UNITED STATES OF AMERICA

FOOD AND DRUG ADMINISTRATION

+++

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

+++

ANESTHESIOLOGY AND RESPIRATORY THERAPY DEVICES PANEL OF THE

MEDICAL DEVICES ADVISORY COMMITTEE

+++

FEBRUARY 2, 2024

9:00 a.m. EST

Via Web Conference

Transcript Produced By:

Translation Excellence

3300 South Parker Road, Aurora, CO 80014

https://translationexcellence.com/

Participants

Voting Chair	Hugh A. Cassiere, M.D., F.C.C.P., F.A.C.P.	Director, Critical Care South Shore University Hospital
Voting Members	Lonny B. Yarmus, D.O., M.B.A.	Professor of Medicine, Division of Pulmonary and Critical Care Medicine Johns Hopkins University School of Medicine
	Naresh M. Punjabi, M.D., Ph.D.	Mary Jane and Lino Sertel Professor of Pulmonary Medicine, Chief Division of Pulmonary, Critical Care, and Sleep Medicine University of Miami Miller School of Medicine
	Anne Whitney Brown, M.D..	Transplant Pulmonologist, Inova Fairfax Advanced Lung Disease and Transplant Program, Senior Director of Clinical Affairs, Cystic Fibrosis Foundation
	Cheryl K. Gooden, M.D.	Associate Professor of Anesthesiology and Pediatrics, Yale School of Medicine
Temporary Non-Voting Members	Ben Saville, Ph.D.	Associate Professor of Anesthesiology and Pediatrics, Yale School of Medicine
	Karla V. Ballman, Ph.D..	Consultant, Division of Clinical Trials & Biostatistics Professor of Biostatistics, Mayo Clinic College of Medicine and Science, Associate Director of Quantitative Health Sciences, Mayo Clinic Comprehensive Cancer Center Director, MCCCC Biostatistics Shared Resource
	Raymond J. Lanzafame, M.D., M.B.A.	General Surgeon
	James S. Taylor, M.D.	Clinical Professor of Dermatology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University
	Tamorah R. Lewis, M.D., Ph.D.	Associate Professor, Department of Pediatrics Divisions of Neonatology and Pediatric Clinical Pharmacology, University of Toronto
	Thomas E. Wiswell, M.D.	Neonatologist, Kaiser Permanente Moanalua Medical Center
	Julian M. Goldman, M.D., FASA	Anesthesiologist, Department of Anesthesia, Critical Care and Pain Medicine Massachusetts General Hospital
Industry Representative	Wilson C. Wilson, M.D., M.A.	Executive Vice President, Clinical Operations Masimo Corporation
Consumer Representative	Rachel Brummert, M.S.	Communications Lead, American Society of Pharmacovigilance
Patient Representative	Joseph P. O'Brien, M.B.A.	President & CEO, National Scoliosis Foundation Patient Representative

Designated Federal Officer	Candace Nalls	Food and Drug Administration
Office Director	Malvina B. Eydelman, M.D., Ph.D..	Food and Drug Administration, CDRH/OPEQ/OHTI
Division Director	James J. Lee, Ph.D.	Food and Drug Administration, CDRH/OPEQ/OHTI/DHTIC

Contents

Call to Order

Dr. Cassiere: Good morning, everyone. I would like to call this meeting of the Anesthesiology and Respiratory Devices Panel to order. I am Dr. Cassiere, the chairperson of this panel. I am an Associate Professor of Medicine and Cardiovascular thoracic surgery at the Zucker School of Medicine. My area of expertise relevant to this meeting is in respiratory therapy devices, patient monitoring, and critical care. I am the director of Critical Care Services at South Shore University Hospital in Bayshore, New York, a part of Northwell Health.

I note for the record that the members present constitute a quorum as required by [indiscernible – no audio] 2118. I would like to also add that the panel members participating in today's meeting have received training in device law and regulations.

For today's agenda, the panel will discuss an approach to improve quality premarket studies and associated methods used to evaluate the performance of pulse oximeters submitted for premarket review, taking into consideration the patient's skin pigmentation and patient reported race and ethnicity. Before we begin, I would like to remind the public and panelists that this is a non-voting meeting, and I ask our distinguished committee members and FDA attending virtually to introduce themselves. Can members please turn on your video monitors if you have not already done so and unmute your microphone before you speak.

Panel Introductions

I will call your name. Please state your area of expertise, your position, and your affiliation. First up is Dr. Yarmus.

Dr. Yarmus: Good morning. Lonnie Yarmus. I am the director of interventional pulmonology and have expertise in complex airway management, critical care, and central airway obstruction.

I'm a professor of medicine and oncology at Johns Hopkins University School of Medicine and vice chairman of the Department of Medicine.

Dr. Cassiere: Thank you. Dr. Punjabi.

Dr. Naresh M. Punjabi: Good morning, everybody. Naresh Punjabi. I'm a professor and chief of pulmonary critical care and sleep medicine at the University of Miami Miller School of Medicine, my area of expertise is in the area of overactive pulmonary disorders, critical care medicine, and sleep medicine.

Dr. Cassiere: Dr. Brown.

Dr. Anne Whitney Brown: Hi, good morning. Whitney Brown. I'm a pulmonary critical care physician. I have expertise in CF and lung transplantation. I work in a shared capacity at Inova Fairfax Hospital, and I'm senior director of clinical affairs at the Cystic Fibrosis Foundation.

Dr. Cassiere: Thank you. Dr. Saville.

Dr. Saville: Yeah, hi. Ben Saville. Good morning. I'm the president and lead statistical scientist at Adaptix Trials. I'm a biostatistician by trade, and my specific expertise is in Bayesian and adaptive clinical trial design.

Dr. Cassiere: Dr. Ballman.

Dr. Ballman: Hi, Karla Ballman. I am a professor at Mayo Clinic, a professor of biostatistics. My expertise is in clinical trial design as well as study design and analyses.

Dr. Cassiere: Dr. Lanzafame.

Dr. Lanzafame: Hi, I'm Dr. Raymond Lanzafame. I am a general surgeon in Rochester, New York. I have more than 42 years of experience with lasers, energy sources with expertise in light tissue interactions and tissue effects of these devices.

Dr. Cassiere: Dr. Taylor.

Dr. James S. Taylor: Good morning, Jim Taylor. My clinical research interests are occupational and environmental dermatology and medical dermatology, including putative reactions to medical devices, as well as quality and patient safety. I'm a consultant dermatologist at the Cleveland Clinic and clinical professor of Dermatology at the Cleveland Clinic Lerner College of Medicine.

Dr. Cassiere: Dr. Lewis.

Dr. Lewis: Hi, I'm Tamorah Lewis. I'm an associate professor in the departments of Pediatrics and Pharmacology at the University of Toronto Temerty School of Medicine. I am a clinical neonatologist who works in level four neonatal ICUs at the Children's Hospital, Sick Kids. And I also perform research in neonatal clinical pharmacology.

Dr. Cassiere: Dr. Wiswell.

Dr. Wiswell: Hi, I'm Tom Wiswell. I'm a neonatologist. My most recent academic position was as a professor of pediatrics at the University of Central Florida. I have expertise in pulmonary disorders of the newborn infant, and I've actually been studying pulse oximetry since the late 1980s. And my current practice actually is at Kaiser Permanente Moanalua in Honolulu.

Dr. Cassiere: Dr. Goldman.

Dr. Goldman: Hi. Good morning. Julian Goldman. I'm an anesthesiologist at the Massachusetts General Hospital, Department of Anesthesia, Critical Care and Pain Medicine, and the Medical Director of Biomedical Engineering for the Mass General Brigham Health System.

As an anesthesiologist, I've been involved with the clinical use of pulse oximetry since its time of introduction to clinical practice. In my hospital biomedical engineering role, I've helped evaluate and source low-cost pulse oximeters to support COVID-19 care. As a researcher, I've been involved with testing and evaluating pulse oximeters for quite a while in different venues, and I've worked on developing pulse oximetry and related standards. In particular, I introduced

the section on fidelity and delays in the pulse oximeter standard to support clinical usability and interpretation of the otherwise quite complex standards and specifications.

Dr. Cassiere: Dr. Feldman.

Dr. Feldman: Good morning, Jeffrey Feldman. I'm a practicing anesthesiologist at the Children's Hospital of Philadelphia. Also a professor of anesthesiology at the Perlman School of Medicine University of Pennsylvania. I also chair the Committee on Technology for the Anesthesia Patient Safety Foundation. My areas of expertise include anesthesia delivery systems, both mechanical ventilation in the operating room as well as anesthetic delivery and patient monitoring. And I'm pleased to join the panel this morning.

Dr. Cassiere: Dr. Wilson.

Dr. Wilson: Good morning. My name is William Wilson. My area of expertise is cardiovascular anesthesiology and critical care medicine. I am a past professor of anesthesiology, medicine, and surgery at the University of California, Irvine Medical Center. Where I was the chief medical officer and chief of critical care medicine. I'm currently the executive VP for clinical research and operations at Masimo, and I'm serving today as the industry representative on this committee.

Dr. Cassiere: Rachel Brummert.

Ms. Brummert: Good morning, my name is Rachel Brummert. My area of expertise is in medical device safety and medication safety. I'm the communication lead for the American Society of Pharmacovigilance, and today, I'm serving as the consumer representative.

Dr. Cassiere: Mr. O'Brien.

Mr. O'Brien: Morning, Joe O'Brien, president, and CEO of the National Scoliosis Foundation. My area of expertise is spinal deformity and patient education and support, and I am the patient representative.

Dr. Cassiere: Dr. Eydelman.

Dr. Eydelman: Hi, everybody. My name is Malvina Eydelman, and I'm the director of the Office of Health Technology 1, which is ophthalmic, anesthesia, respiratory, ENT, and ventral devices here at the FDA. Welcome, everyone, and thank you for joining us today.

Dr. Cassiere: And Dr. Lee.

Dr Lee: Good morning, I'm James Lee. I'm the Division Director for Sleep Disorder Breathing, Respiratory, and Anesthesia Devices here at OPEQ OHT1 FDA. Thank you.

Dr. Cassiere: Great, thank you, everyone, for those introductions. Candice Nalls, the designated federal officer for today's Anesthesiology and Respiratory Therapy Devices panel, will make some introductory remarks.

Conflict of Interest Statement

Ms. Nalls: The Food and Drug Administration, FDA, is convening today's meeting of the Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee under the authority of the Federal Advisory Committee Act, FACA, of 1972. With the exception of the industry representative, all members, and consultants of the panel are special government employees or regular federal employees from other agencies and are subject to federal conflict of interest laws and regulations.

The following information on the status of this panel's compliance with federal ethics and conflict of interest laws covered by, but not limited to, those found at 18 USC Section 208 are being provided to participants in today's meeting and to the public. FDA has determined that members and consultants of this panel are in compliance with federal ethics and conflict of interest laws.

Under 18 USC Section 208, Congress has authorized the FDA to grant waivers to special government employees and regular federal employees who have financial conflicts of interest

when it is determined that the agency's need for a particular individual's services outweighs his or her potential financial conflict of interest.

Related to the discussion of today's meeting, members and consultants of this panel who are special government employees or regular federal employees have been screened for potential financial conflicts of interest of their own, as well as those imputed to them, including those of their spouses or minor children and for purposes of 18 USC Section 208, their employers. These interests may include investments, consulting, expert witness testimony, contracts, grants, credits, teaching, speaking, writing, patents and royalties, and primary employment. For today's agenda, the panel will discuss ongoing concerns that pulse oximeters may be less accurate in individuals with darker skin pigmentations.

The advisory panel 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 the patient's skin pigmentation and the patient reported race and ethnicity. The committee will discuss the type and amount of data that should be provided by manufacturers to the FDA to evaluate the performance of pulse oximeters submitted for premarket review, including prescription and over-the-counter indications and labeling considerations.

Based on the agenda for today's meeting and all financial interests reported by the panel members and consultants, a conflict-of-interest waiver has been issued in accordance with 18 USC Section 208B.3 to Dr. Jeffrey Feldman. Dr. Feldman's waiver addresses his personal financial interests in a healthcare sector mutual fund that contain underlying asset shares in potentially affected or competing firms. The aggregate value of his holdings in the fund is between $75,000 and $125,000. The waiver also addresses Dr. Feldman's employer pending federally funded related grant, with the amount to be awarded is between $250,000 and $600,000

annually. Dr. Feldman reported that he will be involved with the study, but he will receive no personal compensation from the study's funds. The waiver allows this individual to participate fully in the panel deliberations.

FDA's reasons for issuing the waiver are described in the waiver documents, which are posted on the FDA's website http//www.fda.govs/advisorycommittees/default.htm. Copies of the waiver may also be obtained by submitting a written request to the agency's Division of Freedom of Information at 5630 Fishers Lane, Room 1035, Rockville, Maryland, 20857.

Dr. William Wilson is serving as the industry representative, acting on behalf of all related industry. He is employed by Masimo Corporation. For the record, the agency notes that the following invited guest speakers with us today have acknowledged interest in affected firms at issue that are related to the matter before the panel. Jeffrey Feldman has acknowledged interest in the form of consulting in a past speaking event with an effective or competing firm at issue. Philip Bickler has acknowledged his employer's interest with multiple affected or competing firms at issue in the form of research studies. Christopher Almond has acknowledged his employer's related research study for which he serves as an investigator that is federally funded. Steven Barker has acknowledged that he owns stock and is employed by an affected or competing firm at issue. Linus Park has acknowledged that he owns stock and is employed by an affected or competing pulse oximeter manufacturer. Meghan Lane-Fall has acknowledged interest with an affected or competing firm at issue in the form of a one-time speaking event. Garrett Burnett has acknowledged his employer's interest in the form of a research grant for which he serves as the principal investigator and that is funded by an affected or competing firm at issue.

We would like to remind members and consultants that if the discussions involve any other products or firms not already on the agenda for which an FDA participant has a personal or

imputed financial interest, the participants need to exclude themselves from such involvement, and their exclusion will be noted for the record.

FDA encourages all other participants to advise the panel of any financial relationships that they may have with any firms at issue. A copy of this statement will be available for review and will be included as part of the official transcript. Thank you.

For the duration of the anesthesiology and respiratory therapy devices panel meeting on February 2nd, 2024, Tamorah R. Lewis, MD, has been appointed to serve as a temporary non-voting member. For the record, Dr. Lewis serves as a consultant to the Pediatric Advisory Committee in the Office of Pediatric Therapeutics, Office of the Commissioner. This individual is a special government employee who has undergone the Customary Conflict of Interest Review and have reviewed the materials to be considered at this meeting. The appointment was authorized by Rachel Bressler, Acting Director, Advisory Committee Oversight and Management Staff, on January 4th, 2024.

Before I turn the meeting back over to Dr. Cassiere, I'd like to make a few general announcements. In order to help the transcriber identify who is speaking, please be sure to identify yourself each and every time that you speak. The press contact for today's meeting is Carly Kempler. Thank you very much. Dr. Cassiere?

Dr. Cassiere: Good morning. Dr. Gooden, I was negligent in my chairman duties by not introducing you. If you could introduce yourself to the committee, I apologize.

Dr. Gooden: I think you're probably referring to me.

Dr. Cassiere: Yes.

Dr. Gooden: I'm Cheryl Gooden, and my area of expertise is anesthesiology with a subspecialty of pediatric anesthesiology. I am an associate professor of anesthesiology and pediatrics at Yale

University School of Medicine, and I'm a pediatric anesthesiologist at Yale New Haven Children's Hospital.

Dr. Cassiere: Thank you. Thank you, Ms. Nalls, for that introduction comment. At this time, Dr. Shuren, Center Director for the Center of Devices and Radiologic Health, and Dr. Lee, Acting Associate Commissioner for Minority Health, will provide the opening remarks. Dr. Shuren.

FDA Opening Remarks – Dr. Jeff Shuren

Dr. Shuren: Thank you for joining us for today's virtual public meeting of the Anesthesiology and Respiratory Therapy Devices Panel of the Medical Devices Advisory Committee.

Pulse oximetry is a vital public health tool that plays a role in everyday health care for many individuals. Disparate performance of pulse oximeters across people with different amounts of skin pigmentation, racial and ethnic groups, is of great importance to public health and the FDA. Although pulse oximetry is useful for estimating blood oxygen levels, pulse oximeters have limitations and may be less accurate under certain circumstances that should be considered.

Today's meeting will include discussions about potential approaches 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. The committee will also discuss the type and amount of data that should be provided by manufacturers to the FDA to evaluate the performance of pulse oximeters for premarket review, including prescription and over-the-counter devices and labeling considerations to ensure pulse oximeters are safe and effective for all patients and used appropriately.

In 2022, the FDA's Center for Devices and Radiological Health made advancing health equity a strategic priority, and we have prioritized actions to better understand the needs of different populations and the challenges they face in accessing health care and health technologies. We are committed to the continued evaluation of the safety, effectiveness, and availability of medical

devices, including with respect to how devices may perform differently across racial and ethnic groups.

The FDA has taken proactive steps to improve premarket evaluation strategies and equitable device performance to help improve the accuracy of pulse oximeters across all US patient populations and demographics. We provided a 24-hour summary of the advisory committee meeting we held in November 2022 to keep the public informed about the discussion topics and the panel's recommendations at that time and issued several public communications in June, September, and November of 2022 to continue efforts to keep the public informed of the FDA's actions related to pulse oximeters.

Additionally, in advance of today's public meeting, we published a discussion paper this past November titled Approach for Improving the Performance Evaluation of Pulse Oximeter Devices Taking into Consideration Skin Pigmentation, Race, and Ethnicity. This paper was informed by the discussions from the November 2022 meeting of this advisory committee, where stakeholders shared information and perspectives about ongoing concerns that pulse oximeters may be less accurate in individuals with darkest skin pigmentation. We've also worked closely on two FDA-funded real-world evidence studies at the UCSF Stanford Center for Excellence in Regulatory Science and Innovation.

The purpose is to prospectively evaluate the performance of pulse oximeters in adults and children using simultaneous oximetry measurements and objective skin pigmentation measurements. This work aims to address some of the limitations of previously published real-world studies. The FDA has engaged stakeholders, gathered input from ongoing clinical research, and evaluated all available information pertaining to factors that may affect pulse oximeter accuracy and performance to inform the content of our November 2023 discussion paper and today's meeting.

The feedback we receive will inform changes we may make to our Pulse Oximetry Guidance document. Pulse oximeters, premarket notification submissions, guidance for industry and Food and Drug Administration staff. We will keep the public informed as significant new information on recommendations becomes available.

In the interim, the public should continue to refer to the FDA's recommendations in our November 2022 safety communication titled Pulse Oximeter Accuracy and Limitations. I'd like to thank the members of the Advisory Committee for their participation. We look forward to today's discussion and your important feedback. Thank you.

FDA Opening Remarks – Dr. Christine Lee

Dr. Lee: Good morning. I'm Christine Lee, Acting Director for the Office of Minority Health, and Health Equity. I'm glad to be able to join you today for this meeting on such an important topic. The ongoing efforts to better understand and evaluate the relationship between race, ethnicity, skin pigmentation, and oximeter accuracy is closely related to advancing health equity and addressing disparities among racial and ethnic minority populations, which are priorities across the FDA, including CDRH and the Office of Minority Health and Health Equity. We know that variations associated with race and ethnicity have been correlated with risks for certain diseases, conditions, and responses to regulated products. Understanding these factors remains important for identifying and addressing health disparities.

The FDA Office of Minority Health and Health Equity has been engaged broadly across the agency and as well as with external stakeholders in protecting and promoting the health of racial and ethnic minorities and other diverse populations through research and communication. Among some of our most recent activities include the Enhance Equity Initiative, which highlights research projects and communication resources to enhance equity in clinical trials by supporting efforts to advance diversity in clinical trials; equity of data efforts by increasing

research studies about diverse groups, including but not limited to ethnicity, race, age, disability, and geography; and equity of voices by amplifying FDA's communication with diverse groups to ensure stakeholders, including consumers, are informed about FDA's efforts and to understand diverse patients perspectives, preferences, and unmet needs.

Throughout the FDA, information on race and ethnicity has and continues to be integral to our understanding of the health issues affecting the US population and support of improving population health outcomes. We are committed to the continued evaluation of the safety, effectiveness, and availability of regulated products, including how medical devices perform across racial and ethnic populations.

Today's discussion on the performance of pulse oximeters across racial, ethnic, skin pigmentation groups and considerations for pulse oximeters intended for OTC use will inform the FDA as we consider the regulation of these medical devices. Advancing health equity is a top priority for the FDA, CDRH, and the Office of Minority Health and Health Equity, and the performance of the medical devices needs to be well understood to mitigate any negative unintended consequences for all populations. With that, again, I'm glad to be able to join you today, and I look forward to this important discussion. Thank you.

FDA Presentation: Pulse Oximeters: Technology, Accuracy Limitations and Regulations

Dr. Cassiere: Thank you. At this time, we will view the FDA's presentation on today's meeting topic. I would like to remind public observers at this meeting that while this meeting is open for public observation, public attendees may not participate except at the specific request of the panel chair. FDA, you may now begin your presentation.

Dr. Lee: Good morning. I am here to present regarding the agency's regulations of pulse oximeters, which includes the reviewing of the technology, its limitations on accuracy, and how the agency regulates the device type. My name is James Lee. I am the division director here in OPEQ OHT1, Division for Sleep Disordered Breathing, Respiratory and Anesthesia Devices, with my co-authors, Mr. Bradley Quinn, the assistant director, and Mr. Neil Patel, our senior reviewer in the anesthesia devices team. Pulse oximetry is an essential tool that indirectly measures arterial oxygen saturation in real-time monitoring. Pulse oximeters are widely used by many types of healthcare providers and consumers to obtain an indirect measurement, SpO_2, of arterial blood oxygen saturation, SaO_2.

It is an expedient measurement of oxygen level in the hemoglobin, where SaO_2 measurements are obtained by actual arterial blood collection through punctures, which is the gold standard for measuring oxygen levels. In comparison, SpO_2 estimates the oxygen levels in the blood hemoglobin and is expressed as a percentage.

The COVID-19 pandemic resulted in a significant increase in the use of pulse oximeters in both hospital and home settings. Pulse oximetry is based on two physical principles: the presence of a pulsatile signal generated by arterial blood and the fact that oxygenated hemoglobin and reduced hemoglobin have different absorption spectra. As mentioned on the prior slide, SpO_2 is estimated as a percentage of oxygenated hemoglobin over oxygenated and deoxygenated hemoglobin.

Oxygenated and deoxygenated hemoglobin have different absorption spectra, which allows the use of optical techniques using two or more different wavelengths of light to measure differences in absorption. For example, in typical wavelengths of light used to obtain SpO_2 measurements, red light at 660 nanometers and infrared light at 940 nanometers, the total absorption is relatively low, allowing enough light to pass through from the emitting LEDs to the

photodetector as shown in the upper right-hand figure. Oxygenated, defined as O_2Hb in the lower figure absorbs or attenuates the amount of infrared light more than compared to the red light. The opposite is true for deoxygenated or reduced hemoglobin, which absorbs more red light and allows relatively more infrared light to pass.

The second basic principle of pulse oximetry is the presence of a pulsatile arterial signal, which allows changes in light absorption to be measured. The upper right-hand figure shows the changing absorbance of light during an arterial pulse. The portion that changes is often referred to as the AC portion of the signal. The portion that remains constant due to the residual arterial blood, venous blood, and other tissue is referred to as the DC portion of the signal with a ratio of the AC to DC signal for both red and infrared light. A ratio of rate ratios, referred to as an r value, which again is a ratio of AC to DC for the red signal value or the ratio of AC to DC for the infrared signal, can be calculated with the calculated r value.

You can estimate the SpO_2 value using empirical calibration data, as shown in the lower figure. The relations of pulse oximeters are based on a moderate device level of risk, meaning these devices are reviewed under the FDA CDRH 510(k) program and cleared under the basis of substantial equivalence as described in our 2013 FDA guidance document for pulse oximeters. These devices undergo clinical testing, bench testing, and other standardized tests like biocompatibility to review the device's relative safety and effectiveness against predicates with the same intended use. These devices are considered in terms of spot checking or trending tools and not for diagnosis of a disease.

A cornerstone of the FDA guidance document and the review is the use of internationally recognized standards for the evaluation of pulse oximeters. This standard establishes the basic safety testing needed and defines protocols and tests that are utilized in the evaluation of pulse oximeters to find performance.

We would recommend the use of updated consensus standards and the FDA guidance on oximeter validation activities when using SpO_2 either as a standalone physiological measurement or within the context of a multi-parameter monitor or where a device may be using SpO_2 information in a custom device. Pulse oximeters intended for medical purposes are labeled for trending or spot-checking for oxygen saturation levels of patients in hospitals and doctor's offices, although they may sometimes be prescribed for home use or available over the counter. Regarding what is considered general wellness, regulation of these devices is under exemptions that fall under product codes like PGJ or OCH. These are specifically for general wellness, which apply to sports and aviation uses, but are not intended for medical purposes. For this reason, they do not undergo FD premarket review. Please see our reference here at the link for our general wellness guidance on low-risk devices.

Pulse oximeters intended for medical purposes are class two devices intended to measure blood oxygen saturation levels, and the FDA reviews medical technology like oximetry under established regulations under the Code of Federal Regulations or CFR. Here, we would like to provide the established definitions of these regulations.

Furthermore, the FDA buckets or groups within these regulations by product codes, which allow for a more granular definition of products and allow for detailed assignment according to both established regulations and sorting of these devices by product code. What should be noted is that a symmetry device utilizes both visible and invisible wavelengths of light and relies on absorption by tissues where the remaining signal can be transmitted to an opposing sensor.

Innately, configurations like this, while effective in picking up physiological signals, have limitations due to the technology and configurations of the sensor system. Highlighted here are several confounding factors that may contribute to either the inaccuracy of the signal or

increase the variability in relation to ground truth or an absolute oxygen saturation in the hemoglobin. One of the main considerations is a level of skin pigmentation, which is a major part of our discussions. In addition, there are patient dispositions that may raise or challenge the accuracy of the SpO_2 signal due to the ratio of oxygenated and deoxygenated hemoglobin, which does not include consideration for other types of hemoglobin. Therefore, disorders of the hemoglobin molecule, or anemia, may cause a reduction in pulse oximeter accuracy.

In addition to other factors, like intravascular death eyes, tattoos, or nail polish, an ambient light may interfere with the sensor and contribute to lower accuracy levels. 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 that reported for DQA, DPZ, and NLF were searched for reports received between January 1st, 2002, October 25th, 2023, which are solely for prescription use. Then, a text search was used to identify any report with the term skin, and a code search was used to identify any reported submitted as a death report and then evaluated. Each report was identified through the text search, and each death report was then reviewed to determine if it was relevant to an inaccurate SpO_2 readings and assess the potential for sources for inaccurate readings.

The initial search by the three product codes yielded 12,248 adverse event reports. The figure here represents the number of total reports and the number of adverse event reports submitted through the MDR systems for product codes DQA, DQZ, and NFL from January 1st, 2000 through October 25th, 2023. Most adverse event reports were classified as malfunctions at 91.5%, followed by serious injury reports at 4.4% and death reports at 2.5%. 40% 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 total, while oximetry has provided an excellent tool in evaluating non-invasively important vital signs like oxygenation levels, the limits of utility do exist in measurements, which have clinical implications for use. Thank you.

FDA Presentation: A Systematic Literature Review of the Real-World Performance of Pulse Oximeters

Dr. O'Neill: Good morning. My name is Allison O'Neill, and I'm the Associate Director for Post market Surveillance in OHT1.

I will be presenting a systematic literature review of the real-world performance of pulse oximeters. First, I will begin by explaining the purpose of the review. The scientific community has been aware for some time that certain patient factors may impact the accuracy of pulse oximeters, and these factors may include skin pigmentation.

In 2013, the FDA published a guidance document, which made recommendations for manufacturers, specifically that each premarket study should have subjects with a range of skin pigmentation, including at least two darkly pigmented subjects or 15% of the subject pool, whichever is larger. Between 2013 and 2020, there was some limited literature on this subject. In 2020, the COVID-19 pandemic brought increased awareness of the importance of pulse oximetry, both in the hospital and home settings. In December 2020, Sjoding et al. published a correspondence in the New England Journal of Medicine, which reported that Black ICU patients had approximately three times the rate of occult hypoxemia compared to white patients; the study received media attention and spurred additional interest and related research studies.

The FDA wanted to provide the public with information on the interpretation and limitations of pulse oximetry. In February 2021, the FDA released a safety communication with recommendations for providers and patients. At that time, the FDA committed to the continued

evaluation of the safety and effectiveness of pulse oximeters. A systematic literature review was performed for the advisory committee meeting in 2022 and updated in 2023. In November, the FDA published a discussion paper about a new approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters with a request for stakeholder feedback.

Today, I will briefly present a high-level overview of relevant literature published since 2013. This review includes literature previously presented at the 2022 Advisory Committee meeting, as well as articles published between 2022 and 2023. We searched the PubMed database to identify relevant articles. A total of 223 potential publications were identified from the search, plus additional cross-referencing. We excluded articles published before the 2013 FDA guidance document. We also excluded publications that were not relevant to the topic, did not include any clinical data, or did not include skin pigmentation or race-ethnicity data.

A total of 46 publications were selected for inclusion in the review. I will now briefly describe the overall body of literature, but in the interest of time, I will not cover each individual study. For more details, please see Appendix 1 of the Executive Summary. Of 46 selected articles, there were seven systematic reviews and eight lab studies. There were also 31 publications that used real-world data from hospital patients, usually in patients from the ICU or surgical unit. Of the 31, nine were considered cross-sectional studies that prospectively collected data, and 22 were retrospective studies that relied on electronic health record data. Most studies were conducted using US patients, although there were some from Europe, Asia, Africa, and Australia.

Seven systematic reviews pertaining to the review topic were published between 2022 and 2023. There was some overlap in the articles that were included in each review. Six of the

seven total reviews concluded that there were overestimations in people with darker skin pigmentation based on the evidence reviewed.

Systematic reviews typically represent a higher level of evidence than one study alone. Of course, it should be noted that literature reviews inherently have the same limitations as the studies that are included in the review. Also, some of these reviews contain articles going back to the 1970s and 80s, which is a larger time frame than considered for the FDA's review. Regarding the nine cross-sectional studies, each study reported either the bias between SpO_2 and SaO_2 or an odds ratio for occult hypoxemia. Seven of nine papers reported a positive bias or statistically significant odds ratio for patients with darker skin pigmentation or African American race compared to the reference group.

The 22 retrospective studies were more varied in design and chosen endpoints. Overall, most studies reported a statistically significant association between race and either occult hypoxemia or SpO_2 bias. Additionally, some individual studies also reported an association with certain clinical endpoints, including treatment probability, organ dysfunction, and in-hospital mortality. All studies relied on self-reported race and ethnicity data from electronic health records.

The Lit Search identified seven laboratory studies using healthy adult volunteers who underwent controlled desaturation. The eighth was a pooled retrospective analysis of nine other desaturation studies. Overall, there was evidence that PulseOx performance was affected by motion, perfusion index, skin pigmentation, and degree of hypoxemia.

Now, I will discuss the important limitations that should be considered when assessing the published literature, and especially real-world data. First, study variables were defined differently by study. There was no standardized cutoff for what SpO_2 and SaO_2 levels merit classification as occult hypoxemia. This may be defined differently by site based on clinical

need. Skin pigmentation data may be captured by different tools, although race and ethnicity was also often used as a proxy. Secondly, PulseOx accuracy often appears worse in real-world studies than in desaturation studies. Using real-world data means that the paired measurements of SaO_2 and SpO_2 are usually not simultaneous, which may lead to larger disagreements due to normal fluctuations and treatment effects on SaO_2.

Also, researchers must often rely on self-reported race ethnicity, as skin pigmentation data such as Fitzpatrick scores are typically not collected for hospital patients, and there may be residual confounding from variables not captured. Also, in a controlled lab study, researchers may try to set up ideal testing conditions, such as warming the participant's hands and waiting for a 30 second stable plateau.

Thirdly, the real-world population is more heterogeneous than a controlled lab study, typically including very sick patients from an ICU population rather than healthy volunteers. The prevalence of hypoxemia will vary between such groups, impacting the positive predictive value and negative predictive value and confounding comparisons of the occult hypoxemia rate. Fourth, heterogeneity of technology used. Pulse Ox accuracy will most likely vary depending on which device or product is used, and this data was not provided in every article. It can be assumed that brands, models, and the use of reprocessed sensors varied across hospitals and even within the same hospital. Additionally, technology advances, and thus, pulse Ox accuracy may have changed over time.

Finally, there is inherent publication bias in any literature review, which means that, in general, results showing statistically significant associations are more likely to be submitted and accepted for publication compared to results that do not show significant associations. Overall, despite the limitations noted, there is real-world evidence from studies capturing tens of thousands of patients that suggest the pulse oximeter accuracy varies by self-reported race and or

skin pigmentation. Eighteen additional relevant articles were published since the 2022 Advisory Committee meeting, demonstrating the recent increasing interest in this topic.

However, there is still a need for additional 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 such as perfusion index in order to have more robust evidence about the impact of skin pigmentation on real-world pulse oximetry.

Thank you to our panel members for participating in today's meeting. Next, you will hear the FDA's proposed approach to improve the premarket clinical study of pulse oximeters.

Proposed Approach to Premarket Clinical Study

Dr. Hendrix: Good morning. I'm Kumudhini Hendrix, an anesthesiologist and currently OHT1 chief medical officer within OPEQ CDRH. Today, I'll be presenting the agency's proposed approach to improve premarket clinical study. First, I'll give a brief overview of the current premarket study. Next, I'll present the agency's proposed approach to address non-disparate performance of prescription and over the counter for medical purposes pulse oximeters.

While clinical study and trial are used interchangeably in the current pulse oximeter guidance, today's meeting and the discussion paper, for discussions today, we're referring to the premarket desaturation clinical performance study for Annex EE of ISO standards 80601-2-61 for pulse oximeters within the scope of our 2013 FDA pulse oximeters premarket notification submission guidance, in vivo premarket desaturation testing for SpO_2 accuracy is recommended under laboratory conditions for all new pulse oximeters, as well as for all prior cleared pulse oximeters with significant electrooptical sensor modifications and or SpO_2 algorithm modifications. The objective of in vivo premarket desaturation studies is to verify the SPO two

accuracy of pulse oximeter device performance in comparison to the gold standard measurements of blood SaO_2 by an oximeter over the specified SaO_2 range of 70 to 100%.

Therefore, typical labeling is a general indication for noninvasive measurement of blood oxygen saturation. If a manufacturer wishes to seek a specific clinical indication for the use of a pulse oximeter, for example, to screen for or diagnose a disease or condition, then additional clinical safety and effectiveness data is requested to be submitted.

Guidance currently recommends that the premarket desaturation study should include a sufficient number of participants to attain statistical significance necessary to demonstrate a specified SpO_2 accuracy. A minimum of 200 pool data pairs—that's SpO_2 and SaO_2—distributed evenly over the tested range of 70 to 100% from a minimum of ten healthy volunteers is recommended.

As for FDA guidance, the study should have participants that range in age, gender, and skin pigmentation. FDA guidance recommends at least two participants, or 15% of the participant pool, whichever is greater, to be darkly pigmented. Per ISO standards, carboxy hemoglobin should be less than 3%, methemoglobin less than 2%, and total hemoglobin greater than ten grams per deciliter.

Testing conditions for premarket desaturation studies can include the application of warming techniques to improve circulation and pulse amplitude at the pulse oximeter probe site, covering of pulse oximeter probes with opaque material to prevent optical interference, and the addition of carbon dioxide gas to an inspired gas mixture to maintain normocarbia and to prevent respiratory alkalosis secondary to hyperventilation caused by hypoxemia.

A catheter is placed within the artery on all test participants prior to desaturation. A cleared pulse oximeter is placed as a reference to detect stable plateaus. The fraction of inspired oxygen is varied in a stepwise manner to achieve a series of targeted steady-state saturation

periods. When the reference pulse oximeter stabilizes for 30 seconds or more, arterial blood is sampled for comparison of simultaneous data pairs; that is premarket-device SpO_2 and SaO_2.

Per FDA guidance, all the following data as it pertains to premarket desaturation testing is requested to be submitted for FDA review: Pertinent test apparatus used; inclusion-exclusion criteria; number of samples taken per subject; specific conditions of testing such as motion, low pulse amplitude, and laboratory conditions; type and frequency of motion testing if applicable, criteria and methods for determining stability of reference arterial blood oxygen saturation at the pulse oximeter sensor site; desaturation profile, including target plateaus and ranges; as well as formulae used for determination of root mean square difference—that is ARMS.

Additionally, individual data pairs, as well as pool data pairs, are requested to be plotted in a modified bland Altman plot to show arterial saturation on the x-axis, as it is considered ground truth. These plots, as on the right side of the slide, have SaO_2 on the x-axis and the difference between SpO_2 and SaO_2 on the y-axis.

FDA guidance also requests population mean bias, that is, mu zero; between-subject variance, that is, sigma mu squared; and within-subject variance, that is, sigma squared; as well as upper 95% and lower 95% limits of agreement.

Some of the limitations of the current premarket clinical study are the following. Subjects are healthy volunteers selected from a pool of limited volunteers. Sample size is not large, typically ten to 20 subjects. Skin pigmentation is generally qualified subjectively, such as light, medium, or dark, and not specific to any particular anatomical sites, for example, the dorsum of the hand or ventral aspect of the forearm. Additionally, SpO_2 accuracy is pooled point estimate accuracy, that is, ARMS across the entire tested range of SpO_2 without threshold accuracy analysis. Importantly, it is not powered to determine significant differences between cohorts, for example, pigmentation levels.

On November 1, 2022, the FDA convened the anesthesiology device panel to discuss ongoing concerns that pulse oximeters may perform disparately in individuals with darker skin pigmentation and to garner feedback on ways to improve premarket clinical studies. The panelists concluded that current available evidence does demonstrate disparate performance in patients with darker pigmentation.

They recommended that the full spectrum of skin pigmentation be assessed with subjective and objective pigmentation methods. They recommended that self-reported race and ethnicity be included in premarket clinical studies. They asked for an adequate sample size to ensure equitable performance. They recommended that overall performance be tightened as much as feasible and suggested within 2% and asked for a more clinically meaningful accuracy metric.

They raised concern over the inaccurate and unverified performance of wellness pulse oximeters and recommended clear labeling for over-the-counter devices. The agency's proposed approach was informed by prior anesthesia device panel recommendations, statistical modeling of desaturation lab data on current and accurate pulse oximeters, and UCSF Stanford Center for Excellence in Regulatory Science and Innovation Real World Studies.

As outlined in the discussion paper, our approach recommends to increase the minimum sample size of study participants from ten to 24 and to increase diversity representation by including the entire range of skin pigmentation and accounting for race and ethnicity. Additionally, we also propose to improve the overall accuracy, assure non-disparate performance across the entire range of pigmentation, as well as recommend diagnostic performance analyses at important clinical thresholds, such as SpO_2 at 90%.

We are proposing the same approach for all pulse oximeters for medical purposes. Currently, accuracy or ARMS has been evaluated using a point estimate of the root mean square

error pooled over the full range of SaO_2 measurement. The uncertainty in how well the point estimate Or ARMS represents the population—that is, the true ARMS—has not been considered to increase certainty in our evaluation of accuracy. We propose that a 95% confidence interval for ARMS be reported. A 95% confidence interval on ARMS will include the true population of ARMS and 95% of applications.

Additionally, we're proposing a tighter accuracy. In the 2013 pulse oximeter guidance, pulse oximeter accuracy was considered acceptable if the point estimate for ARMS is 3% or less. We now propose that pulse oximeter accuracy be considered acceptable if ARMS is significantly less than 3%. For our proposed sample size of at least 24 subjects, only pulse oximeters with a true population ARMS of 2.1% or less are likely to have a 95% CI of less than 3%. A true population ARMS of 2.1% or less, aligns closely with the advisory panel recommendations that ARMS be tightened to within 2%. Given the limits of current technology, ARMS cannot be expected to be very much lower than 2.1%.

Additional to tightening up accuracy and performance, we're harmonizing both transmittance and reflectance pulse oximeters to the same accuracy performance metric. The 2013 guidance recommends to have at least ten subjects varying by age, gender, and pigmentation. It specifies that at least two or 15% of the participant pool, whichever is larger, to be darkly pigmented. Based on our statistical modeling of desaturation lab data on current and accurate pulse oximeters, we propose at least 24 healthy subjects with a minimum of 480 data pairs. We recommend that at least 40% of participants be of each gender. Importantly, the agency is recommending pigmentation assessment using subjective and objective pigmentation methods.

