U.S. FOOD & DRUG ADMINISTRATION

Center for Devices and Radiological Health

Approach for Improving the Performance Evaluation of Pulse Oximeter Devices Taking Into Consideration Skin Pigmentation, Race and Ethnicity

Discussion Paper and Request for Feedback

Approach for Improving the Performance Evaluation of Pulse Oximeter Devices Taking Into Consideration Skin Pigmentation, Race and Ethnicity

Discussion Paper and Request for Feedback

This discussion paper is intended for discussion purposes only and does not represent draft or final guidance. It is not intended to propose or implement policy changes regarding evaluation of devices intended for non-invasive measurement of arterial blood oxygen saturation. This document is not intended to communicate the FDA's proposed (or final) regulatory expectations but is instead meant to seek early input from groups and individuals outside the Agency and advance a broader discussion among stakeholders on this topic.

The objective of this discussion paper is to obtain public comment and stakeholder feedback on a series of questions related to an approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters taking into consideration a patient's skin pigmentation, and patient-reported race and ethnicity. For purposes of this discussion paper, improvement is defined as reducing the disparate performance of pulse oximeters for patients of different skin pigmentation, races and ethnicities. Please submit your comments regarding this discussion paper to https://www.regulations.gov, Docket No. FDA-2023-N-4976. FDA intends to consider all comments timely submitted to this docket (FDA-2023-N-4976) before issuing draft guidance related to this topic.

CDRH will hold a virtual public meeting of the Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee on February 2, 2024. Among other topics, FDA intends to ask the panel to consider the approach offered in this discussion paper. FDA intends to provide to the panel a summary of comments timely submitted in response to the docket referenced above.

I. Introduction

The Center for Devices and Radiological Health (CDRH) is committed to ensuring that patients and healthcare providers have timely and continued access to safe, effective, and high-quality medical devices and safe radiation-emitting products. The Coronavirus Disease 2019 (COVID-19) pandemic resulted in increased use of pulse oximeters not only in the hospital setting, but at home as well. The performance of pulse oximeters has generated substantial interest from stakeholders. The objective of this discussion paper is to provide an initial perspective to stimulate discussion and to solicit feedback from a variety of stakeholders on how to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters taking into consideration a patient's skin pigmentation, race and ethnicity.

II. Background

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 is measured using a blood sample obtained from arterial puncture. SpO_2 is an estimate of how much oxygen the hemoglobin contains compared with how much it could contain, expressed as a percentage. Pulse oximetry is a quick and non-invasive alternative to arterial blood sampling for estimating oxygen saturation. This discussion paper is intended to cover pulse oximeters intended for medical purposes that require premarket

authorization, including standalone pulse oximeters and pulse oximeter functions as part of multi-parameter devices.[1]

The COVID-19 pandemic resulted in an increase in the use of pulse oximeters. In a safety communication issued on February 19, 2021, the FDA informed patients and health care providers that although pulse oximetry is useful for estimating blood oxygen levels, pulse oximeters have limitations and a risk of inaccuracy which, under certain circumstances, should be considered.[2] FDA's safety communication stated that multiple factors may affect the performance of a pulse oximeter's readings, such as poor circulation, skin pigmentation, skin thickness, skin temperature, current tobacco use, and use of fingernail polish.

The February 2021 Safety Communication stated that:

> "In many cases, the level of inaccuracy may be small and not clinically meaningful; however, there is a risk that an inaccurate measurement may result in unrecognized low oxygen saturation levels. Therefore, it is important to understand the limitations of pulse oximetry and how accuracy is calculated and interpreted.
>
> FDA-cleared prescription pulse oximeters are required to have a minimum average (mean) accuracy that is demonstrated by desaturation studies done on healthy patients. This testing compares the pulse oximeter saturation readings to arterial blood gas saturation readings for values between 70-100%. The typical accuracy (reported as Accuracy Root Mean Square or A_{rms}) of recently FDA-cleared pulse oximeters is within 2 to 3% of arterial blood gas values. This generally means that during testing, about 66% of SpO_2 values were within 2 or 3% of blood gas values and about 95% of SpO_2 values were within 4 to 6% of blood gas values, respectively."[3]

