

NBER WORKING PAPER SERIES

CLINICIAN BEHAVIOR WHEN SKIN-TONE AFFECTS TEST RESULTS

Marcella Alsan
Liran Einav
Amy Finkelstein
Jonathan Zhang

Working Paper 34168
<http://www.nber.org/papers/w34168>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025, Revised February 2026

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. The contents do not represent the views of the U.S. Department of Veterans Affairs, the United States Government, or the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w34168>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Marcella Alsan, Liran Einav, Amy Finkelstein, and Jonathan Zhang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Clinician Behavior When Skin-Tone Affects Test Results
Marcella Alsan, Liran Einav, Amy Finkelstein, and Jonathan Zhang
NBER Working Paper No. 34168
August 2025, Revised February 2026
JEL No. I11, I14, J14, J15

ABSTRACT

We compare care for Black versus white patients following pulse oximetry, a widely-used device for measuring blood oxygen levels which over-estimates 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.

Marcella Alsan
Stanford University
Department of Economics
and NBER
malsan@stanford.edu

Amy Finkelstein
Massachusetts Institute of Technology
Department of Economics
and NBER
afink@mit.edu

Liran Einav
Stanford University
Department of Economics
and NBER
leinav@stanford.edu

Jonathan Zhang
Duke University
Sanford School of Public Policy
and U.S. Department of Veterans Affairs
Office of Mental Health
and also NBER
jonathan.zhang@duke.edu

1 Introduction

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., Obermeyer et al. 2019; Vyas, Eisenstein, and Jones 2020; Sjoding et al. 2020; Diao et al. 2024). In all these instances, the algorithms are biased toward under-diagnosing 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 over-state the kidney function of Black patients (Vyas, Eisenstein, and Jones 2020).

There is considerably less evidence, however, on the downstream consequences of these algorithmic biases for patient treatment. The concern is that, by under-estimating 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) de-bias 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 de-bias 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 de-bias the pulse oximetry readings, we should observe, on average, *higher* rates of subsequent interventions for Black patients than white patients with the same reading.

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., Tuttle 2019; Arnold, Dobbie, and Hull 2022; Baron et al. 2024; Corredor-Waldron, Currie, and Schnell 2024; Philip and Ozkaya 2025), 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, Hull, and Imas (2022) 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 (Smedley, Stith, Nelson, et al. 2003) and updated in 2024 (National Academies of Sciences and Medicine 2024). The causes of these gaps are multi-faceted, 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.¹ 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 2 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 3 describes the VA data, and Section 4 presents our findings. The final section concludes.

¹Currie, MacLeod, and Musen (2024) provide a recent review, and Philip and Ozkaya (2025) provides a recent example.

2 Setting

2.1 Pulse oximetry

Pulse oximetry is a non-invasive optical technique used ubiquitously in clinical medicine to gauge the oxygen saturation of blood (Cleveland Clinic [2022a](#); Yale Medicine [2022](#); University of Iowa Department of Otolaryngology [2024](#)). 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 (Torp et al. [2023](#); Witt [2014](#)). 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 (American Thoracic Society and American College of Chest Physicians [2003](#); Cleveland Clinic [2022b](#)). 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 (Thoracic Society of Australia and New Zealand [2015](#)).

The clinical use of pulse oximetry to diagnose and treat acute illness is thus similar to that of many other non-invasive 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 (Conner et al. 2025). Our focus here, however, is not on the lack of precision in the measure, but 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.

2.2 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 non-invasive medical screening devices that use optical sensing, such as temporal artery (forehead) thermometers (Bhavani et al. 2022) and electronic transcutaneous bilirubin (TcB) devices used to screen for neonatal jaundice (Hulzebos et al. 2021), as well as wearable technologies such as fitbit tracking of heart rates (Chen 2019; Hailu 2019; Koerber, Khan, Shamsheer, et al. 2023)

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 (Al-Halawani et al. 2023; Bierman, Benner, and Rea 2024).

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 (Bierman, Benner, and Rea 2024). Because calibration curves were primarily calibrated using paired pulse-oximeter and ABG readings on lighter-skinned individuals (Moran-Thomas 2020; Al-Halawani et al. 2023), pulse oximetry readings therefore deliver an over-estimate of the oxygenated hemoglobin for patients with darker skin tones.

Recently, Starnes et al. (2025) conducted a prospective study to carefully measure the extent of this mis-calibration. 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.²

2.3 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 (Alsan et al. 2024) as well as the market-based pull-factors that affect the development and commercialization of medical technologies (e.g., Acemoglu and Linn 2004; Finkelstein 2004) can affect clinical care.

The fundamental observation that oxygenated blood is more transparent to red light than de-oxygenated blood has been known since the 1930s. But this insight was not sufficient for building a practical, continuous, non-invasive 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 (Merrick and Hayes 1976). 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 (Merrick and Hayes 1976; Bajaj 2022).

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” (Merrick and Hayes 1976). However, their expensive and somewhat cumbersome over-the-ear design was not further commercialized (Moran-Thomas 2021).

Meanwhile, Takuo Aoyagli, an engineer in Japan, overcame the issue of noise introduced

²A limitation is that the sample is comprised of mostly children and under-represents 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 (Patwari, Huang, and Bonetta-Misteli 2023) suggests a similar bias correction.

by the pulsatile nature of arterial blood by creating the “Ratio of Ratios” concept described above (Miyasaka et al. 2021; Cables & Sensors 2025). 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 (Cecil et al. 1988). Although subsequent refinements were made to reduce measurement error created by patient movement, little attention was paid to skin tone in model design improvements (Bajaj 2022), and it is unclear how salient the skin-tone bias was for medical care providers during our 2014-2018 study period.

2.4 Existing literature

In the decades subsequent to the wide-spread diffusion of the modern pulse oximeter in the 1980s, the skin-tone bias in pulse oximetry received some, but relatively limited attention (e.g. Bickler, Feiner, and Severinghaus 2005; Feiner, Severinghaus, and Bickler 2007). 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., Sjoding et al. 2020; Fawzy et al. 2022; Valbuena et al. 2022; Winny and Jurmo 2024), as well as articles in the broader media (e.g., Moran-Thomas 2020, 2021) documented and discussed the skin-tone bias in pulse oximetry.³ This upsurge of attention likely stemmed from both the increased prominence of pulse oximetry during the pandemic (Moran-Thomas 2020), 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., Alsan, Chandra, and Simon 2021; Finkelstein et al. 2024; Foster, Porter, and Pharris-Ciurej 2024). In February 2021, the FDA issued a safety communication about pulse oximetry’s accuracy in pigmented skin, among other situations (U.S. Food and Drug Administration 2021). Prospective studies on the difference in accuracy across racial lines were launched and committees were convened by the FDA to discuss performance-related benchmarks for oxygen measurement devices moving forward, culminating in a January 2025 draft guidance from the FDA for future blood oximetry device testing (U.S. Food and Drug Administration 2025).

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

³Indeed, 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 (Kress and Hall 2015; Baron and Massaro 2025).

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. (2024) 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. 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 (Fawzy et al. 2022); this study attracted considerable attention (e.g. McFarling and Gilyard 2022; Science Staff 2022). 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., Henry et al. 2022; Gottlieb et al. 2022). 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.⁴

3 Data

Our data come from electronic health records at the Veterans Health Administration (VA)—the largest integrated healthcare system in the US, treating over 6 million veterans in 2018 (Bernard, Woodroffe, and Lui 2022). 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

⁴Matos et al. 2025 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.

