
Research Article / Special Article

Pulse Oximeter Performance, Racial Inequity, and the Work Ahead

<https://doi.org/10.4187/respcare.09795>

Cite as: RESPCARE 2021; 10.4187/respcare.09795

Received: 10 November 2021

Accepted: 11 November 2021

This Fast Track article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any supplemental data.

Alerts: Sign up at rc.rcjournal.com/alerts to receive customized email alerts when the fully formatted version of this article is published.

Pulse Oximeter Performance, Racial Inequity, and the Work Ahead

Olubunmi Okunlola^{1,2}, Michael S. Lipnick^{1,2,3}, Paul Batchelder⁴, Michael Bernstein^{2,5}, John
Feiner^{1,2,3}, Philip Bickler^{1,2,3}

Affiliations:

1. Center for Health Equity in Surgery and Anesthesia, University of California San Francisco
2. Hypoxia Research Laboratory, University of California San Francisco
3. Department of Anesthesia and Perioperative Care, University of California San Francisco
4. Clinimark Labs
5. Physio Monitor LLC

Corresponding Author:

Michael S Lipnick MD

1001 Potrero Avenue

San Francisco, CA 94110

E-mail: Michael.lipnick@ucsf.edu

COI Statement

The UCSF Hypoxia Research Laboratory, Clinimark and Physio Monitor LLC charge pulse oximeter manufacturers for performing validation studies, but no companies were involved with the writing or data analysis presented in this paper.

Abstract**Introduction****Proposed Changes**

Recognize That Pulse Oximeter Errors in Patients With Darkly Pigmented

Skin Are Real and Require Further Examination

Understand “Sjoding’s Outliers” and Conduct Real-World, Patient-Centered

Trials of Performance

Ensure Adequate Diversity for Study Subjects in Testing Protocols

Harmonize Standards for Accuracy

Address the Need for More Evaluation and Transparency for Inexpensive or

Non-FDA-cleared Oximeters

Conclusions

Abstract

It has long been known that many pulse oximeters function less accurately in patients with darker skin. Reasons for this observation are incompletely characterized and potentially enabled by limitations in existing regulatory oversight. Based on decades of experience and unpublished data, we believe it is feasible to fully characterize, in the public domain, the factors that contribute to missing clinically significant hypoxemia in patients with darkly pigmented skin. Here we propose five priority areas of inquiry for the research community and actionable changes to current regulations that will help improve oximeter accuracy. We propose that leading regulatory agencies should immediately modify standards for: measuring accuracy and precision of oximeter performance, analyzing and reporting performance outliers, diversifying study subject pools, thoughtfully defining skin pigmentation, reporting data transparently, and accounting for performance during low perfusion states. These changes will help reduce bias in pulse oximeter performance and improve access to safe oximeters.

Introduction

The majority of the world's poor populations have no access to reliable pulse oximetry. While this challenge is often thought of as a matter of device availability in low and middle-income countries (LMICs), the inequity in pulse oximetry is more complex. Pulse oximetry is receiving unprecedented attention for its essential role in COVID-19 patient management as well as new studies reconfirming the technology's inaccuracy in patients with darkly pigmented skin. Retrospective data from work published earlier this year by Sjoding et al. identified large enough

errors in pulse oximeter performance, for patients who self-identified as Black, to miss clinically significant hypoxemia.^{1,2} As a result, the US Food and Drug Administration (FDA) received a Congressional Inquiry requesting a “review of the interaction between a patient’s skin color and the accuracy of pulse oximetry measurements”.³

Based on decades of experience with pulse oximeter testing and development at the UCSF Hypoxia Research Laboratory and at Clinimark Laboratories, we believe it is feasible and necessary to fully characterize, in the public domain, the factors that may contribute to missing clinically significant hypoxemia in patients with darkly pigmented skin. Here we propose five priority areas of inquiry, actionable changes to current regulations, and initial steps that will improve access to safe oximeters.

Proposed Changes

Recognize That Pulse Oximeter Errors in Patients With Darkly Pigmented Skin Are Real and Require Further Examination

Skin pigment has long been observed to impact oximeter performance. The earliest studies had significant limitations in design and reported mixed findings. To help clarify the impact of skin pigment on oximetry, in 2005, the UCSF Hypoxia Research Laboratory undertook a controlled, prospective study which reported a positive bias for pulse oximeters in darkly-pigmented healthy test subjects in instruments from three different manufacturers.⁴ This study raised concerns that confirmatory tests used by the FDA to approve devices, which chiefly use light-skinned volunteers, may be insufficient. Although there is renewed interest in this issue due to the high

prevalence of hypoxemia in patients with COVID-19, we are disappointed that in over 15 years since this study, the magnitude and clinical importance of these errors remain neglected and incompletely defined.

