# FDA Executive Summary

# Review of Pulse Oximeters and Factors that can Impact their Accuracy

Prepared for the
November 1, 2022, meeting of the
Anesthesiology Devices Advisory Committee
Center for Devices and Radiological Health (CDRH)
United States Food and Drug Administration

## Contents

# I. Introduction and Purpose of the Advisory Committee Meeting

Pulse oximeters are widely used by many types of healthcare providers and consumers to obtain an indirect measure ($SpO_2$) of arterial blood oxygen saturation ($SaO_2$). $SaO_2$ measurement is considered the gold standard for assessment of blood oxygen saturation levels. It involves an invasive procedure, requiring an arterial puncture (blood draw) to directly measure the blood oxygen saturation. $SpO_2$ is an estimate of how much oxygen the hemoglobin contains compared to how much it could contain, expressed as a percentage. Pulse oximetry is a non-invasive and quick alternative to arterial puncture for estimating oxygen saturation. Pulse oximeters can be categorized as:

- Prescription Use Pulse Oximeters: Regulated under product codes DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter), are reviewed by the FDA, receive 510(k) [i] clearance, and are available only with a prescription. The FDA requires that these pulse oximeters undergo clinical testing to confirm their accuracy. These are used to monitor (i.e., trending or spot checking) oxygen saturation levels of patients, most often in hospitals and doctors' offices, although they may sometimes be prescribed for home use. **These medical devices are the focus of the November 1, 2022 Anesthesiology Panel meeting.**

- Over-the-Counter (OTC) Pulse Oximeters: Regulated under product codes PGJ and OCH. Most commonly these products are intended for general wellness[ii] or sporting/aviation uses and not intended for medical purposes. Such products do not generally undergo FDA premarket review. They are often sold directly to consumers in stores or online and may utilize mobile medical apps intended for estimating oxygen saturation for non-medical purposes. See FDA's guidance document *General Wellness: Policy for Low-Risk Devices* for additional information. **Although the use of pulse oximeters for general wellness or sports/aviation is not the focus of the panel meeting, the Panel will be asked to comment on the potential inaccuracies due to skin pigmentation for OTC pulse oximeters when consumers use them for medical purposes.**

As part of its premarket evaluation of pulse oximeters, FDA has long required premarket data assessing pulse oximeter performance in subjects with different skin pigmentation. However, a December 2020 published report suggests that pulse oximeters may be less accurate in patients with darker skin pigmentation.[1] To further assess this issue, the Agency reviewed the published literature, Medical Device Reporting (MDR) data, and clinical evidence from desaturation studies on the accuracy of pulse oximeters[iii]. In February of 2021, the FDA issued a *Safety Communication* to inform patients and health

[i] https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-notification-510k

[ii] See FDA's guidance document *General Wellness: Policy for Low-Risk Devices*, available at https://www.fda.gov/media/90652/download

[iii] FDA had access to pre-print in 2021, Okunklola et.al. *Respiratory Care* 2022, 67 (2)

care providers that although pulse oximeters are useful for estimating blood oxygen levels, these devices have limitations and a risk of inaccuracy under certain circumstances that should be considered.[iv] The safety communication provided recommendations for patients using pulse oximeters at home regarding how to take and interpret an $SpO_2$ reading as well as when to contact a health care provider. It also provided recommendations to health care providers to be aware of factors that can affect the accuracy of pulse oximeter reading, to refer to device labeling to understand the accuracy of the specific devices and sensors they are using, and to consider accuracy limitations when using these devices to assist in diagnosis and treatment decisions. During the year 2021, the FDA embarked on several initiatives to address the concerns related to pulse oximeters inaccuracies, as described below.

An assessment of the U.S. pulse oximeter market demonstrates the use of these devices increased during the first year of the COVID-19 pandemic.[v] In addition, the Agency completed an evaluation of pulse oximeter 510(k) submissions (under product code DQA) that have been cleared for marketing since 2000, to better understand the assessment and reporting of skin pigmentation in premarket desaturation studies. The evaluation also included a review of the labeling for the pulse oximeters captured in the analysis. The assessment shows that after the FDA guidance for prescription use pulse oximeters[vi] was issued in 2013, there was an increase in the submission of line-level data, use of plots describing accuracy, submission of skin pigmentation data; and there was use of different skin color scales and categorizations by different manufacturers. For details on this study methods, results, and discussion, see Section VI. Given the importance of objective assessment of skin pigmentation in evaluating its impact on pulse oximeter accuracy, a review of published literature on objective approaches to assess skin pigmentation was completed, and a summary is presented in Section X.

Since the December 2020 publication[1], additional real-world studies have been published suggesting increased risk for missed diagnosis of hypoxemia[vii] (i.e., *"occult hypoxemia"*), delays in treatment eligibility decisions and worse patient outcomes among subjects with darker skin pigmentation.[2-9] Recently, FDA completed a systematic literature review of the real-world performance of pulse oximeters. See Section VIII for details on the systematic literature review methods, summary of findings, and additional details on the evidence assessment.

[iv] https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication

[v] https://www.businesswire.com/news/home/20200827005591/en/Global-Pulse-Oximeter-Market---Outlook-and-Forecast-to-2025---ResearchAndMarkets.com

[vi] https://www.fda.gov/media/72470/download

[vii] An abnormal low concentration of oxygen in the blood

FDA is holding a public meeting of the Anesthesiology Panel of the Medical Devices Advisory Committee on November 1st, 2022. The Panel will be asked to discuss the concerns related to pulse oximeters being less accurate in individuals with darker skin pigmentations. Specifically, the Panel will discuss the available real-world evidence on accuracy of pulse oximeters, factors that may affect pulse oximeters accuracy, and will make recommendations for health care providers, patients labeling, and study design and analyses. The Panel meeting will allow an opportunity for patients, patients' organizations, professional organizations, academia, and industry to share their perspectives on complex issues involving the regulation and accurate performance of pulse oximeter devices.

## II. FDA Regulation of Pulse Oximeters

Prescription use pulse oximeters are Class II devices[viii] intended to measure blood oxygen saturation levels and are regulated under:

- **21 CFR 870.2700:** Oximeter (product codes: DQA and NLF). An oximeter is a device used to transmit radiation at a known wavelength(s) through blood and to measure the blood oxygen saturation based on the amount of reflected or scattered radiation. It may be used alone or in conjunction with a fiberoptic oximeter catheter.

- **21 CFR 870.2710:** Ear Oximeter, product code DPZ. An ear oximeter is an extravascular device used to transmit light at a known wavelength(s) through blood in the ear. The amount of reflected or scattered light as indicated by this device is used to measure the blood oxygen saturation level.

A manufacturer that intends to market a new pulse oximeter for medical purposes in the U.S. must submit a premarket submission for market authorization.[ix,x] Per 21 CFR 807.87(g), a pulse oximeter that has undergone a significant change or modification from its currently cleared configuration that can significantly affect the safety or effectiveness of the device, requires submission of a new 510(k). Although the review standard for 510(k) submissions is comparative, the principles of safety and effectiveness underlie the substantial equivalence determination. The standard for a determination of substantial equivalence in a 510(k) review is set out in section 513(i) of the FD&C Act.

Currently, all prescription use pulse oximeters cleared for market distribution in the U.S. are labeled with a general indication for non-invasive measurement of blood oxygen saturation (i.e., continuous

[viii] https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device
[ix] https://www.fda.gov/media/82395/download
[x] https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/de-novo-classification-request

monitoring or spot checking for trending). To date, FDA has not cleared any OTC pulse oximeters for medical purposes.

# III. Pulse Oximetry Technology

## A. Current Technology

Pulse oximetry is based on two physical principles: the presence of a pulsatile signal generated by arterial blood, and the fact that oxygenated hemoglobin ($HbO_2$) and reduced hemoglobin (HHb) have different absorption spectra.[10] The optical techniques that have been developed for the assessment of $SaO_2$ are based on these different light-absorption spectra. In the red and infrared wavelength regions (typically 660 nm and 940 nm, respectively), the absorption is relatively low, allowing for measurement of light transmission.[11] $HbO_2$ absorbs more infrared light and allows more red light to pass through; whereas, HHb absorbs more red light and allows more infrared light to pass through. The ratio of the red to infrared light measurement is calculated by a processor and converted to $SpO_2$ (expressed as a percentage). An important component of pulse oximetry is photoplethysmography (PPG), which measures changes in light absorption due to variations in arterial blood volume. [11]

The regulations 21 CFR 870.2700 and 870.2710 include devices using reflectance, transmittance, and fiber optic technologies. Prescription use pulse oximeters measure the amount of transmitted, reflected, and scattered light through various application sites (e.g., finger, ear, foot, hand, forehead, back, and nose).

Some of the factors that can impact the accuracy of pulse oximeters include (but not limited to):

- Skin pigmentation
- Dyshemoglobinemias: disorders in which the hemoglobin molecule is functionally altered and prevented from carrying oxygen.
- Severe anemia: disorder in which the blood has reduced ability to carry oxygen. Anemia occurs when there are not enough healthy red blood cells to carry oxygen to the body organs.
- Low perfusion: reduced peripheral blood flow and subsequent reduction in the detectable signal at pulse oximeter sensor site.
- Dyes
- Nail polish
- Ambient light
- Motion artifacts

The November 1st Anesthesiology Advisory Panel will discuss the impact of skin pigmentation on the accuracy of pulse oximeters.

### B. Developments in Pulse Oximetry Technology

While currently cleared 510(k) pulse oximeters are limited to the technology described above, there is published literature describing technology developments, including lasers or other narrow-bandwidth light sources, which are being developed to improve the accuracy of pulse oximeters. While any differences in device performance correlated with race/ethnicity could potentially be caused by a variety of factors, skin pigmentation is typically considered the most likely cause for optical medical devices. This is due to: (a) the strong, spectrally-varying optical absorption of epidermal melanin [12], and (b) evidence that variations in skin pigmentation have a significant impact on visible to near-infrared reflectance signals, and thus impact the performance of optical devices (e.g., regional oximeters, bilirubinometers). [13] There is also evidence that the optical properties of the sensor and/or bandage surface may impact oximeter accuracy [14,15] thus, for example, avoiding sensors with highly reflective or colored bandages may reduce sensitivity to pigmentation. Finally, more advanced optical approaches, such as frequency domain near-infrared spectroscopy [16] and photoacoustic imaging [17] have also been used to measure arterial oxygen saturation and may enable greater accuracy. However, these technologies are not inherently immune to melanin optical absorption and, like reflectance-based approaches, may benefit from melanin correction algorithms based on novel methods such as deep learning.

## IV. FDA Guidance on Pulse Oximeters

### A. Premarket Evaluation of Pulse Oximeters

The "*Pulse Oximeters- Premarket Notification Submissions [510(k)s]: Guidance for Industry and FDA Staff*" [xi] was issued in final version March 4, 2013. The scope of the guidance document is limited to Class II pulse oximeters regulated under 21 CFR 870.2700 and 21 CFR 870.2710.

The FDA pulse oximeter guidance recommends that desaturation studies include 10 or more healthy subjects that vary in age and gender, include 200 or more data points (i.e., paired observations of $SpO_2$-$SaO_2$), and for the study subjects to have a range of skin pigmentation, including at least 2 darkly pigmented subjects or 15% of the study group, whichever is larger. The recommendation for distribution of darkly pigmented subjects was based on census data for distribution of race in the U.S. and not based on the distribution of skin tone of the general U.S. population.



[xi] https://www.fda.gov/media/72470/download

The FDA guidance also includes recommendations for testing pulse oximeter's accuracy under motion conditions and low perfusion. For motion performance claims, it is recommended to describe the characteristics of each motion, e.g., amplitude, type, and frequency of motion. For low perfusion performance claims, it recommends a functional tester, set to the signal amplitude defined as low perfusion for the system.

Other topics covered in the guidance include recommendations for evaluation of accuracy in neonates, alarm testing, display values, outputs and indicators, saturation pulse signal, software, electrical, mechanical, and environmental safety, electromagnetic compatibility, biocompatibility, cleaning, disinfection, and sterilization.