Current scientific evidence from lab desaturation studies[4,5] suggests that there are some accuracy differences in pulse oximeter performance, especially in lower arterial saturations (SaO_2) between lightly and darkly pigmented participants. In correspondence published in the *New England Journal of Medicine* by Sjoding, et al. (2020)[6] the authors reported that Black patients had nearly three times the frequency of occult hypoxemia (low oxygen level in the blood) as detected by blood gas measurements not

---

[1] Pulse oximeters for medical purposes may include both prescription and over-the-counter (OTC) indications.
[2] See Pulse Oximeter Accuracy and Limitations: FDA Safety Communication (February 2021), www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication.
[3] See Pulse Oximeter Accuracy and Limitations: FDA Safety Communication, www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication. The February 2021 safety communication also stated that "Over-the-counter (OTC) oximeters are sold directly to consumers in stores or online and include smart phone apps developed for the purpose of estimating oxygen saturation. Use of OTC oximeters has increased as a result of the COVID-19 pandemic. These products are sold as either general wellness or sporting/aviation products that are not intended for medical purposes, so they do not undergo FDA review. OTC oximeters are not cleared by the FDA and should not be used for medical purposes." This statement is in reference to standalone OTC pulse oximeters not intended for medical purposes.
[4] Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005;102.4:715-719.
[5] Okunlola OE, Lipnick MS, Batchelder PB, Bernstein M, Feiner JR, Bickler PE. Pulse Oximeter Performance, Racial Inequity, and the Work Ahead. *Respir Care*. 2022;67(2):252-257.
[6] Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. *N Engl J Med.* 2020;383(25):2477-2478.

detected by pulse oximetry, when compared with White patients. While this retrospective study has limitations, these findings highlighted a need to further evaluate and understand the association among skin pigmentation, race and ethnicity, and pulse oximeter performance. An association of a variable with pulse oximeter accuracy does not always imply causation and may be observed for many reasons. For example, a real-world association of race and ethnicity with a summary metric of pulse oximeter accuracy may mean that race and ethnicity is associated with another variable that is the actual cause of the variation in pulse oximeter accuracy.

FDA has issued guidance titled "Pulse Oximeters – Premarket Notification Submissions [510(k)s]"[7] (hereinafter, the guidance), with recommendations for the content of information to be submitted to FDA in a premarket submission. In particular, the guidance recommends that the devices undergo performance testing in a controlled desaturation laboratory study as described in *ISO 80601-2-61:2011 Medical Electrical Equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment* or equivalent method. The testing is conducted on healthy adult participants, whose fraction of inspired oxygen concentration is decreased in a series of targeted steady-state desaturation periods. During these periods, repeated SaO_2 measurements are obtained from blood samples taken from an indwelling arterial catheter and paired with SpO_2 readings from the investigational pulse oximeter. The targeted range for the SaO_2 measurements is 70-100%.

In the guidance, FDA recommends that participants in the laboratory study have a range of skin pigmentations, including at least 2 darkly pigmented participants or 15% of the participant pool, whichever is larger. The guidance does not recommend methods to measure pigmentation. Apart from age and gender recommendations,[8] these clinical studies are neither recommended to enroll demographic groups nor be statistically powered to detect differences in performance between such cohorts. This discussion paper is among FDA's efforts to review the effects of skin pigmentation on the performance of these devices, examining data from controlled desaturation laboratory studies and data from real world settings.

On November 1, 2022, FDA convened the Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee[9] (hereinafter referred to as "Anesthesiology Devices Panel") to discuss ongoing concerns that pulse oximeters may perform disparately in individuals with darker skin pigmentations. The panel was convened to discuss factors that may affect pulse oximeter accuracy and performance, the available evidence about the accuracy of pulse oximeters, recommendations for patients and healthcare providers, and the amount and type of data that should be provided by manufacturers to assess pulse oximeter accuracy and performance and to guide other regulatory actions as needed.