While the current guidance recommends a range of skin pigmentation with at least two darkly pigmented or at least 15% of the participant pool, whichever is greater, the agency now

proposes a two-tiered pigmentation evaluation using both a subjective Monk Skin tone scale with ten values, referenced as MST from this point on, be evaluated on the forehead. After an even enrollment of participants across the entire MST scale, the agency recommends that the objective individual typology angle, referenced as ITA from now on, be measured at the amateur sensor site. I'll be presenting the details of this approach in the next few slides.

MST is a ten-valued subjective scale, which has been validated to capture race and ethnicity diversity and pigmentations within the US. It has been found to be more inclusive of the spectrum of skin tones seen within the US demographics than other subjective scales such as, for example, the Fitzpatrick scale. Additionally, MST is standardized to color scales such as the CIE Lab. This relationship is demonstrated on the right side of the slide.

Of note, MST has been shown to have a high interclass coefficient of being within one point across the entire MST scale, even among a global pool of raters. Colorimetry is the most common and well-standardized approach for objective evaluation of pigmentation. Standard colorimetry methods, such as by the CIE Lab, are used to measure colorimetric parameters, where L star is luminance, A star is the red-green component, and B star is the yellow-blue component.

The objectively measured variables can then be used to calculate ITA by using the formula on the slide. ITA provides an objective, continuous, quantitative measure of skin pigmentation. The validity of ITA as a strong correlate of melanin content has been confirmed in clinical studies using histological analysis of biopsied skin samples.

After considering the totality of best and current evidence on an inclusive and representative pigmentation assessment method and feedback from stakeholders, we recommend that during recruitment of study participants, MST be evaluated on the forehead. During recruitment, we recommend that participants from diverse racial and ethnic groups be

maximized. Next, participants are recommended to be grouped into three MST cohorts, namely MST values one through four in one, MST values five through seven in the next, and MST eight through ten values in the final cohort.

By assigning at least 25% of total participants per MST cohort, we hope to ensure an even spread of participants across the MST scale. To further ensure that not all subjects within each cohort come from only one MST value, we propose there be at least 15% of participants of a cohort or one subject for each value, whichever is larger. An example of distribution is provided in blue. Assuming a sample size of 24 participants, enrollment could be six participants, that is 25%, for MST cohort one through four, nine participants for the other two MST cohorts, that's five through seven and eight through ten. And within, say, the MST one through four cohort, by requiring at least one participant for each value, there will be at least one participant with an MST value of 1.

After steady participants are recruited and evenly distributed across the entire MST values, we recommend that ITA be measured at the emitter sensor site. For fingertip pulse oximeters to allow for the widest range of emitter site pigmentation, we recommend measuring ITA on the pigmented skin at the midline of the dorsal distal phalanx, notated here in the slide with an orange circle. The purpose of this objective measurement is to use it as a continuous variable to assure non-disparate performance.

In addition to the primary analysis of ARMS, we now propose a new co-primary analysis for non-disparate performance assurance. We propose that pulse oximeter performance is considered non-disparate when the absolute value of the maximum difference in SpO_2 bias across both continuous ITA and categorical MST levels is less than 1.5% for SaO_2 ranges above 85% and less than 3.5% for SaO_2 ranges between 70 and 85%. The performance goals of 1.5 and

3.5% were considered the smallest achievable values for current and accurate pulse oximeters based on statistical modeling of desaturation lab data.

Desaturation lab data suggests that for accurate pulse oximeters, the true maximum differences in bias will be about 0.5% and 2% across the two So_2 ranges, respectively. And that the estimates of these two differences will meet the performance goals of 1.5% and 3.5% with 80% power in a study with 24 subjects and 480 total data pairs of SaO_2 and SpO_2 distributed evenly across the SaO_2 uniformly from 70 to 100%. You will hear the details of the statistical plan in the next FDA presentation.

It remains uncertain whether pulse oximeter performance is disparate between individuals from different race and ethnicity groups but with the same level of sensor site pigmentation. Until there's certainty on whether racial and ethnic differences within the same level of pigmentation give rise to non-disparate pulse oximeter performance, the agency considers it important for premarket clinical studies to include participants from diverse racial and ethnic groups. We recommend maximizing racial and ethnic diversity during enrollment and reporting per FDA guidance on the collection of race and ethnicity data in clinical trials. FDA notes that MSD is not a proxy for race and ethnicity diversity. However, MSD has been validated to capture race and ethnicity diversity and pigmentation. FDA notes that an even distribution of participants across the entire MST scale will ensure a certain amount of racial and ethnic diversity. Therefore, meeting acceptance criteria across the MST scale may ensure the evaluation of non-disparate performance across those race and ethnicity groups.

Noting that certain SaO_2 values are more clinically relevant, the agency is considering recommending additional analysis for non-disparate performance at such thresholds and recommend reporting receiver operating characteristic, that is, ROC curve; and area under that

ROC curve, that is, AUC; for overall, for MST cohorts, and for ITA values. An example of a target condition of SaO_2 less than 90% across ITA values is shown on the right side of this slide. The FDA is considering the same premarket clinical study design and non-disparate performance evaluation for over-the-counter pulse oximeters for medical purposes as for prescription use pulse oximeters. We recommend that labeling include information about the premarket clinical study design and a description of the device performance.

The panel will be asked to discuss the 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 devices used for medical purposes. Thank you.

Assessment of Skin Pigmentation

Dr. Pfefer: Good morning. I'm Josh Pfefer, a biomedical research engineer at CDRH. The focus of this presentation is the assessment of skin pigmentation in pulse oximetry studies using objective and subjective methods.

Pulse oximeters operate based on the principles of tissue optics, and the scientific literature indicates that nearly all transdermal optical sensing technologies are impacted by epidermal melanin in one way or another. As light propagates through tissue, it is scattered by components such as cells and absorbed by constituents such as hemoglobin, water, and melanin. While melanin is limited to the epidermis layer, it's a dominant absorber that can reduce detected light by up to 70% and alter both the spectral distribution of light and pathways taken by light propagating through tissue. Anatomic variations in melanin are also significant, with sites such as the fingernail and palmar finger containing relatively low levels compared to sites such as the forehead and the skin proximal to the fingernail.

In determining the optimal approach for assessing skin pigmentation in pulse oximetry studies, it's worthwhile to first consider the potential mechanisms at play. Perhaps the most widely discussed mechanism involves optical absorption by melanin, leading to changes in light intensity and or spectral content. If this is the true dominant mechanism. Then, evaluating robustness with respect to epidermal melanin content at the sensor site likely provides the best opportunity to detect any performance disparity that exists. Dr. Ellis Monk's research has indicated that perceived colorism may impact patient health and, in turn, device performance. Thus, evaluating skin tone based on the forehead may provide the best way to detect disparity. A third mechanism that has been proposed is genetically correlated physiological traits, such as differences in vascular [indiscernible] In this case, a metric based on melanin content at a constitutive pigmentation site may be optimal. Other factors may also degrade pulse oximeter performance.

FDA has considered a variety of skin pigmentation assessment methods based on subjective and objective techniques, starting with subjective approaches that have been used in pulse oximetry studies. There is racial ethnic self-identification, skin color descriptors, the well-known Fitzpatrick approaches, as well as more well-standardized color scales. Then there are the more rigorous objective optical methods based on spectroscopy and colorimetry, as well as gold standard validation approaches based on skin biopsy to quantify melanin content.

Over the past 30 years, instruments that use visible to near-infrared light to objectively assess skin pigmentation have been implemented extensively for the development of cosmetics, as well as scientific research studies in anthropology and medical disciplines such as dermatology. A variety of commercial systems based on spectroscopy and colorimetry are marketed in the US and abroad for research. These instruments are often portable and can provide data on specific tissue sites, such as the palmar or dorsal finger. However, they

implement a range of illumination, detection, and processing approaches and generate different metrics, some of which are not well standardized.

Colorimetry represents the most widely used approach for quantifying skin pigmentation. Standards developed by CIE describe best practices for colorimetry, including color-matching functions used to simulate the human visual system. The CIE Lab is a commonly used color space that defines variables, including L star for light-dark levels, A star for red-green levels, and B star for yellow-blue levels. Variables L star and B star are then used to calculate a parameter called the Individual Typology Angle, or ITA, which quantifies pigmentation.

Human subject studies have provided extensive evidence that colorimeters exhibit a high degree of intra and inter-observer repeatability, as well as inter-instrument repeatability, with many studies showing inter-class correlation coefficients of 0.9 or greater. In several studies, clinical ITA values have been compared with gold standard measurements of epidermal melanin content. Histological stain methods such as Fontana Masson enable quantification by assessment of melanin volume fraction in the epidermis and basal layer or through spatially integrated transmittance.

Another approach is high-performance liquid chromatography, or HPLC, which provides high, highly sensitive, and specific quantification of key eumelanin and pheomelanin degradation byproducts in homogenized tissue samples. Several studies have demonstrated a strong linear correlation between ITA and these gold-standard measurements of melanin content, which indicates that colorimetry should provide an effective tool for evaluating skin pigmentation in pulse oximetry studies.

How do we ensure that a specific colorimeter is accurate? First, we can standardize key system parameters. As recommended by colorimetry standards, devices would ideally use D65

illumination, simulating average daylight and aperture sizes of approximately 3 to 8 millimeters while rejecting specular reflection.

Colorimeter validation should be performed based on human studies with a diverse population, in comparison to a gold standard measurement system, such as a high-quality spectroscopy-based system. However, it may also be acceptable to use established color charts to perform this validation. Preliminary research findings appear to support the idea that skin pigmentation effects play a primary role in the disparities documented in clinical studies. Numerical modeling at the FDA and other institutions indicates that increasing levels of melanin can lead to similar levels of bias, as shown in clinical studies. Additionally, a preprint manuscript by Fauzy et al. provides clinical evidence that recruiting and analyzing data based on ITA measurements alone can reveal significant SpO_2 measurement disparities. If epidermal melanin is proven to be the primary cause of SpO_2 disparities, this may have important implications for the design of future clinical studies.

In terms of subjective assessment approaches, one of the most well-known is the Fitzpatrick Skin Phototype Scale. However, no well-standardized color chart for each phototype has been established, and this method tends to show poor correlation with pigmentation, particularly in subjects with darker skin. Other partially standardized approaches have included the Von Luschan Scale, subjective methods with high quality, and commercially available color charts, including Munsell and Pantone. The recently developed Monk Skin Tone scale is standardized to multiple color systems, including the CIE Lab.

The Monk scale was developed to evaluate the role of perceived colorism through the use of facial skin tone as an indicator of patient health. One of the primary advantages of MST is that it's standardized to CIE Lab and RGB parameters, thus facilitating accurate reproduction of the color scale. Studies have indicated that MST provides better-perceived representativeness than

the Fitzpatrick scale and a high level of inter-rater reliability. It's also worth noting several shortcomings of this approach, including a lack of commercially available color charts, minimal color differences between MST levels at very low and very high pigmentation ends of the scale, and potential susceptibility to inter-observer bias and errors due to variable lighting. We will hear from Dr. Monk later today about his scale, and we will also hear from our UCSF colleagues about a real-world pulse oximetry study involving the MST scale.

Another important consideration is the anatomical sites used for pigmentation assessment. The MST scale was developed to assess the forehead. This site provides a wide range of pigmentation levels, enabling strong visualized differentiation. In colorimetry measurements, the intent is to assess pigmentation in locations that directly interact with light from the pulse oximeter, such as the fingernail and palmar finger. However, the range of pigmentation levels in these sites is relatively small, which complicates the differentiation of subjects with different skin tones. Since light delivered by the pulse oximeter interacts with other regions of the skin, sites adjacent to the fingernail, where the pigmentation range is much greater, can be used. Overall, the agency believes that ITA and MST approaches should provide a high degree of effectiveness in studies evaluating pulse oximeter robustness to skin pigmentation. This afternoon, the panel will be asked to discuss and make recommendations about the assessment and reporting of skin pigmentation data in studies evaluating the accuracy of pulse oximeters.

Statistical Considerations

Dr. Pennello: Good morning. I'm Gene Pennello, a mathematical statistician at the Center for Devices and Radiological Health at FDA. My presentation is on statistical assessment of pulse oximeter performance.

Many aspects of pulse oximeter performance may be considered. After an introduction to the concept of an individual agreement, I will discuss descriptive plots, the proposed co-primary analyses of accuracy root mean square error, and non-disparate performance assurance, and other potential objectives. First, individual agreement. The error, the pulse oximeter value, SpO_2, and estimating the arterial oxygen saturation value SaO_2 is the difference D equals SpO_2 minus SaO_2. SpO_2 bias is the expected value of D. In other words, the mean of D across repeated measures. SpO_2 imprecision is the standard deviation of D across repeated measures. SpO_2 bias and imprecision are not necessarily constant but could vary depending on, for example, the level of SaO_2 or the levels of other covariables. Descriptive plots for studies comparing two methods for measuring the same quantity, Bland and Altman recommended plotting the observed difference between the two measurements, y and x, versus their mean, x plus y over 2, to examine whether the difference exhibits a trend in location or variability across the mean.

In pulse oximeter studies, y and x are the SpO_2 and SaO_2 values, respectively. The Bland Altman plot shown here is for real-world data. Superimposed on the plot is the mean difference or bias line, the dotted line; the zero-difference line for reference, the solid line; and the lower- and upper-95% limits of agreement, the dashed lines, which is an interval within which mean differences should lie in 95% of applications when the assumptions underlying the construction of the interval hold.

The modified Bland-Altman plot is the difference y minus x versus x, which has been suggested when x is a reference value or has much less measurement error than y. The modified Bland-Altman plot for SpO_2 versus SaO_2 is shown here for the same data as in the last slide. A slight negative trend appears due to regression to the mean because SpO_2 and SaO_2 are not perfectly correlated. The more highly correlated y and x, the more similar the Bland-Altman plot and the modified Bland-Altman plot will appear. The quantile or q plot is a plot of the quantiles

or percentiles of one variable against the corresponding quantiles of another variable. In the left Q plot of SpO_2 versus SaO_2, the points lie close to the dashed line of identity, indicating that SpO_2 has very little bias for most SaO_2 values. In the right Q plot for subjects in the Fitzpatrick 5 and 6 dark skin categories, the points mostly lie above the identity line, indicating positive SpO_2 bias. It appears to increase with decreasing SaO_2.

Accuracy root mean square error, or ARMS has been the primary performance measure for evaluating pulse oximeters in premarket-controlled desaturation studies. The result shows that ARMS is a function of both bias and imprecision of SpO_2. In the FDA discussion paper, the agency proposes that the overall ARMS of pulse oximeters is shown to be less than 3% with statistical significance. In other words, the upper limit of the 95% confidence interval on should be less than 3%.

If the true overall ARMS is less than or equal to 2.5%, then the power is expected to be at least 80% to show ARMS is less than 3% with statistical significance. In a controlled desaturation study of 24 subjects with an average of 20 repeated pairs measurements of SaO_2 and SpO_2 per subject for a total of 480 paired measurements, statistical significance may be demonstrated either with hypothesis testing at the one-sided significance level of 2.5% or with the upper limit of the 95% confidence interval on ARMS being less than 3%. However, if the mean of SpO_2 minus SaO_2 varies substantially between subjects, then the sample size of subjects may need to be larger than 24 to achieve 80% power and non-disparate performance assurance.

In the FDA discussion paper, in addition to the primary objective of demonstrating less than 3% with statistical significance, the agency proposes a new co-primary objective of demonstrating non-disparate performance across skin pigmentation levels. For the individual typology angle, ITA, measurement of constitutive skin pigmentation, non-disparate performance is declared when the estimated maximum difference in SpO_2 bias between ITA values is less

than 1.5% per SaO_2 in the interval 85 to 100% and less than 3.5% for SaO_2 in the interval 70 to 85%. For the Monk Skin tone or MST scale, non-disparate performance is defined similarly.

For sample sizes of 24 subjects and 20 paired repeated measures per subject on average, simulations suggest that for SaO_2 in the interval 70 to 85%, the power is 80%, for the point estimate to be less than 3.5% when the true maximum difference in SpO_2 bias equals 2%, and for SaO_2 in the interval 85 to 100%, the power is 80% for the point estimate to be less than 1.5% when the true maximum difference in SpO_2 bias is 0.5%.

This plot depicts data that are consistent with SpO_2 bias being linear in ITA. When SpO_2 bias is linear in ITA, the maximum difference in SpO_2 bias between any two ITA values within an ITA interval occurs between the lower and upper limits of that interval. For example, for the ITA interval of minus 50 to 50, the maximum difference in SpO_2 bias between any two ITA values occurs between the ITA values of minus 50 and 50.

To evaluate whether SpO_2 bias is non-disparate across an ITA interval, the maximum difference is compared with the acceptance limit. For the essay on two intervals, 70 to 85 and 85 to 100%, the proposed acceptance limits for the maximum difference are 3.5% and 1.5%, respectively, as already mentioned. To evaluate whether the bias is non-disparate across an interval, a linear mixed effects model of the difference D equals to minus was proposed in the FDA discussion paper. Recall that SpO_2 bias is defined as the expected value d. The model assumes SpO_2 bias is linear in SaO_2, ITA, and their interaction. The model assumption should be checked for goodness of fit to the data. In particular, the assumption that SpO_2 bias is linear in ITA should be checked.

The model assumption that SpO_2 bias is linear in ITA may be checked visually by plotting the mean of SpO_2 minus SaO_2 per subject versus their ITA value. These plots were created for a

particular pulse oximeter device evaluated in a controlled desaturation study of 21 subjects with an average of 23 paired repeated measures per subject.

The ITA was measured at the mid-dorsal distal phalanx on the right hand. The left-hand plot is for the interval SaO_2 less than 85%. The right-hand plot is for the interval SaO_2 greater than 85%. Locally estimated scatter plot smoothing, or LOESS, was used to fit a nonlinear curve to the data. The red lines are the fitted curves, which suggest some degree of non-linearity. The two large circles on each plot indicate the ITA values at which the LOESS estimated difference in SpO_2 bias is maximized. In the left-hand plot for SaO_2 less than or equal to 85%, the lowest estimated maximum difference in SpO_2 bias is 1.2%, which is less than the acceptance limit of 3.5%; that is, the acceptance limit of 3.5% is met for this SaO_2 interval. In the right-hand plot, for So_2 greater than 85%, the lowest estimated maximum difference is 1.4%, which is less than 1.5%, the acceptance limit for this interval; that is, the acceptance limit of 1.5% is met for this SO_2 interval.

The distribution and range of ITA varies with the location at which it is measured. On the left, box plots of the distribution of ITA for nine locations are shown for ICU patients. On the right, the twenty-fifth and seventy-fifth percentiles of ITA are shown per location. This interquartile range varies substantially by location. In this table and plot, the range of ITA is given for nine different locations in a controlled desaturation study of thirty-four healthy subjects. Again, the ITA range varies substantially by location.

Another potential study objective is to evaluate the diagnostic accuracy of SpO_2 to detect the presence or absence of a target condition, for example, SaO_2 less than 90%. For the condition, SaO_2 less than 90%, a test positive result may be defined as SpO_2 is less than some value, for example, 90%, 92%, and 94%. The true positive fraction, or sensitivity, is the probability that a subject with the condition tests positive for the condition. The false positive

fraction, or FPF or one minus specificity, is the probability that a subject without the condition tests positive for the condition.

The receiver operating characteristic, or ROC curve, is the plot of TPF versus FPF for every possible cutoff in a measurement that could be used to define a test-positive result for the target condition. Here, SpO_2 is the measurement, and SaO_2 less than 90% is the target condition. AUC is the area of the ROC curve.

ROC curves overall and for each of the six ITA groups are shown for a controlled desaturation study of three pulse oximeter devices in 21 healthy subjects. SaO_2 was restricted to the interval of 85 to 95% in an attempt to mimic the SaO_2 distribution in real-world hospitalized patients. ITA was measured at the mid-dorsal distal phalanx on the left hand. The legend shows the area on the ROC curve or a UC by ITA group and is the number of subjects per group. N is the total number of measurement pairs of SaO_2 and SpO_2 per group. Because N is small for each group, the per-group OCS and AUCs are imprecise.

Another potentially useful analysis is to characterize the uncertainty of the SaO_2 value given an SpO_2 measurement. Inverse prediction may be used to obtain an interval of plausible SaO_2 values given the SpO_2 value. Inverse prediction is based on fitting a regression of SpO_2 on SaO_2. Based on the fitted regression, a 95% pointwise inverse prediction band for SaO_2 may be constructed. It is shown here in blue. For example, in a horizontal dashed line plotted for SpO_2 equals 92%, the intersection of the horizontal dashed line with the blue prediction band indicates that the 95% prediction interval plausible SaO_2 values are 87 to 96% approximately. This plot could be useful to pulse oximeter users for characterizing the uncertainty of the unknown SaO_2 value given the SpO_2 value.

In summary, for premarket-controlled desaturation studies conducted to support FDA clearance of pulse oximeters, FDA proposes sample sizes of, N equals 24 subjects, and M equals

480 total number of simultaneous measurements of SaO_2 and SpO_2. FDA proposes a co-primary analysis of ARMS and disparity in SpO_2 bias across ITA and MST levels. For the ITA analysis, a question of ITA interval should be used for evaluating the maximum difference in SpO_2 bias between ITA values.

Other analyses may be worth considering, for example, the diagnostic accuracy of SpO_2 for classifying a target condition, for example, SaO_2 less than 90%; prediction interval of plausible SaO_2 values given the SpO_2 value; repeatability of the SpO_2 value, for example, the standard deviation or coefficient of variation of among repeated measures taken under the same or different conditions of measurement; and association of race and or ethnicity with SpO_2 bias. Thank you for your attention.

Questions for FDA Presenters

Dr. Cassiere: Thank you. This is Dr. Cassiere. Do the panel members have any brief clarifying questions for the FDA for all the presentations that we just heard? And remember to raise your hand. And I will call on you for questions. Dr. Saville.

Dr. Saville: Yeah, I had a question on slide 17 of Dr. Pennello's presentation. Just questions of clarification here. These points on this graph, are these the means of each individual? So, if you have 20 observations for each individual, are we plotting the mean of each individual? And number two. The criteria that you're using to define the maximum, minimum range of the ITA; is that maximum based on the model? So you have a model-based estimate of the max difference, and that's what the hypothesis test or the criteria that you're testing is doing?

Dr. Cassiere: Not to put pressure on the FDA, is it possible to show that slide for everyone to view?

Dr. Malvina B. Eydelman: This is for Dr. Pennello; why don't you start answering while our IT folks see if they can go back?

Dr. Pennello: Okay, this is Gene Pennello, FDA. Thanks for the questions. That particular plot is plotting the mean difference for each subject against their ITA value. That's correct. And the way we had proposed to evaluate the non-disparate performance assurance in terms of the maximum difference in bias between two ITA values in an interval was based on the linear mixed-effects model. That's correct.

Dr. Saville: Okay. Thank you.

Dr. Cassiere: Dr. Goldman, you had your hand up.

Dr. Julian Goldman: Yes, thank you. Julian Goldman here. I wonder if the FDA could clarify. We covered a lot of ground today, and it's it seems like some of the topics relate to identifying a means to assure non-disparate performance. Some of them seem to provide a means for manufacturers to characterize or disclose more information about a product's performance, and the third thing we touched on was the notion that perhaps a larger body of data could eventually be acquired or gathered for analysis by the FDA so that we could better understand pulse oximetry performance and behavior because there still remain quite a few questions as they were presented by the various FDA presenters in terms of the methods and cutoff points and so forth. So, I'm hearing three different buckets, and in order to best interpret the rest of today's information, it might be helpful to understand those goals.

Dr. Malvina B. Eydelman: Thank you. This is back there. Thank you. Dr. Goldman, for your excellent question. Our goal is to maximize performance and process symmetry for all patients in the United States. And to that end, we're trying to gather information from all sources, including today's panel. We look forward to hearing from all of the panel input, which will then be taken and incorporated into our further steps to maximize the performance.

Dr. Cassiere: Great. Thank you. Miss Brummert.

Rachel Brummert: Rachel Brummert here. I have a question for the FDA. Where did you get the number of 24 for the sample size, and why such a low number?

Dr. Malvina B. Eydelman: Dr. Hendricks, would you want to start?

Dr. Hendrix: Yes. Hi, this is Dr. Hendricks, chief medical officer, OHT1 within OPEC CDRH. The sample size was derived from statistical analysis modeling based on desaturation lab data. At hand, and I'm going to turn the floor over to Dr. Pennello, my colleague, to explain a little bit more about the sample size derivation.

Dr. Pennello: Sure, yeah, this is Gene Pennello again. Per the recommendations of the panel at the first meeting, we wanted to tighten that criterion for ARMS, and our proposal to do that was to use the 3% as the performance goal. In the existing guidance, but we wanted to tighten that up to show that the sponsor of the device needed to show that it was less than 3% with statistical significance. That means the true value would have to be a lot less than 3% to show that with 80% power for the particular sample size. So we did some sample size calculations, and we came up with 24 because 24 will give you 80% power for accurate pulse oximeters, those that have an RMS less than two per 2.5%, say, or less than that.

And we have to consider the limits of the technology. It'd be difficult for ARMS to be any less than the true ARMS. But we also had considered sample size for the non-disparate performance assurance as well.

Dr. Cassiere: Yeah. Thank you. I just want to remind the panel members that this is specific questions for the presentation, and there's a lot to discuss in terms of general discussions, which we're going to save for the panel deliberations. This is specific questions regarding the material that was presented and any clarifying questions that was confusing. With that said, Dr. Taylor

Dr. Taylor: Jim Taylor. This could be given offline. I guess the basic question about the composition of the devices, both the medical and the wellness devices; is there criteria that the

FDA provides, and what composition information is provided by the manufacturers to the FDA? Dr. Lee could address that offline if you need to.

Dr. Cassiere: Thank you. All right. Great. We'll put that table. Dr. Saville.

Dr. Saville: Yeah, just a couple of real quick follow-up questions. For the power calculation, I imagine there were assumptions that went into between- versus within-subject variance. I didn't see anything about that on those slides. So, maybe a question for Dr. Pennello. Does that calculation, that power depend on that between- versus within-subject variance? And if so, is that going to be shown to companies so they know what should be expected? And then the second question is on slide 13, and maybe this is a question for later, but I'd love to hear more justification about how you came up with those criteria for the max difference for non-disparate performance to 1.5% and 3.5%. And why does that differ by the SaO_2 to just try to understand the rationale for that? Thanks.

Dr. Eydelman: So this is Dr. Iman again. Dr. Pennello, if you can answer that in one minute, please do. Otherwise, I will ask that we continue with the agenda to stay on time, and then we can come back with more answers during panel deliberation.

Dr. Pennello: Great. Yes. Gene Pennello, FDA. The sample size of 24 does depend on the between and within-subject variances and the bias. Those values, we didn't purport those, but those are based on real on looking at real data in desaturation studies.

Dr. Cassiere: Dr. Lanzafame. Got a question for us.

Dr. Lanzafame: Yes, the examples used discussed the location of a finger. Are there similar calculations made for other sites, such as the earlobe? And how does that affect the thickness of the site relative to the data collected?

Dr. Cassiere: I'm going to make a comment. That sounds like something right for the panel discussion coming up, Dr., if that's okay. And then Dr. Wilson, quick question.

Dr. William C. Wilson: Yes. Very quickly. In the November 2022 meeting, there was a recommendation for decreasing the ARMS, and as shown in the slide deck by Dr. Hendricks, there was even a proposed accuracy metric of decreasing the ARMS to 2.1. However, in Dr. Pennello's presentation, there was no discussion of decreasing the ARMS. So I'm wondering, can the FDA be more clear? Is there still currently a recommendation to decrease the ARMS from 3% to 2 or 2.1%?

Dr. Malvina B. Eydelman: Dr. Hendricks, please?

Dr. Hendrix: Yes. Hi. This is Kumu Hendricks again, the chief medical officer from OHT1. I'm going to turn the floor to Dr. Pennello to answer this question because it goes into statistical expertise. Thank you.

Dr. Pennello: So, the previous requirement was that the point estimate of ARMS had to be less than 3%. Now, it has to be met with statistical significance. So that means that the true ARMS, which we will never know for a particular device, has to be, but it's going to have to be a lot less than 3% in order to show that it's less than 3% with statistical significance.

And in the slides, the true ARMS would have to be at least 2.5% or maybe as low as 2. 1% to get 80% power to show that it's less than 3% with statistical significance. I hope that makes sense.

Dr. Cassiere: All right. Thank you. So we're going to have a brief 6-minute break. We're going to reconvene at 1050 a.m. And at this time, we can take a break for the allotted six minutes.

Dr. Cassiere: This is Dr. Cassiere. It is now 10.50 am, and I would like to resume this panel meeting. We will proceed with the guest speaker presentations portion of the meeting. Each speaker has been granted 10 minutes each to speak. The first speaker is Dr. Ellis Monk. Dr. Monk, you may begin.

MST Scale Pigmentation

Dr. Ellis Monk: My name is Dr. Ellis Monk, and I'm a professor of sociology at Harvard University. Before I begin, I'd first like to state that I have no financial disclosure or conflicts of interest with the presented material in this presentation.

The title of my presentation is "Skin Tone and Medical Devices: Why Measurement Matters." Much of my research focuses on colorism, a system of discrimination based on the lightness or darkness of someone's skin. It's important to note that census categories of race, ethnicity, and skin tone are not the same characteristics.

In fact, there is considerable heterogeneity in skin tone within and across census categories of race and ethnicity. Evidence shows globally that skin tone is significantly associated with inequalities in education, earnings, employment, health, and much more. And research shows, including my own, that discrimination is one factor that helps explain these inequalities.

While much attention is often given to race, ethnicity, and racism, relatively little attention tends to be paid to skin tone and colorism, even though it has massive effects on inequality all around the world. For example, one of my studies showed that, among African Americans, there's actually more educational inequality along the skin tone continuum ranging from light to dark, using nationally representative data, than there is between blacks and whites as a whole. But the manifestation of colorism I'd like to focus on today is in the realm of technology. What I often refer to as colorblind technology, forms of technology that don't work equally well, or at all, for anyone who isn't light or skin.

On the slide, I've put a picture of a Shirley card, which is something that Kodak used to calibrate skin tones, lighting, and shadows in their photo printing system. The thinking was that Shirley was the standard. So, if it looked good for Shirley, it looked good for everybody. It

probably should come as no surprise to us today that this practice was eventually tweaked by Kodak after receiving decades of criticism for excluding a properly diverse range of skin tones to calibrate their photo printing system.

This form of color blindness and technology seems to persist to this very day. For example, researchers have found that autonomous or self-driving cars have a lot of trouble recognizing darker skin pedestrians. The thinking is that, as the researchers have argued, that these cars were not trained in environments that had adequate representation of darker-skinned people, and as a consequence, these cars have a lot of trouble recognizing darker-skinned people as human beings who obviously should be avoided in the road.

This is, of course, one reason why we have pulse oximetry regulations that require that two subjects, or 15% of the study pool, are required to be darkly pigmented. The thinking is that by including darker skin patients in these validation studies, we can ensure that pulse oximeters work equally well for everyone, regardless of their skin tone. But of course, this regulation raises a few questions. First of all, is the number of subjects correct? But also the much more profound question of who is darkly pigmented in the first place.

The last few decades, it's been the Fitzpatrick scale that served as the standard for skin tone classification. Especially when it comes to validating pulse oximeters. The Fitzpatrick scale was designed in 1975 by Dr. Thomas Fitzpatrick of Harvard Medical School. It was actually intended to categorize how Caucasian skin, in particular, reacts to UV during phototherapy for treating various skin conditions. That is to say, it was not actually intended as this gold standard of skin tone classification, and in fact, initially, it only had four tones represented, and ten years after its initial launch, two more skin tones were added to the Fitzpatrick scale. This is a picture of the Fitzpatrick scale.

One of the criticisms of the scale is that it excludes the majority of blacks and yields data that tends to overestimate the percentage of the black population that would be rated a four on the skin tone scale. It's also worth noting that some studies find that the Fitzpatrick scale performs poorly, even when used as intended, especially for ethno-racial minorities. Ironically, given a selection of skin tones which live in a very restricted intermediate zone, the Fitzpatrick scale may simultaneously be too dark for many lighter-skinned people while not being dark enough for many darker-skinned people. All of which means, for the purposes of recruiting people into our studies, we may exclude people who are lighter skin and darker skin by using the Fitzpatrick scale.

These limitations, in addition to many more, led the Department of Homeland Security some years ago to begin searching for an alternative to the Fitzpatrick scale for the purposes of classifying skin tone. And as you can see from some of the slides from their presentation, they came to very similar conclusions about the Fitzpatrick scale. One thing I want to highlight in particular that they noted was the conflation of any six-point skin tone classification scheme with the Fitzpatrick scale, and one of the reasons this occurs is because the Fitzpatrick scale was never really standardized, which means that any version of the Fitzpatrick scale that you download online may not match with another version that someone else downloaded online.

That leads me to introduce the Monk Skin Tone Scale, which I explicitly designed to measure skin tone in diverse populations. It's intended to be an easy-to-use, reliable, and cost-effective means of measuring skin tone. It's open source, which means that anyone can use it free of charge, and the main way that the Monk Skin Tone Scale is designed to mitigate biases relative to prior visual scales, such as the Fitzpatrick scale, is by including a wider range of carefully selected skin tones to better represent the dynamic range of skin tones we see in the United States and beyond.

The color selection for the scale was based on extensive field work I conducted in the US and Brazil, computer software that creates facial stimuli for social psychological experiments using skin reflectance spectrum scores, consulting maps of the distribution of UV exposure in human skin tone around the world, and it was initially validated through cognitive interviewing I did along with the National Social Life Health and Aging Project, which is a gold standard epidemiological health and aging focused study, in addition to nationally representative surveys I'll refer to later on, and it was adopted for data collection in 2021 in Wave 4 of NSHAP. Here's a picture of the Monk Skin Tone Scale. By 2022, Google announced to the world that they had adopted the scale to improve skin tone representation across a wide suite of their products. Google has adopted the Monk Skin Tone Scale to improve a whole wide array of different products, including their search engine. They've highlighted that they've used the Monk Skin Tone Scale a lot to improve the skin tone representation in Google Photos and to test their Real Tone filters on their Pixel camera phone really to fix issues similar to what Kodak was experiencing with skin tone calibration in their photo printing systems some time ago.

The foundation for launching the Monk Skin Tone Scale at Google was some of the US scale validation research that we've conducted, as well as some of the global research that we're continuing to conduct, to ensure that the Monk Skin Tone Scale is optimally representative. These studies have already been published, and I'll just quickly point to some of the findings here. One of these studies, for instance, finds that the Monk Skin Tone Scale is not only perceived to be much more inclusive than the Fitzpatrick Scale but it's also perceived to be as inclusive as a 40-point scale from the cosmetics industry, which is quite a remarkable finding given that scale is four times larger than the Monk Skin Tone Scale.

At US-based research, we've also conducted global research on annotation, and this is a paper that was accepted for presentation at NeurIPS, Neural Information Processing Systems

Conference, in 2023. This paper basically shows high levels of consensus using annotator pools from all around the world using the Monk Skin Tone Scale.

Quickly, I'd also like to point out that Meta.ai or Facebook has also adopted the Monk Skin Tone Scale. In this case, they adopted it in Casual Conversations V2, which is a large data set they use to measure algorithmic bias in machine learning and artificial intelligence. To conclude, I'd like to highlight that findings from our ongoing research show that the Monk Skin Tone Scale is as easy to use as the Fitzpatrick Scale, that is in addition to being significantly more representative and inclusive than the Fitzpatrick Scale and the global research that I pointed to before shows that there's a high level of consensus globally using expert and crowdsourced, or non-expert, annotators.

Lastly, I'd like to say that it's important to collect both subjective and so-called objective measures of skin tone. There's still a lot to learn about the potential role of skin tone in pulse oximetry, and using different measures, both subjective and so-called objective measures may help tap into different mechanisms through which skin tone may produce these inaccuracies that some researchers are finding with respect to pulse oximetry. Using the subjective measures, in particular, will not only help us communicate who is included in these validation studies but also help us continue to consider the social determinants of health.

Real-World Evidence and Pulse Oximetry

Dr. Philip Bickler: Good morning. I'm Dr. Philip Bickler, and I'm here with Dr. Carolyn Hendrickson to report on progress on the EquiOx study, which is a prospective clinical study of pulse oximeter errors in hospitalized patients. Today, we have an update after enrolling about 480 patients. Carolyn and I are from the departments of anesthesia and perioperative care and department of medicine division of critical care medicine at the University of California at San Francisco. We've been supported by CERSI and the US Food and Drug Administration.

Dr. Carolyn Hendrickson: The primary aim of the EquiOx study is to measure the bias and pulse oximeter SpO_2 measurements across a range of skin pigmentations among critically ill hypoxemic patients.

Dr. Bickler: The EquiOx secondary aims are to compare subjective skin pigment scales to objective spectrophotometer data to inform pulse oximeter performance studies, to relate skin pigmentation and race among patients hospitalized in San Francisco, to test if previously described differences in bias between race categories is explained by skin pigment differences, to determine if pulse oximeter performance in clinical use is similar to performance measured in controlled laboratory studies, and finally, to test low perfusion as a mediator that explains differences in pulse oximeter performance.

Why are we doing the EquiOx study? It began in 2020 with reports of occult hypoxemia, a phenomenon that involved pulse oximeter readings that were greater than 92% but SaO_2 values that were less than 90% in black patients. There were a number of questions about the validity of these conclusions because of problems relating to imprecise pairing of saturation measurements and blood gas readings and hemoximeters perhaps set to fractional and not functional saturation, issues with self-reported race, lack of skin pigment data, presence of interfering pigments, anemia, low perfusion, and motion not to mention other problems.

In the hypoxia research lab at UCSF, our data on pulse oximeter performance recapitulated with Sjoden and others had seen that there is missed hypoxemia in darkly pigmented patients. Our work found that misdiagnosis of hypoxemia was more probable under conditions of low perfusion, that is, a perfusion index less than one.

The EquiOx study then was sought to relate the studies that we had done in the hypoxia lab with a real-world clinical study. So our study strategies for the EquiOx studies and advantages over retrospective studies included synchronous paired samples, measurement of

functional saturation with a hemoximeter, quantification of skin pigment with Monk, von Luschan, Fitzpatrick, and quantitation with the skin colorimeter. We have quantified perfusion; measured optical signals; and involved an inclusive network of collaborators, stakeholders, and statisticians; and we have enrolled a population with broad skin pigment range.

Our subjective measurement scales are treated as categorical variables, and they include the von Luschan scale, Fitzpatrick, and Monk scales depicted here. These parameters are measured at the forehead, ear, inner arm, and fingers. Our objective measurement assessing skin pigment is the Konica Minolta spectrophotometer. It produces a value called the ITA, which is a representation of color based on brightness and hues in the yellow and blue and red and green spectrum.

Dr. Carolyn Hendrickson: Our patient enrollment to date is shown here with a breakdown by race categories. These are self-identified races abstracted from the electronic medical record. We have a broad representation of race identities, including a large proportion of patients who self-identify as other race.

We are showing you here the distribution of patients who have been enrolled using the Monk Skin Tone Scale at the dorsal finger. You can see that our population has weighted heavily towards the middle, and we have lower enrollment in both extremes of the skin pigment categories. And notably, we've had lower enrollment in the darkest skin pigment categories labeled I and J here.

We are showing here the distribution of Monk Skin tone categories across the x-axis, with A being the lightest skin tone and I being the darkest, with a proportion of race breakdown among each of those categories, showing that the social construct of race actually has broadly overlapping skin pigmentation as assessed by the Monk Skin tone scale. A notable exception to

that is in our lower represented enrollment categories of H and I, which are predominantly showing black or African American self-identified race in those darkest pigments.

Here, we show the perceived Fitzpatrick skin tone group across the x-axis with ITA, the objective measure of lightness to darkness scale obtained from the spectrophotometer on the y-axis. What we're showing here is measurements at the cheek and the forehead for our enrolled subjects. And I'd like to call your attention to the fact that in each Fitzpatrick skin tone group, there is broad overlap in the objective measurement of ITA. We're trying to make the point here that the perceived Fitzpatrick skin tone group is a very imprecise way to categorize dark skin pigment that's assessed through an objective measurement.