### B. Statistical Analyses Considerations

The guidance recommends 510(k) submissions include line listing data, Bland-Altman plots, error plots for both individual subjects and pooled subjects, and rationale for any data points excluded from the analysis. Summary statistics are recommended such as population mean bias ($\mu_0$), between-subject variance ($\sigma_\mu^2$ ), within-subject variance ($\sigma^2$ ), upper 95% and lower 95% limits of agreement, and the root mean square (Arms). The guidance recommends acceptable Arms by type of sensor, as presented below:

*Table IV-1 Acceptable Arms by Sensor Type*

| Sensor Type | Typical Arms |
|---|---|
| Transmittance, wrap and clip | $\leq$ 3.0% |
| Ear clip | $\leq$ 3.5% |
| Reflectance | $\leq$ 3.5% |

In premarket desaturation studies, the primary performance metric is *Arms*, the root mean square of the difference between simultaneous paired measurements of $SpO_2$ and $SaO_2$ pooled across all measurements from all subjects. If *Arms* = 3%, then the probability that an $SpO_2$ value is within 3% of the $SaO_2$ value is roughly 68%. The Bland-Altman (BA) scatterplots [18] of the difference $SpO_2 - SaO_2$ vs. the mean $(SpO_2 + SaO_2) / 2$ and error plots are useful for examining if the location or spread of the differences exhibits a pattern across the mean. Patterns may vary across individual subjects. BA plots include horizontal lines for the mean difference across all paired observations, called the mean bias, and the 95% limits of agreement (LoA), which is the mean bias plus or minus twice the standard deviation of the differences. If the differences are normally distributed with constant mean and variance, then 95% of individual differences are expected to be in the 95% LoA. A variant of the BA plot that is often preferred is the modified BA plot of SpO2 – $SaO_2$ vs. $SaO_2$, although in this plot differences may tend to

decrease with increasing $SaO_2$ due to regression to the mean. [19-24] Intrasubject correlation among repeated pairs $SaO_2$ and $SpO_2$ measurements leads to larger standard errors and wider confidence intervals on performance metrics.[19-24]

Intrasubject correlation among repeated pairs $SaO_2$ and $SpO_2$ measurements leads to larger standard errors and wider confidence intervals on performance metrics (e.g., Arms, mean bias, LoA) than if the paired measurements were independent (e.g., if each came from a different subject).[25,26]

There are other summary measures for pulse oximeter accuracy such as mean bias (average difference or deviation between $SpO_2$ and $SaO_2$ across all $SaO_2$ values), mean absolute deviation (MAD) of the differences, total deviation index (a specified quantile of the absolute deviation) and coverage probability (probability that an absolute deviation does not exceed a pre-specified acceptable deviation).[27] Some studies evaluate coverage probability, calling it the acceptable agreement rate.[27-31] Some studies evaluate coverage probability, calling it the acceptable agreement rate.[28-31]

BA plots may reveal that the $SpO_2$ - $SaO_2$ differences tend to decrease with increasing $SaO_2$. In that case, the mean difference (i.e., mean bias) across all $SaO_2$ values will depend on the distribution of observed $SaO_2$ values. A difference in mean bias between two groups will be difficult to interpret when the groups differ in $SaO_2$ distribution, which can happen in the real-world because the $SaO_2$ distribution is uncontrolled, unlike desaturation studies, and may be affected by demographic factors, physiological factors, and disease characteristics associated with $SaO_2$ that may be distributed differently between groups. For the same reason, real-world differences between groups in other measures that summarize across $SaO_2$ values are difficult to interpret.

A quantile-quantile scatterplot of $SpO_2$ vs. $SaO_2$ can be useful for visualizing the bias of $SpO_2$ for $SaO_2$ by examining departure of the points from the 45-degree line.[xii]

Many studies report a box plot of the $SaO_2$ mean and interquartile range by $SpO_2$ level for each skin color group.[5,9,32,33] However, if lower $SaO_2$ values tend to occur more frequently in group 1 than group 2, then for any given $SpO_2$ value the $SaO_2$ mean and interquartile range (i.e., the box) will tend to be lower for group 1 than group 2 even when $SpO_2$ has the same accuracy in both groups for every $SaO_2$.

A linear regression of $SpO_2$ on $SaO_2$ can be used to model the bias of $SpO_2$ for a particular value of $SaO_2$, provided that the model fits the data well. For example, $SpO_2$ bias may be evaluated at $SaO_2$ = 88% or

[xii] National Institute of Standards and Technology (NIST). *Engineering Statistics Handbook.* https://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm

another threshold below which patients are defined to have hypoxemia. If the regression line is $SpO_2$ = a + b*$SaO_2$ + error, then the $SpO_2$ bias at $SaO_2$ = x is bias(x) = a + (b – 1)x [CLSI, 2013][xiii]. Note that if a = 0 and b = 1, bias(x) = 0 for any x. Suppose for two groups of subjects g = 1,2, e.g., light, and darker skin groups, the regression lines are $SpO_2 = a_g + b_g*SaO_2$ + error. Then the difference in $SpO_2$ bias between the groups at $SaO_2$ = x is $bias_2(x) – bias_1(x) = a_2 - a_1 + (b_2 – b_1)x$.

Finally, linear regression of $SpO_2$ on $SaO_2$ can be adjusted for covariates such as skin color, age, sex, body mass index (BMI), and finger size (for finger clips). In real-world studies, the covariates can include clinical factors, medical conditions, and treatments, e.g., APACE score, cardiovascular SOFA score, vasoactive infusion score (VIS), diabetes mellitus, carboxyhemoglobin or methaemoglobin level, smoking status, body temperature, on vasopressors on inotropes, capillary refill, and local factor interference. If skin color is a categorical factor, then the multivariable regression model is an example of analysis of covariance (ANCOVA), with skin color effects on $SpO_2$ adjusted for $SaO_2$ and the other variables. The skin color effects represent differences between the skin color groups in $SpO_2$ bias. Interaction of skin color with $SaO_2$ can be modeled with a separate coefficient (slope) on $SaO_2$ for each skin color level.

### C. Pulse Oximeters Premarket Evaluation

Desaturation studies are usually conducted on a small sample of healthy volunteers and pose an acceptable risk to healthy adult study participants even though their fraction of inspired oxygen concentration [$FiO_2$] is decreased to low levels and $SaO_2$ measurements are obtained from a blood sample taken with an indwelling arterial catheter. The $FiO_2$ delivered to test subjects is varied to achieve a series of targeted steady-state saturation periods detected by a reference pulse oximeter. Arterial blood samples are periodically taken from an indwelling arterial catheter for use in the comparison. Multiple simultaneous pairs of $SpO_2$ and $SaO_2$ observations are taken per subject over a specified range (e.g., 70 % to 100 % $SaO_2$). Motion testing is required for all motion performance claims. Bench testing is required for low perfusion claims and recommended under motion and low perfusion conditions.

After receiving FDA clearance based on verification of device performance in healthy subjects, pulse oximeters are used on hospitalized as well as outpatient adults, and pediatric populations including neonates for a variety of endpoints such as but not limited to triage, and initiation, escalation, or weaning of therapy. Oximeters are typically used to maintain target $SaO_2$ ranges on critically ill populations experiencing conditions such as sepsis, cardiac arrest, and respiratory failure. Though

[xiii] Clinical and Laboratory Standards Institute. *EP09-A3: Measurement Procedure Comparison and Bias Estimation Using Patient Samples, 3rd Edition.* Wayne, PA: 2013. Accessed at https://clsi.org/standards/products/method-evaluation/documents/ep09/

additional convenience arterial sampling, clustered around 90%, is recommended for neonatal populations, this is currently not required for other populations. Extrapolation of device performance verified on healthy subjects under controlled study conditions to critically ill populations who are often on vasoactive medications remains a challenge. While in clinical practice pulse oximeters are used to detect hypoxemia, none of the legally marketed pulse oximeters in the U.S. currently have a cleared indication for diagnosis of hypoxemia; rather, all of them are cleared with the general indication as a tool for non-invasive measurement of blood oxygen saturation.

*The Panel will be asked to discuss and make recommendations about how accurate pulse oximeters should be, pulse oximetry performance across sub-groups of subjects with different skin pigmentation, and measures for pulse oximeters accuracy.*

## V. Recognized International Standard

Device Standard for the Pulse Oximeter: Tool for Assuring Safety

FDA recognizes ISO 80601-2-61 Second Edition 2017-12 (Corrected version 20180-02)[xiv,xv] ; and generally, pulse oximeter 510(k) submissions reference the standard. It is a joint standard[xvi] between ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission).[15] It applies to the basic safety and essential performance of pulse oximeter equipment intended for use on humans.[15] Basic safety[xvii] is protection from physical hazards (e.g., shock, burn, crushing), while essential performance[xviii] is the performance of a clinical function that must be maintained in the presence of a disturbance, the loss of which is determined to be unacceptable by the manufacturer (e.g., $SpO_2$ and pulse rate accuracy).

The standard covers pulse oximeter equipment intended to estimate arterial oxygen hemoglobin functional saturation and pulse rate for patients in professional healthcare institutions, as well as in the

[xiv] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/detail.cfm?standard__identification_no=37508
[xv] ISO 80601-2-61 Medical Electrical Equipment- part 2-61: Particular Requirements for Basic Safety and Essential Performance of Pulse Oximeter Equipment (2nd ed., 2017)
[xvi] Jointly developed by ISO TC 121/SC3 JWG10 (lead), and IEC TC62/SC 62D JWG5, https://www.iso.org/obp/ui/#iso:std:iso:80601:-2-61:ed-2:v2:en
[xvii] From IEC 60601-1 defined as freedom from unacceptable RISK directly caused by physical hazards when Medical Electrical (ME) equipment is used under normal condition and single fault condition. See AAMI CR500 for a more complete explanation. AAMI CR500:2019 - Basic Introduction to the IEC 60601 Series (ansi.org)
[xviii] From IEC 60601-1 defined as performance of a clinical function, other than that related to basic safety, where loss or degradation beyond the limits specified by the manufacturer results in an unacceptable risk

home healthcare and emergency medical services environments. Hazards inherent in the clinical use of pulse oximetry are *not* covered by specific requirements in the standard.

The standard includes specific test methods and acceptance criteria, including requirements for safe surface temperature of sensor-tissue interface, alarm response time, electronic data interface specifications, test methods for demonstrating accuracy, and disclosure of motion and low perfusion performance.

The first edition (ASTM F1415) was published in 1992. Succeeding versions (ASTM F1415:2000, ISO 9919:2005, ISO 80601-2-61:2011, 2017) have tried to incorporate relevant issues that emerged as pulse oximetry gained broad clinical use. The standard serves as a resource to users and manufacturers about how oximeters work, the physical and physiological assumptions that are made including tissue perfusion and signal inadequacy, response time, and implications for accuracy and performance testing.

Currently, pulse oximeters are designed to provide acceptable clinical accuracy in specific patient populations such as neonatal, pediatric, and adult. The standard provides a recommended clinical protocol for verifying the accuracy performance in an idealized test environment (performed on healthy adults in a controlled laboratory setting) and does not require real-world evidence collected in clinical settings. As of the current edition, the standard does not specify the number of subjects, nor the demographics of this test population. In contrast to the FDA guidance document, the standard does not attempt to distinguish between the performance of pulse oximeters used for "spot checking" or "continuous monitoring" because the difference between these applications relates to the presence of an alarm. The standard is currently being updated.

## VI. Overview of Desaturation Studies from Historical 510(k) Submissions

Given concerns about the potential impact of skin pigmentation on pulse oximeter accuracy, the FDA performed an initial examination of prior cleared 510(k) submissions under product code DQA to better understand skin pigmentation assessments and reporting in regulatory submissions. The evaluation examined premarket clinical study characteristics such as reporting of skin pigmentation classification and reporting of factors potentially impacting device accuracy (e.g., skin pigmentation) in device labeling.

## A. Methods

Out of a pool of 420 510(k) submissions received between January 1, 2000 and December 31, 2020, an approximate 10% sample (44 submissions) with equal numbers pre- and post-2013 FDA guidance was identified through an iterative process of random selection and review for eligibility (Figure VI-1). Submissions were excluded if clinical data were not required, the device was not a pulse oximeter, the device was for pediatric use only, or if the data was not relevant. Pulse oximeters that did not require clinical data included 510(k)s with well-defined manufacturer modifications of their own marketed devices (i.e., special 510(k) submissions) and submissions leveraging predicate device data for modifications that that did not affect the oximetry technology or performance.

