Pulse Oximeters: Technology, Accuracy Limitations, and Regulation

James J. Lee Ph.D., Division Director
Bradley Quinn, Assistant Division Director
Neel Patel M.Eng. Senior Reviewer

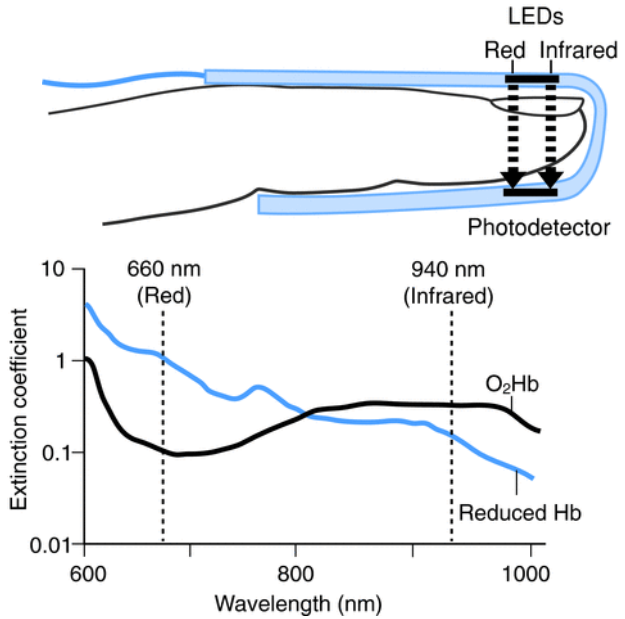
DHT1C: Division of Sleep Disordered Breathing, Respiratory and Anesthesia
OHT1: Ophthalmic, Anesthesia, Respiratory, ENT and Dental Devices

Background - Fingertip Pulse Oximetry

- Essential use to obtain an indirect measure (SpO_2) of arterial blood oxygen saturation (SaO_2) in real time monitoring
  - Expedient surrogate measurement of oxygen level in hemoglobin
  - SaO_2 measurement, obtained via arterial puncture, is considered the gold standard for assessment of blood oxygen saturation levels
  - SpO_2 is an estimate of how much oxygen the hemoglobin contains compared to how much it could contain, expressed as a percentage
  - The Coronavirus Disease 2019 (COVID-19) pandemic resulted in increased use of pulse oximeters in the hospital and home settings

Principles of Operation

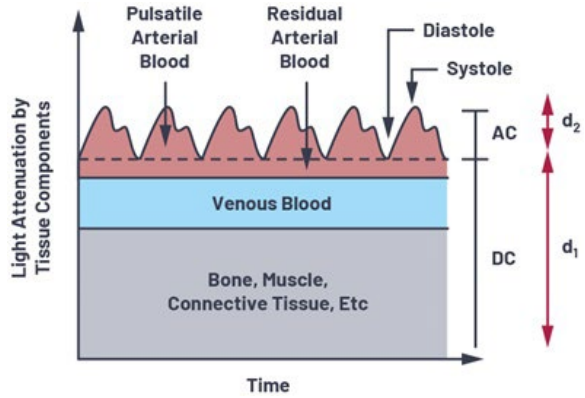
- Oxygenated and deoxygenated hemoglobin have different absorption spectra
- Photodetector used to measure the differential absorption of 2 or more wavelengths of light (typically red 660 nm and infrared 940 nm)
- SpO_2 estimated as percentage of oxygenated hemoglobin to oxygenated + deoxygenated hemoglobin
  - Does not account for presence of dysfunctional hemoglobin



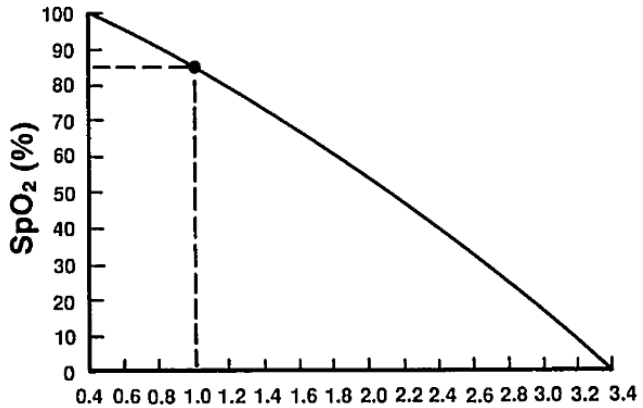
Xiong, Z., & Kodali, B. (2011). Pulse oximetry and capnography. In C. Vacanti, S. Segal, P. Sikka, & R. Urman (Eds.), *Essential Clinical Anesthesia* (pp. 186-190). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511842306.028)

Principles of Operation

- Absorption ratios change with pulse with arterial blood
- Ratio of absorption ratios converted to SpO_2 using algorithm and lookup table developed with calibration data



https://www.analog.com/ru/technical-articles/how-to-design-a-better-pulse-oximeter.htmld text



$$R = \frac{AC_{660}/DC_{660}}{AC_{940}/DC_{940}}$$

Deshmane, Anagha. (2009). False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram.

- Pulse oximeters can be categorized as:

  - Pulse Oximeters Intended for Medical Purposes: Regulated under product codes including, but not limited to, DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter). Following are considerations for certain pulse oximeter devices that FDA regulates:

    - Reviewed by the FDA through 510(k) pathway;
    - Clinical testing used to confirm accuracy
    - Used to monitor (i.e., trending or spot checking) oxygen saturation levels of patients, in hospitals and doctors' offices, although they may sometimes be prescribed for home use, or available over the counter.
    - Recognized Consensus Standards 1-139 ISO 80601-2-61 Particular requirements for basic safety and essential performance of pulse oximeter equipment
    - *Pulse Oximeters - Premarket Notification Submissions [510(k)s]: Guidance for Industry and Food and Drug Administration Staff (2013)*

  - General Wellness: Regulated under product codes PGJ (Oximeter, Wellness) and OCH (Oximeter, Sporting, Aviation)

    - Most commonly intended for sporting/aviation and general wellness uses and are not intended for medical purposes.
    - They are often sold directly to consumers in stores or online and may utilize mobile medical apps intended for estimating oxygen saturation for non-medical purposes.
    - See FDA's guidance document *General Wellness: Policy for Low-Risk Devices* for additional information.

Pulse Oximeters Intended for Medical Purposes are Class II devices intended to measure blood oxygen saturation levels and are regulated under:

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

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

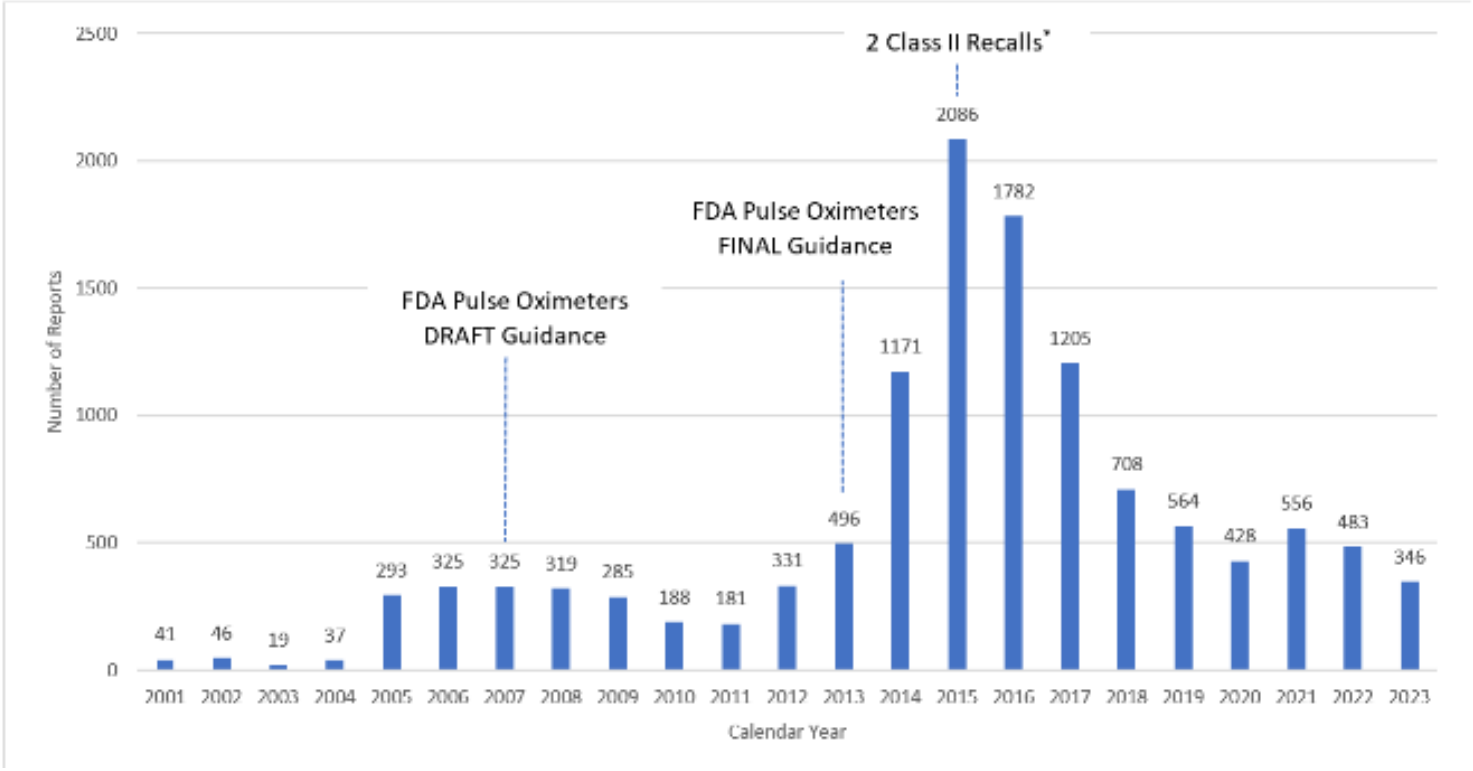
- Multiparameter devices with pulse oximeter functionality have been regulated under other classification regulations.