[7] See Pulse Oximeters – Premarket Notification Submissions [510(k)s], Guidance for Industry and Food and Drug Administration Staff (March 2013), www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug.

[8] See Pulse Oximeters – Premarket Notification Submissions [510(k)s], Guidance for Industry and Food and Drug Administration Staff, www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug.

[9] See November 1, 2022: Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee Meeting Announcement (November 2022), www.fda.gov/advisory-committees/advisory-committee-calendar/november-1-2022-anesthesiology-and-respiratory-therapy-devices-panel-medical-devices-advisory#event-materials.

In preparation for the Anesthesiology Devices Panel meeting, FDA issued an Executive Summary,[10] which summarized three systematic reviews that addressed the topic of potential bias of pulse oximetry in people with darker skin pigmentation, all published in the year 2022. The authors' conclusions are described below:

- Cabanas, et al. (2022)[11] 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 performance 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" and also that "a more accurate method for classifying the research participants into categories by degree of skin pigmentation should be employed in these studies."

- Shi, et al. (2022)[12] 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 SaO_2 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."

- Poorzargar, et al. (2022)[13] 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.

---

[10] See FDA Executive Summary: Review of Pulse Oximeters and Factors that can Impact their Accuracy, Prepared for the November 1, 2022, meeting of the Anesthesiology Devices Advisory Committee (November 2022), www.fda.gov/media/162709/download.

[11] Cabanas AM, Fuentes-Guajardo M, Latorre K, León D, Martin-Escudero P. Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. *Sensors.* 2022;22(9):3402.

[12] Shi C, Goodall M, Dumville J, Hill J, Norman G, Hamer O, Clegg A, Watkins CL, Georgiou G, Hodkinson A, Lightbody CE, Dark P, Cullum N. The accuracy of pulse oximetry in measuring oxygen saturation by levels of skin pigmentation: a systematic review and meta-analysis. *BMC Med*. 2022;20(1):267.

[13] Poorzargar K, Pham C, Ariaratnam J, Lee K, Parotto M, Englesakis M, Chung F, Nagappa M. Accuracy of pulse oximeters in measuring oxygen saturation in patients with poor peripheral perfusion: a systematic review. *J Clin Monit Comput*. 2022;36(4):961-973.

The Anesthesiology Devices Panel concluded that the currently available clinical evidence for prescription pulse oximeters showed disparate performance in patients with dark skin pigmentation, which leads to increased risk and worse outcomes for patients. The panel also indicated that factors other than pigmentation, such as perfusion and perhaps demographic factors in terms of the width of the finger, the breadth of the finger, and obesity, explain some of the disparate performance and should be examined by FDA. To address these concerns, the panel recommended standardization of skin pigmentation assessment. The panel recommended that overall, pulse oximeters for clinical use should be more accurate and proposed reducing the Accuracy Root Mean Square (A_{RMS}) threshold.

Following the Anesthesiology Devices Panel meeting, CDRH has been considering how to improve the evaluation of pulse oximeter accuracy and performance by taking into greater consideration differences in skin pigmentation and considering how race and ethnicity may impact the performance of a pulse oximeter's measurement of oxygen saturation.

CDRH will hold a virtual public meeting of the Anesthesiology Devices Panel on February 2, 2024 to consider, among other potential topics, a new approach to improve the assessment of pulse oximeter performance. While a variety of factors may impact the accuracy of pulse oximeters, the objective of this discussion paper is to provide an initial perspective to stimulate discussion and solicit feedback from stakeholders on how to improve performance evaluation of pulse oximeters taking into account skin pigmentation, race and ethnicity.