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 (Valbuena et al. 2022).

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 under-represents Hispanic patients (Starnes et al. 2025)). 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 (Patterson et al. 2015). 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. Of them, 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.⁵

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

⁵For 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.

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 pre-existing Elixhauser comorbidities (Elixhauser et al. 1998) 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.

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 (Appendix Table A.1). 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 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 years 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.

Figure 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.

4 Results

Racial differences in follow-up care Figure 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 minutes of the first pulse oximetry reading across race, and Panel (c) plots the fraction of patients to receive supplemental oxygen, again within 300 minutes 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.⁶

These differences are statistically significant – Panels (b) and (d) plots the racial differences and 95% CI constructed using heteroskedastic robust standard errors – 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).⁷ Moreover, these results are unaffected by controlling for observable differences across Black and white patients (Appendix Figure A.2), or measuring follow-up care through 150 minutes rather than 300 (Appendix Figure A.3), and are similar to what we find if we use ED visits in all VA facilities (Appendix Figure A.4).

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 Appendix Figure A.5 panels (a) and (b) shows 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 minutes) 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) Appendix Figure A.5 show that the

⁶And, naturally, also unconditionally, given that – as we saw in Figure 1 – Black patients are associated with higher initial measurement.

⁷In addition, consistent with prior work both at the VA (Valbuena et al. 2022) and in many other settings (Martin et al. 2024), 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. Appendix Figure A.1 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.

probability of any follow up (ABG, supplemental oxygen, or additional pulse oximetry) is 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.

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 [Figure 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. \(2025\)](#) which, as discussed in [Section 2.2](#), compare pulse oximetry measures of blood oxygenation to ABG measures *taken at the same time*. [Starnes et al. \(2025\)](#) 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 [Figure 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 is now about 19 percentage points lower (31.0% compared to 49.5%).

5 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, non-invasive tool to measure blood oxygenation, pulse oximeters over-estimate

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 (World Health Organization 2024). 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.

References

- Acemoglu, Daron, and Joshua Linn. 2004. “Market size in innovation: theory and evidence from the pharmaceutical industry.” *The Quarterly journal of economics* 119 (3): 1049–1090.
- Alsan, Marcella, Amitabh Chandra, and Kosali Simon. 2021. “The great unequalizer: initial health effects of COVID-19 in the United States.” *Journal of Economic Perspectives* 35 (3): 25–46.
- Alsan, Marcella, Maya Durvasula, Harsh Gupta, Joshua Schwartzstein, and Heidi Williams. 2024. “Representation and Extrapolation: Evidence from Clinical Trials.” *The Quarterly Journal of Economics* 139 (1): 575–635.
- American Thoracic Society and American College of Chest Physicians. 2003. “ATS/ACCP Statement on cardiopulmonary exercise testing.” *American Journal of Respiratory and Critical Care Medicine* 167 (2): 211–277.
- Arnold, David, Will Dobbie, and Peter Hull. 2022. “Measuring Racial Discrimination in Bail Decisions.” *American Economic Review* 112 (9): 2992–3038.
- Bajaj, Simar. 2022. “Perspective — Racial Bias Is Built into the Design of Pulse Oximeters.” Accessed: 2025-06-10. <https://www.washingtonpost.com/made-by-history/2022/07/27/racial-bias-is-built-into-design-pulse-oximeters/>.
- Baron, E. Jason, Joseph J. Jr. Doyle, Natalia Emanuel, Peter Hull, and Joseph P. Ryan. 2024. “Unwarranted Disparity in High-Stakes Decisions: Race Measurement and Policy Responses.” *NBER Working Paper No. 33104*.
- Baron, Rebecca M., and Anthony F. Massaro. 2025. “Approach to the Patient with Critical Illness.” Chap. 311 in *Harrison’s Principles of Internal Medicine*, 22nd ed., edited by Dennis L. Kasper, Anthony S. Fauci, Stephen Hauser, Dan Longo, J. Larry Jameson, Joseph Loscalzo, Steven M. Holland, and Carol A. Langford. New York: McGraw Hill.
- Bernard, Didem M., Abigail Woodroffe, and Siying Lui. 2022. “Understanding Veterans’ Healthcare Use and Experience, 2018–2019.” AHRQ MEPS, Accessed: 2025-07-27. https://meps.ahrq.gov/data_files/publications/rf50/rf50.shtml.
- Bhavani, Sivasubramaniam V, Zanthia Wiley, Philip A Verhoef, Craig M Coopersmith, and Ighovwerha Ofotokun. 2022. “Racial differences in detection of fever using temporal vs oral temperature measurements in hospitalized patients.” *Jama* 328 (9): 885–886.
- Bickler, Philip E, John R Feiner, and John W Severinghaus. 2005. “Effects of skin pigmentation on pulse oximeter accuracy at low saturation.” *Anesthesiology* 102 (4): 715–719.
- Bierman, Andrew, Kevin Benner, and Mark S. Rea. 2024. “Melanin Bias in Pulse Oximetry Explained by Light Source Spectral Bandwidth.” *British Journal of Anaesthesia* 132 (5): 957–963.
- Bohren, J Aislinn, Peter Hull, and Alex Imas. 2022. “Systemic discrimination: Theory and measurement.” *NBER Working Paper No. 29820*.

- Cables & Sensors. 2025. “History of Pulse Oximetry.” Accessed: 2025-06-10. <https://www.cablesandsensors.com/pages/history-of-pulse-oximetry>.
- Cecil, William T, Kevin J Thorpe, Eugene E Fibuch, and Gerald F Tuohy. 1988. “A clinical evaluation of the accuracy of the Nellcor N-100 and Ohmeda 3700 pulse oximeters.” *Journal of clinical monitoring* 4:31–36.
- Chen, Nicholas. 2019. “Fitbit and Other Wearables May Not Accurately Track Heart Rates in People with Darker Skin. That’s a Problem.” STAT News, Accessed 2025-06-10. <https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/>.
- Cleveland Clinic. 2022a. “Arterial blood gas (ABG): What it is, purpose, procedure & levels.” Accessed: 2025-05-28. <https://my.clevelandclinic.org/health/diagnostics/22409-arterial-blood-gas-abg>.
- . 2022b. “Blood Oxygen Level: What It Is, Purpose & Normal Range.” Accessed: 2025-06-05. <https://my.clevelandclinic.org/health/diagnostics/22447-blood-oxygen-level>.
- Conner, Peter, Liran Einav, Amy Finkelstein, and Petra Persson. 2025. “Targeting Precision Medicine: Evidence from Prenatal Screening.” *Journal of Political Economy* 133 (2): 604–651.
- Corredor-Waldron, Adriana, Janet Currie, and Molly Schnell. 2024. *Drivers of Racial Differences in C-Sections*. NBER Working Paper w32891. National Bureau of Economic Research.
- Currie, Janet, W Bentley MacLeod, and Kate Musen. 2024. “First Do No Harm? Doctor Decision Making and Patient Outcomes.” *NBER Working Paper No. 32788*.
- Diao, James A, Yixuan He, Rohan Khazanchi, MJ Nguemeni Tiako, Jonathan I Witonsky, Emma Pierson, Pranav Rajpurkar, Jennifer R Elhawary, Luke Melas-Kyriazi, Albert Yen, et al. 2024. “Implications of race adjustment in lung-function equations.” *The New England journal of medicine* 390 (22): 2083.
- Elixhauser, Anne, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey. 1998. “Comorbidity measures for use with administrative data.” *Medical Care* 36 (1): 8–27.
- Fawzy, Ashraf, Tianshi David Wu, Kunbo Wang, Kevin E. Sands, Alexander M. Fisher, Sarah Arnold Egloff, Joseph D. DellaVolpe, Theodore J. Iwashyna, Yiqiong Xu, and Brian T. Garibaldi. 2022. “Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19.” *JAMA Internal Medicine* 182 (7): 730–738.
- Feiner, John R, John W Severinghaus, and Philip E Bickler. 2007. “Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender.” *Anesthesia & Analgesia* 105 (6): S18–S23.
- Finkelstein, Amy. 2004. “Static and dynamic effects of health policy: Evidence from the vaccine industry.” *The Quarterly Journal of Economics* 119 (2): 527–564.