Pulse oximeters intended for medical purposes include devices using reflectance, transmittance, and fiber optic technologies. Pulse oximeters measure the amount of transmitted, reflected, and scattered light through various application sites (e.g., finger, ear, foot, hand, forehead, back, and nose).

- Some of the factors that can impact the accuracy of pulse oximeters include (but not limited to):
  - Skin pigmentation
  - Dyshemoglobinemias: disorders in which the hemoglobin molecule is functionally altered and prevented from carrying oxygen
  - Severe anemia: disorder in which the blood has reduced ability to carry oxygen. Anemia occurs when there are not enough healthy red blood cells to carry oxygen to the body organs
  - Low perfusion: reduced peripheral blood flow and subsequent reduction in the detectable signal at pulse oximeter sensor site
  - Dyes
  - Nail polish
  - Ambient light

All Adverse Event Reports

- MDR search was conducted on October 25, 2023 going back to January 1, 2000
- MDR data for reported product codes DQA (Oximeter), DPZ (Oximeter, Ear) and NLF (Reprocessed Oximeter)
- Then a text search for reports with the term “skin”, and a code search was used to identify any report submitted as a death report.
- Each report identified through the text search and each death report was then reviewed to determine if it was relevant to inaccurate SpO_2 readings and to assess potential sources for the inaccurate reading.



Adverse Event Reports received between January 1, 2000 to October 25, 2023
Reported Product Codes DQA, NLF, and DPZ, n = 12,248

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

*Out of 681 reviewed reports (381 with search term "skin" and 305 death reports, note 5 death reports also mention skin and were only counted once)
** indicates % estimated out of 101; the categories are not exclusive of each other; numbers can add up to more than 101.

Summary

- Pulse oximeters provide real-time noninvasive estimates on oxygen saturation through continuous and spot-checking monitoring
- Limits of utility exist in measurements which have clinical implications on use


FDA
U.S. FOOD & DRUG
ADMINISTRATION

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

Allison O'Neill, PhD, MA
Associate Director for Postmarket Surveillance
Office of Health Technology 1
Center for Devices and Radiological Health

PURPOSE OF REVIEW

2013 … 2020 2021 2022 2023 2024

Beginning of COVID-19 Pandemic

Sjoding et al (2020), Media attention

Pulse Oximeter Accuracy and Limitations: FDA Safety Communication

Nov 2022 Advisory Committee Meeting

Nov 2023 Discussion Paper and Request for Feedback

Feb 2024 Advisory Committee Meeting

Pulse Oximeters - Premarket Notification Submissions [510(k)s]

Guidance for Industry and Food and Drug Administration Staff

Document issued on: March 4, 2013

This document supersedes Non-invasive Pulse Oximeter General Guidance Document, September 7, 1992
The draft of this document was issued on July 19, 2007.

For questions regarding this document contact Neel Patel at 301-796-5580 or neel.patel@fda.hhs.gov.

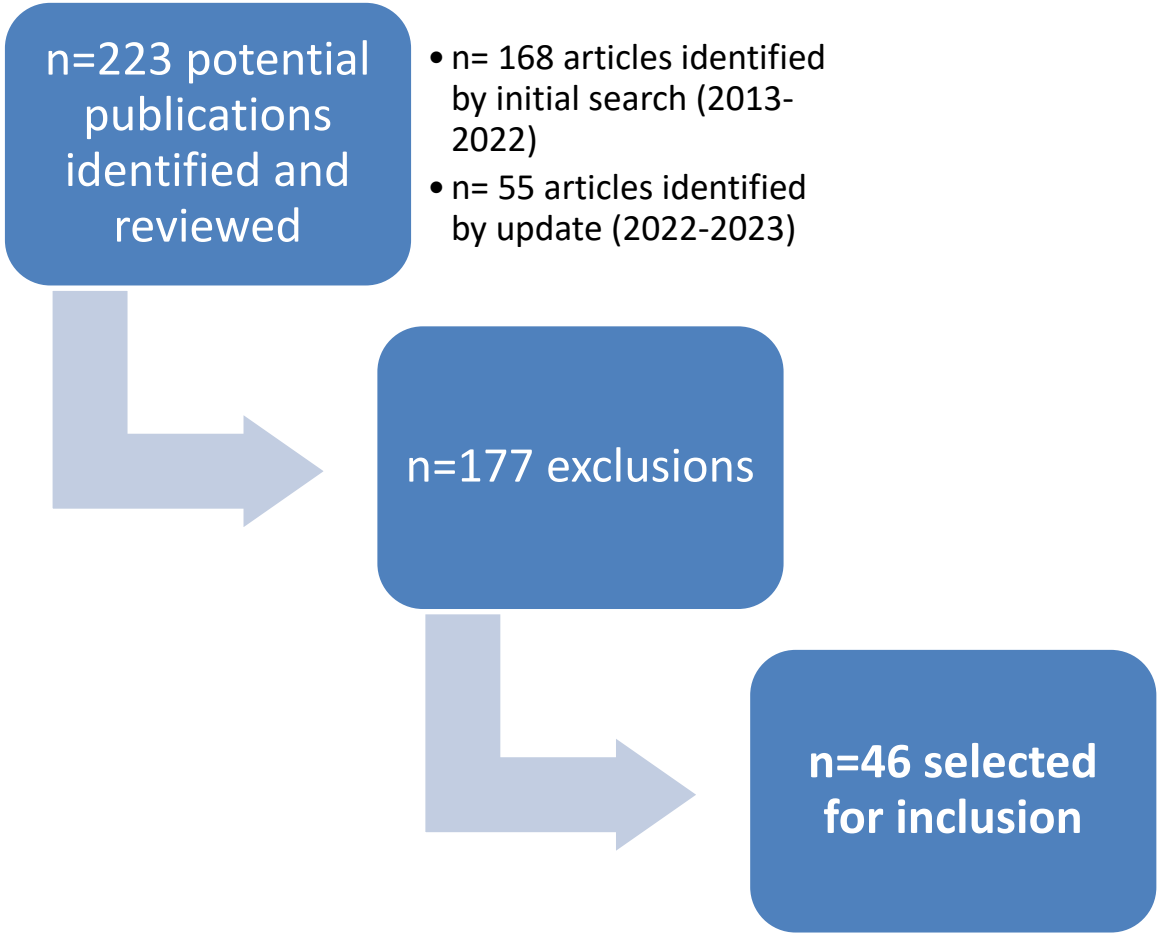
CDRH Center for Devices and Radiological Health

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of Device Evaluation
Division of Anesthesiology, General Hospital, Infection Control, and Dental Devices
Anesthesiology and Respiratory Devices Branch