III. Approach for Discussion

In an effort to reduce the disparate performance of pulse oximeters and thereby increase the overall performance estimates of such devices, the Agency is considering the utility of a clinical trial design that includes the entire range of skin pigmentation and also accounts for race and ethnicity. For purposes of this discussion paper, improvement is defined as reducing the disparate performance of pulse oximeters for patients of different skin pigmentation, races and ethnicities. CDRH is seeking stakeholder feedback on the following clinical study design and considerations:

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)[14] using an assessment that has been validated to capture race and ethnicity diversity in pigmentations within the United States (US), and is intended to result in a more inclusive and representative study of the intended patient population. This is designed to increase diversity in skin pigmentation, race and ethnicity of participants in a clinical study used to evaluate the performance of pulse oximeters reviewed by the Agency.

For example, one approach is to have a clinical study that 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

[14] 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.

capture race and ethnicity diversity in pigmentations within the US.[15] The forehead, a common location for perception and assessment of pigmentation,[16,17,18] has a wide range of pigmentation levels that allow for MST assessment and stratification of participants into 1 of 10 MST values.[19] 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
- 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 sensor placement – where the pulse oximeter is placed (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, the Agency is considering the utility of the following to analyze the performance of the pulse oximeter being evaluated. Non-disparate performance is defined by lack of variation in SpO_2 bias across ITA and MST levels, where SpO_2 bias is the mean of the difference $D=SpO_2 - SaO_2$. Given current limits of the pulse oximeter technology, FDA believes it would be appropriate to consider a pulse oximeter to have non-disparate performance when the estimate of the absolute difference in SpO_2 bias across ITA and MST levels is < 1.5% when $SaO_2 > 85\%$, and < 3.5% when $70\% < SaO_2 \leq 85\%$.

The first step in assessing non-disparate performance would be to utilize a mixed effects model on data using ITA and SaO_2 as predictors (hereinafter, called the ITA-derived performance analysis).

- The mixed effects model would include as predictors ITA, SaO_2, and ITA by SaO_2 interaction, and random effects for participant, participant by ITA interaction, and participant by SaO_2 interaction (see Section B.iv, below).

[15] 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; Accepted for publication by ACM Journal on Responsible Computing]. 2022. Available at https://scholar.harvard.edu/sites/scholar.harvard.edu/files/monk/files/heldreth_et_al._which_skin_tone_measures_are_the_most_inclusive_vpreprint.pdf.

[16] Hsiao JH, Cottrell G. Two fixations suffice in face recognition. *Psychol Sci*. 2008;19(10):998-1006.

[17] Hannon L, DeFina R. Reliability Concerns in Measuring Respondent Skin Tone by Interviewer Observation. *Public Opin Q*. 2016;80(2):534-541.

[18] Hugenberg, Kurt & Wilson, John. (2013). Faces are central to social cognition. *Handbook of Social Cognition*. 167–193.

[19] Hugenberg, K. Faces are central to social cognition. 167–193.

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

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 ≤ 85%.

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

i) MST Scale

By considering 25% or more participants per each MST cohort (i.e., 1-4, 5-7, 8-10), the methodology being considered 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.

ii) ITA Scale

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.

iii) Pigmentation Assessment

Though the guidance recommends that participants in the controlled desaturation laboratory studies have a range of skin pigmentation, it does not specify how or where to assess skin pigmentation. As mentioned, the forehead is a common location for perception and assessment of pigmentation,[20,21,22] and while such an anatomical site may capture the greatest range of pigmentation, a quantitative assessment of skin pigmentation at the sensor site would be most appropriate to optimize assessment of pulse oximeter performance accuracy.

Human epidermal melanin exhibits substantial absorption through ultraviolet, visible, and NIR (near-infrared) wavelengths, and that light absorption decreases exponentially with wavelength. Of importance, melanin concentration varies widely in humans due to both genetics and sun exposure and can strongly impact light attenuation and visualized pigmentation. Therefore, sensor site pigmentation – specifically, the level of melanin in the epidermis at the sensor site – may affect pulse oximeter accuracy.