To begin to understand if the current generation of pulse oximeters continues to read high in subjects with dark skin, we analyzed data from 18 hypoxemia studies completed over the last 3 years and including 9 pulse oximeter brands, employing both transmission and reflectance, with three sensor types (reusable finger-clip, disposable adhesive- finger, disposable adhesive-forehead). The studies were conducted in accordance with the International Organization for Standardization (ISO), FDA guidelines on pulse oximeters, and more detailed techniques as described previously.⁵ Subjects breathed an air/nitrogen mixture controlled to attain multiple stable levels of hypoxemia between 67% and 100% arterial oxygen saturation (SaO_2). This included 3,778 data pairs (849 darkly pigmented pairs, 2929 lightly pigmented pairs) from 491 subjects (108 darkly pigmented skin, 383 lightly pigmented skin). Each subject contributed similar numbers of points in the range of interest (± 2 per subject). These data are shown in Figure 1.

Two significant observations are evident in the comparison presented in Figure 1. The first is that a small positive bias in oximeter readings still exists in individuals with dark skin pigmentation, similarly observed by Bickler and Feiner in 2005 and 2007.^{4,6} The second is that the paucity of data points within the red rectangle in the Clinimark laboratory data, as compared to the number of data points within the red rectangle in Sjoding's data, supports the notion that the performance of pulse oximeters in clinical environments is different from that in ideal laboratory conditions.⁷ Laboratory hypoxia study subjects are screened for good health. Additionally, the incidence of many confounding issues for pulse oximetry including low perfusion, motion, marginal probe

positioning, irregular heartbeat, or unusual pulse morphology is low compared to a clinical setting. It is also important to note that, to date, the effects of skin pigment on pulse oximeter performance have been published for only a relatively small number of pulse oximeter models. More studies in both the clinical and laboratory setting are needed to understand the impact of skin pigmentation on SpO₂ in currently manufactured oximeters.

In pulse oximetry, the ratio of absorbance of two different wavelengths of light, red and infrared (IR), by oxyhemoglobin and deoxyhemoglobin is measured and mathematically translated into an oxygen saturation reading (SpO₂). This empiric calibration requires hypoxia testing in human subjects and may fail when calibration is based on a non-representative study population. Errors generated in this way are likely amplified by low perfusion, motion and interfering factors.

While there is greater absorbance of red light by melanin in darker skin, potentially resulting in positive bias in saturation estimates (i.e. pulse oximeters reading higher than the true blood oxygen saturation), this effect is incompletely characterized.⁸ Further studies are necessary to quantify this effect, determine if dark skin pigment is correlated with other factors that may impact oximeter performance, and to determine the impact of dark skin pigment when other factors that affect signal quality and signal processing may co-exist (e.g. low perfusion or low quality oximeter software and hardware).

Understand “Sjoding’s Outliers” and Conduct Real-World, Patient-Centered Trials of Performance

Based on the data presented in Figure 1 we suggest that factors including but not limited to oximeter design may explain what Sjoding reported. First, further characterizing the effect of skin pigmentation on the accuracy of pulse oximeter saturation readings during hypoxemia

requires a well-controlled testing environment. Arterial oxygen saturation is prone to fluctuations over short periods of time, necessitating that the blood sampling and the pulse oximeter reading be done as close together as possible for accurate comparisons (i.e., seconds, not minutes). The ability to control and maintain a steady state of hypoxemia from which to record measurements is a benefit of the laboratory environment. If a researcher in the clinical environment is unable to record an SpO₂ reading from the pulse oximeter and a corresponding arterial oxygen saturation within a narrow window, then the data should be interpreted with great caution. In the Sjoding study, some paired recordings were documented up to 10 minutes apart from each other.

It is also important to acknowledge that discordance between SpO₂ and SaO₂ values can be due to many factors besides skin pigmentation. Such factors include variations in breathing, excessive motion, incorrectly applied probes, anemia, temperature, and low perfusion.⁹⁻¹¹ All studies, whether in the clinical or laboratory setting, should record and account for these factors. Establishing standard data sets that include sufficient information to account for potential impact of outliers on clinical performance is necessary for improving oximetry performance.