Another important feature of our data is that in this real-world clinical study saturations of less than 90% are rare. We are studying ICU patients who have normal staffing ratios with nurses, respiratory therapists, and physicians who are not under the stress of the pandemic, and they rescue their patients or attend to them before desaturation events happen. So, it's quite uncommon for us to observe SpO_2 in the lower ranges below 90%.

Dr. Bickler made the point that we think the perfusion index is a really important variable to study. And in the earlier data from the controlled laboratory subjects, we saw that when a cutoff value of a perfusion index of one or less is used that missed hypoxemia is more common. So, in this distribution, we see that the median perfusion index is three for the observations we've made to date, but about a quarter of our observations are in the perfusion index, less than one range, which was important in the lab data that was shown in the introductory slides.

Here, we're showing that pulse oximeter probe location is variable among our ICU patients. Most of our observations are from oximeters on fingers, but you can see that occasionally, we observe oximeters in an unusual location like an ear and that earlobe oximeters are actually fairly common. We know that our clinicians put probes on fingers first, and if they

can't get a reliable reading and there's a poor tracing, they'll move the oximeter to a different location.

In summary, the EquiOx study is a real-world prospective study of pulse oximeter accuracy with detailed data collection. We're collecting a wealth of information, including self-identified race, on subjective and objective skin pigment measurements at a variety of locations using several different skills. We're collecting data on the perfusion index and the stability of the SpO_2 tracing at the time of the ABG draw. We're also collecting comorbidities and things like vasopressor administration at the time of the sample. Most of our blood samples are in the SpO_2 range of greater than 90%, and the perfusion index is low in many of our observations.

The probe location varies in the real-world setting, and this may impact some of the interpretation of our findings. So, with that, we will close with some gratitude to all of the people who are involved in this work and all of our supporters. This is our team shown here and our funding supporters down below. Thank you for your time and attention.

Dr. Almond: Hi, my name is Chris Almond from Stanford University, and it's a pleasure to present a prospective clinical study to evaluate the accuracy of pulse oximeters in pediatric patients with increased skin pigmentation. We have no financial disclosures. This project is supported by a grant from the Food and Drug Administration to the UCSF Stanford CERSI.

This is a quick look at our study team. This project originally began as a conversation between Desireé Conrad, myself, and Michelle Tarver at the FDA and has really been a fascinating project since that time. What I'd like to cover is to review the study design and rationale and to provide an update on the status of the current pediatric study to review the baseline characteristics of the study cohort to explore some preliminary correlations between pigment scales and colorimetry and then to review some of the challenges and lessons learned from conducting a prospective pediatric study and pulse oximetry.

Like adults, pulse oximetry is widely used to determine whether a child or infant is adequately oxygenated, and studies suggest that pulse oximetry systematically overestimates the true oxygen saturation in children with darker skin pigment. This error or bias puts children with darker skin pigment at considerable risk by failing to detect important levels of hypoxemia that drive critical treatment decisions like medication usage, hospital admission, timing of surgery, ICU transfer, intubation, and ECMO.

Prior studies have a number of limitations that have already been reviewed, so we won't repeat them here. The purpose of this prospective real-world study is to address the limitations of retrospective studies. This issue is known almost 20 years ago after Bickler had reported the initial study, but really came to the public's attention in the COVID era after several research letters were published prominently. This one happens to look at the pediatric disparity, and importantly, this era may be magnified and low perfusion states. You can see here in this Bland-Altman that at low perfusion, there seems to be somewhat of a hockey stick upward where there appears to be bias at low perfusion.

This is a prospective originally single, now multicenter study that involves not-anemic children under 21 years of age who happen to have an arterial line. The setting is children's hospitals, mostly in the cardiac cath lab and cardiac ICU. Skin pigment was measured by the von Luschan scale, the Monk Skin test, and the Fitzpatrick scale and colorimetry was measured using the Konica and Delfin colorimeters. The primary outcome was the difference in the SpO_2 and the SaO_2. Secondary variables include perfusion index, age, saturation, and self-reported ethnicity. The total sample size originally was 154 subjects.

The Stanford IRB and FDA approved the study protocol, which involved written informed consent from research participants, and a DMC was planned to look at the data to reestimate a sample size based on the incoming data. Where are we? We've come a long way.

Originally, we were invited to submit the proposal in June of 2022, and with UCSF's help, we were able to submit the proposal three weeks later. We received the notice of grant award around seven weeks later. Somewhat surprisingly, we were able to get Stanford IRB approval within about a week. This reflected that the Stanford IRB recognized that this was a priority study to review and approve. We hired a research coordinator about five weeks later and ultimately received FDA IRB approval. We completed our contracts and enrolled our 1st patient by the end of January of 2023.

It was pretty clear early on that there was a skew in the enrollments toward the lighter skin-pigmented patients. So, we implemented a variety of measures. But most importantly, we added Michelle Williams as a coinvestigator, who has expertise in minority recruitment. This seemed to improve things over time, and we ultimately completed enrollment of the original 154 subjects within the year. And this is essentially what our enrollment looked like during that first year; we actually didn't have that many consent declines or consent failures, around 9%, and there didn't seem to be any significant difference across racial groups. But we didn't have enough von Luschan categories of patients to really answer the question. So, the study was expanded to 312 subjects again with a 20 to 40 von Luschan four category of patients as the primary target. We essentially added to institutions the children's health care of Atlanta hospital at Emory University and Oakland Children's Hospitals, who have very diverse populations.

Somehow both hospitals were able to get up and running in absolutely record time. Emory was up and running in about seven weeks and enrolled their first patient and Oakland not long thereafter. We're up to 228 patients, and actually we now have 14 (one-four) patients. We are also pleased as we just received an award of $10,000 to help subsidize the study in the second year. We hope to have the trial completed around July and plan to present at least some of the information from the trial at the National Black Nurses Associated Conference that will be

held in San Francisco here in July. Just to review some of the baseline characteristics of the study cohort in the pediatric study, we have a median age of around five years with a pretty good distribution across age groups: 50% are female, 34% are Hispanic or Latino ethnicity, 34% are white, and the remainder are nonwhite with African Americans representing 17% of the current population, which is an improvement from when we first started.

A notable advantage of the pediatric study is because of the prevalence of congenital heart disease. We have a fair number of subjects who have saturations at rest that are below 90%, and you can see here, about one in three fall into that category. If we look at enrollment overall by von Luschan categories, we see that there is a leftward tilt to the enrollment. If we limit it to those who self-report is African American, you can see that it's shifted toward the right, but this is a good illustration of how self-report of race is not a very good predictor of skin pigment. I think another implication here is that for this one von Luschan 4 category, only about one in three African Americans by self-report qualify for von Luschan four, which has important implications for study planning.

We saw something very different, very similar with the Monk Skin tone test with a leftward skew in the overall population, a rightward skew amongst African Americans, and again, less than 20% fall into the IHJ category, at least among children. And again, we see that self-report of race is not a very good predictor of skin pigment.

We looked initially at the correlation between pigment scales and color imagery with ITA, and overall, we see perhaps the best correlation between the Monk Skin Test and von Luschan with 36 of its categories and less with von Luschan 4 and Fitzpatrick 6. If we look at the correlation between various pigment interests and, we see that the overall correlation is relatively good, perhaps slightly better with the Delfin, although these data are quite preliminary.

Just to look at some of the challenges and lessons learned from conducting a pediatric pulse oximetry study, by far and away, the single most important one is that strategies to enhance minority recruitment can be highly effective with a multimodal strategy. Initially, we prioritized von Luschan 4 to approach and enroll, and where possible, we favored minority concordance between the investigators and the patient's family. We developed a weekly dashboard to track minority recruitment and to study our consent failures. We recruited Michelle Williams to the team, who has been just a wonderful resource, and she helped us to develop a brochure that targeted a diverse audience, helped us with our scripts, and approaching patients, highlighting the importance to identify racial bias. We expanded recruitment to include electronic consent. This is important because many families can't actually be at the bedside in the hospital with their children, either because of work or having other children. And lastly, we expanded the study to other regions, including Atlanta and Oakland.

In addition to this first lesson learnt, we also found that because there's no widely accepted pigment scale available. We essentially ended up collecting data with all three, along with ITA, for comparisons later. Because the colorimeter can be somewhat intimidating for younger children, we involved Child Life to help us to approach those patients. To reduce the impact of pre-analytical factors or noise in the system on co-symmetry measurements, we used a validated portable blood gas analyzer that was provided as a loan by Masimo. For this, we found the cardiac OR wasn't a very good place to enroll patients because they're too dynamic and don't have a pulse while on bypass so we shifted enrollment to favor more the cath lab and the cardiac ICU. And lastly, as we mentioned, because parents may not be able to be at the bedside, we obtained remote consent from those where it was needed.

So, in conclusion, the SPOT-Bias study is designed to determine whether racial bias exists in contemporary FDA-approved oximeters and children across a wide range of ages. The

study has enrolled over 225 children across three sites and is projected to complete enrollment in the coming months when we'll be able to reveal the symmetry data. The correlation between the Monk Skin Tone test and other pigment scales appears reasonably strong, and initial studies suggest that there's moderate correlation with ITA.

And lastly, while we've encountered a variety of challenges in conducting the study, most have been addressable and should allow us to answer the study question while also helping to inform the design and conduct of future pediatric oximetry studies. I just want to thank everyone at Stanford, especially Desireé and Rohan at UCSF, the team there at Harvard Medical School that serves as our DMC, FDA Atlanta, and the overall CERSI program. Thank you again for your time.

Patient Perspective – Adult

Mr. McClure: Hello, and thank you for your time today. I appreciate you letting me share my experiences with my pulse oximeter. In 2013, I was diagnosed with emphysema, the more severe form of COPD. In 2019, I finally stopped smoking after a total of 40 years. In an effort to find relief for this COPD, I had some lung valves implanted into the lower right lobe of my right lung in October 2020. Then, in January of 2021, one of those valves had to be removed and replaced because it was not functioning properly. It was at that time in January of 2021 that my pulmonologist prescribed 24/7 oxygen therapy for me.

Once that was prescribed, then came the need for the pulse oximeter to be able to monitor my O_2 level. The first oximeter that I received was a gift from someone who worked in the industry. But the challenge with that one was that it was very slow to record information and that meant that by the time I got a reading, it was I had already started to improve or panic even more.

Then, the letters were very small and hard to read, and it also was upside down to me. It was designed where it would be right side up to the person standing in front of me looking down at it. So it was a real challenging one to use effectively. I kept it for almost a year, though, and then I ordered one off of Amazon; this one by Metene, and this one has worked a lot better.

It was priced at about $50, about halfway in range of what I saw offered on the day that I was looking for them. It has larger letters. It's a bright white readout, and the letters are right side up to me when I look at it. So it's much easier to read, but it's still challenging in that sometimes, when I get a reading, the pulse rate might say 27, and the O_2 rating might say 90. Then I know that it's not correct because I know my pulse is not 27. So then I'll do it again, and it might be closer to what I feel is right. Sometimes, I have to do it up to three times where I'm convinced that this is the right reading. Which ties into the concern, which is what I think this meeting is all about is the fact that these pulse oximeters don't always read accurately for people who have melanated skin or heavily melanated skin like myself, and maybe that's the reason why these readings either take so long, or it takes multiple readings.

And in the process, some things can go wrong, and trying to wait to get the right reading because these panic attacks are real. And I've learned how to control them with personal breathing and my mental state, but some of that sometimes can be based on the information that I'm getting from my pulse oximeter.

I learned about that problem from my daughter. My daughter had read about a study that said that people with melanated skin were not often read accurately. As I mentioned that to my many medical professionals, I have a lot of different health issues; none of them, except for my African American primary care physician, had even heard of the problem.

So that is a big problem in terms of the risks that people are being placed in, so I'm happy to know that this workshop is being done today, and I'm hoping that the technology will quickly catch up with the need before anything tragic happens to any of the patients out there.

Thank you very much for your time. I really appreciate it. I will be around to answer any questions that you might have at the end of the program today.

Patient Perspective – Pediatric

Ryan Jolly: My name is Ryan Jolly, and I am the mom of two children that use pulse oximeters, and we live in Lenexa, Kansas, a suburb of Kansas City. I'm going to talk to you about one of my children that is African American, and we've been using pulse oximeter with her in our home for ten years. She came to me at a little bit over two years old, and she's been using that the entire time she's been in my home. We use it when she's asleep. She has a tracheostomy. She has esophageal laryngeal insufficiency. She has dysphagia and a paralyzed left vocal cord, all attributable to a chromosomal abnormality that affects one in 94 million people.

In my case, we didn't have treatment options before we began using pulse oximetry in the home. Her respiratory issues started at birth. She had a traumatic birth and immediate intubation. And so, by the time she came to me at two, she was already established as a pulse oximetry user. We get our Pulse EX-Q equipment through a DME company, and we don't have a ton of choices in what equipment we get, which is whatever manufacturer the DME company is contracted with.

We've used two of the most prominent manufacturers over the course of time my kids have been in my home, and each of them has their pluses and their minuses. For my kiddo, who's African American, her skin tone is much darker than my other child. He is much paler, and his is much more accurate, much more consistent.

What we see with my daughter, when I talk about consistency with a darker skin tone, is that the machines tend to lose her reading for no reason that we can detect. She might be wearing it for hours at a time, and we're getting good readings, and everything is fine. And then no one moves, nothing happens, and suddenly we're getting an alarm message. When it comes to using an over-the-counter device, I'm a nurse by training, so I was familiar with how they work in theory as a medical equipment device, and I find them to be fairly straightforward, but I can see where those without my medical background might find it difficult to troubleshoot and maneuver.

In that, all of the devices we have used use an icon system and not words. And so, an icon system is great, but if you don't know what the icons mean, then you're lost. In the middle of the night, when it's alarming, and there's an icon of some kind, if you don't know instinctively what that icon means, you have to pull out a manual and try and figure it out. And that's not exactly user-friendly for families in a crisis because their child's PulseOx is going up. I feel that I was trained well in the use of the pulse oximeter. We have a great DME company that we work with and they have been really helpful as we've changed machines or as we've troubles had to troubleshoot various issues. But it is certainly a learning curve; even though I do have medical training, it's new for me every time a different alarm goes off, or it's been months since we had training of any kind, and then suddenly we're having difficulty. It's almost like you have to relearn what this particular machine means.

When it comes to the benefits and opportunities or risks for the pulse oximeter in the home, I don't see downsides for us. It enables us to keep my daughter at home. As I said, she is trach dependent, and she requires eyes on care 24/7 because she has an intellectual disability as well and doesn't recognize her trach as a lifesaving device. So, if she were not on a pulse oximeter at night, we would need to staff a night shift that watched her while she slept. But because we have the use of a pulse oximeter in the home, she can sleep in her own bed in my

room. And I can get sleep as well, which is critical to being able to care for her and stay alive. Everybody needs sleep.

I became aware of the issues associated with accuracy based on skin pigmentation just through trial and error. Initially, when she would have an error that we couldn't resolve as I described before, where everything would be fine and then suddenly her pulse ox would drop for reasons we couldn't identify, her oxygenation would plummet; I always thought it was the probe. And so, I would replace the probe that became problematic because we're only allowed a few under our insurance plan per month. And it took a little bit of a learning curve to realize that maybe it wasn't the probe. Maybe the probe just couldn't see at the moment And so we've had to adjust the way we determine that. So that's how we manage troubleshooting, at least in our home.

When it comes to establishing appropriate parameters for the PulseOx itself, we have a great pulmonologist that we work with that establishes those for us but we've definitely seen that change as well as my daughter has grown. As I said, she came to me at two, and now she's 12, and her window for readings has certainly expanded. Previously, we got an alarm when her heart rate dropped below 75. Now, as a 12-year-old, when she's in deep sleep, it might fall to 50. So we've had to work with our pulmonologist to really set a parameter based on her needs, establishing a baseline for her, and then responding accordingly in both the settings and when we have alarms.

I would say that the Pulse Ox allows my daughter to have a quality of life that would otherwise be denied her. It also allows our family to have a quality of life that would otherwise be denied us. Between my two children, we have a nurse in our home 16 hours a day. So that eight-hour window where it can be just my daughter and I and not a trusted nurse but a stranger in our home, allows us to have a little bit of normalcy that otherwise would be denied.

Patient Perspectives – Industry

Tara Federici: Hello, my name is Tara Federici, and I'm the Vice President of Technology and Regulatory Affairs for AdvaMed.

AdvaMed is the world's largest trade association representing medical device and diagnostic manufacturers. Our companies produce the innovations that transform health care through earlier disease detection, less invasive procedures, and more effective treatments. We have more than 400 member companies, ranging from the largest to the smallest innovators.

AdvaMed advocates for a legal, regulatory, and economic environment that advances global healthcare by assuring worldwide patient access to the benefits of medical technology. We promote policies that foster the highest ethical standards, rapid product approvals, and appropriate reimbursement. I want to take this opportunity to thank FDA for the opportunity to share with this committee the many actions and steps AdvaMed, and our members are taking to improve diversity in Medical Device Studies.

The focus of today's meeting is to discuss an approach to improve the quality of premarket studies and associated methods used to evaluate the performance of pulse oximeters submitted for FDA premarket review while taking into consideration skin pigmentation, race, and ethnicity.

AdvaMed strongly supports efforts to diversify medical device clinical trials, and we have a number of efforts underway to tackle this challenging issue. These include AdvaMed's health equity initiative to promote inclusion and equity in healthcare, where we partner in education with stakeholders, promote research equity in the medical device industry, and work to facilitate access by diverse patients to innovative medical devices. We also have AdvaMed's Take Her Health to Heart initiative to increase enrollment and retention of women in cardiovascular clinical trials. This important initiative has been underway since 2015.

In addition, AdvaMed and our many industry partners are also active participants in various external efforts to improve diversity in clinical studies, including the MedTech Color Collaborative Community and clinical society initiatives. As part of AdvaMed's efforts, we've developed recommendations for industry to improve recruitment and retention of women and diverse participants in clinical trials, including creating a checklist of actions for company sponsors to improve recruitment and retention of women in cardiovascular device trials, developing recommendations and key takeaways from a series of workshops that we conducted in collaboration with Meharry Medical College in 2021. And we've also issued a report outlining approaches to increasing diversity in clinical research and addressing health inequities. I encourage interested individuals to visit AdvaMed's website to learn more about these programs.

In addition to these ongoing programs, many of our companies have identified diversity in clinical trials as a high priority and are investing significant resources in this effort. Based on feedback and input from our companies, these programs include conducting patient engagement activities with diverse patients prior to developing trial designs to both inform and facilitate optimal trial designs to recruit diverse participants, reviewing the study design and inclusion and exclusion criteria for protocols and related follow-up criteria to ensure that they are not inadvertently excluding diverse participants. They are building diverse research networks and creating partnerships with new investigators and health care organizations that serve diverse patients.

They are encouraging clinical trial sites to develop specific recruitment programs for women and minorities, and they include language in the research agreement to set expectations for investigators with respect to enrolling women and minorities. And they are actively seeking women and diverse principal investigators, co investigators, and site research staff because we know participants respond more favorably to individuals who look like them.

They are expanding screening logs to capture data around why participants choose not to enroll in trials, along with their perceptions around research, to help inform existing and future studies. Importantly, they are identifying alternative follow-up requirements for trials that will encourage the participation of women in diverse patients by minimizing the number of follow-up visits. These include establishing phone follow-up or home visits by the nurse coordinator, allowing telehealth follow-up visits, and allowing the participant's primary care doctor to perform some of the follow-up requirements. We know that follow-up is a key challenge in device trials.

Our companies are also actively seeking opportunities to share information on medical device trials at public and clinical events. They are working to understand the limitations of current data sources, such as prevalence data, which may reflect the health inequities of the current healthcare system. And lastly, they are acknowledging the history of abuse in clinical research and communicating to participants the policies and clinical trial safeguards that are intended to protect them.

In closing, given the reliance on pulse oximeters during the COVID-19 pandemic and the important role pulse oximeters play in our healthcare system, AdvaMed supports efforts to ensure that pulse oximeter studies include diverse patients based on skin tone, race, and ethnicity. I'm happy to respond to any questions. Thank you.

Dr. Stephen J. Barker: Hello, this is Stephen J. Barker. I am a professor emeritus of anesthesiology at the University of Arizona and also the chief science officer at Masimo. I'd like to talk to you today to give an industry perspective on the FDA's discussion paper on the effects of skin pigmentation upon pulse oximetry, a very important topic.

The FDA and industry are well aligned in the importance of equity in medical device performance. We applaud the FDA for addressing the importance of pulse oximetry in this issue. The FDA's discussion paper on pulse oximeters and skin pigmentation shows careful thought

and extensive research. The discussion paper takes a significant step in accounting for race and ethnicity accuracy issues in pulse oximetry.

The FDA has requested feedback on seven separate questions. Four of these questions are on the use of the Monk Skin tone, which I will call the MST scale, and the individual typology angle or ITA assessment to address different skin pigments within the US Population. One question requests feedback on the performance criteria among varying skin pigments. Two questions ask for additional measures the agency can take to address the diversity of skin pigmentation, and I will try to address all of those.

The use of the Monk Skin Tone MST scale for the initial assessment of skin pigmentation, followed by an objective measure using the Individual Typology Angle, ITA, is satisfactory. The MST, the Monk Skin Tone, provides good resolution and range to skin tones from very light to very dark. The ITA provides an objective measure to detect subjective bias in the assessment of skin tone, so our recommendations include that the FDA could provide more specific guidance regarding the location for obtaining ITA measurements, identify preferential assessment sites as being the forehead or the back of the hand. The nail beds have very large variability and measurement at the finger sensor site may be more challenging than the forehead or the hand. Require at least three ITA samples at the designated target area to decrease error from a single measurement.

Here is a comparison of these different skin tone measurement methods. On the y-axis, you see the ITA or the individual typology angle, and on the x-axis, you see the three subjective assessments: the Massey scale is the red bars at the bottom, the Monk Skin tone on the forehead is the blue, and the Monk Skin Tone on the back of the hand is the yellow. They're all fairly similar across the pigmentation range. These are the pigment distributions of a study that we recently did and that I coauthored a paper that is now published in the Journal of Clinical

Monitoring where 74 subjects divided themselves into either black or white. And you see the Massey scale on the x-axis and the number of subjects in each group on the y-axis. Self-identified white are the blue bars; self-identified black are the orange bars. You see, the two peaks are at Massey scales of two and three for the whites and six and seven for the blacks. But the other thing to note about this figure is that Massey Scale values of one and nine are very rare. And in fact, ten is nonexistent; we had nobody at a ten. So, the two extremes are rare.

So, our recommendations are to move forward with stratifying skin pigments into three MST cohorts: one to four, five to seven, and eight to ten. These cohorts can detect performance differences and trends with skin pigmentation. We agree that at least 25% of participants should come from each MST cohort, one to four, five to seven, and eight to ten. Balancing these cohorts will help avoid under or over weighting of a particular cohort or skin tone range. Require at least one subject with MST values of two and nine. Inclusions of MST two and nine, that's Monk Skin tone, assure diverse subject enrollments and help avoid recruitment constraints due to lack of available subjects, and remember, one and ten are very rare in the US population.

Our feedback criteria for non-disparate performance, the proposed bias requirement or mean error requirement for non-disparate performance of plus or minus 1.5% per SaO_2 values greater than 85% is acceptable, but the definition of the lower SaO_2 between 70 and 85% SaO_2 of 3.5% plus or minus raises clinical concerns. We think that is too broad. SaO_2 values below 85% present more significant hypoxia risks then values above 85%. Disparate results in this SaO_2 range are also what has been identified as occult hypoxemia in several recent publications.

So, our recommendations on these criteria are to use the Monk Skin Tone Scale for the main assessment of non-disparate performance. The use of two different skin pigment measures will be unnecessarily complicated and confusing. Define non-disparate performance for maximum bias difference in the range SaO_2 70 and 85% as plus or minus 2% bias rather than the

proposed 3%. The plus or minus 2% threshold is similar to the SaO_2 greater than 85 range of 1.5% while accounting for the potentially greater bias that exists at lower SaO_2 values. For proposals for further FDA actions: Tighten the accuracy requirements of laboratory validation studies from ARMS; now remember ARMS is the root mean square error, so that includes the random error as well as the bias or mean error. Tighten it from the 3% ARMS in the document to 2% ARMS. Reducing the error will lessen the clinical impact of disparate performance due to skin pigmentation.

Apply these new requirements to reprocessed sensors for pulse oximeters. Reprocessed sensors are significant modifications of FDA-cleared disposable sensor devices. The impact of the reprocessing should be assessed for its effect on accuracy for disparate performance due to skin pigmentation. Track the 510(k) submission and clearance rates over the next two years to assess adherence to these new requirements. So, to summarize our recommendations, we agree with the proposed use of MST and ITA to quantify the skin pigmentation diversity of the US population. The preferred locations of measurements should be on the forehead and the back of the hand.

Require subjects in the range of MST2 and MST9 and require at least 25% of participants in each of the three cohorts that we defined: one to four, five to seven, and eight to ten. The criterion for non-disparate performance is recommended to be a bias of 1.5% plus or minus for SaO_2 between greater than 85% and plus or minus 2% for SaO_2 between 70 and 85%. Tighten the SpO_2 accuracy requirements, overall accuracy, from 3% ARMS to 2% ARMS, and apply the new requirements to reprocessed sensors. Follow-up after two years to assess the consistent adherence to these requirements and the burden to manufacturers. That's the end of my presentation. I thank you very much and I hope we can continue this discussion. Thank you.

Dr. Cassiere: I'd like to thank our speakers. This is Dr. Cassiere. Do the panel members have any brief questions for our guest speakers? Dr. Feldman.

Dr. Feldman: Yeah, this is a question for Dr. Barker. He talked about measuring at the back of the hand, and I've forgotten the other location now, but neither were at typical locations where the sensor would be applied. And I'm curious if he could address why he's recommending that the location be measured differently from the sensor site and the implications for that.

Dr. Barker: Thanks, Jeff. That's an excellent question. I hope you can hear me. The back of the hand, rather than the fingertip, because the fingertip is highly variable between finger locations, and it doesn't represent the skin absorbance very well on the rest of the hand. There's nothing against measuring the on the fingertip, but to overall assess the skin pigment of the subject of the patient, you need something other than the fingertip. Fingertips tend to be, especially if you're going through the fingernail, tend to be a lighter color. And the forehead is a good location to assess overall pigmentation level of a given subject, and that has been shown in other studies.

I think this discussion of the different assessments of how to measure skin tone is an excellent one. And I'm glad to hear it and see it continue because it's obviously more than just darkness, for example, the Massey scale is primarily about darkness and pigmentation color, which you measure in the ITA, so it's complicated. And this is an excellent discussion. Thanks, Jeff.

Dr. Cassiere: Dr. Lewis.

Dr. Lewis: Hi, this is Tamorah Lewis. My question is for the two patient representatives, Mr. McClure and Ms. Jolly. In addition to moment accuracy of pulse oximetry, I heard from both of you that other performance aspects of the PulseOx are very important, specifically the time from application to getting an accurate read and also the ability of the PulseOx to maintain an accurate read over certain amounts of time.

Can you please speak a little bit more about the importance of these two aspects in addition to just the ability for it to be accurate at a moment in time and how these may be incorporated into premarket testing of devices?

Ryan Jolly: I'll jump in. This is Ryan Jolly. There's a lot of questions in that one question. But the way that really impacts us, because my experience is different than the other representative in that we only use it during sleep for both of my children; the way it impacts us primarily is what I mentioned in my video is sleep and the quality of life that gives us to be able to have sleep and not have additional support in our home to monitor her. I don't know the answer to premarket testing, but I don't have a follow-up for that. I don't know the answer.

Dr. Lewis: I'll clarify. I'm sorry. Ms. Jolly. For example, right now in premarket testing, is it accurate? But I think your really important point was that it needs to stay accurate over periods of time. For example, it may be that it would, as a parent, it would be useful for you to know that it stays accurate for 30 minutes, 60 minutes, or something like that.

Ryan Jolly: It would, absolutely. If we had some guidelines. I think I mentioned in my video that prior to figuring out that it was maybe her skin tone was a contributing factor in our state. We're allowed through our Medicaid program four probes a month. And so we were going through four probes a week, trying to figure out what our problem was, if there was premarket testing that said, it's accurate for 30 minutes, it's accurate for four hours, it's accurate for whatever I'd have known then to turn the machine off, give it a minute, observe her, turn it back on, and give it another try and that would have been helpful for us in the early stages because that's what we do now is we stop, we let the machine take a moment, and then we restart or remove the probe and place it somewhere else.

Mr. McClure: And then my concern is different from Miss Jolly's in the sense that, whereas hers needs to be consistent over long periods of time, mine needs to quickly give a result because I

need to know what I need to do in response, whether I need to utilize my rescue inhaler. Or if it's if it gets below a certain point, do I need to call EMS? Those are the kinds of those are the reasons that I need mine to react quickly and mine because I have two different types, but I only use one. But while this was going on, I put the two of them on my fingers to test them, and one was very different. The reading was very different by, say, maybe six points on one versus the other, and I wondered what the cause was that one of them was less effective on my darker skin, or is one of them just less effective? I don't know, but mine needs to be quick because I need to respond, and I need to keep myself from going into a state of panic, which then just makes my shortness of breath increase even more, thus lowering my O_2 level even more.

Dr. Cassiere: Dr. Bickler, are you responding to a specific question?

Dr. Bickler: No, I was going to make a comment to Dr. Barker's presentation, if that's possible.

Dr. Cassiere: Actually, this is for panel members to ask questions to the speakers. And for time's sake it was, you can only respond to specific questions that were asked.

Dr. Bickler: Very good. Thank you.

Dr. Cassiere: No problem. Dr. Goldman.

Dr. Goldman: Yes. Thank you. Julian Goldman. It's a question for both Ms. Jolly and Mr. McClure. I wonder if either had ever had an opportunity to contact or speak with the manufacturers of their pulse oximeters to try to figure out how they could improve the performance and to better understand why they were having that difficulty.

Mr. McClure: I have not, but that is a that is a great suggestion, and I'm not opposed to doing so. Thank you.

Dr. Goldman: You're welcome.

Ms. Jolly: Yeah, I have not either. Mr McClure, I'm not opposed to doing so, but in the moment found usable workarounds and moved on to the next urgent need.

Dr. Goldman: If I could ask, perhaps it's difficult to figure out how to do that. I'm wondering, did you try, and was it hard to figure that out, or you just didn't have a chance to an opportunity to do that.

Mr. McClure: I didn't reach out to the manufacturer. It didn't even cross my mind until you mentioned it.

Dr. Goldman: Great. Thank you very much.

Panel Deliberations

Dr. Cassiere: All right. Thank you. This is Dr. Cassiere again. We will now begin the panel deliberations. Although this portion is open to public observers, public attendees may not participate except that the specific request of the panel chair. Additionally, we request that all persons who are asked to speak identify themselves each time. This helps the transcriptionist identify speakers. During the next hour, we will open up the floor to questions for the FDA. Is the FDA prepared to respond to panel questions poised today?

Dr. Eydelman: Thank you, yes.

Dr. Cassiere: And does any panel member have any questions or comments for the FDA? And this is what's going to be the section for our deliberations today. And I see Dr. Ballman has her hand up.

Dr. Ballman: Yes, I have a question with respect to this study design. I understand it's powered appropriately for 24 patients. But it wouldn't be powered for the six and each individual sort of bucket. And so I'm wondering if the coprimary endpoint for the absolute difference is supposed to capture that perhaps, one bucket not doing as well as the other.

Dr. Eydelman: Hi, this is Dr. Eydelman. Dr. Cassiere, if you don't mind, I have asked my team to answer first the questions that we didn't get a chance to answer fully from this morning.

Dr. Cassiere: Oh yes, that's true. There were two questions. Dr. Saville and Dr. Lanzafame had very good questions for the FDA, and I think this is the time to deliberate on that. Thank you for reminding me, Dr. Eydelman.

Dr. Eydelman: Thank you. If I can ask Dr. Hendricks and Dr. Pennello to come on camera, and if I can ask our IT support to project the relevant slides. Thank you.

Dr. Hendricks, do I need to start while IT is getting the slides up?

Dr. Hendrix: Sure, I'll go ahead first. Could you project slide 43, please? I believe there was a question asked about what happens when sensor sites are at a different place other than the fingertip. So, I want to address that question. While the agency considered the strength and level of evidence and uncertainties and the benefit-risk consideration to arrive on the proposed pigmentation measurement methodology, we considered data from desaturation labs and ongoing real-world studies, and we noted that there is the difference in the amount of pigmentation or melanin distributed across the body, for example, forehead versus the inner aspect of the upper arm or the palm or surface of the hand. So, by enrolling MST 1 through 10 from the forehead, where there is a very wide pigmentation range, we hope to enroll participants across the entire range of skin pigmentation. And the assumption is, therefore, subsequent ITA at sensor sites, wherever that is, it'll represent the gamut of sensor site pigmentation relevant to that anatomical site, which then allows for analysis to ensure non-disparate performance. I hope that answered that question.

Dr. Eydelman: Please move on to the next.

Dr. Hendrix: Yes, there was another question on why 1.5% and 3.5% were derived on for acceptance criteria. Again, the agency considered feasibility. That is the limits of current technology as well as clinical relevance to arrive at the lowest absolute difference in SpO_2 bias that we believe is clinically relevant acceptance criteria for non-disparate bias performance. As

we presented this morning, our proposed acceptance criteria will translate to a true maximum difference of bias of 0.5%. That's within 1% for SaO_2 ranges above 85% and 2% for SaO_2 between 70 and 85%, and that is with 80% power in the proposed sample size of at least 24 participants with 480 data pairs distributed evenly across the SaO_2 range of 70 to 100%. Since the differences in mean bias typically increase with decreasing SaO_2 with current technology, the agency chose a clinically relevant cutoff of 85% to include most of the important evidence-based clinical decision points or thresholds that is important for interventions. And we hope today to hear from the panel members on our proposed acceptance criteria for non performance. Thank you. And I'm going to ask Dr. Pennello to respond to the very important question about ARMS.

Dr. Pennello: Yes, this is Gene Pennello. Could you present slide 37, please? There was a question this morning about what I characterize as the FDA tightening the ARMS goal or not, and I would like to elaborate a little bit about that. FDA is tightening the ARMS goal relative to the 2013 FDA guidance. The success criterion for ARMS in that guidance is that the sample value of ARMS in the study be less than 3%. For example, the ARMS value of 2.9% is less than 3%, so it meets the success criteria in the 2013 guidance. However, this sample of the value of ARMS is an uncertain estimate of the true value of ARMS. The 95% confidence interval on ARMS quantifies the uncertainty of the true value of ARMS. The true value of ARMS will lie in the 95% confidence interval in 95% of applications. In the discussion paper, the new success criterion for RMS is that the sample value of ARMS be significantly less than 3%. To meet this success criteria in the upper limit, the 95% confidence interval has to be less than 3%. For example, a sample value of 2.9%, having a 95% confidence interval of 2.7 to 3.3%, say, would not meet the new success criteria because the upper limit of that 95% confidence interval is greater than 3%.

For a study to have 80% power to show that the RMS is significantly less than 3%, the true value of ARMS has to be considerably less than 3% to account for the uncertainty, that's of the sample value. If the true ARMS is 2.1% or less, then the power should be at least 80% that ARMS will be significantly less than 3% of the study and thus pass the new success criterion. But it's important to note that the true value of ARMS is actually a function of three parameters. It's the within and between subject variances of the difference SpO_2 minus SaO_2 and the SpO_2 bias, and different values for these parameters affect the sample size, even if they lead to the same true value of air mess. However, for most configurations of the parameter values leading to a true ARMS less than 2.1%, the power is greater than 80%, and ARMS is significantly less than 3% with a sample size of 24 subjects and 480 total measurements.

Dr. Cassiere: Thank you for that clarification. I'm going to go around and Dr. Lewis, I didn't forget about you. You've had your hand up for a while. Question?

Dr. Lewis: Thank you. This is Tamorah Lewis. Excuse me if this question was discussed in the 2022 meeting because I was not part of that one, but based on the published literature and data we've been shown today, it seems like perfusion or blood flow is a significant interactor with skin pigmentation. And so, my question to the FDA is, have you considered incorporating some variation in local blood flow into the current premarket study recommendations? And are there ways to experimentally change local blood flow safely and healthy volunteers so that variables could be incorporated into current premarket study recommendations?

Dr. Eydelman: Thank you for your excellent question. I'm going to ask Dr Hendricks to give the response.

Dr. Hendrix: Thank you for that important question. The agency considered the quality and strength of evidence and uncertainties regarding this issue. Percent modulation or perfusion

index is currently an optional, unstandardized measure of signal strength for ISO standards 80601-2-61.

As presented earlier today, the premarket desaturation lab studies allow for the warming of hands and optimizing the percent modulation per reference pulse oximeter to control for this variable confounder. The proposed approach does not account for low percent modulation or perfusion index currently. The agency is carefully reviewing well-designed real-world studies such as CERSI UCSF and Stanford studies that evaluate pulse oximeter error and performance due to percent modulation variation across the entire skin pigmentation to study these important signals. And today we do hope to hear from the panel on this important topic. I will ask Dr. Weininger about whether he would like to add to my response in terms of the question you asked about if there are any clinical studies that are being done for the perfusion index. Thank you.

Dr. Weininger: Thank you. Dr Hendricks, Sandy Weininger, FDA. I'll say briefly that there are currently two methods under study to try to address your issue. One is involving cooling one side of the body. And that's your test side and making measurements on the other side of the body to try to get a a handle on what happens under low perfusion. And the other effort is to try to use a simulator to imbalance the red and infrared and to lower the percent modulation. Both of those methods are pretty complex and remain to be validated. Thank you.

Dr. Cassiere: Dr., I just want to put you on hold for a second. I just have a follow-up to Dr. Lewis's important question. The possibility of doing real-world tests since some of the laboratory data does not show occult hypoxemia, but in the real world, it does, and that ties to this perfusion index and low perfusion with oximetry. That seems to be the problem with highly pigmented individuals; is it possible? And again, I'll open this up to the panel members in a moment that there should be some real-world testing that does besides the lab. And that could

help with with answering this question. I don't mean clinical trials; I mean real-world use of the that's going to be approved and ICU patients.

Dr. Malvina Eydelman: Thank you very much. Dr. Hendricks, can you give a short version of the answer to this? We've had a lot of discussions.

Dr. Hendrix: Yeah, I think this is a really important question. Thank you for that, Dr. Cassiere. The agency carefully considered the strength and quality of evidence, uncertainties, and benefit-risk assessment when proposing the improved premarket clinical study, and the agency's proposed approach considers pulse oximeter performance across the entire range SaO_2 of 70 to 100% to validate accuracy and demonstrate non-disparate performance across the entire skin pigmentation. We believe that the control desaturation lab studies on healthy participants as most appropriate to safely and accurately demonstrate this since healthy volunteers are considered a less risky patient population, and therefore, the proposed approach has a more favorable benefit-risk ratio.

Conducting this type of validation in patients with multiple co-morbidities, non-healthy patients may pose additional and necessary risks to volunteers. Additionally, we note that the controlled desaturation lab studies allow for control of confounders such as the prevalence of hypoxemia, non-simultaneous SaO_2-SpO_2 data pairs, changes in percent modulation motion, and other factors, and thereby, these desaturation lab studies allow for accurate validation of overall error, as well as non-disparate performance assurance.

And just wanted to also go in a little bit; even though you have well-designed real-world pulse oximeter studies, we note the following important challenges: non-randomized comparisons of skin color, groups on SpO_2 bias and occult hypoxemia, the need to account for hypoxemia prevalence since cult hypoxemia rate increases with hypoxemia, and the paucity of

data of SaO_2 below 90%, where SpO_2 bias is greater and this is due to the standard of care of SpO_2 to targeted treatments by clinicians.

Importantly, how an individual performs across the full range of SaO_2, 70 to 100%, has not really been will get studied. But despite these limitations, it's important to consider how such free market, non-disparate performance translates to real-world performance, and the agency will closely monitor whether the new approach will reduce disparate pulse oximeter performance in sick hospitalized patients. Additionally, the agency encourages well-done prospective real-world studies to investigate this and regularly monitors and assesses published literature with regard to this subject for your time. And rigorous real-world studies such as the CERSI studies will provide important information on the interactions of pulse oximeter skin pigmentation and clinical factors such as percent modulation that will inform the agency's approach. And we hope today to hear feedback from the panelists on this very important topic. Thank you for your question.

