FDA Executive Summary

Performance Evaluation of Pulse Oximeters Taking into Consideration Skin Pigmentation, Race and Ethnicity

Prepared for the
February 2, 2024, meeting of the
Anesthesiology and Respiratory Therapy Devices Panel of the
Medical 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. From a regulatory perspective, pulse oximeters can be categorized as prescription use or over-the-counter (see Section II).

As part of the premarket evaluation of pulse oximeters, FDA has required clinical data assessing pulse oximeter performance in participants with different skin pigmentation. However, a December 2020 published report suggested that pulse oximeters may be less accurate in Black patients compared with White patients.[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, with a focus on skin pigmentation. In February of 2021, the FDA issued a *Safety Communication* to inform patients and health 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.[i] 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 readings, 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. Since then, the FDA has 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 demonstrated that the purchase and use of these devices increased during the first year of the COVID-19 pandemic.[ii] In addition, the Agency completed an evaluation of pulse oximeter 510(k) submissions (under product code DQA) that have been cleared for

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

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

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 showed that after the FDA guidance for prescription use pulse oximeters[iii] was issued in 2013, there was an increase in the submission of line-level patient 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. Details on the study methods, results, and discussion were described in the 2022 Executive Summary.[iv] 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 VII.

Since the December 2020 publication,[1] additional real-world studies have been published suggesting increased risk for missed diagnosis of hypoxemia[v] (i.e., *"occult hypoxemia"*), delays in treatment eligibility decisions and worse patient outcomes among patients with darker skin pigmentation.[3-10] FDA completed a systematic literature review of the real-world performance of pulse oximeters. See Section V for details on the systematic literature review methods, summary of findings, and additional details on the evidence assessment.

FDA is holding a public meeting of the Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee on February 2, 2024. The Committee will be asked to discuss the performance evaluation of pulse oximeters, taking into consideration skin pigmentation, race, and ethnicity. Specifically, the Committee will discuss an approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters submitted for premarket review, taking into consideration a patient's skin pigmentation, and patient-reported race and ethnicity. The Committee will discuss the type and amount of data that should be provided by manufacturers to FDA to evaluate the performance of pulse oximeters submitted for premarket review, including prescription and over-the-counter indications, and labeling considerations. The Advisory Committee 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 oximeters.

[iii] https://www.fda.gov/media/72470/download
[iv] https://www.fda.gov/media/162709/download
[v] An abnormal low concentration of oxygen in the blood

II. FDA Regulation of Pulse Oximeters

Pulse oximeters can be categorized as:

- Prescription Use Pulse Oximeters: Regulated under product codes DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter), these pulse oximeters are reviewed by the FDA, are regulated under the 510(k) process,[vi] and are available only with a prescription. The FDA requires that these pulse oximeters undergo clinical testing to confirm their accuracy. The clinical testing results are reviewed during FDA's premarket assessment, prior to 510(k) clearance. These devices 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.

- Over-the-Counter (OTC) Pulse Oximeters: Regulated under product codes PGJ and OCH. Most commonly these standalone pulse oximeters are intended for general wellness[vii] 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. Of note, multi-parameter OTC devices that include pulse oximetry for medical purposes are regulated under different primary product codes. OTC oximeters intended for medical purposes undergo review by the FDA and require premarket authorization.

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.

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

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

[viii] https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device

A manufacturer that intends to market a new pulse oximeter for medical purposes in the U.S. must submit a premarket submission (i.e., 510[k]) 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 monitoring or spot checking for trending).

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.[11] 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 where pulse oximeters operate (typically 660 nm and 940 nm, respectively), the absorption is relatively low, allowing for measurement of light transmission.[12] 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 for systolic and diastolic phases, and the ratio of these ratios is calculated 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.[12]

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 and reflected 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 are not limited to):

- Skin pigmentation, i.e., epidermal melanin content;

[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

- 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;
- Motion;
- Ambient light;
- Dyes, such as tattoo ink or medications with optical absorption properties;
- Nail polish.

B. Potential Device Technology Approaches for Bias Mitigation

Scientific literature describes a range of developments with potential to mitigate ethno-racial disparities in performance of optical devices. Typically, mitigation approaches focus on reducing the impact of epidermal melanin content – which exhibits strong, spectrally-dependent optical absorption – on detected signals.[13] Variations in skin pigmentation have been shown to impact signals detected by other optical device technologies, such as regional oximeters and transcutaneous bilirubinometers.[14] In these devices, mitigation approaches have included the use of multi-wavelength measurements to quantify melanin content and mathematically compensate for its impact on detected signals, as well as the use of multi-distance collection geometries to reduce sensitivity to superficial tissue absorbers.

Cleared pulse oximeters incorporate a "ratio of ratios" approach (systolic/diastolic and red/NIR wavelengths) in order to cancel out the impact of static absorbers like melanin. However, this approach appears to be insufficient to account for effects of skin pigmentation. There are a number of relatively simple approaches to improve the robustness of pulse oximeters to skin pigmentation that have been described or implied by prior research studies:

a. More rigorous monitoring of signal quality. A recent clinical study performed at UCSF indicates that racial disparities in performance occur when the pulse oximeter percent modulation (i.e., perfusion index) is low (e.g., <1.0).[15] Thus, devices that only report results when percent modulation is sufficiently high should be more robust to pigmentation.
b. Avoiding sensors with highly reflective or colored bandages. There is evidence that the optical properties of the sensor and/or bandage surface may impact oximeter accuracy.[16,17] Light colored bandages may enhance sensitivity to epidermal absorbers.
c. The use of lasers or other narrow-bandwidth light sources in pulse oximeters.[18,19] Because melanin absorption spectrum decreases strongly with wavelength, the red LED spectrum may red-shift slightly as it interacts with the epidermis (upon entry, exit and other boundary

reflections), which would cause the calibration curve for devices to be pigmentation-dependent. A narrow bandwidth source should mitigate this effect.

Somewhat more complex approaches for improving robustness to skin pigmentation may include:

a. Multi-wavelength pulse oximetry with algorithms to compensate for melanin absorption. Pulse CO-oximeters used for real-time detection and monitoring of dyshemoglobins such as methemoglobin (metHb) and carboxyhemoglobin (COHb)[20] incorporate LEDs at several wavelengths beyond the typical red and infrared values. It may be possible to use this type of technology, along with algorithms that are analogous to those used in bilirubinometry[21] to extract the effect of melanin from measured signals.
b. Implementing quantitation of skin pigmentation level in conjunction with algorithms that change calibration curves (determined empirically from participants with specific skin pigmentation levels) based on this value.[22] This type of approach may be useful even if the underlying mechanisms of pigmentation-dependent bias are not understood.

Finally, more advanced optical technologies, such as polarization,[23] frequency domain near-infrared spectroscopy[24] and photoacoustic imaging[25] have also been used to measure arterial oxygen saturation and may enable greater accuracy and robustness to skin pigmentation. However, these technologies are not inherently immune to melanin optical absorption and may require alternate mitigation approaches.[26]

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 is currently considering updating the 2013 Guidance's recommendations and seeking input from the Committee and other stakeholders. Section IV outlines the current guidance recommendations; see Section X for highlights of key clinical considerations for potential updates to FDA's guidance document.

The FDA pulse oximeter guidance contains recommendations for in vivo testing for pulse oximetry accuracy under laboratory conditions. Desaturation studies are usually conducted on a small sample of

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

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 participants 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). The FDA guidance recommends that desaturation studies include ten 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. rather than the skin tone distribution of the general U.S. population.

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

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 patient-level line listing data, Bland-Altman plots, error plots for both individual patients and pooled patient data, and rationale for any data points excluded from the analysis. Summary statistics are recommended such as population mean bias (μ_0), between-subject variance (σ_μ^2), within-subject variance (σ^2), upper 95% and lower 95% limits of agreement, and the average root mean square (Arms). The guidance recommends acceptable Arms by type of sensor, as presented below in Table IV-1.

Table IV-1 Acceptable Average Root Mean Square (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 has been Arms, the root mean square of the difference between simultaneous paired measurements of SpO_2 and SaO_2 pooled across all measurements from all patients. If Arms = 3%, then the probability that an SpO_2 value is within 3% of the SaO_2 value is roughly 68% (assuming the distribution is normal). The Bland-Altman (BA) scatterplots[27] of the difference D = SpO_2 – SaO_2 vs. the mean A = (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 patients. 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 D vs. SaO_2. However, in the latter plot, D may tend to decrease with increasing SaO_2 due to regression to the mean (RTM), a pattern caused not by SpO_2 bias but by random variation between SpO_2 and SaO_2.[28-33]

The bias of SpO_2 at particular values of SaO_2 can be evaluated visually using a quantile-quantile scatterplot of the ordered values of SpO_2 vs. the ordered values of SaO_2.[34] The distance of a point from the 45-degree line is an estimate of SpO_2 bias for that SaO_2 value. Intrasubject correlation among repeated pairs of 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 patient).[28-33,35,36]

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).[37] Some studies evaluate coverage probability, calling it the acceptable agreement rate.[37-41]

A linear regression of SpO_2 on SaO_2 can be used to model the expected difference of D = SpO_2 – SaO_2 for a particular value of SaO_2, provided that the model fits the data well. For example, the expected difference may be evaluated at SaO_2 = 88% or another threshold below which patients are defined to

have hypoxemia. If the regression line is $SpO_2 = a + b*SaO_2 + error$, then the expected value of D at $SaO_2 = x$ is $E(D) = a + (b - 1)x$ [CLSI, 2013].[42] Under the model, the expected value of D given the value of SaO_2 can be shown to be the sum of SpO_2 bias and RTM.[30,33] Note that if a = 0 and b = 1, E(D) = 0 for any x. Suppose for two groups of patients, 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 the expected difference between the groups at $SaO_2 = x$ is $E_2(D) - E_1(D) = a_2 - a_1 + (b_2 - b_1)x$.