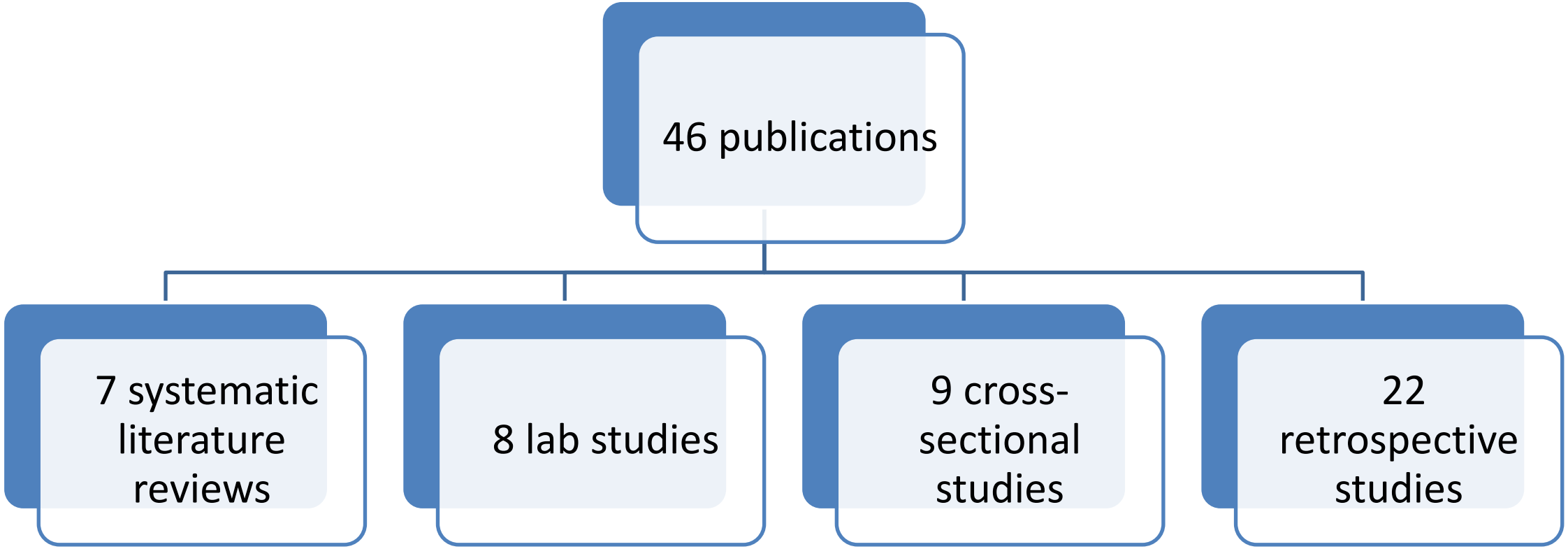
SYSTEMATIC LITERATURE REVIEW

Methodology

- PubMed search strategy: pulse oxim* AND (race OR racial OR pigment*)
- Inclusion criteria:
  - Publication date after FDA guidance document (2013)
  - Must include clinical data, with at least 9 participants
  - Must include measurement of either skin pigmentation or race/ethnicity
  - Literature reviews must be performed in systematic way, and report reproducible methodology



Results


46 publications
7 systematic literature reviews
8 lab studies
9 cross-sectional studies
22 retrospective studies

Systematic Reviews (n=7)

Reference	Description
Cabanas 2022	• 41 articles published bet. Jan 1976 - Feb 2022 related to skin pigmentation • Of 11 studies at low risk of bias, 8 studies found inaccuracies due to skin pigmentation, and 3 reported no loss of accuracy • Reported overestimations in populations with darker skin pigmentation, especially at low saturation or hypoxemia conditions
Shi 2022	• 32 articles published before Dec 2021 related to skin pigmentation • Meta-analysis: Reported overestimations in people with high level of skin pigmentation (pooled mean bias 1.11%; 95% confidence interval 0.29 to 1.93%) and people described as Black/African American (1.52%; 0.95 to 2.09%) (moderate- and low-certainty evidence)
Poorzargar 2022	• 22 articles published before Dec 2020 related to poor perfusion • Only 1 study controlled for skin pigmentation, and none strictly followed FDA recommendations
Jamali 2022	• 7 nonclinical and 15 clinical studies published between 1985-2022 • Reported higher frequency of technical difficulties, increased mean bias, and higher rates of occult hypoxemia for subjects with dark skin tones, that may then be associated with less aggressive clinical management and increased mortality • Recommendation: the data required for approval should be disaggregated for different skin tones and oxygen saturation ranges

Systematic Reviews (cont.)

Reference	Description
Al-Halawani 2023	• 22 out of 28 references identified found that SpO_2 was overestimated in those with darker skin relative to reference SaO_2 measurements obtained by blood gas analysis. • Over half of the studies reported an increase in bias for subjects from all racial subgroups as they became less saturated.
Aoki 2023	• 8 out of 10 references identified "statistically significant higher pulse oximeter readings in darker-skinned patients with hypoxia compared to their arterial blood gas measurements. • Occult hypoxia was more prevalent in Black and Hispanic patients than in White patients. Minority patients overall (Black, Asian, and American Indian) were more likely to have a $SaO_2 < 88\%$ that was not detected by pulse oximetry (occult hypoxemia) during hospitalization. • With greater levels of hypoxemia, the differences between SpO_2 and SaO_2 were greater. If SaO_2 was < 90%, then SpO2 was overestimated in all ethnicities but worse in minorities.
Parr 2023	• 34 studies and 1 systematic review • Pulse ox likely overestimate Black/African American patients' blood oxygenation levels; some degree of bias and considerable imprecision (moderate strength of evidence) • Occult hypoxemia likely more common among Black/African American patients (moderate strength of evidence) • Evidence insufficient to draw conclusions about clinical outcomes attributable to race/ethnicity biases

Real World Studies: Cross Sectional (n=9)

- Statistically significant association between skin pigmentation/race and occult hypoxemia or mean bias in 7 of 9 studies
- Patient populations: hospital inpatients (including ICU, surgical unit, mechanical ventilation, COVID-19)
- 8 studies with adult patients, 1 study with infants (with Cyanotic Congenital Heart Disease)
- Pigmentation:
  - Fitzpatrick: 5
  - Race/ethnicity: 2
  - ITA: 1
  - Munsell: 1
- See Table XIII-1 in Executive Summary for results

Real World Studies: Retrospective (n=22)

- Statistically significant association between race and either occult hypoxemia (OH) or mean bias in 19/22 studies
- Patient populations: hospital inpatients (including ED, ICU, NICU, COVID-19, ECMO)
- Four studies limited to pediatric age range; others enrolled adults only, or reported no age cutoff
- All relied on self reported race/ethnicity (available in electronic health records) rather than objective methods of pigmentation measurement
- See Table XIII-2 in Executive Summary for results

Lab Studies (n=8)

- 7 lab studies and 1 meta-analysis
- Healthy adult volunteers undergo controlled desaturation
- Pigmentation:
  - Fitzpatrick: 5
  - Race/ethnicity: 2
  - Dark, medium, light: 1
- Pulse ox performance affected by motion, perfusion index, skin pigmentation, and degree of hypoxemia
- See Table XIII-3 in Executive Summary for results

Limitations of Literature

- Variable definitions
  - Occult Hypoxemia
  - Skin pigmentation vs. race/ethnicity
- Real world and mostly retrospective data
  - Time between paired measurements
  - Rely on self reported race/ethnicity
  - Residual confounding from unmeasured variables
  - Testing conditions are less ideal than in a controlled lab study
- Heterogeneity of population
  - Sick patients vs. healthy volunteers
  - Prevalence of hypoxemia impacts the positive and negative predictive values (e.g., see Pennello 2022)
- Heterogeneity of technology
  - Brand of oximeter not always reported
  - Technology changes over time
- Publication bias

Summary

- Real-world evidence from literature suggests that pulse oximeter accuracy may vary by self-reported race, and skin pigmentation
  - 18 relevant articles published since the 2022 advisory committee meeting
- Need for prospective studies that:
  - utilize **standardized** measurement of skin pigmentation
  - capture **simultaneous** measurement of SaO_2 and SpO_2 paired data
  - systematically collect data on important **confounders**


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Proposed Approach To Improve Premarket Clinical Study

Kumudhini Hendrix, MD

Chief Medical Officer

OHT1/OPEQ/CDRH

Outline

- Overview of *Current* Recommendations* for Premarket Clinical Study
  - Submissions
  - Data
  - Limitations
- FDA's Proposed Approach** to address non-disparate performance of pulse oximeters*:* Prescription and Over the counter (OTC) for medical purposes
  - Sample size
  - Diversity in pigmentation, race and ethnicity
  - Performance

**2013 FDA Pulse Oximeters- Premarket Notification Submissions Guidance*
***Approach for Improving the Performance Evaluation of Pulse Oximeter Devices Taking into Consideration Skin Pigmentation, Race and Ethnicity*

Recommendation for Premarket Clinical Study: 2013 FDA guidance*

For pulse oximeters within the scope of the guidance:

- *In vivo* testing for SpO_2 accuracy
  - All new pulse oximeters intended for medical purposes
  - All prior cleared pulse oximeters with following modifications:
    - Significant electro-optical sensor modifications
    - SpO_2 algorithm modifications
- *Objective*
  - To verify the SpO_2 accuracy in comparison to the gold-standard measurements of blood SaO_2 by a CO-oximeter over the specified range (70-100%)
- For devices with *specific clinical indication(s),* additional clinical safety and effectiveness data is recommended