- Finkelstein, Amy, Geoffrey Kocks, Maria Polyakova, and Victoria Udalova. 2024. "Heterogeneity in Damages from a Pandemic." *Review of Economics and Statistics*, 1–45.
- Foster, Thomas B., Sonya R. Porter, and Nikolas Pharris-Ciurej. 2024. "Excess Mortality During COVID-19 Varied by Race, Ethnicity, Geography." U.S. Census Bureau, Accessed: 2025-06-10. <https://www.census.gov/library/stories/2024/03/excess-mortality-during-covid-19.html>.
- Gottlieb, Eric Raphael, Jennifer Ziegler, Katharine Morley, Barret Rush, and Leo Anthony Celi. 2022. "Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit." *JAMA internal medicine* 182 (8): 849–858.
- Hailu, Ruth. 2019. "Fitbits and other wearables may not accurately track heart rates in people of color." Stat News, Accessed: 2025-06-05. <https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/>.
- Al-Halawani, Rasha, Peter H. Charlton, Mohamed Qassem, and Panayiotis A. Kyriacou. 2023. "A Review of the Effect of Skin Pigmentation on Pulse Oximeter Accuracy." *Physiological Measurement* 44 (5): 05TR01.
- Henry, Nicole R, Andrew C Hanson, Phillip J Schulte, Nafisseh S Warner, Megan N Marento, Timothy J Weister, and Matthew A Warner. 2022. "Disparities in hypoxemia detection by pulse oximetry across self-identified racial groups and associations with clinical outcomes." *Critical care medicine* 50 (2): 204–211.
- Hulzebos, Christian V, Libor Vitek, Carlos D Coda Zabetta, Aleš Dvořák, Paul Schenk, Eline AE van der Hagen, Christa Cobbaert, and Claudio Tiribelli. 2021. "Screening methods for neonatal hyperbilirubinemia: benefits, limitations, requirements, and novel developments." *Pediatric research* 90 (2): 272–276.
- Koerber, Dana, Sadia Khan, Tina Shamsheer, et al. 2023. "Accuracy of Heart Rate Measurement with Wrist-Worn Wearable Devices in Various Skin Tones: a Systematic Review." *Journal of Racial and Ethnic Health Disparities* 10:2676–2684.
- Kress, John P., and Jesse B. Hall. 2015. "Approach to the Patient with Critical Illness." Chap. 321 in *Harrison's Principles of Internal Medicine*, 19th ed., edited by Dennis L. Kasper, Anthony S. Fauci, Stephen Hauser, Dan Longo, J. Larry Jameson, and Joseph Loscalzo. Archived version. New York: McGraw Hill.
- Martin, Daniel, Chris Johns, Lexy Sorrell, Eugene Healy, Mandeep Phull, Segun Olusanya, Mark Peters, and Jeremy Fabes. 2024. "Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: a systematic review." *British Journal of Anaesthesia* 132 (5): 945–956.
- Matos, João, Mahmoud Alwakeel, Sicheng Hao, Dana Naamani, Tristan Struja, Judy Wawira Gichoya, Leo Anthony Celi, Timothy McMahon, Heather A King, Christopher E Cox, et al. 2025. "Differences in arterial blood gas testing by race and sex across 161 US hospitals in four electronic health record databases." *American Journal of Respiratory and Critical Care Medicine* 211 (6): 1049–1058.

- McFarling, Usha Lee, and Katherine Gilyard. 2022. “Faulty oxygen readings delayed Covid treatments for darker-skinned patients, study finds.” Stat News, Accessed: 2025-06-05. <https://www.statnews.com/2022/05/31/faulty-oxygen-readings-delayed-covid-treatments-darker-skin-patients/>.
- Merrick, Edwin B., and Thomas J. Hayes. 1976. “Continuous, Non-Invasive Measurements of Arterial Blood Oxygen Levels.” Accessed: 2025-05-27, *Hewlett-Packard Journal* 28 (2): 2–9.
- Miyasaka, Kazuyuki, Kirk Shelley, Satoru Takahashi, Hiroyuki Kubota, Koji Ito, Ikuto Yoshiya, Atsushi Yamanishi, et al. 2021. “Tribute to Dr. Takuo Aoyagi, Inventor of Pulse Oximetry.” Epub 2021 Aug 2, *Journal of Anesthesia* 35 (5): 671–709.
- Moran-Thomas, Amy. 2020. “How a Popular Medical Device Encodes Racial Bias.” Boston Review, Accessed: 2025-06-05. <http://bostonreview.net/science-nature-race/amy-moran-thomas-how-popular-medical-device-encodes-racial-bias>.
- . 2021. “Oximeters Used to Be Designed for Equity. What Happened?” Accessed: 2023-12-02. <https://www.wired.com/story/pulse-oximeters-equity/>.
- National Academies of Sciences, Engineering, and Medicine. 2024. *Ending Unequal Treatment: Strategies to Achieve Equitable Health Care and Optimal Health for All*. Edited by Georges C. Benjamin, Jennifer E. DeVoe, Francis K. Amankwah, and Sharyl J. Nass. Washington, DC: The National Academies Press.
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. “Dissecting racial bias in an algorithm used to manage the health of populations.” *Science* 366 (6464): 447–453.
- Patterson, Olga V., Makoto Jones, Yiwen Yao, Benjamin Viernes, Patrick R. Alba, Theodore J. Iwashyna, and Scott L. DuVall. 2015. “Extraction of Vital Signs from Clinical Notes.” *Stud Health Technol Inform* 216 (6): 1035.
- Patwari, Neal, Di Huang, and Francesca Bonetta-Misteli. 2023. “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*, 143–147.
- Philip, Minu, and Ozde Ozkaya. 2025. “Disparate Treatment and Outcomes in Emergency Departments: Evidence from Florida.” Job Market Paper, New York University, *Unpublished Manuscript*.
- Science Staff. 2022. “‘Racially biased’ devices caused delayed treatment for Black COVID-19 patients.” Accessed: 2025-06-05. <https://www.science.org/content/article/racially-biased-devices-caused-delayed-treatment-black-covid-19-patients>.
- Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. 2020. “Racial Bias in Pulse Oximetry Measurement.” *New England Journal of Medicine* 383 (25): 2477–2478.

