Title: Absence of Racial Bias In Pulse Oximetry Saturation Measurement

References

1. NEJM, 383(25), 2477–2478, 2020
2. Anaesthesia;77(2):143-152, 2021
3. The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities. Princeton: Princeton University Press, 2003

Authors: Neal Fleming, MD, PhD; Audrey Marlar, MS2; Bradley Knabe, MS2, Yasamin Taghikhan, MD; Fatima Yusuf, BA, Richard Applegate, MD

Affiliations: Dept. of Anesthesiology & Pain Medicine, UC Davis School of Medicine

Introduction

Pulse oximetry is a critical monitor for assessing oxygenation status and guiding oxygen therapy. A retrospective comparison of arterial oxygen saturation (SaO_2) with peripheral pulse oximetry (SpO_2) found a nearly 3-fold increase in in the incidence of hypoxemia undetected by pulse oximetry in Black patients compared to White patients.[1] This report triggered an FDA Safety Communication emphasizing the limitations and potential inaccuracies of pulse oximetry in certain situations including home monitoring of patients with COVID-19. This possible racial bias in measurement could contribute to inequities in healthcare. However, race is not binary. Furthermore, the variable of interest is actually skin color. Race does not uniquely characterize skin color. There is a wide range of graded colors within any group of individuals with a common racial identity. A subsequent study focused on SpO_2 measurements in ethnically diverse patients hospitalized with COVID-10 pneumonitis failed to demonstrate any effect of ethnicity on SpO_2 accuracy.[2] Our study reported here retrospectively examined SpO_2 accuracy using data from unrelated clinical trials that included a standardized, graded assessment of skin color in addition to racial identity.

Methods

Following Human Subjects Research Committee approval, electronic medical records and previous clinical trial data were used to collect basic demographic information, including reported race and ethnicity as well as skin color assessed using the NIS Massey and Martin Skin Color Scale.[3] This scale ranges from 1 (very light skin) to 10 (very dark skin) and was recorded at the time of enrollment by trained personnel who referred to a standardized color image of the scale values. Arterial blood gas (ABG) PaO_2 and SaO_2 values and the corresponding SpO_2 values were extracted from clinical laboratory reports and vital signs flowsheets. For all PaO_2 values ≤125 mmHg, the corresponding SaO_2 and SpO_2 values were compared. Differences were grouped by the patient's Massey score. Group comparisons were performed with the Kruskal-Wallis test using GraphPad Prism version 9.2.0 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com.

Results

Massey skin color assessments were available from 742 patients. Within this group, patients who identified as Black had Massey scale vales ranging from 5 to 9 and those who identified as White had values ranging from 1 to 5. For each patient, all ABG values for the initial and any subsequent hospitalizations were extracted from the medical records. 579 patients had ABG PaO_2 values ≤125 mmHg. A total of 4,030 individual comparisons were available for analysis. The number of measurements at each Massey scale rating is summarized in Table 1. The average errors for each group are presented as a violin plot in Figure 1. There was a statistically significant difference among the groups ($p<0.0001$).

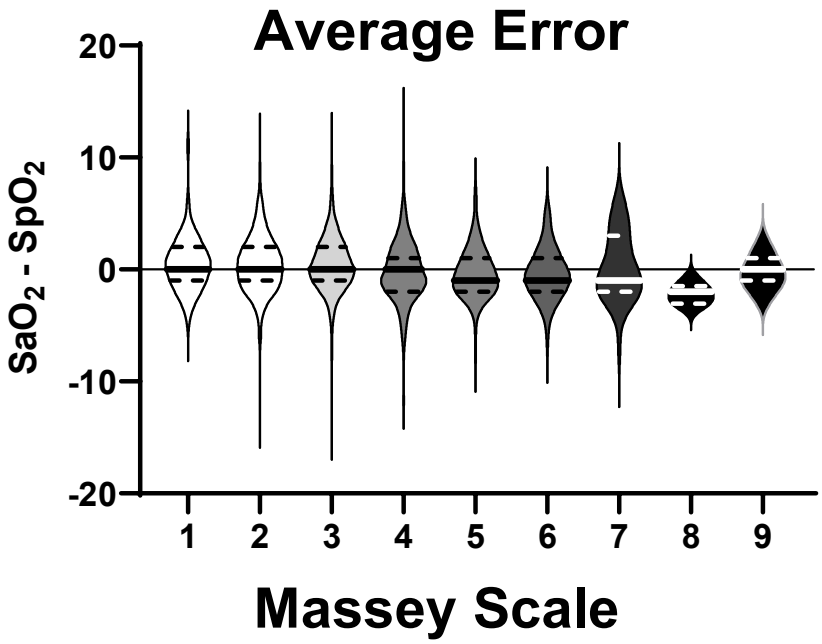
Conclusions

The potential for a clinically significant influence of skin tone on the accuracy of SpO_2 measurements merits careful evaluation. These differences are better characterized using a standardized graded scale of skin color rather than patient identified race. This review failed to demonstrate any bias due to race or skin color in this measurement. Despite statistical significance, the observed errors are all within the stated and expected accuracy of the SpO_2 monitor. Two limitations must be highlighted. The distribution of Massey scores is skewed to lower values. In this data set, the statistical differences are driven by the values for the Massey scale 8 which includes only 5 measurements from 3 patients. Second, the retrospective design in this review limits the confidence in the SpO_2 measurements, especially at lower PaO_2 values. These frequently occur in clinically dynamic periods that may not be as accurately captured by the lower time resolution data stored in the available vital signs flowsheets. These results do, however, demonstrate the need, and provide a template for subsequent prospective evaluations of skin color as a confounding influence in SpO_2 measurement.

Table 1.

Massey Score	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Number of patients	27	221	196	80	30	14	7	3	1
Number of values	123	1471	1281	764	184	137	38	5	27
Median Error	0.00	0.00	0.00	0.00	-1.00	-1.00	-1.00	-2.00	0.00
Mean Error	0.53	0.46	0.38	-0.2	-0.35	-0.39	0.21	-2.2	-0.3
Std. Deviation	2.5	2.6	2.6	2.9	2.3	2.3	3	0.86	1.6

Figure 1.


Average Error
20
10
0
-10
-20
SaO2 - SpO2
1
2
3
4
5
6
7
8
9
Massey Scale