Low perfusion is overdue for special consideration. Published data demonstrate that low perfusion causes error and dysfunction for some oximeters, and unpublished data show this error can be significant especially for some low-cost, fingertip devices.¹⁰ Furthermore, the combination of low perfusion and darkly pigmented skin also produces a significant degree of discordance between SpO₂ and SaO₂ and is likely a factor in the clinical environment of Sjoding's study. Currently, there are no requirements that low perfusion be accounted for when assessing pulse oximeter performance for clinical approval. We propose development of standardized protocols for measuring, reporting, and accounting for perfusion that can be incorporated into certification standards such as those by the FDA. This should include reporting

performance thresholds during low perfusion. As an interim step, we propose inclusion of available perfusion data in all studies of pulse oximetry accuracy. For example, one might report the ‘perfusion index’ or a comparable but standardized value, as measured by the study device or a reference device. Better characterizing the effect of low perfusion on oximeter performance is an essential step toward rectifying bias caused by skin pigment and may in fact be fundamentally related.

Ensure Adequate Diversity for Study Subjects in Testing Protocols

Standard practice in clinical trials of pulse oximeters involves pooling data from all subjects to derive overall performance data, including the mean bias and the root mean square error (A_{RMS}). Such an approach may hide outliers of performance in darkly pigmented individuals. While all data are typically presented to regulatory agencies for device clearance, individual examiners may be tasked with raising concerns about these performance outliers.

The current standard for FDA approval of pulse oximeters follows the 510k clearance process, last updated in 2013, by which a manufacturer proves that their medical device meets the standards previously approved for constructions in the same category.¹² At present, the approval process for pulse oximeters requires at least two volunteers or 15% of the study group, whichever is higher, to be “darkly-pigmented,” but does not provide specific details about exactly how to meet this requirement.¹² Many pulse oximeter studies, like the UCSF study in 2005, have employed the use of the Fitzpatrick phototype scale, a method of classifying human skin by phototype based on the presence of melanin. However, despite commonly being used for oximetry validation studies, the Fitzpatrick classification was not designed for this purpose and has many limitations, including potential for inter-operator variability and subjectivity.¹³ Using

race to define patients with dark skin pigmentation, as was done in the Sjoding et al. study, is also problematic on many levels. As such, we propose utilizing standardized and less subjective methods for classifying research subjects into categories by degree of skin pigmentation for the purpose of understanding the impact of skin pigment on oximeter performance. These methods should also be used to determine skin pigment at the site of oximeter measurement (e.g., ear or finger), including both surfaces of measurement for transmittance devices (e.g., the nail bed and ventral aspect of the fingertip). There are several such methods employed by the dermatology community that can quantify skin color at the site of measurement.¹⁴ By objectively quantifying skin color, researchers can better distribute study subject representation to accurately account for the spectrum of skin pigmentation. Additionally, we propose establishing standards that require manufacturers to demonstrate that a pulse oximeter meets performance standards in patients with darker skin when data are analyzed alone for patients with dark skin pigmentation, and not only when combined in an A_{RMS} fashion with data from lighter skin patients. Currently, pooled analyses are the common practice and only requirement for FDA 510k approval.

The inclusion of more study subjects with darkly pigmented skin in pulse oximetry research must be done thoughtfully for two additional reasons. First, the majority of device testing is done in high-income countries with citizens that may be predominantly lighter skinned. Additionally, we acknowledge that due to the history of medical injustice as well inequities that persist today in the United States, recruitment of Black subjects for research is negatively affected by mistrust of medical research establishments.¹⁵

Harmonize Standards for Accuracy

The most widely accepted performance standards for pulse oximetry are specified by the FDA and ISO (Table 1). The FDA requires A_{RMS} of less than 3% for transmittance devices and 3.5% for reflectance devices, and the ISO 80601 requires A_{RMS} of 4%.^{12,16} During the COVID-19 pandemic, consumer grade pulse oximeters have been widely used for clinical care decision making, both in the home and in hospital settings. We propose that all devices, whether reflectance or transmittance, should meet 3% A_{RMS} to receive FDA 510k or compliance with ISO.