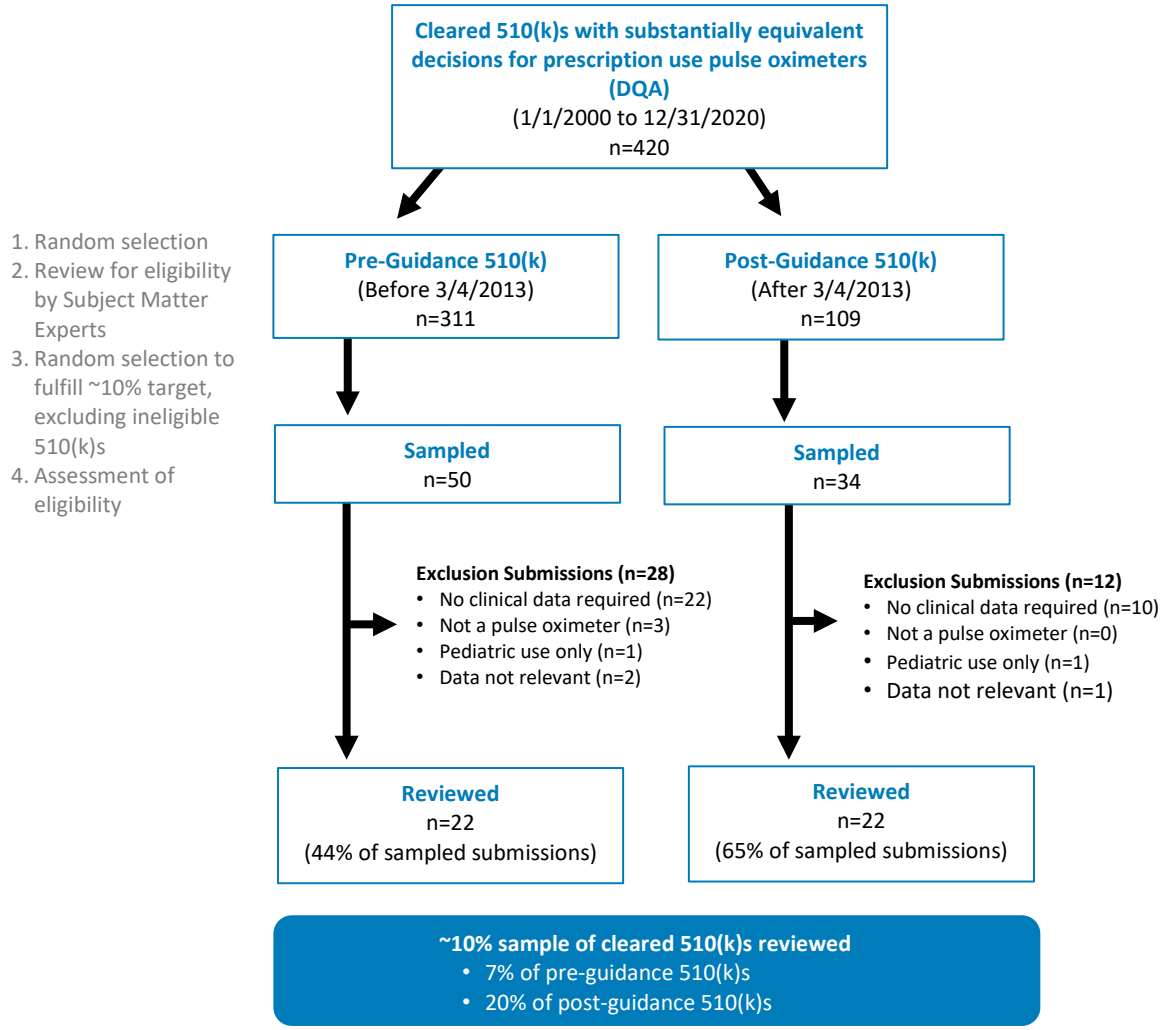



*Figure VI-1. Flowchart for Ascertainment of the 510(k) Submissions Sample*

Data elements were extracted and descriptively analyzed, including skin pigmentation scale and classification, number of subjects, number of paired oxygen measures per subject, availability of patient line-level data, inclusion criteria, reporting of participant characteristics from clinical desaturation studies as well as factors that were indicated as potentially impacting device accuracy from device labeling.

## B. Results

Table VI-1 presents data describing the premarket studies. Desaturation studies in the pre- and post-guidance submissions had a median of 12 study subjects (ranging from 6 to 33) and a median of 24 paired $SaO_2$ and $SpO_2$ measurements (ranging from 13 to 35) per subject. Reporting of patient line-level data increased from 54.5% of pre-guidance submissions to 95.5% of post-guidance submissions per the 2013 guidance recommendations.

In characterizing the clinical study population, all pre- and post-guidance desaturation studies listed being healthy as an inclusion criterion for subject participation. 43.2% of studies specified in their inclusion criteria that participants must be non-smokers, and 50.0% indicated that participants must not have specific medical conditions for eligibility, such as hypertension, cardiovascular disease, and respiratory disease. The percentage of studies reporting various participant characteristics were as follows: 86.4% for age, 97.7% for sex, 72.7% for race/ethnicity, 11.4% for nationality, 27.3% for weight and height, and 9.1% for finger size. Regarding analyses, a large increase was observed in the number of submissions including Bland-Altman plots per the 2013 guidance from 36.4% of pre- to 100% of post-guidance submissions.

As opposed to 50.0% of pre-guidance submissions, skin pigmentation information was only not provided by all but 1 post-guidance 510(k)s. Submissions used inconsistent methods for categorization with many of these classifications lacking clear and standardized definitions. Specifically, among the clinical studies, 2 (4.5%) indicated subjects with darker skin pigmentation but did not report skin pigmentation data. For others, 8 (18.2%) used 2 categories, 13 (29.5%) used 3 categories, 6 (13.6%) used 4 categories, 1 (2.3%) used 5 categories, 1 (2.3%) used the Von Luschan scale, 1 (2.3%) used the Fitzpatrick scale. Of note, 12 (27.3%) studies did not specify skin pigmentation categories, 11 of which were pre-guidance. Although heterogenous, skin pigmentation reporting post-guidance improved with 95.5% of submissions specifying skin pigmentation categories as compared to 50.0% of pre-guidance.

*Table VI-1 Characteristics of Sampled Studies Supporting 510(k) Submissions for Prescription Use Pulse Oximeters: January 1st, 2000 through December 31st, 2020*

| | Pre-Guidance [a] (n=22) | Post-Guidance [b] (n=22) | All Submissions (n=44) |
|---|---|---|---|
| | n (%) | n (%) | n (%) |
| **Number of study subjects** Median (min, max) | 12 (6, 33) | 12 (10, 25) | 12 (6, 33) |
| **Number of paired oxygen measurements per subject** Median (min, max) | 23 (13, 31) | 24 (21, 35) | 24 (13, 35) |
| **Patient level data available** | 12 (54.5) | 21 (95.5) | 33 (75.0) |
| **Study Inclusion Criteria** [c] | | | |
| Healthy | 22 (100.0) | 22 (100.0) | 44 (100.0) |
| Non-smokers | 10 (45.5) | 9 (40.9) | 19 (43.2) |
| No listed medical conditions [d] | 10 (45.5) | 12 (54.5) | 22 (50.0) |
| **Demographic Data Reported** [c] | | | |
| Age | 20 (90.9) | 18 (81.8) | 38 (86.4) |
| Sex | 21 (95.5) | 22 (100.0) | 43 (97.7) |
| Race/Ethnicity | 15 (68.2) | 17 (77.3) | 32 (72.7) |
| Nationality | 2 (9.1) | 3 (13.6) | 5 (11.4) |
| **Other Participant Characteristics** [c] | | | |
| Weight | 4 (18.2) | 8 (36.4) | 12 (27.3) |
| Height | 4 (18.2) | 8 (36.4) | 12 (27.3) |
| Finger size | 1 (4.5) | 3 (13.6) | 4 (9.1) |
| **Use of Bland-Altman plots** | 8 (36.4) | 22 (100.0) | 30 (68.2) |
| **Skin Pigmentation Classification** | | | |
| Singular | 2 (9.1) | . | 2 (4.5) |
| Binary [e] | 3 (13.6) | 5 (22.7) | 8 (18.2) |
| Ternary [f] | 3 (13.6) | 10 (45.5) | 13 (29.5) |
| Quaternary [g] | 2 (9.1) | 4 (18.2) | 6 (13.6) |
| Quinary [h] | . | 1 (4.5) | 1 (2.3) |
| Von Luschan scale [i] | 1 (4.5) | . | 1 (2.3) |
| Fitzpatrick scale [j] | . | 1 (4.5) | 1 (2.3) |
| Not specified [k] | 11 (50.0) | 1 (4.5) | 12 (27.3) |

[a] Received between January 1, 2000 through March 3, 2013
[b] Received on or after March 4, 2013 through December 31, 2020
[c] Not mutually exclusive
[d] Examples of indicated medical conditions include hypertension, cardiovascular disease, and respiratory disease
[e] Example categories: light, dark
[f] Example categories: light, medium, dark
[g] Example categories: light, medium, medium dark, and dark
[h] Example categories: light, medium light, medium, medium dark, dark
[i] 1-36, from the lightest to darkest category
[j] Type I-VI, from the lightest to darkest category
[k] Includes studies with no specified skin pigmentation categories or indication of only "light to dark"

Table VI-2 presents factors that may impact accuracy of pulse oximeters that were listed in device labeling. When examining device labeling in pre- and post-guidance submissions, the three most frequently listed factors that may impact device accuracy included use of intravascular dyes (90.9%), excessive light (88.6%), and presence of electromagnetic or electrosurgical sources (84.1%). The 2013 FDA guidance indicates that labeling should include "all applicable safety information, warnings, cautions, and notes". Subsequently, increased inclusion of many of these factors in the labeling was observed post- as compared to pre-guidance. For example, among the evaluated device labeling post-guidance, 27% indicated that darker skin pigmentation may influence device accuracy. This was an increase from only 9% of device labels among pre-guidance submissions.

*Table VI-2 Factors that may Impact Pulse Oximeters Accuracy, included in Device Labeling*

| Potential Factors Impacting Accuracy | Pre-Guidance [a] (n=22) | Post-Guidance [b] (n=22) | All Submissions (n=44) |
|---|---|---|---|
| Intravascular dyes | 22 (100.0) | 18 (81.8) | 40 (90.9) |
| Excessive light | 20 (90.9) | 19 (86.4) | 39 (88.6) |
| Presence of electromagnetic or electrosurgical sources | 19 (86.4) | 18 (81.8) | 37 (84.1) |
| Dysfunctional hemoglobin | 15 (68.2) | 18 (81.8) | 33 (75.0) |
| Excessive patient movement | 16 (72.7) | 17 (77.3) | 33 (75.0) |
| Fingernail polish or artificial nails | 15 (68.2) | 17 (77.3) | 32 (72.7) |
| Poor circulation | 15 (68.2) | 15 (68.2) | 30 (68.2) |
| Incorrect sensor, or incorrectly applied sensors | 15 (68.2) | 14 (63.6) | 29 (65.9) |
| Extreme moisture (e.g., sterilization, rain) | 14 (63.6) | 14 (63.6) | 28 (63.6) |
| Venous pulsation | 14 (63.6) | 10 (45.5) | 24 (54.5) |
| Severe anemia | 9 (40.9) | 14 (63.6) | 23 (52.3) |
| Low perfusion | 4 (18.2) | 16 (72.7) | 20 (45.5) |
| Improper finger placement | 4 (18.2) | 10 (45.5) | 14 (31.8) |
| In cardiac arrest or shock | 4 (18.2) | 9 (40.9) | 13 (29.5) |
| **Darker skin pigmentation** | **2 (9.1)** | **6 (27.3)** | **8 (18.2)** |
| Poor pulse quality | 3 (13.6) | 2 (9.1) | 5 (11.4) |
| Elevated levels of bilirubin | 1 (4.5) | 2 (9.1) | 3 (6.8) |
| Diagnostic testing | 1 (4.5) | 2 (9.1) | 3 (6.8) |
| Venous congestion | . | 2 (9.1) | 2 (4.5) |
| Extreme finger size | 1 (4.5) | . | 1 (2.3) |

[a] Received between January 1, 2000 through March 3, 2013
[b] Received on or after March 4, 2013 through December 31, 2020

### C. Conclusions Regarding Impact of 2013 FDA Guidance

The 2013 FDA guidance publication resulted in increased use of skin pigmentation data and reporting of classifications used, availability of patient line-level data, use of Bland-Altman plots to describe accuracy, and inclusion of factors that may impact accuracy in the labeling for prescription use pulse oximeters. However, among desaturation clinical studies, a wide variety of skin pigmentation categories was observed, many of which did not include clear or standardized definitions. In addition, the potential impact of darker skin pigmentation on device accuracy was noted in only a limited number labeling for the examined post-guidance 510(k) submissions (27%).

