Pulse Oximetry Forum: advancing best practices & device standards

(a joint meeting of experts and the ISO Oximeters JWG)

March 14-15, 2023

Hybrid Meeting
To attend virtually via Zoom: Register here
To attend in person in San Francisco: email us (space is limited)
Mission Hall, UCSF 550 16th St., San Francisco, CA 94158 (Map)

Draft Agenda

Day 1, March 14, 2023 (Mission Hall, 2nd floor, room 2103)

Theme I - What are the problems?

Time (PST)	Activity	Speakers
08:00- 08:30 am	**Sign-in & Coffee**	
08:30- 09:15 am	**Welcome**, overview and goals for this meeting from Co-Hosts	Mike Lipnick Odi Ehie James Lee Bob Kopotic Sandy Weininger
09:20- 10:45 am	**Real world challenges in pulse oximetry**	Michael Lipnick (moderator) Liz Igaga Michael Sjoding Carolyn Hendrickson Desiree Conrad and Rohan Taneja Louise Mulroy
10:50- 11:45 am	**Confounding factors in pulse oximeter performance** and their impact on device performance, controlled desaturation study design, and health/healthcare disparities	Phil Bickler (moderator and speaker) Mike Bernstein Ellis Monk Mark Ansermino
11:45- noon	**Brief Q&A/Discussion**	
noon- 1:00 pm	**Lunch** (Provided by UCSF OpenOximetry)	
1:00- 2:30 pm	**Strengths & shortcomings of controlled desaturation studies** Challenges in ensuring adequate diversity (not only how we define skin color, but also recruiting adequate diversity in skin color)	Bob Kopotic (moderator and speaker) Paul Batchelder John Feiner Bunmi Okunlola David Macleod
2:30- 3:30 pm	**Statistical analyses** and sample size for controlled desaturation study design (different objectives for design versus qualification)	John Feiner (moderator and speaker) Gene Pennello David Milkes
3:30- 4:00	**Open Forum**	Zach Vesoulisch François Lellouche

4:00- 5:00 pm	**Day 1 review**, discussion and theme outputs	Sandy Weininger

Day 2, March 15, 2023 (Mission Hall, 3rd floor, room 3700)

Theme II - What are the solutions?

Time (all PST)	Activity	Speakers
08:30- 9:00 am	**Sign-in & Coffee**	
09:00- 11:00 am	**Optimized skin color quantification techniques** to ensure adequate diversity of skin color (pros and cons of current methods & defining optimal methods)	Jenna Lester (moderator) Ellis Monk Leo Shmuylovich Sandhya Vasudevan Wim Verkruijsse Dr. Adia Benton and Dr. Amy MoranThomas
11:00 am- noon	**Discussion**	Jenna Lester and Kelvin Moore (moderators)
noon- 1:00 pm	**Lunch** (Provided by UCSF OpenOximetry)	
1:00- 2:00 pm	**How to account for challenging clinical scenarios and potential roles for phantom in vitro testing**	Sandy Weininger (moderator) Mike Bernstein Chetan Patel Josh Pfeffer Geoff Mathews and Veronica Hickson
2:05- 3:00 pm	**Opportunities to utilize Standards for improving communication of device performance** to minimize potential misdiagnosis and/or health disparities from pulse oximeter misuse	Mike Lipnick
3:05- 4:00 pm	**Open forum**	Kelvin Moore (moderator) Gauri Singh
4:00-5:00 pm	**Day 2 review** & discuss proposal	Bob Kopotic

List of Challenges and Priorities (identified in prior ISO and OpenOximetry meetings)

- Harmonizing/improving lab data set collection
- Skin color quantification techniques
- Improving global procurement standards to account for device quality and lifetime cost
- Perfusion quantification in regulatory frameworks
- Harmonizing clinical trial data set collection
- How to accelerate adoption of standards for routine use of pulse oximeters
- How to accelerate adoption of standards for routine use of pulse oximeters for neonates infants, and children
- Implications for home monitoring
- How to ensure transparency and adequate consumer/user awareness and communication
- Updates to oxygen therapy guidelines
- Education (Clinician, patient, policymaker etc)
- How to define adequate performance, not letting best be the enemy of good
- How to ensure adequate diversity in regulatory committees
- Need real world and useful data on patients with dark skin pigment
- To what extent will future studies measure skin color in more than one dimension, since there are multiple melanins with different absorption spectra? Any one-dimensional skin tone scale may be reductive to the extent that it does not capture the differences between demographic groups.
- Explore approaches to reducing the cost of pulse oximeters and making them more available to general public
- Identify new approaches to maintenance of equipment used to treat pneumonia, such as oxygen
- What are the opportunities for redeployment and improved device availability and coverage for oxygen commodities post COVID-19?
- Determine how best to assess new automated pneumonia-diagnosis technology (e.g. multi-modals), including accuracy, demand, usability and appropriateness, at different levels of the health system
- Precision of observations. We have typically used SpO2 to one decimal place but this should be specified in the protocol. Averaging (mean, median, mode) and time window should be reported. We typically think of HR as counts (rounding down) – most monitors do not do this!
- Repeatability (ability to achieve the same observation under the same repeatability conditions). This should be evaluated in the context of each experiment. If observations are not repeatable this will introduce noise (uncertainty) in the comparison. I think we should evaluate and report this for each (reference/ test) device.
- Within subject variation. This can be a major contributor to poor repeatability. Different fingers on the same hand at the same time do give different measurements! Sick

patients (and especially newborns!) have much more variability. Variability increases at lower saturations.

- Signal quality is critically important (I know you have reported this). I think it is more than perfusion index.
- Observer bias. Digital number preference is very well described entity. Electronic recording would be ideal to overcome this!
- We would ideally model all causes of uncertainty to fully understand how much of this is due to performance of the test device and how much is noise.
- Defining the gold standard - there is variability between hemoximeters and variability in how they are used e.g. functional vs fractional saturation. Should there be standards for this?
- Neonatal and pediatric performance data are limited
- Post market quality assurance is lacking, especially in LMICs
- Characterization of intra-patient site-to-site SpO_2 performance variability
    - BLUF - Intra-patient site-to-site variability is important since it a) probably contributes to the uncertainty of real-world data, b) is important to interpret spot-check (intermittent) trend data, and c) reflects - and should inform - real-world POx usage.
    - Clinical pulse oximetry usage necessitates moving the probe to alternate sites for various reasons (e.g. to follow manufacturer's instructions to avoid prolonged application of clip style sensors, repositioning patients, etc). As we know, a not uncommon reason to move a probe is to find the “best” (highest) SpO2. This begs the question of what metric is available to characterize site-to-site performance variation when a sensor is removed and reapplied.
    - A potential answer is that intra-patient performance is “baked into” the current Arms statistics. But from a practical perspective, a clinician might accept an SpO2 error of, say, 4%, and attribute that to the published Arms value related to inter-patient (population) variability. But if you were to remove and replace a probe on a different digit of the same (stable) patient, would YOU accept a 4% variation? Probably not. That probe would be tossed into the trash bin b/c we expect intra-patient variability to be less than the population-derived Arms. Is that assumption valid? Where is the data?

Links to Useful/Referenced Resources

- About The Open Oximetry Project
- Collaborator sign-up (please sign up if you have not, and share with others)
- Collaborator list (this includes only those who have chosen to release their info)
- Open Oximetry Study protocols and device testing list
- UNICEF Pulse Oximetry Target Product Profile
- FDA Open Panel Hearing Pulse Oximetry (November 1, 2022)
- Federation of American Scientists Meeting on racial bias in medical devices (November 2, 2022, recording here)
- https://www.gov.uk/guidance/the-use-and-regulation-of-pulse-oximeters-information-for-healthcare-professionals
- The FDA guidance is at the following link (section 10 speaks to labeling): https://www.fda.gov/media/72470/download
- The 'real-world" clinical research protocols : https://openoximetry.org/study-protocols/
- The use and regulation of pulse oximeters (information for UK healthcare professionals):
- https://www.gov.uk/guidance/the-use-and-regulation-of-pulse-oximeters-information-for-healthcare-professionals#:~:text=Pulse%20oximeters%20intended%20for%20clinical,not%20approve%20medical%20devices%20directly
- Guidance on Regulating medical devices in the UK:https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk#:~:text=will%20be%20regulated.-,The%20role%20of%20the%20MHRA,of%20UK%20conformity%20assessment%20bodies
- Info on oximeter testing under previous symposium: https://iampov.org/about/
- https://academic.oup.com/poq/article-abstract/80/2/534/2588836
- Interestingly the albinism case had falsely low values, but the patient was taking an herbal extract that functions as a UV sensitizer, similar to psoralen I believe: https://journals.lww.com/anesthesia-analgesia/fulltext/2004/02000/oculocutaneous_albinism_and_spurious_pulse.60.aspx
- Learn more about the role that von Luschan scales have played in histories of scientific racism at the US Holocaust Museum: https://collections.ushmm.org/search/catalog/irn564926
- Linearity is defined in the CLSI harmonization database: https://htd.clsi.org/default.asp as ability (within a given interval or range) to provide results that are directly proportional to the concentration (or amount) of the analyte (measurand) in the test sample.
- https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=870.2700
-

if you are unable to view these links due to your organization's firewall, please email - fekir.negussie@ucsf.edu or michael.lipnick@ucsf.edu

MEETING MINUTES (LIVE)

THIS SECTION IS FOR GROUP NOTE TAKING

Time (PST)	Activity	Speakers
08:00-08:30 am	**Sign-in & Coffee**	
08:30-09:15 am	**Welcome**, overview and goals for this meeting from Co-Hosts	Bob Kopotic Sandy Weininger Mike Lipnick Odi Ehie James Lee
	1. Dr. Odi Ehie, UCSF Associate Professor Pediatric Anesthesiology **a. Pulse oximetry, skin color, and health equity** **i. Race/Ethnicity/Skin Color definitions and differences** **ii. Updating Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals (JAMA 2021)** **1. Race and ethnicity are social constructs, without scientific or biological meaning** **2. Race is not equal to biological construct** **b. What is structural racism?** **i. Importance in clinical research trials with respect to racial and ethnic disparities in study populations** **ii. Reconsidering the Use of Race Correction in Clinical Algorithms** **iii. Why Does DEI Matter?** **1. Make a conscience effort to address bias** **2. James Lee, FDA CDRH** **a. Classification of Pulse Oximeters and US Regulations** **i. Prescription Use Pulse Oximeters** **1. Reviewed through 510(k) process by FDA**	

2. Used to monitor oxygen saturation levels of patients (trend, spot checking)

ii. Over the Counter (OTC) Pulse Oximeters
   1. Not intended for medical purposes without FDA approval process

iii. Factors that can impact the accuracy of pulse oximeters including but not limited to:
   1. skin pigmentation, hemoglobin disorders, severe anemia, low perfusion dyes, nail polish, ambient light

iv. 510(k) verify SpO2 accuracy in comparison to the gold-standard measurements of blood SaO2 by co-oximeter over the specified SaO2 range (70-100%)
   1. Significant number of subjects to attain statistical significance to demonstrate a specified SpO2 accuracy
   2. Healthy adult volunteers and range of demographics
   3. Allowable Testing Conditions
   4. Testing Procedure for desaturation profile
   5. Data Submission for FDA Review
   6. Analyses
      a. Pooled accuracy Arms across the entire tested range of SpO2
      b. No threshold accuracy analyses
      c. Not powered to compare subgroups statistically
      d. Bland Altman, population mean bias, between and within subject variance, Upper and Lower LOA
      e. Outliers Discussion and factors that may have affected performance

v. Reconsidering Premarket Clinical Denaturation Studies
   1. Indications - currently cleared for use for monitoring trends, spot checking
   2. Should certain target ranges require a higher degree of accuracy?

