J. Currie, W. B. MacLeod, K. Musen, "First do no harm? Doctor decision making and patient outcomes" NBER Working Paper No. 32788 (2024).
13. Cleveland Clinic, Arterial blood gas (ABG): What it is, purpose, procedure & levels (2022). https://my.clevelandclinic.org/health/diagnostics/22409-arterial-blood-gas-abg. Accessed 28 May 2025.
14. Yale Medicine, Pulse oximetry (2022). https://www.yalemedicine.org/conditions/pulse-oximetry. Accessed 27 May 2025.
15. University of Iowa Department of Otolaryngology, Pulse oximetry: Basic principles and interpretation (2024). https://iowaprotocols.medicine.uiowa.edu/protocols/pulse-oximetry-basic-principles-and-interpretation. Accessed 27 May 2025.
16. K. D. Torp, P. Modi, E. J. Pollard, L. V. Simon, Pulse oximetry. StatPearls (2023). https://www.statpearls.com/point-of-care/28070. Accessed 27 May 2025.
17. C. Witt, "Vital signs are vital: The history of pulse oximetry." ACP Hospitalist (2014). https://www.acpjournals.org/doi/10.7326/acph-201405-newman. Accessed 5 June 2025.
18. , American College of Chest Physicians, ATS/ACCP Statement on cardiopulmonary exercise testing. *Am. J. Respir. Crit. Care Med.* **167**, 211–277 (2003).
19. Cleveland Clinic, Blood oxygen level: What it is, purpose & normal range (2022). https://my.clevelandclinic.org/health/diagnostics/22447-blood-oxygen-level. Accessed 5 June 2025.
20. R. Beasley *et al.*, Thoracic Society of Australia and New Zealand oxygen guidelines for acute oxygen use in adults: 'Swimming between the flags'. *Respirology* **20**, 1182–1191 (2015).
21. P. Conner, L. Einav, A. Finkelstein, P. Persson, Targeting precision medicine: Evidence from prenatal screening. *J. Polit. Econ.* **133**, 604–651 (2025).
22. S. V. Bhavani, Z. Wiley, P. A. Verhoef, C. M. Coopersmith, I. Ofotokun, Racial differences in detection of fever using temporal vs. oral temperature measurements in hospitalized patients. *Jama* **328**, 885–886 (2022).
23. C. V. Hulzebos *et al.*, Screening methods for neonatal hyperbilirubinemia: Benefits, limitations, requirements, and novel developments. *Pediatr. Res.* **90**, 272–276 (2021).
24. R. Hailu, Fitbits and other wearables may not accurately track heart rates in people of color. STAT News (2019). https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/. Accessed 10 June 2025.
25. D. Koerber *et al.*, Accuracy of heart rate measurement with wrist-worn wearable devices in various skin tones: A systematic review. *J. Racial Ethn. Heal. Disparities* **10**, 2676–2684 (2023).
26. R. Al-Halawani, P. H. Charlton, M. Qassem, P. A. Kyriacou, A review of the effect of skin pigmentation on pulse oximeter accuracy. *Physiol. Meas.* **44**, 05TR01 (2023).
27. A. Bierman, K. Benner, M. S. Rea, Melanin bias in pulse oximetry explained by light source spectral bandwidth. *Br. J. Anaesth.* **132**, 957–963 (2024).
28. A. Moran-Thomas, How a popular medical device encodes racial bias. Boston Review (2020). https://www.bostonreview.net/articles/amy-moran-thomas-pulse-oximeter/. Accessed 5 June 2025.
29. J. R. Starnes *et al.*, Pulse oximetry and skin tone in children. *New Engl. J. Med.* **392**, 1033–1034 (2025).
30. N. Patwari, D. Huang, F. Bonetta-Misteli, "Short: Racial disparities in pulse oximetry cannot be fixed with race-based correction" in *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, S. Wang, Ed. (Association for Computing Machinery, New York, NY, 2023).
31. M. Alsan, M. Durvasula, H. Gupta, J. Schwartzstein, H. Williams, Representation and extrapolation: Evidence from clinical trials. *Q. J. Econ.* **139**, 575–635 (2024).
32. D. Acemoglu, J. Linn, Market size in innovation: Theory and evidence from the pharmaceutical industry. *Q. J. Econ.* **119**, 1049–1090 (2004).
33. A. Finkelstein, Static and dynamic effects of health policy: Evidence from the vaccine industry. *Q. J. Econ.* **119**, 527–564 (2004).
34. E. B. Merrick, T. J. Hayes, Continuous, non-invasive measurements of arterial blood oxygen levels. *Hewlett-Packard J.* **28**, 2–9 (1976).
35. S. Bajaj, Ra,cial bias is built into the design of pulse oximeters. The Washington Post (2022). https://www.washingtonpost.com/made-by-history/2022/07/27/racial-bias-is-built-into-design-pulse-oximeters/. Accessed 10 June 2025.