**2013 FDA Pulse Oximeters- Premarket Notification Submissions Guidance*

Participants – Current recommendations for Premarket Clinical Study

- Sufficient number of participants to attain statistical significance to demonstrate claimed SpO_2 accuracy
  - Minimum of pooled 200 data pairs (SpO_2, SaO_2) evenly distributed over the tested range (70-100%)
- Healthy adult volunteers who vary in physical characteristics to the greatest extent possible
  - 10 or more healthy participants with a range of skin pigmentation, age and gender
  - At least 2 darkly pigmented participants or 15% of participant pool, whichever is greater
  - COHb <3%, MetHb <2%, ctHb> 10g/dL

Testing Conditions

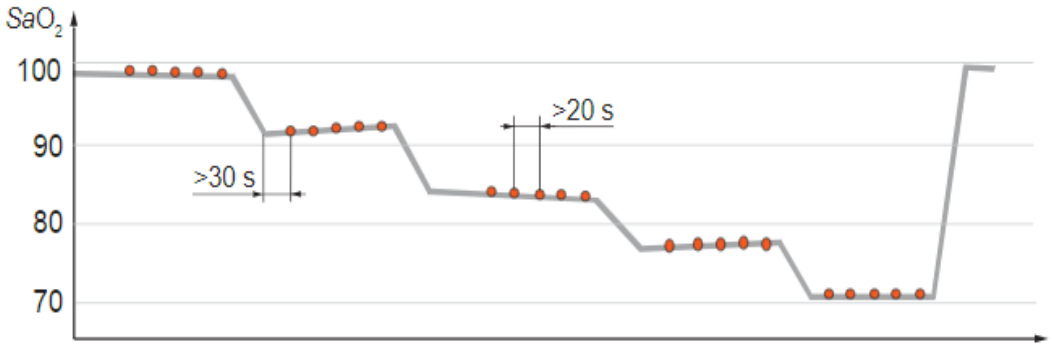
- Application of warming techniques can be utilized to improve circulation and pulse amplitude at a pulse oximeter probe site
- Pulse oximeter probes can be covered with opaque material to prevent optical interference
- CO_2 gas can be added to inspired gas mixture to maintain normal carbon dioxide levels

Methodology

- Fraction of inspired oxygen [FiO_2] delivered to test subjects is varied in a stepwise manner to achieve a series of targeted steady-state saturation periods
- When reference oximetry blood saturation stabilizes (≥30s) to an acceptable plateau, arterial blood samples from an indwelling arterial catheter is sampled for comparison of simultaneous data pairs (premarket device SpO_2, SaO_2)

Table EE.1 — Example of target plateaus and ranges

SaO_2 plateau range %	Target number of samples
100 to 97	5
97 to 92	5
92 to 85	5
84 to 78	5
77 to 70	5
Total	25



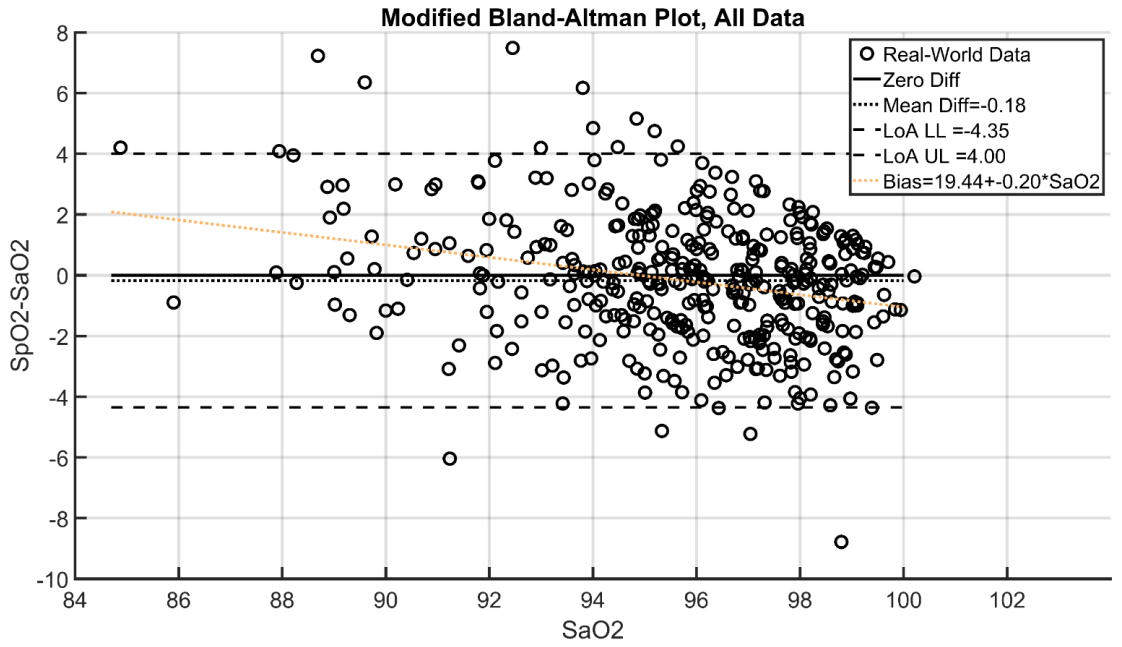
Points are SaO_2 values at the time of the blood draws.

Data

- Pertinent test apparatus used
- Inclusion/exclusion criteria
- Number of samples taken per subject
- Specific conditions of testing [laboratory conditions, subject motion, low amplitude]
- Type and frequency of motion for testing, if applicable
- Criteria and methods for determining stability of reference SaO_2 at the pulse oximeter sensor site
- Desaturation profile [target plateaus, and ranges]
- Formula used for determination of root mean square difference (A_{rms})

Analyses

- Individual and pooled modified Bland-Altman Plots (SpO_2-SaO_2 versus SaO_2)
- Population mean bias (μo)
- Between-subject variance $(\sigma\mu^2)$
- Within-subject variance (σ^2)
- Upper 95% and lower 95% limits of agreement



Ebmeier et al, 2018

Limitations

- Participants
  - Limited number
  - Subjective skin pigmentation assessment
- Accuracy
  - Pooled point estimate accuracy (A_{rms}) across the entire tested range of SpO_2
  - No threshold accuracy analyses
  - Not powered to determine significant difference between cohorts

Key Recommendations: Advisory Committee Meeting (11/01/2022)

- Subjective and objective pigmentation assessment
- Inclusion of self-reported race and ethnicity
- Adequate sample size
- Tighten performance accuracy
  - Overall A_{rms} within 1.5%-2% [68% CI]
  - Addition of a more clinically meaningful accuracy metric
- Improve labeling for OTC devices

FDA's Proposed Approach

- Increased overall sample size
  - N ≥ 24 subjects (≥ 480 measurements)
- Increased diversity representation
  - include the entire range of skin pigmentation
  - account for self-reported race and ethnicity
- Improved performance
  - increased accuracy
  - non-disparate performance assurance across entire range of pigmentation
    - diagnostic performance analyses at clinically relevant thresholds (e.g., SpO_2 at 90%)
    - applies to all pulse oximeters for medical purposes

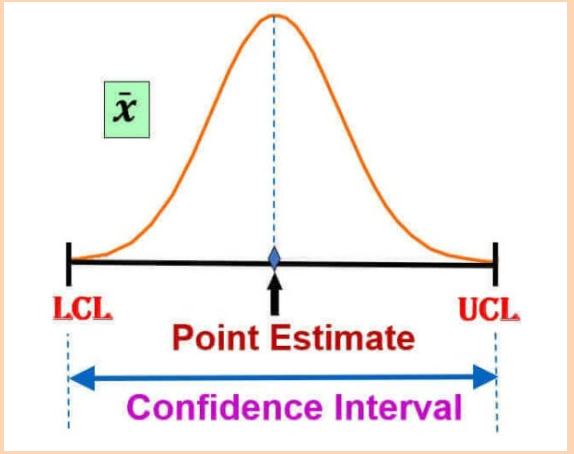
Clinically Meaningful Accuracy Metrics
Accuracy root mean square error (A_{rms}) over full measurement range

2013 guidance

- A_{rms}: performance evaluated with a point estimate
  - Uncertain how well the point estimate represents the population
  - Precision of estimation not characterized
  - Assuming SpO_2 has a normal distribution,
    - $SpO_2 \pm 1$ A_{rms} has probability 68% of including the SaO_2 value.
    - $SpO_2 \pm 2$ A_{rms} has probability 95% of including the SaO_2 value when SpO_2 bias is small, imprecision large.