3. Sandy Weininger, FDA

a. **Overview of ISO 80601-2-61 Standard for Pulse Oximeter**
   i. **Establish minimum internationally accepted safety floor by providing consensus objective test methods and acceptance criteria**
   ii. **Safety and Performance requirements for manufacturers**
   iii. **Process for identification of unsafe conditions**
   iv. **What is an engineering standard and difference to clinical perspective in terms of safety and effectiveness?**
   v. **Identify credible reference devices for performance evaluation and clinical significance of requirements**
   vi. **Standards map - General Standard 60601-1 and Collateral Standards discussed**
   vii. **Scope of Pulse Oximeter Standard - Estimation of arterial oxygen hemoglobin saturation and pulse rate**
   viii. **Augmenting required labeling of device performance under review**
   ix. **SpO2 accuracy (<= 4%) defined in terms of Root-mean square difference**
   x. **Controlling variability in collection of verification data**
   xi. **The Qualification of a pulse oximeter defined in ISO 80601-2-61 is not an exhaustive set of requirements on how to calibrate an oximeter**
   xii. **Factors affecting Performance**

Chat Notes:

Julian Goldman: I added "Characterization of intra-patient site-to-site SpO2 performance variability" to Google doc in "List of Challenges and Priorities". This topic has been discussed at some length with ISO and OpenOx SMEs in lead-up to this meeting.

Sandy Weininger: There is a (relatively) new ISO standard that specifies what should go into instructions for use. FDA spends a lot of effort to assess the quality of documentation. (remember that labeling is the least effective form of risk control but its the easiest to update)

The FDA guidance is at the following link:
https://www.fda.gov/media/72470/download

09:20-10:45 am	**Real world challenges in pulse oximetry**	Liz Igaga Michael Sjoding Carolyn Hendrickson Desiree Conrad and Rohan Taneja Louise Mulroy
	1. Dr. Liz Igaga, Makerere University Anesthesia and Critical Care Director Safety and Quality, Smile TRain **a. Global operation theatre distribution and pulse oximetry supply (Funk, Weiser, et al, 2020)** **b. Challenges in Real World** **i. Access to appropriately sized probes in the real word - pediatric challenge specifically** **ii. Durability** **iii. Maintenance** **iv. Training (valuable tool, misused, under utilized)** **v. Reliability (extreme conditions, shock, movement in transport)** **vi. Skin Pigmentation? Clearly define the unknown extent of issue and curate solutions** **c. Access and Performance require attention** **2. Dr. Michael Sjoding, University of Michigan Associate Professor of Internal Medicine, Pulmonary and Critical Care** **a. COVID-19 Surge at University of Michigan** **b. Change in practice using pulse oximeters versus the way they have been tested for safety and effectiveness** **c. Evaluated pulse oximeter accuracy collected through Electronic health Record data for self-reported White and Black patients hospitalized in 2020** **d. Analyses performance showed bias between patient populations** **e. Retrospective data did have delay in measurements and reference** **f. How often was hypoxemia missed? (SaO2 < 88%) when the pulse oximeter was 92-96%** **g. Prevalence of hypoxemia difference in groups not as much of concern in published data analysis** **h. Small pulse oximeter differences can translate into significant population disparities in hidden hypoxemia** **i. problem is in the "tails" of statistical distribution of defined hypoxemia** **i. Opportunities and limitation of retrospective data** **j. Actions to consider**	

i. report pulse oximeter performance across racial groups
ii. power studies to detect small, but clinically important performance differences
iii. ensure pulse oximeter testing aligns with use in practice
iv. routine collection of real world post market surveillance data to evaluate in clinical practice

3. Carolyn Hedrickson ,MD MPH UCSF
   a. ICU Clinical Research Study Challenges
      i. Barriers to Data Capture
      ii. Covariates, Confounders
      iii. Power, Testing Relevant …
   b. Data Analysis Plan
      i. compare pulse oximeters errors across skin pigment categories
      ii. Key covariates
   c. Screening and enrollment
   d. Recruitment Process
   e. Challenges to collect transient low saturation data in patients.
   f. Study Subject Self-Identified Race distributions
   g. Capturing Desaturation Events
      i. Stability of saturation <90% is rare
      ii. Improving staffing
   h. VL Dorsal Fingertip Pigmentation Distribution
   i. Unknown to current data how to model SpO2 and SaO2 relationship
   j. Perfusion Index: half enrolled have abnormal perfusion in hospitalized state
   k. What are the drivers of pulse oximeter errors and bias in critically ill patients?
      i. Skin pigment only? Structural Racism chronic comorbidities and differential care, perfusion, tissue edema interactions?

4. Dr. Desiree Conrad, Chief Resident Stanford Children's Health and Tohan Taneja, Stanford
   a. FDA sponsored pediatric study
      i. Study Design - Population and Variables
      ii. Single Center prospective clinical trial (observational)
      iii. Skin Pigment Distribution: Lighter end of scale is current majority using Fitzpatrick and Monk scales

iv. 19 subjects and 38 paired measurements so far, r2=.79 from SpO2 to SaO2 linear regression
v. Challenges and Lessons Learned
    1. Using multiple pigmentation scales
    2. Gold standard for skin pigmentation does not exist
    3. can take up to 1 hour per patient to assess skin pigmentation
    4. Colorimeters were not designed for pediatrics
    5. Dynamic changes in oxygen saturations in the operative room where enrollment was initially anticipated, enrollment changed to CVICU for easier and more stable saturation measurements
    6. Parent Anxiety hindering enrollment
    7. Child Anxiety hindering enrollment
        a. movement artifacts PTSD from multiple past procedures
        b. Colorimeter may be intimidating
vi. Anticipated to complete enrollment July 2023 (154 subjects)

5. Dr. Noha Aboleta
    a. Review of systematic overestimation of oxygen saturation for identified racial population in Dr. Sjoding's published data.
    b. In clinical practice pulse oximeters more than only trending and spot checking monitor as indicated for device use
    c. Used for diagnosis and decision making in clinical care
    d. Unacceptable for unequal performance across skin tones to be addressed
    e. Honest communication to health care providers and community most affected by this issue is critical moving forward

6. Dr. Louise Mulroy, MHRA (Regulator of medications, medical devices in the UK)
    a. Independent review of equity in medical devices ongoing
    b. MHRA Guidance and Regulatory Activity
        i. Address bias in medical devices effort in a wider sense not only pulse oximeters

	ii. **Currently reviewing pulse oximeter evidence with differently pigmented skin. Not aware of any reported incidents of harm to use across populations.** **Chat:** **DS (?) : regarding a previous comment about using adult probes (sensors) on infant feet - off label usage is a potential source of error unless such a sensor is designed to do this and has been tested. Following manufacture directions for use (DFUs) is essential in research and real world applications.** **Josh Pfefer: perhaps I missed it, but what are the recruiting targets for this study, particularly for very dark skin subjects? Are they based on an objective or subjective method?** **Josh Pfefer: What do you mean by 4 von Luschan groups, as it's a scale of 30+ values?** **William Wilson: How will you adjust for outliers?** **Michael Lipnock: Manufacturers to make it easier to identify models of approved devices in research - continued discussion later this week.** **Carolyn Hendrickson: For those who are interested in using the 'real-world" clinical research protocols. The documents can be found here: https://openoximetry.org/study-protocols/. Also feel free to email me at carolyn.hendrickson@ucsf.edu if materials or consultation would be helpful AND if you have feedback on how we are collecting data and planning analyses.**	
10:50-11:45 am	**Confounding factors in pulse oximeter performance** and their impact on device performance, controlled desaturation study design, and health/healthcare disparities	Phil Bickler (moderator and speaker) Mike Bernstein Ellis Monk Mark Ansermino
	1. **Dr. Phil Bickler, UCSF** a. **Why the problem is more than skin deep** b. **Pulse oximeter performance bias a complex issue** c. **Fundamental issue that all pulse oximeters are the same requires education to address** d. **Pulse oximeter biased high: carboxyhemoglobin, skin pigmentation**	

e. Pulse oximeter biased low: interfering substances, increased venous pulsations
f. Other accuracy and precision factors: anemia, low perfusion motions
g. Distinguishing between pulsatile and nonpulsatile interferences is important
h. Anemia: Less pulsatile signal and this condition can now be simulated to evaluate device performance and low signal situations
i. Understanding Sjoding: UCSF Hypoxia Lab
    i. The effects of skin pigmentation on pulse oximetry accuracy at low saturation have been seen in the past from controlled desaturation studies at UCSF lab.
    ii. What factors affect bias?
    iii. Skin pigmentation classification - Is Fitzpatrick scale really the best assessment of pigmentation?
    iv. Multivariate model of factor possibly related to bias: skin class (Fitzpatrick), perfusion index, hypoxia, sex, age.
        1. Both oximeters evaluated would pass FDA Clearance
        2. Bias as a function of perfusion and skin pigment (light, medium, dark)
            a. darker pigment subjects show greater bias at lower levels of perfusion
        3. Bias including hypoxia
            a. bias correlated to levels of hypoxia and perfusion across skin pigment range
    v. Missed Hypoxemia discussion of lab study data
    vi. Low perfusion is the factor amplifying bias in real world retrospective data analyses
    vii. Accounting for skin color alone may not be addressing full solution to problem

2. Mike Bernstein
    a. Design Issue Most Overlooked by Pulse Oximeter Manufacturers
        i. Electrical Noise
            1. All electric systems had noise and in most case random
            2. In pulse oximetry the signal is the pulse plethysmogram not the entire optical signal

3. Why is this a problem? Sensitivity of common algorithm for SpO2 to noise
4. Understanding difference between a "good" and "not so good" pulse oximeter in the presence of electrical noise
5. Dark skin and low perfusion
    a. Dark skin and reduced light signals result in worse SNR.
    b. Lowe Perfusion subjects SNR worse as perfusion decreases as a component of pulsatile signals
6. Advanced devices can report accurate SpO2 at SNR around 1:1
7. Many marketed devices report wrong values with a weak signal in the presence of noise
8. Other factors to be investigated other than physiology.

3. Ellis Monk, Associate Professor Department of Sociology, Harvard University
    a. Many pulse oximetry qualification clinical studies did not include a diverse subject population contributing to the apparent bias in performance
    b. Disentangling effect of race through factors associated with pulse oximeter inaccuracy
    c. Health disparities between black and whites often remain even after controlling for SES and health behaviors (Das 2013)
    d. Skin tone stratification and disparity examined
    e. Social Inequality: A Matter of Categories?
    f. Social stress and effect on emotional and physical health disparities
    g. Race and skin tone are NOT the same characteristic
    h. Potential confounders for pulse oximetry are associated with skin tone within and across ethnoracial categories

4. Dr. Mark Ansemino Professor Anesthesiology, University of British Columbia, Vancouver, Canada
    a. Measurement Uncertainty
        i. This means that there might be a mistake we we measure something, and that we are not 100% on the accuracy of measurement
        ii. What is the truth or true value of a physiological parameter?

iii. What is the uncertainty range of a measurement?
iv. Clinical Decision Uncertainty
    1. Lack of understanding and guidance using measurements as a threshold, how we define thresholds and education on use of thresholds with this information in decisions
    2. Understanding context of uncertainty

Chat:

T (?): What was the measure of SNR in the anemia ex-vivo study?

Jim Stafford: Is there reason to believe that skin tone effect would be different between transmissive and reflective pulse oximeter. ?
Phil Bickler: This is important question and be studied

T(?): Note that when multiple sensors are placed together on a study without proper shielding, it can lead to significant errors that are not representative of typical clinical scenarios.

Cameron McGuire: Can you use the Perfusion Index as a correction factor to predict the SaO2 based on the SpO2? Have you looked at this prospectively?
Phil Bickler: Complex interaction, need to analyze raw signals not processed SpO2.

T (?): Can you comment on the stability of SpO2 and the perfusion, when these data pairs were measured?

Jim Stafford: Many if not most pulse oximeters change the current of the LED to maintain a "minimum" level. Doesn't this "fix" the problem? Does this impact SNR the same way for both dark skin and low perfusion?