36. A. Moran-Thomas, *Oximeters Used to be Designed for Equity. What Happened?* (Wired, 2021).
37. K. Miyasaka *et al.*, Tribute to Dr. Takuo Aoyagi, inventor of pulse oximetry. *J. Anesth.* **35**, 671–709 (2021).
38. Cables & Sensors, History of pulse oximetry (2025). https://www.cablesandsensors.com/pages/history-of-pulse-oximetry. Accessed 10 June 2025.
39. W. T. Cecil, K. J. Thorpe, E. E. Fibuch, G. F. Tuohy, A clinical evaluation of the accuracy of the nellcor n-100 and ohmeda 3700 pulse oximeters. *J. Clin. Monit.* **4**, 31–36 (1988).
40. P. E. Bickler, J. R. Feiner, J. W. Severinghaus, Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* **102**, 715–719 (2005).
41. J. R. Feiner, J. W. Severinghaus, P. E. Bickler, Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: The effects of oximeter probe type and gender. *Anesth. Analg.* **105**, S18–S23 (2007).
42. A. Fawzy *et al.*, Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with covid-19. *JAMA Intern. Med.* **182**, 730–738 (2022).
43. V. S. Valbuena *et al.*, 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**, e069775 (2022).
44. A. Winny, N. Jurmo, The problem with pulse oximeters: A long history of racial bias. Johns Hopkins Bloomberg School of Public Health (2024). https://publichealth.jhu.edu/2024/pulse-oximeters-racial-bias. Accessed 27 May 2025.
45. J. P. Kress, J. B. Hall, "Approach to the patient with critical illness" in *Harrison's Principles of Internal Medicine*, D. L. Kasper *et al.*, Eds. (McGraw Hill, New York, ed. 19, 2015).
46. R. M. Baron *et al.*, "Approach to the patient with critical illness" in *Harrison's Principles of Internal Medicine*, D. L. Kasper *et al.*, Eds. (McGraw Hill, New York, ed. 22, 2025).
47. M. Alsan, A. Chandra, K. Simon, The great unequalizer: Initial health effects of covid-19 in the United States. *J. Econ. Perspect.* **35**, 25–46 (2021).
48. A. Finkelstein, G. Kocks, M. Polyakova, V. Udalova, Heterogeneity in damages from a pandemic. *Rev. Econ. Stat.* **8**, 1–45 (2024).
49. T. B. Foster, S. R. Porter, N. Pharris-Ciurej, *"Excess Mortality" During Covid-19 Varied by Race, Ethnicity, Geography* (U.S. Census Bureau, 2024).
50. U.S. Food and Drug Administration, Pulse oximeter accuracy and limitations: FDA safety communication (2021). https://content.govdelivery.com/accounts/USFDA/bulletins/2c276cb. Accessed 27 May 2025.
51. US Food and Drug Administration, Pulse oximeters for medical purposes-non-clinical and clinical performance testing, labeling, and premarket submission recommendations. Draft Guidance for Industry and Food and Drug Administration Staff (2025). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-medical-purposes-non-clinical-and-clinical-performance-testing-labeling-and. Accessed 27 may 2025.
52. D. Martin *et al.*, Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: A systematic review. *Br. J. Anaesth.* **132**, 945–956 (2024).
53. U. L. McFarling, Faulty oxygen readings delayed covid treatments for darker-skinned patients, study finds. Stat News (2022). https://www.statnews.com/2022/05/31/faulty-oxygen-readings-delayed-covid-treatments-darker-skin-patients/. Accessed 5 June 2025.

54. R. Perez, 'Racially biased' devices caused delayed treatment for black covid-19 patients. Science (2022). https://www.science.org/content/article/racially-biased-devices-caused-delayed-treatment-black-covid-19-patients. Accessed 5 June 2025.
55. N. R. Henry *et al*., Disparities in hypoxemia detection by pulse oximetry across self-identified racial groups and associations with clinical outcomes. *Crit. Care Med.* **50**, 204–211 (2022).
56. E. R. Gottlieb, J. Ziegler, K. Morley, B. Rush, L. A. Celi, Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit. *JAMA Internal Med.* **182**, 849–858 (2022).
57. J. Matos *et al*., Differences in arterial blood gas testing by race and sex across 161 US hospitals in four electronic health record databases. *Am. J. Respir. Crit. Care Med.* **211**, 1049–1058 (2025).
58. D. M. Bernard, A. Woodroffe, S. Lui, Understanding veterans' healthcare use and experience 2018–2019. Agency for Healthcare Research and Quality (2022). https://meps.ahrq.gov/data_files/publications/rf50/rf50.pdf. Accessed 27 July 2025.
59. O. V. Patterson *et al*., Extraction of vital signs from clinical notes. *Stud Heal. Technol Inf.* **216**, 1035 (2015).
60. A. Elixhauser, C. Steiner, D. R. Harris, R. M. Coffey, Comorbidity measures for use with administrative data. *Med. Care* **36**, 8–27 (1998).
61. World Health Organization, Oxygen (2024). https://www.who.int/health-topics/oxygen. Accessed 1 February 2026.
62. M. Alsan, L. Einav, A. Finkelstein, J. Zhang, Outpatient, Vital Signs, Lab Results Datasets. VA Corporate Data Warehouse (CDW). Accessed 5 June 2025.