*The Panel will be asked to discuss and make recommendations about impact of skin pigmentation on the accuracy of pulse oximeters (prescription use and OTC).*

*The Panel will be asked to discuss and make recommendations about labeling for prescription use pulse oximeters.*

## VII. Medical Device Adverse Events Reports

The Medical Device Reporting (MDR) regulation (21 CFR 803) specifies mandatory requirements for manufacturers, importers, and device user facilities to report certain device-related adverse events and product problems to the FDA.[xix] Periodic review of MDR data is one of the surveillance tools the FDA uses to monitor the performance of medical devices. The section below describes analysis of MDR reports for prescription use pulse oximeters.

Importantly, although MDRs are a valuable source of information, this passive surveillance system has limitations. The incidence, prevalence, or cause of an event cannot typically be determined from this reporting system alone due to under-reporting of events, inaccuracies in reports, lack of verification that the device caused the reported event, and lack of information about details such as frequency of device use. Because of these limitations, MDRs comprise only one of the FDA's several important postmarket surveillance data sources. These reports, along with data from other sources, can contribute important information to a medical device's benefit-risk assessment.

[xix] https://www.fda.gov/medical-devices/postmarket-requirements-devices/mandatory-reporting-requirements-manufacturers-importers-and-device-user-facilities

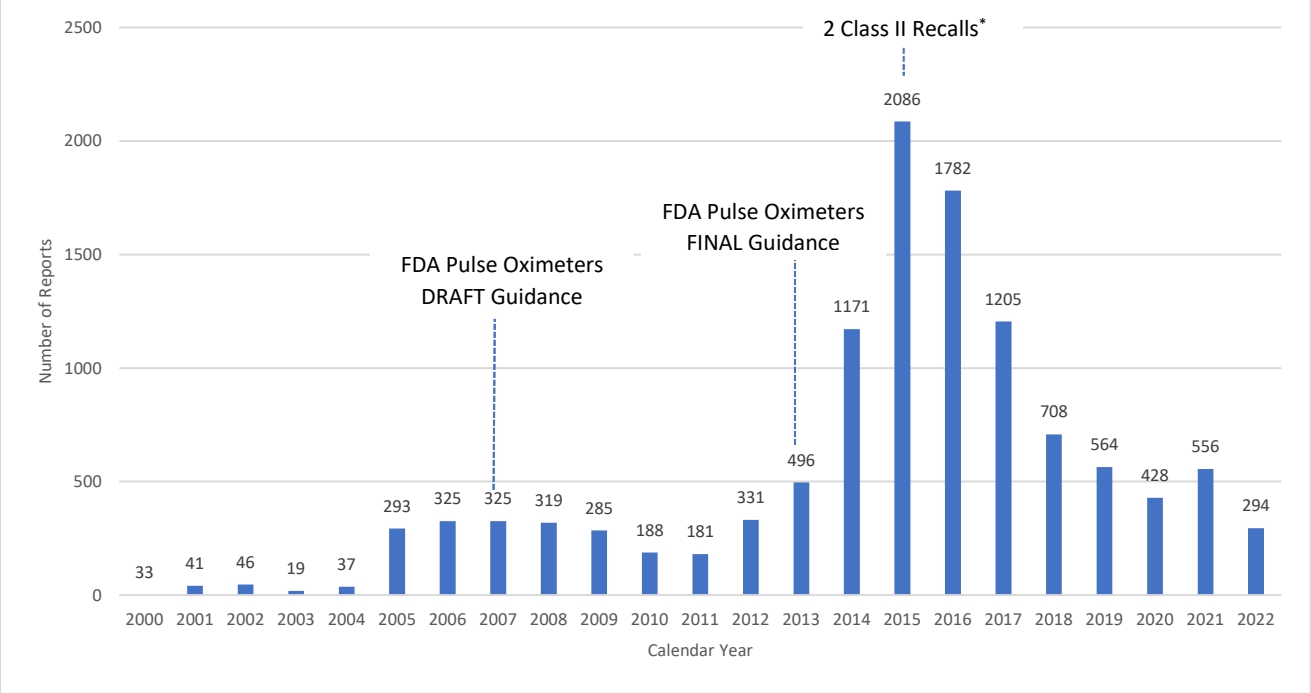
## A. Search Methodology

MDR data for product codes DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter) were pulled for reports received between January 1, 2000 to August 29, 2022. We started with a review of reports with the term "skin" and those submitted as death reports.

Each report identified through the text search was then reviewed to determine if it was relevant to inaccurate $SpO_2$ readings and to assess potential sources for the inaccurate reading.

## B. Results

The initial search by the three product codes yielded 11,713 adverse event reports. Figure 2 presents the number of adverse event reports per year.




* Jul 27, 2015: Coviden Class II recall due to potential missing segments on the display that could result in misinterpretation of data, impacted 317,257 distributed units; and Nov 25, 2015: Massimo Class II recall due to sensors manufactured with incompatible configurations that could result in sensors that will provide either no readings or inaccurate readings, impacted 3,476 distributed units worldwide.

*Figure VII-1 Number of Adverse Event Reports Submitted through the MDR System for Product Codes DQA, DQZ, and NFL from January 1, 2000 through August 29, 2022*

Most of the adverse event reports were for malfunctions (91%), followed by serious injury reports (4%), and death reports (2%) (Table VII-1). Forty percent of the death reports mentioned issues with the alarm system, either not alarming at all or having the volume set too low. It is important to note that these

reports were from critically ill patients and a causal association between the use of the pulse oximeter and the death cannot be established. In some cases, the patients were found unresponsive without the pulse oximeter attached to them, some reported the death and the user's request for pulse oximeter trending data even when the reporter did not consider the death to be associated with the use of the pulse oximeter. Several death reports stated that the pulse oximeter continued to provide $SpO_2$ readings after the patient's death.

Ninety-nine reports were identified as potentially related to inaccurate $SpO_2$ readings. Most (84%) do not contain sufficient information to ascertain the potential source for the inaccuracy (Table VII-1). Only three mentioned the patients were African American males, and these were death reports. These reports do not contain sufficient information to determine if the $SpO_2$ readings prior to the death were within the pulse oximeter accuracy specifications. In addition, three reports mentioned skin pigmentation, as: "dark skin", "light brown skin", and "light skin" (these were malfunction reports). There were two reports that mentioned low or poor perfusion.

*Table VII-1 Adverse Event Reports received between Jan 1, 2000 to Aug 29, 2022 for Product Codes DQA, NLF, and DPZ: N = 11,713*

| | n | % |
|---|---|---|
| **Adverse Event Report Type** | | |
| Malfunction | 10,712 | 91.45 |
| Serious Injury | 512 | 4.37 |
| Death | 296 | 2.53 |
| Other | 137 | 1.17 |
| *Missing* | *56* | *0.48* |
| **Potentially related to Inaccurate $SpO_2$ Reading** | **99** | **15.59*** |
| Mentions African American Race | 3 | 3.03** |
| Mentions Skin Pigmentation | 3 | 3.03** |
| Mentions Low Perfusion | 2 | 2.02** |
| Reports device continued to provide readings after patient's death | 15 | 15.15** |
| Insufficient Information to Assess Reason for Inaccuracy | 83 | 83.84** |

* Out of 635 reviewed reports (344 with search term "skin" and 296 death reports, note 5 death reports also mention skin and were only counted once); ** % estimated out of 99; the categories are not exclusive of each other; numbers can add up to more than 99

Table VII-2 presents the top 10 reported patient and device problems for reports relevant to $SpO_2$ inaccurate readings. Each adverse event report may include more than one problem code, the data in the table does not represent unique adverse events.

*Table VII-2 Top 10 Patient and Device Problems Reported in Adverse Event Reports related to Inaccurate Pulse Oximeter Readings, n = 99*

| Health Effect Clinical Code (ProCode) | Count* | Device Problem Code (ProCode) | Count* |
|---|---|---|---|
| Death (1802) | 53 | Incorrect Measurement (1383) | 29 |
| No Known Impact Or Consequence To Patient (2692) | 13 | High Readings (2459) | 25 |
| Low Oxygen Saturation (2477) | 13 | False Reading From Device Non-Compliance (1228) | 9 |
| No Consequences Or Impact To Patient (2199) | 8 | Low Readings (2460) | 8 |
| Insufficient Information (4580) | 7 | Unable to Obtain Readings (1516) | 5 |
| Cyanosis (1798) | 4 | Device Operates Differently Than Expected (2913) | 4 |
| Skin Discoloration (2074) | 4 | Incorrect, Inadequate or Imprecise Result or Readings (1535) | 4 |
| No Information (3190) | 3 | Material Integrity Problem (2978) | 3 |
| Therapy/non-surgical treatment, additional (2519) | 3 | Invalid Sensing (2293) | 3 |
| Loss of consciousness (2418) | 3 | High Test Results (2457) | 3 |

*Categories are not exclusive of each other; numbers can add up to more than 99

## C. Evidence Assessment

As noted above, there are several limitations that need to be considered when assessing evidence from MDRs. First, the submission of an adverse event report does not necessarily mean that there is a causal association between the use of a pulse oximeter and the adverse event being reported. Through review of the subset of MDRs for pulse oximeters, we found these reports lacked sufficient information and the potential reason for the inaccurate reading cannot be ascertained. The large amount of missing data on race, ethnicity and skin pigmentation hinders our ability to assess their potential impact on the pulse oximeter performance. The MDR system is a passive surveillance system limited by lack of denominator data; therefore, rates for adverse events cannot be estimated.

MDR data is best used for qualitative snapshot assessment of adverse events for a device or device type, for trending of adverse events over time, and for safety signal detection (i.e., unexpected events, change in severity or frequency of known events, user error/human factors issues).

# VIII. A Systematic Literature Review of the Real-World Performance of Pulse Oximeters

## A. Search Methodology

A PubMed search was conducted on August 9, 2022, using the strategy "pulse oxim* AND (race OR racial OR pigment)" to identify clinical studies published since 2013 pertaining to the topic of real-world performance of pulse oximetry in individuals with darker skin pigmentation.

## B. Results

The search identified 162 articles. Of these, 140 were excluded for the following reasons: publication date (n=74), relevance (n=43), no clinical data (n=16), did not include all variables of interest (n=5), and fewer than nine participants (n=2). Additionally, six articles were identified via cross referencing and additional searches after the initial search date. After the inclusion/exclusion process, 28 articles were selected for inclusion. Of 28 articles, 24 included clinical data derived from the following study designs: seven cross sectional studies, 15 retrospective studies, and three lab studies (shown in the Appendix 1). Of note, three articles were preprints published online before final publication,[34,35] one of which was published online before peer review.[36]

Of the 24 clinical studies, two specified that the participants were infants,[8,37] and one specified children 17 and under;[2] the rest specified adults or had no age cutoff.