- Smedley, Brian D, Adrienne Y Stith, Alan R Nelson, et al. 2003. “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*. National Academies Press (US).
- Starnes, Joseph R., Wendi Welch, Christopher C. Henderson, Stephen Hudson, Scott Risney, George T. Nicholson, Thomas P. Doyle, et al. 2025. “Pulse Oximetry and Skin Tone in Children.” *New England Journal of Medicine* 392 (10): 1033–1034.
- Thoracic Society of Australia and New Zealand. 2015. “Oxygen Guidelines for Acute Oxygen Use in Adults: ‘Swimming between the flags’.” Accessed: 2025-05-27. <https://thoracic.org.au/wp-content/uploads/2021/06/TSANZ-AcuteOxygen-Guidelines-2016-web.pdf>.
- Torp, Klaus D., Pranav Modi, Elizabeth J. Pollard, and Leslie V. Simon. 2023. “Pulse Oximetry.” [Updated 2023 Jul 30]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, Accessed: 2025-05-27. <https://www.ncbi.nlm.nih.gov/books/NBK470348/>.
- Tuttle, Cody. 2019. “Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums.” Available at SSRN: <https://ssrn.com/abstract=3080463>.
- U.S. Food and Drug Administration. 2021. “Pulse Oximeter Accuracy and Limitations: FDA Safety Communication.” Accessed: 2025-05-27. <https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication>.
- . 2025. “Pulse Oximeters for Medical Purposes – Non-Clinical and Clinical Performance Testing, Labeling, and Premarket Submission Recommendations.” Draft Guidance for Industry and FDA Staff.
- University of Iowa Department of Otolaryngology. 2024. “Pulse Oximetry: Basic Principles and Interpretation.” Accessed: 2025-05-27. <https://medicine.uiowa.edu/iowaprotocols/pulse-oximetry-basic-principles-and-interpretation>.
- Valbuena, Valeria S.M., Michael W. Sjoding, Emily Gilbert, Thomas S. Valley, Colin R. Cooke, and Hallie C. Prescott. 2022. “Racial Bias and Reproducibility in Pulse Oximetry Among Medical and Surgical Inpatients in General Care in the Veterans Health Administration 2013–19: Multicenter, Retrospective Cohort Study.” *BMJ* 378.
- Vyas, Darshali A, Leo G Eisenstein, and David S Jones. 2020. “Hidden in plain sight—reconsidering the use of race correction in clinical algorithms.” *New England Journal of Medicine* 383 (9): 874–882.
- Winn, Annalies, and Nicole Jurmo. 2024. “The Problem with Pulse Oximeters: A Long History of Racial Bias.” Johns Hopkins Bloomberg School of Public Health. Accessed: 2025-05-27. <https://publichealth.jhu.edu/2024/pulse-oximeters-racial-bias>.
- Witt, Chance. 2014. “Vital signs are vital: The history of pulse oximetry.” *ACP Hospitalist*.
- World Health Organization. 2024. “Oxygen.” Accessed February 1, 2026. Accessed February 1, 2026. <https://www.who.int/health-topics/oxygen>.

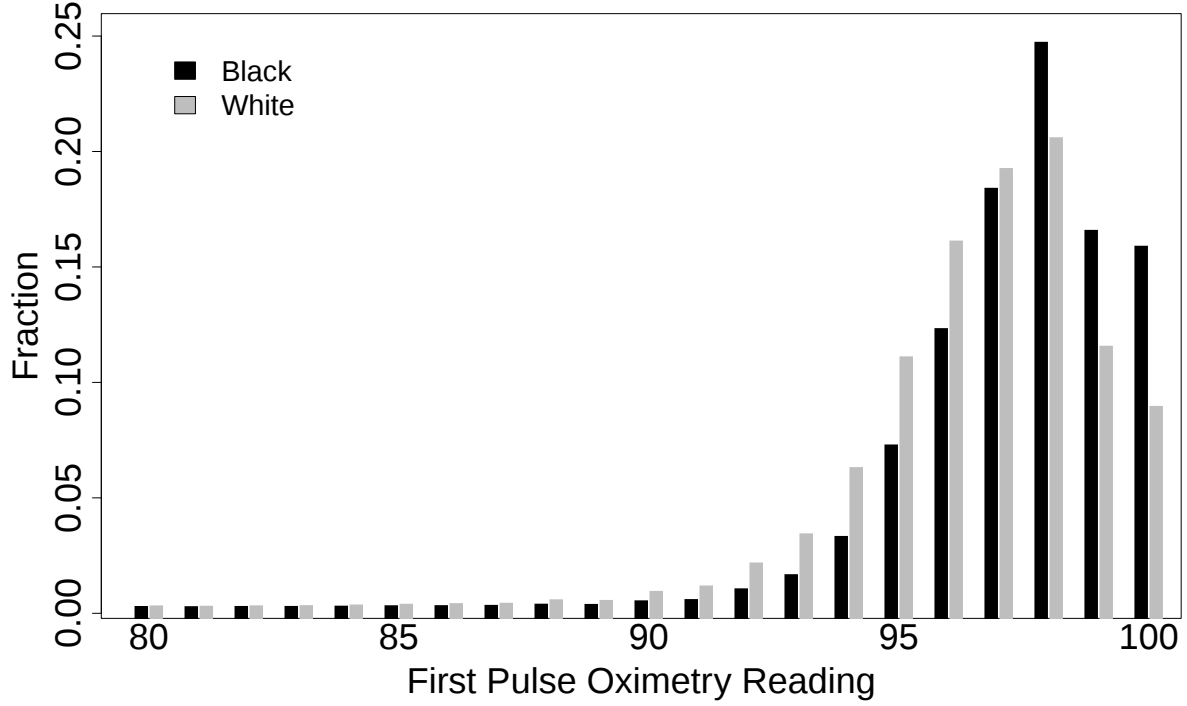
Yale Medicine. 2022. "Pulse Oximetry." Accessed: 2025-05-27. <https://www.yalemedicine.org/conditions/pulse-oximetry>.

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: 18-34	0.08	0.10	-0.02
Age: 35-49	0.17	0.13	0.04
Age: 50-64	0.45	0.29	0.16
Age: 65-79	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	