Linear regression of SpO_2 on SaO_2 can be adjusted for covariates such as skin color, (e.g., individual typology angle or ITA[43]), pulsatility (i.e., perfusion index[15]), race, ethnicity, gender, age, sex, body mass index (BMI), finger size (for finger clips), etc. 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 methemoglobin level, smoking status, body temperature, use of vasopressors or 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 the effects of 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. Other Considerations

After receiving FDA clearance based on verification of device performance in healthy patients, pulse oximeters are used on hospitalized and outpatient adults, as well as pediatric populations including neonates for a variety of purposes 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 additional convenience arterial sampling, clustered around SaO_2 values of 90%, is recommended for neonatal populations, this is currently not required for other populations. Extrapolation of device performance derived from healthy patients 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, they 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 pulse oximeters accuracy and performance across sub-groups of patients with different skin pigmentation.

V. Systematic Literature Review of the Real-World Performance of Pulse Oximeters

A. Search Methodology

A PubMed search was conducted on December 6, 2023, to update the systematic literature review presented in the 2022 executive summary. The search used the strategy "pulse oxim* AND (race OR racial OR pigment)" to identify clinical studies published between August 9, 2022 – December 6, 2023, pertaining to the topic of pulse oximetry performance in individuals with darker skin pigmentation. Additional articles were identified via cross referencing and review articles.

B. Results

The search update identified 51 articles published between 2022-2023, and four additional articles were identified through cross referencing. Of these 55, 32 were excluded for the following reasons: no clinical data (n=22), did not include all variables of interest (n=6), and relevance (n=4). Additionally, five articles were already presented in the 2022 executive summary as preprints;[44-48] these articles were reviewed for any new or changed information, which was added to Appendix 1 as appropriate. After the inclusion/exclusion process, 18 new articles were selected for inclusion.

The evidence assessment includes the 18 new articles, added to the 28 articles included in the 2022 executive summary literature assessment, for a total of 46 articles. The study designs are specified as follows: cross-sectional studies (n=9 total, including 2 new articles), retrospective studies (n=22 total, including 7 new articles), laboratory studies (n=8 total, including 5 new articles), and systematic literature reviews (n=7, including 4 new articles). The new evidence was added to the tables in Appendix 1, to reflect the entire body of evidence currently available; new articles are marked with an asterisk before the first author's name. Of note, three articles were pre-prints published online before final publication.[15,49,50]

Of the 39 clinical studies, there were five that focused on pediatrics.[3,9,51-53] The other studies either focused on adults or did not specify an age range.

There were seven relevant systematic literature reviews identified which addressed the topic of potential bias in pulse oximetry for people with darker skin pigmentation, all published in the years 2022-2023. The authors' conclusions are described below.

- 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).[5]

- 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).[47]

- 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.[54]

- Jamali, et al (2022) identified 7 nonclinical and 15 clinical studies published between 1985-2022, looking at pulse oximetry accuracy for adults with a variety of skin tones. They concluded that "*a review of relevant literature indicates a higher frequency of technical difficulties, increased mean bias, and higher rates of occult hypoxemia for subjects with dark skin tones, that may then be associated with less aggressive clinical management and increased mortality… solving this disparity requires improved regulatory requirements for the approval of pulse oximeters; in addition to the increased inclusion of individuals with dark skin tones during device calibration, the data required for approval should be disaggregated for different skin tones and oxygen saturation ranges*" (page 1962).[55]

- Al-Halawani, et al (2023) reported that 22 out of 28 references identified found that SpO_2 was overestimated in those with darker skin relative to reference SaO_2 measurements obtained by blood gas analysis. Over half of the studies reported an increase in bias for subjects from all racial subgroups as they became less saturated. They concluded the following: "*We have analysed the literature on the effect of skin pigmentation on pulse oximeter accuracy, and which reports that SpO_2 is frequently overestimated in Black adults and infants, and in subjects with darker skin. As a result, these patients are more likely to experience occult hypoxaemia than White subjects, which may lead to delayed medical attention… We propose potential areas to investigate in the near future, such as the immediate identification of inaccurate pulse oximeters, the investigation of multi-wavelength pulse oximeters in subjects with different skin tones, obtaining more data from darkly skinned subjects to implement in-built calibration options, to objectively quantify skin pigmentation, and the development of computational models to predict differing bio-optical outcomes. Future work must include the development of pulse oximeter design and technology to eliminate bias associated with skin pigmentation, as well as all other known limitations*" (pages 18-19).[56]

- Aoki, et al (2023) reported that 8 out of 10 references identified "*statistically significant higher pulse oximeter readings in darker-skinned patients with hypoxia compared to their arterial blood gas measurements. Occult hypoxia was more prevalent in Black and Hispanic patients than in White patients. Minority patients overall (Black, Asian, and American Indian) were more likely to*

have a SaO2 < 88% that was not detected by pulse oximetry (occult hypoxemia) during hospitalization. With greater levels of hypoxemia, the differences between SpO2 and SaO2 were greater. If SaO2 was < 90%, then SpO2 was overestimated in all ethnicities but worse in minorities. In conclusion, the bias found in pulse oximeter readings in the skin of color broadly impacts patients with hypoxemia. The failure of SpO2 measuring devices to detect occult hypoxemia can delay the delivery of lifesaving treatment to critically ill patients requiring respiratory rehabilitation and supplemental oxygen therapy. This may lead to adverse health outcomes, increased in-hospital mortality, and complications such as organ dysfunction. An improvement in pulse oximeter detection mechanisms that would include all skin pigmentations is therefore much desired to optimize individual healthcare status and minimize disparities in treatment." (page e46078)[57]

- In a Management eBrief published by the Department of Veteran Affairs, Parr, et al. (2023), the authors concluded that "*Pulse oximeters likely overestimate Black or African American patients' blood oxygen saturation level, increasing the risk for unrecognized or 'occult' hypoxemia. Occult hypoxemia occurs to some degree in all races/ethnicities but is likely more common among Black or African American patients compared with White patients. Clinicians should be aware of the risk for occult hypoxemia in patients with darker skin pigmentation. Evidence from hospital and health system settings relevant to VA also suggests that the amount of bias in pulse oximeter readings could vary substantially from patient to patient regardless of their race/ethnicity*."[58]

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, often

approximately ten minutes apart, 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. One study directly compared multiple device models.[59] 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, 2022[4]). 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). Other samples were focused on patients with COVID-19, which may be associated with decreased perfusion.[60] The prevalence of hypoxemia may vary between such groups, confounding comparisons of the occult hypoxemia rate.

Skin pigmentation has been postulated as a contributing factor to pulse oximeter error. However, in RWE studies proper adjustment is difficult because skin pigmentation level is not systematically measured as part of routine care or captured in the EHR systems, which are used as the 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. It is also used inconsistently, most notably whether Hispanic ethnicity is combined with race variable (e.g., White, Black, or Hispanic) or considered separately (e.g., non-Hispanic White, Hispanic White, non-Hispanic Black, Hispanic Black). 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. Since the last literature review in 2022, there were more new real world evidence studies identified; seven out of nine of these were reporting retrospective data. Therefore, there is still 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.

VI. Recognized International Standard

Note: This section is unchanged from the 2022 FDA Executive Summary.

Device Standard for the Pulse Oximeter: Tool for Assuring Safety

FDA recognizes ISO 80601-2-61 Second Edition 2017-12 (Corrected version 20180-02);[xii,xiii] and generally, pulse oximeter 510(k) submissions reference the standard. It is a joint standard[xiv] between ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission).[17] It applies to the basic safety and essential performance of pulse oximeter equipment intended for use on humans.[17] Basic safety[xv] is protection from physical hazards (e.g., shock, burn, crushing), while essential performance[xvi] 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 home healthcare and emergency medical services environments. Hazards inherent in the clinical

[xii] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/detail.cfm?standard__identification_no=37508
[xiii] 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)
[xiv] 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
[xv] 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)
[xvi] 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

interpretation and use of pulse oximetry readings 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 adequacy), 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 patients, 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.

VII. Assessment of Skin Pigmentation

There is no consensus on the best approach to assess skin pigmentation for medical device development. Neither the current FDA guidance nor the recognized ISO standard for pulse oximeters recommends a particular methodology to assess skin pigmentation. In studies supporting 510(k) submissions, skin pigmentation classification has been exclusively subjective. Different scales have been utilized by different manufacturers, most often the Fitzpatrick Skin Type Scale and von Luschan Chromatic Scale. There has also been inconsistent categorization of skin pigmentation data.

The Fitzpatrick Skin Type Scale (FST) is a numerical classification for skin types. It was developed by a dermatologist, Dr. Thomas B. Fitzpatrick, to determine how different skin types react to ultraviolet (UV)

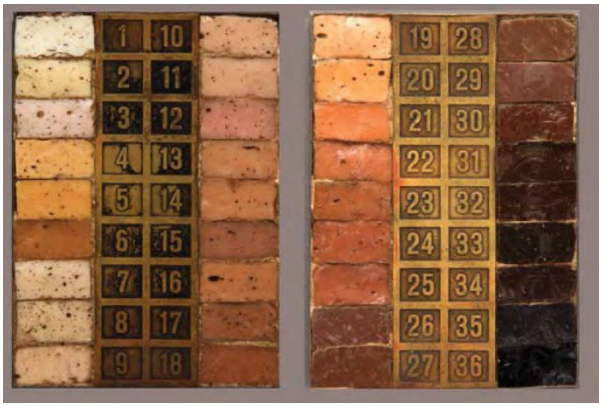
light (i.e., ability to tan when exposed to sunlight).[61] Although this scale was not developed for assessment of skin color, it is commonly used for that purpose. It ranges from Skin Type I (Fair) through VI (Dark Brown), and it is usually assessed visually or self-reported. In contrast, the von Luschan Chromatic Scale was developed by an anthropologist, Felix von Luschan, and was used extensively to establish racial classifications of populations according to skin color.[62] The scale consists of 36 opaque glass tiles which are compared with the patient's skin color, ideally on a site on the body that is less exposed to sun, such as under the arm.