Dr. Cassiere: All right, Dr. Hendricks. Thank you, Dr. Hendricks. Appreciate that. We'll have more to say about that in the afternoon. Dr. Feldman, sorry to keep you waiting.

Dr. Jeffrey Feldman: Not at all. Thank you. It's a great discussion. And I want to compliment the FDA in general on the approach this morning in the organization. It's been great. My question goes to the proposed sample size and subgroups. But before I ask the particular question, I just want to offer a little perspective that motivates the question. When the Sjoden et al. study and subsequent studies came out, I think what it hit us in the face or hit me in the face was that, is there a real-world function of pulse oximeters that's different from what we had documented previously in the laboratory and with prior FDA requirements? And so if we take that data at face value, we can see that laboratory data is not going to be predictive ultimately of real-world experience. And I think that's going to be true going forward.

With regard to the question of non-disparate performance. I think the goal, a primary goal, is to put to rest the question of whether or not race, ethnicity, more generally, and skin tone, specifically, plays a role in pulse oximeter performance. So that, to me, is an overriding goal in putting this together. And I do think there are a few concerns that I would love to hear more discussion about with regard to the proposals about whether or not it will really put that question to rest with the new requirements.

With regard to the patient population definitions, and I think that's probably the biggest factor, 24 feels like a short, a relatively small number, but more specifically, the subgroups of six to nine patients in each subgroup feels small to me. And especially when you start to think about applying more inferential statistics than maybe have been discussed so far. So, for example, I would like to see an inferential statistical comparison between groups of patients with skin tone to document that, indeed, with the data provided, there is no apparent difference in performance between those groups.

By pooling data on ARMS across all of the subjects, I think you run the risk of hiding problems with performance in certain groups. And I think that was certainly true with the previous criteria where there were only a couple of patients of dark skin tone that probably the performance was not apparent given the large number of patients that didn't have that criteria.

So I think my specific question is group size, but also in particular around inferential statistics, comparing groups with different skin tones and appropriately powering the groups in those categories for those statistical analyses. Thank you.

Dr. Cassiere: Dr. Lanzafame

Dr. Lanzafame: Yes. Raymond Lanzafame and, thank you for the the discussion on perfusion. There's still, though, a question of, I think, other issues with real-world performance, which I think may have in part, been reflected in the two patient presentations. Specifically, there is some

degree of red shift of the devices as the devices heat and likely in the real world as the site of the read is compressed. There may be some differences in that. I wonder if the study designs by allowing for warming of the hand also are taking into account a time course or red shift of the optical component as well as the thickness of the tissue being measured question.

Dr. Malvina Eydelman: I believe it's a combination between Dr. Hendricks and Dr. Pennello. Dr. Hendricks, I don't know which one of you wants to start because there's still a statistical question.

Dr. Cassiere: Just one second. Dr. Eydelman, your audio seems to be a little off. I'm not sure if it's just me, or is it anyone else that hears that?

Mr. : It is slightly lower. So you may need to project a little bit just to help everybody out. Sorry.

Dr. Cassiere: Thank you. I'm sorry. Not to point that out, but it was just hard to hear some of the comments that you were making.

Dr. Eydelman: How about now? Is it better now?

Dr. Cassiere: Much better. Thank you.

Dr. Eydelman: Let's take it away while I play with my microphone.

Dr. Cassiere: Wonderful.

Dr. Malvina Eydelman: I'll go ahead and answer the last question. Dr. Pennello is going to take the question on statistics for Dr. Feldman. What we do know is in terms of, I think your question was how long the warming techniques allowed in desaturation labs and the length of where the pulse oximeters for desaturation testing versus, real-world where the pulse oximeters can be placed for a very long time, while I'm not an expert on red wave shifts, and maybe Dr. Pfefer can follow that up if he has additional input on that response, I will say that the desaturation lab study typically for an individual, because it's a step-wise decrease in FiO_2 with stable plateaus

that allow at least 30 seconds or more for five samples to be taken at every, stable plateau level. It is about two hours, plus or minus one hour, typically at least in many desaturation lab studies.

So the remaining question is, is that enough to account for differences in the red light wavelength? I'm not sure about that. But I hope that you give some input on that. There are considerations. I think there are real confounders in the real world in physiology, maybe in the application. Sometimes, as a clinician, you know where you put the pulse oximeter. If it doesn't give me a number that I like, I sometimes will change fingers. I will put it in different parts of the anatomical area, maybe on the earlobe with a finger probe. So, all that varies, and certainly, a desaturation lab study does not capture all of those. Dr. Pfefer, if you have anything to add, please go ahead. Otherwise, we'll turn the question to Dr. Pennello to answer Dr. Feldman's question.

Dr. Pfefer: Okay Josh Pfefer from OPEQ CDRH. I would just briefly comment on the redshift idea and I'm not sure if that is referring to a thermally induced redshift. I believe that probably won't happen in the devices until very high temperatures. There is another redshift effect that's been proposed due to melanin content. And that certainly still is an outstanding potential mechanism that I think has not been thoroughly vetted. So, we're still trying to answer some of these questions about the fundamental mechanism.

Dr. Malvina Eydelman: Thank you Dr. Pfefer. Dr. Pennello, did you want to go back to the statistical issue?

Dr. Pennello: Yes, I wanted to respond to Dr Feldman's question about subgroup analysis and certainly agree with a sample size of 24; there won't be enough power to show differences between subgroups may be defined by statistical significance.

And what we had opted in the discussion papers not to think about is a continuous variable and then a statistical model so that we can compare the SpO_2 bias. Or an ITA value versus another

value, but use the data all of the data to estimate the SpO_2 bias at any ITA value, so thereby increasing power, getting away from dichotomizing of the data, and using continuous ITA for the Sp. Yeah, for one thing that could be considered, it's a normal variable. So, you can consider ordinal data statistical models to exploit that ordinal nature to maximize power to see differences.

Dr. Cassiere: Great. Thank you. Dr. Goldman.

Dr. Goldman: Yes. Thank you. Julian Goldman here. First of all, I just would like to say that I really applaud the efforts of the FDA to reduce disparate bias and pulse oximetry. This has clearly been a very heavy lift for a long time, and we need to address it. So thank you.

My question is about the proposed acceptance criterion. Given that the exact mechanism of disparate bias is still under investigation by many groups, and it was just underscored by Dr. Pfefer, does the FDA have data or a sense of the ability of manufacturers to achieve the new success criterion of the new performance criterion? And related to that, will the additional testing and acceptance criteria affect the market availability, especially of lower cost pulse oximeters? As we saw with the COVID-19 public health emergency and the surging need for pulse oximeters, it was very difficult to find those, and naturally, during an evolving, testing protocol, evolving criteria, and in a time where we're still trying to sort out the root cause of some of these performance issues, I wonder if that would affect us affect the market and wondering what the FDA is thinking on this is at the moment and whether there is any data available.

Dr. Eydelman: Thank you, Dr. Goldman. So, while the team has thought about all of this for a long time, for the purposes of today's meeting, I would like to turn the question around and see if the panel can give their thoughts and recommendations on these topics for us so that we can incorporate them into the path going forward.

Dr. Goldman: That really is the essence if we incorporate a performance criterion, which we would all love to see at a time when we are uncertain of the root cause of the disparate performance and if we don't have a good sense of what the impact will be on the market, we could end up having a marginal increase in performance and suddenly severely affect the market. I'm eager, I think, as probably are others, to have much more data. If I had a magic wand, we would have much more real-world evidence on more patients, under more controlled conditions vastly more data points, and really continue to be digging into this very complex issue.

And I understand we're trying to build the plane while we're flying it. And that's the way life is totally agree and accept that. But I think perhaps I'm raising the point that we have to be cognizant of what the downside might be. So, even though we all want the same thing, we don't want to cause a problem unintentionally.

Dr. Eydelman: And we definitely agree with your sentiment, and we're trying to maximize all the information available, including today's panel input, in order for us to proceed in the most scientific, evidence-based manner.

Dr. Julian Goldman: Yeah, and thank you for doing that.

Dr. Cassiere: Great. Dr. Wilson, you've had your hand up.

Dr. Wilson: Yes, I do. I would like to further address the discussion point of laboratory versus real-world data. And firstst, I want to commend the remarks that Dr. Hendricks made, I think, particularly in terms of what's ethically and appropriate and what is actually safe. It's important that we do get measurements between 70 and 90, not just 90 and above, and we have to just, recall that pulse oximetry is a very dynamic measure. It changes quickly, and some of the problems with some of the real-world data have been a lack of contemporary areas measures, but also not ensuring that there's a stable plateau as we ensure in the laboratory. When individuals do that real-world work, they'll want to, going forward, ensure that there's a stable plateau

before making measurements. Other factors include the various sensor site locations that was demonstrating the equidox data again, that many individuals in the clinical domain may take a probe that was meant for a finger and place it on the air to get a measurement, but it's not developed for that site. You'll get a signal, but the accuracy is going to be in question because it was not engineered for that location.

There are a number of other confounders that need to be considered. And when one wants to look just at skin pigmentation, then it is very important that you eliminate the cofounders. And that can be best done in a laboratory setting.

Dr. Eydelman: Thank you very much, Dr. Wilson.

Dr. Cassiere: Dr. Yarmus.

Dr. Yarmus: Yeah, thanks, Lonny Yarmus. So maybe a somewhat brief comment just focusing on the same arena of real-world data. Back to Dr. Hendricks's comments earlier speaking from a clinical trialist perspective and an intensivist. I think the risk question is interesting because I would foresee this as an opportunity to really get this real-world data without significant risk. So patients, for example, in an adult, I see you test these types of metrics that we're talking about. There's not really an inherent risk, right? There's a baseline. Equipment that could be utilized and that is what would be used for the clinical parameters. And then we have that opportunity to test these newer devices that are on the market, and really, I think, focus on the bulk of the issues that we're talking about in a real-world scenario with all of the different parameters, including the variants and different oximetry levels. Thank you.

Dr. Cassiere: I just want to step off on what Dr. Yarmus just said, that the risk that the patients in the ICU, they've inherited their risk by being in the ICU. And I guess we'll talk about it later in the afternoon. At least my read, the bulk of the data shows low perfusion, dark pigmented equals higher occult hypoxemia. That's the real meat. And I guess what others have said, and Dr.

Yarmus said pretty good, And so did Dr Goldman. We'll go through all of this to find out that we have really good accuracy with dark-pigmented and normally healthy adults with normal perfusion and not fix the problem. And I think part of our discussion in the afternoon for everyone is how we mix up a perfusion index into maybe the healthy laboratory.

And again, we have some, I have a couple of ideas, we'll talk about later. And again, to translate to what Dr. Yarmus said, why do device companies not just have relationships with hospitals and health systems and ICU providers to do a quick, not a study, just to take a look at their devices under low perfusion? And again, I guess we'll talk about that a little bit more later on, but I don't want to take away from Dr. Saville; you have your hand up.

Dr. Saville: Yeah, thank you, Ben Saville. So again, I'm the statistician here, so I'm going to try not to get in the weeds here. But I've been thinking about the models and the analysis that have been presented and they're all talking about regressing the SpO_2, doing linear regression, for example, on covariates and the SaO_2. And in my head, I feel like it should be the other way around. I feel like the endpoint is really the arterial oxygen saturation. And what you're trying to do is predict that using observed data. We're not going to get it from everyone, right? We're not going to take a blood assay, but we're going to take this pulse oximeter.

Has the FDA given any thoughts? And certainly, the inverse prediction that Dr. Pennello showed on, I think, slide 25 kind of hinted at this. But in my mind, I was thinking, this feels like a prediction model problem to me, where you have the oximeter reading, you have the list of covariates, and you want the best prediction of what the real oxygen saturation level is and there'd be uncertainty.

You're going to try to quantify that, put an interval around it, and then make decisions on what that interval is. And I don't know how complex these oximeters can get. I don't know if they can have more complex algorithms built in where they can give you a predicted value, but it

just feels like right now, the assumption that the oximeter exactly equals the real value. It just seems like a stretch to me. I feel like there should be a prediction model. So, I'm curious if the FDA has thoughts on that.

Dr. Eydelman: Dr. Pennello?

Dr. Pennello: Yes, Gene Piniello, FDA. Thanks for that question. Yeah, the most clinically relevant metric here is the prediction of the SaO_2 value. The problem with having statistical models of SaO_2 with SpO_2 as a predictor is that SpO_2 has measurement errors, and these statistical models assume that the predictors are measured without error, so it could distort the estimates of the parameter values.

And another limitation of that is that it will depend on the SP two distribution. So, to get away from both of those limitations, it's better to have, and really, if you think about it, SaO_2 is the reference value, and SpO_2 is the estimate of it, and it responds to it. So, to me, statistically, it's more natural the outcome plus SpO_2 is not measured with error, at least much less error than SpO_2, so the regression models are more appropriate. I hope that responds to your question.

Dr. Saville: That helps. I'm still racking my head in terms of, I feel like you could do it either way, but I just feel like there's potential here to potentially look at the other way around. Thank you.

Dr. Cassiere: Dr. Wilson.

Dr. Wilson: Yes. Thank you. I wanted to further elaborate on the excellent point that Dr. Yarmus made about using individuals in the intensive care units with already preexisting arterial lines as convenient samples. Of course, that's excellent. One of the concerns there, however, is that we rarely allow individuals in that setting, in the ICU, to have saturations that are below 90% for very long. Certainly not in the 70 to 80 range. Now, occasionally, that does happen, but typically we are making changes, we're addressing it, we're doing something often actively to the

individual in order to get the saturations up. So, in order to get adequate numbers in those ranges, it's quite difficult, and of course, in order to do a simultaneous arterial blood sample and pulse oximetry value, one needs to make sure that you have a stable plateau. And then I just wanted to further emphasize that in real-world settings, your reference devices, including the cooximeters, can have variability multi, sometimes cooximeters can vary by 1%, and that's a sign significant range. So, you need to make sure that your reference is stable and that other pre-analytical and then post-analytical errors are not at play. So, there's a number of factors that are more difficult to control in those settings.

And just one final point, which is in the pediatric population, it's a little bit easier. In the cath lab, if you're dealing with congenital heart disease children who will have SaPs that are down in that range so, that's a population that individuals should consider. And with the increasing numbers of adults, with some still right to left side mixing, special populations could be found. But in general, it's quite rare in the intensive care unit.

Dr. Cassiere: I'm going to push back a little on that, Dr. Wilson. I think what Dr. Yarmus is; I see him shaking his head; I'm speaking for you, though. But what I'm concerned about is having an O_2 set of 95% on a patient who has a low perfusion index, and that O_2 sat of 96% really is arterial saturation of 88 or 86. That's what we're talking about. We're not talking about looking for patients in the ICU that have O_2 sats of 85% consistently. That's a separate issue. We're looking for this occult hypoxemia. And that will answer, I think, and again, we'll talk about this later on, that will answer if there's a cold hypoxemia in real-world patients who have low perfusion. And I don't think that's a big ask. And I think that would help answer some of the questions that are still outstanding about this perfusion occult hypoxemia connection.

Dr. William C. Wilson: Agree on that one. No dispute there. That's very important. We would, of course, need to ensure stable plateau values in order to get a contemporary measurement.

Dr. Cassiere: Great. Thank you—Dr. Feldman, your hands up again.

Dr. Jeffrey Feldman: Yes, thank you. Jeff Feldman. So, another issue I wanted to raise is the way we do skin tone assessment. I think there's been a lot of conversation already about the challenges of doing that reliably. The ITA Is a new one for me to learn about, and it strikes me that with all these skin tone assessments, none were developed with the purpose that is in mind for what we're trying to do with pulse oximetry. And that is to determine the influence of skin tone on the transmission of light from the emitter to the detector at the sensor site.

So, the ITA perhaps gives some insight into that, but in reality, that was validated based upon a histological assessment of melanin content, not actually on how the light is transmitted through a digit or wherever the sensor site is. And if we were going to come up with a better technique, I think it would be something that looked more specifically at how each frequency of light was transmitted at the sensor site and how it was impeded on the patient who's being tested. Getting back to this notion of how we put to bed the question of non-disparate performance in the laboratory setting, knowing that real-world experience will become different, I do have some concerns even about relying on it because I don't think it really gives us the information we need to understand. It may turn out to be the most practical approach, but it may fail to do what we want it to do at the same time.

Dr. Cassiere: Thanks, Dr. Feldman. I guess that's going to be very ripe for our panel discussion later on when we answer the FDA questions. So please keep that on the forefront. Dr. Saville, your hand's up again.

Dr. Saville: Yeah, thank you. Another stat question. It's on a different topic, but so statistically, I like the idea of using the upper bound of 95% credible interval or confidence interval for that first criterion, the ARMS. I'm still struggling a bit with the criteria for the non-disparate performance. And that this idea of the 1.5% and 3.5% thresholds being the maximum model

based bias. That feels like an ad hoc criterion to me. It is model-based, at least. But since you have a model, why has the FDA given thought to using more inferential statistics? P values, composite of holes for that interaction term, or even a Bayesian model with a Bayesian posterior probability, where you can actually get the probability that there's an association and the bias depends on the level of the skin pigmentation. I'm curious about his thoughts on that.

Dr. Eydelman: Yeah, so I'm going to ask Dr. Pennello again. I understand he has a backup slide. I don't know, Dr. Pennello, if you wanted to take a stab at this slide because IT would need a little warning before you can fill it out.

Dr. Pennello: Yes. Dr. Eydelman, so I have backup slides that respond to Dr Feldman's question about data or whether the non-disparate performance insurance would actually be met or not. I'm sorry. You're muted.

Dr. Eydelman: Perhaps we can go to the next question, and we can ask IT to pull up that slide, and we can circle back.

Dr. Cassiere: Dr. Saville.

Dr. Saville: Yeah, that's not my question. My question was something different. If we could answer that one, that'd be great.

Dr. Pennello: The 1.5%, 3%, and 3.5%. To have the power to show those in a sample size of 24, your true maximum differences in bias will have to be a lot less than that, even for the point estimate to be less than those acceptance limits. And our analyses have shown that you'll have power if the maximum difference in bias is 2% to show the 3.5%, and you'll have power to show the maximum difference is 1.5% if the true maximum difference is 0.5%. So, the 0.5 and 2% true values. I think we felt that those are probably the current limits of technology that's probably the limit of the best true values that could be expected for accurate pulse oximeters and the limits for

having power to show that the point estimates are below 3% were determined on that basis. At least, that's my recollection.

Dr. Saville: So I understand you chose those thresholds based on what you had statistical power for, but my question really is more about have you thought about using inferential statistics and even looking at the power of some sort of criteria based on inferential statistics, such as a Bayesian posterior probability or a confidence interval, that sort of thing. That's my real question.

Dr. Pennello: Yeah, so that's actually for the international standards organization, who's also revising their standard for essential performance of pulse oximeters. We had looked at Bayesian models and desaturation studies for them. That was a while ago, and then that hasn't been considered, but we can certainly consider Bayesian models. I don't think, regardless of the statistical analysis with a sample size of 24, it would be difficult to show those limits of 3.5% and 1.5% with statistical significance because the sample size will be too small. But, the basic models offer a very flexible way of modeling the data. Thank you for asking that.

Dr. Saville: You could drive some decision criteria on a model that has nothing to do with statistical significance. So you could certainly look at that.

Dr. Pennello: Yes.

Dr. Saville: Thank you.

Dr. Cassiere: Dr. Punjabi.

Dr. Punjabi: Naresh Punjabi. I'm also looking at the same concept that has been discussed by Dr. Saville on this issue of proposed acceptance criteria for nondisparate performance, focusing on 3.5. While much of this discussion has focused on potential implications, what happens in the ICU, there's a whole world of oximetry usage that happens when the lights go off, which is in the area of sleep and breathing. And in fact, this 3.5 criteria has huge implications.

Specifically, when we think about sleep-disordered breathing at nighttime, our definition for the disease process requires a 3% desaturation. So I'm very surprised at this. It's a comment and a question, and I'm wondering what the implications are because we often have patients that will desaturate into this zone at nighttime. And if you're telling us that 3.5 will be acceptable, boy, we have a lot of misclassification going on and diseases and misdiagnosis or over diagnosis that will run crazy stuff in our field. So this 3.5 is because that is the definition for our disease, a 3% desaturation, and people will go down to that level. So I'd love to get the FDA's comment on this and think about how this will impact a classification of disease if these are the criteria and disease, that's very common. Thank you.

Dr. Eydelman: Before we answer this, I understand it is ready with this slide. So perhaps Dr. Pennello, you can answer the question you were trying to answer before, and Dr. Hendricks will ask the question that was just asked after Dr. Pennello finishes.

Dr. Pennello: Can I clarify first before I go into these slides. The 3.5% refers to the difference in bias between two skin color levels as measured by ITA. So, if you had a bias of 2% at the dark end of ITA and then an SpO_2 bias of minus 1.5% at the light end of ITA, that difference is 3.5%. That's right at the boundary of the performance goal. But neither of those two values is in absolute value greater than 3% or 3.5%. So, I hope that clarifies what we're trying to do. It's the difference in bias, not the absolute bias.

So, I wanted to respond to Dr Goldman's question about do we have any data on whether the non-disparate performance criteria can be met. So, we did look at a data set. This is just one data set of 12 devices in nine locations. And so these are 12 pulse oximeters and I've listed them in order of ARMS, and device 11 has air at 2.1, so it's pretty darn good for a sample size of 14 subjects. I just wanted to use the real data. I didn't want to simulate 24 or anything like that. And you can see the blue values indicate the non-disparate performance criteria of 3.5% in the

interval of 70 to 85%, which are being estimated at the midpoint 77.5 of O_2. So the difference in biases are all being met for this device at all nine locations, and for the other interval of 85 to 100, where we evaluate at the midpoint, 92.5, the differences in bias are all met, except for location 3. Now, the differences in bias being less than 1.5% at the bottom here.

So, if you go to the next slide, this is another device and all the, and on this device we had data on 13 subjects and 376 total pairs. And you can see that the performance criteria are met at all locations for this device. And then we go to the next one. This has an RMS of 2.2. The next one is 2.3. This one actually had 34 subjects. And all the performance criteria are met. And so you can keep going. But as you go, you advance a couple more slides, you'll start to see is actually that it doesn't always work. So, this is 2.5 in the study, but for this device, it's not meeting the non-disparate performance criteria. So it's not guaranteed that if you're ARMS is good enough to pass the first coprimary objective of being less than 3% with statistical significance, and it is here because the confidence intervals 2.3 to 2.8 doesn't necessarily mean you're going to meet this non-disparate performance assurance.

But for most of the devices they have pretty good ARMS. They're meeting the non-disparate performance criteria at most locations, and for ones where ARMS is pretty large, these are often not met. And there are more slides, and I don't know if you want to advance one more. This one, it's meeting them all. Let's go to the next one. Even this device 4, with an ARMS of 2.7, is meeting all the criteria at all locations. Maybe one more. See, this one is missing a few and the next one. As the ASMR gets larger, the trend is that the non-disparate performance criteria created tend to be missed at some locations. So that's what I wanted to present just to respond to Dr. Goldman's question about data.

Thank you very much. We appreciate you sharing that. May I ask a question about the data that Gene just shared with us?

Dr. Cassiere: Yes.

Dr. Goldman: I realize that this might be too complex of a deep dive, but based upon what you've shared, it does make me wonder if we will be able to see reasonably good performance in terms of meeting a non-disparate performance goal in the lab, and yet still find, emerging and new data sets in the real-world evidence data collection that has the outliers that have raised the questions that we're contending with today. So Gene, based on the data that you analyzed, do you think that might still be a challenge that we'll see good quality data in the lab and the clinical studies on volunteers, but perhaps still be wrestling with these outliers in clinical environments?

Dr. Pennello: Yes, thanks. Dr. Goldman, for the question. Gene Penelope. That's certainly possible. [crosstalk] But that's all I can say about that.

Dr. Cassiere: Thank you. Dr. Saville, before we go on, the FDA needs to respond to Dr. Punjabi's out-of-the-box question about, uh, reclassification of sleep apnea patients based upon the standard.

Dr. Eydelman: Yes. Dr. Hendricks, Dr. Punjabi.

Dr. Hendrix: Good afternoon. This is Kumudhini Hendricks here from OHT1 CDHR OPAQ. I just want to say we modeled from well-performing pulse oximeter desaturation lab data to arrive at what is feasible, but also clinically relevant. And I understand; I hear what you say about the 3.5% SpO_2 bias below 85% with respect to especially sleep apnea patients. One high-level comment that I will say is while we're talking about discreet SpO_2 differences versus trending, it's a little different. So for a given patient and how they desaturate from 92 to, let's say 87, the whole time, they could be like, if they're within that range, if they transverse, of course, the 85%, it pops you to a different SpO_2 bias difference. But that is different from what we're trying to ensure as SpO_2 points. Whether it's 85 or 83 or 87. But broadly, what I want to say is, in our talks, I believe we said even though the difference in absolute maximum SpO_2 bias is for SeO_2

levels less than 85% is 3.5%, the true difference is 2% with 80% power for the sample size. I just lost my train of thought for the other part that I wanted to say, but we hope to get feedback from the panelists on the acceptance criteria. Whether a 3.5% or the 2% true difference is something that really is not clinically relevant. This is important feedback from the panelists today. And we look forward to that. Thank you.

Dr. Cassiere: Great, we have four minutes till our break. Dr. Saville, you have your hand up?

Dr. Saville: Yes, Ben Saville, I'll try to make this quick. So there's mounting evidence that the bias and these oximeter readings depend on skin pigmentation, and I'm going through in my head, trying to figure out different different scenarios of how are these trials are going to play out. And suppose Dr. Pennello just showed an example where you have some, you have a new device, and it meets the criteria for the ARMS, but it doesn't meet the criteria for the non-disparate performance analysis. And in that case, is that any different than where we currently are? And what would be the FDA perspective of that kind of device where it's accurate? It's generally accurate, but there's a disparate reading, and there's certain skin pigmentations where it's not as accurate. What would be the interpretation of that kind of result, considering that it may not be even different from what we currently have?

Dr. Eydelman: Dr. Saville, I would like to ask you what would be your recommendations. We're here to listen to you. One thing that I want to assure everybody on the panel is for any device that does reach a market; we work very hard to make sure that the labeling is very transparent. So, any recommendations that you might make about what does and doesn't reach the market and what kind of communication and labeling is helpful to both physicians and patients would be extremely useful for us.

Dr. Saville: Yeah, and I think, in other words, there is a lot more to talk about. Yeah, that's the answer. It is going to have to be in the label because I can see lots of gray areas where there's a lot of different scenarios that could play out here.

Dr. Cassiere: Dr. Feldman. You're you're going to round us into lunch.

Dr. Feldman: Thank you, Jeff Feldman. I'll try to be brief then. I think relative to that, just the last commentary, I think it's important to emphasize that these clinical devices are designed with lookup tables. They are not discrete devices that base their measurements on physics. The principles are based on physics, but what comes to the bedside is a device that's manufactured with a lookup table based on a population that the manufacturer studies. And then creates a device that will work for as many patients as possible.

So, I think the presumption in this new effort is that perhaps these new criteria might influence those lookup tables and that future devices may perform better across a broader range of patients, but that the legacy devices are going to be stuck with the way they are, unless their software can be upgraded. So, the way these devices are designed is inherently a big challenge to what we're trying to do because it ultimately relates to creating a process that influences the manufacturers to design something that's more generally applicable than they have to date.

Dr. Cassiere: Great, thank you. With that comment, we're going to take a 30-minute lunch. Panel members, please do not discuss the meeting topic during lunch, amongst yourselves, or with anyone attending virtually, and we will resume at 1.30 PM. Thank you, Everybody.

Open Public Hearing

Dr. Cassiere: This is Dr. Cassiere, it is now 1:30 PM and I would like to resume this panel meeting. We will proceed with the open public hearing portion of the meeting. Public attendees are given an opportunity to address the panel to present data, information, or views relevant to

this meeting agenda. Ms. Nalls will read the Open Public Hearing Disclosure Process Statement. Ms. Nalls?

Ms. Nalls: Both the Food and Drug Administration, FDA, and the public believe in a transparent process for information gathering and decision making to ensure such transparency at the Open Public Hearing session of the Advisory Committee meeting. FDA believes that it is important to understand the context of an individual's presentation. For this reason, FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement to advise the committee of any financial relationship that you may have with any company or group that may be affected by the topic of this meeting. For example, this financial information may include a company's or a group's payment of your travel, lodging, or other expenses in connection with your attendance at the meeting.

Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do not have any such financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

Dr. Cassiere: Thank you, Ms. Nalls. FDA has received five requests. Each speaker will be given five minutes to speak. We will begin our presentations with the speakers who are live, and I'd like to welcome Dr. Ajizian. You can start.

Dr. Ajizian: Thank you. Hi, everyone. I'm Dr. Sam Ajizian. I'm the chief medical officer of the Medtronic patient monitoring business.

My medical practice was in pediatric critical care for over 20 years. I've used pulse oximetry throughout my career, and I'm intimately familiar with the important role it plays in safe patient care. I'd like to thank the FDA for their work on this important issue and for this

opportunity to continue sharing our perspective and recommendations for continuing to enhance the performance of pulse oximeters. I am a full time employee of Medtronic and a stockholder.

I want to start by saying that Medtronic strongly supports the ongoing efforts bringing together industry, FDA, healthcare practitioners, the community, and others to continue to research and advance the performance of pulse oximetry devices. We are very proud of our Nellcor pulse oximetry devices. At the same time, we acknowledge that numerous variables can affect their performance in the real world, including skin pigmentation. We continue to work internally to evaluate and further improve the performance of our Nellcor pulse oximeters. Our primary mission is improving public health. And patient health and our goal is to ensure that our pulse oximeters like all our medical devices. are as accurate and as effective as possible for all patients, regardless of skin pigmentation level.

At the same time, we cannot do this alone. Advances in pulse oximetry require partnerships with all stakeholders, including FDA, industry, organizations like ISO, scientists, patients, local communities, and healthcare practitioners at the bedside. Health equity cannot be achieved without listening, exchanging ideas, and working collaboratively on solutions.

At the last advisory committee, we committed to several areas to drive equity and we remain committed. We consider these four critical pillars to be foundational areas for continued investment to advance equity in pulse oximetry. We believe community engagement helps us to better respond to real world needs and concerns around diversity in industry verification studies. We built our new Medtronic patient monitoring physiology lab in a neighborhood of racial and ethnic diversity and we have forged ties to community leaders to raise awareness and build trust. This positions us well to ensure more inclusive subject enrollment for our studies. And we continue to look for other ways to enroll diverse populations. We also continue to develop new educational tools and offerings to help clinicians understand our current pulse oximeter devices.

We've trained hundreds of clinicians across several events over the past year on Nellcor technology to enable better care for all patients. Innovation requires investment and Medtronic is committed to doing what is needed to continue to innovate and further enhance the performance of our pulse oximetry devices.

And lastly, myself and other team members have personally worked closely with ISO, industry, and regulators on updating and enhancing guidelines and standards for the testing and development of all pulse oximetry devices with the mission of enhancing equity. This work is focused on a wide variety of issues, including diversity and clinical studies. And the use of standardized scales to assess skin pigmentation, among others.

Turning now to the questions raised by FDA, I want to begin by stating that we agree with most of the recommendations made by the agency. We support efforts to harmonize FDA's proposal with the ongoing work of the ISO. As to specifics, we agree with FDA's approach to increase the minimum sample size in studies, including specifically the number of participants who fall in the darkest skin pigmentation categories. We also agree with FDA's approach to standardize the inclusion of skin tone diversity with the use of validated scales such as the Monk Skin Tone Scale, or MST. At the same time, we believe that a tolerance of plus or minus one on the MST could be allowed. This would help facilitate patient enrollment without sacrificing diversity or adversely affecting the value of the study.

Finally, non-disparate bias appears to be a suitable method to evaluate the performance of a pulse oximeter with respect to skin pigmentation. However, we recommend further inquiry and data collection before setting performance thresholds for this parameter.

There are many steps before we can celebrate the change we all want to see. But each step, including today's discussion. Is important in creating a better future. We look forward to the continued momentum of strong partnerships between the stakeholders present here today to

achieve our unified health equity purpose. We're committed to continuing to work together to advance pulse oximetry for all patients and to foster health equity across all aspects of our industry. Thank you.

Dr. Cassiere: Great. Our next speaker up is Dr. Michael Abrams. Dr. Abrams.

Dr. Abrams: Hi there. Good afternoon, everyone. I'm Michael Abrams of Public Citizen's Health Research Group and I have no financial conflicts of interest on this matter.

Public Citizen is a consumer advocacy group with a 50 year history of monitoring FDA activities, including those pertaining to Class II III medical devices. In 2022, we testified before this very Advisory Committee about the inaccuracies of pulse oximeters related to race, ethnicity, and differences in skin pigmentation. Our concerns on that issue remain. Documented pulse oximetry devices in persons with darker skin pigmentation date back to at least 1991 when Zeballos and Weissman observed in 33 black men, the oximeter measured oxygen levels were several percentage points higher than corresponding levels measured by direct blood gas readings.

By the time of this committee's 2022 proceedings, the FDA had collected at least 15 additional studies revealing similar concerns about the accuracy of pulse oximeters related to skin pigmentation. These studies included thousands of subjects, several studies found that needed hospital-based respiratory care was delayed or denied because of low sensitivity pulse oximeter readings. The FDA materials for today's meeting add 13 studies to those collected earlier. Seven of these studies at least confirm bias related to race and ethnicity.

For example, the Fawzy et al. 2023 study of over 24,000 COVID-19 hospitalizations observed that occult hypoxemia rates that is SpO2 minus SaO2 discrepancy of at least 4 percent were evident in nearly 20 percent of black or Hispanic patients compared to 13 percent of white patients. Premarket analysis from the FDA shows that most applications for clearance of pulse

oximeters include little information about race, ethnicity, or skin pigmentation effects, including application cleared as late as 2022. I confirmed this deficiency by reviewing the last 19 cleared applications of 2023 in the FDA's 510(k) database. Only nine of those applications mentioned pigmentation categories more frequently characterized simply as light and/or dark, while two used the Fitzpatrick scale. Only one application disclosed point estimates regarding bias that clearly contrasted in that case just light and dark categories. All 19 devices were cleared because they were deemed substantially equivalent to prior devices, which likely involves even less data on skin pigmentation. Surprisingly, adverse event reports or recalls for pulse oximeters have not recently increased despite this committee's activities and a 2021 FDA communication about device inaccuracies. The FDA's briefing materials for today's meeting confirm the lack of an expected increase in adverse event reports using data compiled by Madris Kinnard with her device event software. I independently confirmed that there has been no noticeable increase in adverse event reports or recalls. The device events data show not only that the recent adverse event signal was low, but also that recalls were infrequent. Moreover, race and ethnicity was not reported in essentially all of the 12,000 plus adverse event reports isolate.

This advisory committee meeting is considering more rigorous clearance standards for pulse oximeters, including more racially and ethnically representative subject pools. Such changes will only suffice if the required subsamples are large enough to generate between group point estimates that give clinicians a true sense of the pigmentation bias inherent in the devices that they're using. Through your deliberations we urge the committee to clarify that underlying statistical power consideration.

And finally, enhanced pulse oximeter testing standards should be implemented for new, all new, and all existing devices. At the same time, all pulse oximeter labeling should be immediately revised to educate and caution clinicians about skin pigmentation bias, as called for

in a November 23, 2023 letter from 25 attorney generals to the FDA commissioner. Devices that demonstrate unreasonable error or otherwise fail to comply with testing requirements should be recalled or removed from the market as efficiently as possible. Thank you very much.

Dr. Cassiere: Next up is our pre recorded presentations with, I believe, Dr Lucas.

Dr. Lucas: Hello, I'm Scott Lucas, Vice President of Device Safety at ECRI. This pre-recorded video is intended to air at the February 2nd, 2024 public hearing convened by the FDA's Medical Devices Advisory Committee. We applaud at ECRI the FDA for convening this meeting on this important topic of healthcare equity, mainly the disparity of performance of pulse oximetry among patients of different skin tones. It is our understanding that a recommendation in the November 2022 Advisory Committee meeting on this topic was standardization of skin pigmentation assessment during clinical trials and development of pulse oximeters. We've provided written comments accordingly in response to FDA's discussion paper in preparation for this meeting today. I'd like to share a few highlights mainly related to the human factors associated with the proposed design approach for skin assessment techniques. I'm not bringing in accuracy of pulse oximetry in general in this discussion. Although that is in our purview as we've been evaluating medical device safety and human factors for over 50 years and pulse oximeters since 1989.

The FDA proposal we're talking about today includes two methodologies to determine skin tone, both of which have merit abbreviated as MST and ITA. But it's important to note the limitations of both techniques and work to mitigate them.

First, MST is based on human interpretation, which is subjective. Second, using colorimetry for ITA requires consistent and accurate use of the colorimeter device. And third, both MST and ITA methodologies rely on reflective light. However, pulse oximeters measure transmitted light, and we must consider how light interacts with the melanin. We recommend

using both methods, as suggested in the FDA proposal, until we better understand their potential disparities.

Based on human factors experience, we offer three specific recommendations to increase the objectivity of these methodologies.

Number one, we suggest a pilot test of the proposed MST protocol to assess skin tone and skin variations across many patients. It's also critical to train users on this tool and evaluate if they can use it consistently. For example, assess if different users can get the same reading on the same patient. Then address any discrepancies through cognitive interviewing. Update training programs until we achieve a predetermined level of consistency across users. A good goal, for example, would be to reach 80 percent agreement of each skin tone rated.

Number two. MST is typically performed by looking at skin pigment on a patient's forehead, but pulse oximeters are most commonly applied on the finger. For most people, the skin pigment differs between the face and the hands. So we recommend identifying MST at the pulses ox sensor location, which is consistent with the pigmentation assessment recommendation in the FDA proposal for quantification of skin tone.

Number three, for calculating ITA. We recommend choosing colorimetry devices validated for reliable and consistent use across a diverse group of users. Some device vendors may have completed testing that shows they can provide consistent readings, but without the validation data. We suggest conducting a pre-study validation trial to assess if one, individuals can use colorimeters consistently, such as at the same angle or the same pressure against the skin, and two, if different users get the same readings on a single patient.

In general thought, this level of rigor in clinical trials is a great start, but it should not stop when the device is cleared. We recommend integrating post market performance reviews to track incidents with the devices and to capture performance disparities in previously cleared devices.

Vendors should continue to submit data about device performance among patients with different skin tones to identify and reduce care disparities.

Finally, I'd like to close with one final thought. It wasn't long ago, in the height of the pandemic, before COVID tests were readily accessible, when these pulse ox devices were critical in determining when a patient was in grave danger, and device performance issues led directly to patient harm. So we must keep in the forefront of our minds that this is a significant health equity issue. The color of a patient's skin should never degrade the quality or the effectiveness of tools that healthcare providers use to give life saving care.

On behalf of ECRI, thank you again for prioritizing this important patient safety initiative.

Dr. Wickerson: Hello, my name is Grace Wickerson and I am the Health Equity Policy Manager at the Federation of American Scientists. I have no conflicts of interest to disclose today.

Thank you so much for this opportunity to speak to the panel on this topic of pulse oximetry and how to handle skin pigmentation in the testing of pulse oximeters. I am excited to see that this conversation is returned to the advisory committee and is under deep consideration by the food and drug administration. The Federation of American scientists has been focused on this issue since January of 2022. And we have been focusing on a continuum of policies that are necessary to integrate health equity into the design, development, and regulation and monitoring of medical products. Whether they be drugs, devices, or algorithms. And our focus on pulse oximetry is because it is a perfect case study for what happens when we don't consider a diverse range of participants of racial and ethnic backgrounds, gender, and other key attributes in our design of clinical studies. Thus, creating products that can have massive repercussions on health outcomes once they're used in the real world.

And I'm excited that the recommendations that we proposed in our public comment, as well as an event that we held the day after November 1st, 2022 advisory committee meeting are under consideration in this draft guidance including a focus on a broader way to analyze skin pigmentation and actually putting into place a way to do so that is both it recruits a diverse range of participants and also introduces an objective metric to guide the design of pulse oximeters. I think this is an, a critical advancement upon the current guidance, including the way that this guidance also increases the number of people that should be included within this clinical study. I think it's important to note that these devices should still be considered to work well in all populations and not just work good enough in marginalized populations such as black and brown Americans.