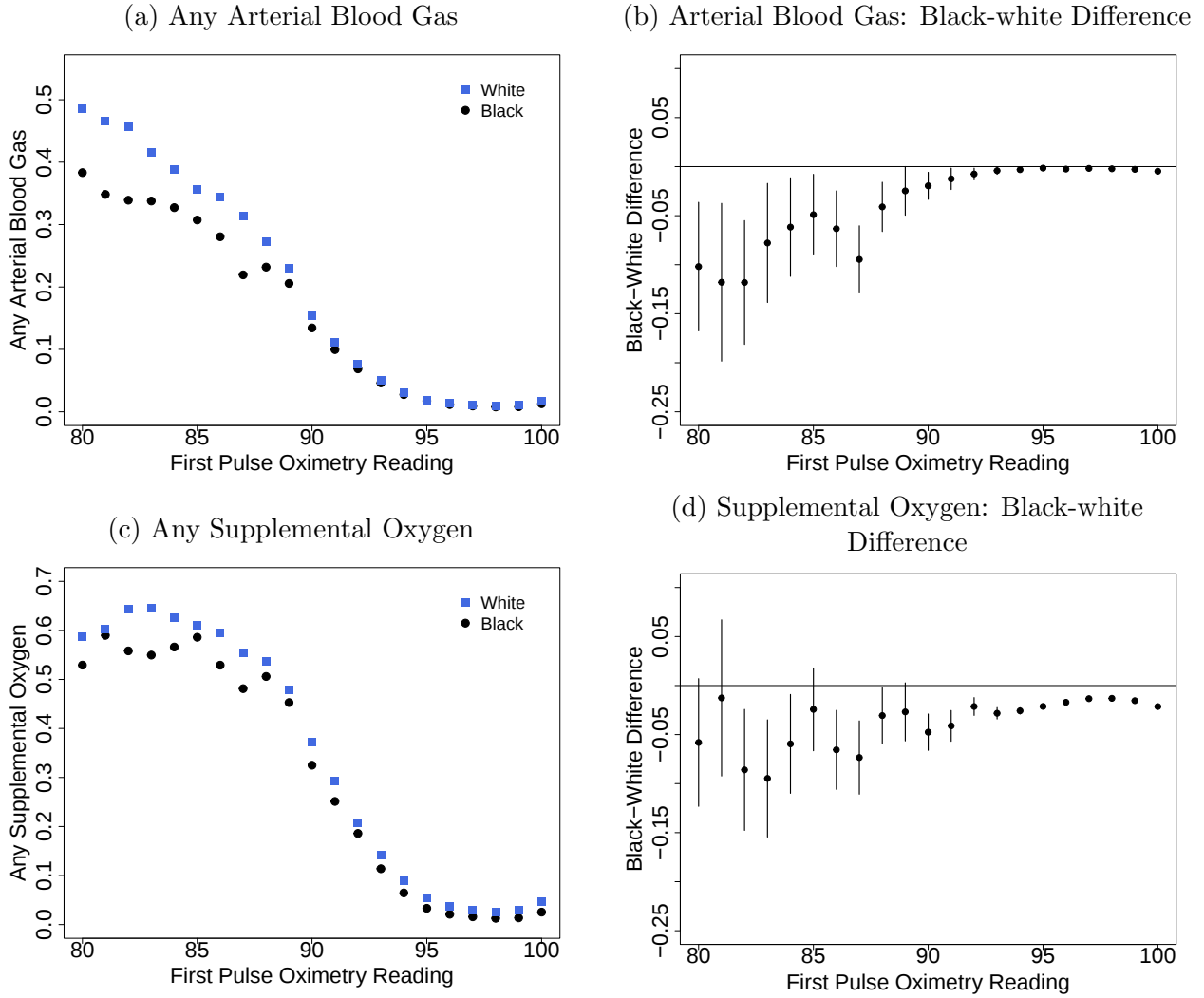
Notes: This table presents summary statistics at the ED visit level. Comorbidities are diagnoses in the year prior to the ED visit. There are a maximum of 30 Elixhauser comorbidities; we report the average number, as well as rates for the top 5 most common ones in our sample. Column 1 and 2 shows the 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)).

Figure 1: Distribution of First Pulse Oximetry Measurements, by Race



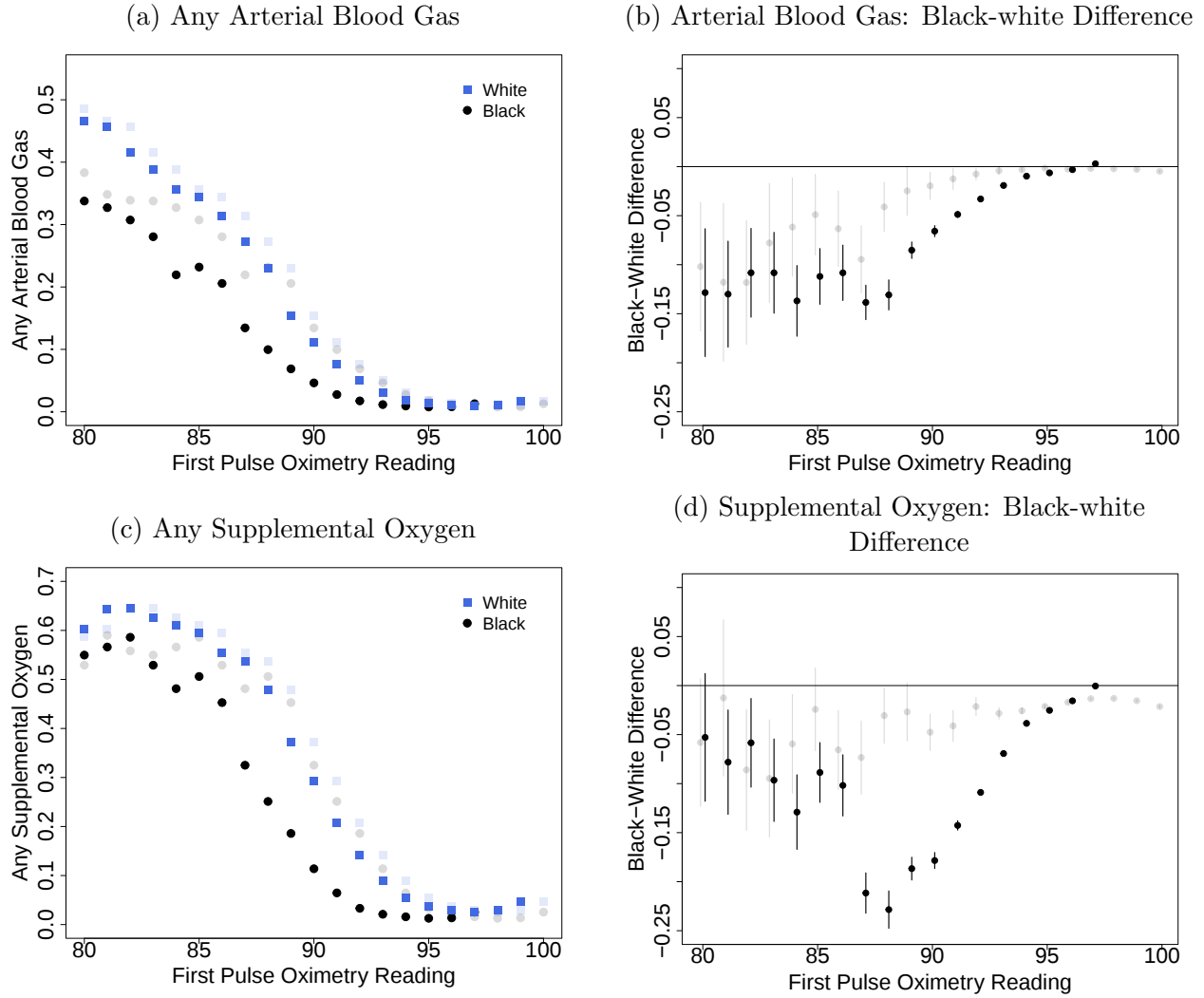
Notes: This figure plots the distribution of the first pulse oximetry reading taken in the emergency department for Black non-Hispanic (black bars) and white non-Hispanic patients (gray bars) in our primary analysis, which consists of 1,147,893 ED visits and 2,366,044 ED visits with pulse oximetry readings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