Figure VII-1 below presents both scales, and Table VII-1 presents the 36 von Luschan skin categories in relation to a version of the Fitzpatrick Skin Type Scale.

A. Fitzpatrick Skin Type Scale[63]

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

B. von Luschan Chromatic Scale – Color Tiles[62]



C. Reproduction of von Luschan Chromatic Scale[64]

	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 VII-1 The Fitzpatrick Skin Type Scale and the von Luschan Chromatic Scale

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

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
28 to 34	V: Brown
35 to 36	VI: Dark Brown

The Monk Skin Tone Scale (MST) is a recently introduced 10-value subjective scale (Figure VII-2). The MST has been validated in US populations to capture racial and ethnic diversity in pigmentation,[65] thereby promoting inclusivity. It also has been standardized to color scales such as Commission Internationale d'Eclairage L*a*b* (CIELAB).[66] Additionally, regarding interrater reliability, MST has been shown to have a high intraclass correlation coefficient (ICC) of 0.90-.96, even among a global pool of

raters.[67] MST has been recently adopted by Google[xvii] and other digital platforms for uses including research.

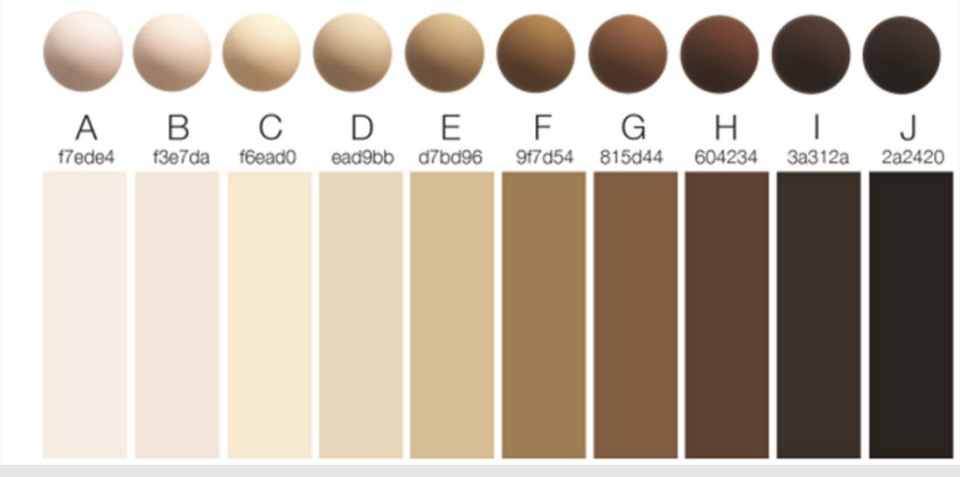


Figure VII-2 The Monk Skin Tone Scale

Narrative Literature Review of Optical Melanometers

Table VII-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 to date. Since the release of the FDA guidance in 2013, the majority of these studies have used binary or ternary descriptive methods (light, medium, dark) without a standardized color scale for differentiating patient populations (for further detail, see the 2022 FDA Executive Summary). 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 skin pigmentation. Standardized color systems such as the Fitzpatrick Skin Type Scale (FST),[68-71] von Luschan,[62] Massey,[38] and Munsell[53] 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.[2,72]

[xvii] https://blog.google/products/search/monk-skin-tone-scale/

Table VII-2 Methods of estimating skin pigmentation to increase quantitative capability (listed in order from least to 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 Type Scale Types I-VI)	Subjective (questionnaire-based), quantitative (categorical); sometimes used as a non-standardized color scale
Color scale (von Luschan, Massey, Monk)	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)

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,[13] 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).[21,73] 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.[2,5,9]

Further complicating the ability to differentiate patient populations based on skin pigmentation is the variation in skin pigmentation at different anatomical sites. The most common pulse oximetry site is the finger – specifically the fingernail of the distal palmar finger, which contains minimal melanin especially when compared to anatomical sites with high pigmentation levels such as the arm or torso. 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.[74] Because a patient'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.[75] 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,[76,77] melasma[78] and psoriasis.[79,80] 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.[81-90]

Colorimetry is the most common and well-standardized approach for objective evaluation of pigmentation. Standard colorimetry approaches are used to measure Commission Internationale d'Eclairage L*a*b* (CIELAB) colorimetric parameters, where L* is luminance, a* is the red/green component, and b* is the yellow/blue component. These objectively measured variables can then be used to calculate ITA, defined as $ITA=180/\pi \times \arctan((L^* - 50)/b^*)$.[43] ITA was developed to provide an objective, continuous, quantitative measure of skin pigmentation. The validity of ITA as a strong correlate of melanin content ($|R|>0.90$) has been confirmed in clinical studies using histological analysis with Fontana-Masson,[91] high performance liquid chromatography,[92] and spectrophotometry[91] of biopsied samples. Overall, the ITA scale ranges from

< -80° (very dark skin) to > 80° (very light skin), but the actual ITA range of human skin depends highly on the body site where it is measured and the practical range tends to be more limited (e.g., -80° to 55° on the cheeks).[93]

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

VIII. Medical Device Adverse Event 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.[xviii] 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 an 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.

A. Search Methodology

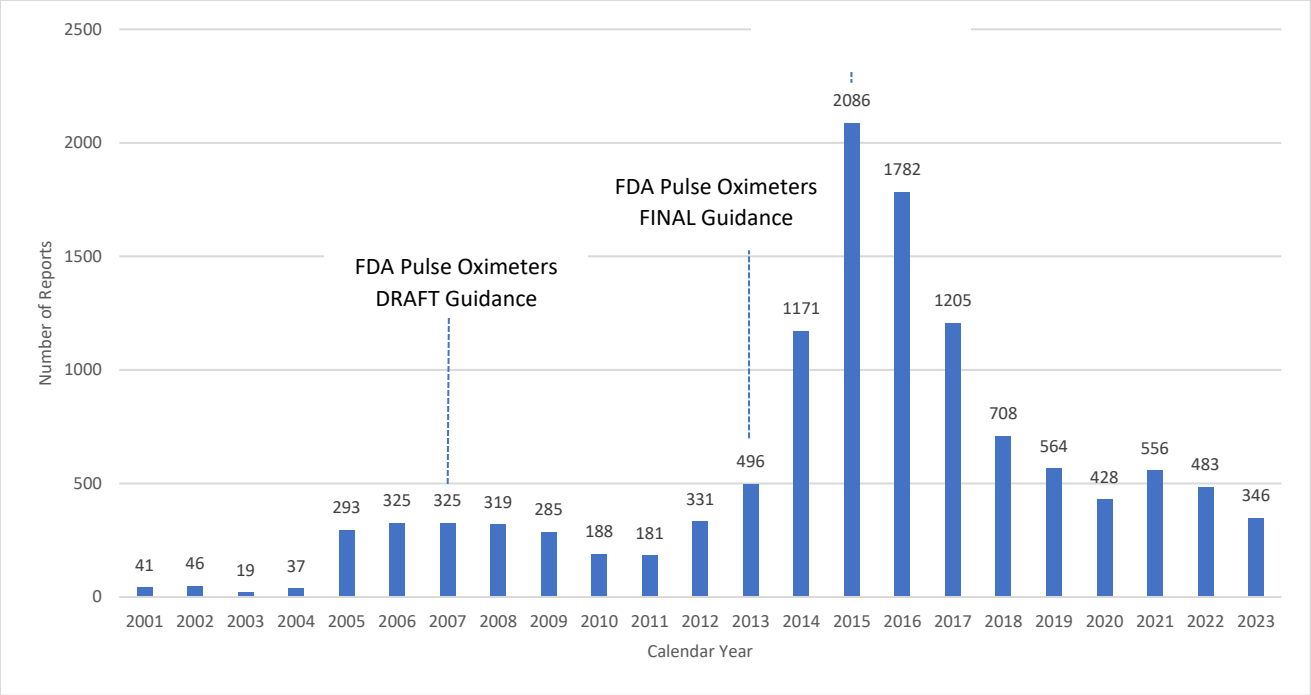
An MDR search was conducted on October 25, 2023, to update the MDR analysis presented in the 2022 executive summary. MDR data for product codes DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter) were searched for reports received between January 1, 2000, to October 25, 2023. Then a text search was used to identify any report with the term “skin”, and a code search was used to identify any report submitted as a death report.

Each report identified through the text search and each death report 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 12,248 adverse event reports. Figure VIII-1 presents the number of reports per year.

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



* Jul 27, 2015: Covidien 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: Masimo 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 VIII-1 Number of Adverse Event Reports Submitted through the MDR System for Product Codes DQA, DQZ, and NFL from January 1, 2000 through October 25, 2023

Most of the adverse event reports were classified as malfunctions (91.5%), followed by serious injury reports (4.4%) and death reports (2.5%) (Table VIII-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.

One hundred and one reports were identified as potentially related to inaccurate SpO_2 readings. Most (84.2%) did not contain sufficient information to ascertain the potential source for the inaccuracy (Table VIII-1). Only three mentioned the patients were African American, and these three were death reports. These reports did 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 VIII-1 Adverse Event Reports received between January 1, 2000 to October 25, 2023 for Product Codes DQA, NLF, and DPZ, n = 12,248

	n	%
Adverse Event Report Type		
Malfunction	11,210	91.5
Serious Injury	539	4.4
Death	305	2.5
Other	138	1.1
Missing	*56*	*0.5*
Potentially related to Inaccurate SpO_2 Reading	101	**14.8***
Mentions African American Race	3	**3.0****
Mentions Skin Pigmentation	3	3.0**
Mentions Low Perfusion	2	2.0**
Reports device continued to provide readings after patient’s death	16	15.8**
Insufficient Information to Assess Reason for Inaccuracy	85	84.2**

*Out of 681 reviewed reports (381 with search term “skin” and 305 death reports, note 5 death reports also mention skin and were only counted once)

** indicates % estimated out of 101; the categories are not exclusive of each other; numbers can add up to more than 101.