That being said, it's important to consider whether these thresholds that are being introduced to prevent bias are actually evidence based and ensure that we don't just sort of allow the existing problem to persist. It was important to understand where the 1.5 percent and 3.5 percent thresholds being considered for non-disparate performance are coming from and to consider whether a tighter range would be necessary to actually prevent the clinical negative outcomes that we're seeing. Some recommendations would be to consider the clinically important range for that is where hypoxemia is known to start between 85 and 94 SPO2, and ensure that within that working range, we know these devices can work, within a much because it's an example of a lesser degree of bias than what is proposed. And it's still unsafe for real world outcomes.

I think another important point to consider, and I think something that requires partnership and collaboration with other federal agencies, hospitals, and other key entities, is to actually review these devices once they're out on the market. So, this pulse oximeter question has been open since the 90s and just got its attention because of a concerted effort of academics,

healthcare systems, and other partners. That doesn't need to be the case. With as the FDA expands its use of real-world evidence, we can use real world evidence to detect how these devices are working out in the clinical environment. Part of this has been, as the Federation of American Scientists has explored, is that the Veterans' health administration procures plenty of devices that have been approved by the FDA. There's a potential collaboration there to use that line of communication. Skin pigmentation to define when a problem arises with a medical device.

In closing, I think it's important to think about the skin pigmentation standard being scalable to everywhere across the FDA that uses skin pigmentation to assess a medical device. As more and more devices use light, we are requiring standards that allow us to assess the ways that skin pigmentation might interfere with measurements and thus introduce more health inequities, and we need a systematic approach. Thank you so much.

Dr. Fawzy: Good afternoon. I'm Ash Fawzy from Johns Hopkins University, here to speak about the clinical consequences of racial bias on pulse oximetry. I share strategies for gathering high quality clinical data to evaluate pulse oximeter accuracy. I have no financial interests or professional relationships to disclose.

During the height of the COVID-19 pandemic, several clinicians, myself included, began noticing inaccuracies in pulse oximeters. Indeed, retrospective clinical studies have shown consistent overestimation of true oxygen saturation by pulse oximeters. In two studies of COVID-19 patients, pulse oximeters on average overestimated oxygen saturation by 0.93 to 1.2 percent among black compared to white individuals. Despite this seemingly small error, based on our modeling, black patients accounted for over half of the patients whose need for COVID-19 treatment was never recognized. And among the patients who eventually had their eligibility for COVID-19 treatment recognized, black patients were 29 percent less likely to receive timely

therapy. and Hispanic patients 23 percent less likely because of systematic overestimation of oxygen saturation by pulse oximeters.

In the larger study that looked at actual delivery of COVID-19 therapy, patients with an unrecognized need for COVID-19 therapy due to overestimation of oxygen saturation by pulse oximeters were 10 percent less likely to receive COVID-19 therapy regardless of race and their odds of hospital readmission was 2.4 times higher. We then set out to investigate. whether differences in pulse oximeter performance among Black and Hispanic patients could be due to skin pigmentation by conducting a prospective study that looked at 350 paired arterial blood gas samples and pulse oximeter readings from 12 Johns Hopkins Hospital intensive care unit patients. Skin tone was measured objectively using a colorimeter and five patients were darkly pigmented with a forehead ITA of negative 30 degrees or less. For this study, we exclusively used arterial blood gases obtained as part of routine clinical care. We identified the exact time of blood draw by manually reviewing the arterial line tracing and matching it with pulse oximeter readings from a high frequency vital sign data repository. By leveraging accurate clinical data rather than relying on research blood samples, we are more likely to capture occult pulse oximeter errors not evident to clinicians or have a more accurate representation of real world pulse oximeter performance and reduce the burden to both participants and research staff.

After controlling for the effect of oxygen saturation level, pH, heart rate, and mean arterial pressure, the pulse oximeter, on average, significantly overestimated the true oxygen saturation. For patients who had dark, brown, and tan skin on the forehead or forearm. This overestimation was also seen among patients with brown and tan finger pads among darkly pigmented patients. The median overestimation error was 2.5 percent at saturations of 90 to 94 percent, and 10 percent of saturations below 90 percent, whereas median error was near zero for lighter pigmented participants. Regardless of saturation level. Overall, the average root mean

square error was 2.75 percent. However, device performance was significantly poorer in darkly pigmented patients with an average root mean square error of 4.15 percent, a value that would not meet the current FDA standard for device clearance and is far above the proposed thresholds.

In summary, in our prospective study, an FDA cleared pulse oximeter performed poorly and darkly pigmented patients had been tested on a group of diverse, critically ill patients. Rather than relying on a less diverse sample of healthy volunteers, this device would not have met criteria for regulatory clearance. As we have shown, there are real clinical consequences to pulse oximeter error that disproportionately impacts patients of color and exacerbates healthcare disparities. As clinicians, my colleagues and I always aspire to do what's best for our patients over the past 30 years, we have fallen short because the tools at our disposal have failed us. We call on this panel and the manufacturers of pulse oximeters to give us better tools so that we can provide excellent equitable care to all our patients.

Additionally, we urge the panel and FDA to require pulse oximeters be tested on a diverse population using objective measurements of skin tone and actual clinical data in real world settings to ensure adequate and equitable performance when making important clinical decisions. I would like to thank my collaborators and the patients who contributed to our studies. Thank you for your attention.

Questions from Panel

Dr. Cassiere: This is Dr. Cassiere panel chair. Does anyone on the panel have any questions for any of the open public hearing speakers? Ms. Brummert.

Ms. Brummert: If the person from Medtronic is still here, I'm wondering, you admit that there's a lot of faulty data coming out. And I'm just wondering if you've considered a voluntary recall on these devices.

Dr. Ajizian: Yeah, thank you for the question. We have a very robust quality and safety process that is continuous over the entire life cycle of our products. So we're continually looking at signals and putting them through our process to see if field corrective actions need to be done or communication with physicians or other providers and that can take many forms. The issues that we were talking about today around pigmentation and pulse oximetry have not triggered any of those levels within Medtronic.

In other words, despite continually evaluating this and putting this through rigorous analysis to examine patient risk. We believe with 100 percent certainty that our devices conform to current FDA standards. We've seen objective data presented by some of the other speakers here today that support that. And currently, we think that if there was a recall by any manufacturer around this, we would undermine public safety, because this is a foundational device in operating rooms and ICUs and ERs and ambulances and everywhere. And, you know, one of the things you've heard me say is we've really doubled down on education here. To help users understand what these issues are around pigmentation. So when you're faced with a patient in respiratory distress in an emergency room or another environment who is darkly pigmented, we hope that we can arm you with some education to help you understand the strengths and limitations of that pulse oximeter.

What we really don't encourage is for users to use the pulse oximeter of any brand, actually, speaking more as a doctor now, as a threshold for doing something medically or not doing it. That's not the way medicine is practiced. It should be looked at as one number in the context of the entire patient and a complete exhaustive medical evaluation. So, it has not triggered our quality process. We continue to evaluate and we do feel that the benefits of the products in the field far outweigh the risks that we're talking about today.

Dr. Cassiere: Dr. Lewis.

Dr. Lewis: Yes. My question is also for the representative from Medtronic. Thank you for your presentation. You mentioned that the company has invested in innovation in this space. And I was hoping that you could give the panel one or two examples of innovations in technology that you guys have been working on.

Dr. Ajizian: Sure. Obviously, I can't talk about proprietary pipeline things, Dr. Lewis. I'm sure you understand that. But if you just look back at our behavior since the Schrodinger article came out. While I'll get to technological innovation in a second, we've reviewed and shared publicly our own internal data that was recalculated completely around performance in light and dark skin healthy subjects and validation trials. We've also looked at the objective data that was presented by the Bickler lab, for example. And we fully believe that our pulse oximeters are functioning as specified and well within specifications to enable safe patient care. As far as innovation goes we have a new neonatal sensor on the market for the last year that's got a variety of technical features and algorithm enhancements that, that actually help see across many, many confounding variables in the field. We also continually revise our algorithms around our digi-cal digital AI based neural network algorithm that is always being refined. And before the pandemic started, we have been working on our oximeters that, of course, will help improve performance in what we call corner cases, which includes deep skin pigmentation, low perfusion, thick skin and other confounders. It's a continuous improvement process. It has been for 30 years in no court and it will continue to be so.

Dr. Cassiere: Before I go on to talk to him, I just want to kind of step off to what Ms. Brummert had mentioned. I understand your answer, but it almost sounded like you were blaming the medical community for using pulse oximetry as a treatment threshold, during Covid, that was my interpretation of your answer. That we should have not used pulse oximetry to either treat with remdesivir or dexamethasone, and that it's not the fault of the device, which has since the 1990s

shown to have issues with occult hypoxemia, because of skin pigmentation. Could you, could you clarify that for me, please?

Dr. Ajizian: Yeah, Hugh, you bring up a great point. The pandemic really pushed this technology into the forefront in many ways, including triage of people in overwhelmed emergency departments, wards, ICUs, and even home, where we had no resources and places to put our patients. And as a result, we had to develop protocols quickly that were based in well-meaning practice and trying to optimize outcomes for as many patients as possible. I think we were all trained as physicians to realize that, that even a non-invasive blood pressure, even an ECG, even a pulse oximeter value should never be used in isolation. It should be used as part of a complete. exam and a complete approach to the patient. The pandemic pointed out that there are weaknesses here around skin pigmentation in these devices and that's why we're all here today, but in no way am I trying to say that the medical community is at fault here. We were all doing the best we possibly could with an overwhelming situation that even drove many of us to use over the counter pulse ox, and have patients obtain these things doing the best we can, well intended care or a mass casualty type triage situation.

Dr. Cassiere: Great, thank you. Dr. Goldman.

Dr. Goldman: Yes, thank you. Julian Goldman here. My question is for Dr. Fawzy. First of all, thank you very much for the presentation and for your excellent work, in this area. you know, looking at the slides you showed and aware also of your publications. I'm wondering if you could provide any additional insight into what have may caused, what appear to be outlier readings as opposed to, you know, a homogenous bias, which in a sense might be easier to address. But it looks like there really are outliers and what did you learn about that?

Dr. Fawzy: Yeah. thank you so much Dr. Goldman for that question. That's a very astute pickup and, you know, there has been publications about this. I think the most important one was by Dr.

Valbuena, and her colleagues, she's from the University of Michigan, who showed that even in the same patient, the pulse oximeter error is different throughout the day. So even if the pulse oximeter error in a black patient or a darkly pigmented patient, is present at one point, it might not be present in the same patient at a different time. So there are multiple things happening here and several different things that are probably interacting together. We tried in our study, although the sample size was low and it was a pilot study even though we did meet that FDA threshold of 10 patients and, 200 samples. We did try to use some of the clinical characteristics that we have like pH and mean arterial pressure to try to control for some of these things. But there are certainly several things that are interacting here that need more investigation to understand what they are so that we can really get a good understanding of what needs to be changed about the technology of the pulse oximeter to make it more accurate.

Dr. Goldman: Just for clarification, I don't recall, did you say you also recorded signal strength? Or perfusion, and a measure of that with the pulse ox signal.

Dr. Fawzy: So, we did a visual assessment, you can see that in the preprint that we put up. We did a visual assessment of how good the tracing was for the pulse oximeter and then did a sub analysis of that and it didn't really show much difference in terms of the results, but that's also a great point. We don't have access to the raw data coming from the monitor or from the pulse oximeter and if the pulse oximeter manufacturers can give us access to that, we can certainly do much better research, in the clinical setting, with that information and use that to understand the bias better.

Dr. Goldman: Thank you for that.

Dr. Cassiere: Dr. Gooden.

Dr. Gooden: Yes. Hi, this is Cheryl Gooden. my question is for Dr. Ajizian from Nellcor. We've had discussions about location of the probe, and I think this is probably reflective more so of the

adult patient population in pediatrics, and I'm sure many may be aware of this that or not that in pediatrics many times the location of the probe may be on the toe or even the foot. And I noticed in one of your slides that it was on the foot. Do we have data algorithms that are used by Nellcor, and perhaps other companies? And I guess my question related to that is how does that impact pigmentation when we're looking at that in our discussions?

Dr. Ajizian: Yeah. Great question, Dr. Gooden. And first of all, as a pediatric provider myself, I'd like to commend FDA for the significant neonatal and pediatric presence on this committee. It's wonderful to see our patients who are often, on the fringe of recognition, have such a spotlight on them. So, to answer your question, yes, we obviously validate for all positions that are approved for in our instructions for use. That said, at the bedside, we often, we, not Medtronic, we providers will often move things around and try to find other sites. You're perfectly fine using that on the neonatal foot, with our products, but it's imperative that we as providers understand the differences in capability and specifications from one manufacturer to another. And there are often significant differences, not only in performance, but positioning as well.

As far as the adult population goes, I will say that we frequently see opportunities to improve on positioning, both how it's put on and where, large at a high level. And so again, I come back to, as a very, very important part of this issue, not only making devices that work better across skin pigmentation spectrums, but making education easily available and continuous for users to not have drift in their practice, have a place they can go to get gold standard information for manufacturers like ours, and other major manufacturers. So, a great question, but in general, you're absolutely fine. I think as things go forward, we look forward to being able to work with this group and other members, of ISO, for example, and the researchers to try to make sure performance continues to improve every time we bring a new iteration out, whether it's a new site or a new probe or algorithm.

Dr. Cassiere: Great. Any other, any other questions from the panelists to our speakers? Okay. I, now pronounce the open public hearing, to be officially closed, which brings us into our invited speakers portion. At this time, we will continue with our guest speaker presentations, each speaker will be granted 10 minutes to speak. The first speaker is going to be Dr. Joseph Wright. Dr. Wright, you may begin.

Invited Speaker: Dr. Joseph Wright

Dr. Wright: Good afternoon and thank you for the opportunity to speak with you, at this meeting. My name is Dr. Joseph Wright. I am the Chief Health Equity Officer and Senior Vice President for Equity Initiatives at the American Academy of Pediatrics. I have no financial conflicts to disclose, and in terms of professional context, I am a pediatric emergency physician who has used pulse oximetry as a ubiquitous clinical practice tool throughout my career.

The American Academy of Pediatrics is a professional organization of 67,000 members representing primary care pediatricians, pediatric medical specialists, and pediatric surgical specialists. The AAP has published policy on the issue of pediatric medical devices. The policy statement discusses how off label or physician-directed use of medical devices is often necessary and appropriate because medical devices are often not designed and or tested with the unique needs of children in mind. The statement is clear that pediatric-specific device development is the goal. Unfortunately, children are too often an afterthought in terms of device development.

AAP also championed the Pediatric Medical Device Safety and Improvement Act, which resulted in new incentives for pediatric device development in rare diseases and created an important grant program to stimulate pediatric device development.

Unfortunately, however, we have not made as much progress in advancing pediatric device development as we have, for instance, in advancing pediatric drug development. The Academy is also on a concerted journey rooted in policy to ensure equitable care for all children

based on identifying and mitigating inequities that are embedded in clinical practice that can potentially result in or exacerbate outcome disparities.

So the question is, why study children? You've heard earlier today, and I think the committee certainly appreciates that children are not little adults. There are certainly unique anatomical, physiologic in growth and development patterns that are important to understand, vis a vis this work. We cannot extrapolate directly from adult experience. And there are also discovery opportunities in terms of this work as well. Being able to understand performance in low oxygenation pathologies, like children with cyanotic congenital heart disease, for instance, and also being able to understand performance in special population cohorts unique to pediatrics, such as neonates.

So, while there has been incremental progress, and this particular study, which was published last spring, was a cath lab study that did align SaO2 and SpO2 and showed that consistent with other studies, pulse oximetry overestimated oxygen saturation in children of Black or African American race.

However, these are just, pardon the pun, baby steps and necessary foundation for the work to be done. And that said, while the FDA does not have the authority to require that medical devices be studied in children, FDA can encourage companies to equivalently study their devices in children as they do in adults. So, that we don't leave children behind. The AAP appreciates the steps that have been taken by FDA through the Centers of Excellence in Regulatory Science program to include funding for pulse oximetry research in children. And we certainly heard earlier in the meeting the progress that's been made by Dr. Christopher Altman and his investigators at Stanford and, and now at other sites across the country to investigate the use of pulse oximeters in, in children. However, we also heard the challenges of getting adequate enrollment of enough patients. I believe Dr. Allman mentioned that, the initial N of 228 needed

to get to 312 in order to enroll adequate numbers of patients at the von Luschan scale category 4 of the, the darkest melanated individuals. So, the Academy encourages FDA to fully support completion of studies that are currently underway to equitably address unique needs of children, of all melanation, as innovative technologies and devices are developed that mitigate current inequities.

So with that, again, I thank you for the opportunity to speak with you this morning and this afternoon and truly appreciate what the FDA through its ESRI program has done to support this work and encourage the FDA to continue that support. Thank you very much.

Invited Speaker: Collaborative Community

Dr. Lipnick: Thank you for the opportunity to speak today about the open oximetry collaborative community. I'm Michael Lipnick, anesthesiologist and intensivist at the University of California, San Francisco Center for Health Equity in Surgery and Anesthesia. I have no financial disclosures or conflicts of interest of note, while manufacturers do participate in the collaborative community that I'll be talking about today, I want to highlight that none of them provide financial support for this initiative. And it's probably been made clear by other speakers today, as well as numerous publications and prior meetings over the past couple of years, pulse oximetry is a significant global health issue.

For many regions, access to oximeters is a big challenge, followed by issues of performance. The issue of worst device performance in people with darker skin pigment is a significant global health concern with evolving data. These data are not always easy to access, or interpret, or translate into clinical practice, and given recent links between pulse oximeter performance and health and healthcare disparities, these issues require urgent attention. This has largely been the impetus for the OpenOximetry project. The project has multiple components, including a prospective real world clinical trial supported by the FDA's CERSI, as well as

multiple studies in the hypoxia lab here at UCSF. But one key component of the project, which I'll be talking about today, is the creation of a collaborative community. The collaborative community is structured as a perpetual working group with several objectives. And the initial focus of the collaborative community is on how to, how to improve performance of pulse oximeters in patients with darker skin pigment. But the community is trying to do, quite a bit more than that.

Some of the objectives include identifying the challenges that are going to be, that must be addressed to improve equitable performance of pulse oximeters, improving best practices and standards for both research and regulatory purposes. This is going to require a lot of data sharing and translation and communication of these research findings into practice changes. This also is going to require a lot of advocacy, working very closely with regulatory bodies. And in order to achieve any of these goals, we're going to need a diverse group of global stakeholders and experts.

The community has grown organically over a series of stakeholder meetings over the past couple of years. These meetings have included two in person workshops and hundreds of participants. There are over 150 members from numerous disciplines who are currently collaborators in the community. These include experts from industry, non profits, health, engineering, anthropology, and more from more than 18 countries. Here are just some of the organizational members represented and a full list can be found on the website, which continues to be updated with more who are joining.

Based on the challenges and priorities identified to date, a handful of subgroups have formed around these four areas. I'll briefly go through each.

The first, the clinical trials group, brings together researchers from multiple labs and institutions where pulse oximeter performance testing are ongoing or launching. This involves

about eight different clinical studies from multiple countries. The group has worked together to provide input on multiple protocols for, data collection across study sites, trying to harmonize data collection and develop best practices where previously few existed.

One focus of this group and these studies is trying to understand why devices in the lab may perform relatively well, but in the real world, they perform quite differently in this figure of published data, comparing data from the laboratory setting on the left and the clinical setting on the right, you can see considerably more performance problems, data outliers in the clinical setting on the right. We don't understand why, and we need to, and we need more data from clinical trials in the lab to figure this out. Harmonization and expansion of these data sets hopefully will provide the data that are needed to answer these questions, and hopefully that will be available for the first time in the coming months.

The next group, the education and communication group, is just forming, with its first formal meeting planned for later this month. Based on feedback from community members, this group will try to develop communication and educational material to support clinicians and patients in the correct use of pulse oximeters, maximizing performance accuracy based on currently available data with existing devices and known existing limitations. The plan would be to update this as data evolve and as new devices hit the market. So the information generated will likely take the place of an open access living infographic that can be rebranded and tailored to various local contexts and again, updated as the data evolve.

Lastly, all of this needs to be easy to understand. For example, terms like individual typology angle, root mean square error, non disparate bias. These are not easy to understand. Yet, understanding these terms is essential for ensuring that pulse oximeters are not used in a manner that causes health and healthcare disparities. Better education and communication materials could possibly go a long way to addressing these challenges in the near term.

The third subgroup, the Skin Color Diversity Group, has been working to clarify best practices for ensuring diversity of skin color and pulse oximeter studies. This has included sharing and reviewing data and identifying research priorities. Some of the members of this group are also working on creating an online toolkit to disseminate curated and novel resources that include best practice recommendations, key terminology, frequently asked questions on this topic to make it easier for others to find, easier for others who are working in this space to improve equity in their studies. Here's an example of some data that's been shared by this working group, with regulatory bodies already on how well data can be measured at different parts of the body, data for quantifying skin pigment, how pigment, correlates across different, anatomical sites and how reproducible these findings can be. For example, as shown here in the top two plots, the inner arm, forehead, dorsal finger may correlate reasonably well with each other depending on sun exposure of the population, but measurement at the back of the ear and the fingernail shown in the bottom plots, these are hard to reproduce and differ significantly from other parts of the body.

Data shared also suggests that subjective skin quantification scales alone can lead to underrepresentation of people with darker skin pigment. Here's a visual representation of cohorts of healthy volunteers used to test one pulse oximeter with current regulatory frameworks. Each bar here represents the percent of subjects in this cohort with light, medium, or dark skin pigment as color coded here. And as you can see, if you use a subjective scale like the perceived Fitzpatrick, the cohort with dark skin pigment on the far right exceeds the green 15 percent threshold of current regulatory guidance. However, if we use the individual typology on objective measure to categorize the same group, we have fewer individuals who meet the criteria for having dark skin pigment. And this cohort does not meet the threshold of 15 percent required by existing regulatory. This is true when we've looked at cohorts used for testing many devices

and performance validation studies. You can see the majority are not meeting the indicated threshold, the intended threshold. when using subjective versus objective methods. So there's a lot that can be done in this area right now.

The group's also trying to better understand how subjective scales and objective scales relate. Also what categories of binning or cutoff should be used to accurately represent the world's population. One upcoming study by collaborative community members in East Africa and hopefully South Asia will soon be able to help shed light on some of these questions with data never before available.

The last collaborative community subgroup has been focusing on ways to leverage data to accelerate progress data sharing. This is focused on creating common data models to standardize definitions and the structure of data, as well as the standardization of data collection protocols, as I've already mentioned. The group is trying to leverage data networks, create new repositories at scale, including repositories that have raw pulse oximeter signal, something that has been not widely available. In addition to the new data sources, the group is trying to standardize and share data analytic techniques. And the goal ultimately is that these larger data sets hopefully will allow further characterization of the root causes of these performance problems we're talking about and facilitate investigations that otherwise would not be possible using smaller disparate data sources. The first versions of these data sets concurrent can be accessed from the open oximetry website, and we have some significant updates planned in the coming weeks and months. So please continue to check back.

To wrap up, based on the discussions today. There have been some key themes and lessons learned that have emerged and are shaping future directions for the collaborative community.

The first is that challenges differ by setting. In some places, access is still the leading concern when it comes to pulse oximeters, though device performance is also a significant concern. As work is done to improve equitable device performance, it must be done in a manner that does not reduce access in such a way that patient safety and health disparities are made worse. Balancing access and performance is key.

Second, while several key challenges can be made now to improve diversity in medical device studies, data, and best practices for ensuring diversity of skin pigment are going to continue to evolve in the near future and are certainly going to remain an area of active exploration in the community.

And finally, more data from more diverse settings are needed. This is going to require new strategies for data collection and data sharing. It's exciting to see many new studies emerging and joining the collaborative community. It also will be very exciting to see studies from low and middle income countries, which can not only improve applicability of findings of all studies and broader representation in study populations, but can also expand global research and development capacity that are needed to address many of the challenges facing medical device manufacturers currently.

With that, I'll conclude and a big thanks to the open oximetry project core team shown here and to our collaborators, supporters, and community members. Thanks very much for your time.

Invited Speaker: Professional Societies' Perspectives on Pulse Oximetry

Dr. Gurubhagavatula: I am Indira Gurubhagavatula. I'm representing the American Academy of Sleep Medicine. I'm a professor of medicine at Penn. I have no conflicts to disclose.

Recent data suggests that oximetry is less accurate in those with darker skin pigments who experience more occult hypoxemia and more variability in readings that are checked

repeatedly over time. These studies were done in critically ill or hospitalized patients. We don't know if similar disparities exist in sleep center patients. We see a lot of sleep apnea, a condition with repeated brief airway closures that occur when a person falls asleep. If the airway closes completely, it's called an apnea and it gets scored. Partial airway closure causes partial drops in airflow called hypopneas, which are only scored if the desaturation that follows is big enough.

So, oximetry is a core signal for us. Our entire scoring system depends on it. We don't do studies in the ICU. We study patients in outpatient labs and in homes, and these portable studies can be done in acutely ill, hospitalized patients as well. Our devices have a high enough sampling rate to capture these desaturations. Here, if there are four desaturations and sampling is infrequent, you would only pick up two of them. To score a hypopnea, the size of the desaturation matters. Our clinical standard is at least 3 percent magnitude, but Medicare and other payers require at least a 4 percent desaturation to happen for the event to be counted. So the ability to detect drops in saturation to within that 1 percent margin can be the difference between whether a diagnosis is made or a case gets missed. We need high sensitivity and low variance when the same person is tested repeatedly over the course of the night.

In this home sleep study, there are oscillations in snoring volume, airflow, and oximetry. In this example, only the first two hypopneas can be scored because they are the only ones associated with an at least 4 percent desaturation. The others cannot be scored. The questions then are, were these smaller desaturations underestimates, and are these underestimates more likely with darker skin color? So approved devices need to be able to pick up desaturations, to within a 1 percent error within subject variance.

Under the current FDA standard for calibration, only 10 healthy subjects need to be tested. Our patients are often unhealthy. BMIs can exceed 70. Our patients have low perfusion states with heart failure and atrial fibrillation being common. We see patients who have smoking,

advanced lung disease, neuromuscular disease. Our black patients often have more severe symptoms and have worse outcomes, so including more of them in calibration studies is critical, rather than just the larger of 2 or 15 percent of the calibration sample. The typical approved device has arms within 3 percent of blood gas values, which translates to a 95 percent confidence interval of up to 6 percent around a point estimate.

Sleep studies, however, need greater accuracy within 1 percent to avoid missing cases with low variance during repeated measures. More and more patients and our researchers are using wearables. Flawed data there can propagate exponentially and find itself into big data sets. Our in lab studies include transcutaneous CO2 monitoring, which relies on a similar technology and may also be prone to some of the same color biases. If sleep apnea is missed, the consequences can be severe. Daytime sleepiness, higher vehicular crash risk, cognitive, emotional, physical, and occupational effects are all possible. Long term health risks are cardiovascular disease, Alzheimer's, dementia, and even death. Undiagnosed sleep apnea imposes a large annual financial burden in the U.S. There's also public health risk in the form of severe or fatal crashes, especially if commercial vehicles are involved. Our recommendations are to have calibration studies that reflect sleep center patients and are more inclusive across the spectrum of skin color, race, gender, and health conditions. To have a diverse pool of scientists doing these studies and to power the studies to pick up small desaturations with low intra subject variance over time. The rigor of the approval standards has to reflect these diagnostic needs. We need post market surveillance of devices. already released and labeling needs to include relevant metrics for us like the size of bias and within subject variance.

And finally, we need to educate all stakeholders including clinicians, patients, payers, and researchers. We commend the FDA for taking on this issue, which is of course significant to our field. Thank you for this opportunity.

Dr. Baugh: Hello, my name is Aaron Dorian Baugh. I'm currently assistant professor of medicine at the University of California, San Francisco. I want to thank the Food and Drug Administration for the opportunity to comment on this important issue and wish you all a happy Martin Luther King Day. Now turning to the issue at hand, I want to stress that I strongly support the two-tiered approach of pursuing both the Monk Skin Tone scale and the individual typology angle as mechanisms to interrogate this question.

In thinking about how we could further improve, I want to highlight two special issues. To understand the first, please look with me at these Bland Altman plots of agreement between oxygen saturation as measured by pulse oximetry and arterial blood gas as reported by Christopher Chesley and colleagues. Now, the first point I want to make is that this correlation is dynamic. Even within the same individual within this study, you had the relationship changing over time and condition. And this is important because, what we see as opposed to the original qualifying studies for, pulse oximetry is that the Tom Bailey study at University of Michigan that really brought this into debate and contention was done in ICU patients. And so, to get the right answer, we should also try to replicate some of these experimentally, imperfect conditions that might arise in an ICU, hands that are not warmed, other simulations of hypoperfusion or poor flow. So that's the first.

The second, let's look a little bit and think about the shape of this Bland Altman plot, where the x axis is the mean agreement, or in this case perfect agreement. Things above it are overestimation, and below it are underestimation. This is important because, based on this hypothesis that different skin tones will cause differential light absorption, we would expect that in dark-skinned individuals, or those with dark-skinned pigment, you would only have overestimation of values by false oximetry.

However, what we see in this study from Dr. Chesley is that in all races, black, white, and Hispanic, that he reports in this graph, there is bilateral error, over and underestimation, which suggests that there could be some different mechanism at play, or at least one additional one. Now, the mechanism that came up the most in the media is race and specifically, people that identify as Black race. And you see those in these major headlines here. Now, for many people that read or heard about this, the nuance about skin pigment, could have been or was lost. Indeed, I say that even in many of my colleagues in other specialties, they have an insufficient understanding of pulse oximetry to appreciate this skin pigmentation hypothesis.

So we have an interesting emergent situation where the best hypothesized, explanation scientifically of what's happening diverges from the public's understanding of what's going on. And especially with this misidentification around one racial ethnic group. and in such a circumstance, I would argue it's our obligation to resolve, both of those questions simultaneously. For example, what could happen when we don't consider the negative case of COVID-19 vaccine safety where quite a gulf emerged between the scientifically robust evidence demonstrating it was appropriate for use and that failure to assuage all public anxieties about the newness of the product.

In contrast, consider the work of A. Philip Randolph. Through his March on Washington movement, he created public pressure to address segregation in the defense industry. Now, although the Roosevelt administration began to reach out privately to the responsible corporations, Randolph did not relent until an executive order was promulgated. He correctly recognized that such a public action would send reassurance to blacks and warning racists that our nation would pursue its aspiration to equity found in its establishing documents. Similarly, the FDA can, through the structure of this trial, send a message about our commitment to equity.

Now, whether this is done through having a minimum number of Black identifying patients in every stratum or having an entire trial replicated only among black identifying, participants. I think it's going to be essential that we answer at least the racial question with regards to black identifying persons about do pulse oximeters work safely for them or not, and trying to include as many races as possible.

In conclusion, I'm glad to live in a republic that considers the needs and concerns of all people. And I remind you that that word republic, *res publica*, the common or public things, exhorts us to answer questions in a way that not only satisfies those who have advanced scientific training, but is also comprehensible and satisfying to all those who may be affected. Thank you so much.

Dr. Rizzo: Good afternoon. My name is Ann Rizzo and I am a triply boarded surgeon that is representing the American College of Surgeons. I am here to talk today about the FDA's proposal on pulse oximetry. If nothing more, the COVID-19 epidemic showed us that we did indeed have some inaccuracy in the technology that we were using to measure pulse oximetry in patients with darker skin. We had a study that showed that patients of darker skin had lower saturations than other patients. But what this study did not show, is that these patients were treated any differently, nor did not get the treatment that they indeed needed. We know from our studies that there are inaccuracies in this technology, not just for dark skin, but for several other things, such as blood dyscrasias, tattoos, nail polish, poor blood flow, or modeling, and things like thicker skin, tobacco usage, and the temperatures of the patients. What most physicians know who treat patients with low blood saturation is that there could be inaccuracies, and so most physicians that are treating patients critical or not always correlate the inaccuracies of the pulse oximeter with a blood reading of oxygen, which is a saturation of oxygen taken from a blood test.

The question if a study should be done to look at these correlations based on skin color and diffusion of the spectroscopy of pulse oximetry is yes, but frankly, the attention, research, and development should also be put into pulse oximetry to use new technologies that will be more accurate and thus not need a correlation. Something such as near infrared spectroscopy technology, which is not cognizant whatsoever of the depth of penetration nor of the coloration of the skin. Thus, using the finger, we know, is much less accurate than using an earlobe for measurement. We should be aiming some of our treatments at looking for the new technology that will give us the most accurate reading of blood oxygenation versus just the pulse oximetry, which is a noninvasive method known to have flaws. This method has been used by many non medical devices that were widely used during COVID that were non-correlated and non-tested. So doing a study to look at skin tone and its effect is very useful, but I would ask that the FDA also fund additional research and development of newer and better technologies to help our non invasive methodology treat patients to the best of our ability and also help us providers from having to order additional testing to look for correlation. So we would support the testing to look at skin tones and its effects, but we would also support testing, research, and development of new technologies to improve pulse oximetry across the board. Thank you very much.

Dr. Ehrenfeld: On behalf of the physician and medical student members of the American Medical Association, I appreciate the opportunity to speak to you today.

Pulse oximeters are painless, non invasive medical devices used as a part of the clinical standard of care. As we know, pulse oximeters are widely used for rapid estimation of blood oxygen saturation and pulse rate. Depending on the reading, some patients may require immediate medical attention. Unfortunately, the accuracy of pulse oximeters for some patients has come into question. The devices have been found to be less accurate for patients with darker skin tones. Studies have shown that pulse oximeters are three times more likely to provide

misleading readings for patients with darker skin pigmentations, leading to missed critical diagnoses of low blood oxygen levels.

During the COVID-19 pandemic, patients with inaccurate pulse oximetry readings were less likely to get extra oxygen, 10 percent more likely to experience delays in receiving advanced treatment, and more likely to be readmitted to the hospital. Additional studies have identified disparities with pulse oximetry readings throughout the pandemic for patients of color, worsening health gaps that previously existed.

As a result, our AMA has adopted policy recognizing this issue. The AMA urges the FDA to ensure such devices provide accurate and reliable readings for patients with diverse skin pigmentation. The AMA Also, calls for healthcare personnel and the public to be educated on the limitations of pulse oximeter technology so they can account for measurement error. Our AMA acknowledges the FDA's attention to this serious matter that is disproportionately impacting certain racial and ethnic groups. We applaud the FDA's efforts to convene key stakeholders to address these public health concerns. The development of new devices that ensure accuracy across all skin tones is imperative. The challenge for patients and physicians are not just technical, but evidence of a system deeply rooted in bias and racism. We should have known from other industries that bias is built into our light sensing technology with examples going back almost 100 years in Kodak film and persisting in digital photography. Hewlett Packard recognized the calibration issues with pulse oximeters decades ago. Equitable pulse oximeter sensors are possible as evidenced by research on crosstalk free or reflectance sensors.

Health equity means everyone receives the care they need, regardless of their skin color. Health equity is about acknowledging disparities and actively working to bridge the gaps. Physicians need to be confident that their tools, rather than perpetuating unequal care, can help them provide high quality care to every patient. Patients need to be reassured that they are not

being exposed to potential harm. By tools they or their physician are using. Today's meeting seeks to address these concerns. And as such, the AMA would like to offer a few recommendations for consideration. At a regulatory level, the Food and Drug Administration should require quantitative data on device performance across a range of skin pigmentations in clinical studies, particularly devices with color sensing technology. To identify and mitigate the impact on historically minoritized communities. To ensure statistical comparisons of performance are possible in smaller study samples, quantitative data should include oversampling of people with darker skin tones. The devices should be studied across a broad range of skin tones in a reliable scale, such as the Monk Skin Tone Scale or Reflectance Colorimetry.

Self-identified rays only should be used to report on disparate impact rather than as a proxy variable for skin tones. Studies illuminating the varying medical treatment responses based on color of skin warrant the design of regulatory pulse oximetry accuracy standards. Pulse oximeters shown to reproduce racial bias, particularly those that are FDA approved, should require a warning label noting risks of inaccurate readings. At an educational level, healthcare providers must be made aware of limitations of pulse oximetry technology and trained to account for systematic measurement error when developing diagnosis and treatment plans. The in-full health principles detail the need for understanding and dismantling systemic racism and bias in health innovation.

At a coverage level, healthcare industry leaders must advocate for reconfiguring Medicare reimbursement guidelines on oxygen therapy at home, accounting for the racial biases encoded into current pulse oximeter readings. Pulse oximeters with evidence of comparable performance across the spectrum of skin tones should be prioritized on payer formularies.

This aims at addressing system level barriers. These recommendations are aligned with the In Full Health Initiative principles co developed by the AMA and our External Equity and Innovation Advisory Group, which aim to advance equitable health innovation. The limited ability of pulse oximeters to provide accurate results for patients with darker skin pigmentations is a critical issue demanding our attention.

As we strive for innovation, let's continue to advocate for equity, ensuring that the devices we rely on for vital health information do not contribute to disparities, but rather serve as tools for equitable and just healthcare. Thank you and thank you for your time and consideration.

Dr. Bhakta: Hi, my name is Nirav Bhakta. I'm a physician and on the faculty at the University of California in San Francisco, where I take care of patients in the intensive care units, outpatient pulmonary clinics, and the pulmonary function laboratory. I'm also the vice chair of the American Thoracic Society pulmonary function testing committee through which I've co-chaired a workshop and statement on the use of race in the interpretation of PFTs. I have no financial disclosures to report.

The ATS community, which includes clinicians and investigators in pulmonary, critical care, and sleep medicine, needs to be confident the medical devices they use make measurements accurately and consistently. We are encouraged to see the FDA working with manufacturers to rectify this urgent situation in a timely manner. I propose five points for you to consider.

Point number one, medical devices should be authorized based on evidence from high quality studies showing accurate and reliable readings. We request review of the proposed bias limits, which can be informed by statistical and safety considerations, especially within a range of saturations approximately 88 to 92 percent, where many important clinical decisions are made. The proposed 1.5 percent limit for average bias for saturations greater than 85 percent is close to or larger than what was found in two studies showing differential treatments and

outcomes across race. The data for one of those studies is shown here. This limit may need to be smaller and extended to lower saturation levels. There is also a need to set a standard for precision. Because beyond average difference, the variability is large, is known to differ by race, and will affect clinical outcomes. We suggest transparent reporting of bias and precision across the range of saturation levels tested.

Point number two, the evidence submitted to support authorization of medical devices should be based on data collected from large and diverse study populations representative of the patients we take care of. The sample size should be larger for monk skin tone scale categories with greater variability. The proposal to include at least 1 study participant per category will be insufficient to establish confidence within the medical community. The study design should also appreciate how pigmentation varies at different sites across the body, as well as on the emitting versus receiving sides of transmitted sensors. A study of 24 participants is insufficient to capture the diversity of racial and ethnic backgrounds, skin tones, and age ranges. Performance is also affected beyond skin tone alone, including by critical illness and associated variable perfusion. With a sample size of 24, the proposal to conduct mixed effect models with interactions is methodologically problematic.

Point number three, there should be greater transparency with respect to the evidence submitted to support authorization. We anticipate devices will account for skin pigmentation measured by the device at the point of care and sensor site. The algorithms should be based on rigorous designs and follow established transparent reporting of diagnostic models, such as tripod. We strongly discourage algorithms based on population average of people with varying levels of skin pigmentation as a substitute for skin tone specific adjustment, as this will reduce the accuracy for all individuals. In no circumstance should self reported race or ethnicity or operator perception of skin darkness be used to adjust measured values. Race and ethnicity

should, however, be collected with secondary analyses performed to determine residual bias between race categories after adjustment for skin tone. Such data will help identify other sources of error that may promote health disparities.

Point number four, there is a need for ongoing post marketing evaluation, including real world studies. The updated guidance should be accompanied with the ongoing opportunity to identify, fix, and get feedback about problems from the frontline clinicians and biomedical staff. This is imperative for patient safety. Real world data will help us understand performance of devices in cardiac and pulmonary disease across a range of disease severity. Expanding the existing mod platform may require collaboration with the NIH to help fund data collection.

Point number five, there is a need to regulate pulse oximeters sold directly to consumers. The FDA should use this market influence to provide information and improvements on consumer grade pulse oximeters. Many clinicians and patients rely on home based measurements to assess patients health and to dose supplemental oxygen.