Figure 2: Racial Differences in Follow-Up Care



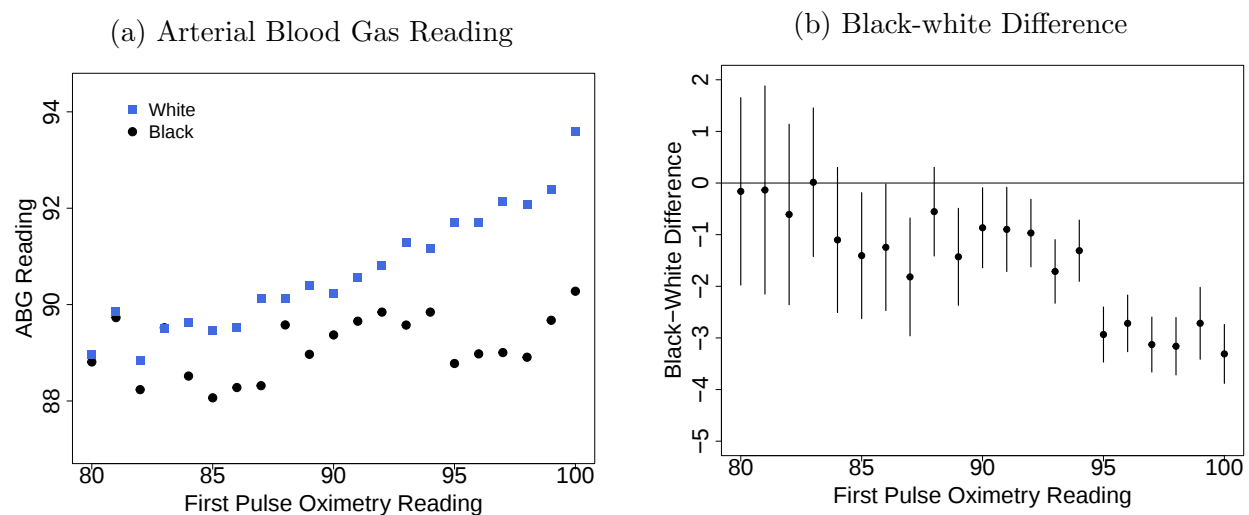
Notes: This figure plots two clinical decisions by first pulse oximetry reading in the ED and by race: any ABG (top row) and any supplemental oxygen (bottom row), both within 300 minutes of the first pulse oximetry measurement. Panel (a) shows the fraction of patients who receive an arterial blood gas within 300 minutes 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 minutes and the 95% confidence interval (calculated using heteroskedastic robust standard errors). Panel (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.

Figure 3: Racial Differences in Follow-Up Care, Corrected



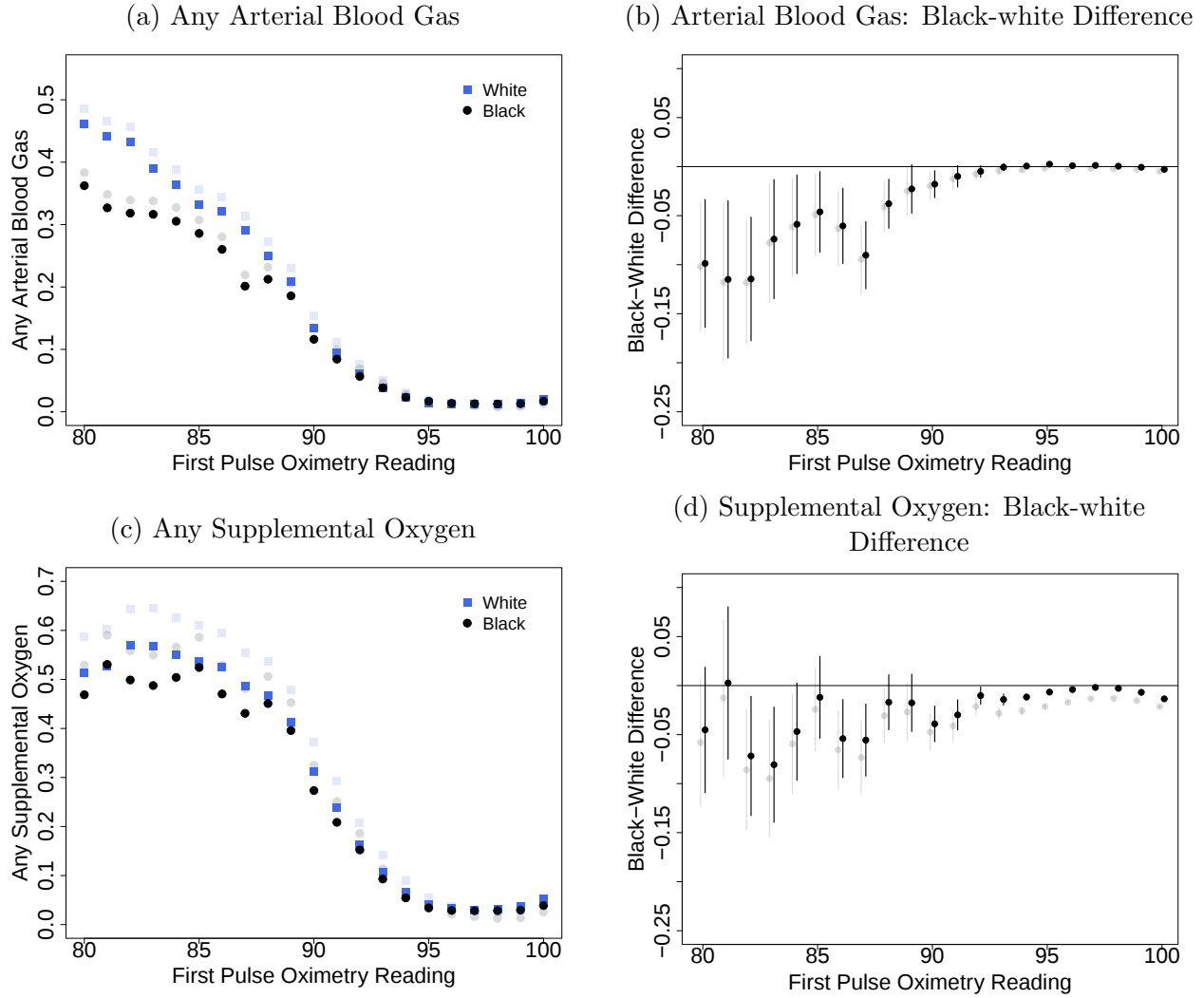
Notes: The “corrected” pulse oximetry readings use the Starnes et al. (2025) correction. Specifically, using their Table 1 results, rounded to the nearest integer, we “correct” pulse oximetry by subtracting 1 from white patients and 3 from Black patients. Across all panels, the lighter transparent points are non-corrected and identical to Figure 2. The darker points are corrected. Blue squares and black circles correspond to white and Black patients, respectively. The 95% confidence intervals are calculated using heteroskedastic robust standard errors. 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.

Appendix Figure A.1: Racial Differences in ABG Reading, Conditional on an Arterial Blood Gas