Table VIII-2 presents the top ten 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 VIII-2 Top Ten Patient and Device Problems Reported in Adverse Event Reports related to Inaccurate Pulse Oximeter Readings, n = 101

Health Effect Clinical Code	Count*	Device Problem Code	Count*
Death (1802)	62	Incorrect Measurement (1383)	61
No Known Impact or Consequence to Patient (2692)	13	High Readings (2459)	41
Low Oxygen Saturation (2477)	23	False Reading from Device Non-Compliance (1228)	9
No Consequences or Impact to Patient (2199)	8	Low Readings (2460)	62
Insufficient Information (4580)	45	Unable to Obtain Readings (1516)	21
Cyanosis (1798)	5	Device Operates Differently Than Expected (2913)	4
Skin Discoloration (2074)	7	Incorrect, Inadequate or Imprecise Result or Readings (1535)	43
No Information (3190)	3	Material Integrity Problem (2978)	8
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 101

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 text search and review of the subset of MDRs potentially related to inaccuracy, we found that these reports generally lacked sufficient information to determine the reason for the inaccurate reading. 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 during these events. 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). The trends in reporting seen in 2023 were similar to the previous year.

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.[xix] Both studies are real-world, prospective clinical trials to address the main study question of determining 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 oximeter 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. Patients with a range of SaO_2 values are considered for enrollment (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, as well as the Fitzpatrick and von Luschan skin color scales. The Monk Skin Tone Scale was also added in early 2023. Data on other factors that may impact the performance of pulse oximeters (e.g., demographics, carboxyhemoglobin levels [adults], comorbidities, supplemental oxygen therapy, core temperature) are captured from the electronic health records.

The adult study[xx] is being conducted in the UCSF Health Medical System. The study population includes patients 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. Enrollment began in August 2022 and is currently ongoing as of December 2023.

The pediatric study[xxi] is being conducted at The Heart Center at Stanford's Lucile Packard Children's Hospital. It will include patients 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. Enrollment began in January 2023 and is currently ongoing as of December 2023.

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

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

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

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

X. Considerations related to FDA's 2013 Guidance Document on Pulse Oximeters

In an effort to reduce the disparate performance of pulse oximeters, the Agency is considering updates to the 2013 Guidance's recommendations for premarket clinical study of these devices. As discussed below, a proposed clinical trial design has been offered for discussion purposes and includes the entire range of skin pigmentation and also seeks to account for race and ethnicity. This approach is intended to cover pulse oximeters intended for medical purposes that require premarket authorization, and would apply to devices intended for both prescription and OTC use. FDA has outlined this approach in the recently published discussion paper entitled "An Approach for Improving Performance Evaluation of Pulse Oximeter Devices Taking Into Consideration Skin Pigmentation, Race and Ethnicity"; the approach is summarized below.

A. Clinical Study Design

The Agency is considering the utility of a clinical study design that includes a larger number of participants (compared to the current guidance)[xxii] using an assessment that has been validated to capture race and ethnicity diversity in pigmentations within the US. Thus, the goal of this approach is to promote a more inclusive and representative study of the intended patient population with respect to skin pigmentation, race and ethnicity.

The proposed clinical study would collect 480 paired datapoints from a minimum of 24 participants that span the entire Monk Skin Tone (MST) scale. MST has been validated to capture race and ethnicity diversity in pigmentations within the US.[65] The forehead, a common location for perception and assessment of pigmentation,[94-96] has a wide range of pigmentation levels that allow for MST assessment and stratification of participants into one of ten MST values. Specifically, the Agency is considering the following elements as part of the overall clinical study approach that may meet the Agency's public health goals:

- ≥ 25% of participants in each of the MST cohorts 1-4, 5-7, and 8-10 (e.g., 6 participants in MST 1-4, 9 in MST 5-7, and 9 in MST 8-10); and

[xxii] See Pulse Oximeters – Premarket Notification Submissions [510(k)s], Guidance for Industry and Food and Drug Administration Staff, which recommends 200 paired datapoints of SpO_2 and SaO_2 in at least 10 study participants with a range of skin pigmentations, including at least 2 darkly pigmented participants or 15% of the pool of participants, whichever is larger, www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug.

- At least 1 participant or ≥ 15% of participants within each cohort, whichever is larger, should be included for each MST value (e.g., ≥ 1 participant for MST 10 for a sample size of 24 participants).

As part of the clinical study, each participant would have their skin pigmentation measured by an instrument at the emitter site of the pulse oximeter sensor placement (hereinafter, sensor site), and the result categorized by Individual Typology Angle (ITA). The objective ITA measurement of sensor site pigmentation will likely improve consistency in skin pigmentation evaluation leading to a better understanding of any performance differences that may be due to skin pigmentation.

Following the categorization by MST and ITA and collection of the paired datapoints, and given the current limits of the pulse oximeter technology, the Agency is considering the following approach to analyze the premarket clinical performance of the pulse oximeters:

- Define non-disparate performance as the lack of variation in SpO_2 bias across ITA and MST levels, where SpO_2 bias is the mean of the difference $D = SpO_2 – SaO_2$.
- Recommend that the estimate of the absolute difference in SpO_2 bias across ITA and MST levels be < 1.5% when $SaO_2 > 85\%$, and < 3.5% when $70\% < SaO_2 \leq 85\%$.
- Assess results utilizing a mixed effects model that would include ITA, SaO_2, as predictors and ITA by SaO_2 interaction, and random effects for participant, participant by ITA interaction, and participant by SaO_2 interaction (see Section B).

To reduce residual variation, a second step in assessing non-disparate performance would be conducted with a similar mixed effects model on data using MST and SaO_2 as predictors (hereinafter, called the MST-derived performance analysis).

Under the approach being considered by the Agency, the estimated maximum absolute difference in SpO_2 bias in each analysis (between ITAs in the ITA-derived performance analysis, and between MSTs in the MST-derived performance analysis) would need to be < 1.5% when $SaO_2 > 85\%$, and < 3.5% when $70\% < SaO_2 \leq 85\%$.

B. Discussion of Considerations for Possible Components of Clinical Study Design

ITA Scale and ITA-derived Performance Analyses

Measuring ITA values at the sensor site is expected to improve the overall analysis of performance across levels of skin pigmentation. Due to concerns about inaccurate binning or grouping of data, ITA data would not be stratified into pigmentation cohorts for analyses of non-disparate performance. The mixed effects model in the ITA-derived performance analysis may be used to estimate the expected value of D (where $D = SpO_2 – SaO_2$) at low and high values of ITA at the midpoints $SaO_2 = 77.5\%$ and $SaO_2 = 92.5\%$, of the intervals 70-85% and 85-100%, respectively. The model should be checked for goodness-of-fit to the data.

MST Scale and MST-derived Performance Analyses

By considering 25% or more participants per each MST cohort (i.e., 1-4, 5-7, 8-10), the proposed methodology is intended to ensure inclusion of a sufficient sample size of participants with light, medium and dark pigmentation. This would result in an overall increase in the number of darkly pigmented individuals from ≥ 2 (in the guidance) to ≥ 6, or ≥ 15% (in the guidance) to ≥ 25% (i.e., MST 8-10); whichever is larger. It would also provide that each MST value has at least one participant and that pigmentation ranges include the entire MST scale.

For the MST-derived performance analysis, MST is included as a predictor of bias ($D=SpO_2 – SaO_2$) in the mixed effects model (See Section A. Clinical Study Design). We note that the power to detect MST effects on bias could potentially be increased by additionally including ITA as a predictor in the MST mixed effects model because this may improve model fit and thereby reduce residual variation. Similarly, in the same expanded model, the power to detect ITA effects on bias may benefit from adjustment for MST effects.

Considering Race and Ethnicity in Performance Analyses

Variation in pulse oximeter performance among racial and ethnicity groups may be difficult to differentiate from variation in performance due to skin pigmentation at the sensor site. At this time, it is uncertain whether pulse oximeter performance is disparate between individuals from different race and ethnicity groups with the same level of sensor site pigmentation. It is important to note that MST,

though validated for capturing race and ethnicity diversity in pigmentations within the US,[65] is not a proxy for racial and ethnic diversity.

Race and ethnicity categories may also be included as predictors in the mixed effects model. However, we note that some race and ethnicity categories or some combinations of them may not be represented equally in some datasets (or not at all). In other words, some care may be needed in implementing a statistical model for evaluating race and ethnicity effects on pulse oximeter bias.

One hypothesis that has been proposed to help explain why race and ethnicity may give disparate pulse oximeter performance is that these variables are correlated with melanin pigment, which can absorb significant levels of light at the sensor site and impact the spectral content of detected signals. This interaction may cause the oximeter to generate inaccurate results.[2,15,97] Variations in the number, size, and aggregation of melanosomes within the melanocyte and keratinocyte contribute to racial and ethnic differences in pigmentation.[98] However, ranges and levels of pigmentation can vary not only within one race and ethnicity group but can also overlap across race and ethnicity groups. Among the subjective pigmentation scales, MST is known to agree with a person's self-identified skin tone better than other subjective scales, which makes MST more inclusive across race and more representative of US demographics.[65] However, it remains uncertain whether pulse oximeter performance is disparate between individuals from different race and ethnicity groups with the same level of sensor site pigmentation.

The Panel will be asked to discuss FDA's proposed approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters submitted for premarket review, including:

- *A more inclusive and representative trial design;*
- *Defining non-disparate performance; and*
- *Considerations for studies of Over the Counter (OTC) devices used for medical purposes.*

XI. Summary

Although pulse oximeters provide clinical benefit for patients through the non-invasive estimation of blood oxygen saturation, the risks associated with inaccurate pulse oximeter readings must be well understood. The COVID-19 pandemic, which resulted in increased use of pulse oximeters in the hospital and home settings, highlighted the limitations and risks associated with this technology, particularly in

patients with darker skin pigmentation. 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 these limitations to avoid delays in treatment and adverse patient outcomes. There is an increasing amount of 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 (but not limited to) nail polish, motion, perfusion index, carboxyhemoglobin, and prevalence of hypoxemia across different groups. Furthermore, standardization of skin pigmentation assessment, reporting and categorization of skin pigmentation data for analysis of SpO_2 bias, remain as challenges for the assessment of pulse oximeter performance across skin pigmentation, racial and ethnic groups.