Skin pigmentation can be assessed using subjective and objective methods. Colorimetry is the most common and well-standardized approach for objective evaluation of pigmentation. Standard colorimetry methods 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

[20] Hsiao JH. Two fixations suffice in face recognition. 998-1006.
[21] Hannon L. Reliability Concerns in Measuring Respondent Skin Tone by Interviewer Observation. 534-541.
[22] Hugenberg, K. Faces are central to social cognition. 167–193.

yellow/blue component. These objectively measured variables can then be used to calculate ITA[23] - defined as ITA=180/ π × arctan((L* - 50)/b*) – which 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, [24] high performance liquid chromatography,[25] and spectrophotometry[26] 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).[27]

MST is standardized to color scales such as CIELAB. Additionally, MST has been shown to have a high intraclass correlation coefficient of 0.90-0.96, even among a global pool of raters.[28] It should be noted, however, that best practices for performing evaluations (e.g., lighting conditions) and high-quality printing source for color charts have not yet been established for MST or other subjective methods.

iv) ITA-derived performance analysis

The mixed effects model in the ITA-derived performance analysis may be used to estimate the expected value of D 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%. The model should be checked for goodness-of-fit to the data.

v) 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,[29] is not a proxy for race and ethnicity diversity.

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, above). We note that the power to detect MST effects on bias could

---

[23] Del Bino S, and Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br J Dermatol*. 2013;169 Suppl 3:33-40.

[24] Del Bino S, Ito S, Sok J, Nakanishi Y, Bastien P, Wakamatsu K, Bernerd F. Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. *Pigment Cell Melanoma Res*. 2015;28(6):707-717.

[25] 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.

[26] Del Bino S. Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. 707-717.

[27] Del Bino S. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. 33-40.

[28] Schumann C, Olanubi GO, Wright A, Monk Jr E, Heldreth C, Ricco S. Consensus and Subjectivity of Skin Tone Annotation for ML Fairness. [Preprint; Accepted for publication by Conference on Neural Information Processing Systems (NeurIPS) 2023]. Available at https://www.researchgate.net/publication/370814466_Consensus_and_Subjectivity_of_Skin_Tone_Annotation_for_ML_Fairness.

[29] Heldreth CM. Which Skin Tone Measures are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence. [Preprint; Accepted for publication by ACM Journal on Responsible Computing]. Available at https://scholar.harvard.edu/sites/scholar.harvard.edu/files/monk/files/heldreth_et_al._which_skin_tone_measures_are_the_most_inclusive_vpreprint.pdf.

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. 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, which could lead to non-existence of maximum likelihood estimates under the model. 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.[30,31,32] Variations in the number, size, and aggregation of melanosomes within the melanocyte and keratinocyte lend to racial and ethnic differences in pigmentation.[33] 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.[34] 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.

vi) Summary

FDA is considering the utility of a potential clinical study design that incorporates the entire range of skin pigmentation found in US racial and ethnic groups (i.e., the entire MST scale) to evaluate pulse oximeter performance to better determine whether there is any evidence of non-disparate performance based on an ITA-derived performance analysis as well as an MST-derived performance analysis. While the definition for non-disparate performance remains the same for both methods (i.e., estimated maximum absolute difference in SpO_2 bias between low and high ITA or MST values is < 1.5% when SaO_2 > 85%, and < 3.5% when 70% < SaO_2 ≤ 85%, where SpO_2 bias is defined as the expected value of D=SpO_2 – SaO_2), the analysis remains distinct for each. Since an MST value may not be unique to a certain racial and ethnic group due to the variation of range and levels of pigmentation within each group, a diverse representation of racial and ethnic groups in a clinical study may not be assured for the ITA-derived or MST-derived performance analysis. Until there is certainty on whether racial and ethnicity differences within the same level of pigmentation give rise to non-disparate pulse oximeter performance, the Agency posits that it remains important for premarket clinical studies to include participants from diverse racial and ethnicity groups, as well as diverse pigmentation levels, to improve generalizability of

[30] Gudelunas MK, Lipnick M, Hendrickson C, Vanderburg S, Okunlola B, Auchus I, Feiner JR, Bickler PE. Low perfusion and missed diagnosis of hypoxemia by pulse oximetry in darkly pigmented skin: A prospective study. [Preprint]. 2022. Available at www.medrxiv.org/content/10.1101/2022.10.19.22281282v1.

