Pulse Oximeter Bias and Inequities in Retrospective Studies—Now What?

Errors in pulse oximeter performance associated with dark skin pigment have been observed since the inception of the modern pulse oximeter and were first reported in the 1980s.[1] Higher rates of pulse oximeter errors in hospitalized Black patients compared with White patients were first reported in clinical studies in the 1990s[2] and in controlled healthy subject laboratory studies in the early 2000s.[3,4] The magnitude of these errors and their impacts on health and health-care delivery were not widely studied or discussed for nearly 2 more decades.

Pulse oximeter readings drive treatment decisions, including the administration of oxygen, the timing for hospitalization, and the prescription of life-saving therapies (Table 1). With increased attention to hypoxemia and health-care disparities during the severe acute respiratory syndrome coronavirus 2 pandemic, the issue of bias in oximetry is now receiving long overdue attention. More studies have been published on this topic in the past 2 years than in the preceding 2 decades.[1] Several retrospective studies highlight the correlation between dark skin and worse performance in some pulse oximeters and also link this correlation to disparities in health and health care.[1]

In this issue of RESPIRATORY CARE, Chesley et al[5] report findings aligned with the growing body of evidence that pulse oximeters do not perform equitably across race and ethnicity categories. In this large retrospective, multi-center cohort of hospitalized adults in Pennsylvania, race was identified by using chart review. Occult hypoxemia, defined as an arterial oxygen saturation (S_{aO_2}) of <88% when S_{pO_2} was between 92 and 96%, was disproportionately more common among minorities.[5] The minority comparison group included Black, Latinx, Asian, and Pacific Islander subjects compared with White subjects. Moreover, the investigators found that the discordance between S_{pO_2} and S_{aO_2} measurements widened at lower S_{pO_2} values and that the magnitude of bias changed over time among the subjects with serial samples.[5]

Inconsistent bias is to be expected in the clinical setting in which dynamic variables impact oximeter performance (eg, anemia, motion, tissue edema, and peripheral perfusion). These findings support and complement data presented in recent publications.[6,7] The investigators call for improved pulse oximeter technology and offer 3 recommendations in response to these data. First, the investigators discourage clinicians from relying on a single paired sample of S_{pO_2} and arterial blood gas analysis performed early in a patient's hospitalization to assess the magnitude of bias in pulse oximetry for an individual. Second, the investigators note limited evidence to support a more resource intensive strategy of frequently sampling arterial blood gas S_{aO_2}. Third, they suggest targeting a higher S_{pO_2} goal, of 94–98%,

SEE THE ORIGINAL STUDY ON PAGE 1499

to decrease the prevalence of hypoxemia in patients with darker skin pigment.

On the Topic of Pigment

Chesley et al[5] state that future studies and regulatory bodies should explicitly focus on patients' race and ethnicity to mitigate disparities. The social construct of race certainly impacts health, but skin pigment and race should not be conflated. Future work in pulse oximetry must move toward characterizing skin pigment rather than using self-reported or chart-documented race. Race is not monolithic and within any given race and ethnicity, individuals have a range of skin tones. In a 2022 meta-analysis of 382 articles, only 15 articles assessed the impact of skin pigmentation on pulse oximeter accuracy.[8] The qualitative categories were not specific: light, medium, and dark.[8]

Colorimeters and spectrophotometers, which measure erythema and melanin indices in addition to other relevant pigment parameters, are highly useful in parsing the mechanisms that drive the observed associations between skin pigment and pulse oximeter errors. Most importantly, these quantitative data can inform efforts to improve pulse oximeter technology and hone regulatory approaches to address the observed biases. On a global scale, quantifying pigment rather than using mutable and culturally informed race categories will allow for interventions that more effectively remediate the bias in pulse oximetry. In particular, we at the University of California San Francisco Hypoxia Research

The authors have disclosed no conflicts of interest.

Correspondence: Kelvin L Moore Jr BA, University of California San Francisco, School of Medicine, 533 Parnassus Ave, Box 0846, Floor 01, Room 001 San Francisco, CA 94143. E-mail: kelvin.moorejr@ucsf.edu.