After hearing from patients, regulators, researchers, and industry, the Committee will be asked to discuss the proposed approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters for medical purposes submitted for premarket review, taking into consideration a patient's skin pigmentation, and patient-reported race and ethnicity. The Committee will discuss the type and amount of data that should be provided by manufacturers to FDA to evaluate the performance of pulse oximeters submitted for premarket review, including prescription and over-the-counter indications, for labeling considerations, and ways to help guide other regulatory actions as needed.

XII. References

1. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. *N Engl J Med*. Dec 17 2020;383(25):2477-2478. doi:10.1056/NEJMc2029240
2. Okunlola OE, Lipnick MS, Batchelder PB, Bernstein M, Feiner JR, Bickler PE. Pulse Oximeter Performance, Racial Inequity, and the Work Ahead. *Respir Care*. Feb 2022;67(2):252-257. doi:10.4187/respcare.09795
3. Andrist E, Nuppnau M, Barbaro RP, Valley TS, Sjoding MW. Association of Race With Pulse Oximetry Accuracy in Hospitalized Children. *JAMA Netw Open*. Mar 1 2022;5(3):e224584. doi:10.1001/jamanetworkopen.2022.4584
4. Burnett GW, Stannard B, Wax DB, et al. Self-reported Race/Ethnicity and Intraoperative Occult Hypoxemia: A Retrospective Cohort Study. *Anesthesiology*. May 1 2022;136(5):688-696. doi:10.1097/ALN.0000000000004153
5. 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)doi:10.3390/S22093402
6. 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*. May 31 2022;182(7):730-738. doi:10.1001/jamainternmed.2022.1906
7. 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*. Feb 1 2022;50(2):204-211. doi:10.1097/CCM.0000000000005394
8. 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*. Apr 2022;161(4):971-978. doi:10.1016/j.chest.2021.09.025
9. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J Perinatol*. Jan 2022;42(1):79-85. doi:10.1038/s41372-021-01230-3
10. 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*. Nov 1 2021;4(11):e2131674. doi:10.1001/jamanetworkopen.2021.31674
11. Jubran A. Pulse oximetry. *Crit Care*. 1999;3(2):R11-R17. doi:10.1186/cc341
12. Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. *Med Devices (Auckl)*. 2014;7:231-9. doi:10.2147/MDER.S47319
13. Jacques SL. Quick analysis of optical spectra to quantify epidermal melanin and papillary dermal blood content of skin. *J Biophotonics*. Apr 2015;8(4):309-16. doi:10.1002/jbio.201400103
14. Feiner JR, Bickler PE, Mannheimer PD. Accuracy of methemoglobin detection by pulse CO-oximetry during hypoxia. *Anesth Analg*. Jul 2010;111(1):143-8. doi:10.1213/ANE.0b013e3181c91bb6
15. Gudelunas MK, Lipnick M, Hendrickson C, et al. Low perfusion and missed diagnosis of hypoxemia by pulse oximetry in darkly pigmented skin: A prospective study. *medRxiv*. 2022;doi:10.1101/2022.10.19.22281282
16. 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*. May 1 2022;13(5):2909-2928. doi:10.1364/BOE.454020
17. ISO, IEC. Medical Electrical Equipment: Particular Requirements for Basic Safety and Essential Performance of Pulse Oximeter Equipment. 80601-2-61. Switzerland 2017.
18. Bickler P, Tremper KK. The Pulse Oximeter Is Amazing, but Not Perfect. *Anesthesiology*. May 1 2022;136(5):670-671. doi:10.1097/ALN.0000000000004171

19. Pologe JA, Arnold DH, Delianides TP. Multi-analyte calibration and verification of a multi-parameter laser-based pulse oximeter. *J Clin Monit Comput*. Apr 2022;36(2):579-586. doi:10.1007/s10877-021-00704-1
20. Feiner JR, Rollins MD, Sall JW, Eilers H, Au P, Bickler PE. Accuracy of carboxyhemoglobin detection by pulse CO-oximetry during hypoxemia. *Anesth Analg*. Oct 2013;117(4):847-858. doi:10.1213/ANE.0b013e31828610a0
21. 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*. Aug 2000;106(2):E17. doi:10.1542/peds.106.2.e17
22. Guo C, Huang W, Chang H, Hsieh T. Calibrating Oxygen Saturation Measurements for Different Skin Colors Using the Individual Typology Angle. *IEEE Sensors Journal*. 2023;PP(99):1. doi:10.1109/JSEN.2023.3288151
23. Jakachira R, Diouf M, Lin Z, et al. Single-wavelength, single-shot pulse oximetry using an LED-generated vector beam. *Opt Express*. Jul 18 2022;30(15):27293-27303. doi:10.1364/OE.461871
24. 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*. Jun 15 1999;24(12):829-31. doi:10.1364/ol.24.000829
25. Kirchner T, Jaeger M, Frenz M. Machine learning enabled multiple illumination quantitative optoacoustic oximetry imaging in humans. *Biomed Opt Express*. May 1 2022;13(5):2655-2667. doi:10.1364/BOE.455514
26. Mantri Y, Jokerst JV. Impact of skin tone on photoacoustic oximetry and tools to minimize bias. *Biomed Opt Express*. Feb 1 2022;13(2):875-887. doi:10.1364/BOE.450224
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. Feb 8 1986;1(8476):307-10.
28. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. Feb 2005;34(1):215-20. doi:10.1093/ije/dyh299
29. 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*. Oct 2015;44(5):1748. doi:10.1093/ije/dyv161
30. Bland JM, Altman DG. Regression towards the mean. *BMJ*. Jun 4 1994;308(6942):1499. doi:10.1136/bmj.308.6942.1499
31. Bland JM, Altman DG. Some examples of regression towards the mean. *BMJ*. Sep 24 1994;309(6957):780. doi:10.1136/bmj.309.6957.780
32. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. Oct 21 1995;346(8982):1085-7. doi:10.1016/s0140-6736(95)91748-9
33. Johnson WD, George VT. Effect of regression to the mean in the presence of within-subject variability. *Stat Med*. Aug 1991;10(8):1295-302. doi:10.1002/sim.4780100812
34. (NIST) NIoSaT. *Engineering Statistics Handbook*.
35. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. Jun 1999;8(2):135-60. doi:10.1177/096228029900800204
36. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17(4):571-82. doi:10.1080/10543400701329422
37. Lin LI. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Stat Med*. Jan 30 2000;19(2):255-70. doi:10.1002/(sici)1097-0258(20000130)19:2<255::aid-sim293>3.0.co;2-8
38. 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*. Apr 2016;17(4):315-20. doi:10.1097/PCC.0000000000000660

39. Harris BU, Stewart S, Verma A, et al. Accuracy of a portable pulse oximeter in monitoring hypoxemic infants with cyanotic heart disease. *Cardiol Young*. Aug 2019;29(8):1025-1029. doi:10.1017/S1047951119001355
40. Ross PA, Newth CJ, Khemani RG. Accuracy of pulse oximetry in children. *Pediatrics*. Jan 2014;133(1):22-9. doi:10.1542/peds.2013-1760
41. 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*. Mar - Apr 2018;47(2):93-99. doi:10.1016/j.hrtlng.2017.12.002
42. Institute CaLS. *EP09-A3: Measurement Procedure Comparison and Bias Estimation Using Patient Samples, 3rd Edition*. 2013.
43. Del Bino S, Sok J, Bessac E, Bernerd F. Relationship between skin response to ultraviolet exposure and skin color type. *Pigment Cell Res*. 2006;19(6):606-14. doi:10.1111/j.1600-0749.2006.00338.x
44. Gadrey SM, Mohanty P, Haughey SP, et al. Overt and occult hypoxemia in patients hospitalized with novel coronavirus disease 2019. *medRxiv*. Jun 16 2022;doi:10.1101/2022.06.14.22276166
45. 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;
46. Seitz KP, Wang L, Casey JD, et al. Pulse Oximetry and Race in Critically Ill Adults. *Crit Care Explor*. Sep 2022;4(9):e0758. doi:10.1097/CCE.0000000000000758
47. 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*. Aug 16 2022;20(1):267. doi:10.1186/s12916-022-02452-8
48. 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*. Jun 3 2022;doi:10.4187/respcare.09769
49. Leeb G, Auchus I, Law T, et al. The Performance of 11 Fingertip Pulse Oximeters During Hypoxemia in Healthy Human Subjects with Varied, Quantified Skin Pigment [Preprint]. *Lancet*. 2023;
50. Fawzy A, Ali H, Dziedzic PH, et al. Skin Pigmentation and Pulse Oximeter Accuracy in the Intensive Care Unit: a Pilot Prospective Study. *medRxiv*. Nov 17 2023;doi:10.1101/2023.11.16.23298645
51. Ruppel H, Makeneni S, Faerber JA, et al. Evaluating the Accuracy of Pulse Oximetry in Children According to Race. *JAMA Pediatr*. May 1 2023;177(5):540-543. doi:10.1001/jamapediatrics.2023.0071
52. Savorgnan F, Hassan A, Borges N, Acosta S. Pulse Oximetry and Arterial Saturation Difference in Pediatric COVID-19 Patients: Retrospective Analysis by Race. *Pediatr Crit Care Med*. Jun 1 2023;24(6):458-462. doi:10.1097/PCC.0000000000003208
53. 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*. Mar 2017;182:375-377 e2. doi:10.1016/j.jpeds.2016.11.043
54. 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*. Aug 2022;36(4):961-973. doi:10.1007/s10877-021-00797-8
55. Jamali H, Castillo LT, Morgan CC, et al. Racial Disparity in Oxygen Saturation Measurements by Pulse Oximetry: Evidence and Implications. *Ann Am Thorac Soc*. Dec 2022;19(12):1951-1964. doi:10.1513/AnnalsATS.202203-270CME
56. Al-Halawani R, Charlton PH, Qassem M, Kyriacou PA. A review of the effect of skin pigmentation on pulse oximeter accuracy. *Physiol Meas*. Jun 1 2023;44(5)doi:10.1088/1361-6579/acd51a
57. Aoki KC, Barrant M, Gai MJ, Handal M, Xu V, Mayrovitz HN. Impacts of Skin Color and Hypoxemia on Noninvasive Assessment of Peripheral Blood Oxygen Saturation: A Scoping Review. *Cureus*. Sep 2023;15(9):e46078. doi:10.7759/cureus.46078