Howard Saft: An early oximeter by Hewlett Packard used more than 2 wavelengths. I understand that this approach fixed the issue. Might that be a design feature that needs to be brought back into use?

Gorm Greisen: Could the software know, what is signal and what is noise, and therefore be realistic about its own accuracy, and even report its results as a confidence interval/a range of plausible true values?

	Mike Bernstein: Telling the difference between signal and noise is most of the development of a pulse oximeter. **Josh Pfefer: Given that red light (660 nm) is absorbed by melanin 3x stronger than NIR light (940 nm), wouldn't a PI based on red light be more useful for assessing pulse ox signal quality reduction due to skin tone, as opposed to the NIR approach typically used by pulse oximeters?** **T(?): Excellent presentation Dr. Monk. An earlier presentation today briefly touched on differential treatment as a possible confounder in pulse oximetry . How big of an issue if at all do you believe this is in the US healthcare system? And if so what impact might this have on studies that rely on Electronic Medical Record data where factors such as simultaneous arterial and SpO2 measurements are not well controlled?** **Ellis Monk: I think it is quite likely a very important factor. In fact, I am in the midst of designing a set of audit studies to dig deeper into how not only race, but also skin tone, may play a role in bias in patient/clinician interactions.** **Geoff Matthews: Melanin causes an increase in the transmitted red wavelength. Effect on reflective sensors can be quantified by checking the wavelength of the reflected light.** **Julian Goldman: Concept that a pulse oximeter can do a better job letting us know borderline information on suspect data at limits of performance is ideal but comes with challenges** **Francois Lellouche: Really important issue of uncertainty for Oxygen therapy, considering the uncertainty of the measurement combined with the uncertainty of the SpO2 target recommendations (90-94 to 94-98%) !** **Gorm Greisen: As far as I can see, it is even worse: some pulseoximeters gives a standardised plethymorgraphy signal, taking away the possiblity of the clinician to judge the pulse-syncroneous waveform** **Howard Saft: Gorm-Great comment. Clinicians get used to waveforms when they think back to physiology and understand how things work. Giving the perfusion signal should be required for clinical decision making. This, embedded within an educational program, can go a long way to making a difference. (I will defer to the engineers on the degree of filtering, etc.)** **Carolyn Hendrickson: To build on Dr. Monk's point, I think the impact of structural racism on differential health care practices is well**	

	documented and extends beyond models that invoke skin pigment. In medical school many are trained in problem based learning exercises that open with "A 65 year old African American woman comes to a primary care clinic to discuss low back pain." We know that learners and providers are already using heuristics to think about diagnostic and therapeutic that differ depending on if a race cue is provided and differ between race cues that are provided. This is a pervasive and insidious problem for healthcare.	
11:45-noon	**Brief Q&A/Discussion**	
	1.	
noon-1:00 pm	**Lunch** (Provided by UCSF OpenOximetry)	
1:00-2:30 pm	**Strengths & shortcomings of controlled desaturation studies** Challenges in ensuring adequate diversity (not only how we define skin color, but also recruiting adequate diversity in skin color)	Bob Kopotic (moderator) Paul Batchelder John Feiner Bunmi Okunlola David Macleod
	1. Paul B, Chief Clinical Officer, Clinimark Lab, Colorado **a. Controlled Clinical Laboratory Hypoxia Accuracy Trial** **i. Demonstrate Essential Performance or maximum accuracy capability by isolating the device from confounding conditions** **ii. Best-case accuracy capability of a device, or essential performance** **iii. No movement, normal perfusion correct sensor placement, sensor cross-talk, normal levels of physiological variables** **iv. Multifactorial study design** **1. Motion Test** **2. Low Perfusion Test** **v. Pre analytical and Analytical Techniques - Critical Controls** **vi. Reference equipment accuracy, co-oximeter is 1%** **1. inaccuracy of reference is important when published accuracy specification of many pulse oximeters in 2 or 3%** **2. There is no standard reference material for oxygen saturation measured by co-oximeter** **vii. Co-oximeter Calibration Curve Differences between brands of equipment widely used**	

viii. Sampling stability methods during desaturation profile is required. SpO2 measurement and stability at time of drawing arterial samples is critical.
ix. UCSF hypoxia study stability techniques discussed
x. Clinimark Lab saturation stability method discussed

2. John Feiner, UCSF
   a. Desaturation plateau stability and timing
   b. Laboratory techniques explained
   c. Demonstration of software analysis of hypoxia desaturation subject case
   d. Quality check and objective criteria to evaluate study data
   e. Biological data is noisy and challenging even with measurements using gold standard co-oximeter.

3. Bunmi Okunlola, UCSF
   a. Challenges for achieving health equity
      i. Elucidate the role of skin color with more studies
      ii. Catalog device data and make it freely available
      iii. Intention and Reality
         1. Role of skin color in device function must be clearly defined
         2. Methods of categorizing skin color must be updated
         3. Must be intentional about representation in clinical studies
      iv. Barriers to recruitment in clinical studies
         1. Complex historical barriers but can implement solutions
            a. Awareness of studies
            b. Access to studies
            c. Unclear communications
            d. Perception of benefit
            e. Historical context and mistrust
         2. Positive aspects to study enrollment
            a. Altruism
            b. To be a part of something big and positive
            c. Desire to find solutions for themselves or family
            d. Help improve technology
         3. Engaging Communities
            a. Recruit actively

b. Diversity of your teams
c. Involve community leadership
d. Practice effective ongoing communication

4. UCSF Hypoxia Lab Equity Team Solutions
    a. Building diverse team
    b. Background research and education
    c. Updated recruitment materials
    d. Community feedback survey
    e. Idea-sharing and collaboration
5. How can we move forward ethically and responsibly?
    a. clear communication of risks and benefits
    b. language services
    c. health screening for safety
    d. well-planned incentives
    e. Be well educated about the communities working with

Chat:

Bob Kopotic: There was a time when tonometry devices were available for testing/verifying CO-oximetry SaO2 value. Be certain to use Functional SaO2 values from a CO-oximeter to compare to SpO2 values.

Sandy Weininger: note these are characteristics of idealized studies - designed to have ONLY ONE parameter (SaO2) varying so that we can compare the performance of this new device to a standard (or a predicate …)

https://iampov.org/about/ … look for previous symposiums - lots of info on oximeter testing.

Any techniques to measure (get an absolute value) the flow in the digit under the probe?
Gorm Greisen: plethysmography

Bob Kopotic: The adult subjects are healthy (ASA 1), i.e., no COPD subjects are tested in a controlled desaturation study

Zachery Vesoulis: understand the need for control in lab setting--does this method of screening healthy subjects select out a population that has a distinct cardiovascular phenotype that may behave differently than some/much of the general population?

	Bob Kopotic: real-time inhaled & exhaled gas analysis is essential to closely titrating the gases needed for managing hypoxia and isocarbia. **Prescreening subjects for the ability to perform well under hypoxia.** **Sandy Weininer: Is the population tested and to which the pulse oximeters are calibrated represent the clinical population? Why does low perfusion cause this bias in SpO2 accuracy?** **What is the gold standard for characterizing low perfusion? What is the physiological variable for perfusion index?** **Phil Bickler: Need more real world studies to confirm performance as developed using controlled lab hypoxia study. How the assumptions and controls in lab setting apply to real world clinical use.** **S (?): I am really interested in the term stability. Is it per blood draw or per Plateau - how is it defined?** **Jim Stafford: Is it an assumption that improved sampling in studies conducted in the US is representative of other populations that may have a large portion of darker skin tone?** **Gorm Greisen: As I see it: An average result from a study on a representative sample from a population of distinct groups will clearly not be relevant for populations with different proportions of groups. Either results should be group specific or some form of equalizing of results is needed** **Carolyn Hendrickson : Barriers to enrollment thoughts. UCSF has seen support based on current public perception of pulse oximeter equity in health care, being part of the solution. Hesitation can be alleviated with understanding how participation improves healthcare, a moral imperative as research increases benefit to those historically unrepresented.** **Bob Kopotic: Exclusion criteria different for healthy subject hypoxia lab study and real-world clinical study designs. Outlier subjects are real data and may be in the lab study population, tolerating hypoxia, and normal variability in pulse oximeter values is part of the population using technology.** **What is the general population as tested, excluding or including confounders to pulse oximeter measurement accuracy.**	

	Jonas Pologe: Perfusion Index is simply the peak to peak amplitude of the PPG waveform divided by the total amplitude of the waveform. It is only representative of the signal amplitude. The IR wavelength (~900nm) is used for this measurement. **Julian Goldman: @Jonas - So should we use the term optical signal strength in this context (since perfusion is not being measured)? Or another term?** **Sandy Weininger: what are the clinical and engineering uses of PI? if PI is too low, then no confidence in SpO2 estimate?** **Gorm Greisen: In principle PI is a surrogate measure of cardiac output, a hard-to-get clinical variable. Low PI has been shown to predict complications such as severe infection in newborn infants.** **Julian Goldman: Two examples of PI not aligned with flow: 1) On non-pulsatile CPB flow can be high, and PI low. 2) Raise hand above heart (head-height while standing/sitting) and PI will increase - hard to argue that flow/perfusion has increased with that maneuver.** **Jim Stafford: As scientific as this measurement looks , isn't the "calibration" determined by fitting data to subjects in the lab (to produce the lowest Arms) . Then there is a subsequent testing also in the lab (verification). So there seem to be two actual populations that impact the accuracy of the pulse oximeter which may or may not be important to distinguish.** **Sandy Weininger: calibration and verification (qualification) populations should be different** **Howard Saft: Real world studies, may help to optimize both interpretation of the measurement with uncertainty, leading to optimizing clinical decision-making.** **As long as we collect the broad range of data/characteristics, then we can provide information about association. This can continuously improve our understanding of the uncertainty with these devices. It may lead to more specific oximeters for use case scenarios. This may also provide clinicians with continuous improvement of educational material to consider the whole—optimizing interpretation of an uncertain measurement and applying it to optimizing clinical decision making.**	
2:30-3:30 pm	**Statistical analyses** and sample size for controlled desaturation study design (different objectives for design versus qualification)	John Feiner (moderator and speaker) Gene Pennello David Milkes

1. John Feiner, UCSF
    a. Pulse oximeter statistics
        i. Essential Descriptive Statistics
            1. Bias (or error) SpO2-SaO2
            2. Accuracy: Mean Bias
            3. Precision Stand Deviation of the Bias (SD)
            4. Root Mean Square Error (Arms)
            5. Limits of Agreement (LOA)
                a. 1.96* SD
                b. Represents 95% of the data
                c. Adjustment for repeated measurements per subject
        ii. Descriptive statistics include per hypoxia decade groups (60-70, 70-80, 80-90, 90-100) and full saturation range (70-100)
        iii. Modified Bland-Altman plots
            1. SpO2 - SaO2 plotted against SaO2 not average of test and reference
            2. Individual Subject can be shown separately
            3. Mean bias, +/- LOA
            4. Linear regression fit included
        iv. Research studies (Different than verification study requirements)
            1. 95% CI for Arms and limits of agreement
            2. Confounding and interaction
                a. Gender and PI are related
            3. Repeated measures (SaO2, motion, PI)
            4. Repeated measures are extremely robust and may show statistical differences that are not even clinically significant
        v. Receiver Operator Characteristics (ROC) Curves
            1. Describes sensitivity and specificity
            2. Describes overall predictive ability
            3. Can define threshold
        vi. Sample Size
            1. Need to add power analysis
            2. Refines the estimates of descriptive statistics
            3. Increase power of the comparisons
            4. What is the clinically important difference in bias between groups?