There were three relevant systematic reviews identified which addressed the topic of potential bias in pulse oximetry for people with darker skin pigmentation, all published in the year 2022. The authors' conclusions are described below.[4,38,39]

- Cabanas, et al (2022) identified 41 references published between 1976-2022, which included 34 prospective and retrospective studies. Nine studies were considered at high risk of bias due to unstandardized classification of skin pigmentation such as "dark", "black", "light", or "white". The authors reported that there was a considerable upsurge of publications in 2021, due to the COVID-19 pandemic as well as increased concern about pulse oximeter accuracy across skin types. They concluded that "*there is growing evidence that pulse oximeters are less accurate in dark-skinned individuals at lower saturation (<80%) resulting in overestimations.*" (page 15 of 20), and also that "*a more accurate method for classifying the research subjects into categories by degree of skin pigmentation should be employed in these studies*" (page 16 of 20).[4]

- Shi, et al (2022) identified 32 references published between 1985-2021. Meta-analysis of 15 studies using skin pigmentation levels and 22 studies using race/ethnicity showed that pulse

oximetry probably overestimates oxygen saturation in people with high level of skin pigmentation (pooled mean bias 1.11%; 95% CI 0.29 to 1.93%) and people described as Black/African American (1.52%; 95% CI 0.95 to 2.09%), although this evidence was considered moderate to low certainty. The authors concluded that "*Pulse oximetry may overestimate blood oxygen saturation levels for people with dark skin in hospital settings compared with gold standard SaO2 measures. The evidence for the measurement bias identified for other levels of skin pigmentation or ethnicities is more uncertain. Whilst the extent of measurement bias and overall accuracy meet current international thresholds, the variation of pulse oximetry measurements appears unacceptably wide. Such a small overestimation may be crucial for some patients: particularly at the threshold that informs clinical decision-making*" (page 11 of 14).[39]

- Poorzargar, et al (2022) identified 22 references published between 1988-2020, looking specifically at pulse oximetry accuracy under poor perfusion conditions (including hypothermia, vasoactive drug use, or other factors not reported). Only one study controlled for skin pigmentation, by excluding participants with darker skin. *The authors reported that most oximeter models were accurate in patients with poor perfusion, newer models were more accurate than older models, and earlobe placement was more accurate than fingertip. They also concluded that more trials are needed that incorporate FDA guidelines for a diverse range of skin pigmentation.*

Evidence tables sorted by type of study are included in Appendix 1.

## C. Evidence Assessment

There are important limitations that should be considered when assessing the published literature and especially real-world data. There is considerable heterogeneity in study designs, study populations (i.e., healthy volunteers, hospitalized patients, and adult vs. pediatric populations), measurement of skin pigmentation and/or race/ethnicity, and study endpoints. Many used the endpoint of bias or occult hypoxemia with varying definitions; others used an endpoint such as ≤ 94% $SpO_2$ to qualify for treatment. Another point of variability between studies is the amount of time elapsed between $SpO_2$ and $SaO_2$ paired measurements; simultaneous measurements are preferred but not always possible, especially when extracting retrospective data from electronic health records. Pulse oximeter accuracy often appears worse in real-world studies than in desaturation studies. However, in retrospective real-world studies the paired measurements of $SaO_2$ and $SpO_2$ are usually not simultaneous, which may lead to larger disagreements due to normal fluctuations in $SaO_2$ and treatment effects on $SaO_2$.

In each real-world study using electronic health records, hospital grade pulse oximeters were used, although it can be assumed that brands, models, and the use of reprocessed sensors varied across hospitals and even within the same hospital. Some authors provided information regarding which pulse oximeter(s) were used, but others did not. Additionally, technology has advanced over time, and thus

pulse oximetry accuracy may have changed over time. Some authors may have adjusted for year of procedure as a covariate to account for this (e.g., Burnett, et al[3]). We excluded articles published before the 2013 FDA guidance document, which outlined more stringent recommendations for pulse oximeter accuracy and inclusion of individuals with darker skin pigmentation in the testing sample.

Real-world evidence is observational in nature and therefore comparison groups may differ by baseline characteristics such as health conditions and demographics. Some authors have attempted to control for confounding variables such as age or Sequential Organ Failure Assessment (SOFA) score, but residual confounding factors are likely. For example, real world samples often included very sick patients, such as those about to undergo Extracorporeal Membrane Oxygenation (ECMO).[7] The prevalence of hypoxemia may vary between such groups, confounding comparisons of the occult hypoxemia rate.[40]

Skin pigmentation has been postulated as a contributing factor to pulse oximeter error. However, in RWE studies it is hard to adjust for it because skin pigmentation level is not systematically measured as part of routine care, and it is not captured in the EHR systems, which are used as data source for RWE studies. Self-reported race and/or ethnicity is an inaccurate proxy for skin pigmentation but is used frequently due to convenience and availability in medical records. Most studies were conducted using U.S. patients, although some were from Europe, Asia, Africa, and Australia. Studies in other countries may use different types of pulse oximeters, may have different classification systems for race/ethnicity, and may have different prevalence of darker skin pigmentation in the population compared to the US. These factors may also lead to residual confounding. More research is needed with standardized measurement of skin pigmentation.

Finally, there is often publication bias in any review of published literature, where statistically significant results are more likely to be submitted and accepted for publication.

Overall, despite these limitations, there appears to be mounting real-world evidence that pulse oximeter accuracy varies by self-reported race, and skin pigmentation. However, there is a need for prospective studies that utilize standardized measurement of skin pigmentation, capture simultaneous measurement of $SaO_2$ and $SpO_2$ paired data and systematically collect data on important confounders, to have more robust evidence about the impact of skin pigmentation on real-world pulse oximetry.

*The Panel will be asked to discuss the clinical evidence from the scientific literature about inaccuracies in pulse oximetry among subjects with darker skin pigmentation.*

# IX. FDA-Funded RWE Efforts Assessing the Performance of Pulse Oximeters in the US

Two grants have been awarded to the University of California (UCSF)-Stanford *Center for Excellence in Regulatory Science and Innovation* (CERSI), to conduct studies assessing the real-world performance of pulse oximeters in the adult and pediatric populations.[xx] Both studies are real-world, prospective clinical trials to address the main study question of mean bias in $SpO_2$ readings for a given value of $SaO_2$, across skin pigmentation levels. Secondary study questions ask whether real-world pulse oximeters accuracy by different levels of skin pigmentation is within recommended accuracy parameters per FDA guidance; and what other factors (such as low perfusion or low light transmittance) may have an impact on the accuracy of pulse oximeters. Subjects with a range of $SaO_2$ values will be included (e.g., high 80% to low 90% by $SaO_2$). Both studies have been approved by the respective Institutional Review Boards (IRBs) at UCSF and Stanford.

The studies are designed to capture skin pigmentation, simultaneous $SpO_2$ and $SaO_2$ paired measurements, patients' peripheral perfusion (the flow of blood to the extremities of the body), type of pulse oximeter, probe, and site of probe placement, and will include a balanced, diverse patient population. Skin pigmentation will be captured by colorimetry tools, and the Fitzpatrick and von Luschan skin color scales. Data on other factors that may impact the performance of pulse oximeters (e.g., demographics, carboxyhemoglobin levels (adults), comorbidities, supplemental oxygen therapy, core temperature) will be captured from the electronic health records.

The adult study [xxi] is being conducted at the UCSF Health Medical System. The study population includes subjects 22 years old and older, with different skin pigmentation levels who are being treated in the Intensive Care Unit (ICU), operating room, or the emergency room, and who have an arterial blood line in place. The pediatric study[xxii] is being conducted at The Heart Center at Stanford's Lucile Packard Children's Hospital. It will include subjects 21 years old and younger, who are undergoing cardiac catheterization, cardiac surgery, or are hospitalized in the Cardiovascular ICU, and have an arterial blood line in place.

[xx] https://www.fda.gov/science-research/advancing-regulatory-science/cersi-research-projects

[xxi] https://www.fda.gov/science-research/advancing-regulatory-science/prospective-clinical-study-pulse-oximeter-errors-adult-hospitalized-patients-varying-skin

[xxii] https://www.fda.gov/science-research/advancing-regulatory-science/prospective-clinical-study-evaluate-accuracy-pulse-oximeters-children

These studies address a gap in regulatory science, i.e., the need for prospective evaluation of the real-world performance of pulse oximeters.

# X. Assessment of Skin Pigmentation

There is no consensus on the best approach to assess skin pigmentation for medical device development. Neither current FDA guidance nor the recognized ISO standard for pulse oximeters recommend a particular methodology to assess skin pigmentation. As seen from the assessment of historical 510(k) submissions in Section VI, different scales have been utilized by different manufacturers, i.e., Fitzpatrick and von Luschan scales; and there was also inconsistent categorization of skin pigmentation data.

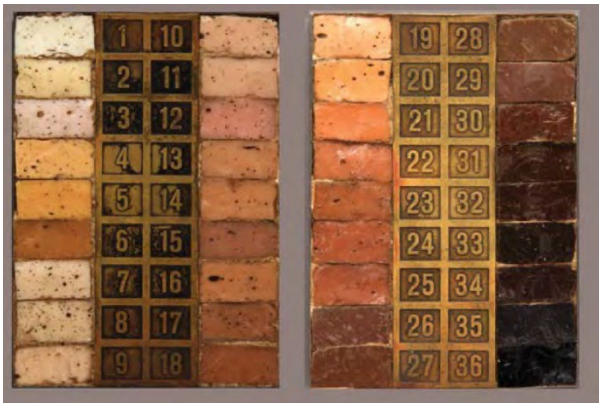
The Fitzpatrick skin type (FST) scale is a numerical classification for skin types. It was developed by dermatologist Thomas B. Fitzpatrick to determine how different skin types react to ultraviolet (UV) light (i.e., ability to tan when exposed to sunlight).[41] Although it was not developed for assessment of skin color, it is commonly used for that purpose. It ranges from Skin Type I (Fair) through VI (Dark Brown), and it is usually assessed visually or self-reported. In contrast, the von Luschan Chromatic Scale was developed by anthropologist Felix von Luschan, it was used extensively to establish racial classifications of populations according to skin color.[42] It consists of 36 opaque glass tiles used to compare with skin color in ideally a place that is not exposed to sun such as under the arm.

The figure below presents both scales and Table X-1 presents the 36 von Luschan skin categories in relation to a version of the Fitzpatrick skin type scale.

**A. Fitzpatrick Skin Type Scale** [43]


doi: https://doi.org/10.1371/journal.pone.0241843.g001

**B. von Luschan Chromatic Scale – Color Tiles** [42]



**C. Reproduction of von Luschan Chromatic Scale**[44]

| | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 10 | | 19 | 28 | |
| | 2 | 11 | | 20 | 29 | |
| | 3 | 12 | | 21 | 30 | |
| | 4 | 13 | | 22 | 31 | |
| | 5 | 14 | | 23 | 32 | |
| | 6 | 15 | | 24 | 33 | |
| | 7 | 16 | | 25 | 34 | |
| | 8 | 17 | | 26 | 35 | |
| | 9 | 18 | | 27 | 36 | |

*Figure X-1 The Fitzpatrick Skin Type Scale and the von Luschan Chromatic Scale*

*Table X-1 The von Luschan Chromatic Skin Scale in Relation to the Fitzpatrick Skin Type Scale*

| Felix von Luschan Scale | Fitzpatrick Scale |
|---|---|
| 0 to 6 | I: Very Fair |
| 7 to 13 | II: Fair |
| 14 to 20 | III: Medium |
| 21 to 27 | IV: Olive skin |
| 28 to 34 | V: Brown |
| 35 to 36 | VI: Dark Brown |

Although not typically used in studies assessing pulse oximeters accuracy, there are alternative tools for skin classification as reported by Ware et.al.[45], typically used in the field of dermatology.

To identify objective approaches to assess skin pigmentation a review of published literature was conducted, and further evaluation of more quantitative methods to assess skin pigmentation is ongoing.