58. Parr NJ, Beech EH, Young S. *Differential Pulse Oximeter Accuracy, Occult Hypoxemia Prevalence, and Clinical Outcomes by Patient Race/Ethnicity: A Systematic Review*. Vol. VA ESP Project #09-199. 2023.
59. Blanchet MA, Mercier G, Delobel A, et al. Accuracy of Multiple Pulse Oximeters in Stable Critically Ill Patients. *Respir Care*. May 2023;68(5):565-574. doi:10.4187/respcare.10582
60. Fox TH, Mazalewski WR, Tran HS, et al. Discordance Between Invasive and NonInvasive Oxygen Saturation in Critically Ill COVID-19 Patients. *J Intensive Care Med*. Jul 7 2023:8850666231186947. doi:10.1177/08850666231186947
61. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol*. Jun 1988;124(6):869-71. doi:10.1001/archderm.124.6.869
62. 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*. Jun 2013;151(2):325-30. doi:10.1002/ajpa.22274
63. 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. doi:10.1371/journal.pone.0241843
64. 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. doi:10.1155/2021/8899931
65. Heldreth CM, Monk EP, Clark AT, Ricco S, Schumann C, Eyee X. Which Skin Tone Measures are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence [preprint]. 2023;
66. ISO. 9-127 CIE ISO 11664-4 First edition 2019-06 Colorimetry - Part 4: CIE 1976 L* a* b* colour space. 2019.
67. Schumann C, Olanubi GO, Wright A, Monk Jr. E, Heldreth C, Ricco S. Consensus and Subjectivity of Skin Tone Annotation for ML Fairness [preprint]. 2023;
68. 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*. May 2018;46(3):297-303. doi:10.1177/0310057X1804600307
69. 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*. Sep 2021;8(1)doi:10.1136/bmjresp-2021-000939
70. 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*. Feb 26 2019;109(3):154-158. doi:10.7196/SAMJ.2019.v109i3.13633
71. 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*. Feb 2022;9(1)doi:10.1136/bmjresp-2021-001159
72. Norton HL. Variation in pulse oximetry readings: melanin, not ethnicity, is the appropriate variable to use when investigating bias. *Anaesthesia*. Mar 2022;77(3):354-355. doi:10.1111/anae.15620
73. Mendenhall MJ, Nunez AS, Martin RK. Human skin detection in the visible and near infrared. *Appl Opt*. Dec 10 2015;54(35):10559-70. doi:10.1364/AO.54.010559
74. Gunes P, Goktay F. Melanocytic Lesions of the Nail Unit. *Dermatopathology (Basel)*. Jul-Sep 2018;5(3):98-107. doi:10.1159/000490557
75. 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*. Jul 1996;35(1):1-10. doi:10.1111/j.1600-0536.1996.tb02258.x
76. Lee KC, Dretzke J, Grover L, Logan A, Moiemen N. A systematic review of objective burn scar measurements. *Burns Trauma*. 2016;4:14. doi:10.1186/s41038-016-0036-x

77. 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*. Mar 2004;30(2):103-7. doi:10.1016/j.burns.2003.09.029
78. Kanellis VG. Objective quantification of melasma severity using melanometers to quantify melanin pigmentation. *Biophys Rev*. Oct 2020;12(5):1139-1140. doi:10.1007/s12551-020-00760-y
79. 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*. Aug 1995;1(3):140-4. doi:10.1111/j.1600-0846.1995.tb00034.x
80. 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*. Apr 23 2009;73(4):936-8. doi:10.1271/bbb.80725
81. 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.
82. Khalid AT, Moore CG, Hall C, et al. Utility of sun-reactive skin typing and melanin index for discerning vitamin D deficiency. *Pediatr Res*. Sep 2017;82(3):444-451. doi:10.1038/pr.2017.114
83. 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*. Nov 2020;26(6):923-931. doi:10.1111/srt.12895
84. 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*. Mar 10 2017;9(3)doi:10.3390/nu9030260
85. 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*. Sep 2017;30(5):488-492. doi:10.1111/pcmr.12598
86. 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*. Apr 2018;43(3):274-280. doi:10.1111/ced.13365
87. 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*. Summer 2019;29(3):505-512. doi:10.18865/ed.29.3.505
88. 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.
89. 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*. Jul 2020;140(7):1418-1426 e1. doi:10.1016/j.jid.2019.11.019
90. Bailey SH, Oni G, Brown SA, et al. The use of non-invasive instruments in characterizing human facial and abdominal skin. *Lasers Surg Med*. Feb 2012;44(2):131-42. doi:10.1002/lsm.21147
91. Del Bino S, Ito S, Sok J, et al. Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. *Pigment Cell Melanoma Res*. 2015;28(6):707-17.
92. Ito S, Del Bino S, Hirobe T, Wakamatsu K. Improved HPLC Conditions to Determine Eumelanin and Pheomelanin Contents in Biological Samples Using an Ion Pair Reagent. *Int J Mol Sci*. 2020;21(14):5134.
93. Del Bino S. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br J Dermatol*. 2013;169(Suppl 3):33-40.
94. Hsiao JH, Cottrell G. Two fixations suffice in face recognition. *Psychol Sci*. 2008;19(10):998-1006.
95. Hannon L, DeFina R. Reliability Concerns in Measuring Respondent Skin Tone by Interviewer Observation. *Public Opin Q*. 2016;80(2):534-41.

96. Hugenberg K, Wilson J. Faces are central to social cognition. In: Carlston D, ed. *The Oxford Handbook of Social Cognition*. Oxford University Press; 2013:167-193.
97. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. Apr 2005;102(4):715-9. doi:10.1097/00000542-200504000-00004
98. Taylor SC. Skin of color: biology, structure, function, and implications for dermatologic disease. *J Am Acad Dermatol*. Feb 2002;46(2 Suppl Understanding):S41-62. doi:10.1067/mjd.2002.120790
99. 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*. Jun 2022;48:101428. doi:10.1016/j.eclinm.2022.101428
100. 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*. Apr 2022;59(4)doi:10.1183/13993003.03246-2021
101. Crooks CJ, West J, Morling JR, et al. Differential pulse oximetry readings between ethnic groups and delayed transfer to intensive care units. *QJM*. Feb 14 2023;116(1):63-67. doi:10.1093/qjmed/hcac218
102. Fawzy A, Wu TD, Wang K, et al. Clinical Outcomes Associated With Overestimation of Oxygen Saturation by Pulse Oximetry in Patients Hospitalized With COVID-19. *JAMA Netw Open*. Aug 1 2023;6(8):e2330856. doi:10.1001/jamanetworkopen.2023.30856
103. 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*. Aug 1 2022;182(8):849-858. doi:10.1001/jamainternmed.2022.2587
104. Kalra A, Shou BL, Zhao D, et al. Racial and ethnical discrepancy in hypoxemia detection in patients on extracorporeal membrane oxygenation. *JTCVS Open*. Jun 2023;14:145-170. doi:10.1016/j.xjon.2023.02.011
105. Kalra A, Wilcox C, Holmes SD, et al. Characterizing the Racial Discrepancy in Hypoxemia Detection in VV-ECMO: An ELSO Registry Analysis. *Res Sq*. Nov 17 2023;doi:10.21203/rs.3.rs-3617237/v1
106. 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;
107. 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*. Jul 6 2022;378:e069775. doi:10.1136/bmj-2021-069775
108. 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*. Feb 2022;77(2):143-152. doi:10.1111/anae.15581
109. 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. doi:10.1155/2018/3521738
110. Barker SJ, Wilson WC. Racial effects on Masimo pulse oximetry: a laboratory study. *J Clin Monit Comput*. Apr 2023;37(2):567-574. doi:10.1007/s10877-022-00927-w
111. Giuliano KK, Bilkovski RN, Beard J, Lamminmaki S. Comparative analysis of signal accuracy of three SpO(2) monitors during motion and low perfusion conditions. *J Clin Monit Comput*. Dec 2023;37(6):1451-1461. doi:10.1007/s10877-023-01029-x
112. Khanna AK, Beard J, Lamminmaki S, Narvainen J, Antaki N, Yapici HO. Assessment of skin pigmentation-related bias in pulse oximetry readings among adults. *J Clin Monit Comput*. Oct 26 2023;doi:10.1007/s10877-023-01095-1

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
*Blanchet 2023[59]	Canada, 2021, ≥18yo, ICU	193 1,055	Fitzpatrick: I n=85 II n=118 III n=6 IV n=1 V n=0 VI n=1	Bias: SpO_2-SaO_2 Hypoxemia: SaO_2 < 90%	Bias (95% CI): Nonin **-3.1 (-3.4, -2.8)** Nellcor -0.3 (-0.5, 0) Masimo -0.2 (-0.5, 0.1) Philips **0.9 (0.7, 1.2)** Detection of hypoxemia: Nonin 92% Nellcor 33% Masimo 42% Philips 17% Not enough data to assess impact of skin pigmentation
Ebmeier 2018[68]	Aus/NZ, 2015, ≥16yo, ICU	394 394	Fitzpatrick: I-II ("light") III-IV ("medium") V-VI ("dark") n's and %s not reported	Bias: Bland-Altman	Unadjusted regression coefficient (95% CI) Light: Reference Medium: 0.9 (0.4, 1.3) Dark: **2.4 (1.2, 3.6)**
*Fawzy 2023 [pre-print][50]	US, 2022, adults (range 34.7-79.4 yrs.), ICU	12 400	Light (ITA ≥ 30^o) Dark (ITA < -30^o)	Bias: SpO_2-SaO_2 Arms	Bias: Light 0.34% Dark 1.05% Difference after adjustment = 1.0% (**95% CI: 0.25, 1.76%)** Arms: Light 1.97% (95% CI: 1.76, 2.17%)