5. Cost of increasing sample size, and even the ability to perform studies is a large barrier

2. Gene Penello, Senior Medical Statistician, FDA
    a. Clinical Performance Studies
        i. Accuracy Study
            1. Conventional Design
                a. All subjects have blood draw for SaO2 reference to SpO2
                b. Sample size table presented for different number of groups and allocation of bias for groups
            2. Hybrid Design
                a. Use of secondary (silver) standard pulse oximeter as a comparator
                    i. highly accurate well-calibrated pulse oximeter
                    ii. FDA cleared pulse oximeter
                    iii. The predictive distribution of SaO2 can be learned from initial subjects with blood reference
                    iv. Potentially less burdensome that the conventional design
                    v. Need to adequately power a hybrid study
                    vi. Key to Hybrid Study Success - silver standard must add value to predict SpO2
                    vii. Effective sample size simulation to demonstrate additional information and narrowed confidence intervals on estimate of bias

        ii. Agreement Study
            1. Agreement of a device measurement with a non-reference method is not an evaluation of device accuracy

	iii. Endpoints and Biomarkers for both study designs b. Options for statistics i. ITA Bins for skin pigmentation classification 1. ANCOVA Analysis of Covariance Real World Data, Adults in ICU 2. Linear Regression - detect a linear change in mean bias across continuous ITA 3. Trend Test ii. Continuous skin color analysis iii. Quantile-Quantile (Q-Q) Plot for SaO2 1. Stratified by Ftizpatrick 1-4, 5-6 groups 2. Pure assessment of bias not conflated by regression to the mean c. Accuracy Summaries d. Coverage Interval e. Inverse Prediction Coverage Interval f. Disaggregation of Accuracy g. ROC Plot h. Regression to the Mean i. Account for measurement uncertainty	
3:30-4:00	Open Forum	Zach Vesoulis Valencia Koomson François Lellouche
	1. Zachary Vesoulis, WUSM a. Disparities in optical device performance i. How are infants different? 1. Hemoglobin 2. Prolonged respiratory support 3. Excessive supplemental oxygen exposure is associated with adverse outcomes 4. The role of development on melanocytes 5. Intermittent hypoxia in preterm infants is common 6. Preterm brain injury and hypoxia 7. Implications of pulse ox inaccuracy a. oxygen is an important life-saving drug in the NICU ii. Pulse ox performance disparities in preterm infants	

	1. replicated Sjoding study in <32 weeks GA population iii. Racial disparities in response to hypoxemia iv. Racial disparities in NIRS 1. limited data on race based differences in NIRS performance 2. Different wavelengths used by manufacturers 3. Jugular bulb reference limits population that can be tested v. Conclusions 1. Preterm and term infants have unique developmental characteristic that make them more vulnerable to adverse outcomes from hypoxemia 2. Francois Lelloouche, MD PhD, Heart and Lung Institute Research Centre, Full Professor Laval University, Canada a. Oxygen-Free Days and the Confounders of Clinical Practice - Discussion of the parameter that may influence oxygen utilization b. International guidelines = SpO2 target for oxygen therapy i. What is the impact of 4% differences on oxygen flow? 1. Study titrating O2 flow for different levels of oxygenation (SpO2 from 90-98%) 2. 4% differences in oxygenation -> increased O2 flow by 3.5 3. Impact of the combination SpO2 Target/Oximeter on O2 flow c. Recommend SpO2 targets for hypoxic patients d. Mean difference between pulse oximeters may be up to 4% in patients e. combination of oximeter differences and SpO2 target difference may have major clinical impart	
4:00-5:00 pm	**Day 1 review**, discussion and theme outputs	Sandy Weininger

1. **Open discussion on secondary (silver) standard and hybrid clinical study design**
    a. **Sandy W: Value in using a well performing reference oximeter**
    b. **Mike Bernstein: Bias in oximeter reference would affect results**
    c. **Gene P: Issue with offset taken care of in predicted analysis of SaO2 from blood sampling data**
    d. **Mark Ansemino: Appealing part of hybrid design is evaluating performance in children and specifically preterm infants.**
    e. **Are additional subjects using silver standard different than the initial subjects used to develop it? Limitations are time and money, so do additional subjects still undergoing hypoxia using secondary standard provide benefit?**
    f. **Gorm Greisen: Compare measures and find agreement in readings, physiological comparisons and build confidence by giving up perceived gold standard. Experience basis builds confidence in performance.**
    g. **Paul Batchelder: Hybrid model developed for adult healthy subject lab study but can be extended to pediatric data collection. Less burdensome approach for adult use of secondary standard reduces risk, complexity of arterial invasive procedure. Start off with a transfer standard verified to a large sample size of blood reference using an oximeter with high accuracy**
    h. **Phil Bickler: Not concerned that the traditional study design is overly risky, time consuming or over burdensome, That system is not broken and an additional approach is needed.**
    i. **Michael Lipnick: Can there be certainty using a secondary standard does not add to uncertainty which we are continuously trying to minimize. How is elevating the secondary standard in this way going to improve pulse oximeter performance? Too much we don't know to make such big assumptions for use of secondary standard**
    j. **Gene Penello: Hybrid design is still an accuracy study**
    k. **Linus Park: There are populations hard to enroll subjects (such as preterm infants), does it make sense to use an alternate method? How to incorporate more of the population into the study?**
    l. **Gorm Greisen:**
    m. **Paul Batchelder: Benefits in recruiting more of all populations (eg. deeply pigmented subjects). May**

	very well end up with a controlled lab hypoxia healthy subject study combined with some real world clinical study not controlling for confounding factors. 2. Low Perfusion Discussion Continued a. Jonas Pologe: Perfusion index is not a measure of accuracy, only a size of pulsatile signal to overall size of optical signal. Useful to determine if SpO2 data is potentially of value. Motion affects perfusion index. PI is not a flow indicator or substitute for laser doppler. b. John Feiner: Measure of signal the pulse oximeter is getting. c. Phil Bickler: Need access to raw pulse oximeter data d. Perfusion Index as a term in a clinical way, we should be confident it is measuring something clinically. If it is biased by pigmentation, is the PI accurate for all populations? e. Josh Pfefer: Review ISO 0601-2-61 current signal quality approach f. Julian Goldman: Support changing term Perfusion Index. Recommend to review language being used. g. Wim V: Perfusion Index does not necessarily represent perfusion. h. Bob Kopotic: Agree issue with term of perfusion index and use of pulsatile index in other technology i. Paul B: Percent Modulation is preferred to PI j. How is PI used by clinicians? It should help decide to keep sensor placement, move it, and believe saturation measurement. Is there some pulsatile flow to the sensor? k. Phil Bickler: Variability in PI is an important monitoring parameter during surgery. Indices are being used and important. l. Input from Jonas Pologe on Perfusion Index 3/15/2023: i. Perfusion Index Discussion: There was some discussion about whether or not a change in name is warranted. My sense is that this would not provide any real value. Perfusion Index (PI), is just that, an index. It is related to perfusion but not a direct measure of perfusion. Think about Body Mass Index (BMI) by way of comparison. Not a direct measure of Body Fat, but a general indicator thereof assuming that you know the sex of the person, the musculature of the person, and the age of the	

		person, the bone density, then it will give you a sense of the body fat. ii. One can easily show that PI is an indicator of perfusion. It goes down as the tissue-under-test (i.e. the finger) cools. It goes to zero if you occlude the radial artery and the ulnar artery and the sensor is on a finger in the ipsilateral hand. iii. PI has also been validated as an indicator of perfusion in many studies against various reference measurement modalities. A couple of these references include: iv. Perfusion index as an objective alternative to the Allen test, with flow quantification and medico legal documentation. July 2014 Anaesthesia, Pain and Intensive Care 18(3):245-249 v. This article concludes: 1. Using Doppler ultrasonography as a reference test, perfusion index is a reliable objective test for diagnosis and prediction of abnormal CHC[collateral hand circulation], with high sensitivity, specificity, and accepted positive predictive value. vi. Lima AP, Beelen P, Bakker J. Use of a peripheral perfusion index derived from the pulse oximetry signal as a noninvasive indicator of perfusion. Crit Care Med. 2002 Jun;30(6):1210-3. doi: 10.1097/00003246-200206000-00006. PMID: 12072670. vii. This article concludes: 1. …peripheral perfusion index measurements can be used to monitor peripheral perfusion in critically ill patients. viii. A few additional comments about PI: ix. PI is affected by the oxygen saturation and the wavelength used to make the PI measurement. The infrared wavelength was selected because	

	PI at the infrared wavelength is less impacted by changes in oxygen saturation. **x. A quick literature search shows that PI already has a number of physiological and clinical uses and applications.** **xi. PI is a very good indicator of photoplethysmographic (PPG) signal strength. But it is not the only indicator of PPG signal strength. One must also assess the waveform characteristics, and the instrument must, itself, determine whether the received light intensity is sufficient to potentially provide accurate readings.** **3. What is the clinical range of SaO2 of interest to test? Clinically important difference saturation?** **a. Carolyn Hendrickson:** **b. Sandy Weininger: Is the trend most important?** **c. Bob Kopotic: The actual SpO2 value does have high clinical value not only trending of saturation** **d. Bill Wilson: Focus on what can really affect ability to measure performance in dark pigmentation groups. What can we do now?** **e. Phil Bickler: Advocate keeping less than 80% range that has forced technology improvements and likely created better performance at higher saturations range.** **4. Michael Lipnick - Tomorrow: What are your top 3 changes recommended to make to ISO 80601-2-61 Standard?**	

Day 2, March 15, 2023 (Mission Hall, 3rd floor, room 3700)

Theme II - What are the solutions?

Time (all PST)	Activity	Speakers
08:30- 9:00 am	**Sign-in & Coffee**	
09:00- 11:00 am	**Optimized skin color quantification techniques** to ensure adequate diversity of skin color (pros and cons of current methods & defining optimal methods)	Jenna Lester (moderator)

		Adia Benton and Amy MoranThomas Ellis Monk Leo Shmuylovich Wim Verkruijsse Sandhya Vasudevan
	1. **Ellis Monk:** a. **Review of study population currently required in ISO and FDA Guidance** b. **Review of skin pigmentation subjective scales** i. **Issues with standardization of visual rating scales** ii. **VL, Fitzpatrick, Massey-Martin, Monk Skin Tone scale differences** iii. **Recommend update to FDA Guidance on deeply pigmented subgroup requirement** iv. **Recommend a more inclusive and validated subjective skin tone scale method** 2. **Leo Shmuylovich** a. **Connecting dermatology and pulse oximetry** i. **Dyspigmentation and pulse oximetry** ii. **Would disruptions in melanin distribution change the demonstrated bias in pulse oximetry?** iii. **Review of Individual Typology Angle (ITA) as a continuous measure** iv. **Do ITA and Fitzpatrick scales correlate?** v. **ITA limitations** 1. **ITA can change without melanin changing** vi. **Review of Spectrophotomter -diffuse reflectance** vii. **Limitations to think about** viii. **Access to the raw data could help** 3. **Dr. Sandhya Vasudevan** a. **Ethnicity/Race and Skin Pigmentation in Pulse Oximetry** i. **Epidermal melanin** ii. **Anatomical site considerations** iii. **Subjective and Qualitative pigmentation classification methods**	

- iv. Access to standardized method is a limitation
- v. Rigor in Skin Pigment Assessment as limitation
- vi. Literature review to evaluate feasibility of melanometry for assessing skin pigmentation in pulse oximeter clinical studies
- vii. Objective vs. Subjective Methods
  1. moderate, but variable correlation to FSP
- viii. Inter-Melanometer Agreement
  1. strong correlation between colorimeters
- ix. Repeatability
- x. Melanin/Hemoglobin Crosstalk
- xi. Melanometry vs Gold Standard
  1. invasive gold standard to be used to validate non-invasive methods for clinical trials
  2. Need consensus methods for melanometer characterization with respect to gold standard (validation)
- xii. Validation bis Color Charts
- xiii. Melanometeres provide a non-invasive estimate of epidermal melanin content and enable more accurate evaluation of skin pigmentation than subjective methods
  1. improving melanometer performance
  2. melanometer validation and standardization