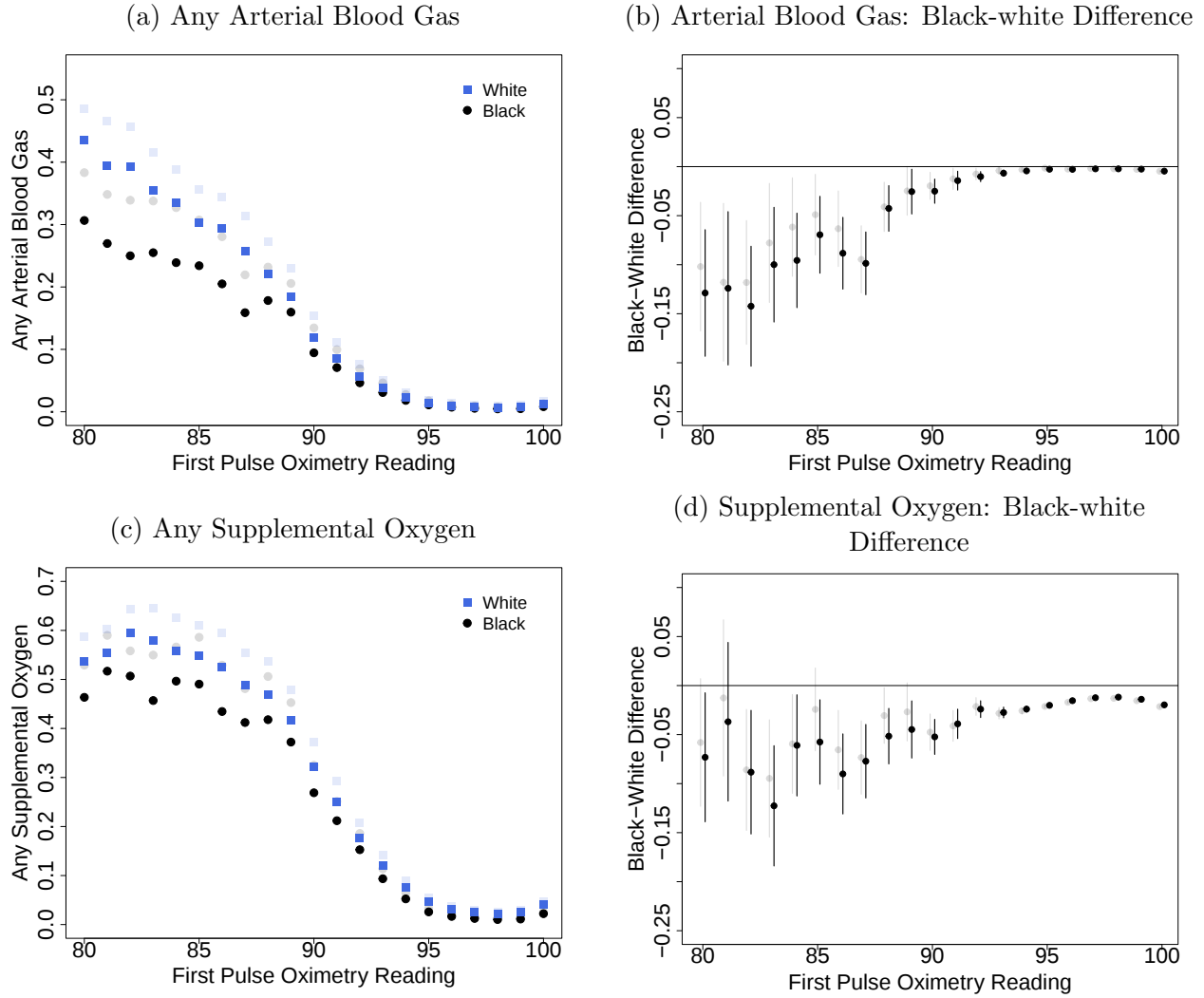
Notes: This figure plots the average arterial blood gas (ABG) reading conditional on receiving an ABG within 300 minutes of the first pulse oximetry measurement in the ED, by first pulse oximetry and by race. Panel (a) shows the mean ABG reading for white (non-Hispanic) patients (blue squares) and Black (non-Hispanic) patients (black circles). Panel (b) plots the average Black-white difference in ABG reading and the 95% confidence interval (calculated using heteroskedastic robust standard errors). 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.

Appendix Figure A.2: Racial Differences in Follow-Up Care, Controlling for Observables



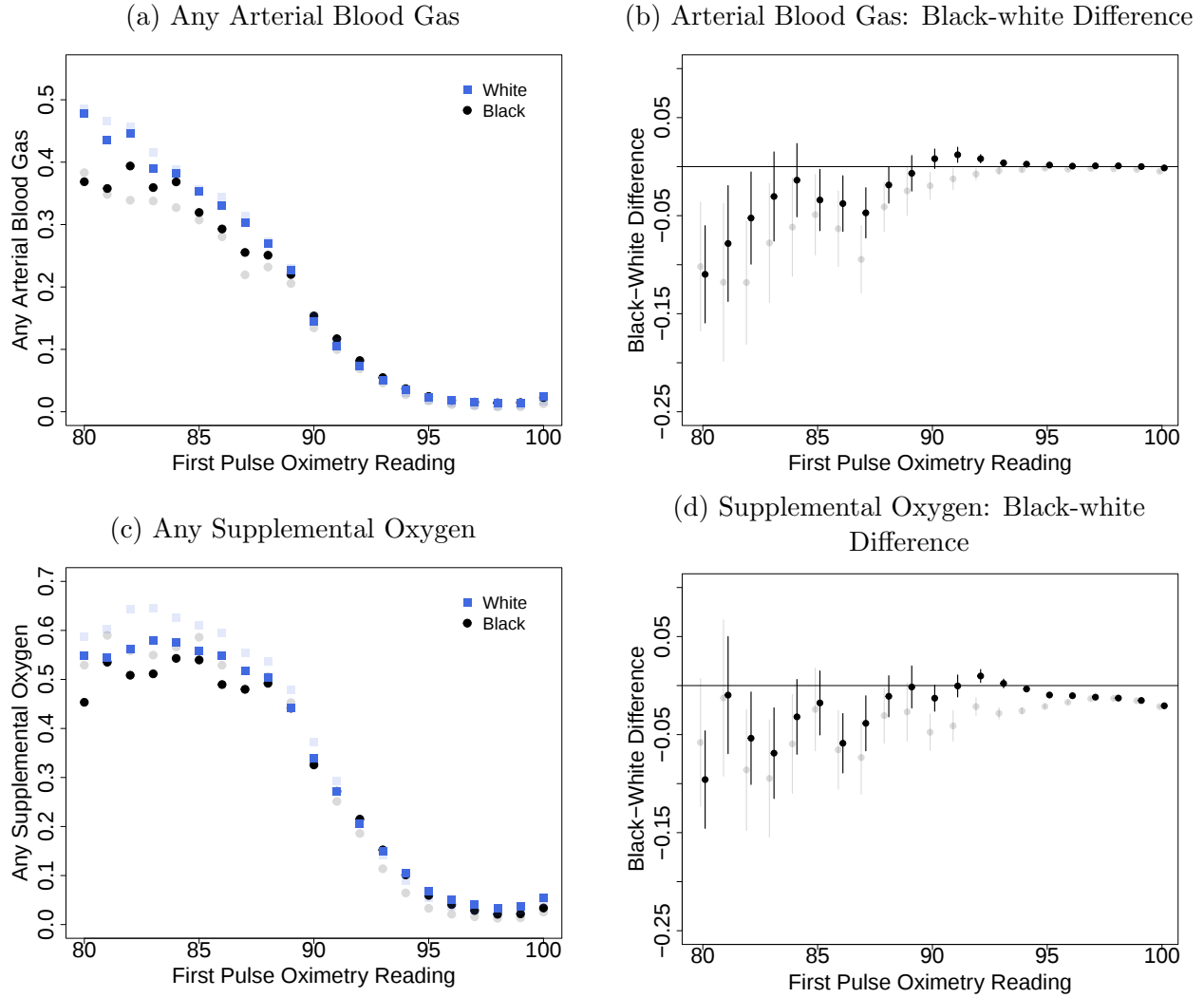
Notes: This figure reproduces our main analyses from [Figure 2](#) after controlling flexibly for age, sex, and prior comorbidities. Specifically, each outcome (any arterial blood gas or any supplemental oxygen within 300 minutes of the first pulse oximetry measurement) is regressed on an indicator for female, age (each year) fixed effects, and indicators for each of the 29 Elixhauser comorbidities. Panels (a) and (c) shows the mean of the residualized outcome, after adding back the mean. Panels (b) and (d) shows the Black-white difference for each outcome. Across all panels, the lighter transparent points are non-corrected and are identical to [Figure 2](#). The darker points are after controlling for age, sex, and prior comorbidities. Blue squares and black circles correspond to white and Black patients, respectively. The 95% confidence intervals are calculated using heteroskedastic robust standard errors. 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.