Narrative Literature Review of Optical Melanometers

The Anesthesia Patient Safety Foundation (APSF) recently highlighted the need for objective measurement of skin tone in pulse oximetry studies.[xxiii] Table X-2 shows a list of different methods currently in use based on their quantitative accuracy in differentiating different skin pigmentation populations. Skin pigmentation classification in pulse oximeter studies has been exclusively subjective in studies supporting 510(k) submissions (Table VI-1). Since the release of the FDA guidance in 2013, the majority (15) of these studies have used binary or ternary descriptive methods (light, medium, dark) without a standardized color scale for differentiating patient populations. In literature reports, even less quantitative subjective methods such as the use of self-identification of racial/ethnic categories such as Black, Asian, Hispanic, and White/Caucasian are used as surrogates for determining skin pigmentation. Standardized color systems such as the Fitzpatrick skin type (FST)[32,46-48], Von Luschan, and Massey [28] and Munsell [37] scales are much less prevalent in either data supporting 510(k) submissions or the general literature. The most common approach, FST, was originally designed to assess susceptibility to sunburn, but has increasingly been used to categorize skin color. In recent years, these subjective approaches for assessing skin pigmentation have come under criticism for poor inter-operator variability, the use of arbitrary categorical bins rather than a continuous measurement scale, and the fact that they do not quantify a biological ground truth, such as epidermal melanin concentration.[33,49]

*Table X-2 Methods of estimating skin pigmentation to increase quantitative capability (top least quantitative, bottom most quantitative)*

| Skin Pigmentation Assessment Method | Degree of Objectivity/Quantitation |
|---|---|
| Racial/ethnic Self-identification (Black, White, Hispanic, Asian, etc.) | Subjective with limited/no skin pigmentation information |
| Skin Color Descriptive Terms (light, medium, medium dark, dark) | Subjective due to lack of a standardized scale. Large variance within skin color groupings. |
| Sunburn susceptibility (Fitzpatrick skin phototype scale Types I-VI) | Subjective (questionnaire-based), quantitative (categorical); sometimes used as a non-standardized color scale |
| Color scale (Von Luschan, Masey) | Subjective but with lower variance due to the use of standardized color categories |
| Optical Melanometry Methods (Spectroscopy, Colorimetry) | Quantifiable information that is not dependent on a subjective evaluation; but some metrics not standardized. |
| Biopsy with histological/optical processing or high performance liquid chromatography | Quantitative melanin content, but can involve reader-dependent steps (e.g., layer identification) |

[xxiii] APSF Newsletter, page 57: https://www.apsf.org/wp-content/uploads/newsletters/2021/3602/APSF3602.pdf

While any differences in optical based medical device performance correlated with race/ethnicity could potentially be caused by a variety of factors, skin pigmentation is typically considered the most likely cause. This is due to: (a) the strong, spectrally-varying optical absorption of epidermal melanin[12], and (b) evidence that variations in skin pigmentation have a significant impact on visible to near-infrared reflectance signals, and thus impact the performance of optical devices (e.g., regional oximeters, bilirubinometers).[50,51] To address the shortcomings of subjective skin assessment methods, there has been growing interest in the use of optical sensing devices incorporating automated algorithms to assess pigmentation such as melanometry.[4,8,33]

Further complicating the ability to differentiate patient populations based on skin pigmentation is the variance in skin pigmentation at different anatomical sites. The most common pulse oximetry site is the finger – specifically the fingernail and distal palmar finger, which contain minimal melanin especially when compared to anatomical sites with high pigmentation levels such as the arm or torso. Furthermore, the fingernail bed has a melanocyte content that is approximately 5% that of normal skin and, unlike melanocytes in the skin, these cells do not produce melanin.[52] Since a subject's pigmentation level should ideally be determined at the pulse oximeter measurement site, the ability of a method to accurately differentiate patient populations needs to reliably quantify small differences in melanin content.

Optical methods for measuring skin melanin content measurements should be accurate, repeatable, and reproducible. Optical methods have a variety of factors that can affect their accuracy. As an example, melanometer errors are due, in part, to variables such as blood content, contact pressure, anatomical site, and ambient temperature.[53] Given the variability in methods and results, it will be important to establish consensus methods for characterization of optical methods for measuring skin pigmentation performance relative to a high-quality ground truth.

Melanometers are one optical method specifically designed to provide a quantitative measure of the epidermal skin melanin content. While most melanometers are not cleared or approved by FDA as medical devices (although some have been cleared as accessory components for laser therapy devices), they are commercially available as research tools. In dermatology, these devices have been implemented to study scars[54,55], melasma[56] and psoriasis[57,58]. Several commercially available melanometers have been compared with well-established conventional subjective skin color classification methods, such as the FST scale. Although moderate to good correlation was observed

between several device outputs and FST, results were overall inconsistent and demonstrate that these devices and skin classification systems require more evaluation to fully harness their potential.[59-68]

Finally, it should be noted that ongoing research in other optical methods with depth-selective capability, like photoacoustics[69] and spatial frequency domain imaging[70], may lead to devices with better accuracy in measuring melanin content non-invasively in vivo[52]. Additional work is needed to establish standardized best practices for validation and clinical implementation of optical skin pigmentation measurement devices such as melanometers, as well as to realize their potential as tools for assessing skin pigmentation.

*The Panel will be asked to discuss and make recommendations about assessment and reporting of skin pigmentation data in studies evaluating the accuracy of pulse oximeters.*

## XI. Summary

Pulse oximeters are non-invasive tools used to estimate blood oxygen saturation. Although pulse oximetry presents benefits for patients, the limitations and risks associated with inaccurate pulse oximeters readings must be well understood. The COVID-19 pandemic resulted in increased use of pulse oximeters not only in the hospital setting, but at home as well. Clinicians recommending the use of pulse oximeters to patients to monitor health conditions at home, and consumers using OTC pulse oximeters for medical purposes, should be aware of the limitations. A case of hypoxemia that goes undetected may result in delays in treatment and adverse patient outcomes.

There is mounting evidence from real-world studies that suggest performance of pulse oximeters can be affected by skin pigmentation. Additionally, there are factors other than skin pigmentation that can also impact the accuracy of a pulse oximeter such as (not limited to) nail polish, motion, perfusion index, carboxyhemoglobin, time between paired $SpO_2$ and $SaO_2$ measures, and prevalence of hypoxemia across different groups. RWE studies may not be able to control for all important confounders. Furthermore, standardization of skin pigmentation assessment, reporting and categorization of skin pigmentation data for analysis of $SpO_2$ bias, remains a challenge.

Finally, pulse oximeters' accuracy in diagnosing hypoxemia (i.e., sensitivity, specificity, positive predictive value, and negative predictive value) is not well described. All prescription use pulse

oximeters available on the U.S. market are cleared with a general indication for use (i.e., to monitor oxygen saturation and spot check for trending), and to date no OTC pulse oximeter has been cleared for medical use.

After hearing from patients, regulators, researchers, and industry, the Panel will deliberate on four important topics:

1. Impact of skin pigmentation on accuracy of pulse oximetry.
2. Standardization of skin pigmentation assessment, reporting and analysis of skin pigmentation data.
3. How accurate pulse oximeters should be, and methodologies to evaluate accuracy of pulse oximeters and $SpO_2$ bias due to skin pigmentation.
4. Modifications to labeling for prescription use pulse oximeters to better communicate their limitations to clinicians and patients using pulse oximeters at home for medical purposes.

# XII. References


1. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. *N Engl J Med.* 2020;383(25):2477-2478.
2. Andrist E, Nuppnau M, Barbaro RP, Valley TS, Sjoding MW. Association of Race With Pulse Oximetry Accuracy in Hospitalized Children. *JAMA Netw Open.* 2022;5(3):e224584.
3. Burnett GW, Stannard B, Wax DB, et al. Self-reported Race/Ethnicity and Intraoperative Occult Hypoxemia: A Retrospective Cohort Study. *Anesthesiology.* 2022;136(5):688-696.
4. Cabanas AM F-GM, Latorre K, Leon D, Martin-Escudero P. Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. *Sensors.* 2022;22(9).
5. Fawzy A, Wu TD, Wang K, 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.
6. Henry NR, Hanson AC, Schulte PJ, et al. Disparities in Hypoxemia Detection by Pulse Oximetry Across Self-Identified Racial Groups and Associations With Clinical Outcomes. *Crit Care Med.* 2022;50(2):204-211.
7. Valbuena VSM, Barbaro RP, Claar D, et al. Racial Bias in Pulse Oximetry Measurement Among Patients About to Undergo Extracorporeal Membrane Oxygenation in 2019-2020: A Retrospective Cohort Study. *Chest.* 2022;161(4):971-978.
8. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J Perinatol.* 2022;42(1):79-85.
9. Wong AI, Charpignon M, Kim H, et al. Analysis of Discrepancies Between Pulse Oximetry and Arterial Oxygen Saturation Measurements by Race and Ethnicity and Association With Organ Dysfunction and Mortality. *JAMA Netw Open.* 2021;4(11):e2131674.
10. Jubran A. Pulse oximetry. *Crit Care.* 1999;3(2):R11-R17.
11. Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. *Med Devices (Auckl).* 2014;7:231-239.
12. Jacques SL. Quick analysis of optical spectra to quantify epidermal melanin and papillary dermal blood content of skin. *J Biophotonics.* 2015;8(4):309-316.
13. Feiner JR, Bickler PE, Mannheimer PD. Accuracy of methemoglobin detection by pulse CO-oximetry during hypoxia. *Anesth Analg.* 2010;111(1):143-148.
14. Afshari A, Saager RB, Burgos D, et al. Evaluation of the robustness of cerebral oximetry to variations in skin pigmentation using a tissue-simulating phantom. *Biomed Opt Express.* 2022;13(5):2909-2928.
15. ISO, IEC. Medical Electrical Equipment: Particular Requirements for Basic Safety and Essential Performance of Pulse Oximeter Equipment. In. *80601-2-61*. Switzerland 2017.
16. Franceschini MA, Gratton E, Fantini S. Noninvasive optical method of measuring tissue and arterial saturation: an application to absolute pulse oximetry of the brain. *Opt Lett.* 1999;24(12):829-831.
17. Kirchner T, Jaeger M, Frenz M. Machine learning enabled multiple illumination quantitative optoacoustic oximetry imaging in humans. *Biomed Opt Express.* 2022;13(5):2655-2667.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.
19. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol.* 2005;34(1):215-220.
20. Barnett AG, Van Der Pols JC, Dobson AJ. Correction to: Regression to the mean: what it is and how to deal with it. *Int J Epidemiol.* 2015;44(5):1748.
21. Bland JM, Altman DG. Regression towards the mean. *BMJ.* 1994;308(6942):1499.

22. Bland JM, Altman DG. Some examples of regression towards the mean. *BMJ.* 1994;309(6957):780.
23. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet.* 1995;346(8982):1085-1087.
24. Johnson WD, George VT. Effect of regression to the mean in the presence of within-subject variability. *Stat Med.* 1991;10(8):1295-1302.
25. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(2):135-160.
26. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat.* 2007;17(4):571-582.
27. Lin LI. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Stat Med.* 2000;19(2):255-270.
28. Harris BU, Char DS, Feinstein JA, Verma A, Shiboski SC, Ramamoorthy C. Accuracy of Pulse Oximeters Intended for Hypoxemic Pediatric Patients. *Pediatr Crit Care Med.* 2016;17(4):315-320.
29. Harris BU, Stewart S, Verma A, et al. Accuracy of a portable pulse oximeter in monitoring hypoxemic infants with cyanotic heart disease. *Cardiol Young.* 2019;29(8):1025-1029.
30. Ross PA, Newth CJ, Khemani RG. Accuracy of pulse oximetry in children. *Pediatrics.* 2014;133(1):22-29.
31. Schallom M, Prentice D, Sona C, Arroyo C, Mazuski J. Comparison of nasal and forehead oximetry accuracy and pressure injury in critically ill patients. *Heart Lung.* 2018;47(2):93-99.
32. Smith RN, Hofmeyr R. Perioperative comparison of the agreement between a portable fingertip pulse oximeter v. a conventional bedside pulse oximeter in adult patients (COMFORT trial). *S Afr Med J.* 2019;109(3):154-158.
33. Okunlola OE, Lipnick MS, Batchelder PB, Bernstein M, Feiner JR, Bickler PE. Pulse Oximeter Performance, Racial Inequity, and the Work Ahead. *Respir Care.* 2022;67(2):252-257.
34. Sudat SEK, Wesson P, Rhoads KF, et al. Racial Disparities in Pulse Oximeter Device Inaccuracy and Estimated Clinical Impact on COVID-19 Treatment Course. *Am J Epidemiol.* 2022 Sep 29:kwac164. doi:10.1093/aje/kwac164. Epub ahead of print. PMID: 36173743.
35. Chesley CF, Lane-Fall MB, Panchanadam V, et al. Racial disparities in occult hypoxemia and clinically based mitigation strategies to apply in advance of technological advancements. *Respir Care.* 2022 Jun 3:rescare.09769. doi: 10.4187/rescare.09769. Epub ahead of print. PMID: 35679133.
36. Gadrey SM, Mohanty P, Haughey SP, et al. Overt and occult hypoxemia in patients hospitalized with novel coronavirus disease 2019. *medRxiv* [Preprint]. 2022 Jun16:2022.0614.22276166. doi: 10.1101/2022.06.14.22276166. PMID:35734082; PMCID: PMC9216725.
37. 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 e372.
38. Poorzargar K, Pham C, Ariaratnam J, et al. Accuracy of pulse oximeters in measuring oxygen saturation in patients with poor peripheral perfusion: a systematic review. *J Clin Monit Comput.* 2022;36(4):961-973.
39. Shi C, Goodall M, Dumville J, 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.
40. Pennello G. In Real-World Studies, Comparisons of Predictive Values Are Confounded by Prevalence. *Chest.* 2022;162(2):e103.
41. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol.* 1988;124(6):869-871.