DOI: 10.4187/respcare.10654

Table 1. Conditions with Therapy Determined by S_{pO_2} Targets and Potential Susceptibility to Disparities

Condition	S_{pO_2} Target, %	Therapy
COVID–19 pneumonia	< 90	Corticosteroids
Community-acquired pneumonia	≤ 90	Hospital admission and titration of supplemental oxygen, antibiotic and/or antiviral administration
ARDS	< 88	Lung-protective ventilation titration with ARDSNet protocol[14]
Pneumocystis jirovecii pneumonia	< 92	Corticosteroids
Interstitial lung disease	≤ 91	Corticosteroids; long-term oxygen therapy
Pulmonary hypertension	< 90	Home oxygen
Chronic lung disease	< 88	Home oxygen
Sleep-related breathing disorders	< 88	Home oxygen in addition to positive airway pressure therapy

COVID-19 = coronavirus disease 2019

Laboratory strongly encourage investigators to characterize skin pigment at the site of oximeter measurement, including nail pigment, because results of studies have shown that individuals with darker skin are more likely to experience melanonychia or benign cases of fingernail hyperpigmentation.[9,10]

Mitigation Strategies for Pulse Oximetry Bias

We agree with the recommendations by Chesley et al[5] to avoid assessing oxygenation status based on a single arterial blood gas performed early in a patient's hospitalization. Moreover, we caution against guidelines and critical clinical decision-making on the basis of an absolute S_{pO_2} cutoff given the observed biases in pulse oximeters. There currently exist numerous therapies whose administration is solely determined by S_{pO_2} cutoffs (Table 1). These cutoffs often dictate whether a patient is admitted and/or is administered corticosteroids and supplemental oxygen. Using a rigid S_{pO_2} cutoff may have significant clinical implications for individuals on which pulse oximeters overestimate their oxygenation status. We argue that the studies that inform these clinical practices and guidelines ought to be re-evaluated or repeated with strict attention to methods ensuring guidelines serve people equitably across a spectrum of skin pigments. We encourage clinicians to be flexible with S_{pO_2} cutoffs and use them in the context of patient presentation until the issue of pulse oximeter bias is resolved. We also caution against frequent arterial blood gas sampling because the clinical benefits are unclear and frequent sampling imposes additional (and potentially unneeded) burdens on clinicians, health systems, and patients alike. Furthermore, this is not an achievable solution in resource-limited settings.

The authors' argument for increasing pulse oximetry targets from 92–96% to 94–98% may be a clinically pragmatic short-term solution for some settings. However, purposefully overestimating oxygen saturations to account for hypoxemia fails to address the fundamental errors within pulse oximeters and can have unintended consequences. First, a change in the S_{pO_2} target is not without a pronounced impact on oxygen consumption, which can be a major challenge in many resource-limited settings. Adjusting the standard of care workflow to accommodate unpredictably flawed technology perpetuates the inequities in the system that gave rise to this problem.

Universally higher oxygen saturation targets might also prolong hospitalizations or create unnecessary demands on out-patient infrastructures. We suggest that the highest priority for investigators and regulators begins with rigorous, standardized prospective studies. Data collection protocols must facilitate performance improvement interventions for next-generation pulse oximeters. Although the suggestion by Chesley et al[5] to increase pulse oximetry targets is intended as a stopgap, clinicians and regulators must not lose sight of the overarching systematic technological flaws. If existing technology underperforms, then new device technology development is indicated.

Despite the issues seemingly inherent in the development of pulse oximeters, we want to underline that clinicians should not abandon oximeter measurements because they are an immensely useful tool. Instead, we advise that clinicians consider the role of measuring S_{aO_2}, when available, if a critical medical decision is being made based on a patient's oxygenation status (Table 1). We urge cautious consideration of adjudicating resources and therapies in cases in which an S_{pO_2} error may drive health-care disparities. We also urge clinicians to better understand the limitations of pulse oximeters as they use them and lend their voices to organized calls for improved technology.

A Call for More Collaborative Research and Data Harmonization

The phenomenon of data harmonization has been carried out in multiple public health disciplines, namely, pharmacoepidemiology as well as social and behavioral sciences.[11-13] The purpose of this harmonization is 3-fold: (1) investigators can streamline data and research dissemination to global networks, (2) pragmatic datasets can be used in tandem with more comprehensive datasets to standardize analysis, and (3) study procedures are more widely shared through an open-source network.

Data harmonization in which key study parameters and procedural methods are clearly defined allows investigators

to collect standardized data that is clinically meaningful. For example, all future prospective studies that observe skin pigment bias in pulse oximetry should operate with the aim to simultaneously collect S_{pO_2} and S_{aO_2} measurements, quantify skin color, measure other important variables (eg, perfusion), and consistently define hypoxemia. The standardization of data collection and analysis would allow the global research community to move beyond observations of bias and to begin to make progress on identifying root causes and potential solutions.