Address the Need for More Evaluation and Transparency for Inexpensive or Non-FDA-Cleared Oximeters

In recent years, there has been a surge in the number of pulse oximeters on the consumer market for less than US\$100 and many for less than US\$30. While a lower price point may increase access, in our opinion it also increases the number of unsafe oximeters being used for clinical care, especially in LMICs.

Some inexpensive oximeters perform well enough in laboratory studies to meet FDA accuracy requirements for clinical care, but many do not.¹⁷ Poor performance of many inexpensive devices may be especially true under conditions of low perfusion and dark skin pigmentation among others. Consumers, clinicians, aid organizations and healthcare systems may be unaware of the shortcomings of these devices due to lack of transparency by some manufacturers and laxity in certification requirements. This is especially noteworthy with the recent uptick in donations and utilization of oximeters during the COVID-19 pandemic.

We call for more testing of consumer-marketed pulse oximeters in human subjects in laboratories capable of rigorous, transparent analysis and compliance with ISO 17025 and ISO

14155, as well as sharing of such data. To promote this goal, we are working with multiple collaborators to launch the OpenOximetry.org Project to accelerate testing and dissemination of oximeter performance data. This includes independent testing of devices that do and do not currently have FDA 510k, and is based on our finding that some FDA approved devices do not meet performance standards for FDA 510k certification.

A summary of our recommendations for improving performance and equity in pulse oximetry is presented in Table 1.

Conclusions

We propose the following steps to reduce bias in pulse oximeter performance and improve access to safe oximeters: standardize the characterization of pulse oximeter errors in patients with darkly pigmented skin, closely examine the performance outliers and conducting patient-centered trials, standardizing diversification of study subjects in testing protocols, harmonize standards for accuracy among leading regulatory agencies, and create mechanisms for evaluating and transparently sharing performance data for inexpensive or non-FDA-cleared oximeters. Such efforts will also help address barriers to adequate representation of populations with dark skin pigment in the conduct of medical research, a longstanding and complex challenge not unique to pulse oximetry. Further, we solicit the commitment of device manufacturers, regulatory bodies, and the medical community to complete this work not because it is profitable, but because it is morally just. Doing so will advance equitable healthcare across the globe.

References

1. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. *NEJM*. 2020;383(25):2477-2478.
2. Valbuena VSM, Barbaro RP, Claar D, et al. Racial Bias in Pulse Oximetry Measurement Among Patients About to Undergo ECMO in 2019-2020, A Retrospective Cohort Study. *Chest*. Published online September 27, 2021. doi:10.1016/j.chest.2021.09.025
3. Warren, E., Wyden, R, Booker, C. Letter to FDA RE: Bias in Pulse Oximetry Measurements.pdf. Accessed February 12, 2021.
<https://www.warren.senate.gov/imo/media/doc/2020.01.25%20Letter%20to%20FDA%20re%20Bias%20in%20Pulse%20Oximetry%20Measurements.pdf>
4. Bickler PE, Feiner JR, Severinghaus JW. Effects of Skin Pigmentation on Pulse Oximeter Accuracy at Low Saturation. *Anesthesiology*. 2005;102(4):715-719.
5. Batchelder PB, Raley DM. Maximizing the laboratory setting for testing devices and understanding statistical output in pulse oximetry. *Anesth Analg*. 2007;105(6 Suppl):S85-94.
6. Feiner JR, Severinghaus JW, Bickler PE. Dark Skin Decreases the Accuracy of Pulse Oximeters at Low Oxygen Saturation: The Effects of Oximeter Probe Type and Gender. *Anesth Analg*. 2007;105(6):S18.
7. Foglia EE, Whyte RK, Chaudhary A, et al. The Effect of Skin Pigmentation on the Accuracy of Pulse Oximetry in Infants with Hypoxemia. *J Pediatr*. 2017;182:375-377.e2.