42. Swiatoniowski AK, Quillen EE, Shriver MD, Jablonski NG. Technical note: comparing von Luschan skin color tiles and modern spectrophotometry for measuring human skin pigmentation. *Am J Phys Anthropol.* 2013;151(2):325-330.
43. Charlton M, Stanley SA, Whitman Z, et al. The effect of constitutive pigmentation on the measured emissivity of human skin. *PLoS One.* 2020;15(11):e0241843.
44. Charoenngam N, Sriussadaporn S. Darker Skin Color Measured by Von Luschan Chromatic Scale and Increased Sunlight Exposure Time Are Independently Associated with Decreased Odds of Vitamin D Deficiency in Thai Ambulatory Patients. *J Nutr Metab.* 2021;2021:8899931.
45. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial limitations of fitzpatrick skin type. *Cutis.* 2020;105(2):77-80.
46. Ebmeier SJ, Barker M, Bacon M, et al. A two centre observational study of simultaneous pulse oximetry and arterial oxygen saturation recordings in intensive care unit patients. *Anaesth Intensive Care.* 2018;46(3):297-303.
47. Harskamp RE, Bekker L, Himmelreich JCL, et al. Performance of popular pulse oximeters compared with simultaneous arterial oxygen saturation or clinical-grade pulse oximetry: a cross-sectional validation study in intensive care patients. *BMJ Open Respir Res.* 2021;8(1).
48. Stell D, Noble JJ, Kay RH, et al. Exploring the impact of pulse oximeter selection within the COVID-19 home-use pulse oximetry pathways. *BMJ Open Respir Res.* 2022;9(1).
49. Norton HL. Variation in pulse oximetry readings: melanin, not ethnicity, is the appropriate variable to use when investigating bias. *Anaesthesia.* 2022;77(3):354-355.
50. Bhutani VK, Gourley GR, Adler S, Kreamer B, Dalin C, Johnson LH. Noninvasive measurement of total serum bilirubin in a multiracial predischarge newborn population to assess the risk of severe hyperbilirubinemia. *Pediatrics.* 2000;106(2):E17.
51. Mendenhall MJ, Nunez AS, Martin RK. Human skin detection in the visible and near infrared. *Appl Opt.* 2015;54(35):10559-10570.
52. Gunes P, Goktay F. Melanocytic Lesions of the Nail Unit. *Dermatopathology (Basel).* 2018;5(3):98-107.
53. Fullerton A, Fischer T, Lahti A, Wilhelm KP, Takiwaki H, Serup J. Guidelines for measurement of skin colour and erythema. A report from the Standardization Group of the European Society of Contact Dermatitis. *Contact Dermatitis.* 1996;35(1):1-10.
54. Lee KC, Dretzke J, Grover L, Logan A, Moiemen N. A systematic review of objective burn scar measurements. *Burns Trauma.* 2016;4:14.
55. Draaijers LJ, Tempelman FR, Botman YA, Kreis RW, Middelkoop E, van Zuijlen PP. Colour evaluation in scars: tristimulus colorimeter, narrow-band simple reflectance meter or subjective evaluation? *Burns.* 2004;30(2):103-107.
56. Kanellis VG. Objective quantification of melasma severity using melanometers to quantify melanin pigmentation. *Biophys Rev.* 2020;12(5):1139-1140.
57. Pershing LK, Bakhtian S, Wright ED, Rallis TM. Differentiation of involved and uninvolved psoriatic skin from healthy skin using noninvasive visual, colorimeter and evaporimeter methods. *Skin Res Technol.* 1995;1(3):140-144.
58. Ahmad A, Kaji I, Murakami Y, et al. Transformation of Arabidopsis with plant-derived DNA sequences necessary for selecting transformants and driving an objective gene. *Biosci Biotechnol Biochem.* 2009;73(4):936-938.
59. Isa ZM, Shamsuddin K, Bukhari NBI, et al. The reliability of Fitzpatrick Skin Type Chart Comparing to Mexameter (Mx 18) in measuring skin color among first trimester pregnant mothers in Petaling District, Malaysia. *Malaysian Journal of Public Health Medicine.* 2016;16:59-65.
60. Khalid AT, Moore CG, Hall C, et al. Utility of sun-reactive skin typing and melanin index for discerning vitamin D deficiency. *Pediatr Res.* 2017;82(3):444-451.

61. Linde K, Wright CY, du Plessis JL. Subjective and objective skin colour of a farmworker group in the Limpopo Province, South Africa. *Skin Res Technol.* 2020;26(6):923-931.
62. Richard A, Rohrmann S, Quack Lotscher KC. Prevalence of Vitamin D Deficiency and Its Associations with Skin Color in Pregnant Women in the First Trimester in a Sample from Switzerland. *Nutrients.* 2017;9(3).
63. Robinson JK, Penedo FJ, Hay JL, Jablonski NG. Recognizing Latinos' range of skin pigment and phototypes to enhance skin cancer prevention. *Pigment Cell Melanoma Res.* 2017;30(5):488-492.
64. Sharma VK, Gupta V, Jangid BL, Pathak M. Modification of the Fitzpatrick system of skin phototype classification for the Indian population, and its correlation with narrowband diffuse reflectance spectrophotometry. *Clin Exp Dermatol.* 2018;43(3):274-280.
65. Sommers MS, Fargo JD, Regueira Y, et al. Are the Fitzpatrick Skin Phototypes Valid for Cancer Risk Assessment in a Racially and Ethnically Diverse Sample of Women? *Ethn Dis.* 2019;29(3):505-512.
66. van der Wal MBA, Bloemen MCT, Verhaegen P, et al. Objective Color Measurements: Clinimetric Performance of Three Devices on Normal Skin and Scar Tissue. *Journal of Burn Care & Research.* 2013;34:e187–e194.
67. Young AR, Morgan KA, Ho TW, et al. Melanin has a Small Inhibitory Effect on Cutaneous Vitamin D Synthesis: A Comparison of Extreme Phenotypes. *J Invest Dermatol.* 2020;140(7):1418-1426 e1411.
68. Bailey SH, Oni G, Brown SA, et al. The use of non-invasive instruments in characterizing human facial and abdominal skin. *Lasers Surg Med.* 2012;44(2):131-142.
69. Attia ABE, Balasundaram G, Moothanchery M, et al. A review of clinical photoacoustic imaging: Current and future trends. *Photoacoustics.* 2019;16:100144.
70. Phan T, Rowland R, Ponticorvo A, et al. Quantifying the confounding effect of pigmentation on measured skin tissue optical properties: a comparison of colorimetry with spatial frequency domain imaging. *J Biomed Opt.* 2022;27(3).
71. Seitz KP, Wang L, Casey JD, et al. Pulse Oximetry and Race in Critically Ill Adults. *Crit Care Explor.* 2022;4(9):e0758.
72. Bangash MN, Hodson J, Evison F, et al. Impact of ethnicity on the accuracy of measurements of oxygen saturations: A retrospective observational cohort study. *EClinicalMedicine.* 2022;48:101428.
73. Crooks CJ, West J, Morling JR, et al. Pulse oximeter measurements vary across ethnic groups: an observational study in patients with COVID-19. *Eur Respir J.* 2022;59(4).
74. Gottlieb ER, Ziegler J, Morley K, Rush B, Celi LA. Assessment of Racial and Ethnic Differences in Oxygen Supplementation Among Patients in the Intensive Care Unit. *JAMA Intern Med.* 2022;182(8):849-858.
75. Sudak SEK, Wesson P, Rhoads KF, et al. Racial Disparities in Pulse Oximeter Device Inaccuracy and Estimated Clinical Impact on COVID-19 Treatment Course. *Am J Epidemiol.* 2022.
76. Valbuena VSM, Seelye S, Sjoding MW, 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.
77. Wiles MD, El-Nayal A, Elton G, et al. The effect of patient ethnicity on the accuracy of peripheral pulse oximetry in patients with COVID-19 pneumonitis: a single-centre, retrospective analysis. *Anaesthesia.* 2022;77(2):143-152.
78. Baek HJ, Shin J, Cho J. The Effect of Optical Crosstalk on Accuracy of Reflectance-Type Pulse Oximeter for Mobile Healthcare. *J Healthc Eng.* 2018;2018:3521738.
79. Mantri Y, Jokerst JV. Impact of skin tone on photoacoustic oximetry and tools to minimize bias. *Biomed Opt Express.* 2022;13(2):875-887.

# XIII. Appendix 1: Evidence Tables for Systematic Literature Review on RWE Performance of Pulse Oximeters

*Table XIII-1 Literature Describing Real-World Evidence from Cross-Sectional Studies*

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| Ebmeier 2018[46] | Aus/NZ, 2015, ≥16yo, ICU | 394<br>394 | Fitzpatrick:<br>I-II ("light")<br>III-IV ("medium")<br>V-VI ("dark")<br>n's and %s not reported | Bias: Bland-Altman | Unadjusted regression coefficient (95% CI)<br>Light: Reference<br>Medium: 0.9 (0.4, 1.3)<br>Dark: **2.4 (1.2, 3.6)** |
| Foglia 2017[37] | US, 2013-2015, infants, Cyanotic Congenital Heart Disease | 35<br>35 | Munsell System Soil Color Chart:<br>Light n=21 (60%)<br>Dark n=14 (40%) | Bias:<br>$SpO_2$-$SaO_2$ | N.S. |
| Harskamp 2021[47] | Netherlands, 2020, ≥18yo, ICU | 35<br>234 | Fitzpatrick:<br>I-III n=30 (85.7%)<br>IV-VI n=5 (14.3%) | Mean bias in $SpO_2$<br><br>Accuracy measured by Arms, and Mean Absolute Error (MAE)<br><br>Diagnostic accuracy for hypoxemia ($SaO_2$ < 90%) | Mean bias range: -0.6 to -4.8<br>None of the pulse oximeters met Arms < 3% in $SaO_2$ range of 70-100%.<br>MAE range: 2.3 to 5.1 and 5 of the pulse oximeters met < 3%<br>Darker skin complexion associated with poorer $SpO_2$ performance<br><br>Negative Predictive Value (NPP) for hypoxemia diagnosis: 98% to 99%<br><br>Positive Predictive Value (PPV): 11% to 30% |
| Henry 2022[6] | US, 2018-2020, ≥18yo, ICU or surgical | 26,603<br>128,285 | Race:<br>White n=24,493<br>Black n=1,263<br>Asian n=574<br>American Indian n=273 | OH: $SaO_2$ < 88% despite concurrent $SpO_2$ > 92% | Adjusted OR (95%CI)<br>White: Reference<br>Black: **1.65 (1.28, 2.14)**<br>Asian: 1.53 (0.95, 2,47)<br>American Indian: 1.31 (0.80, 2.16) |

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| | | | | | OH associated with mortality **OR=2.96 (1.20, 7.28)** |
| Seitz 2022[71] | US, 2018-2021, adults (no age cutoff reported), patients receiving invasive mechanical ventilation in medical ICU (excluding COVID-19) | 1,024<br>5,557 | Race:<br>White n=4,788 pairs (86%)<br>Black n=769 pairs (14%) | OH: $SaO_2$ < 88% despite $SpO_2$ between 92%-96%<br><br>Hyperoxemia: $PaO_2$ > 150mmHg despite $SpO_2$ between 92%-96% | OH:<br>White: 1.1% (0.7, 1.7)<br>Black: 3.5% (1.6, 6.6)<br><br>Hyperoxemia:<br>White: 2.4% (1.8, 3.2)<br>Black: 4.7% (2.5, 8.1) |
| Smith 2019[32] | South Africa, years NR, ≥18yo, surgical | 220<br>220 | Fitzpatrick:<br>I n=12 (5.5%)<br>II n=28 (12.7%)<br>III n=69 (31.4%)<br>IV n=45 (20.5%)<br>V n=28 (12.7%)<br>VI n= 38 (17.3%) | Bias: Bland-Altman | N.S. |
| Stell 2022[48] | UK, years NR, ≥19yo, COVID-19 | 50<br>915 | Fitzpatrick:<br>I n=6<br>II n=21<br>III n=9<br>IV n=5<br>V n=7<br>VI n= 2 | OH: $SpO_2$ (reference model) < 92% despite concurrent $SpO_2$ (portable model) > 92% ("false negative") | Skin tone significant predictor of bias, especially for darker skin (FSP 5 or 6), for 2 of 5 devices |