Proposed Accuracy Metric

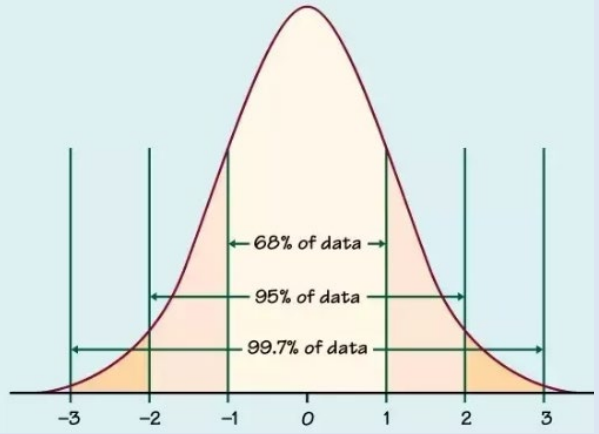
- A_{rms}: performance evaluated by an interval estimate
  - 95% CI on A_{rms} characterizes uncertainty in estimating true A_{rms}.
  - Increased certainty: If upper limit of 95% CI is < 3%, then statistically, A_{rms} *is significantly* < 3%.



Tighter Arms

2013 guidance

- Arms goal: 3% for transmittance, 3.5% for reflectan
- Assuming SpO_2 has a normal distribution,
  - 1 Arms ≈1 SD: SpO_2± Arms includes SaO_2 68% of time
  - to calculate 95% CI or 2 SD – double Arms
    - +/-6% for transmittance
    - +/-7% for reflectance



Proposed accuracy metric

Arms goal now defined by 95% CI

- 3% for transmittance and reflectance
  - With N=24, 80% power to meet 95% CI < 3% if true Arms ≤ 2.1%
- High accuracy needed to meet 95% CI < 3%

Typical Arms Specification by Sensor Type

Sensor Type	Arms [95% CI]
Transmittance, wrap and clip	< 3 %
Ear clip	< 3 %
Reflectance	< 3 %

Increased Sample Size: Non-disparate Performance Assessment

2013 guidance

- 10 or more healthy subjects
  - 200 or more data points
- Vary in age and gender
- Range of skin pigmentation
  - At least 2 darkly pigmented subjects or 15% of subject pool (whichever is larger)

Proposed sample size

- ≥ 24 healthy subjects
  - 480 or more data points
  - Approximately equal distribution of data pairs across pigmentation
- ≥ 40% of each gender
- Subjective and Objective Pigmentation assessment

Pigmentation Assessment

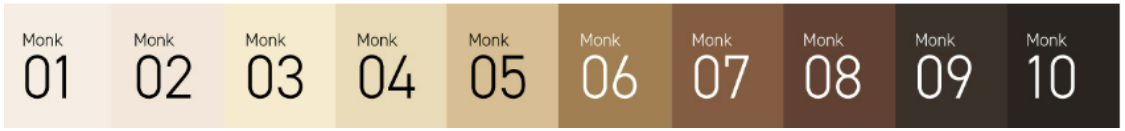
2013 guidance

- Range of skin pigmentation
  - ≥ 2 darkly pigmented individuals or ≥ 15% of participant pool; whichever is greater

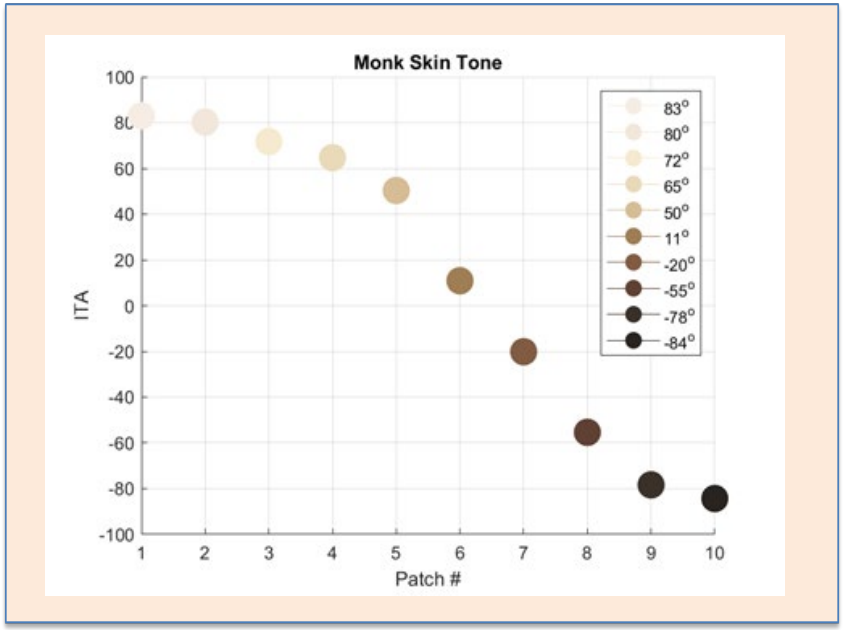
Proposed approach

- Evaluate Monk Skin Tone (MST) at forehead (10 MST Values)
- Measure objective Individual Typology Angle (ITA) at emitter sensor site for non-disparate performance analyses

Monk Skin Tone (MST)



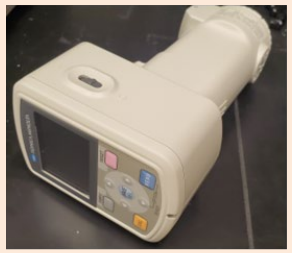
- Subjective skin color scale
  - 10 values
- Validated to capture race and ethnicity diversity in pigmentation within the US
  - More inclusive than Fitzpatrick Scale
- Standardized to color scales (CIELAB)
  - Objectivity to assignment
- High inter-class correlation (ICC:0.9-0.96)
  - Validated with US and OUS raters

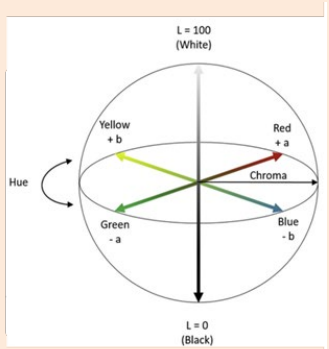


Heldreth et al., ACM J of Responsible Computing, 2023
Schumann et al., [Preprint; accepted for publication by conference on Neural information processing system], 2023

Individual Typology Angle (ITA)

- Individual Typology Angle
  - Standardized CIELAB color space approach (L*a*b*)
  - Objective (reproducible, accurate)
  - Continuous
  - Correlates with epidermal melanin content (gold standard)

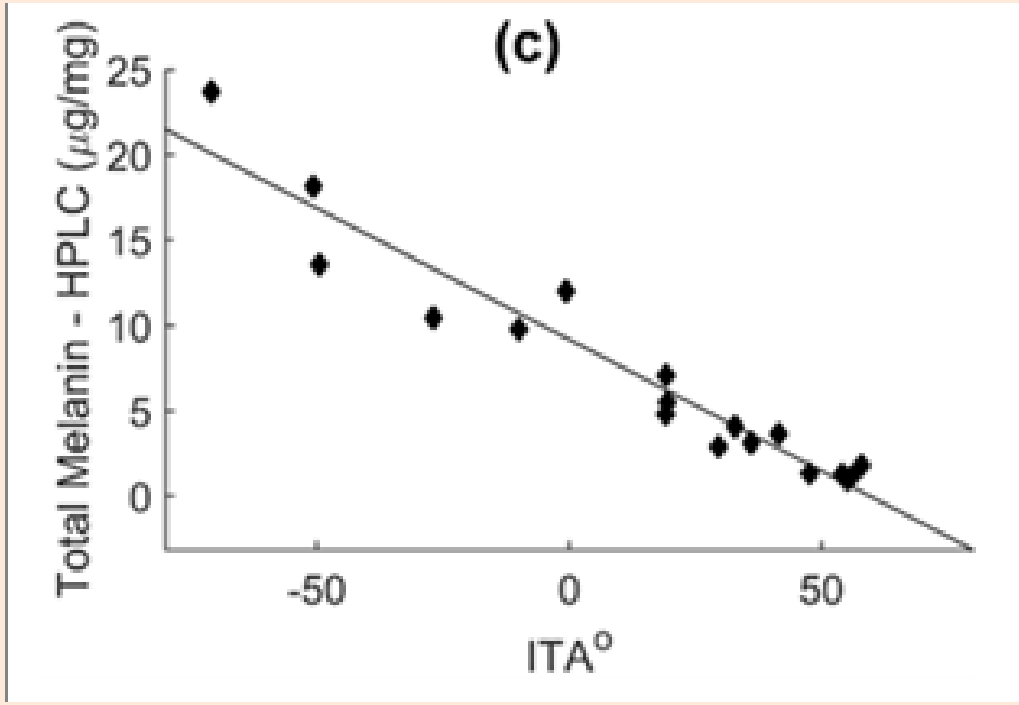




L* a* b*

(intensity, red/green, blue/yellow)

$$\mathrm{ITA}^{\circ} = \tan^{-1}\left(\frac{L^{*}-50}{b^{*}}\right) \times \frac{180}{\pi}$$