4. Wim Virkruijsse
   - a. PMM requirements
     - i. accurate, precise, sensitive
     - ii. Unambiguous, standardizable
     - iii. Practical
   - b. PMM Available
     - i. Subjective
     - ii. Objective
   - c. Matamerism
     - i. two objects have same color under one illumination source but different colors with another lamp
   - d. Two Studies
     - i. Comparing methods

	ii. Repeatability/reproducibility, precision e. Color Scale Matching i. within operator better than between operator consistency f. Conclusions i. Pigmentation across body varies strongly ii. Color matching scales investigated iii. EM meters iv. colorimeter v. What is the root cause of bias? vi. Where to measure skin color? 5. Adia Benton, Associate Professor of Anthropology, Northwestern University a. A Social History of Measuring Color in Skin: recommendations to avoid past errors i. Review of Skin Color Scales/Charts 1. Access to standardized methods a limitation ii. Effect of Skin Pigmentation on Pulse Oximetry in Emergency Department iii. Problems with dose-response iv. Earlier studies showing racial bias dismissed for not quantifying color v. The unceasing significance of colorism vi. Sample innovations of device focused design vii. Self Identifications of Race in clinical studies supported **Chat:** **<u>Relevant sections in standard and FDA Guidance–</u>** **Sandy Weininger: current standard says:** **201.12.1.101.4 Characteristics of the clinical study population for determination of SpO2 accuracy** **a) The summary of the clinical study report used to assess SpO2 accuracy shall:** **1) state whether the test subjects were sick or healthy; and** **2) describe their skin colour, age and gender.** **b) This information shall be disclosed in the accompanying document.** **Check compliance by inspection of the accompanying document.**	

	FDA guidance doc: **We recommend you conduct the study described in Clause 201.12.1.101.2 and Annex EE.2 of ISO 80601-2-61:2011 on 10 or more healthy subjects that vary in age and gender. Your data should include 200 or more data points (paired observations: pulse oximeter, co-oximeter). These data should be distributed as described in Annex EE.2.3.4(g).** **Your study should have subjects with a range of skin pigmentations, including at least 2 darkly pigmented subjects or 15% of your subject pool, whichever is larger.** **<u>Comments:</u>** **From David Busch to Everyone 12:36 PM** **This may be a minor point, but color shifts between printers/monitors can be quite significant. Are printed cards depicting the Monk scale from calibrated printers available?** **From Michael Lipnick to Everyone 12:39 PM** **Hi David, Ellis will likely have an answer for progress on that… FYI we are printing on multiple printers, including photo drugstore and calibrated printers and comparing with colorimetry and in practice how the variability impacts implementation/scoring** **From Wei-Chung Cheng to Everyone 12:40 PM** **How did you measure the color difference dE between the print and the reference scale?** **From Michael Lipnick to Everyone 12:41 PM** **Please continue to share questions in the chat - we will follow this session with Q&A for all speakers at same time** **From TBD to Everyone 12:42 PM** **Since the Monk scale traces to specific Pantone values, color charts can be purchased from X-Rite, who owns Pantone, obviating the need to print from local printers.** **From François Lellouche to Everyone 12:43 PM** **In practice how do you use the Monk skin tone scale ? Do you punch holes in the middle of each colors and compare with the skin (this is how other researchers do)? do you use the colors with the shades or without the shades ?**	

From Wei-Chung Cheng to Everyone 12:44 PM
Which 10 of the 110 PanTone SkinTone patches were used to represent the 10 Monk scale patches?

From Megh Rathod to Everyone 12:44 PM
to clarify, how would you then classify "dark-pigment" on the MST? where would you draw the line (groups G-J, H-J, or I-J?) As a group have we settled on where to distinguish this?

From Josh Pfefer to Everyone 12:49 PM
The "dark pigment" line would vary by tissue site, as different sites have different ranges of pigmentation.

From Mike Bernstein Physio Monitor to Everyone 12:50 PM
All these scales are visible. Should we consider infrared?

From Cameron McGuire - UCSD to Everyone 12:53 PM
I'm struck by the notion that an exquisite amount of data is lost when transforming a continuous variable to a categorical variable

From Megh Rathod to Everyone 12:54 PM
Replying to "I'm struck by the no..."

I second that

From Sandhya Vasudevan - FDA CDRH to Everyone 12:55 PM
How were the Von Luschan and Massey scales expressed in Lab/ITA space?

From Sandy Weininger FDA/ISO to Everyone 12:56 PM
2 objectives of assessing : 1) be inclusive - make sure we capture a cross section of the intended use population and 2) assess melanin (or other aggrevating elements) content to determine any bias that it might introduce (note this is the current hypothesis)

From William Wilson to Everyone 12:58 PM
simple, repeatable, and across the skin color spectrum with even distribution (Massey and Monk come closest)

From TBD2 to Everyone 12:58 PM
CIE color spaces and ITA are based on human perception of color (tristimulus functions), which may not be the best for quantification of melanin, or account for other chromophores. Should diffuse reflectance spectra be the standard instead?

From Cameron McGuire - UCSD to Everyone 12:59 PM
Reacted to "CIE color spaces and..." with 👍🏾

From Michael Lipnick to Everyone 01:01 PM
Replying to "2 objectives of asse..."

objective>subjective
From Cameron McGuire - UCSD to Everyone 01:02 PM
Which software do you recommend for the CM 700d? The CM-SA Skin Analysis Software or the SpectraMagic NX Software?

From François Lellouche to Everyone 01:06 PM
In the first studies we have conducted in patients, we used the Fitzpatrick. In the next studies, we will add the Monk skin tone scale and probably the Delfin device (Ithink this is the one recommended in the OpenOximetry webpage. Other devices should be used ?

From Michael Lipnick to Everyone 01:09 PM
OpenOx not making recommendations… we are using delfin and Konica devices in the clinical trials but only the Konica 700d in the lab studies at present (FP, VL and Monk in both clinical and lab studies)
OpenOx protocols here, including scales being used and downloadable version of Monk - https://openoximetry.org/study-protocols/

From Howard Saft to Everyone 01:10 PM
Thank you. How might others help you obtain access to the raw data?

From Wei-Chung Cheng to Everyone 01:10 PM
Leo: Great presentation. Which CIELAB dataset of Fitzpatrick did you use to convert into ITA?

From François Lellouche to Everyone 01:10 PM
For each method to measure/evaluate skin color, how many measures should be taken and reported (facial/fingers/exposed/non exposed skin areas?)

From Phil Bickler to Everyone 01:11 PM
Really eye-opening Leo! Thank you.

From Cameron McGuire - UCSD to Everyone 01:11 PM
This is an aside but I wonder if some of the dermatologists can weigh in about melanin expression in the subungual region as compared to other body sites in dark-pigmented individuals

	and how this is tied to increased prevalence of subungual melanoma in darker-pigmented individuals. We discussed yesterday that perfusion might be a confounder and that it might not all be melanin. Are there other chromophores we should be considering? **From Michael Lipnick to Everyone 01:12 PM** **Francois - we outline what we are doing in the protocol online. 9 sites at present, 3 measurements per site. We hope to refine that after interim data analysis to a more practical approach (likely forehead and site of oximeter application - e.g. both sides of finger at site where probe applied)** **From Michael Lipnick to Everyone 01:12 PM** **Francois - we outline what we are doing in the protocol online. 9 sites at present, 3 measurements per site. We hope to refine that after interim data analysis to a more practical approach (likely forehead and site of oximeter application - e.g. both sides of finger at site where probe applied)** **From François Lellouche to Everyone 01:17 PM** **Thank you Mike, 2 measures will be certainly more practical! As discussed by Ellis Monk, it would be good to clearly describe how to use the different scale to have an homogeneous utilization in the different groups. Some groups punch holes in the center of the colors to compare with skin of the patients but I do not know if this is well described somewhere.** **From Leo Shmuylovich to Everyone 01:23 PM** **Reacted to "CIE color spaces and..." with** 👍🏾 **From Josh Pfefer to Everyone 01:32 PM** **Seems like the sensitivity of ITA to deoxyHb shouldn't be a significant factor in clinical studies where healthy subjects are used since SaO2 is quantified by CO-ox to ensure that the subject starts at a high level.** **From David Busch to Everyone 01:43 PM** **KM has options for smaller (3mm?) and window-covered apertures (easier cleaning)** **From François Lellouche to Everyone 01:48 PM** **is there a chance that one day, these color sensors could be miniaturized and fit into an oximeter to allow automatic adjustment of the calibration based on color ? alternatively, there should be a switch on the oximeters to manually change**	

	the calibration based on skin pigmentation as suggested in the chat yesterday by Sandy Weininger ? **More date on patients are probably necessary to go there…** **From Leo Shmuylovich to Everyone 01:48 PM** **Francois, if color sensors indeed contain the information necessary to make a calibration, then I think the answer is yes** **From Gene Pennello to Everyone 01:48 PM** **I wonder if the precision SD of the repeated measures varies with skin color or not.** **From Leo Shmuylovich to Everyone 01:50 PM** **The sensors for the micro spectrometer are tiny: https://learn.adafruit.com/adafruit-as7341-10-channel-light-color-sensor-breakout** **Look at the zoomed in picture - and Daniel Franklin is an expert at miniature wearable sensors** **From William Wilson to Everyone 01:50 PM** **why do they study obviously flawed scales like Fitzpatrick and Von Luschan?** **From Wei-Chung Cheng to Everyone 01:50 PM** **Which CIELAB dataset of von Luschan was used in the study?** **From Leo Shmuylovich to Everyone 01:51 PM** **David yes KM has smaller apertures** **From David Busch to Everyone 01:51 PM** **@leo we have purchased the smaller apertures- do you have experience in using them and the window covered version?** **From Sandhya Vasudevan - FDA CDRH to Everyone 01:55 PM** **@Wim - What exactly was SPA?** **From Leo Shmuylovich to Everyone 01:55 PM** **@David yes, we have all 4 sizes,** **From Sandhya Vasudevan - FDA CDRH to Everyone 01:57 PM** **@Wim - You mentioned that Mexameter seems more robust to changes in blood. Is this a newer device that was purchased?** **—** **From David Busch to Everyone 02:04 PM** **Sound is inconsistent- w/ background noise, can this be improved?**	

	From Desireé Conrad to Everyone 02:04 PM **There's a lot of background noise that makes it hard to hear. I don't know it it's just my computer though.** **From Amy Moran-Thomas to Everyone 02:04 PM** **You can learn more about the role that von Luschan scales have played in histories of scientific racism at the US Holocaust Museum:** **https://collections.ushmm.org/search/catalog/irn564926** **From Cameron McGuire - UCSD to Everyone 02:06 PM** **Not just your computer Desiree. I think there are some unmuted laptops in the room where the slides are playing that are picking up and then broadcasting the noise, cutting off the other "speaker" (MH3700@ucsf.edu)** **From Wei-Chung Cheng to Everyone 02:06 PM** **@Wim: On the slide "On color scale matching" you show the metamerism problem of color matching. Did you measure the spectra of the "lab light" and "Verivide"?** **From David Busch to Everyone 02:07 PM** **Could the persons present in person please mute their microphones?** **From Julian Goldman to Everyone 02:10 PM** **(Don't let zoom hear you criticize it)** **From Howard Saft to Everyone 02:11 PM** **Reacted to "(Don't let zoom hear..." with 👍🏼** **From Josh Pfefer to Everyone 02:14 PM** **According to this interesting dissertation, von Luschan's legacy is complicated.** **https://digitalcommons.mtu.edu/etdr/488/** **From Cameron McGuire - UCSD to Everyone 02:14 PM** **Reacted to "(Don't let zoom hear..." with 😆** **From Wim Verkruijsse to Everyone 02:20 PM** **@sandhya, thank you for your questions, the SPA is the SPA 99 model by C&K** **https://www.courage-khazaka.de/en/?view=article&id=196:spa99-e&catid=23:alle-produkte** **@Sandhya, I am not sure what caused why the current Mx18 meter does not find the same impact of blood on MI as I found > 15 years ago. Really no idea.** **From Leo Shmuylovich to Everyone 02:23 PM** **Thank you for sharing the oculocutaneous albinism case. I don't know what the colorimeter devices would show for color measurements for people with albinism**	