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
					Dark 4.15% (95% CI: 2.35, 5.72%) OH: Light 0.4% Dark 2.1%
Foglia 2017[53]	US, 2013-2015, infants, Cyanotic Congenital Heart Disease	35 35	Munsell System Soil Color Chart: Light n=21 (60%) Dark n=14 (40%)	Bias: SpO_2-SaO_2	N.S.
Harskamp 2021[69]	Netherlands, 2020, ≥18yo, ICU	35 234	Fitzpatrick: I-III n=30 (85.7%) IV-VI n=5 (14.3%)	Mean bias in SpO_2 Accuracy measured by Arms, and Mean Absolute Error (MAE) Diagnostic accuracy for hypoxemia (SaO_2 < 90%)	Mean bias range: -0.6 to -4.8 None of the pulse oximeters met Arms < 3% in SaO_2 range of 70-100%. MAE range: 2.3 to 5.1 and 5 of the pulse oximeters met < 3% Darker skin complexion was associated with poorer SpO_2 performance. Negative Predictive Value (NPP) for hypoxemia diagnosis: 98% to 99% Positive Predictive Value (PPV): 11% to 30%
Henry 2022[7]	US, 2018-2020, ≥18yo, ICU or surgical	26,603 128,285	Race: White n=24,493 Black n=1,263 Asian n=574 American Indian n=273	OH: SaO_2 < 88% despite concurrent SpO_2 > 92%	Adjusted OR (95%CI) White: Reference Black: **1.65 (1.28, 2.14)** Asian: 1.53 (0.95, 2,47) American Indian: 1.31 (0.80, 2.16) OH associated with mortality **OR=2.96 (1.20, 7.28)**
Seitz 2022[46]	US, 2018-2021, adults (no age cutoff reported), patients receiving invasive mechanical ventilation in medical	1,024 5,557	Race: White n=4,788 pairs (86%) Black n=769 pairs (14%)	OH: SaO_2 < 88% despite SpO_2 between 92%-96% Hyperoxemia: PaO_2 > 150mmHg despite SpO_2 between 92%-96%	OH: White: 1.1% (0.7, 1.7) Black: 3.5% (1.6, 6.6) Hyperoxemia: White: 2.4% (1.8, 3.2)

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
	ICU (excluding COVID-19)				Black: 4.7% (2.5, 8.1)
Smith 2019[70]	South Africa, years NR, ≥18yo, surgical	220 220	Fitzpatrick: I n=12 (5.5%) II n=28 (12.7%) III n=69 (31.4%) IV n=45 (20.5%) V n=28 (12.7%) VI n= 38 (17.3%)	Bias: Bland-Altman	N.S.
Stell 2022[71]	UK, years NR, ≥19yo, COVID-19	50 915	Fitzpatrick: I n=6 II n=21 III n=9 IV n=5 V n=7 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

Asterisk = published since August 9, 2022 (cutoff date of literature search presented in the 2022 executive summary); 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[3]	US, 2015-2020, children (≤17yo), inpatient	1,061 9,023	Race: White n=878 (82.8%) Black n=183 (17.2%)	OH: SaO_2 < 88% despite SpO_2 > 92%	Adjusted OR (95% CI) White: Reference Black: **2.16 (1.36, 3.44)*** Black: **1.79 (1.07, 3.02)**** *paired measures as unit of analysis **patients as unit of analysis
Bangash 2022[99]	UK, 2017-2021, adults, inpatient	16,818 20,231	Race: White n=13,649 (81.2%) Asian n=1,965 (11.7%) Black n=674 (4.0%) Other n=530 (3.2%)	OH: SaO_2 < 94% despite SpO_2 ≥ 94%	Adjusted OR (95% CI) White: Reference Asian: NR Black: **1.47 (1.09, 1.98)** Other: NR
Burnett 2022[4]	US, 2008-2019, ≥18yo, patients receiving anesthetic	46,253 151,070	Race/ethnicity: White n=22,089 Black n=5,177 Asian n=2,612 Hispanic n=6,304 Other n=10,071	OH: SaO_2 < 88% despite SpO_2 ≥ 92%	Adjusted OR (95% CI) White: Reference Black: **1.44 (1.11, 1.87)** Asian: 0.77 (0.51, 1.17) Hispanic: **1.31 (1.03, 1.68)** Other: 1.24 (1.00, 1.53)
Chesley 2022[48]	US, 2019-2021, adults (no age cutoff reported), ICU	7,693 105,467	Race/ethnicity: White n=4,621 (60%) Black n=1,919 (25%) Latinx n=226 (3%) Asian/PI n=239 (3%) Indigenous n=17 (0.2%) Other n=220 (3%) Unknown n=451 (6%)	OH: SaO_2 < 88% despite SpO_2 between 92%-96%	Adjusted OR (95% CI) White: Reference Black: **2.84 (2.44, 3.30)** Latinx: **1.69 (1.22, 2.34)** Asian/PI: **1.64 (1.15, 2.34)** Indigenous: 0.51 (0.07, 3.72) Other: 1.41 (0.99, 1.99)
Crooks 2022[100]	UK, 2020-2021, no age limit reported,	2,997 5,374	Race: White n=3,946 Black n=151	Mean difference SpO_2-SaO_2	Adjusted White: Reference 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 Mixed n=36 NR n=995		Asian: **+1.9% (+0.6, +3.2%)** Mixed: +3.2% (-0.1, +6.6%)
*Crooks 2023[101]	UK, 2020-2021, adults, inpatients with COVID-19	748	Race: White n=420 Unrecorded n=205 Black/mixed n=48 Indian/Pakistani n=53 Other n=22	Median SpO_2 (Interquartile range) Median SaO_2 (Interquartile range)	Median SpO_2 (IQR): White: 94 (92, 26) Unrecorded: 94 (91, 96) Black/mixed: 94 (91, 95) Indian/Pakistani: 94 (92, 96) Other: 94 (89, 95) p=0.51 Median SaO_2 (IQR): White: 94.4 (90.4, 97.5) Unrecorded: 94.2 (90.2, 97.3) Black/mixed: 93.0 (89.8, 95.5) Indian/Pakistani: 91.6 (89.3, 94.4) Other: 92.0 (89.3, 94.8) **p=0.005**
Fawzy 2022[6]	US, 2020-2021, no age limit reported, Emergency department visit or hospitalized for COVID-19	1,216 32,282	Race/ethnicity: Non-Hispanic White n=460 Black n=478 Asian n=63 Non-Black Hispanic: n=215	OH: SaO_2 < 88% despite concurrent SpO_2 of 92% to 96% Treatment Initiation: SpO2 ≤94% or use of supplemental oxygen	OH: White: 17.2% Black: 28.5% Asian: 30.2% Non-Black Hispanic: 29.8% Treatment Initiation (hazard ratio, 95%CI) Black: 0.71 (0.63, 0.80) Non-Black Hispanic: 0.77 (0.66, 0.89) Asian: 0.97 (0.62, 1.5)
*Fawzy 2023[102]	US, 2020-2021, no age limit, hospitalized for COVID-19	24,504 213,229	Race/ethnicity: Non-Hispanic White (NHW) n=10,133 Black n=3,922 Hispanic n=7,895 Other n=2,554	OH: SaO_2 < 88% despite 92% ≥ SpO_2 ≥ 96% Unrecognized need for treatment: SaO_2 < 94% despite SpO_2 ≥ 94%	OH: NHW: 13.0% Black: 18.3% Hispanic: 20.9% Other: 19.7%

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
					Unrecognized need for treatment [Adjusted OR (95% CI)]: NHW: Reference Black: **1.46 (1.23, 1.72)** Hispanic: **1.18 (1.01, 1.39)** Other: 1.23 (1.00, 1.52)
*Fox 2023[60]	US, 2020, ≥18yo, ICU	263 (179 COVID+, 84 COVID-) 484	Race: Black n=202 non-Black n=61	Discordance= $\lvert SpO2-SaO2 \rvert > 4\%$ Saturation difference: $SaO_2 - SpO_2$	Discordance: COVID+ cohort: 27.9% COVID- cohort: 16.7% The saturation difference between COVID+ and – cohorts (average difference -1.22%, 95% CI: -2.08, -0.35) was no longer significant after adjusting for race (average difference -0.38%, 95% CI: -1.40, 0.64%).
Gadrey 2022[44]	US, 2020-2021, ≥18yo, Emergency department visit or hospitalized for COVID-19	5,319 1,909,867	Race/ethnicity: White, non-Hispanic: n=1,433 Black n=2,835 Hispanic n=612 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[103]	US, 2008-2019, no age limit reported, ICU	3,069 n/a	Race/ethnicity: Asian n=83 Black n=207 Hispanic n=112 White n=2,667	Time-weighted average supplemental oxygen rate	Model 2: regression coefficient Asian: **-0.291 (-0.546, -0.035)** Black: **-0.294 (-0.460, -0.128)** Hispanic: **-0.242 (-0.463, -0.020)** Model 3 (after controlling for discrepancy between average SpO_2 and average Hb oxygen saturation) Asian: -0.144 (-0.386, 0.098) Black: -0.081 (-0.239, 0.077) Hispanic: -0.092 (-0.301, 0.118)
*Kalra 2023a[104]	US, 2016-2021, ≥18yo, patients who received	196 (139 VA ECMO, 57 VV ECMO)	Race/ethnicity: VA ECMO cohort:	OH: $SaO_2 < 88\%$ despite $SpO_2 \geq 92\%$	VA ECMO cohort: Asian: OH= 1%, bias= 0.2%