Joshua Pfefer PhD, Sandhya Vasudevan PhD, OSEL

Ito, S., et al., International journal of molecular sciences, 2020

Proposed Pigmentation Assessment

Recruitment

- **Evaluate Monk Skin Tone (MST) at forehead (10 MST Values)**
  - *Maximize* participants from diverse racial and ethnic groups
  - Divide values into 3 MST cohorts (MST 1-4, MST 5-7, MST 8-10)
    - ≥25% in each cohort *e.g. for N=**24**: **6** (MST 1-4), **9** (MST 5-7), **9** (MST 8-10)*
    - At least 1 subject or ≥15% participants in cohort for each value (whichever is greater) *e.g. **≥1** for MST 1*

Sensor Site Measurement

- **Measure objective Individual Typology Angle (ITA) at emitter sensor site for non-disparate performance analyses**
  - For fingertip - pigmented skin at the midline of the dorsal distal phalanx
    - Allows for widest range of emitter-site pigmentation

Proposed Acceptance Criteria: Non-disparate Performance

- Across ITA *and* MST, the estimate of the maximum difference in SpO_2 bias* is

$< 1.5\%$ for $SaO_2 > 85\%$

$< 3.5\%$ for $70\% \leq SaO_2 \leq 85\%$

**Where SpO_2 bias is the mean of the difference D (where $D = SpO_2 - SaO_2$)*

Race and Ethnicity

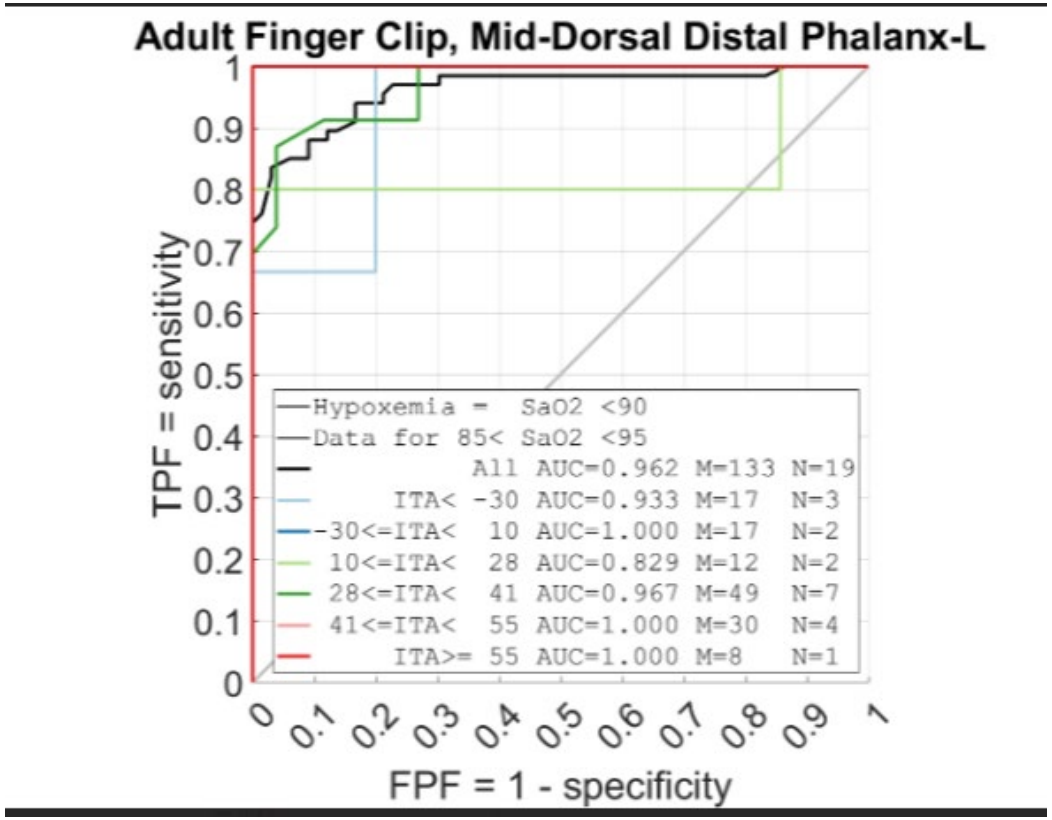
- Disparity of pulse oximeter performance in individuals from different race or ethnicity with same level of sensor site pigmentation is unknown
- Maximize racial and ethnic diversity of participants and collect race and ethnicity data*
  - MST is not a proxy for race and ethnicity diversity
  - MST validated to capture race and ethnicity diversity in pigmentation
    - Even distribution across entire MST ensures some racial and ethnic diversity
      - Meeting acceptance criteria across MST ensures non-disparate performance across those respective race and ethnicity groups

**Collection of Race and Ethnicity Data in Clinical Trials, 2016*

Diagnostic Performance Analyses: ROC Plot

- Additional analyses at clinically relevant thresholds (e.g., 90%)
  - Receiver operating characteristic (**ROC**) curve and area under the ROC curve (**AUC**)
    - Overall
    - Across MST
    - Across ITA

Target condition: SaO_2 < 90%



ROC Plots, Controlled Desaturation Study Data

OTC Pulse Oximeters for Medical Purposes

- Same premarket clinical study design as prescription use
- Same acceptance criteria for non-disparate performance as prescription use
- Labeling
  - Information about premarket clinical study design
  - Description of the device performance

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

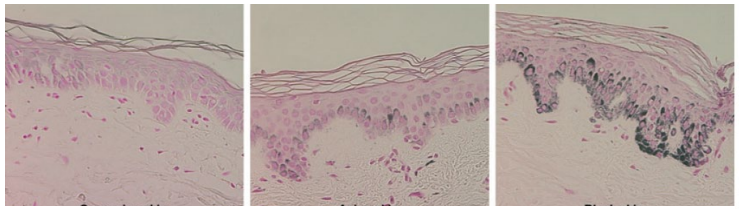
- *A more inclusive and representative trial design;*
- *Defining non-disparate performance; and*
- *Considerations for studies of Over the Counter (OTC) devices used for medical purposes.*

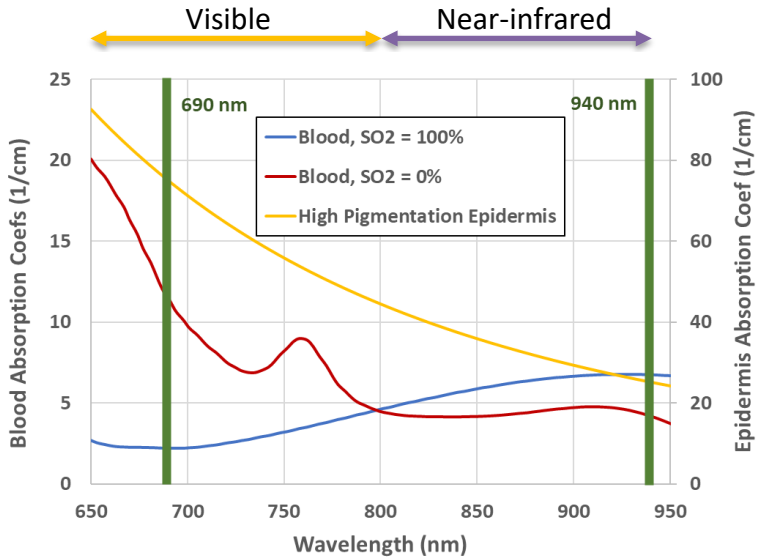
Assessment of Skin Pigmentation

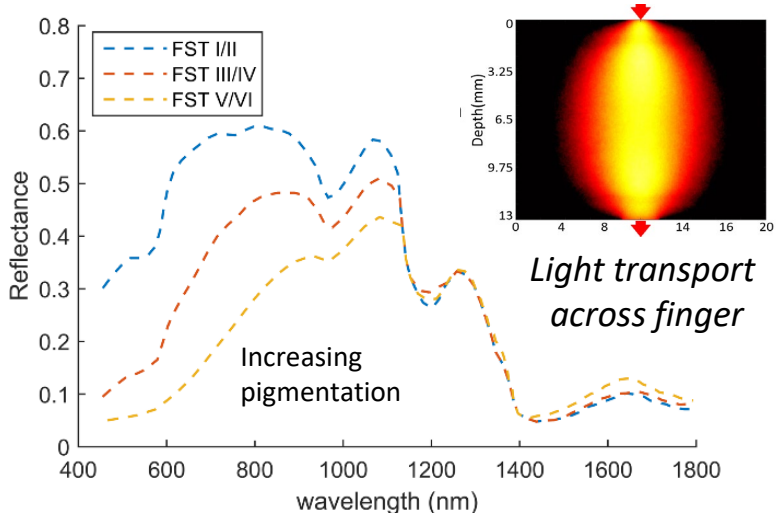
Joshua Pfefer, PhD
Division of Biomedical Physics
Office of Science and Engineering Laboratories
Center for Devices and Radiological Health