In conclusion, we appreciate that setting a more rigorous standard will increase the time and effort required to bring medical devices to market. The FDA and other federal agencies such as the NIH can work together to support partnerships between academic institutions and industry to minimize barriers to conducting high quality research. We encourage the FDA to expand these standards to all 510(k) approvals, improve standards, increase diversity, and transparent reporting stand to improve all medical devices and increase clinicians and the public's trust in the medical system. As the largest body of pulmonologists and intensivists, the ATS is ideally positioned to provide timely guidance for the community, and we are committed to a workshop on this important topic, which will include participants from industry. We welcome FDA participation in the workshop.

I'm grateful to the committee for your consideration of these important matters, and I thank you for the opportunity to address you today.

Dr. Lane-Fall: This is a video commentary for the anesthesiology devices panel of the FDA Medical Devices Advisory Committee focused on pulse oximetry and skin pigmentation. This video commentary represents the views of the Anesthesia Patient Safety Foundation, the APSF. It was prepared by me, Megan Lane-Fall, Vice President of the APSF. I disclose the following conflicts or possible perceived conflicts relevant to the topic of this presentation. I serve as the co-investigator for a grant funded by the NIH that's focused on evaluating the relationship between skin color and pulse oximeter accuracy in children and I serve as a speaker for Medtronic on the topic of health equity and the pulse oximeter paradox.

The Anesthesia Patient Safety Foundation or the APSF appreciates the opportunity to provide commentary as the United States Food and Drug Administration considers steps to ensure the accurate and adequate performance of pulse oximeters used in clinical settings. The Anesthesia Patient Safety Foundation is a nonprofit corporation founded in 1985 to advance patient safety in the practice of anesthesia and perioperative care. The vision of the APSF is that no one shall be harmed by anesthesia care, which is aligned with the FDA's mission of protecting public health by ensuring the safety, efficacy, and security of medical devices like the pulse oximeter. The APSF is a coalition of perioperative professionals. The APSF family, as we like to call it, includes anesthesiologists, certified registered nurse anesthetists, anesthesiologist assistants, paranesthesia nurses, surgeons, and other proceduralists. The APSF works closely has worked closely with industry partners, academics, regulatory agencies, and others to advocate for safe patient care and to inform the development of safe practices, standards, guidelines, and devices.

Relevant to the FDA's interest in the pulse oximeter, the APSF took a keen interest in recent reports of pulse oximeter measurement bias related to dark skin pigmentation. Anesthesia professionals use the pulse oximeter every day in practice across the continuum of perioperative care, including during preoperative evaluation, during surgery and other procedures, during the recovery from surgery and other procedures, during interventional pain procedures, and in the intensive care unit.

Importantly, pulse oximeter values inform multiple clinical decisions made by the anesthesia professional, including the decision to proceed with surgery, the decision to provide supplemental oxygen, which oxygen modality and fraction of inspired oxygen are selected decisions to intubate or extubate a patient, the titration of positive end expiratory pressure, and decisions to admit to specific sites of care, such as the intensive care unit or to the hospital following ambulatory procedures. The pulse oximeter is a powerful tool that allows us to provide safer and more responsive care than we could using arterial blood gas sampling alone. For that reason, the APSF would like to see the FDA adjust approval standards for pulse oximeter devices to ensure accurate performance across a range of skin pigmentation values in clinically-relevant oxygen saturation ranges.

While the recent attention on pulse oximeters arose in race-based analyses, scientific evidence suggests that pulse oximeter biases related to skin pigmentation, a measurable parameter, race, while important, is a social category that is not reliably associated with skin pigment. Race should therefore not be used as a basis for subject selection in device testing studies. At the same time, approval standards should also test other conditions known to affect pulse oximeter accuracy such as perfusion.

We at the APSF appreciate the opportunity to share our perspective with the FDA and we appreciate the renewed attention on safe patient care in the United States. Thank you for the time to provide these remarks.

Dr. Burnett: Hello, my name is Garrett Burnett. I'm an anesthesiologist at the Icahn School of Medicine at Mount Sinai in New York City and I'll be presenting on behalf of the Society for Technology in Anesthesia.

Before we move on, I wanted to mention that I was recently awarded the Anesthesia Patient Safety Foundation Medtronic Research Award: the study listed here. This grant was provided by the APSF and Medtronic was not involved in the study design or any future foundation Medtronic Research Award. As part of my brief talk, I wanted to update the panel on emerging evidence as well as address some of the potential solutions that were proposed by the panel.

First study I wanted to mention was recently published in Anesthesia and Analgesia and was a laboratory based deoxygenation study using 146 volunteers with diverse skin color as measured by the Fitzpatrick scale. This study found that low peripheral perfusion combined with darker skin pigmentation led to significant errors and misdiagnosis of hypoxemia.

The next study is a preprint study, but it comes out of Johns Hopkins. And was a study looking at 12 critically ill patients in the ICU. These patients had objectively measured skin pigmentation as measured by their ITA or Individual Typology Angle and found that their pulse oximeters performed below FDA criteria for clearance in patients with darkly pigmented skin.

Finally, in an abstract presented at the Society for Technology and Anesthesia's annual meeting in January of 2024, found that melanin was a strong differential absorber of red and near infrared light, both of which are vital to pulse oximeter functionality.

In terms of our recommendations, it's important that we recognize that pulse oximeters will remain vital throughout the perioperative period to anesthesiologists. And we believe making steps to ensure that that premarket testing is equitable and pulse oximeters function appropriately for all patients is vital.

The panel's proposal to incorporate objective measures of skin pigmentation through the Monk Skin Tone Scale or through ITA measurement are a step in the right direction. Potential future studies could link the Monk Skin Tone Scale to ITA measurements in order to minimize the need for multiple objective measures of skin pigmentation during the pre-market testing. Regardless, incorporating subjects of all races and ethnicities using objective measures is vital to the pre-market testing of pulse oximeters. I believe the changes to the acceptable absolute differences listed here are reasonable. This will improve the accuracy in the common oxygenation ranges that are encountered clinically.

Overall, as a society, we want the panel to recognize how important equitable and accessible pulse oximeters are to anesthesiologists. We hope that finding a balance between equitable patient care and the demands of pre-market testing can be found to ensure that all patients are treated fairly. We believe incorporating the objective measures of skin pigmentation and in the form of the Monk scale or ITA measurements are promising in terms of addressing this problem.

Finally, I hope that this panel and the FDA considers providing research funds to support investigators working to find solutions to this problem. Collaborative work through open oximetry is ongoing and support to this group or other groups investigating this important topic would be valuable to future patients.

Thank you to the panel, thank you to the FDA, and on behalf of myself and the Society for Technology and Anesthesia, we appreciate you having us. These are my references, and thanks.

Dr. Davis: Hi, I'm Dr. Terry Davis, President of the American Association of Critical Care Nurses. Thank you for the invitation to provide testimony. AACN is the largest specialty care nursing organization providing education and advocacy for acute and critical care nurses in the United States. This issue is vitally important to us. Nurses are at the front line of health care using pulse oximetry to evaluate patients' clinical status. AACN is also a member of the Critical Care Societies Collaborative, a partnership of four professional societies whose members care for America's critically ill and injured patients. AACN was a signatory to the CCSC letters urging the FDA to direct developers of pulse oximeters to test all devices to ensure accurate and reliable readings for all patients, regardless of their degree of skin pigmentation. We continue to advocate for this action. I have no financial disclosures or conflicts of interest with the material in this presentation.

I'd like to thank the FDA for their attention to this issue and for including AACN as a key partner. Collectively, we've made progress in raising awareness of the disparity in pulse oximetry accuracy. This disparity results from inequities in the way the technology is evaluated. Because skin pigmentation is related to an individual's racial and ethnic identity, the inaccuracy of the pulse oximetry in darker skin tones exacerbates existing racial inequities in our healthcare system. Given widespread reliance on pulse oximetry data in clinical decision making, addressing this inequity is imperative. The clinical study design outlined in the discussion paper site specific tools such as the MST and to ensure inclusion of a broad range of skin tones in study participants and the ITA to measure skin pigmentation at the sensor site. Use of validated tools

and objective measures can more effectively reveal accuracy gaps. This is a step forward, and there is further work to be done.

Raising awareness of disparities in pulse oximetry readings is essential, as changes in study design will not improve the accuracy of existing technology for some time. AACN's commitment to bridge this gap in patient care is demonstrated in our webinar on pulse oximetry and skin color, presented by Dr. John Gallagher, a doctorally prepared nurse. The webinar explains how pulse oximetry works and reviews the data demonstrating higher rates of occult hypoxemia among patients who identified as black, substantiating that the degree of skin pigmentation affects accuracy. That webinar is open access and is available to all healthcare providers.

Key points highlighted in the webinar include to minimize the risk of inaccuracy, use the correct sensor with the correct placement. Be aware of disparities in pulse oximetry accuracy in patients with dark skin pigmentation, as well as in patients experiencing changes in regional perfusion, anemia, and edema. Consider other assessment data in conjunction with pulse oximeter readings when making clinical decisions.

Pulse oximetry is a ubiquitous technology and the evidence demonstrates a disparate risk for occult hypoxemia among patients with darker skin pigmentation. Because of its widespread use, we must develop processes to ensure Pulse oximetry accuracy, including consumer grade pulse oximeters. Create and disseminate content that clearly defines the risk of relying on pulse oximeters as a single data point to confirm adequate oxygenation. This situation draws attention to the profound gaps in our healthcare system and offers an opportunity for the FDA to lead the implementation of equitable practices in clinical study designs for all diagnostic tools and products. Skin pigmentation is a measurable physical attribute that can and should be considered in testing any technology. In taking steps to address this one inequity, the FDA can establish new

standards where inclusivity is the rule. Not the exception, so that when assessments and algorithms are developed, they are appropriate for everyone who seeks our care. The FDA can set this precedent through its powerful role as a convener, bringing together the scientific and professional communities. AACN, as a leader in critical care, stands ready to support this pa—

Dr. Cassiere: All right. Thank you. This is Dr. Cassiere, panel chair. Do any of the panel members have any brief questions for the guest speakers? Okay. I will take that as a no. I want to thank our guest speakers for a very informative presentations. I'd like to, at this time, announce that we're going to take a five minute break. Panel members, please do not discuss the meeting topics during the break amongst yourselves or anyone attending virtually. We will resume at approximately 3:17. Thank you.

FDA Questions to the Panel

Dr. Cassiere: This is Dr. Cassiere, Panel Chair. It is now 3:20 and I would like to resume this panel meeting. At this time, we're going to focus our discussion on the FDA questions. Panel members, copies of the questions are in your panel packets. I would ask that each panel member identify him or herself each time he or she speaks, to facilitate the transcription. And I ask if we could please show the first question. Great. I will read the question.

Question 1

Dr. Cassiere: The question one that the FDA proposes for us, for deliberation, is the Agency is proposing a more inclusive and representative clinical trial design, to improve the quality of pre-market studies to evaluate the performance of the pulse oximeters, taking into consideration a patient's skin pigmentation and patient-reported race and ethnicity. Some of the key elements of this proposal are: number one, inclusion of at least 24 participants that span the entire Monk Skin Tone. The MST has been validated to capture race and ethnicity diversity in pigmentation within

the United States. This will improve generalizability of study results. Second item. An initial assessment of skin pigmentation with the Monk Skin Tone Scale, followed with an objective pigmentation measurement, the Individualized Topology Angle, ITA, at the sensor site. Next slide, please. Question 1, Part A. Please discuss the advantages and challenges to the proposed clinical trial design, including specific consideration and discussion of the following: A, pigmentation measurement approach, and whether it will provide approximate diversity of race, ethnicity, and pigmentation in the clinical pre-market study. I'm going to stop there because there's a couple of other parts of this question. So, for our panel deliberations, let's focus on that part A. Do we believe, or the pigmentation measurement approach proposed by the FDA, is appropriate for the clinical pre-market studies? And I open it up to the panel, and I have a bunch of hands up already. And Dr. Lewis, you had your hand up? Was that a mistake?

Dr. Lewis: I was ready for part B, but if someone wants to comment on Part A--

Dr. Cassiere: Oh, sure. Dr. Lanzafame.

Dr. Lanzafame: Sorry, talking to myself momentarily, [unintelligible]. Yes, I think there certainly is an advantage to using both the MST and the ITA approaches, as well as gaining subjective information regarding race and ethnicity. I do believe though, that inclusion of a greater proportion of individuals at the darker skin types, the darker MST levels, is something that should be seriously considered. And, similarly, regarding the sites of measurement, if I understood the strategy, the location quote at the sensor site, what was actually more the Palmer aspect of the hand, which really is not truly the site of the sensor. I think that still introduces some experimental variability. I realize there are challenges relative to the size of the measurement devices, that are being used to gather the data. So perhaps one might consider actually taking multiple values, over those larger sites, in order to obtain an average.

Dr. Cassiere: Great, thank you. Dr. Goldman.

Dr. Goldman: Thank you. Julian Goldman here. So, three points. Number one, I generally support the approach that's being proposed, in terms of looking at both the visual assessment with MST and then an objective assessment with ITA. And then, but I'd like to address two aspects of that. One of them is, I think that a number of presenters and the data proposing the binning, I think is pretty-- is not unreasonable, given the challenge of finding sufficient patients in every single MST slot. So, I think there's some value to that binning, and that's been discussed at length. So I'd like to call that out and support it. The second point I'd like to make is the measurement of ITA on the dorsal aspect of the distal most aspect that's accessible on a finger, seems like a good idea. But let's keep in mind, it isn't actually the measurement site for pulse oximetry, as has been mentioned. So, I think it's important, as with most of what we're proposing or is being proposed, and discussed, that we keep eyes open and recognize that it may or may not be successful, and therefore it's important to monitor and collect data, and reassess whether these ideas were solid, as more data is collected. Thank you.

Dr. Cassiere: So, Dr. Goldman, let me just-- you agree with the bucketing of one to four, five to seven, and eight to 10?

Dr. Goldman: That bucketing, or similar bucketing. I don’t know what the magic numbers are, but yes. As opposed to trying to make sure there are numbers in every single category, I think it's just gonna be not practical at all.

Dr. Cassiere: Yeah, okay. And I’m interested to hear from other panel members about that in a minute, but let's, Dr. Feldman.

Dr. Feldman: Yeah, just to extend on those comments. I also agree with having the MST in there, because of the concern of race and ethnicity in this question is going to be important. But I'll come back to my comments earlier, in that I think the overriding goal here is to try to put aside the question of whether or not skin tone is introducing bias in devices that are approved or not.

And I do have some concerns that the proposed testing methods and sample sizes, which we'll get into later, I guess, we'll actually put that question to rest. And in particular, because, ITA, even though it's objective, is really not going to tell us what happens to light as it's transmitted through the sensor site. And we saw a few minutes ago that there's a recent abstract from the STA (phonetic) meeting, that I have to admit I haven't seen. Came out of the Medtronic lab, actually looking at the absorption by skin tone of specific frequencies of light. And that may be the better approach, if we really want to put this question to the rest.

Dr. Cassiere: So, Dr. Feldman, before I go on, let me probe you a little bit more than that. So, you're discussing more than the ITA location, but just the ITA in general, using it as an objective measure of skin pigmentation?

Dr. Feldman: Correct. I agree that it is an objective measure. My concern, though, is that it's only related to the melanin content of the skin, as determined by histologic studies. It doesn't actually tell us how light is interfered with, or interacted with, in the skin of a dark finger. And I think that really is the information that we'd want in an objective measure.

Dr. Cassiere: So, again, I'm going to ask again, to probe a little bit more. So, if it was not the finger, but the forehead or the forearm, with the ITA, would that be something that you would think would be applicable?

Dr. Feldman: Actually, I think that if it's not at the measurement site, then I think the value of that as an objective measure for really achieving what we want to achieve, which starts to go down. Not saying it's not valuable at all, but it really may not answer the question of skin tone and the interaction with most (phonetic) of the similar measurements.

Dr. Cassiere: Okay, yeah, we'll discuss that a little further. Dr. Wilson.

Dr. Wilson: Yes, thank you. The points that were made by Doctors Goldman and Feldman, I think are right on. The problem with where the sensor is, is typically over the nail bed, which

often has very little melanin. But we need to decide on an area for the ITA, so the distal fingertip, on the dorsal aspect, just proximal to the nail bed, may be the best surrogate, even though there are some issues as Doctor Feldman just mentioned. But I would recommend that you take, or that the FDA would require, a balance average of three measurements at the site, because there still could be some variability there. And I think my other points were already made by the others.

Dr. Cassiere: So, let me just ask again. So, you're in agreement with-- so, I think from what I'm gathering, I haven't spoken to everyone yet, and I'll call out individual names in a moment. We seem to be so far-- the Monk Scale seems to get a pass, in terms of a subjective measure of skin pigmentation. And now we're really talking about is, does anyone disagree with that? Because I haven't heard anyone so far say that the Monk Scale does not seem to be a good, part of this FDA's new evaluation pre-market study. And it seems like our focus now is on the objective measure of scale.

Dr. Feldman: So I definitely agree that the MST does a good job of covering the spectrum. The IT is a very good major, objectively, of the skin pigment. The concern is that the light is moving through a pathway that may actually have less pigment than you're going to determine by the ITA, just proximal of that. Nonetheless, and work that we've looked at, the ITA correlates fairly well with a Monk Skin Scale across the spectrum.

Dr. Cassiere: So let me just-- again, to wrap my head around it. I think one of the one of the issues that jumps out to me, is we're not looking for a way of seeing how the light's going to go through the nail bed. We're looking for a way to support the Monk Scale determination of skin pigmentation objectively, right?

Dr. Goldman: If you're trying to do that, and you're measuring the Monk Skin Tone at the forehead, then the ITA should be at the forehead. But the concern is that skin pigment has an effect, and to put that to bed, measurements as close to the center site would make sense. The

concern is, the light is traveling mostly through the finger bed, which has much less pigment than the dorsal aspect of the finger just proximal of that, where the Agency is now recommending we take that measurement.

Dr. Cassiere: Okay. And you're in support of to have a number in there of how many of those ITA measurements they do?

Dr. Goldman: It's recommended to take three measurements and take an average.

Dr. Cassiere: Okay, thank you doctor. Dr. Taylor. We need help from our dermatologist on the committee.

Dr. Taylor: Well, I'll bow to the technologists, more so. I agree with the use of the other point I want to really want to make is that I think it's important to have the comorbidities. That's really not been addressed. Finger edema was mentioned. So, if you're doing the measurements with the finger, and then the other is sun exposure. Every study we do, we quantitate sun exposure or limit sun exposure if somebody has had acute sun exposure for two or three weeks. So at least, and especially in the forehead, because the forehead would be more likely to, I would think, be pigmented from sun. So, and then there was another group of Skin of Color Society that I'll suggest they contact the FDA. I don't know if they have anything else to add. It's a group of dermatologists that have discussed this for years. They may be able to comment. But definitely sun exposure needs to be, and other comorbidities, need to be included in the study design, I think.

Dr. Cassiere: Let me clarify that a little bit, because a lot of these, they're going to be volunteers that are normal and healthy. Are we recommending to the FDA that they should look for volunteers with diabetes, hypertension, peripheral vascular disease, and other comorbidities? Is that what you're--

Dr. Taylor: No, no, no. No.

Dr. Cassiere: But I understand it could affect the pulse.

Dr. Taylor: Right.

Dr. Cassiere: I think the question that the FDA is posing about the pre-market evaluation, and I guess later on we'll talk about do they need to do real-world testing, where I think your comorbidity question really hits home, in terms of whether they're obese, hypertension, diabetes is going to be maybe a little more pertinent for that patient population than the volunteers.

Dr. Taylor: Well, right. Well, with the volunteers, you could still do-- they're likely to have, have or have not sun exposure-- I quantitate, identify if they've got edema, of the fingers and so forth.

Dr. Cassiere: Great, thank you, Dr. Taylor. Dr. Lanzafame.

Dr. Lanzafame: Yes, hi. I don't want to get too far into the, the weeds here, but there were a couple of statements made that probably need to be clarified a bit. There are spectral curves for melanin, and its variants, both absorption and transmission, so people know what wavelengths are absorbed and to what degree, particularly in the range we're talking about. At issue becomes what the density of that is, in the point specific area that we're attempting to measure in this case, the level of oxygenation. And so, really the matter is to what degree at the specific point that we're trying to get objective data, do we also have objective data about the level of melanin present, by our surrogate, which is both the MST and the ITA. And so I think all of this is a good attempt to do that, but I think we also have to remember, as everybody's been talking about throughout the day, that there are many other confounding factors as well as those.

Dr. Cassiere: Thank you, Dr. Lanzafame. Dr. Gooden.

Dr. Gooden: Yes, hi. Cheryl Gooden. As I listened to much of this, and I think I kind of mentioned a little bit earlier my concerns for the pediatric patients, because I have to put on my pediatric hat as a pediatric anesthesiologist. And a lot of the discussion again has been related to location of the probe, for the fingertip as well as the forehead. And I would say if we're going to

include pediatric patients in these studies moving forward, I think it's necessary to also consider where we're putting the probe in pediatric patients. So, for example, many times we will use the toe, or we'll use the foot itself, for placement of the probe. So, it's just a comment that I make about pediatric patients, and perhaps the FDA may want to consider that in studies moving forward, since pediatric patients will obviously be part of this.

Dr. Cassiere: Great, thank you. Dr. Feldman.

Dr. Feldman: Yeah, just to clarify a couple of comments. The ultimate goal is to know that when a device that's been approved is placed on a patient, it will perform within the accuracy range that the FDA ends up prescribing, irrespective of skin tone, racial designation, et cetera. Clearly, self-described racial designation is not a sufficient way to stratify volunteers. Monk's Skin Tone, to the extent that it's been nicely validated by Dr. Monk, has some value, but again, doesn't necessarily give us information about what's happening at the sensor site, hence the desire to make a measurement at that location. What I was suggesting with regard to actually measuring the transmission of light, I know that there's extinction curves for melanin that are well known, but of course we don't know how those apply to the individual patients. So, if I was reviewing a study, and I've been a reviewer for a long time, where someone said we took each volunteer and we measured at the site of oximeter measurements, first. The transmission at 660 and 940 of these frequencies when we found it to be attenuated by X percent, in that patient, and then stratified patients on that basis, to me, that would be physically much more relevant and likely to yield more reliable predictive results into the future. I recognize that that current approach is not in the literature, it hasn't been established, although the abstract, that STA abstract that was mentioned, is certainly interesting. In trying to provide advice to the FDA, I don't want to fall into the trap of saying, well, we've got these well-established ways of doing things that have

been validated in other domains, therefore let's use them here, when in fact they may not apply as well as we might like.

Dr. Cassiere: Great, thank you for those comments. Any other specific comments about this? Dr. Brown, anything to add or agree or disagree, or just your thoughts?

Dr. Brown: Yes, thank you. Whitney Brown. I like the inclusion of the MST. I think ultimately, if you look at it very pragmatically, what we need to do is enrich the study population for a variety of MSTs, in those buckets. And I did like the idea of one of the speakers that said, make sure you cover two and nine as sort of the extremes, at least of our population, in this country. And I think the other, in terms of really objective measurements of transmission through light, et cetera, those are nice to have, but I think if we include enough samples-- if you take a pragmatic approach, we just really want to see how it performs in actual fact. So I really favor that approach.

Dr. Cassiere: Great, so I'm getting a sense that this bucket approach to the Monk Scale seems to be reasonable, as opposed to picking out individual numbers and having a certain number of patients. That's what I'm gathering from what I'm hearing from everyone. Dr. Wiswell, do you want to agree, disagree, or any comments along this Part A question?

Dr. Wiswell: I have a couple comments. I certainly agree with the two approaches, the MST, the ITA. But I sort of feel, even though with the statisticians' explanations, that the total end of 24 patients just seems low to me, keeping in mind that these are healthy volunteers. And, if it's not what the devices that are going to be used on in the future, healthy volunteers, and so personally I would like higher numbers, recognizing that underpowered studies can lead you astray. And, while the bucket approach sounds reasonable, I'd probably narrow it down to one and two, three and four, five and six, so you have five different buckets, because there may be just such if you

have it only three buckets, as it were, it may not catch everybody. Again, especially at the darkest pigmentation end of the spectrum.

Dr. Cassiere: Great, thank you. Dr. Yarmus, any dissent, agreement, or what do you think?

Dr. Yarmus: Yeah, general agreement. I think Dr. Brown's great use of pragmatic approaches, I think is super relevant, because this is what this is going to come down to. I think the healthy volunteer piece seems like it's going to be needed for base interpretation. But I think we should also be thinking about these separate pockets of pragmatic, I don't know if it's clinical trial or really just pragmatic, clinical observation studies in the scenarios that I think we all have recognized are relevant, and bridge those observation studies into this.

Dr. Cassiere: Great. It's scaring me, both our biostatisticians have their hands up. Dr. Ballman.

Dr. Ballman: No, mine's not statistical. It's just sort of referring to this. When people are talking about buckets, I was interpreting it as within those buckets you have to make sure you're across the range as well. And I think that needs to stay in there, so that it's not just on the low end of every bucket. And also I like the ITA as well, because I think just putting them in these buckets, as mentioned at the site of the sensor, it might differ even for everyone who's like MST six. And so you can get some sense of that variability as well.

Dr. Cassiere: Great, thank you. Dr. Saville. You're muted, Dr. Saville. No…

Dr. Saville: Can you hear me now?

Dr. Cassiere: Oh, perfect. Yep.

Dr. Saville: Okay, I'm not sure what's going on there. So, I don't have a lot to add here, other than that I do think it's a good idea to have both the MST and the ITA. And I think it's a good idea to stratify the enrollment criteria by MST to give you a wider range there, for (phonetic) the skin pigment. I do have some comments on sample size, but I think we were going to save that for the next topic.

Dr. Cassiere: Yeah, we're going to be in part B real soon.

Dr. Saville: So, I'll stop there.

Dr. Cassiere: Before we do a little summary, Mrs. Brummert, any thoughts, any agreement? You like what you're hearing?

Ms. Brummert: I agree with the use of the MST, but I remain uncomfortable with using 24 healthy volunteers. I think we need to include people with medical conditions, hospital setting, healthy patients, just that we're going to get a better baseline if we include more of what we're talking about today. This 24 healthy patient thing doesn't work for me.

Dr. Cassiere: Great. Mr. O'Brien. Agree, disagree, you like what you're hearing?

Mr. O'Brien: No disagreements. I like what I hear, overall.

Dr. Cassiere: Alright, great. Dr. Wilson, you have any other comments before we do maybe a summary?

Dr. Wilson: The only additional comment is related to the real world in terms of the spectrum of skin tones in the United States, and having the three bins makes a lot of sense. But requiring ones and 10s may be exceedingly difficult for anyone to actually get data for submission.

Dr. Cassiere: Dr. Gooden, I apologize if I missed your hand.

Dr. Gooden: Yes, hi. Cheryl Gooden. I would also like to echo that I am a proponent of the MST. I think it's the most diverse scale that we've discussed today. And hope that we can also apply it, not only in the adult population, as we talk about a variety of patients, but also use it for our pediatric patients.

Dr. Cassiere: Great. So, I'm going to try to summarize question 1A. It seems like the committee likes the subjective and objective approach to these pre-market studies. The Monk Scale seems to be on point, which is good for a subjective approach. It seems we have some buy in for the ITA, even though it may not be extensively studied. There may be some variability. To Dr. Brown's

point, sometimes you gotta do it. You gotta be pragmatic. And this is going to be, I think, the first objective measure that the FDA is going to require for pre-market approval. So, it's a start, and I think a good start. I'm also hearing, again, more of this bucketing, and some good points from the panel members on they want the one to fours, the five to sevens, and the eight to 10s, but you don't want them all to be ones and fives and eights. So, I think we like the idea of having representation across the Monk Scale. I think that that's going to be also very, very helpful. And I'm going to ask my FDA colleagues if you're satisfied with this discussion and what's come out of it at this point, Dr. Eydelman?

Dr. Eydelman: Yes, that was a great summary, thank you. Please go ahead.

Dr. Cassiere: Great. So now we're going to do part 1B. If we could put the-- and we already got a lot of hands up for this one. That's great. What additional-- I'm sorry, Part 1AB. Excuse me. B. Proposed sample size and whether it will address appropriate diversity with respect to race, ethnicity, and skin pigmentation. And I think we have a-- there's a lot to say about this. And Dr. Willis, Dr. Lewis, you want to start us off?

Dr. Lewis: Yes, thank you. This is Tamorah Lewis. So I think there's two things I want to discuss. First is that, after listening to all the discussion, and especially some of the external stakeholder groups, I think we have to really talk about powering for between-group differences. And, with the two studies presented by the CERSI groups, I think we actually do have the data on a full range of MST patients, both adult and pediatric, that we can do power calculations that would make-- that would allow us to query between-group differences. The current idea to just look at sort of variation across the entire spectrum starts to get at that, but I think especially in the effort of regaining public trust and regaining trust of clinicians, it's going to be really important that we power these pre-market studies to look at between-group differences. The other comment that I would like to make, and I hope that the panel will help me think through, is this idea of over

sampling the most-- the skin tones that we believe to be most at risk for poor performance. So, instead of sampling equally across the MST, we have, I think, enough cumulative data at this point to understand what are the skin pigment groups that are at risk for poorest performance, and is there merit to considering really focusing and oversampling in those darker skin pigmentation groups? And I really feel like we have to be drawn to doing what's right, and not what's easy or feasible. And there are patient advocacy groups that can be engaged, that can help us in this oversampling of the most at-risk patient groups if we decide that that's the best thing to do.

Dr. Cassiere: So, Dr. Lewis, let me just probe you a little bit with that. So that's also helpful with part AA, where we're hearing that it may be difficult to get the upper end of the Monk Scale. And I guess that's dependent upon where they're doing this volunteer enrollment. And to your point, if you can't find it in your backyard, go out of the neighborhood, and find who's in the Monk Scale for the studies. But in terms of the part B, the sample size, are you, and again I don't mean to pick on you, but it's going to be one of many. That sample size of 24, spread out amongst the MST and the ITA, the FDA is asking us, do you think that's an adequate sample size or should it be modified, or what are your thoughts on that?

Dr. Lewis: Well, I'm a clinician scientist but I'm not a biostatistician by training, but what I understood from what the FDA presented this morning is that the current power calculations are not adequately powered to look for between-group differences. And so I think that right now it's powered to look at the total summary statistic, I forget what it's called at the moment. But I think that's inadequate.

Dr. Cassiere: Okay. Okay, that’s helpful. That’s going to bring us right into Dr. Saville.

Dr. Saville: Yeah, thank you. Ben Saville. So, in terms of the power, if we’re looking at between-group differences, I think certainly what Dr. Pennello said earlier-- instead of just looking at

those categories, looking at, for example, the ITA is an ordinal variable and continuous variable. That's going to be important. I think with 24, I still don't know what that power is going to look like. My hunch is that it is probably insufficient. You may need larger sample sizes to really evaluate whether the interaction exists with skin pigmentation and the bias, and to assess that inferentially with either [unintelligible] or Bayesian methods. I think you would need some virtual clinical trial simulations to really inform what that power is. I'm a huge fan of clinical trial simulation, and I know the FDA did several simulations to get those power calculations. And they use simulations to demonstrate that they had 80 percent power under a certain set of assumptions. And I think it's a little risky to give that sample size. I know they said a minimum of N equals 24 with 20 repeated measurements per individual. But, to me, that gives a false impression that N equals 24 is appropriate in all settings. It's certainly better than what the previous guidance was, back from 2013.

But I think there's a risk there to companies coming in and saying, alright I need 24 and I need 20 observations, repeated observations for each individual, and I just have to figure out how to get those observations of those individuals in the different categories, and that's it. But the reality is, that power calculation, the 80 percent power, that was based on a set of assumptions. And that included not only the SaO2 but it also included the between- and within-subject variability. And, what that means is, given a single set of assumptions, you have 80 percent power. But if I were helping a company design a trial, I'd say let's look at the power across the whole spectrum. What if we're way off on our assumption on that between- versus within-subject variation? What if you're on this side or on this side, or this extreme or that extreme? And so I'd look at the whole power curve, not just a single calculation, but here's the sample size range that you need for 80 percent power under a different set of assumptions. And I don't know how expensive these trials are. I work mostly with kind of with more expensive class 3 devices,

clinical trials. And I don't know what the risk is here to the companies, but in terms of increasing the probability of being a success, having a successful trial, really understanding what those virtual trial solutions could look like, and what the data could look like, and getting the right sample size for each individual sponsor, as opposed to a single recommendation. Because the reality is, N equals 24, that's not a magic sample size. It could be too big, it could be too small, depending on what the objectives are, and the variability that you're going to see in the population that you enroll.

So, FDA has already done a lot of those simulations. It would be very easy and straightforward, I feel like, to feed in different simulations into those to get a power curve. One recommendation I would even have is, FDA, if they really want to help people out in terms of getting that power calculation done, is you could create like an R shiny (phonetic) app, where you basically just create a GUI that the people go in and they input their assumptions, and it can create a curve like that for them. But I think that's something that, it would be helpful if the companies could understand what the risks are, based on the assumptions that fit into that calculation.

Dr. Cassiere: So, Dr. Saville, let me ask you a question along that line. So, they're asking us about the proposed sample size and whether it's going to address diversity. So, if we take it to the bare basics, the diversity is really each one of those Monk Scale categorizations. Is 24 going to assure us that there's enough diversity within that patient population of 24, if we say that's the bare minimum, to assure that the light skin and the darker skin behave the same in terms of pulse oximetry? Or is that going to take a higher power number?

Dr. Saville: So, it depends on what your criteria are going to be, for what's-- is there an interaction here or is there not interaction? So, right now, this criteria, for example, for the second part of the primary analysis is looking at the-- what do you call it? The non-disparate

performance analysis. So, the way that analysis is currently defined, 24 gives you sufficient power to either reject or fail to reject that hypothesis based on if you're less than 1.5 or less than 3.5. So, yes, you have sufficient power to detect differential bias between those different skin pigmentations, with that criteria. Is that the appropriate criteria? I'm not convinced. So I would rather see inferential statistics. I would rather see a Bayesian probability that there is something, there's a meaningful differential bias between groups. And so, I'm guessing, if I were to dig in and do the simulations, I'm guessing I would want to see a larger sample size. But, part of this goes back to my first comment, in that it really depends on what the objectives are, and you use simulations to help inform that.

Dr. Cassiere: So, let me let me ask you again, not to pick on you, but I got you there. Let's say we would recommend doubling that, to 40 or a number of that magnitude. Does that mitigate some of the issues that you brought up?

Dr. Saville: It likely does. But again, I would want to see the numbers and the calculations. I'd want to see, under different assumptions, how does that impact the sample size? On extreme, if I have extreme assumptions, is 48 much more protective against those assumptions? And it's going to provide a more robust answer. I would have to see just how sensitive those power calculations that FDA did are to the assumptions that go in. If they're really sensitive to those assumptions, then I think I'd absolutely encourage a company to go bigger than that. Because I don't think 24 would be sufficient. So I would have to see the numbers, though, to really give you an answer on that.

Dr. Cassiere: Okay, great. Alright, Dr. Feldman.

Dr. Feldman: Yeah, so two comments. I know the desire here is to come up with a particular number, and I'm not a statistician, so I can't do that. But I think the important thing is to have the right question, decide what question we're trying to ask, and answer with a powered study. So, in

my view, the question we want to put to rest is to test the differences between groups based on skin tone, and be able to say with statistical certainty that's appropriately powered, that the differences we observe between those groups are not, not different, within the accuracy range specified. So, after we do all these observations, we can compare the groups, between group comparisons, as has been supported, and that those differences are not different. If it can show that, then I think we've shown non-disparate performance in the data set. The second comment is, there's nothing intrinsic about pulse oximetry that prevents it from working well in people with dark skin tone. The reason I believe that it doesn't, now, is because the populations that informed the look up tables were not sufficiently representative of people with dark skin tone. If it had been the other way, if it had been mostly dark skin tone and few white skin people that informed those look up tables, we would be looking at bias in the other direction. So, fundamentally, that's going to have to change, and whatever processes are put in place will need to need to detect that.

Dr. Cassiere: So, Dr. Feldman, thanks. So that last point that you made is important, we'll highlight it a little bit more. Just the mere fact that skin pigmentation is going to be highlighted, and there's going to be much more diversity, should change those calibration curves and make pulse oximetry more, for lack of a better term, accurate, in our dark pigmented patient population.

Dr. Feldman: I think, ultimately, if I was a manufacturer looking at what I've done in the past and what I need to do going forward, as the requirements are more stringent for performance across skin tones, you're going to have to go back and take a look at how the device was designed. I'm not an engineer to that level of what these guys do, but there may be other tweaks, but that's going to be a fundamental limitation, I believe.

Dr. Cassiere: And I think to your first point, during the FDA presentation, they seem to, at least in my mind, again, I'm not a biostatistician, they seem to prove the statistics behind that 24

number, differentiating intragroup differences, to mitigate them. Anyone disagree with that or have a different interpretation?

Dr. Feldman: Yeah, I don't want to take anything away from Dr. Pennello's nice work on what he's done, but I don't believe it's the appropriate power analysis for what we're trying to accomplish.

Dr. Cassiere: And if you could expand on that, that would be really helpful.

Dr. Feldman: So, again, I may want to call, phone a friend, one of the biostatisticians.

Dr. Cassiere: Absolutely.

Dr. Feldman: But in my view, we want to be able to say at the end that, when we compare the performance between groups of different skin tones, that we can say with inferential statistics that the differences that we detect are not-- I'm losing the right word, but are not important. They're within the accuracy ranges that we're trying to perform. So, that if we say it's two percent and we show with inferential statistics that dark skin and light skin both perform equally within that range, then we've nailed it. So, I see a friend shaking your head, so that's always reassuring, but please weigh in. I don't feel like I have quite the statistical background to be an expert on this.

Dr. Cassiere: I don't mean to jump ahead, Dr. Goldman and Dr. Wilson, but Dr. Ballman, I'd like to hear her comment on this.

Dr. Ballman: Thank you. Karla Ballman. Yeah, I think it goes back to what the other statistician pointed out. It depends upon what your measure is. So it's powered for the measure that they have right now for the disparate, but I do think it's more important, perhaps, like I was thinking maybe each sort of bucket, or however you're going to put the skin tones, that each of those are either within sort of the guidelines that they want for accuracy, and/or show that there's no difference among the different groups. So, I agree with what Dr. Feldman was saying that, I think

that the measure they picked right now is perfectly powered for, but I don't think it's powered for a little bit deeper dive into how different it might be among the different groups.

Dr. Cassiere: Okay, alright. Dr. Wilson.

Dr. Wilson: Yeah. Thank you.

Dr. Cassiere: I'm sorry, I skipped over Dr. Goldman. I apologize. Dr. Goldman, they took your hand down already. I apologize.

Dr. Goldman: That's okay. Thank you very much. Julian Goldman. So, I want to take a step back and recognize that most of the studies and the issues that are being observed, not all of them, but the examples of clinically significant disparate bias are occurring in patients in clinical settings. They're not occurring in the control laboratory desaturations environments, with healthy volunteers, adult volunteers. So it's possible, these numbers, sure we can increase the numbers to 48, to 100, whatever the magic number is, and we will increase the precision of pulse oximetry performance in that population, of healthy adult volunteers that can be desaturated in the laboratory. But the question really still remains, what the heck is going on, in these clinical environments, that are causing the numbers that are raising such concern? And I think it's vital that we continue to attend to that fundamental question. That we're doing our best to address the issue now, without fully understanding the mechanism. I was really struck when Dr. Fawzy, in response to the questioning, pointed out that they observed the same patients with dark skin pigmentation had an error that varied, a bias that varied throughout the day. We can study 100 volunteers in the de-sat lab, and what are we going to do to what's really going on in these clinical settings. So, I don't want to lose sight of the forest for the trees in this. So, we're increasing from 10 subjects to 24. It seems it seems reasonable. I think no one in the deliberations I've heard, for example, in the ISO pulse oximetry committee meetings, there isn't a magic number. This is a movement toward greater precision. But you can keep going. So, I think

we want to put that in perspective. And we still are not requiring data, at all, in these volunteers that are low profusion data. And certainly, the signals are that the disparate performance is probably related to conditions of dark skin pigmentation--

Dr. Cassiere: I'm going to agree with you, but that was going to be a really good discussion for the part B of this question, because part B is exactly what you said. What the FDA is asking us in part B, what other recommendations do we have to help with performance? I want to put that on hold a little bit. You and I are in total agreement. I want to focus on this proposed sample size and whether it addresses diversity.

Dr. Goldman: Well, I also wanted to point out that we don't want to end up with excessive potentially burdensome optimization, without discussing the challenging areas. So thank you.

Dr. Cassiere: I agree one hundred percent.