Appendix Figure A.3: Racial Differences in Follow-Up Care, 150 Minutes



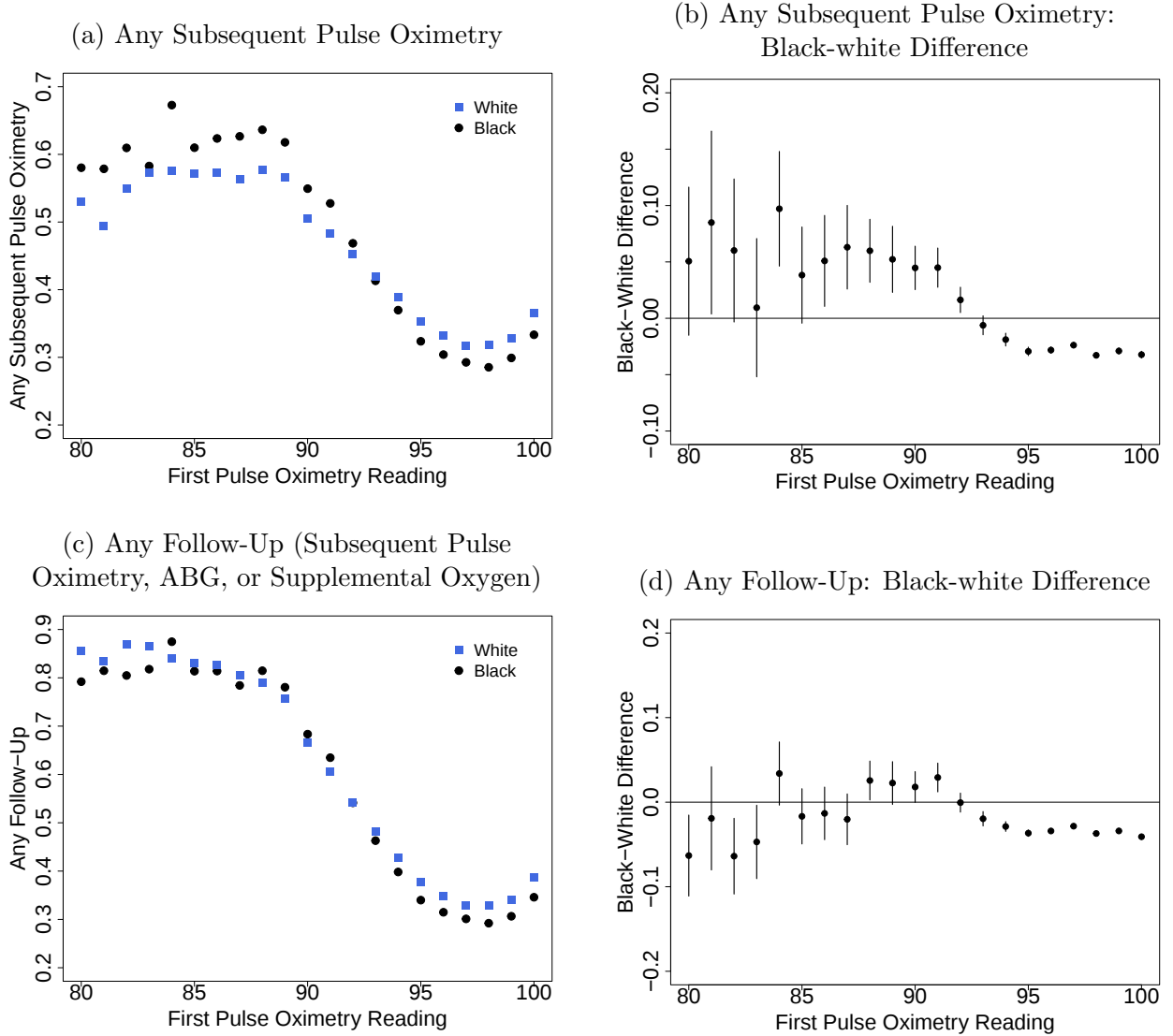
Notes: This figure reproduces our main analyses from [Figure 2](#) after but measuring follow-up care over 150 minutes from the start of the ED visit, instead of 300 minutes. Across all panels, the lighter transparent points are for 300 minutes and are identical to [Figure 2](#). The darker points are for 150 minutes. Blue squares and black circles correspond to white and Black patients, respectively. The 95% confidence intervals are calculated using heteroskedastic robust standard errors. 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.

Appendix Figure A.4: Racial Differences in Follow-Up Care, All VA Facilities



Notes: This figure repeats our main results from [Figure 2](#) using observations from all 105 VA facilities for which we observe an initial pulse oximetry reading. There are 6,114,728 ED visits with pulse oximetry readings of at least 80; 1,922,503 (31.4%) are Black patients. Across all panels, the lighter transparent points are our baseline results (for the 39 facilities that have pulse oximetry measurement for over 90% of ED visits, with the sample consisting of 1,147,893 ED visits and 2,366,044 ED visits for Black and white patients, respectively) and are identical to [Figure 2](#). The darker points are for all 105 facilities. Blue squares and black circles correspond to white and Black patients, respectively. The 95% confidence intervals are calculated using heteroskedastic robust standard errors.

Appendix Figure A.5: Racial Differences in Subsequent Pulse Oximetry and Follow-Up



Notes: This figure plots the likelihood of taking a subsequent pulse oximetry measurement within 300 minutes of the first reading or any follow-up action, by first pulse oximetry reading and by race. Specifically, if the patient has an ABG test within 300 minutes or receives supplemental oxygen, subsequent pulse oximetry is defined as a subsequent measure prior to that ABG test or supplemental oxygen. Panel (a) shows the mean fraction receiving any subsequent pulse oximetry for white patients (blue squares) and Black patients (black circles). Panel (b) plots the average Black-white difference in taking a subsequent pulse oximetry and the 95% confidence interval. Panel (c) shows the mean fraction receiving any follow-up, defined as subsequent pulse oximetry, arterial blood gas, or supplemental oxygen within 300 minutes, provided that pulse oximetry does not occur after ABG or supplemental oxygen. Panel (d) plots the average Black-white difference in any action and the 95% confidence intervals (calculated using heteroskedastic robust standard errors). 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.

Appendix Table A.1: No Differences by Race in Rates of Observing Pulse Oximetry

	<i>Dependent variable:</i>			
	Any Pulse Oximetry?			
	All Facilities		High Pulse Ox Rate Facilities	
	(1)	(2)	(3)	(4)
Black Patient	0.023 (0.031)	0.026 (0.031)	−0.002 (0.003)	−0.001 (0.003)
Mean (White):	0.7	0.7	0.96	0.96
Controls:		Yes		Yes
Number of Facilities:	105	105	39	39
Observations	8,646,045	8,646,045	3,689,401	3,689,401

Notes: This table investigates whether there are differential rates by race in observing a pulse oximetry measure. For this analysis we remove the restriction that the pulse oximetry measure must be 80 or above. We regress an indicator variable for whether the ED visit has at least one pulse oximetry reading on an indicator for Black race, among all emergency department visits between 2014 and 2018 for Black non-Hispanic and white non-Hispanic patients. Columns (1) and (2) estimates this regression using all 105 VA facilities and columns (3) and (4) estimates it on our baseline sample of the 39 facilities that have pulse oximetry measurement for over 90% of encounters. Columns (2) and (4) include controls for sex, age fixed effects, and 3 digit ICD-10 diagnosis fixed effects. Robust standard errors are clustered at the facility level. *p<0.1; **p<0.05; ***p<0.01.