[31] Okunlola OE. Pulse Oximeter Performance, Racial Inequity, and the Work Ahead. 252-257.

[32] Bickler PE. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. 715-719.

[33] Taylor SC. Skin of color: biology, structure, function, and implications for dermatologic disease. *J Am Acad Dermatol*. 2002;46(2 Suppl Understanding):S41-62.

[34] Heldreth CM. Which Skin Tone Measures are the Most Inclusive? An Investigation of Skin Tone Measures for Artificial Intelligence. [Preprint; Accepted for publication by ACM Journal on Responsible Computing]. Available at https://scholar.harvard.edu/sites/scholar.harvard.edu/files/monk/files/heldreth_et_al._which_skin_tone_measures_are_the_most_inclusive_vpreprint.pdf.

study results. One possibility is for non-disparate performance to be shown with ITA-derived as well as MST-derived performance analyses as described in this discussion paper.

IV. Questions for Stakeholders

CDRH has provided an initial perspective to stimulate discussion and to solicit feedback from stakeholders related to a possible approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters taking into consideration a patient's skin pigmentation, race and ethnicity. CDRH is interested in obtaining public comment and feedback on the following series of questions. Please submit your comments regarding this discussion paper to https://www.regulations.gov, Docket No. FDA-2023-N-4976. FDA intends to consider all comments timely submitted to this docket (FDA-2023-N-4976) before issuing draft guidance related to this topic.

1. Do you see benefits to conducting an initial assessment with the Monk Skin Tone (MST) scale to capture race and ethnicity diversity in pigmentation relevant to the US population followed with an objective Individual Typology Angle (ITA) assessment at the sensor site to assess non-disparate performance for enrolled participants?

2. What are your thoughts on stratifying clinical study enrollment such that the entire range of skin pigmentation found in US race and ethnicity groups (i.e., the entire MST scale) is represented?

3. Do you agree with the proposed definition of non-disparate performance assurance that the estimated maximum absolute difference in SpO_2 bias across both ITA and MST levels is < 1.5% for SaO_2 > 85%, and < 3.5% when 70% < SaO_2 ≤ 85%?

4. What are your thoughts on using the proposed ITA-derived performance analysis and MST-derived performance analysis to assess whether the criteria for non-disparate performance assurance are met across skin pigmentation, race and ethnicity?

5. Do you agree that the proposed approach for the clinical trial design achieves adequate race and ethnic diversity in enrolled participants to demonstrate reliable and accurate MST-derived performance analysis?

6. In addition to what has been discussed, what are other ways for the Agency to improve evaluation of the performance of pulse oximeters, while taking into consideration differences in skin pigmentation?

7. In addition to what has been discussed, what are other ways for the Agency to consider how race and ethnicity may impact the performance of a pulse oximeter's measurement of oxygen saturation?

V. Conclusion

The COVID-19 pandemic resulted in an increase in the use of pulse oximeters. Although pulse oximetry is useful for estimating blood oxygen levels, pulse oximeters have limitations and a risk of inaccuracy under certain circumstances that should be considered. While there are multiple factors that may affect the accuracy of pulse oximeters, this discussion paper is focused on skin pigmentation, race and ethnicity as well as an approach for determination of non-disparate performance in premarket studies. CDRH currently is considering how modifications to the design of premarket studies could address these key factors that may affect the assessment of accuracy for pulse oximeters, thereby improving

confidence in the performance outcomes, helping to ensure availability of safe and effective, high-quality pulse oximeter devices. CDRH is interested in obtaining public comment and feedback to the questions posed in this discussion paper. This feedback will be useful when CDRH convenes a virtual public meeting of the Anesthesiology Devices Panel on February 2, 2024 to advise CDRH on ways to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters taking into consideration a patient's skin pigmentation, race and ethnicity.