February 2, 2024

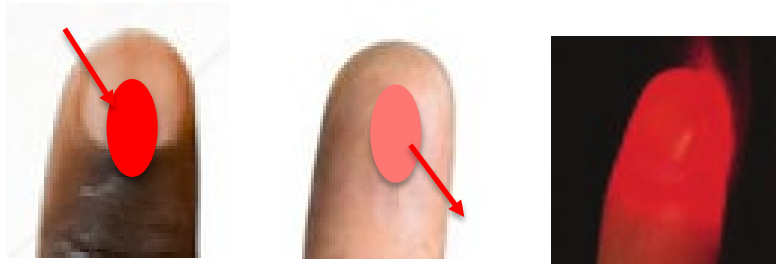
Pulse Oximetry and Light-Tissue Interactions







- **Epidermal melanin**
  - Significant red/NIR absorber
  - Can reduce detected light by up to ~70%
  - May influence light propagation pathways



Dorsal Palmar/Ventral

Kyriacou et al., Academic Press 2021
Chatterjee et al., Sensors 2019

Mendenhall et. al. Appl Opt 2015
New York Times, Dec 22, 2020

Mechanisms and Evaluation

- Mechanism of racial bias in pulse oximetry may determine optimal evaluation approaches

Possible Mechanism	Variable	Sources
Optical absorption by melanin → spectral/intensity changes	Epidermal melanin content at sensor site (ITA)	*Arefin et al., Proc. SPIE 2022, Norton Anaesth 2022*
Perceived colorism	Skin tone at forehead (MST)	*Monk et al. JHSB 2021*
Genetically-correlated physiology (e.g., vascular response)	Epidermal melanin content at constitutive site (minimal sun impact)	*Patterson et al., Am J Hematol 2022, Drew et al., Am J Phys-RICP 2020*

ITA = Individual typology angle

MST = Monk Skin Tone scale

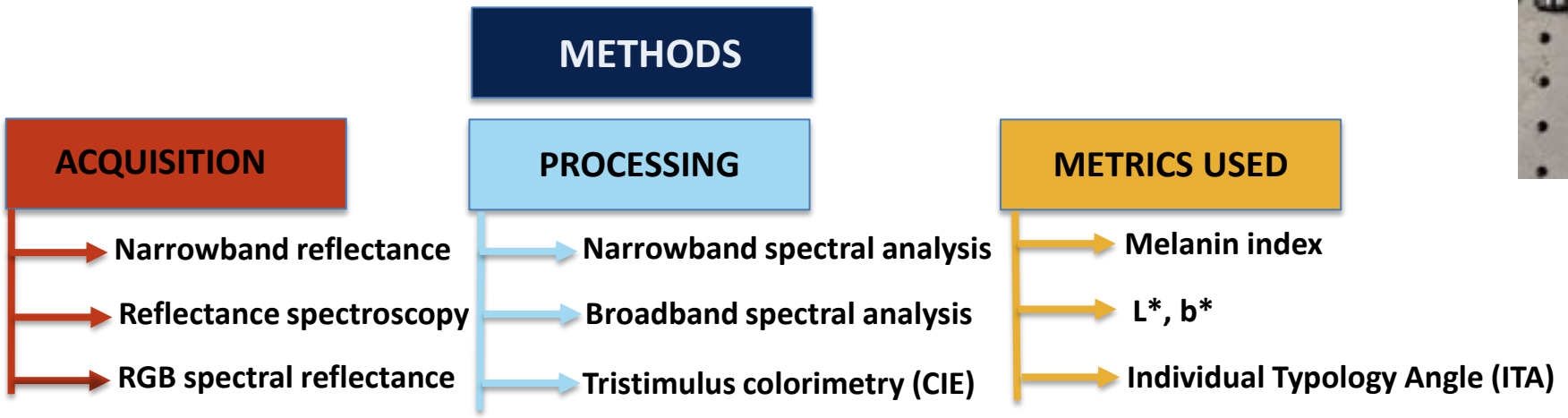
  - Combination of factors: pigmentation & low PI *(Gudelunas et al., Anes Analg 2023)*
- Other factors may degrade performance
  - Nail polish, bilirubin, ambient light/light piping, bandage color

Skin Pigmentation Assessment Methods

Assessment Method	Degree of Objectivity/Quantitation
Racial/Ethnic Self-Identification (e.g., Black, White, Hispanic, Asian)	Subjective with limited/no skin pigmentation information
Skin Color Descriptors (e.g., light, medium, medium dark, dark)	Subjective due to lack of a standardized scale. Large variance within skin color groupings
Sunburn susceptibility / color scale (Fitzpatrick skin phototypes, I-VI)	Subjective (questionnaire-based), quantitative (categorical); non-standardized color scale
Color scales (von Luschan, Massey, Monk)	Subjective but with standardized or semi-standardized color categories
Optical Methods (Spectroscopy, Colorimetry)	Quantitative data from objective measures; but some metrics not standardized
Biopsy *with histological/optical processing or high-performance liquid chromatography*	*Quantitative melanin content, but can involve reader-dependent steps (e.g., layer identification)*

Objective Optical Approaches

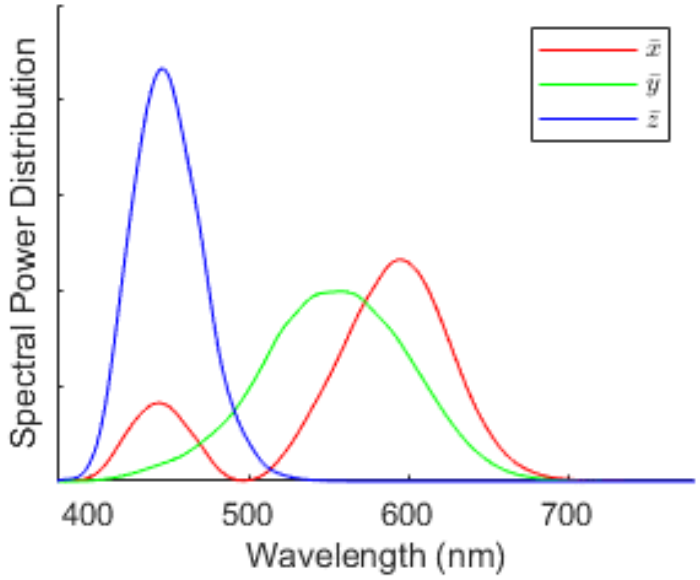
- Based on absorption of visible to near-infrared light in skin
- Applications in cosmetics, anthropology, and medicine
- Commercial, handheld systems using spectroscopy or colorimetry
- Local anatomical site pigmentation (e.g., palmar finger)



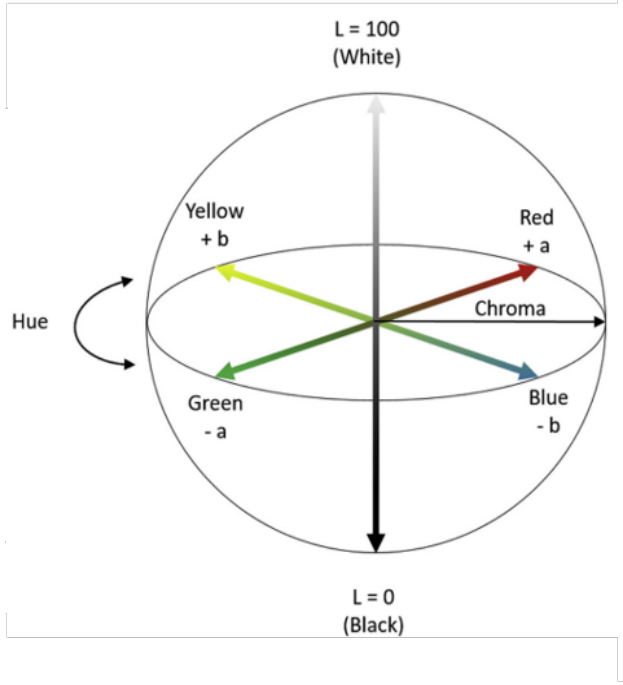
Colorimeters:
CL400 DSMIII SCC CM700d

Colorimetry

CIE standard observer color matching functions

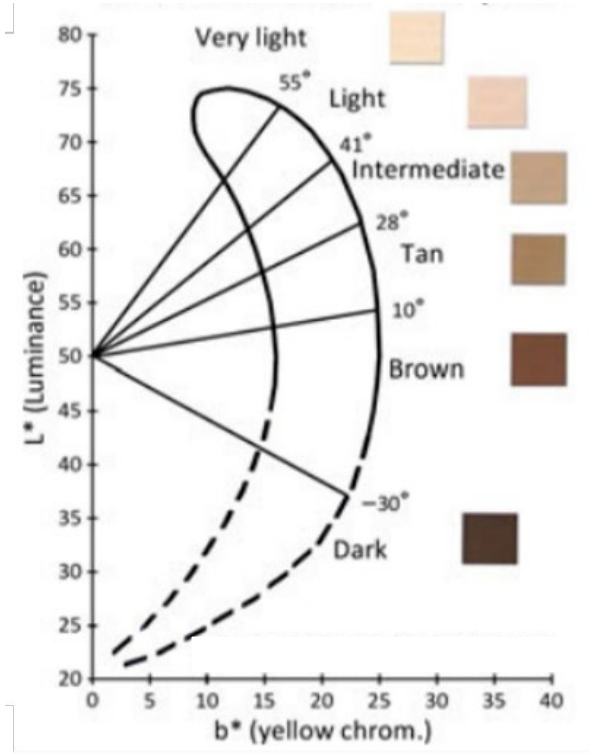


CIE L*a*b* color space



Individual Typology Angle

$$\mathrm{ITA}^{\circ} = \tan^{-1}\left(\frac{\mathrm{L}*-50}{\mathrm{b}*}\right) \times \frac{180}{\pi}$$