Bold font = statistically significant; N.S. = no significant differences between groups; OH = Occult Hypoxemia

*Table XIII-2 Literature Describing Real-World Evidence from Retrospective Studies*

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| Andrist 2022[2] | US, 2015-2020, children (≤17yo), inpatient | 1,061<br>9,023 | Race:<br>White n=878 (82.8%)<br>Black n=183 (17.2%) | OH: $SaO_2$ < 88% despite $SpO_2$ > 92% | Adjusted OR (95% CI)<br>White: Reference<br>Black: **2.16 (1.36, 3.44)***<br>Black: **1.79 (1.07, 3.02)****<br><br>*paired measures as unit of analysis<br>** patients as unit of analysis |
| Bangash 2022[72] | UK, 2017-2021, adults, inpatient | 16,818<br>20,231 | Race:<br>White n=13,649 (81.2%)<br>Asian n=1,965 (11.7%)<br>Black n=674 (4.0%)<br>Other n=530 (3.2%) | OH: $SaO_2$ < 94% despite $SpO_2$ ≥ 94% | Adjusted OR (95% CI)<br>White: Reference<br>Asian: NR<br>Black: **1.47 (1.09, 1.98)**<br>Other: NR |
| Burnett 2022[3] | US, 2008-2019, ≥18yo, patients receiving anesthetic | 46,253<br>151,070 | Race/ethnicity:<br>White n=22,089<br>Black n=5,177<br>Asian n=2,612<br>Hispanic n=6,304<br>Other n=10,071 | OH: $SaO_2$ < 88% despite $SpO_2$ ≥ 92% | Adjusted OR (95% CI)<br>White: Reference<br>Black: **1.44 (1.11, 1.87)**<br>Asian: 0.77 (0.51, 1.17)<br>Hispanic: **1.31 (1.03, 1.68)**<br>Other: 1.24 (1.00, 1.53) |
| Chesley 2022[35] | US, 2019-2021, adults (no age cutoff reported), ICU | 7,693<br>105,467 | Race/ethnicity:<br>White n=4,621 (60%)<br>Black n=1,919 (25%)<br>Latinx n=226 (3%)<br>Asian/PI n=239 (3%)<br>Indigenous n=17 (0.2%)<br>Other n=220 (3%)<br>Unknown n=451 (6%) | OH: $SaO_2$ < 88% despite $SpO_2$ between 92%-96% | Adjusted OR (95% CI)<br>White: Reference<br>Black: **2.84 (2.44, 3.30)**<br>Latinx: **1.69 (1.22, 2.34)**<br>Asian/PI: **1.64 (1.15, 2.34)**<br>Indigenous: 0.51 (0.07, 3.72)<br>Other: 1.41 (0.99, 1.99) |
| Crooks 2022[73] | UK, 2020-2021, no age limit reported, | 2,997<br>5,374 | Race:<br>White n=3,946<br>Black n=151 | Mean difference $SpO_2$-$SaO_2$ | Adjusted<br>White: Reference<br>Black: **+1.8% (+0.2, +3.4%)** |

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| | inpatients with COVID-19 | | Asian n=246<br>Mixed n=36<br>NR n=995 | | Asian: **+1.9% (+0.6, +3.2%)**<br>Mixed: +3.2% (-0.1, +6.6%) |
| Fawzy 2022[5] | US, 2020-2021, no age limit reported, Emergency department visit or hospitalized for COVID-19 | 1,216<br>32,282 | Race/ethnicity:<br>Non-Hispanic White n=460<br>Black n=478<br>Asian n=63<br>Non-Black Hispanic: n=215 | OH: $SaO_2$ < 88% despite concurrent $SpO_2$ of 92% to 96%<br><br>Treatment Initiation: SpO2 ≤94% or use of supplemental oxygen | OH:<br>White: 17.2%<br>Black: 28.5%<br>Asian: 30.2%<br>Non-Black Hispanic: 29.8%<br><br>Treatment Initiation (hazard ratio, 95%CI)<br>Black: 0.71 (0.63, 0.80)<br>Non-Black Hispanic: 0.77 (0.66, 0.89)<br>Asian: 0.97 (0.62, 1.5) |
| Gadrey 2022[36] *preprint before peer review | US, 2020-2021, ≥18yo, Emergency department visit or hospitalized for COVID-19 | 5,319<br>1,909,867 | Race/ethnicity:<br>White, non-Hispanic: n=1,433<br>Black n=2,835<br>Hispanic n=612<br>Other n=439 | Clinical deterioration (either transfer to ICU or in-hospital mortality) | Black patients appeared to have better oxygenation but worse outcomes for comparable degrees of apparent oxygenation (reported graphically as Empirical Cumulative Distribution Functions) |
| Gottlieb 2022[74] | US, 2008-2019, no age limit reported, ICU | 3,069<br>n/a | Race/ethnicity:<br>Asian n=83<br>Black n=207<br>Hispanic n=112<br>White n=2,667 | Time-weighted average supplemental oxygen rate | Model 2: regression coefficient<br>Asian: **-0.291 (-0.546, -0.035)**<br>Black: **-0.294 (-0.460, -0.128)**<br>Hispanic: **-0.242 (-0.463, -0.020)**<br><br>Model 3 (after controlling for discrepancy between average $SpO_2$ and average Hb oxygen saturation)<br>Asian: -0.144 (-0.386, 0.098)<br>Black: -0.081 (-0.239, 0.077)<br>Hispanic: -0.092 (-0.301, 0.118) |
| Sjoding 2020[1] | US, 2014-2015 and 2020, no age limit reported, ICU | 10,001<br>48,097 | Race:<br>Black n=1,326<br>White n=8,675 | OH: $SaO_2$ < 88% despite 92% ≥ $SpO_2$ ≥ 96% | UM cohort: OH (adjusted)<br>Black: 11.4% (7.6, 15.2%)<br>White: 3.6% (2.5, 4.6%) |

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| | | | | | Multicenter cohort: OH (unadjusted)<br>Black: 17.0% (12.2, 23.3%)<br>White: 6.2% (5.4, 7.1%) |
| Sudat 2022[75] | US, 2020-2021, adult (no age limit reported), Cohort 1 (hospital visits with ABG), Cohort 2 (emergency visits with COVID-19) | Cohort 1: 43,753 paired measures<br>Cohort 2: 8,735 paired measures | Race/ethnicity:<br>Cohort 1:<br>Non-Hispanic White (NHW) n=35,127<br>Non-Hispanic Black (NHB) n=8,626<br><br>Cohort 2:<br>Non-Hispanic White n=7,036<br>Non-Hispanic Black n=1,699 | Cohort 1:<br>$SpO_2$ $SaO_2$ pairs, OH<br><br>Cohort 2: clinical and treatment characteristics | Cohort 1:<br>Concurrent $SaO_2$<$SpO_2$ by 2%, and difference was 1% larger for NHB<br>OH:<br>NHW 3.01%<br>NHB 5.50% (p<0.001)<br><br>Cohort 2: lower admission probability, dexamethasone treatment, supplemental oxygen treatment, and increased time to treatment |
| Valbuena 2022a[7] | US, 2019-2020, ≥18yo, patients on ECMO due to ARDS or COVID-19 | 372<br>372 | Race/ethnicity:<br>White n=186<br>Hispanic n=70<br>Asian n=65<br>Black n=51 | Pre-ECMO OH: SpO2 between 92 to 96% despite $SaO_2$ < 88% | Rate (95%CI)<br>White: 10.2% (6.2, 15.3%)<br>Hispanic: 8.6 % (3.2, 17.7%)<br>Asian: 9.2% (3.5, 19.0%)<br>Black: 21.5% (11.3, 35.3%)<br><br>OR (95%CI):<br>White: reference<br>Black: 2.57 (1.12, 5.92) |
| Valbuena 2022b[76] | US, 2013-2019, US veterans (no age limit reported), inpatient excluding ICU | 30,039<br>30,039 | Race:<br>Non-Hispanic White n=21,918<br>Non-Hispanic Black n=6,498<br>Hispanic/Latino n=1,623 | OH: $SpO_2$ ≥ 92% despite $SaO_2$ < 88% | OH: Unadjusted probability (95% CI)<br>Non-Hispanic White: 15.6% (15.0, 16.1%)<br>Non-Hispanic Black: 19.6% (18.6, 20.6%)<br>Hispanic: 16.2% (14.4, 18.1%) |
| Vesoulis 2022[8] | US, 2012-2019, infants <32 weeks gestation, NICU | 294 | Race:<br>White n=170<br>Black n=124 | OH: $SaO_2$ < 85% despite concurrent $SpO_2$ > 90% | OH:<br>White: 7.7% (181 of 2342)<br>Black: 9.2% (188 of 2044) |

| Source | Study Population (location, year(s) of data collection, age, health status) | Sample Size (patients, paired measures) | Skin Pigmentation Measurement or Proxy (categories, n, %) | Outcome Variable Definition | Reported Measure |
|---|---|---|---|---|---|
| | | | | | p = 0.08 |
| Wiles 2022[77] | UK, 2020-2020, ≥16yo, COVID pneumonitis | 194<br>6,216 | Race:<br>White n=135<br>Asian n=34<br>Black n=19<br>Other n=6 | Bias:<br>$SpO_2$-$SaO_2$ | Bias (limits of agreement)<br>White: 0.28% (1.79, 2.35)<br>Asian: -0.33% (-2.47, 2.35)<br>Black: -0.75% (-3.47, 1.97)<br>Other: NR |
| Wong 2021[9] | US, 2014-2021, no age limit reported, ICU | 79,044<br>87,971 | Race:<br>Asian n=1,919 (2.3%)<br>Black n=26,032 (29.6%)<br>Hispanic n=2,397 (2.7%)<br>White n=57,632 (65.5%) | OH: $SaO_2$ < 88% despite $SpO_2$ ≥ 88%<br>Organ dysfunction: SOFA and CVSOFA scores 24 hrs after ABG measurement<br>Length of hospital stay<br>In-hospital mortality | OH:<br>Asian: 4.9%<br>Black: 6.9%<br>Hispanic: 6.0%<br>White: 4.9%<br>p<.001<br>OH was associated with greater organ dysfunction and higher in-hospital mortality. |

*Table XIII-3 Literature Describing Laboratory Studies*

| Source | Study Design | Study Population (location, age, health status) | Sample Size (patients, paired measures) | Proxy for Skin Pigmentation | Comparison | Reported Measure |
|---|---|---|---|---|---|---|
| Baek 2018[78] | Desaturation | Korea, adult, healthy volunteers | 28 | Race:<br>Asian<br>Caucasian<br>African American | Crosstalk sensor vs. crosstalk-free sensor<br><br>Measurement error (bias): $SpO_2$-estimated $SpO_2$ ± precision (SD) | Crosstalk sensor:<br>Asian 0.8258 ± 2.1603<br>Caucasian 0.8733 ± 1.9716<br>African American -3.0591 ± 3.9925<br><br>Crosstalk-free sensor:<br>Asian -0.8824 ± 2.2859<br>Caucasian 0.6741 ± 3.2822<br>African American 0.9699 ±2.2268 |
| Mantri 2022[79] | Case series | US, adult, healthy volunteers | 9 | Fitzpatrick:<br>Type 1 n=3<br>Type 4 n=3<br>Type 6 n=3 | Photoacoustic imaging vs. pulse oximeter | SpO2: no significant differences by skin type<br><br>Photoacoustics: higher PA signal, reduced penetration depth for Type 6 |
| Okunlola 2022[33] | Desaturation | U.S., adult, healthy volunteers | 491<br>3,778 | Skin pigmentation: dark, medium, light | Dark vs. light to medium | Small positive bias in darkpigmentation group (data presented in Bland Altman plots) |