The Open Oximetry Project (www.OpenOximetry.org, San Francisco, CA) is developing and sharing comprehensive data collection methodologies that aim to harmonize and accelerate research to improve access to safe pulse oximetry. Limited dissemination of study procedures and obscure assessments of end points stifle the timely translation of research data to evidence-based clinical decision making tools on a global scale. An open forum invites experts and stakeholders to improve a protocol and to efficiently adapt it to the current state of knowledge and technology. It is our hope that additional disciplines adopt this framework because health disparities are not an issue unique to pulse oximetry.

Conclusions

Pulse oximetry remains one of the most ubiquitous and useful patient safety tools worldwide, even with its current limitations. More prospective studies are needed to collect more meaningful data (eg, characterize skin pigment rather than race, document confounders such as perfusion and oximeter probe positioning) to further describe the biases observed and identify potential solutions. Data harmonization and collaborative research efforts between investigators may be excellent tools to address the global implications of pulse oximeter bias. In addition, these efforts will prevent investigators from perpetuating mistakes in data collection.

Finally, sweeping changes that increase S_{pO_2} targets to account for systematic errors within pulse oximeters should be made with extreme caution and for consideration of clinical setting and the type of pulse oximeter used. It is imperative that clinicians and investigators understand that increasing S_{pO_2} targets in lieu of addressing the structural racism baked into pulse oximetry technology excuses the faulty technology and may have unintended clinical consequences. Additionally, the use of absolute S_{pO_2}cutoff for therapy decisions requires great caution and reconsideration. The fact that bias in oximeter performance has been known for 40 years but relatively little has been done to investigate the potential health and health-care disparities is alarming. Attention to diversity, equity, and inclusion in all aspects of biomedical engineering research and regulatory science, from research staff to subject recruitment, is essential to solving the issues manifested in pulse oximeters and likely other technologies as well.

Kelvin L Moore Jr
University of California San Francisco School of Medicine, San Francisco, California

Hypoxia Research Laboratory
University of California San Francisco
San Francisco, California

Center for Health Equity in Surgery and Anesthesia
University of California San Francisco
San Francisco, California

Koa Gudelunas
Hypoxia Research Laboratory
University of California San Francisco
San Francisco, California

Michael S Lipnick
Philip E Bickler
Hypoxia Research Laboratory
University of California San Francisco
San Francisco, California

Center for Health Equity in Surgery and Anesthesia
University of California San Francisco
San Francisco, California

Anesthesia and Perioperative Care
University of California San Francisco
San Francisco, California

Carolyn M Hendrickson
Hypoxia Research Laboratory
University of California San Francisco
San Francisco, California

Pulmonary Critical Care
Zuckerberg San Francisco
General Hospital and Trauma Center
San Francisco, California

REFERENCES

1. Openoximetry.org. Publications. OpenOximetry. https://openoximetry.org/publications/. *Accessed October 12, 2022*.
2. Jubran A, Tobin MJ. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. Chest 1990;97(6):1420-1425.
3. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology 2005; 102(4):715-719.
4. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg 2007;105(6 Suppl):S18-S23.
5. Chesley CF, Lane-Fall MB, Panchanadam V, Harhay MO, Wani AA, Mikkelsen ME, et al. Racial disparities in occult hypoxemia and clinically based mitigation strategies to apply in advance of technological advancements. Respir Care 2022;67(12):1499-1507.
6. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and ethnic discrepancy in pulse oximetry and delayed

identification of treatment eligibility among patients with COVID-19. JAMA Intern Med 2022;182(7):730-738.
7. Valbuena VSM, Seelye S, Sjoding MW, Valley TS, Dickson RP, Gay SE, 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 2022;378:e069775.
8. Shi C, Goodall M, Dumville J, Hill J, Norman G, Hamer O, et al. The accuracy of pulse oximetry in measuring oxygen saturation by levels of skin pigmentation: a systematic review and meta-analysis. BMC Med 2022;20(1):267.
9. André J, Lateur N. Pigmented nail disorders. Dermatol Clin 2006;24 (3):329-339.
10. Jefferson J, Rich P. Melanonychia. Dermatol Res Pract 2012;2012: 952186.
11. IPUMS. IPUMS INTERNATIONAL: Harmonized International Census Data for Social Science and Health Research. Available at: https://www.ipums.org/projects/ipums-international. *Accessed October 12, 2022.*
12. Wang SV, Pottegård A, Crown W, et al. HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR Task Force. Pharmacoepidemiol Drug 2022;1:22.
13. Doiron D, Marcon Y, Fortier I, Burton P, Ferretti V. Software application profile: Opal and Mica—open-source software solutions for epidemiological data management, harmonization and dissemination. Int J Epidemiol 2017;46(5):1372-1378.
14. Brower RG, Matthay MA, Morris A, et al; Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000;342(18):1301-1308.