	Interestingly the albinism case had falsely low values, but the patient was taking an herbal extract that functions as a UV sensitizer, similar to psoralen I believe https://journals.lww.com/anesthesia-analgesia/fulltext/2004/02000/oculocutaneous_albinism_and_spurious_pulse.60.aspx So whether it's absence of pigmentation or reaction to a photosensitized is unclear - another example of how many different factors may play a role in optical assessment **From Wim Verkruijsse to Everyone 02:24 PM** **@Wei-Chung Chen: I did not measure the illumination spectra of the lab light yet. The spectrum of the Verivide lamp is supposed to be close to D65.** **From François Lellouche to Everyone 02:27 PM** **FDA/ISO will probably provide recommendations on the MINIMAL skin color evaluation (subjective / objective evaluations) required for pulse oximeters evaluation in healthy subjects and in patients (probably not possible to have 4 scales/3devices and 9 measurements for each method, especially in patients)** **From William Wilson to Everyone 02:28 PM** **Monk and Massey very similar and cover the spectrum - key do not allow Von Luschan or Fitzpatrick** **From Michael Lipnick to Everyone 02:28 PM** **Massey has no standardization? Correct? That is a major flaw** **From Ellis Monk to Everyone 02:28 PM** **^I think there are some differences. I will speak to that.** **From Leo Shmuylovich to Everyone 02:29 PM** **I would agree with Ellis that Massey is quite blue at the higher range** **I think Massey is not representative of human pigmentation encountered in practice** **From Josh Pfefer to Everyone 02:31 PM** **What does "too expensive" mean? Colorimeters are not very expensive compared to other costs associated with performing a clinical study, are they?** **From Michael Lipnick to Everyone 02:32 PM** **For regulatory… I don't think any of the solutions on the table is prohibitively expensive**	

	When talking about a multi site clinical study (non regulatory) in low income countries, then buying 20 $10k devices is prohibitive **But that is not the focus here** **From Cameron McGuire - UCSD to Everyone 02:32 PM** **The expense would be generalizing and scaling. If the ideal assessment relies on a 10k spectrophotometer. That is prohibitive for many facilities and/or researchers** **From Vikrant Sharma- Masimo to Everyone 02:32 PM** **Massey and Monk are comparable, in fact Massey as subjective measure is potentially easier than Monk scale (which I think has an advantage in digital space).** **From Ellis Monk to Everyone 02:33 PM** **^Published papers suggest Massey has poor reliability.** **From Wei-Chung Cheng to Everyone 02:33 PM** **The spectral reflectance can change the color detected by cameras and instruments.** **From Sandhya Vasudevan - FDA CDRH to Everyone 02:34 PM** **@Ellis Monk - Poor reliability of Massey - could you point out some papers?** **From Ellis Monk to Everyone 02:35 PM** **https://academic.oup.com/poq/article-abstract/80/2/534/2588836** **From Ellis Monk to Everyone 02:35 PM** **https://academic.oup.com/poq/article-abstract/80/2/534/2588836** **From Sandhya Vasudevan - FDA CDRH to Everyone 02:36 PM** **Thank you Ellis. Have you performed any similar reliability studies on the Monk skin tone scales?** **From Ellis Monk to Everyone 02:37 PM** **Google has, yes.** **Meta/Facebook reports over 90% medium to high confidence using the Monk Scale as well.** **From David Busch to Everyone 02:39 PM** **When discussing clinical studies, I am concerned that the subjective scales are designed for /healthy/ subjects. ICU patients are by definition not healthy; neonates (and especially the very low birthweight premature infants discussed yesterday) have their own issues.**	

	From Jonas Pologe to Everyone 02:40 PM **I confess to being a bit confused. Are we trying to decide what scale to use before we know how we might use it?** **From Julian Goldman to Everyone 02:40 PM** **The word "regulatory" was just used. Not clear what that means in this context. We are discussing science and technology not regs per se. The S&T is needed to improve pulse ox performance and update the standards and regulations** **From Phil Bickler- UCSF to Everyone 02:44 PM** **If you want to take it to anthropologic basics, then a pulse oximeter validation study should reflect the worlds skin color distribution, not a scale. The world is not half "light"** **From Cameron McGuire - UCSD to Everyone 02:48 PM** **If we are attempting to improve pulse oximetry I would propose that SpO2 * (x) = SaO2 reliably "always." (x) could be something as simple as Monk, or ITA, or more complex like a continuous spectral output, or even as complex as x*x'*x''*x'''*x^n.** **From Howard Saft to Everyone 02:48 PM** **I fully agree with Dr. Pfefer--we need to keep the ideal solution as an option. Dr. Monk--this is no reflection on you and your great contributions, but it does need to focus on the real use-case scenario. I will defer to the engineers to sort out which is the best currently, but in the regulation arena--since there is still much work to be done, might we be able to keep the color issue open to some objective measurement, or some group of objective measurements?** **From Victor Ochoa-Gutierrez [University of Glasgow] to Everyone 02:49 PM** **Reacted to "I fully agree with D…" with 👍** **From Gene Pennello to Everyone 02:54 PM** **Linearity is one of the common aspects of analytical validation of a continuous measurement.** **From Cameron McGuire - UCSD to Everyone 02:54 PM** **+1 to that last live comment** **From Howard Saft to Everyone 02:55 PM** **It seems like we are in the process of both creating an ideal (silver standard) without cost limitations. Then, allowing the engineers to find lower cost solutions that make it close to**	

	ideal and more accessible. Doing this and then working our way down from cost/limitations might be help us with a road map to solve this problem in a few years and have clear understanding of the lower cost limitations. **We might want to think of what standards do we need now, based on the information that we have—along with an educational component. Then, consider a plan to readdress the standards in 2-3 years, based on the hope that we make progress based on the silver standard.** **From Gene Pennello to Everyone 02:56 PM** **Linearity is defined in the CLSI harmonization database https://htd.clsi.org/default.asp as ability (within a given interval or range) to provide results that are directly proportional to the concentration (or amount) of the analyte (measurand) in the test sample.** **From Sandy Weininger FDA/ISO to Everyone 03:00 PM** **maybe we can put some proposals on the table and get the sense of where to go. Ex: use Monk scale and collect equal representation in each 3rd (versus equal numbers in each category).** **From William Wilson to Everyone 03:00 PM** **in addition to defining color tone, we need to increase the % of dark skinned individuals in FDA validation studies to >35-40%** **From Ellis Monk to Everyone 03:01 PM** **^There are ITA cutoffs for the Monk Scale that could be very suitable. Leo helped calculate them.** **From Michael Lipnick to Everyone 03:01 PM** **!!! As Sandy proposed - everyone please add to the chat your thoughts on solutions/ways forward** **From David Busch to Everyone 03:01 PM** **@sandy Per earlier discussion, do we need to make sure we include the outermost bins?** **From Victor Ochoa-Gutierrez [University of Glasgow] to Everyone 03:01 PM** **Algorithms need to be improved, ratios of ratios was good in the past but we need to have better optical models, where, light scattering, water, fat, melanin and other compounds in human tissues are included in pulse oximeters devices. The use of more wavelengths (without compromising costs)** **From Daniel Franklin to Everyone 03:02 PM** **Reacted to "Algorithms need to b..." with 👍** **From Adia Benton to Everyone 03:05 PM** **You also need clinical standards that aren't reliant on this metric alone — this was how this problem came to the fore again** **From Odi Ehie- UCSF to Everyone 03:06 PM**	

	I also agree with Adia on this. From Sandy Weininger FDA/ISO to Everyone 03:06 PM can we not make our best educated guess? we can (and must) revisit periodically (standards must be updated at least every 5 years). From Victor Ochoa-Gutierrez [University of Glasgow] to Everyone 03:07 PM Reacted to "can we not make our …" with 👍 Reacted to "You also need clinic…" with 👍 Reacted to "I also agree with Ad…" with 👍 From Gregory Leeb to Everyone 03:07 PM Reacted to "can we not make our ..." with 👍 From Odi Ehie- UCSF to Everyone 03:07 PM Reacted to "Algorithms need to b…" with 👍🏾 From William Wilson to Everyone 03:07 PM Difficult to find people with Monk 1 and 10 Would consider volume of Bins 1 & 10 requiring 0-1 individuals, bins 2 & 9 requiring 2-3 subjects, and bins 3-8 requiring 4-5 subjects each and bins - increase the total number of subjects to at least 30. From Ellis Monk to Everyone 03:08 PM Binning would be fine. We could discuss the fine details. From William Wilson to Everyone 03:09 PM this is inclusive of full spectrum and not quite linear (As the population is not linear across the spectrum From Ellis Monk to Everyone 03:10 PM Reacted to "this is inclusive of..." with 👍 Absolutely, agree. We need ethnoracial and skin tone diversity. From Josh Pfefer to Everyone 03:10 PM agreed From Sandy Weininger FDA/ISO to Everyone 03:10 PM must be able to measure a characteristic to assure equity From Odi Ehie- UCSF to Everyone 03:12 PM To Mike Lipnick's point, the way forward is to be intentional about racial inequity. While there is a lot we don't know, we need to be intentional about addressing a more accurate measurement based on ethnoracial and skin tone diversity as Ellis stated From Wei-Chung Cheng to Everyone 03:14 PM @Leo: Agreed. Targets for subjective assessment and metrics for objective assessment do not need to be the same. From François Lellouche to Everyone 03:17 PM To this point, we do not have enough data in patients to figure out if the skin pigmentation will have minor (within 1% which is not my hypothesis) or major impact on bias (above 2%) for the same oximeter, from I to VI on the Fitzpatrick scale or A to J on the Monk scale. Consequently it is difficult to know if 2 or	

	3 different oximeters calibrations will be required to adjust for skin pigmentation vs a continuous changed calibration… for now we may use Fitzpatrick and Monk as well as one colorimetric methods. And this will have to be regularly reevaluated … with reasonnable number of measures in patients. **From Gene Pennello to Everyone 03:17 PM** **If self-reported race/ethnicity has effects on pulse oximeter accuracy after adjustment for skin color effects, further study may be warranted on underlying factors associated with race/ethnicity that may have causal effects on pulse oximeter accuracy** **From Odi Ehie- UCSF to Everyone 03:17 PM** **Agreed the recruit of subjects need to also represent an ethnoracial and skin tone diversity** **From Gene Pennello to Everyone 03:17 PM** **Kroll and Emancipator method for linearity is coded here. https://analyse-it.com/docs/230/user-guide/method-evaluation/linearity. CLSI EP06 is a standard on statistical evaluation of linearity using goodness-of-fit test statistics. A paper is Kroll MH, Emancipator K, Floering D, Tholen D. An algorithm for finding the linear region in a nonlinear data set. Comput Biol Med. 1999 Sep;29(5):289-301. 10463796.https://www.sciencedirect.com/science/article/pii/S0010482599000116?via%3Dihub. A classic paper is Kroll MH, Emancipator K. A theoretical evaluation of linearity. Clin Chem. 1993 Mar;39(3):405-13. PMID: 8448849. he CLSI standard on linearity is shown at https://clsi.org/media/1437/ep06a_sample.pdf**	
11:00 am- noon	**Discussion**	Jenna Lester and Kelvin Moore (moderators)
	1. Should any subjective scale be used in clinical verification of a pulse oximeter? **a. Study inclusion criteria for ensuring a diverse test population may use subjective scale such ast Monk method, but to evaluate device performance an objective measure of melanin at the correct location of body may be the appropriate method** **b. Lower cost options to spectrophotometers?** **c. Micheal Lipnick: Regulatory approval for device performance should use objective measures**	