Dr. Wilson: Just a few additional points. First is, what we're looking at is the bias difference. What's the bias? The SpO2 minus SaO2, for those with dark skin and light skin. And the ARMS looks at the accuracy by root mean square, looks at both the bias and the precision. If we have a very large sample but we don't narrow the ARMS, then you may not necessarily be doing very much good. So, we do need to make sure that that's being addressed, one way or the other. The second point is, and let me just answer one question that came up, which is how much do these studies cost? And if we expand to 24, then the basic cost will be somewhere in the range of a quarter of a million dollars, about twice what it cost before. But that's okay because we need to make sure that we have representative samples. The most important factor here, from my perspective, is ensuring that we have almost equal number of dark skin and light skin, and that we cover the full spectrum of individuals, and that is being addressed by the agency. Those factors will markedly improve the data, but we also need to narrow the difference in bias. So we

need to decrease that ARMS one way or another. I don't know if it needs to go to 2% or somewhere in that range, but the agency should consider that.

Dr. Cassiere: Great, thank you. Mrs. Brummert.

Ms. Brummert: This is Rachel Brummert here. So, just for perspective, there are 46 participants in this meeting. At one point, we had 58. We had speakers and people coming and going. We had a pretty diverse group, I think. So, the 46 that are in this call are weighing in on a decision that's pretty important. So, using that as like a visual to what point I'm trying to make is that like 24 is half of that, and there's 46 of us making an important decision. So, honestly, I think doubling the sample size would be a benefit. I think 24 is just the bare minimum. And I think we have a responsibility as a panel to get as much information as we can. And I think we can accomplish that by at least doubling it.

Dr. Cassiere: So, I guess if I could add on to that, the proposed 24 patients for a device that's probably going to be used on thousands and thousands of patients, that's also relevant and point well taken. Dr. Yarmus, anything to agree, disagree with, along our thoughts? In terms of the question at hand.

Dr. Yarmus: No disagreement. I completely understand the sort of optics of the sample size. But, and this is probably a little bit later, but I think this all again comes down to what the study is, right? So, if it's just a validation of an existing technology, that's very different than what we've been talking about, which is a full spectrum of individuals, and if that continues to move that way, then yeah maybe it's 24 patients or 48 patients, whatever it is, in each cohort, right? And that, right, that'll exponentially grow, which will have its own issues, but seems to be indicated.

Dr. Cassiere: Dr. Brown, anything to add?

Dr. Brown: I really, regarding the sample size, I like the idea of enriching for the darker skin tones where there's more uncertainty, and there may be less a clear association between the MST

and the ITA. So, I do favor a larger sample size and maybe enriching for the most vulnerable or potentially vulnerable individuals, which may have unpredictable responses based on those two tests.

Dr. Cassiere: Alright, thank you. Dr. Saville.

Dr. Saville: Ben Saville. So yeah, I just want to circle back. I'm hearing a lot of comments that are echoing what I was thinking and what I was trying to express, and I want to make the point that the criteria that was decided upon, the 1.5%, 3.5%, maybe we'll get to that more later, but that criteria was decided on, it was calculated based on what would be feasible with a sample size of 24. And I don't necessarily think that's the right way to do it. What I'm hearing from the panel, what I'm hearing people say is, we want to really learn if there are differences between these different groups with different skin pigmentation, and the right way to phrase that is to figure out okay, how are we going to address that question? And to me, that's asking for inferential statistics, and maybe that requires a bigger sample size. And maybe that perfectly justifies why people are asking for a bigger sample size. So, we don't want a bigger sample size just to be more convinced of some more ad hoc calculated criteria. I feel like what people are really asking for is a more stringent criteria, really a formal evaluation of whether there really is differential bias, based on what skin pigmentation one has.

Dr. Cassiere: Dr. Ballman.

Dr. Ballman: Yeah, I think I'm sort of along those same lines. I have to say, I work in cancer, so it's completely different, and drug approvals, which is different, but we're never given a sample size. We're instead sort of charged with what sort of is a clinically meaningful difference. And, you come up with the sample size. So I think maybe the FDA might think about that, what they would want, sort of what targets for accuracy and whatever. And let the companies come up with what the sample size might be, under their assumptions. Just a thought. There is some literature,

there will be more, about what the variability might be by skin tone, and those companies could use that to sort of power the studies and then convince the FDA that they're within the accuracy limits that have been established across the skin tones.

Dr. Cassiere: Great, thank you. Dr. Feldman.

Dr. Feldman: Yeah, I just have comment about the ARMS discussion a little bit because there was, actually even in the FDA literature, there was some notion that the ARMS, by tightening those limits, it would help to ensure that we don't have, that we have non-disparate performance. And I don't see the ARMS as driving the disparate performance question. To me, picking those limits is important for effective clinical decision making. So, how tight do we need that to be in order to make good decisions for patients? And that should then go on the labeling, and say this device is good with these limits, just like any specifications. The non-disparate performance thing is, once we decide what those limits are, then you just have to see if it performs the same and in all those groups. So, to me, it's two separate things. One is ARMS, what kind of tightness do we need to make good decisions in patients and not miss hypoxemia? And then secondarily, we need to test it to show that it's not disparate between groups with appropriately powered study. So I just wanted to clarify the ARMS. I'm interested in other comments about it. Dr. Wilson mentioned we need to tighten those limits, and I didn't know if that's driven by notions of how it's going to affect clinical decision making, or other thoughts.

Dr. Cassiere: Great. Doctor.

Dr. Yarmus: Yeah, so this might seem a little broad of a question and it's probably more industry focused, but what I'm also wondering is what is, in any of the models that we're talking about, I'm presuming this is a new or improved device, right. Which we don't have yet, to my knowledge. Right? And so, to frame this from my perspective, and maybe I'm off, it would be interesting to understand, and this was approached earlier, what is the technologic scientific

understanding capacity, from an industry perspective, of overcoming the current difficulties? And are we in a situation where we will be able to do the things that we want to do? Because, in my mind, from a device development validation clinical research perspective, is phase one of how to do this, and how to think about it moving forward if that makes sense.

Dr. Cassiere: Definitely, Dr. Yarmus. I'm going to put that open discussion on hold so we can finish the sample size. But yeah, this is again, there's more to come. This is only part one A B. Dr. Wilson.

Dr. Wilson: Just one more comment, just in general, statistically, if the differences in the model are very narrow, in this case, the SpO2 minus SaO2, then one would not need as many individuals, so perhaps 24 and there would have to be some modeling done by the agency, but 24 may be the right number, if the ARMS overall is 1.5% and the ARMS for dark skin is 1.5 or less, and for light skin is 1.5. So the difference between them is like non-significant. Whereas, in contrast, if those values, and let's say that the FDA didn't change the current ARMS of 3%, then if the sample size that came in might need to be a little bit larger, if the ARMS of the data that's presented is higher, like 2.5%, and that there was a lot more difference between the dark skin and light skin. So those variabilities help, and they do associate with what the sample size required.

Dr. Cassiere: Great. Dr. Eydelman, I'm going to try to summarize this. I think you can sense our discomfort with the sample size discussion. A lot of statistics behind it. A lot of what we're looking at. In general, there seems to be a support on what was presented earlier by the FDA, that 24, given that 80 percent power number, can help mitigate some of this diversity using the pre-market evaluation as recommended by the FDA. And I think that's the best we can come up with at this particular point. I that satisfactory? Is there anything else you want us to delve into before I hit part B?

Dr. Eydelman: No, please go ahead. I might have questions after we're done.

Dr. Cassiere: Okay. So we're up to 1B now, and I think a lot of a lot of the panel members have, have hit on this. And this will be a real open discussion. What additional recommendations do you have to improve the evaluation of pulse oximetry performance, while taking in consideration race, ethnicity, and differences in skin pigmentation? And I was going to ask Dr. Goldman to comment on this because he led that. So, Dr. Goldman, lead us into this phase.

Dr. Goldman: Sure, thank you. Yeah. I think we have to, and I think this is happening, we are recognizing that, optimizing performance in the preclinical environment and volunteer adult desaturation laboratory isn't the end goal. It's an important part of testing and comparing pulse oximeters, and it might be a vital part of digging deeper into disparate bias. But if we just focus on that without, recognizing the data that continues to emerge, that the problems are occurring much more so, and to a much greater extent, the data is there, as Dr. Fawzy, to reiterate, really emphasized it, as did others, that we're seeing behavior in a clinical environment that we're just not seeing in the ideal setting of a volunteer subject desaturation.

So, I think it behooves us to put attention on that, and I realize we're not going to solve it today. But we could start to at least figure out what data we need to add, what kind of studies need to be added, to allow industry clinicians and clinical researchers and the FDA to dig deeper. And so towards that, I think that there's enough data to support that low perfusion, especially with darker skin pigmentation, seems to be something that is causing these inaccuracies and the bias. And there are, there have been low perfusion studies performed in laboratories for some time. I think they're more difficult to standardize. They're typically done through cooling, with just cool ambient air, and letting a subject cool down. We don't want to get into that here, now, I'm sure. But I think we want to look into that and better understand whether some of the testing, at a lower perfusion state, is one area. That's one area. The other is, we've touched on higher quality real-world evidence, and the challenges with using studies with patients that have

indwelling arterial catheters for sampling. Part of the problem has to do with the data quality. You know, what kind of sensor was used. Should we have photos to indicate whether it's been applied correctly or too tight? Is the sensor on the same side as the arterial catheter if it's a radial artery catheter and a digital sensor? Was there movement of the hand at the time that the blood draw was being made? What was the averaging time setting of the pulse oximeter module that was used, or the standalone instrument? Could manufacturers help to improve data collection in real-world environments with facilitating the access to real-time wave forms? That was discussed in the presentations. Do we know something about the gain setting of the device? If it's trying to amplify a very weak signal, that could indicate something, for example, about the optical density of the digit. Do we know the signal strength, sometimes called the perfusion index, or the modulation percent? Vitally important, potentially, so perhaps there can be a greater effort to improve the contextual information and the data quality and richness, in the real-world studies, to drill deeper into this.

So, I think those are the points I wanted to make, because we're not getting that data really adequately today, I think. And we're not going to find the-- if we just search, as they say, under the streetlight for our keys, instead of searching where we think the answer lies, we probably won't get to the answer that we need.

Dr. Cassiere: Right. Thank you doctor. So basically, we could change the requirements and pre-market testing and still have the clinical for outcomes in real world environment and not solve anything.

Dr. Goldman: It's quite it's quite possible based on the data that we're seeing today.

Dr. Cassiere: Dr. Lanzafame.

Dr. Lanzafame: Yep. I agree with everything that was said, and I would respectfully suggest that there are a couple of things that can be done in terms of collection of the data, beyond skin

phototype, things like body habitus, the size of the appendage being measured, among the other items that were already mentioned. And I think one of the other questions, really for the group, but also from the perspective of the agency, is at what level of proof does one need to have to be able to market a device? And then what is the agency's perspective on post-release surveillance studies in the real world? And I think the post surveillance study provides an opportunity to gather some of the real-world information that we've all been talking out.

Dr. Cassiere: Great.

Dr. Lanzafame: In addition to the other thing.

Dr. Cassiere: Raymond, great point. You make it through the first hoop and then three to six months, you need to prove in the real world that the device is performing according to specifications. Dr. Lewis.

Dr. Lewis: Thank you. This is Tamorah Lewis. As a neonatologist, I think it's really important to center the patient experience and voice. And so I want to circle back to the two patient representatives that we heard from today. I think a lot of the published literature brings up this issue of disparity in SaO2 versus SpO2. But I think today we heard two other examples of pulse ox performance that are very important to patients. And so, we're talking about the question, what additional recommendations could improve the evaluation of pulse ox devices in the pre-market testing arena? And I bring to the panel, the option of adding potentially a few more secondary outcomes, such as time from placement of device to a steady, accurate reading, because one of the patients worried that his skin pigmentation might be complicating his ability to quickly obtain a steady, accurate reading. And then the second secondary outcome we might discuss is, over a certain time interval, we can talk about is that two hours, is that four hours, what is the percent of time that you have an unreliable waveform? And how does that differ by MST categories? And these are questions that I know haven't been, I don't believe that they have

been queried in the research literature, but I think we heard loud and clear from our patient representatives today that these are consumer concerns. And so if it's not too onerous to add these to the new recommendations, I think it's something we should consider.

Dr. Cassiere: Okay. Dr. Lanzafame.

Dr. Lanzafame: Yes, hi. I want to applaud the last speaker's comments. They were to some degree what I was trying to address earlier this morning, when I was talking about things like red shift. The patient representative with two children, if you recall, said that in order to basically cure the problem in the real world, she had to turn the device off. Aka the device overheated, oversaturated did, whatever. She then had to turn it off for a period of time in order to reset it. The other individual was having a difficulty with the read. I think in both instances, part of the issue is we've come to expect an instant function, and many of these electronic devices do have a burn-in (phonetic) period until they reach a steady state. And then they also have a period of time when you exceed that, and either the sensor oversaturates, or the LED on either end of that overheats, and changes the dynamics. So those are very important considerations that also will weigh into the equation.

Dr. Cassiere: Ms. Brummert.

Ms. Brummert: I thought I was on mute. Never mind, I was going to bring this up later, in the questions about the consumer perspective, but Dr. Lewis opened the door for me so I'm going to bring it up now. I use a pulse oximeter because in 2020, I got the original strain of COVID-19. I almost died, actually. So, this is a vital sign that is used to make medical decisions about whether I need to go to the hospital or what kind of interventions need to be made. So, like the patient representatives, they bought this thinking that it's going to be accurate. Now, I know there's a lot of people like me who have long Covid, basically. So it scarred my lungs. I'm probably going to have this forever. And I know that, consumers everywhere think that this is going to be an

accurate reading and the average consumer doesn't know that there is an issue with racial bias or skin pigmentation. So, five percent of this conversation has been through the doctor's perspective, through people with PhDs, people way smarter than me. But the average consumer doesn't know that this could be a liability in their health care. So, I have two recommendations. One, I think for over the counter, I think it needs to be pulled off the market because we're looking at a dangerous situation here. Or, we should put this behind the counter where pharmacists can adequately explain what the issues are within this, and then the consumer can make an informed decision about it. So, for me, I'm coming from a consumer perspective, and I want to just rope that conversation back in.

Dr. Cassiere: Thank you. Dr. Feldman.

Dr. Feldman: Yeah, I want to come back to this comment about post-market surveillance versus what we require pre-market. I don't want to lose sight of the fact that this device is extremely beneficial to patient care on a day-to-day basis, and has been for decades. So there will be an inevitable tension between the cost and complexity of new pre-market requirements, and the impact on access to the device, either by cost or whatever factors would happen. So, I think we need to be mindful of that. The metaphor I'm thinking about is, the miles per gallon assessment in cars. It's never what the car shows in government testing, and it's because they can't control for the way the driver drives it, how inflated the tires are, all the factors that ultimately-- So, it’s not only probable that there will be a difference in the pre-market results, it's virtually guaranteed when we compare it to real world experience. So, I think the FDA needs to come up with credible techniques, methods that address the disparity based on skin tone. And then introduce post market surveillance requirements to ensure that whatever those performance requirements are shown, actually play out in the real world. It's going to be very difficult to guarantee that real-world performance in pre-market studies.

Dr. Cassiere: So, any other comments? Dr. Taylor, anything along this line you wanted to add? Dr. Wiswell?

Dr. Wiswell: Yeah, just a quick comment. Dr. Fawzy's bringing up the variation that he had in really dark-skinned individuals in their study at Hopkins. That concerns me, and I just think, whether it's oversampling in the same dark-skinned individuals that are part of this study group, the pre-market study group, or expanding the sample size somehow, personally, I think those should be considered.

Dr. Cassiere: Dr. Saville.

Dr. Saville: Ben Saville. I just wanted to say that I agree 100 percent with Dr. Feldman. The idea that you have these studies done in controlled settings and healthy volunteers, and now we're going to use them in the real world, this is the perfect situation in which you have some criteria for pre-market for approval, and then you have some, you have very rigorous post-marketing requirements that allow assessment of whether this actually applies in the real-world setting. So, I agree 100 percent of what you said.

Dr. Cassiere: Dr. Gooden, any additional comments about how they can improve this pre-market evaluation?

Dr. Gooden: I pretty much echo what has been said. I do believe, and I agree with one of the earlier panelists who said that looking at more of the darker spectrum of patients, I think that more focus has to be put on that. So certainly, I think overall increasing the size of the patient population that we look at is necessary. And even if it means going into certain areas where we can get those patients, I think that's something that also has to be considered.

Dr. Cassiere: Thank you. Dr. Brown, any additional comments? Agree, disagree, anything?

Dr. Brown: Agree. No additional comments.

Dr. Cassiere: I'm going to Dr. Adam. I'm going to try to summarize any additional items for these pre market studies. It seems a lot of us obviously, it's been going on for a while, are really focused on this perfusion issue. If there's any way to study that as a part of the pre-market approval process, there are ways of checking perfusion in healthy volunteers, whether it's negative lower pressure devices, whether it's a whole host of things. That was one consideration, to help uncover the disparities in pulse oximetry with dark pigmented patients and low perfusion. That's one thing. The second thing that I think the committee came up with was the mention of, again, maybe skewing the other way in terms of having, favoring more dark pigmented individuals in these studies, pre-market studies, to assure that we have no disparity. And then the other item is, and Dr. Feldman,

I love the using this example of the gas mileage in cars, because I wish my car got the gas mileage that it was promised, but it does not, is to do some type of post-market surveillance, to assure that the new recommendations that come out and go into effect, actually help the patients that it's intended to. And I think that's a fair summary of what the committee had mentioned at this point. Is that enough for you? Would you like us to delve a little bit more? Dr. Goldman has his hand up. I must have missed something.

Dr. Goldman: I just wanted to clarify that we don't want to end up where we've been, which is we want more post-market evidence but then it's not collected in a manner, or can't be collected in a manner, that's informative. It just raises more questions. I just think that has to be part of the thinking here at the same time.

Dr. Cassiere: Great, thank you. And Dr. Yarmus.

Dr. Yarmus: Yeah, Lonny Yarmus. Just to follow up on that, I guess the question to think about is this really post-market surveillance or is this an additional pre-market surveillance, which is, I think where generally the concepts lie here.

Dr. Eydelman: So, let me just make a statement and then another question. So, I think that there is a little bit, perhaps a confusion between the terminology. I think what I heard clearly is a desire from a number of panel members to see a real-world data, or the data in real world use. So, versus post-market surveillance which implies a particular, each particular device. So my question, if you don't mind, especially in light of Dr. Goldman's and Dr. Feldman's comments, just to combine them, can you envision some kind of a real-world evidence study that can address some cross cutting issues? And that will then minimize the need for each manufacturer to potentially do a post-market surveillance.

Dr. Goldman: Yeah, should I respond?

Dr. Cassiere: Yes, Dr. Goldman, please.

Dr. Goldman: I think there have been assumptions, in pulse oximetry, more than assumptions, that there's a body of understanding that has emerged in terms of the physics, these advances that have been understood in terms of the different components and how they absorb, the different absorbers. There's been a body of evidence in pulse oximetry that's been shared, and then manufacturers have applied their special magic in different ways to improve performance or to understand it. It seems that there may be an opportunity to learn more about what occurs in the real world in terms of what might be introducing errors in pulse oximetry. I think that's an assumption right now. and it is important for us to delve into that. If there are phenomena that are occurring in the real-world clinical measurement that end up becoming part of the shared body of science in pulse oximetry, then that would improve instrumentation and patient care across the board.

Maybe there need to be different solutions to it, just as there are today, with different probes, sensors, and so on. So I don't think we can say categorically, but I think that there's an implication from many of those who have been publishing and sharing information, that there

probably are things that need to be understood, that might be affecting all instruments from different manufacturers, maybe not to the same degree. Did I speak to the right, was that the question? I want to make sure I--

Dr. Cassiere: Yeah.

Dr. Goldman: Okay, I tried. Thank you.

Dr. Feldman: I would just add a couple other comments. I certainly agree with Julian and I would be reticent to recommend much in the way of real-world evidence in a 510(k) submission, again, getting back to my point of increasing cost, complexity, et cetera. I don't think that's necessarily the way to go. I do want to commend the FDA on the involvement with the prospective studies, and there are others underway. And I think as those play out there'll be information learned by everybody on how to improve the technology. And so, we'll probably have a better answer to that question in six months, a year. But, if you look at the cost and complexity of those studies, to try to reproduce that and build that into a 510(k) submission, would just, it would I think kill the access to the technology and I wouldn't want to support that.

Dr. Goldman: And I should have said that explicitly. That was not my intent whatsoever, that that be part of pre-market evaluation, at all. It really aligned with what Dr. Feldman is saying.

Dr. Eydelman: We can now move on to the next question.

Dr. Cassiere: So, yeah, so Dr. Saville has his hand up, before we go next.

Dr. Saville: I just want to bring up one concept that may or may not be useful here, but in pivotal trials, we do a lot with what are called master protocols, or platform trials, where instead of having each company do their own separate trials, so if you're going to post market surveillance, instead of having each company do their own thing, you have a single master protocol that allows treatment arms to come in and leave as things become available. And the problem with that is you need someone willing to organize that and run that kind of study. Maybe it's an

academic group that gets funding from NIH. But those kinds of trials do exist, that could add some rigor to a post-market surveillance of devices that do get approved.

Dr. Cassiere: So, before we move on, just to speak to what Dr. Feldman said, before Dr. Eydelman, in a kudos to the FDA for actively getting involved in funding prospective studies that will clarify further pre-market modifications, so be it, to assure that the rules and regs that come out are helping patients long run. So that is kudos to the FDA for that, and that's a plus. So we're going to jump to question two now, everybody. If question two can come up, and I think, if memory serves me correct, we have a part A, a part B, and a part C. Thank you very much for that.

Question 2

Dr. Cassiere: So, question two. FDA is considering defining non-disparate performance as the estimate of the absolute difference of pulse oximetry bias across ITA and MST levels, if the difference is less than 1.5 percent when the oxygen saturation is greater than 85, and less than 3.5 percent when the O2 sat is less than 70 and less than 85. Here's part A. Please discuss the advantage and challenges to the proposed non-disparate performance definition. So, we'll start there, and I think we talked a little bit about this, but Dr. Wiswell. You're muted, Dr. Wiswell.

Dr. Wiswell: Yeah, I'm just saying as a neonatologist, my patient population is clearly different from the adult population, and even the older child population. I have children that will have obviously the anatomic critical cardiac problems and live in the lower saturation range. I have babies with pulmonary hypertension that we're making strong decisions based on saturations, whether it's an emergent kind of surgery for a cardiac patient, whether it's trying to do a cath lab, trying to open up a big hole between your atria to try and get some shunning. Or if I have a bad pulmonary hypertension patient, whether it's to go on ECMO, front lung bypass. And so the saturations' absolute difference just seems too wide for me, if it's less than 3.5 percent, and the

lower saturation range, because I have patients that are not infrequently in the 70 percent to 85 percent range, and sometimes have wide variations and saturations that I am measuring with blood samples and what I'm reading with pulse oximeters. And personally, I would like to see that number go down and be-- unless there's a big technological problem (phonetic), be closer to that 1.5 percent or maybe a 2 percent absolute difference.

Dr. Cassiere: Just a question back at that, Dr. Wiswell. And again, I'm not a neonatologist. Do you commonly correlate peripheral pulse oximetry with an arterial blood gas in each patient, to make sure that the corresponding levels match?

Dr. Wiswell: If we can, but again, we would correlate at-- some patients we have difficulties getting an arterial line in. It would be with a puncture, and so it's after the fact. Some of the patients with the kinds of shunts that they have, whether it's with a cardiac disease or whether it's with the kind of pulmonary hypertension that we can see with sepsis, hypoxia, et cetera, they can have variations that are rapidly occurring and that I don't have easy access to an arterial sample to try and corroborate what I'm seeing on a monitor. And so, because I have those patients in that range, I want to make sure that I'm a little bit closer than-- and have as much comfort that I do if their saturation’s 85 to a hundred percent range.

Dr. Cassiere: So, in your patient population you would like that to be tighter, just like our sleep apnea colleagues mentioned that before, that you'd feel comfortable with-- The 3.5 seems a little too high. You want the lower range, and that was mentioned I think in the Nellcor presentation, or Masimo, I forgot which one. But we'll note that, and we'll discuss that a little further. Dr. Feldman.

Dr. Feldman: I have a question for our FDA colleagues about the terminology in this question. So, it's worded as bias. Bias is an average difference, and doesn't reflect the precision of the device. ARMS is used to encompass both bias and precision. If we use the term bias without

considering precision, then the difference between measurements can be quite a bit larger than what's specified in the question. So, my vote, in rewriting the question, would be to replace bias with ARMS, and then have a discussion around what clinical performance is desirable in different clinical settings. For my practice in anesthesiology, if ARMS was probably under 3%, in almost any category of patient, I can probably manage with that. But I'm at the bedside, kind of minute to minute. Different from the ICU setting, patients are not monitored minute to minute with a bedside care provider, different clinical decision making, the NICU, et cetera. So, but anyway, can we just clarify the terminology in that question, that bias is indeed what's intended, or would ARMS be a better term?

Dr. Eydelman: Sorry, I was muted.

Dr. Cassiere: Yeah, Dr. Eydelman, any comment on that? Because I don't remember reading bias in that question, or did I miss--

Dr. Eydelman [indiscernible] the question again. Bias is what was intended, and I'm going to ask Dr. Pennello to come on camera and explain.

Dr. Pennello: Yeah, this is Gene Pennello. So, for that question, it's bias is what was intended there. And I want to clarify that there's co-primary objectives that both have to be met. One is the overall ARMS, less than three percent with statistical significance. And the second primary objective is to show non-disparate performance in terms of bias, in terms of the difference in bias between skin color levels as measured by ITA or MST.

Dr. Feldman: Dr. Pennello, can you just educate me a bit, if the bias requirement is larger than the ARMS requirement, that doesn't seem consistent to me. Or what am I missing?

Dr. Pennello: Well, there may be some confusion. The goal here is to show that the differential bias is no bigger than three and a half. So, in that particular SaO2 intervals, for example, the bias may be two percent plus two percent in darker skin tone. And it might be negative one and a half

percent in lighter skin tone for a difference of three and a half. Now both of those biases in absolute value are three percent. It's just the difference in bias is two minus one and a half, so that's three and a half. So that's fairly large. And that's exactly the performance goal there. Now, you may wish to have a different criteria than the difference in bias. You may want to suggest that the absolute bias has to be less than some goal, all across the board, of all skin color values.

Dr. Eydelman: Thank you, Dr. Pennello. Perhaps Dr. Hendrix can add a little more on this, as well as on the IFU and the fact that it is not playing a specific clinical indication at the moment.

Dr. Hendrix: So, I can attempt to answer this as how I understand it, as a clinician, Dr. Feldman. So, knowing that SpO2 bias is the mean difference between SpO2 and SAO2, what we're looking for, the difference of differences, is really about the SpO2 difference, between those who have the highest maximum difference, two sets of populations. So, when we're looking at ITA, we said between maximum and minimum. And what we're really looking at is the SpO2 difference, the mean SpO2 difference, between whichever population that happens to measure maximally or minimally. We have seen data where it's not always the darkest and always the lightest who are at those extreme ends. Sometimes it's those with middle tones who have a very positive bias. So, that is why we crafted it across ITA as a continuum. Hopefully, that answers the question. Thank you.

Dr. Cassiere: You're muted, Dr. Feldman.

Dr. Feldman: Feel free, Dr. Cassiere, to terminate this conversation if you think it's going on too long.

Dr. Cassiere: Oh no, not at all. This is exactly the conversation they want and need, and I think the fact that Dr. Saville has his hands up tells us we're on the right track.

Dr. Feldman: Well, so let me just make one more point about it, and what I think about is the Bland-Altman plots, where the bias is the average of the difference. And the precision reflects the

statistical deviation of the measurements in the population that you've studied. So, there's bias, as you've described it, which is a mean difference. But when you study a population of patients, a lot of patients, based on whatever the precision is, are going to be beyond that bias. And so that's why I'm trying to get some clarity around this particular question, because as a clinician I would like to know that two to three percent is what I'm going to get out of the device, not two to three percent on average. But in some patients it's going to be six to seven percent because it's imprecise. And so that's what I'm getting at.

Dr. Cassiere: And I think, Dr. Pennello, the 95% confidence interval, does that help alleviate any of this?

Dr. Pennello: Yeah, Gene Pennello, FDA. So for the non-disparate performance assurance, we're actually only looking for the point estimate to be less than 1.5 or 3.5 percent of the maximum difference in bias. We're not looking at competence intervals there.

Dr. Cassiere: Okay. Dr. Saville.

Dr. Saville: Yeah, Ben Saville. So, I mean that really brings it back to what my point was earlier, is this whole idea of the non-disparate performance analysis is looking at a point estimate of the maximum bias versus the lowest bias of these two extremes of say dark skin or lighter skin. And I think I'd be very interested to see what you could show with confidence intervals and those kinds of estimates. If you can estimate a difference between those, what kind of precision would you have around those? And if it's insufficient, would increasing your sample size help with that? I'd be very interested to see that. One thing I don't love about this definition, these criteria, is that I have the feeling these were calculated. These criteria were decided, essentially back calculated based on what's feasible in this population with a small sample size, to provide 80 percent power. And so it's this kind of circular justification that, okay, well it gives you 80 percent power, but usually it's, okay, what I have an effect size, what sample size do I need for 80 percent power?

So, it's all kind of this circle in terms of the way we're justifying this. And I'd rather see what's the scientific question that we want to define. And I know we're trying to show the non-disparate performance. Is there a better way we could quantify that, in terms of estimating what that disparate performance is, in terms of an estimate, some sort of precision? A Bayesian probability, for example, I think that would be helpful. I'm worried here that you're going to end up with trials that are more in the gray area, meaning inconclusive results. Maybe they meet one of these criteria, but maybe not the other. This kind of arbitrary, well I know it's not arbitrary because I know you came up with these intervals for a reason, but 70, 85 percent, then 85 to 100 percent, you're taking something that's very continuous and you're saying, all right, we've got two different groups, we're going to have two different criteria.

And I worry that you end up with trials that are in this gray area. And where it's just inclusive, where maybe the criteria aren't met exactly. But hey, if you look at it this way, this is based on a linear model and the linearity is suspect here. But if we look at this nonlinear model, look, now we meet that 3.5 percent threshold. And so, and maybe you're not worried about that so much, and maybe you're happy looking at the totality of the evidence rather than sticking strictly to those performance criteria, but I worry that these performance criteria are not going to be adequate to really say, yes these are non-disparate between different levels of skin pigmentation.

Dr. Cassiere: Dr. Goldman.

Dr. Goldman: Yeah, thank you. Julian Goldman. I’m not sure I can shed light, but I am concerned that the way we're phrasing these terms seems to me to be causing quite a bit of confusion. And we, in the comments today, we've heard comments about the importance of a certain level of performance or accuracy, to track a desaturation or to measure saturation accurately. But this is clearly focused on what is being called non-disparate performance, which

in and of itself is a term that I think is hard for a lot of people to grasp because it's a pretty new term. And maybe if we explain that we're talking about reducing bias that's attributable to skin pigmentation, and that's what the whole discussion is about, and we see all the questions in literature repeatedly, that that's what this is about. It's the difference in performance that we're looking at, right? It's that differential bias. Because then some of the discussions and conversations probably don't apply, and others might even be more critical. So, I just want to flag that, that I think this new terminology is adding to what is already a complicated topic, and new material for lots of folks. Like I said, it may not help, but I think that may be part of the problem. You know, non-disparate performance, right? People see performance, now we're thinking about performance of the pulse oximeter overall, as opposed to the difference in performance, across a full range of skin pigmentation. That's it.

Dr. Cassiere: Anyone have any other comments along this line? So far what I've gotten is a lot of uncertainty. But one thing that kind of stands out is a couple of panel members made mention of the breaking up, why should we have a 1.5 with greater than 85 percent, and a higher error threshold, I guess I'd call it, for what some of us would think would be even more clinically significant desaturations. Do we feel comfortable, and excuse me, I forgot who the panel member was who mentioned-- I think it was Dr. Wiswell, you'd mentioned, is there a technological reason why we can't tighten up that lowered desaturation threshold to less than 3.5, to something along the line of 2.5 or two. Does anyone have any comments about that? So in terms of the please discuss the advantages and challenges to the proposed non-disparate performance definition, I think we're-- I'm not sure if we had a satisfactory discussion for the FDA on this. Dr. Eydelman, you want to weigh in on what your thoughts are?

Dr. Eydelman: Yes, thank you. Please move on.

Dr. Cassiere: Okay. We're going to Part 2B. Please discuss alternative acceptance criteria for the agency's consideration. Should they be thinking anything else? Does anyone want to weigh in on this, or do you think we flushed this out and there's no other discussion at all? Dr. Goldman has his hand up.

Dr. Goldman: Yeah, thank you, Julian Goldman. I would like to raise a concern that, in an area that is evolving so quickly, and we've raised so many questions, that disclosure of performance might be equally important, potentially more important, I have to think this through some more, but having acceptance criterion, pass fail criterion, across a range of skin pigmentation, when we, we keep raising questions as to the root cause at the optical level or physiological level, I find a bit concerning. I find it concerning that we may cut off performance of devices that may provide useful information. Now, I don't have enough information to really know for sure what the implication will be, but disclosure of performance is vitally important. And it's often how we manage with a lot of our instrumentation and our laboratory tests. We know to expect decreased performance in some areas. We don't say we won't perform the test. We may say there's higher uncertainty. We do that all the time with a whole range, including glucometry, whole blood glucometry. So, I have a nagging concern that by not emphasizing disclosure, and by going to a pass-fail criterion, of something where the root causes are not well understood yet, we could be throwing out the baby with the proverbial bath water. That’s it.

Dr. Cassiere: Yeah, Dr. Goldman, just a comment on that. Again, hopefully memory is serving me correctly. During Dr. Pennello's, one of his presentations, he had looked at devices that were currently available and if they passed or failed on this new criteria, and I remember, I believe most of them passed this pass or fail. Is that correct? Or am I incorrect?

Dr. Goldman: I think that was very great to hear from him, but I think it was very limited data and some of it may have been modeled. It wasn't as if we looked at the entire-- So, but I'm not sure. That was my impression.

Dr. Cassiere: Dr. Pennello, could you expand on that? That could help us out.

Dr. Pennello: Yeah, Gene Pennello, FDA. That was real data, it wasn't simulated data. On 12 devices, with the ITA measured at nine different locations, and that's just one data set. But that was one analysis that seemed to show that if you had an ARMS that's fairly low in the 2.0 to 2.4 range, you're likely to meet the non-disparate performance goals at most of the locations.

Dr. Cassiere: Does that help out, Dr. Goldman, or no?

Dr. Goldman: To clarify locations, I'm not sure what, what's being referred.

Dr. Pennello: Well, unfortunately, I don't know. These were the different locations at which ITA was measured, so it could have been forehead or inner arm or the digits of the finger at particular locations. But unfortunately with that data set I don't know, those nine locations, where exactly they were.

Dr. Goldman: Again, as I said, I can't speak specifically on this, but it's 12 devices, and we are talking about pass fail criterion, whereas in many other areas of medicine we recognize when there are performance limitations, and we deal with it. We want to know, it's vitally important. And I just wonder how we're approaching this in terms of the pass fail, as we're learning. Again, I may not have all the data. There's a lot out there. But it's different than how we normally deal with things, it seems.

Dr. Cassiere: Let me go to Dr. Eydelman, first. I'm sorry. Dr. Eydelman.

Dr. Eydelman: And on the FDA level, [unintelligible]. But, Dr. Goldman, I just wanted to bring to your attention that all of pulse oximeters do have labeling that conveys the performance. And it is up to you today to recommend any additional labeling, meaning inclusions that you

recommend. So, what our question is, focuses on which devices should be allowed to market. Now, once they are in the market, we have labeling to communicate their performance. So just keep that in mind in light of your comment, that it's the recommendations you're making are about the first step, which is getting to the market. And then there is labeling, and you can suggest how you want the performance communicated in the labeling if you don't believe our current way of communication or performance is adequate. That's all. Thank you.

Dr. Cassiere: Dr. Ballman.

Dr. Ballman: Yeah, a couple comments. I mean, I think it was pretty clear that people aren't understanding this disparate measure. And what it's trying to do is to say how different is the performance and bias between the two groups. Okay, it's not an error thing, it's just how different does it perform in bias between the two groups. But given that everyone's having, and I'm having a hard time understanding it, I think it brings it back to sort of the question that if we really want to know differences between two groups, there are other ways of doing this which we had discussed previously. Now in terms of a yes or no thing, I think if we want to make sure that we have instruments that work on all skin tones within-- and it has acceptable error across all skin tones. I do think that there needs to be an acceptance, yes or no, for the instrument, to force manufacturers to make sure that their instrument performs well across all skin tones. I think if we allow them to say, oh this is the performance that works really well in light skin tones, here's the performance in dark skin tones, it doesn’t work very well, then we’re not going to get an instrument that works well in dark skin tones.

Dr. Cassiere: Dr. Wilson.

Dr. Wilson: Yes, William Wilson. In terms of alternate ways of looking at this, we might want to just consider simplification. And I think it's not just the terminology, but you know, a bias of the bias. It may just be easier to look at what is the actual mean difference for those individuals in

the dark skin category, compared to those in the white skin category. And if that difference is more than some number that the agency believes is clinically significant, then it would be non-disparate performance. And so, what would be a threshold, that the agency would need to think about this and model, but probably something more than 0.25 percent or 0.3 percent, starts to get into the range where it may be clinically significant. Anything less than that, or certainly in that range, is probably not clinically significant. So, but that would be much simpler.

Dr. Cassiere: Dr. Brown.

Dr. Brown: I just had a question about, if we do go with this sort of yes-no, meeting criteria, for devices that are already out on the market or the sort of common devices, not the medical grade, is there a way to add the equivalent of a black box warning to the label, to raise awareness, or is it in within the authority to actually remove non-compliant devices from the market? I know we can probably prevent new devices, but just wondering.

Dr. Eydelman: Yeah, so we would like to collect as much information as for your recommendations today, and then as an agency we will go back and figure out the best way that we can proceed to maximize public health impact.

Dr. Cassiere: So just to-- oh, sorry, Dr. Saville.

Dr. Saville: Yeah, I'd just like to piggyback a little bit on Dr. Wilson's comment, that I think that's exactly what the model is doing. So, I think the model that Dr. Pennello suggested, basically it assumes linearity, and basically what you're going to do is you're going to compare the bias of black patients versus the bias of white patients. And at that absolute difference, if that model-based difference is less than 1.5 percent for a certain oxygen saturation then that criterion is going to be met. Or if it's less than 3.5, it's under the 70 to 85 percent it's met. So, I think that's essentially what the model is doing, is doing it with more granular terms.

Dr. Wilson: Yeah, I was just thinking that it might be to all of our, to the agency's benefit, to have a simplified solution, rather than something that is as complicated as currently being proposed.

Dr. Cassiere: So, Dr. Wilson, I guess what you're trying to say is why do we need the word disparate?

Dr. Wilson: No, the word disparate is something that, it's a new term, but that is easily understood. But, what we're really interested in, is what's the difference between SPO2 and SaO2 on average, the mean bias, in the dark skin versus the light skin. And if it's clinically insignificant, then that, and you have to look at what that threshold is, that's a lot easier to understand. You'll approach it either way.

Dr. Cassiere: I guess what I'm getting at is, these criteria need to be met for every skin tone. So why put in that it's going to be disparate when the expectation is in every category, it needs to meet these criteria. Or, my mistake--

Dr. Wilson: First, in order to do that with-- to show that statistically with every skin tone, you will need larger numbers. And right now, the idea was have three bins, and ensure that there's non-disparate between those three bins. And earlier I was suggesting dark skin tones versus light skins in the less than 0.25, or something like that, would be reasonable. Across bins, maybe that number should be less than 0.2, or something like that. But that would just be, it's just a simpler way of approaching it.

Dr. Saville: So, I'll just add real quick to that, if you don't mind. But the model right now is not saying, hey we're going to take this bin versus this bin. It's a model. Okay, ITA is continuous. It's an ordinal value on the X axis and you have this linear model. here's the bias for high values of ITA. Here's the bias-- I'm sorry, low values of ITA. Here's the bias for high values of ITA. And let's look at the difference between those two things. And so, it's not explicitly saying here's the difference between these two groups. It's saying here's the model-based estimate, which has more

power, certainly in terms-- we're not looking at precision here, but the model is going to have better estimation properties than trying to get a single estimate within each bin by itself. So, that's my comment.