8. Batchelder PB. Ear lobe spectra scans in subjects with varying skin pigment. Unpublished. 1995.
9. Biebuyck JF, Severinghaus JW, Kelleher JF. Recent Developments in Pulse Oximetry. *Anesthesiology*. 1992;76(6):1018-1038.
10. Louie A, Feiner JR, Bickler PE, Rhodes L, Bernstein M, Lucero J. Four Types of Pulse Oximeters Accurately Detect Hypoxia during Low Perfusion and Motion. *Anesthesiology*. 2018;128(3):520-530.
11. Schramm WM, Bartunek A, Gilly H. Effect of local limb temperature on pulse oximetry and the plethysmographic pulse wave. *Int J Clin Monit Comput*. 1997;14(1):17-22.
12. Food and Drug Administration. Pulse Oximeters - Premarket Notification Submissions [510(k)s] Guidance for Industry and Food and Drug Administration Staff. Accessed February 8, 2021. <https://www.fda.gov/media/72470/download>
13. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial Limitations of Fitzpatrick Skin Type. *Cutis*. 2020;105(1):77-80.
14. Baquié M, Kasraee B. Discrimination Between Cutaneous Pigmentation and Erythema: Comparison of the Skin Colorimeters Dermacatch and Mexameter. In: Humbert P, Maibach H, Fanian F, Agache P, eds. *Agache's Measuring the Skin*. Springer International Publishing; 2016:1-12.
15. Corbie-Smith G. The Continuing Legacy of the Tuskegee Syphilis Study: Considerations for Clinical Investigation. *Am J Med Sci*. 1999;317(1):5-8.

16. International Organization for Standardization. ISO 80601-2-61:2017. ISO. Accessed February 15, 2021.
<https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/79/67963.htm>
1
17. Lipnick MS, Feiner JR, Au P, Bernstein M, Bickler PE. The Accuracy of 6 Inexpensive Pulse Oximeters Not Cleared by the Food and Drug Administration: The Possible Global Public Health Implications. *Anesth Analg*. 2016;123(2):338-345.

Figure Legend

Figure 1 - Left panel shows data from recent pulse oximeter performance studies at Clinimark Laboratories and represents simultaneously collected pulse oximeter readings and co-oximeter arterial blood analysis. **Right panel** shows data from Sjoding et al.² The red rectangle is the zone of “occult hypoxia” as described in the Sjoding letter. Data from the controlled laboratory studies show no points in this area. Right panel from Reference 1, with permission.

Table 1 Current and proposed requirements for pulse oximeter performance standards

	*FDA 510k Requirements	**ISO Requirements	Proposed Changes in Requirements
Accuracy and precision	$A_{RMS} < 3\%$ for transmittance oximeters and $< 3.5\%$ for reflectance devices for all paired readings at SpO ₂ 70-100% based on the reference multiwavelength arterial blood oximeter for all test subjects pooled (composite)	$A_{RMS} < 4\%$ for all paired readings at SpO ₂ 70-100% based on the reference multiwavelength arterial blood oximeter for all test subjects pooled (composite)	$A_{RMS} < 3\%$ for all paired readings at SpO ₂ 70-100% based on the reference multiwavelength arterial blood oximeter, for transmittance and reflectance devices, for not only for all test subjects pooled (composite), but also for patients with dark skin pigment when analyzed in isolation
Extended, standardized analysis of outliers	Not required	Not required	Require sufficient information to account for potential impact of outliers on clinical performance
Number and composition of test subjects	10 or more healthy volunteers that vary in age and gender (and 200 data points) and a range of skin pigmentation, including ≥ 2 darkly pigmented subjects or 15% of study pool, whichever is larger	10 total volunteers (and 200 data points) and or 15% of study group with "darkly-pigmented skin"	Consider increasing the total number and/or diversity of study subjects to ensure adequate and balanced representation of varying skin colors (More data analyses are needed to set target numbers for this recommendation)
Definition of dark skin pigmentation	None	None	Race and ethnicity should not be used for assessing the impact of skin pigment on oximeter performance. Objective measurements of skin pigment to quantify and report of range of skin tones should be required (e.g., colorimetry) rather than relying on subjective description of "dark skin" or personal identification of race or ethnicity. Objective methods should be used to determine skin pigment at the site of oximeter measurement (e.g., ear or finger), including both surfaces of measurement for transmittance devices (e.g., dorsal and ventral).
Public and transparent reporting of validation data	None	None	Raw performance testing data with A_{RMS} calculations must be made publicly available by the device manufacturer and should include name and location of validating testing lab
Performance during low perfusion	None	None	Standardized protocols for measuring, reporting, and accounting for perfusion should be incorporated into certification standards. Performance thresholds during low perfusion should be reported by manufacturers. Low signal quality warning for low perfusion should be required for all devices.

* FDA – US Food and Drug Administration; **ISO – International Organization for Standardization

Figure 1 – Bias and outliers in oximeter performance