	d. Paul Batchelder: Subjective scales when not appropriate appropriate to use objective measuring equipment (e.g. neonates) **e. What about binning groups to levels such as light, medium, dark or light, dark for performance analysis using either of the methods?** **f. Josh Pfefer: What is the best scale for this particular purpose of statistical analyses? Current scales were not specifically developed for this purpose, linearity?** **g. Ellis Monk response: goal to be inclusive with scales** **h. discussion re linearity of Monk scale** **i. Sandy W - two objectives of std - to be inclusive and where is the bias coming from..** **j. BK - remember - POX influenced by absorbers and scatter**	
noon- 1:00 pm	**Lunch** (Provided by UCSF OpenOximetry)	
1:00- 2:00 pm	**How to account for challenging clinical scenarios and potential roles for phantom in vitro testing**	Sandy Weininger (moderator) Mike Bernstein Chetan Patel Josh Pfeffer Geoff Mathews and Veronica Hickson
Relevant section of ISO standard (per SW)	**201.12.1.104 Pulse rate accuracy** **a) If equipped, pulse rate accuracy shall:** **1) be disclosed in the instructions for use; and** **2) stated as the root-mean-square (rms) difference between paired pulse rate data recorded with the pulse oximeter equipment and with a reference method.** **b) Pulse rate accuracy shall be stated either:** **1) over the full claimed range of the pulse oximeter equipment; or** **2) as separate pulse rate accuracy specifications over segments of that range.** **c) The reference method for the computation of pulse rate accuracy may be:** **1) an electronic pulse simulator;** **2) ECG heart rate;** **3) palpated pulse;** **4) thoracic auscultation; or** **5) a second pulse oximeter equipment which has been qualified by comparison to one of these references.**	

	d) The reference method for the determination of pulse rate accuracy shall be disclosed in the technical description. Check compliance by inspection.	
	1. **Mike Bernstein:** a. **Simulator Testing** i. **Test the hardware capabilities of a pulse oximeter** ii. **Transmission/Skin Color** 1. **There is no technical definition of dark skin yet. The simulator can position an oximeter's low transmission performance within the range of marketed devices** iii. **Perfusion/Heart Rate** 1. **Perfusion is very difficult to control in a hypoxia test. The simulator can cover the full normal range of perfusion** 2. **The simulator allows determination of the minimum perfusion required for the pulse ox to report a saturation** 3. **Heart rate is difficult to control in human tested** iv. **Challenging a Design** v. **Conclusion** 1. **Cannot use a simulator for accuracy determination** 2. **Dr. Chetan Patil, Optical Diagnostics Research Lab, Temple University** a. **In Silico Simulations of Pulse Oximetry Accuracy and Bias** b. **Original Pulse oximeter theory based on a simplified approximation** c. **Improved understanding of pigmentation dependence of R (Oximetry Ratio)** i. **increased pigmentation in epidermal layer reduces overall signal intensity, and decreases oximetry ration which is used for SpO2 prediction** d. **Expand modeling as a tool to evaluate the entire calibration and test process** e. **To predict SpO2, a calibration equation is necessary**	

f. Review Monte Carlo simulations of pulse oximetry calibration and testing
    i. Transmittance finger clip
    ii. Stochastic sampling of optical properties
g. Effect of pigmentation on Oximeter Ratio (R) and Calibration
    i. R is not independent of pigmentation
h. High melanin test cohorts exhibit overestimation across nearly all saturation levels
i. SpO2 estimate variability increases at low oxygen levels
j. Modeling calibration and test populations can estimate error magnitudes of skin pigmentation
k. In Silico simulations could inform calibration/validation study design

3. Josh Pfefer, FDA
    a. Phantom Based Performance Test Methods for Pulse Oximetry
        i. Review of Optical device phantom work at FDA
        ii. Skin Pigmentation in Optical Diagnostics
            1. Major Impact of melanin on measured signals (reflectance NIR devices)
        iii. Roles for Pulse Oximetry Phantoms
        iv. History of Pulse Oximetry Phantoms
        v. Phantom Requirements
        vi. Phantom Design: Epidermal Absorption Coefficient
        vii. Recent FDA Phantoms with Pigmented Epidermis
        viii. Pulse Oximetry Phantom Work in Progress
4. Geoff Matthews
    a. Lab studies do not account for QI/QC; Challenges in assuring devices meet standards
        i. Vision for the Future
        ii. There is an issue with Pulse Oximeter Sensor Accuracy
        iii. Melanin Pulse Oximeter Bias
        iv. Pulse Ox Sensor Calibration and Accuracy
        v. The magnification of errors
        vi. Clinicians understanding of Performance

	vii. Regulatory Situation - Melanin Bias viii. A Technical Solution to Melanin Bias ix. Regulatory Solution to inaccurate sensors x. Accuracy and Implementation xi. Real World Investigation xii. Recommendations to ISO 5.	
	comments From François Lellouche to Everyone 04:18 PM In the yesterday Chat: Would it be possible to have a switch to change the calibration of the oximeters ? Yes indeed, but to date we do not have enough data IN PATIENTS to recommend this. Based on comparisons of oximeter's in patients and on the literature: Data produced 33 year ago by Amal Jubran and Martin Tobin (Chest 1990) demonstrated the positive bias in patients with dark skin pigmentation (overestimation of the oxygenation in ''black'' patients vs. ''white'' patients) and they recommended to increase by 3% the minimum SpO2 in ICU black patients (95% vs 92% in white patients = Jubran Review in ICM 2004). Data in healthy subjects (2005, 2007 UCSF studies) were reassuring and did not found impact of skin pigmentation for usual SpO2 values (80-100%), but found impact during severe desaturation (60-80%). Recent data during the pandemic demonstrated that there was a significant bias with potential impact of outcome related to occult hypoxemia in patients with dark skin pigmentation (Michael Sjoding study first (letter NEJM 2020) and many other studies since confirmed these differences in SpO2-SaO2 bias). The limitations of the delay between pairs in retrospective studies was acknowledged and underlined by Dr Sjoding. From François Lellouche to Everyone 04:18 PM In patients, recent data show that some oximeters have negative bias (-3%) in light skin patients. Previous data showed that the same oximeter brand had no bias in a study conducted in India (Singh 2017)(this was not evaluated but likely in patients with darker skin) (same brand but not same model = was it the same calibration?). Some other oximeters that have no bias in light skin patients have +4% bias in a study conducted in indian patients (Amalakanti 2016).	

	It seems that the bias are shifted up with darker skinned patients. By how much, this is unclear and probably depend on each oximeter. **From Ravi Durbha to Everyone 04:19 PM** **Does Perfusion index measurement in transmissive and reflective based pulse oximeter differ?** **From François Lellouche to Everyone 04:19 PM** **We will have more data in patients with dark skin pigmentation with studies conducted with the same design as we did (OXYGAP study) in Cincinnati and Montreal = comparison IN PATIENTS of several oximeters with CONCOMMITENT arterial blood gases.** **There are larger bias in patients in comparison with healthy subjects (highlighted by Dr Okunlola and UCSF team in Respiratory care last year).** **They propose and I can not agree more that studies are conducted in PATIENTS in addition to healthy subjects trials.** **It is not possible to use an UNIVERSAL SpO2 targets, for all patients (whatever skin pigmentation and oximeter brand).** **It will be necessary either (from the easiest to the most debatable solutions)** **(i) to use different calibrations specific to the skin pigmentation (this could be done manually by swithching based on the skin pigmentation evaluation (visually or with Delfin-like sensos) or automatically based on the detection of the skin pigmentation bythe oximeters with miniaturized Delfin technology embedded in the oximeters)** **(ii) to change the SpO2 targets based on the patient's characteristics (skin colour +++, perfusion ???) = corrected SpO2 targets** **(iii) To accept SpO2 corrections within oximeters based on the patient's characteristics….** **From Sakari Lamminmaki to Everyone 04:35 PM** **Per my experience there are lots of limitations in commercially available simulators in terms of pulse oximetry performance testing. As an example:** **-simulator noise floor depend a lot on the pulse oximetry LED intensity – they might work not properly with wireless low power pulse oximetry devices using low LED currents and duty cycles** **-simulator response time is too slow to repeat accurately the short LED pulses used in low power wireless pulse oximeters** **From David Busch to Everyone 04:58 PM**	

	@chetan, in the simulations, what range of scattering and anisotropy did you use? **From Josh Pfefer to Everyone 05:01 PM** **Very nice talk Chetan! How did you arrive at your maximum melanin volume fraction value? Your max value is about half that identified by Jacques. Are you estimating the level in the palmar finger?** **From Daniel Franklin to Everyone 05:03 PM** **Hi Chetan, does methemoglobin have different scattering properties than other forms of hemoglobin? If absorption alone doesn't modify R, in the case for melanin?** **From Ravi Durbha to Everyone 05:13 PM** **@chetan, does reflectance based pulse oximeter remove the bias due to nail polish, and also measuring skin tone on the same side, as the placement of sensor minimize the bias due to skin tone.** **From Mike Bernstein to Everyone 05:16 PM** **mainly for testing hospital monitors**	
2:05- 3:00 pm	**Opportunities to utilize Standards for improving communication of device performance** to minimize potential misdiagnosis and/or health disparities from pulse oximeter misuse	Mike Lipnick
	1.	
	From David Busch to Everyone 05:37 PM **@Geoff- could you please give some examples of how big a difference in central wavelength you are finding? Also, FWHM of spectra.** **@geoff Please clarify- bias in your graphs is in SpO2? or Wavelength?** **From Michael Lipnick to Everyone 05:44 PM** **What predicts real world performance in the 80-100% better - 1. A device with Arms<3% 70-100% or 2. A device with Arms <3% 80-100%** **From David Busch to Everyone 05:45 PM** **@michael if we are narrowing the range, how important is accuracy between 95-100 %** **From Julian Goldman to Everyone 05:45 PM** **Seems that "sensors must be on frequency" should be stated as "emitters must be on frequency"** **From William Wilson to Everyone 05:46 PM** **all solutions should increase accuracy across the range. Arms 3 is still way too high . Arms <2% across the**	

range should be goal - may need to be phased in. <2.5% starting now and <2% in one year.

From David Busch to Everyone 05:49 PM

We had the suggestion yesterday to have different devices for congenital heart defect use. Are specialized devices for specific populations the way to go?

From Phil Bickler to Everyone 05:50 PM

I understand why wavelength degradation would be a problem, but what studies in real world situations show this? Sorry to miss that

From William Wilson to Everyone 05:50 PM

note: regarding re-processed sensors, should have calibration information that matches the LEDs. These sensors are just dirty and sloppy

From Howard Saft to Everyone 05:50 PM

Who, How, When, Where do you see the auditing process occurring?

From William Wilson to Everyone 05:51 PM

often have plastic partly covering LEDs and debris that are not allowing optimal characteristics

From michael lipnick to Everyone 05:52 PM

It's also worth mentioning that 3rd party reprocessing of sensors pose an additional issue with performance

From William Wilson to Everyone 05:53 PM

reprocessed sensors commonly do not meet specs

From Gene Pennello to Everyone 05:54 PM

SpO2 will tend to have worse Arms at low SaO2 values in part because of regression-to-the-mean, which is not actually a bias, but a statistical artifact due to random variation. RTM is the reason why Bland and Altman plot recommend plotting the difference Y - X vs the mean (Y+X)/2, not Y - X vs X. See Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet. 1995 Oct 21;346(8982):1085-7. doi: 10.1016/s0140-6736(95)91748-9. PMID: 7564793.