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
	venoarterial (VA) or venovenous (VV) ECMO	16,252	Asian n=9 Black n= 27 Hispanic n=3 Other n=9 White n=91 VV ECMO cohort: Asian n=3 Black n= 19 Hispanic n=13 Other n=2 White n=20	Bias: mean SpO_2 – mean SaO_2	Black: OH= 70%, bias= 0.94% Hispanic: OH= 0%, bias= 0.03% Others: OH= 1%, bias= -0.80% White: OH= 27%, bias= -0.06% VV ECMO cohort: Asian: OH= 6%, bias= 1.0% Black: OH= 66%, bias= 2.9% Hispanic: OH= 11%, bias= 1.1% Others: OH= 1%, bias= -0.53% White: OH= 16%, bias= 0.50%
*Kalra 2023b[105]	US, 2018-2023, ≥18yo, patients who received venovenous (VV) ECMO	13,171	Race/ethnicity: Asian n=1,508 Black n= 1,777 Hispanic n=2,114 White n=7,772	OH: SaO_2 < 88% despite SpO_2 ≥ 92% Compare racial subgroup vs. White (ref) Pre-ECMO: ≤6 hrs before cannulation On-ECMO: 0-30 hrs after cannulation	Pre-ECMO OH: Asian aOR n.s. Black **aOR=1.55, 95% CI: 1.18, 2.02** Hispanic aOR n.s. White (ref) On-ECMO OH: Asian 1.6% (aOR n.s.) Black 3.1% (**aOR=1.79, 95% CI: 1.16, 2.75**) Hispanic 2.5% (**aOR=1.17, 95% CI: 1.15, 2.55**) White 1.7% (ref)
*Ruppel 2023[51]	US, 2016-2021, 1-17yo, cardiac catheterization	774 774	Race: Black n=201 White n=573	OH: SaO_2 < 88% despite SpO_2 ≥ 92% Bias: SpO_2 –SaO_2 Arms: Average root mean square	OH: Black 12% White 4% Bias (adjusted): Black 2.61 (**95% CI: 2.19, 3.04**) White 0.88 (0.63, 1.13) Arms: Black 4.36 White 3.01

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
*Savorgnan 2023[52]	US, 2020-2021, pediatric, admitted for COVID-19	2,713 2,713	Race/ethnicity: Non-Hispanic Black 61.3% Non-Hispanic White 38.7%	Bias: $SpO_2 - SaO_2$ OH: SaO_2 < 88% despite SpO_2 > 92%	Bias: Black patients were associated with more significant bias, which increased with decreasing oxygen saturation (p's<.0001). OH: Black 12% White 4% No association between bias or OH, and length of stay in hospital
Sjoding 2020[1]	US, 2014-2015 and 2020, no age limit reported, ICU	10,001 48,097	Race: Black n=1,326 White n=8,675	OH: SaO_2 < 88% despite 92% ≥ SpO_2 ≥ 96%	UM cohort: OH (adjusted) Black: 11.4% (7.6, 15.2%) White: 3.6% (2.5, 4.6%) Multicenter cohort: OH (unadjusted) Black: 17.0% (12.2, 23.3%) White: 6.2% (5.4, 7.1%)
Sudat 2022[106]	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 Cohort 2: 8,735 paired measures	Race/ethnicity: Cohort 1: Non-Hispanic White (NHW) n=35,127 Non-Hispanic Black (NHB) n=8,626 Cohort 2: Non-Hispanic White n=7,036 Non-Hispanic Black n=1,699	Cohort 1: SpO_2 SaO_2 pairs, OH Cohort 2: clinical and treatment characteristics	Cohort 1: Concurrent SaO_2<SpO_2 by 2%, and difference was 1% larger for NHB OH: NHW 3.01% NHB 5.50% (**p<0.001**) Cohort 2: lower admission probability, dexamethasone treatment, supplemental oxygen treatment, and increased time to treatment
Valbuena 2022a[8]	US, 2019-2020, ≥18yo, patients on ECMO due to ARDS or COVID-19	372 372	Race/ethnicity: White n=186 Hispanic n=70 Asian n=65	Pre-ECMO OH: SpO2 between 92 to 96% despite SaO_2 < 88%	OH Rate (95%CI): White: 10.2% (6.2, 15.3%) Hispanic: 8.6 % (3.2, 17.7%) Asian: 9.2% (3.5, 19.0%)

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
			Black n=51		Black: 21.5% (11.3, 35.3%) OR (95%CI): White: reference Black: 2.57 (**1.12, 5.92**)
Valbuena 2022b[107]	US, 2013-2019, US veterans (no age limit reported), inpatient excluding ICU	30,039 30,039	Race: Non-Hispanic White n=21,918 Non-Hispanic Black n=6,498 Hispanic/Latino n=1,623	OH: SaO_2 < 88% despite SpO_2 ≥ 92%	OH: Unadjusted probability (95% CI) Non-Hispanic White: 15.6% (15.0, 16.1%) Non-Hispanic Black: 19.6% (18.6, 20.6%) Hispanic: 16.2% (14.4, 18.1%)
Vesoulis 2022[9]	US, 2012-2019, infants <32 weeks gestation, NICU	294	Race: White n=170 Black n=124	OH: SaO_2 < 85% despite concurrent SpO_2 > 90%	OH: White: 7.7% (181 of 2342) Black: 9.2% (188 of 2044) p = 0.08
Wiles 2022[108]	UK, 2020-2020, ≥16yo, COVID pneumonitis	194 6,216	Race: White n=135 Asian n=34 Black n=19 Other n=6	Bias: SpO_2-SaO_2	Bias (limits of agreement) White: 0.28% (-1.79, 2.35) Asian: -0.33% (-2.47, 2.35) Black: -0.75% (-3.47, 1.97) Other: NR
Wong 2021[10]	US, 2014-2021, no age limit reported, ICU	79,044 87,971	Race: Asian n=1,919 (2.3%) Black n=26,032 (29.6%) Hispanic n=2,397 (2.7%) White n=57,632 (65.5%)	OH: SaO_2 < 88% despite SpO_2 ≥ 88% Organ dysfunction: SOFA and CVSOFA scores 24 hrs after ABG measurement Length of hospital stay In-hospital mortality	OH: Asian: 4.9% Black: 6.9% Hispanic: 6.0% White: 4.9% **p<.001** 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[109]	Desaturation	Korea, adult, healthy volunteers	28	Race: Asian Caucasian African American	Crosstalk sensor vs. crosstalk-free sensor Measurement error (bias): SpO_2-estimated SpO_2 ± precision (SD)	Crosstalk sensor: Asian 0.8258 ± 2.1603 Caucasian 0.8733 ± 1.9716 African American -3.0591 ± 3.9925 Crosstalk-free sensor: Asian -0.8824 ± 2.2859 Caucasian 0.6741 ± 3.2822 African American 0.9699 ±2.2268
*Barker 2023[110]	Desaturation study	US, adult, healthy volunteers	75 7,183	Race: Black n=39 White n=36	Masimo SET pulse ox vs. SaO_2 Bias: $SpO_2 - SaO_2$ OH: SaO_2 < 88% despite 92% ≥ SpO_2 ≥ 96%	Bias: Black: -0.2 ± 1.40% White: -0.05 ± 1.35% OH: Black: 0% White: 0.2%
*Giuliano 2023[111]	Desaturation study	US, adult, healthy volunteers	28 (14 in motion group, 14 in low perfusion group)	Motion group: Fitzpatrick: Type I n=1 Type III n=6 Type IV n=5 Type V n=1 Type VI n=1 Low perfusion group: Fitzpatrick: Type II n=4 Type III n=5	Perfusion: Normal ≥1.0 Low 0.3≤PI≤1.0 Very low 0.1≤PI≤0.3 Ultra low ≤0.1	Accuracy degraded during motion conditions. Accuracy degraded during low perfusion index ranges. Between group comparisons were not conducted by Fitzpatrick groups.

				Type IV n=3 Type V n=1 Type VI n=1		
*Gudelunas 2022[15] [preprint]	Desaturation study	US, adult, healthy volunteers	146 9,763	Fitzpatrick: Type I and II n=25 Type III and IV n=78 Type V and VI n=43	OH: SaO_2 < 88% despite 92% ≥ SpO_2 ≥ 96%	OH: Type I and II 1.1% Type III and IV 8.2% Type V and VI 21.1% Skin pigment, perfusion index, and degree of hypoxemia significantly contributed to bias
*Khanna 2023[112]	Pooled retrospective analysis of nine desaturation studies	US, adult, healthy volunteers	131 10,800	Skin pigmentation: Light (Fitzpatrick I-III) 75.9% Dark (Fitzpatrick IV-VI) 24.1%	Light vs. dark pigmentation groups	Bias (70%≤SpO2≤80%): Light +0.58% Dark +0.30% Difference = 0.28 (**p=0.0035**) Arms: Light 1.64% Dark 1.71%
*Leeb 2023[49] [preprint]	Desaturation study	US, adult, healthy volunteers	34 4,393	Fitzpatrick: Type I n=1 Type II n=2 Type III n=12 Type IV n=10 Type V n=5 Type VI n=4 ITA (measured at dorsal distal phalanx): grouped into thirds (lightest, medium, darkest)	SpO_2 vs. SaO_2 for 11 OTC pulse oximeters, stratified by skin pigmentation	6 of 11 met ARMS ≤ 3% 9 of 11 met ARMS ≤ 4% 9 of 11 demonstrated worse ARMS in lowest third of ITA values (dark) for SaO_2 between 70-80%
Mantri 2022[26]	Case series	US, adult, healthy volunteers	9	Fitzpatrick: Type 1 n=3 Type 4 n=3 Type 6 n=3	Photoacoustic imaging vs. pulse oximeter	SpO2: no significant differences by skin type Photoacoustics: higher PA signal, reduced

						penetration depth for Type 6
Okunlola 2022[2]	Desaturation	U.S., adult, healthy volunteers	491 3,778	Skin pigmentation: dark, medium, light	Dark vs. light to medium	Small positive bias in dark pigmentation group (data presented in Bland Altman plots)