Dr. Cassiere: Ms. Brummert.

Ms. Brummert: So I just want to piggyback on what Dr. Brown was asking. I don't know if we're referring to over-the-counter products, or medical-grade hospital. But the average consumer is not going to read a black box warning. They don't know that it exists. They don't know where to look for it. So, I don't think that's a viable option, but I do think pulling (phonetic) from the market would be.

Dr. Cassiere: Yes, a great point. So, question three is going to focus on the over-the-counter, and we're going to really delve into that. But your point is well taken. These are for the-- my understanding is the medical grade, that need to do pre-market approving to the FDA. So, I'm going to go around a little bit again. Dr. Gooden, anything else to add to this, or you're satisfied with what everyone's talking about? Or agree or disagree?

Dr. Gooden: I pretty much would say I agree with what has been said.

Dr. Cassiere: Great.

Dr. Gooden: Nothing else to add to at this point.

Dr. Cassiere: And Dr. Taylor?

Dr. Taylor: I agree. No further comments.

Dr. Cassiere: Dr. Lewis?

Dr. Lewis: I agree.

Dr. Cassiere: Dr. Brown? Dr. Wiswell.

Dr. Wiswell: I agree.

Dr. Cassiere: Dr. Feldman.

Dr. Feldman: Yes, I agree, but my only reticence is that I'm still a little bit confused in our conversation, in that at the end of the day, when the clinician puts the device on the patient and the device has gone through this process, you should have some understanding of what the measurement is. Wow good an estimate, not measurement, how good an estimate that is, of what's really going on with the patient. And so, I come back to a little bit what we were talking about before, that I'd like to know that that estimate is pretty likely within two to three percent of the actual SaO2, regardless of skin tone. So, however we arrive at that determination, I think, is important. Thinking about what's going to happen at the bedside after this device is approved.

Dr. Cassiere: Dr. Yarmus, I see you gave us the thumbs up.

Dr. Yarmus: Yep, nothing new to add here.

Dr. Cassiere: Dr. Lanzafame.

Dr. Lanzafame: Agree.

Dr. Cassiere: So, Dr. Eydelman, just to summarize, it appears that alternative acceptance criteria, we're pretty much okay with what was been proposed. And I don't think we have anything to tweak that with. Are you satisfied with that discussion or you wanted to delve into something else?

Dr. Eydelman: No, thank you. I'm satisfied. I just have one tangential question. I heard I heard your recommendation of binning, and I was wondering what would be the delta between the bins. I heard some numbers being thrown around. Is there a consensus on what would be the performance between the bins? In that, if we were to think about that approach, which is obviously not what was proposed at the time, but--

Dr. Cassiere: So in other words, the difference between the one to four and the eight to 10? And the intergroup differences, what would be an acceptable variance?

Dr. Eydelman: I mean, I don't want to spend too much time, but if the panel members have a recommendation, I'd be interested to hear.

Dr. Cassiere: I'm going to lean on Dr. Ballman and Dr. Saville for this, to start out.

Dr. Ballman: Well, I think it's more of a clinical question. I mean, so if you know the mean bias in the light skin is some number and the mean bias in a different bin, a darker skin, is a different bias, and they differ by, I don't know, some number, what number would have you concerned?

Dr. Cassiere: That's throwing it right back at us. Okay.

Dr. Ballman: Well because it's more of a clinical question than it is statistical. We can do the stats around whatever number you give us, but--

Dr. Saville: And I agree with that, in the sense that if your dark skin is a plus two percent bias and the white skin is a minus two percent, that's a four percent difference. Do we care? Well, they're both within two percent of the bias. how important is the conclusion that, hey, there's something different between the dark skin and the white skin? And is that relevant for, they're both off by the same amount, just different directions.

Dr. Ballman: But the direction might matter, in terms of making clinical decisions. So…

Dr. Cassiere: Dr. Lanzafame.

Dr. Lanzafame: I think that's really the concern, that you can demonstrate that there's a statistical difference between two populations, but where the rubber meets the road is, at what point does that difference become clinically relevant to both of those groups? And those are really important. Two very different drivers, in terms of, is it relevant? At what level is it meaningful?

Dr. Cassiere: Dr. Feldman.

Dr. Feldman: I think the question needs to be asked a little bit differently. So, in my view, it's not demonstrating that the difference between the groups is acceptable. I would like to know that for any-- non-disparate, to me, means that for any given patient, irrespective of their skin tone, I will

get a performance within limits that are clinically acceptable. So we define what the clinically acceptable limits are, we study the populations, and then we compare the dark and the light skin, and we prove statistically that the differences are not important within the range that we're trying to test. So they may be wider, perhaps, in the dark skin, but still within clinically acceptable limits. Narrower in the white skin, perhaps, but still within-- But the performance across those groups is non-disparate, or what we want it to the device to do at the bedside. That, to me, is the important question, not the whether the absolute difference between the groups is within a certain margin.

Dr. Cassiere: Dr. Brown.

Dr. Brown: I was just going to throw out two percent as the clinically meaningful difference, not between the groups, but between the real value and what's being reported on the device. Just as a starting point.

Dr. Cassiere: Well, I think Dr. Feldman's point is spot on. I want it-- and in my individual patient, I'm colorblind, I want to know if I can trust that pulse ox. Do I really want to know what the intragroup variability is? With the pertinent thing is clinically, I want to know if that pulse ox is reliable in that particular patient. And looking at different skin tones, and following the criteria that the FDA has, we'll put that on the table as answering yes, hopefully. Does that help, Dr. Eydelman?

Dr. Eydelman: Yes, thank you very much. Dr. Feldman, basically you're proposing the same approach that we proposed originally, just the only difference is with binning. So thank you. Yes, please move on.

Dr. Feldman: Well, binning and also comparing performance between the groups, because I didn't see that in the statistical proposal earlier.

Dr. Eydelman: Thank you.

Dr. Cassiere: Alright, so this brings us up to 2C. If we can have the question up there again, that would be great. Okay. So, Part C is please discuss if there are specific SaO2 thresholds for which the accuracy of the pulse oximeter for detecting hypoxemia should be analyzed. Who would like to start us off? Dr. Lewis.

Dr. Lewis: I'll just share an important example within neonatology. In the New England Journal of Medicine in 2010, a big NICHD study called the SUPPORT Trial was published, and they randomized 1,300 preterm neonates to lower SAT goals and higher SAT goals, 85 to 89 percent, versus 91 to 95 percent. And what they found is that infants randomized through the lower SAT goals had increased mortality, which was really a surprising finding. So I think this speaks to the importance of that critical sort of 87 to 93 range, especially in the NICU, where we are on mechanical ventilators, we're titrating FIO2 continuously, to keep these smallest patients in very tight SAT goal ranges. So, especially in neonatology, I think that range of the SATs is a place where we need especially important focus.

Dr. Cassiere: Great. I'm in agreement. So I'm going to take it back further, to the 1990 study in Chest (phonetic) by Martin Tobin, that looked at patients who had an 02SAT of 92 percent if you were white, 92 percent of those patients had a PO2 of greater than 60. If you were dark skinned, with the same O2SAT, only 50 percent of those patients had a PO2 of 60. And I raise that because, again, clinically we're all trained that the oxygen dissociation curve falls drastically after 90 percent. So, my number when I saw this was talking about the 90 percent threshold, to look at how many percentage of patients with any skin tone have a PO2 that's an acceptable level at that threshold. And I just throw it out there for discussion. And again, it steps off of what, the whole argument, conservative versus liberal oxygen delivery, I think is settled at this point, at least in adults. And I think it may be true for peens and neonates now, that having liberal oxygen doesn't seem to harm. So that 90 percent is really important to keep your PO2 above 60

millimeters of mercury. And I guess what the FDA is asking, should we have a threshold involved where they look at that as well? Anyone want to discuss that? Dr. Wilson.

Dr. Wilson: Just in terms of liberal oxygenation, and particularly in the neonatal population, although low targets increase the rate of mortality and necrotizing enterocolitis (phonetic), the high levels tend to increase retinopathy, prematurity, and BPD. The BPD may be a little more complicated, may be related to their problem, has positive pressure and so forth, but nonetheless, that's what data there is. And there may be problems with high oxygen concentrations in adults that we're not as aware of, other than patients with bleomycin. So I think having, just for the discussion point, we don't want to have excessively high partial pressures of oxygen, higher than necessary to keep the patient in an adequate range.

Dr. Cassiere: Yeah, I'm going to push back a little about that in the adult population. I think the studies are pretty clear now that this liberal versus conservative oxygen, in terms of breaking PO2s down, is a little clearer in the adults. I'll agree with the neonate. Again, that's outside of my expertise, but I think the focus should be looking at an oxygenation threshold where we want to assure that patients with a peripheral O2SAT of X, we know they're in the safe range, for lack of a better term. Dr. Feldman.

Dr. Feldman: Yeah, I think the conundrum for the FDA here is, we're talking about different patient populations and different clinical needs. If you look in the ICUs, if you look in the operating rooms, it's hard to find patients under 90 percent, because we're actively managing them above that level. You go to the cardiac, pediatric cardiac ICU or the NICU, you got a whole different range of saturations now that you're managing too and trying to detect. So, the question is, do you come up with a performance standard that hits all populations? Do you come up with a minimum performance standard and then let the market decide? So, in other words, let's say the minimum performance standard's not quite good enough for the NICU, but a company comes

out and says, hey we got the best device for the NICU, and here's our improved performance, and then the market then picks up that device. And so, I don't know where the FDA wants to live, in whether it's minimum performance standards or trying to hit everything, but unless we can really decide what patient populations we're after, it's a little hard to give firm advice, I think, on this.

Dr. Cassiere: Anyone else? Dr. Goldman.

Dr. Goldman: I actually am looking for clarification of the question. Are we looking for, if we have to pick a sweet spot for performance, can we name the sweet spot? Are we asking what the performance should be in that sweet spot? Are we accepting that there has to be a trade off in pulse oximetry, that if you have optimum performance at 82 you may not have it at 98? I think the question isn't completely clear to me.

Dr. Cassiere: Fair enough. Dr. Eydelman.

Dr. Eydelman: So, it is for the purposes of the labeling. Because as of right now, we're not asking for that. So, Dr. Hendrix, perhaps you can elaborate a little bit further.

Dr. Hendrix: I'm happy to. It's really asking the panelists whether it would be beneficial to provide additional analysis at important clinical thresholds. For example, 90 percent ACCAHA I think has general guidelines to start oxygen therapy. And, as many of the members here have talked, it varies for populations for disease states, even practice of medicine by different clinicians. So, we are garnering information from you today, whether we should ask for additional analysis at important specific clinical thresholds, and what they would be, and knowing that they may differ for pediatrics versus adults. I hope this has helped.

Dr. Goldman: So the label-- almost there. So what might be an example, make up some numbers, but what might be an example of what that label would state, just for clarity.

Dr. Hendrix: I believe Dr. Pennello showed AUC and ROC curves, Dr. Goldman.

Dr. Goldman: It might be an ROC curve that's included in the company documentation, or in the information with the device? Is that…

Dr. Eydelman: Oh, so let me pipe in here. So as of right now, that analysis is not being asked for. And so, what we are requesting your input on, first whether that analysis should be part of every pre-market submission.

Dr. Goldman: Okay.

Dr. Eydelman: And yes, how should then the results of that analysis be communicated in the label?

Dr. Goldman: Okay, thank you. Now I understand exactly what you're looking for.

Dr. Cassiere: Before I keep going, with Dr. Feldman, so, like an example would be like a rule in, rule out for hypoxemia, threshold with an ROC curve?

Dr. Eydelman: So again, right now, that analysis is not required, or not being performed as a routine part of the submission. So what we're asking is, given its clinical implications, should this be an additional analysis that we ask for? And if yes, how should the results of that analysis be communicated in the labeling so it's easily understood, and I saw your hesitance about presenting graphs in the labeling. So, how should it be then communicated so that the average clinician can easily understand what it says?

Dr. Cassiere: Well I was about to say, should they put in positive and negative predictive values of hypoxemia, but I don't think the average clinician would grasp that. But I'm going to ask Dr. Feldman to, he has a hand up, to answer.

Dr. Feldman: Thank you. Jeffrey Feldman. I'll try to answer the question directly. So, I personally think it's much more important to have a performance standard based upon a range. Bias, precision, ARMS, whatever we pick, than it is to have a threshold that we focus on. And the primary reason that I say that is I don't believe clinical decision making should be based upon

a single point estimate of a pulse oximeter data. So, if I'm getting a saturation of 90%, what I do with that is going to be very context-dependent, and it gets to some of the comments that we heard earlier. So I don't think you need to focus on particular thresholds, more so than accuracy. That said, you're raising the question of ROC curves, and I do think there is value to information around diagnostic performance like ROC curves. I personally favor Clark error grids. I think they're a better disclosure, but I don't want to necessarily widen the discussion right now on that. So I think there would be value to including some information about diagnostic performance, but I don't think you need to focus on a particular threshold. I would focus on an accuracy range that you're trying to hit.

Dr. Cassiere: Although, Dr. Feldman, I will say most of COVID therapy's based on a threshold, whether that was correct or not. So, I guess, should we go down that pathway, or ignore, because clinicians use it to initiate therapies?

Dr. Feldman: Well, I mean, that may be the practical consideration, but I think the reality is if you set a threshold, you're going to be wrong in some patients and right in others. And so, you hope that you're right on most of the patients. But I think if you have a device that performs let's say within one or 2%, let's say 2%, and that's well documented in the labeling and it's something you can hang your hat on, then it becomes easier to make clinical decisions. So the patient that's 90 percent and, and you're at that decision threshold, but your resources are overwhelmed and they actually don't look that bad. Maybe you can sit on for a little bit, as opposed to the patient who they were 94 10 minutes ago. Now they're 90, they're gray, they're struggling. Different story. So, I don't know if that answers the question.

Dr. Cassiere: But you're on reliability (phonetic). Not a number, not a threshold. Okay. Dr. Goldman.

Dr. Goldman: Yeah, thank you. Now that I have a better understanding of the question and giving it a little bit of thought, the challenge could be that deriving ROC curves and disclosing information based solely on the volunteer laboratory desaturation trials, knowing that there is more uncertainty in the clinical measurements in patients that are potentially quite ill and whom we would care about applying the ROC curves, it might be compounding-- we might be conveying a level of certainty that doesn't belong there, right? We'd be building something on a somewhat weak foundation. I'm not sure, but it feels that way, and for all the reasons that we've discussed today. So, sometimes when you bury information deeper and you start to build nice graphs and tables and charts, we've seen this all the time, we've all dealt with it, people forget about what the real base of that foundation is, and they don't recognize that the assumptions may no longer be true when they apply the information. Thanks.

Dr. Cassiere: Dr. Lanzafame, you had your hand up, but I see it's down. You have a comment or a question?

Dr. Lanzafame: I think that's already been discussed. At least in terms of the real world reading the labeling, and certainly on the OTC end of it, if there's a major outlier then there's some problem with the technology. But it doesn't make sense to say, okay by God at 90 percent the Delta is 2 percent, at 91 it's 2.1. I don't know.

Dr. Cassiere: Dr. Wilson, what are your thoughts?

Dr. Wilson: I don't really have more to add, I think.

Dr. Cassiere: Alright. Dr. Lewis?

Dr. Lewis: No, I agree. I think for the bedside clinician, they don't really want to know it works great at 90 but not great at 95. So, I agree with the prior comments that we should have a standard that it works well across the ranges.

Dr. Cassiere: Dr. Brown.

Dr. Brown: I agree. Keeping the label simple makes sense if you want to require a little bit more performance data within that critical 87 to 93, for the companies to provide the data to the FDA, that makes sense. But I wouldn't put the details in the label.

Dr. Cassiere: Dr. Wiswell, you're okay with this?

Dr. Wiswell: Yeah, I wouldn't put the details in the label. I wouldn't (phonetic) come up with our own threshold for hypoxia. Stevie Wonder was subjected to liberal oxygen in 1949. He ended up blinded, and they started restricting oxygen in babies in the 1950s. We ended up with a lot more deaths and cerebral palsy. So I would not put labels in our own interpretation of what we want to be hypoxemia from this group of adult volunteers.

Dr. Cassiere: Thank you. Dr. Yarmus.

Dr. Yarmus: I agree. I'm fascinated by the Stevie Wonder story. Thanks.

Dr. Cassiere: Me too. Dr. Gooden.

Dr. Gooden: I would not add anything else to that. I think it would depend on the context of the patient, the age, and what other additional comorbidities the patient might have.

Dr. Cassiere: Dr. Taylor, any comments?

Dr. Taylor: No comments, thank you.

Dr. Cassiere: Our biostatistician group here, Dr. Saville.

Dr. Saville: Nothing to add other than just I think for a range of values is important. I think [indiscernible] a single value show up for a range.

Dr. Cassiere: Dr. Ballman, agreement?

Dr. Ballman: Well, yeah, I agree. I mean, my concern is, these are values and healthy people and it's being applied in the hospital, and I don't think… Yeah, I agree with everyone.

Dr. Cassiere: Ms. Brummert, anything before we move on?

Ms. Brummert: No, I have no additional comments.

Dr. Cassiere: Great. Okay. So, Dr. Eydelman, everyone's okay with the current criteria, with not adding any kind of clinical or other thresholds for the device companies to adhere to, at least at this time. Are you satisfied with that?

Dr. Eydelman: Yes, thank you very much.

Dr. Cassiere: Great, okay, we're up to question three. If we can have question three put up, I'd appreciate it.

Question 3

Dr. Cassiere: Alright, question three for the panel. The agency is considering the same pre-market clinical trial design and definition of non-disparate performance for over-the-counter pulse oximeters, for medical purposes as for the prescription-use devices. A, do you agree with this approach? If not, what do you recommend? Let's start with that. Do I have any takers? So I'm just going to ask a clarifying question. So, if this over-the-counter pulse oximeter is going to be used for a quote medical purpose, it's going to be separate from if I just walk into the pharmacy and pick a pulse ox on that I'm going to use on my treadmill. Is that the intent of this question?

Dr. Eydelman: Dr. Lee, why don't you summarize again what you summarized this morning?

Dr. Lee: Yes, so Dr. Cassiere, you're right. So the over the counter doesn't necessarily mean general wellness, so over the counter means that someone could purchase a medically approved-for-use pulse oximeter. It's not for general wellness or exercising or health.

Dr. Cassiere: Okay, thank you for clarifying that. I have two hands up. I have Dr. Feldman.

Dr. Feldman: Yeah so, I don't see why you would have a different criteria for a performance evaluation if it's going to be for medical purposes. It suggests to me that some sort of, this sounds to me like the warning on the cigarette package or on the alcohol bottle. If it doesn't pass this criteria, there's a big label that says, this has never been tested and is not considered appropriate

for medical purposes and only useful for your wellness, or whatever you want to use it for, and something along those lines. So I think my recommendation would be, same purpose as medical use. Anything that doesn't go through that gets labeled very clearly that it did not go through those criteria.

Dr. Cassiere: Great. Dr. Wilson.

Dr. Wilson: I agree. If it's used for medical decision making it should meet the same criteria. And, in terms of the labeling, you know, you might even be a little bit more severe in the labeling that does not meet criteria, something where they would be warned that it's not at the same level.

Dr. Cassiere: Great. Dr. Goldman.

Dr. Goldman: Thank you. Julian Goldman. Once again, I really would just want to ask a clarifying question. What in the world is this device? are these things that exist today in the market? What are we talking about in terms of an OTC pulse oximeter? Maybe James Lee could clarify that. Or is it the same exact product that just has different labeling, to help someone who is not a medical professional use it. Just maybe clarify the intent.

Dr. Lee: Thank you, Dr. Goldman. It's James Lee, FDA. So yes, there are over the counter indications for some wearables that are for medical use. And so that's where-- and certainly we're thinking of the future as well, for when devices become over the counter and more common, and they're for medical purposes. What type of criteria that you would recommend?

Dr. Goldman: So, my closest analogy that I think about is, are non-invasive blood pressure monitors, or sphygmomanometers, that are widely available over the counter. And when I take out a microscope and I read their little packages, they state that they're the same-- they'd have the equivalent performance, that they've passed standards, and that their FDA cleared. But they're just available at the corner store and online. So, is this, could you help clarify, is that the same idea here? It's the same device, but available more broadly and has clearer instructions?

Dr. Lee: Thank you. One of the things I think that we've been soliciting today too is, what type of instructions and labeling should a particular device have. So that's based on intended use, how its firm or medical device manufacturer intends to market the device, and certainly as we've discussed pretty thoroughly today, the interpretation of that saturation number has to be within context. And lay users may not have that full breadth of medical knowledge. And so, we're looking for some feedback on not just what the performance criteria could be. I heard, in the slide before, we were having discussions about, how the devices could possibly perform differently in different care theaters like the ICU, like at a doctor's office, or at home. And so I think that's what we're looking for. If you find that a one size fits all, high quality type of device design would be the way to go, or is there a flexibility based on, particular intended use, including over the counter, which could be used for a layperson for spot checking at home, for instance.

Dr. Goldman: I understand the focus today is on is on disparate performance, of course, but I think we did hear some useful messages from two of the speakers, that described their own experience with pulse oximeters. The question of the need to stabilize, and how long a measurement requires to be taken. Things that potentially could either be disclosed or performance requirements. Those were pretty good examples. And, is the intended use, or will it support, monitoring your daughter for 24/7 or is it intended for a spot check and can't be kept on? And so on. And we heard some good examples. I'm sure you've heard many, many more at the FDA. Those would seem to be useful inputs, but they don't, of course, relate to disparate performance. Thank you.

Dr. Cassiere: Dr. Brown.

Dr. Brown: Yes, thank you. I think the same standards should be upheld, because as the speaker pointed out earlier, pulse ox in the medical setting is one variable taken into consideration in

concert with the rest of the clinical picture. But for people at home with a heavy reliance on pulse ox, that may be their single factor in making decisions, like, do I go to the ER? And so I feel like the quality, the standard should be the same. And then along those same lines, there's increasing, being forward thinking, there's increasing remote monitoring happening from the home. Streamlining information to care teams or allowing patients to self-monitor and manage their health. And so, I think these need to be held to the same standard.

Dr. Cassiere: Ms. Brummert.

Ms. Brummert: I agree with Dr. Brown. I think the same standard for medical grade needs to be for over the counter. I don't know what it is for medical grade or use in the hospitals. I don't have knowledge about that, but I mean, again, I'm a pulse ox user. I just got this at my local CVS and I use this as a vital sign to determine whether I need to go to the hospital, and I'm sure that's the way it is for a lot of consumers. So I think there needs to be the same standard for both situations.

Dr. Cassiere: Dr. Lanzafame.

Dr. Lanzafame: I agree that there needs to be the same standard. I also want to underscore the fact that the labeling has to be explicit. I think both on the quote medical grade OTC product, and the general wellness product, in terms of when do I need to call someone or do something. Or can the end user, the common man, distinguish how to use one or the other properly, or what's an improper use? I think those are those are critically important considerations beyond the performance issues.

Dr. Cassiere: Dr. Taylor.

Dr. Taylor: Yep. I agree that ideally they should be the same. Two questions. One, for the non cognitive study, are you, or is the FDA proposing a third category of device? Because the wellness devices clearly have been used for monitoring, and would this also put the device out of

the range of most people to purchase? Would it just elevate the cost of it? If it's 1,000, nobody's going to buy it.

Dr. Eydelman: All three are currently available. We're not proposing anything new. All we're doing is discussing the criteria for performance for the medical use of the product. This is not a new category, but just, you know, as we discussed at the last panel, there was a lot of concern about the performance, was always a counter (phonetic) for all sorts of medical purposes. And we're trying to figure out A, which is what you're addressing right now, the performance, and B is the labeling, which I know you're going to get to.

Dr. Cassiere: Dr. Goldman.

Dr. Goldman: Well, I'm not sure if I'm getting ahead with regard to labeling, but I think there's an example that may fit some of the discussion here, with regard to again, I'll go back to the blood pressure monitoring. There's, in a risk-based blood pressure monitor that I purchased, and I think this is not unique, of course it's important that the measuring site is at the level of the heart. And so the intent, the instructions, direct the user to keep the arm bent, elbow on the table, for example. If you're sitting on a chair, bend the elbow, and ensure that the wrist is at heart level. Well, the device indicates if the angle is incorrect. It also, when it transmits the data to your smartphone, it also indicates if there was an error in the angle of your arm, so that it's no longer at your heart level.

Similarly, pulse oximeter, if it requires a certain amount of time for stabilizing a signal, to help a lay user it could say, do not remove, or count down from 10, or, whatever it does, right? There are a lot of ways to do things like this. So, there could be differences not only in the instructions for use, but in the technology to help support correct and accurate use when it's not being performed by, say, a trained medical professional. So, I think those are real-world

examples that can be applied to help ensure people get the best number possible. It may not be required, maybe these are just, optional, or maybe they're market driven.

Dr. Cassiere: Dr. Gooden.

Dr. Gooden: Yeah, Cheryl Gooden. I would like to say that I agree, I think the standard should be the same across the board, no matter in what context the monitor is being used.

Dr. Cassiere: Great. Dr. Yarmus.

Dr. Yarmus: Yeah, I-- sorry, Lonny Yarmus. Just maybe FDA clarification for me, and I see some other comments. So, I go into the pharmacy, the CVS, what is the criteria for segregating an over-the-counter medical device from a health device? Or is there one? Does the consumer know what they're doing? Or, aside from a label, is there any mandate for location differences? Do they have to buy it behind the counter, even though it's over the counter and doesn't require a prescription…

Dr. Eydelman: Dr. Lee, would you like to elucidate further?

Dr. Lee: Thank you, Dr. Yarmus. So, that's actually the basis of our question. So, right now, devices that if you're buying, let's say, most of the time you're buying something on Amazon or any other online company, or you're going to the counter. There are often, if you look closely, caveats about what the device is. They're supposed to be clearly labeled, that they're often forced and you'll see sports and aviation, or general wellness type of claim. So this is like a health and wellness, exercising, or sports and aviation. So what we're looking at here is, as the products in this space expand, are there particular considerations that you would recommend in labeling. And it could go on the labeling as a package insert or it could go on the box, about specifications or instructions for use about the device, particularly if let's say something is for a spot checking versus continuous use. So as technology goes forward, as the innovation happens, we're looking

forward to considering updating labeling and recommendations as we work with sponsors in the future.

Dr. Cassiere: Dr. Feldman.

Dr. Feldman: Yeah, I don't necessarily have an answer to this, but I think there's been some nice points made about the cost and accessibility of the devices to people, particularly folks at home that have chronic medical conditions, who may have a lot of durable medical equipment costs that they're having to bear. And so I don't know if there was a minimum performance requirement that was different from what we're talking about, and whether that would translate into a more accessible device, but if it did, I think it would be worth considering, and then any device that didn't go through any kind of evaluation certainly should be labeled as such. But I think there is an accessibility issue for patients here that's worth considering in deciding what the criteria might or might not be. I can imagine some less stringent criteria that would still be medically useful for patients in their homes. And if that was less expensive, then that'd be great. I did a quick Amazon search. You can buy a pulse oximeter for 15 dollars. That was the lowest cost one that I found, and I suspect that that would have a hard time meeting the criteria we've talked about. And whether or not it would be unsafe for a patient or not, I don't know, but I think it's something that's worth investigating.

Dr. Cassiere: Dr. Wiswell, any thoughts on this?

Dr. Wiswell: Yeah, I did the same thing on Amazon and I found one for 12.59, so that's the one I'm buying. But I'm not going to look at the package, I'm not going to see that it's for wellness or health, or that it's been looked at to be a medical device. And I think it just really, if it's going to be used, quote, medically, and the layperson thinks it's going to be a medical device that's helping them, it's got to be clear that it's met certain criteria. It's not just something that-- there's a

lot of junk sold on Amazon. And I'm just afraid. I've got a 40-year-old daughter who's got long COVID too, and I don't want her to buy something that's not going to be that healthy.

Dr. Cassiere: Ms. Brummert.

Ms. Brummert: I want to reiterate the point that I made earlier, is that I think even the over-the-counter should be behind the counter, because someone with expertise will be able to explain it to the consumer. A consumer is just going to walk into a CVS or just walk into some sort of pharmacy, buy it on Amazon. They're not going to read the labeling. They just know that this is something they can put on their finger and they're going to get some sort of result from that. So, I don't know about labeling but I do think that there needs to be an extra layer of protection for over-the-counter use by having it behind the counter, so to speak, so that these issues can be explained or, what the uses are. Someone can ask their pharmacist, is this the right product for me? And if not, what do I need to look into? So I just wanted to throw that back in there.

Dr. Cassiere: Alright. Dr. Goldman?

Dr. Goldman: Yeah, thank you. Julian Goldman. I'd like to support the importance of thinking about availability and cost, in the, I'll just say, lower end pulse oximeters. I don't want to provide a designation. But certainly in 2020, when our hospital system was trying to identify pulse oximeters that we could potentially buy and provide to patients, or dispense when they left the ER or whatever was happening, it was a tremendous struggle to identify and even determine what the performance level was. And we were fortunate that, a test lab that was based in Colorado, volunteered to test and provide data on a whole range of pulse oximeters. And we used that information and disseminated it. So, I'll just support that these really are important factors. We don't know what will happen in six months or next year. And I would propose that that be part of the thinking from a preparedness and accessibility standpoint for technology like this that's so vital.

Dr. Cassiere: So, Dr. Eydelman, what I'm hearing from the panel is we all agree that there should be a higher standard for these over-the-counter so-called medical grade pulse oximeters. The real question is, are they going to be cost prohibitive if there's an expectation? Again, I think it was Dr. Wilson, mentioned it's going to cost a quarter of a million dollars to do these studies if the new regulations come into effect. And how's that going to translate into over-the-counter pulse oximeter prices if there's an expectation to meet the exact same criteria as the hospital grade. So, I think we agree there should be a higher standard, and I guess we're all scratching our heads what that standard should be. Dr. Feldman brought that up. Could it be a lower standard than the hospital grade? 2013 maybe, FDA requirements. I'm just throwing it out there. And I think that's the gestalt of the panel's recommendations. Although I have Doctor Wilson who has his hand up.

Dr. Wilson: I would just emphasize that if there's true medical decision making on an individual at home, then the device should have medical-grade outputs. Not every pulse oximeter manufacturer that sells over the counter needs to apply for and gain that status. But those that do, the consumer should know whether they're using a device that works fairly well for exercise and well-being, but is not as rigorous for medical decision making.

Dr. Cassiere: Dr. Eydelman, do you want any further discussion on this, or?

Dr. Eydelman: I think we're good, thank you very much. If you could just go around because of labeling. Sure.

Dr. Cassiere: So let's, so we made the decision that they should meet the same criteria. But what type of labeling, and to speak to Ms. Brummert's point, maybe this needs to be behind the counter in terms of access, and what that labeling would be, in terms of medical over the counter pulse oximeters. Anyone want to jump in on that? Dr. Lewis.

Dr. Lewis: I think that-- this is Tamorah Lewis. I think keeping it simple is really important. Having standard language in a clearly defined box, almost like we have nutritional labels, but

obviously something different than what's included in a nutritional label, but that it looks the same every time. Something along the lines of, medical pulse ox devices meet certain standards. This does, in bold, or does not, in bold, meet those standards. And, very, very simple, not what the standards are, patients, I don't think, are going to want that level of detail or be savvy enough, but that just very clearly the agency is, in a standard way, relaying the fact that there are standards and this product does or does not meet them.

Dr. Cassiere: Any other comments on this? Dr. Taylor.

Dr. Taylor: Jim Taylor. I just pulled the pulse oximeter that I bought during COVID. Fortunately, I didn't have to use it, but the material that came with it basically is eight pages, even of which are basically full of very fine print. So, if you're going to label something, that you want people to look at, it's got to be highlighted or colored or something. But I don't know if this is standard for these devices or not. This is an OTC one. Thanks.

Dr. Cassiere: Dr. Feldman.

Dr. Feldman: Yeah, Dr. Feldman. I have a question about what's within the FDA's authority around labeling. In particular, I'm concerned about those devices that will never get submitted for FDA approval. So, clearly we're in a world where people can make very low-cost pulse oximeters, box them up, put them on a shelf. And people are going to buy them. So, you can't buy a pack of cigarettes, you can't buy a bottle of alcohol, without a consumer warning on it in this country. Is that within the purview of the FDA, to require for those devices that do not go through the approval process, or is that some other legislative process that would have to take place in order to get them labeled that way?

Dr. Eydelman: Dr. Lee, why don't you start?

Dr. Lee: So, thank you, Dr. Feldman. So, that is correct, actually. So, for devices, foods, drugs that the agency regulates, there is a clause that there needs to be an exhibit of truthfulness in

labeling. And that's under our Code of Federal Regulations and we govern that, the labeling supplied for product that within our purview is truthful and accurate. And so, that's why, often we do label reviews as part of our review process on the pre-market side, and that's why we're posing the question we are. We're looking for a little bit of guidance or goalposts or criteria that you find important. And certainly, clear labeling, accurate labeling, labeling that's useful, is something that we do support. But we're looking for anything from the panel about anything specific that you've seen, in particular. Be it a prescription, or in this case, if it's direct to patient.

Dr. Feldman: So, every device that goes through the regulatory process gets a nice little pink lung icon on the device, printed somewhere that's visible. Every device that doesn't, and is marked as a pulse oximeter, gets a dark lung with an X through it printed on the device. So when you put it on your finger, it's obvious that something's not right about this thing. Just a quick suggestion.

Dr. Cassiere: Dr. Goldman.

Dr. Goldman: Yeah, thank you. Julian Goldman. I'd like to comment on aspects of accessibility, of both information and of the devices. So I see the intent and value of keeping something behind the counter or ensuring that you provide guidance when someone buys something. But I'm also concerned that someone may go shopping and doesn't speak the local language. And again, it's at the surge of respiratory disease in the community. And I won't comment which pharmacy I normally go to, that the line is so long just to pick up something pretty typical that there really is no one to speak to, most of the time, because of the way things are staffed. So, I just would recommend that we consider balancing those facets of accessibility of equipment. And then, related to that, is accessibility or understanding the information, and needing to take out a microscope to read the eight pages of instructions. a number of years ago, the Anesthesia Patient Safety Foundation, in a workshop, pointed to the value of using QR codes at the point of care to

access relevant information. And all of us have now used QR codes in our daily lives. The FDA has, to the best of my knowledge, now supported the use of QR codes on several devices, including on a ventilator that displays pertinent information when someone scans it. So, that's a technology that's certainly quite mature, has come of age, and could allow users of all different kinds to access information in their language. I don't want to-- not recommending being prescriptive, of course, in a venue like this. Just pointing out that there are good ways that exist to get very useful information at the point of care, including potentially videos and animations, instructions in the language that is needed. And if that's the focus for over the counter, it's really helping to explain to people what to do, how to use it, signs of problems, and when to go to the hospital or get medical help, that can be conveyed even more clearly, potentially.

Dr. Cassiere: Great. Dr. Yarmus.

Dr. Yarmus: Yeah, thanks. Lonny Yarmus. So, at the risk of creating more debate at 6:10, I'm just going to do this. But, so, should we not be thinking of this in the realm of a medication? And if I'm an interstitial lung disease expert physician and want outpatient monitoring of my patient, can I prescribe them an adequate pulse oximeter to use? I mean, that's really what we're talking about. This is going to be physician recommended care monitoring, in an outpatient setting, which, the analogy of prescriptive medication is probably the closest thing. And why should we not consider that bucket?

Dr. Eydelman: So, if I can just interject here. Again, I'm going to ask Dr. Lee to comment further, but I think there's a little bit of confusion. We're talking about OTC, over the counter, not prescription use right now.

Dr. Yarmus: And do those, and I suppose I should know this, does that exist currently? Can I prescribe a pulse oximeter?

Dr. Eydelman: Yes. Go ahead, James.

Dr. Lee: Yeah, Dr. Yarmus, are you asking can you order a prescription use pulse oximeter for a patient? The answer is yes.

Dr. Yarmus: And so why would we separate these two types of devices in this discussion if that's already out there?

Dr. Lee: Well, the second thing is that if patients are able to access, on their own, a pulse oximeter, because either their doctor may not prescribe one, or they're monitoring a chronic disorder and they're interested in advocating for their own health condition and they're able to get over the counter, they may not have direct conversations with their physician in a regularly scheduled paradigm. Or they may be on their own. And so, what we're asking here is, are there any particular considerations, considering a patient may be having to advocate or self-treat or self-monitor themselves. Are there any recommendations that the panel would like to have in such a labeling, for a patient-direct product?

Dr. Eydelman: If I can just suggest-- I know it's 6:11, but Dr. Yarmus, if I can just suggest an analogy of reading glasses. You can go and buy reading glasses over the counter without the prescription. Now, you can also have an optometrist and ophthalmologist give you a prescription. That would be a different way of obtaining those glasses. Also, over the counter, we at FDA, are involved in whether something is over the counter or not. But again, I just wanted to go back to Ms. Brummert's comment about whether it's behind the counter or not, that is not usually-- FDA does not usually get involved in that level. That's more of a state discretion. So, but back to, so the question-- it's 6:12, we have 12 minutes. So back to our question at hand. It was meant to ask for your recommendation for over-the-counter pulse oximeters, not prescription, not wellness, not the ones used in the hospitals. The ones that people can just walk into the store or go to Amazon or some other online entity, and just order because they want to know what their pulse ox is. So, that was what this question was intending to ask.

Dr. Cassiere: Ms. Brummert.

Ms. Brummert: I, too, do not want to extend the conversation more than we have to. It's 6:13. But, we talked earlier about having the same standards for over the counter as for prescription. I'm wondering if FDA, I don't know whether this is a thing, but can FDA reclassify all pulse oximeters to class two?

Dr. Eydelman: I didn't realize I was muted. Sorry. James, I was trying to get you to speak again that they are, so all pulse oximeters are class two. It has nothing to do with classification, but within the class two pulse oximetry rate, we have different types. So sorry if-- I'm not going to ask IT folks to bring up the slide, but after we did have that slide what seems like many hours ago.

Dr. Cassiere: Dr. Wiswell, I think you're going to be the last commenter on this very uncomfortable subject.

Dr. Wiswell: Can't there just be a label that's really in bold, on the outside of the box, not going through the eight pages of microscopic labeling, but just saying, this is not a medical device and it has not undergone scrutiny like medical grade oximeters. Just something like that, but something that's going to grab their attention. And I like Dr. Feldman's description, have black lungs on there with a line through them, or something like that. Yeah, jokingly, but we should be able to make it obvious to the whole population because there's a lot of unsophisticated people that don't read labeling that's in papers on the inside, and they're just going to grab the thing that's the cheapest.

Dr. Cassiere: So, Dr. Eydelman, I know this may be not totally satisfactory, but are you comfortable with the panel's discussion on the last part of this question three?

Dr. Eydelman: Yes, thank you very much.

Dr. Cassiere: Great. so, at this time I'd like to ask our representatives, Ms. Brummert, our consumer representative, and Dr. Wilson, our industry representative, if they have any additional comments. Ms. Brummert. If not, Dr. Wilson.

Dr. Wilson: I want to be very brief but let me just encourage the agency to go forward with the recommendations where there has been a pretty good agreement. So let's not let what we can't do get in the way with what we can do. It's been quite some time since this has been brought up, and I think the public is looking for some action. So, applaud the FDA and I would move forward in areas where there's pretty good agreement.

Dr. Cassiere: Great. At this time, the panel will hear summations, comments, or clarification from the FDA. Dr. Eydelman, you have 10 minutes.

Dr. Eydelman: So, I just wanted to thank a very extensive team at a number of offices for working diligently for a couple of years, to put together all of the knowledge that's currently available to be able to propose what we have brought before the panel today. And I also wanted to thank all of you, our distinguished panel members, for your thoughtful and very extensive deliberations. And we look forward to implementing all of your recommendations to the best of our ability. That concludes my remarks.

Panel Summations

Dr. Cassiere: Great, thank you. At this time we will hear summation comments or clarifications from the panel. We also have 10 minutes, if anyone has anything that they'd like to summarize or highlight at this time. Going once, going twice. Sold. Okay. I'd like to thank the panel, the FDA, the invited speakers, and all of the open public hearing speakers for their contributions to today's panel meeting.

Adjournment

The meeting of the Anesthesiology and Respiratory Therapy Devices Panel is now adjourned.

Thank you.