From Megh Rathod to Everyone 06:04 PM

Would the difference between measurement at the forehead vs finger be attributed to the time it takes for low O2 values to reach the vasculature at the fingers. Curious if those oximeters were compared at the same time points or values changed at all over time 'to account for the lag'

From François Lellouche to Everyone 06:22 PM

ARMS is impossible to understand for most users

From Rogério Pugliesi Almeida Gomes to Everyone 06:23 PM

	Dear colleagues. It was a pleasure to stay with you until now. Thank you for this Forum. I have to leave you now. Good nigth from Brazil. Best Regards for all.Bye!! **From Gene Pennello to Everyone 06:24 PM** **I would communicate to the user an interval of plausible values of Sao2 given the SpO2 value, that is, characterize the measurement uncertainty. The interval is constructed to include the true SaO2 value with a stated probability.** **From François Lellouche to Everyone 06:25 PM** **should we use bias below 1 or 2% ? at least for communication - this could help at least to correct fpr the SpO2 target** **the bias should be provided for at least light/medium/dark skin** **From Sandy Weininger FDA/ISO to Everyone 06:28 PM** **https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm?start_search=1&submission_type_id=&devicename=oximeter&productcode=&deviceclass=&thirdparty=&panel=&regulationnumber=&implant_flag=&life_sustain_support_flag=&summary_malfunction_reporting=&sortcolumn=deviceclassdesc&pagenum=25** **https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=870.2700** **From Louise Mulroy to Everyone 06:34 PM** **In UK a manufacturer may claim compliance with this standard to demonstrate their device is safe, effective but is not mandatory for a manufacturer to comply with this standard.** **From William Wilson to Everyone 06:38 PM** **All pulse oximeters used for medical decision making should be FDA Cleared - Labeling should require only Medical Grade Pulse Oximeters with current Arms requirements to be used for medical care.** **From Howard Saft to Everyone 06:43 PM** **I agree to that last comment and this can be a focus area. QR Code, plus some basic information.** **From Desireé Conrad to Everyone 06:45 PM** **Why would we give them subpar** **From Sandy Weininger FDA/ISO to Everyone 06:47 PM** **i went to Walgreens - the device were silent on clearance or ISO standard conformance** **From jim stafford to Everyone 06:48 PM**	

WHO PQ

From Sandy Weininger FDA/ISO to Everyone 06:49 PM
for procurement, specify product code : DQA
From Gene Pennello to Everyone 06:50 PM
A figure displaying the prediction interval of plausible SaO2 value given every SpO2 could be provided. An example of such a figure is athttps://journal.r-project.org/archive/2014/RJ-2014-009/RJ-2014-009.pdf

From Howard Saft to Everyone 06:50 PM
I agree that we night higher standards. How do we differentiate it from a 'silver standard' device, though? Should they all be 'silver standard' devices?

From David Scott UK to Everyone 06:50 PM
Low resource countries cannot afford the high standard and cost of US medicine, and that is why subpar is good for them

From Louise Mulroy to Everyone 06:51 PM
Having trouble with my microphone

From Sandy Weininger FDA/ISO to Everyone 06:55 PM
i suspect the cost of making a device more robust (e.g. physically able to withstand abuse, prevent battery theft) might be more than the cost of delivering adequate clinical performance
From Michael Lipnick to Everyone 07:01 PM
Replying to "Low resource countri..."

I do not agree with this statement. Dvices must be context appropriate. Multiple initiatives have demonstrated that safe, affordable devices can be procured and implemented in LMICs without compromising costs
From Sandy Weininger FDA/ISO to Everyone 07:01 PM
might want a digital interface for calibration/qualification purposes
From Megh Rathod to Everyone 07:02 PM
Reacted to "I do not agree with ..." with 👍🏽
From Kelvin Moore to Everyone 07:03 PM
Reacted to "I do not agree with ..." with 👍🏽
Removed a 👍🏽 reaction from "I do not agree with ..."
Reacted to "I do not agree with ..." with 👍🏽
From David Scott UK to Everyone 07:05 PM
Replying to "Low resource countri..."

	happy for you to disagree, Michael, but that will not change my mind. One area where the problem is biggest is anaesthetic equipment for low-resource countries. **From Howard Saft to Everyone 07:05 PM** **Julian, Do you address this in your usability work? Interpretation and adequacy?** **From David Scott UK to Everyone 07:07 PM** **Replying to "Low resource countri..."** **I do agree with you about the lack of waveform, Michael, from dave Scott** **From François Lellouche to Everyone 07:07 PM** **Fully agree with Mike, the waveform is mandatory** **From Howard Saft to Everyone 07:08 PM** **Reacted to "Fully agree with Mik..." with 👍🏽** **From Julian Goldman to Everyone 07:09 PM** **We have been addressing the specifics of information to be communicated in our research at MGH. Especially to enable better research data sets, closed loop algorithms, and autonomy. I dealt with the issue of when to display / blank data more so in the past when I worked in the POx industry.**	
3:05- 4:00 pm	**Open forum**	Kelvin Moore (moderator) Gauri Singh
	1. Gauri Singh, LifeBox **a. Challenges to safe and equitable pulse oximetry** **i. LifeBox and Pulse Oximetry** **1. WHO Surgical Safety Checklist** **2. Close the "oximetry gap" by equipping operating rooms and post-anesthesia care units with PO** **3. worked with WFSA on the specifications (durable, lowe-cost, high quality)** **4. COVID pandemic -> demand of pulse oximeters surged**	

	5. Need to diversity supply of reliable devices **ii. Real World Challenges**	
4:00-5:00 pm	**Day 2 review** & discuss proposal	Bob Kopotic
	1.	
	From Sandy Weininger FDA/ISO to Everyone 07:57 PM **50:50 dark/light?** **From Gene Pennello to Everyone 07:57 PM** **Yes! Need to express the measurement uncertainty with an interval of plausible SaO2 values given the SpO2 value. Otherwise, user will take the SpO2 value as the God given true value of SaO2, which it isn't.** **From Sandy Weininger FDA/ISO to Everyone 07:58 PM** **how to communicate realistic ability of pulse oximeters to deliver performance?** **From William Wilson to Everyone 07:59 PM** **3 Requests for ISO enhancement: 1) Increase dark skin tone requirement from 15% (or 2 subjects) to 35% (or 10)** **2) Increase sample size to 25-30 subjects. 3) decrease Arms to <2.0** **From David Scott UK to Everyone 08:01 PM** **The possibility that variation in LED frequency is responsible for the apparent optimistic readings in dark skinned patients needs determined before starting the other studies. From David Scott** **From François Lellouche to Everyone 08:05 PM** **Healthy subject studies are required to better understand physiology, evaluate new technologies, but these studies do not describe adequately the real bias and precison in the real world.** **Patient's study are required and more data are required in patients with medium and dark skin pigmentation.** **Bias and precision in patients should be better communicate to users, rather than ARMS that are not understood by the vast majority of users.** **Knowing the oximeter bias in different groups of skin pigmentations should allow to correct the SpO2 target use for a specific patient with the specific oximeter**	

	used in the unit (usually the same used in a given clinical unit) rather than using an universal SpO2 target (that does not fit all patients and all oximeter). From Michael Lipnick to Everyone 08:06 PM All Panelists/speakers please raise your hand if you would like to add comments From Geoff Mathews to Everyone 08:11 PM Sensor accuracy can be checked by using a spectrometer. From Gene Pennello to Everyone 08:11 PM My statistical wish list: 1. Size study to show clinically acceptable performance goal. 2. Convey SaO2 measurement uncertainty by providing an interval of plausible SaO2 values conferred by the SpO2 value. 3. Provide the ratio of the odds of hypoxemia given the SpO2 value compared to the pre-test odds. For example, an odds ratio of 10 means the chance of hypoxemia has increase 10 times relative to the pre-test odds. The reason to provide the odds ratio is that it is independent of hypoxemia prevalence, which can be vary by population, thus the odds ratio is more portable across populations. From William Wilson to Everyone 08:12 PM If we look at a bias difference (SpO2-SaO2) between dark and light skin individuals 1-2 % is too large and coule lead to errors of care (ie deaths). Bias target should be < 0.5% across the spectrum of 100% -70% From Veronica Hickson to Everyone 08:12 PM Accuracy values for the tolerances of the LEDs to be used in manufacture of sensors need to be determined as within requirements. The system will only perform in the real world if the replacement parts (sensors) are the same as those used in the lab tests. We cannot improve the bias due to melanin levels without improving the overall accuracy of systems used in the real world. From Vikrant Sharma- Masimo to Everyone 08:13 PM Also to emphasize again, a larger sample size is mandated for full representation. From William Wilson to Everyone 08:14 PM Regarding the bins sizes (using Monk scale) consider a sociological survey of universal skin tone spectrum (%	

	of numbers within each Monk scale bin) and fill them accordingly. From Desireé Conrad to Everyone 08:14 PM I advocate for inclusion of children as adult studies do not easily transfer to pediatric population especially with device size From Ellis Monk to Everyone 08:14 PM I think the Fitzpatrick and von Luschan scales are inappropriate for use as visual/subjective skin tone scales for regulatory purposes (i.e. a standard). The 2/15% 'darkly pigmented' standard should be changed We need better representation of darkly-pigmented study participants. That number should be increased, significantly. In fact, we should consider over-sampling both lighter and darker-skinned subjects. This is not, however, to say we should not include those in the 'intermediate' categories. We must absolutely include people across the entire skin tone continuum. The goal in over-sampling is to make sure we have adequate inclusivity and statistical power to detect potential biases. Linked to #1, I would propose the usage of the Monk Skin Tone Scale, when the usage of a visual color scale is appropriate (e.g., study recruitment). It is probably inadvisable to have no agreed upon standard. How would studies be compared? I fully agree, however, that this shift in standards should be re-visited in 5 yea	

Logistics Resources

Airport information
San Francisco is serviced by two international airports:
San Francisco International Airport (SFO) is located approximately 15 miles from San Francisco downtown hotels (30–60-minute drive).
Oakland International Airport (OAK) is approximately 20 miles from San Francisco downtown hotels (30–50-minute drive).

Transportation from the airport
Oakland International Airport ground transportation options and contact information
San Francisco International Airport ground transportation options and pricing
Super Shuttle provides shared rides from SFO or OAK; book at least 24 hours in advance and prepay with a valid credit card
Airport Express Van Service offers SFO and OAK airport shuttle service

Getting Around in San Francisco

Clipper Card allows you to pay for all Bay Area transit-bus, rail, ferry and train

BART is the Bay Area Rapid Transit offers low-cost, fast, and reliable transportation to downtown, nearby airports, surrounding communities and more
San Francisco Municipal Transportation Agency runs a network of fuel-efficient Muni buses, light rail Metro trains, and historic streetcars
Caltrain's Baby Bullet Express Service makes it possible to travel between San Francisco and San Jose in less than an hour, stopping at a few popular stations
UCSF Campus Shuttle

SF restaurants, accomodation and city Information
SF Visitor's Guide

Map of UCSF Mission Bay Area
UCSF has multiple campuses. Please note this meeting will take place at the Mission Bay Campus. The meeting will be located in the **Mission Hall Building at 550 16th Street**, SF, CA, 94158 (the corner of 16th street and 4th street)

UCSF Parking
Parking on campus can be tough. Meters are $5/hr and often a maximum of 2 hours. The garages are $35/day and can fill up. We recommend traveling by rideshare or public transportation to get to campus. Please visit this page for additional information: Mission Bay Parking information

UCSF COVID-19 Guidance
Vaccinated Non-UCSF guests (individuals that are not UCSF faculty, staff and learners including vendor staff) must complete the Digital Screener and follow UCSF screener requirements.

Unvaccinated Non-UCSF guests must receive a negative COVID-19 antigen test performed each morning of the event and complete the Digital Screener and follow UCSF screener requirements.

Mission Hall Building Security and Access

Please check in at the security desk upon arrival at Mission Hall with a government Issued ID each day of the workshop. Security will provide a sticker for the day and access up to the 3rd floor. The conference rooms on the 2nd floor can be accessed via the stairs in the lobby.

Mission Bay Restaurants

Sparks Social food truck lot
SF Kebab
Oda Restaurant
Cafe 24
Esposto's Delicatezza

For questions & correspondence:

- Fekir.Negussie@ucsf.edu for UCSF logistics and in person attendance
- Michael.Lipnick@ucsf.edu for conference schedule
- bob@kopotic.com and Sandy.Weininger@fda.hhs.gov for questions related to ISO JWG