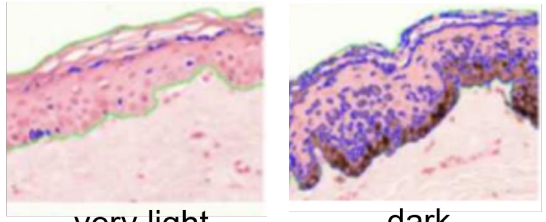
- Colorimeters typically provide high intra-and inter-observer repeatability, as well as inter-instrument repeatability (ICC > 0.9)

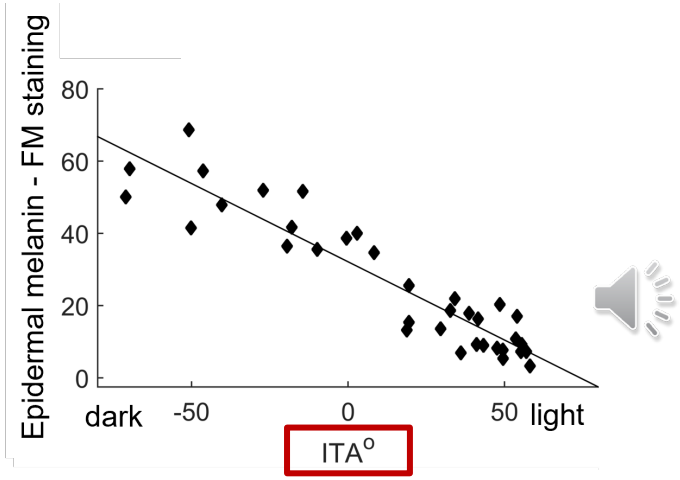
Ly, Bao Chau K., et al. *J Invest Dermatol* 2020
Del Bino, S., et. al. *British J Dermatol* 2013

ITA and Melanin Content

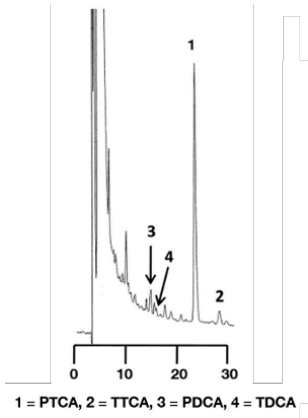
Histologic Area Estimate

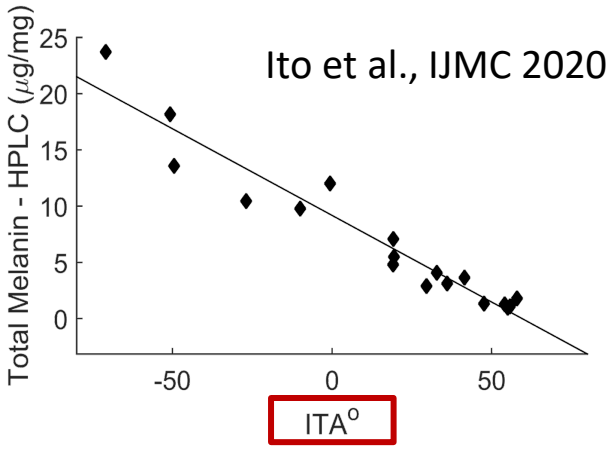
FM staining epidermis





High Performance Liquid Chromatography (HPLC)





- Melanin-specific stains, e.g., Fontana Masson (FM)
- Stain area in total epidermis, basal layer [Del Bino]
- Spatially integrated transmittance / area [Coelho]

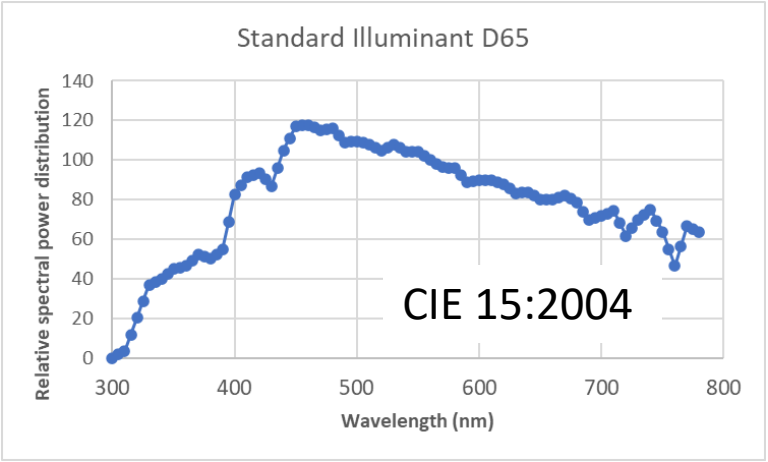
- Quantifies eumelanin (EM) and pheomelanin (PM) degradation byproducts
- Homogenized tissue samples

- Colorimeters measuring ITA typically exhibit strong correlation with melanin content (R = 0.8 - 1.0)

Colorimeter Recommendations

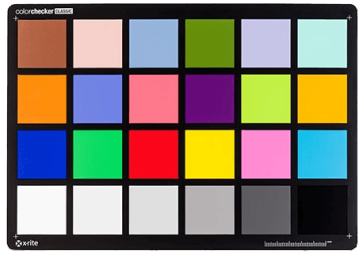
- Specifications: D65 illuminant, aperture 3-8 mm, 10° observer angle, specular reflectance excluded
- Validation should involve comparisons in human subjects covering a wide range of pigmentation levels
- Lab testing (typical approach) may be suitable
  - Standardized color charts used to assess/calibrate colorimeter performance
  - Compare measured values (L*a*b* or ITA) to measurement with reference system

$$\Delta E^*_{ab} = \left[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2 \right]^{1/2}$$



PANTONE SkinTone Guide

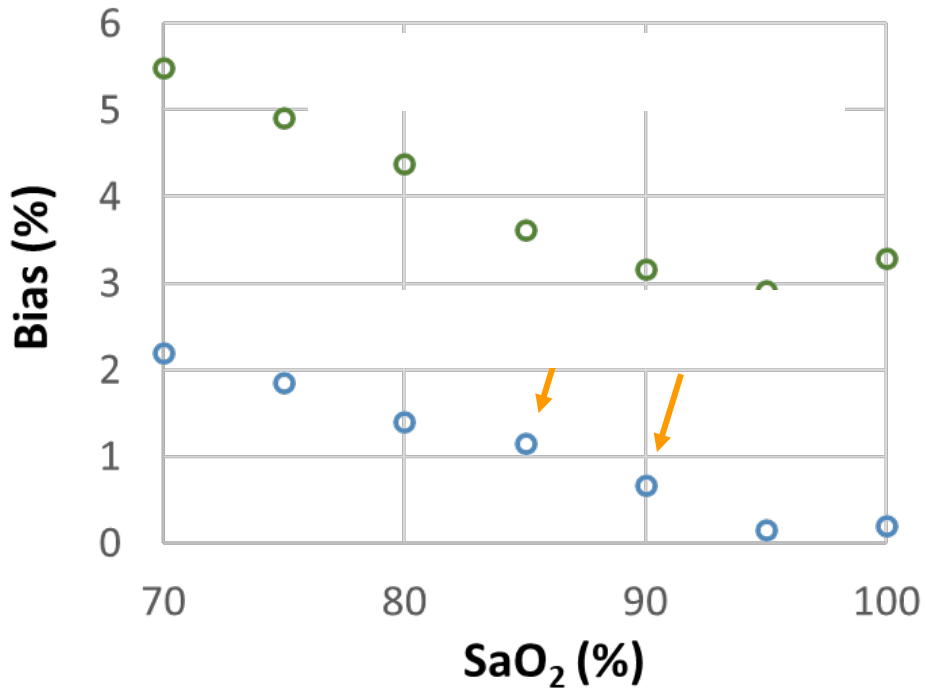
X-RITE colorchecker classic



ISO/CIE 11664-4: , Z., *Colorimetry—Part 4:* 2019

New Evidence on Melanin Content Mechanism

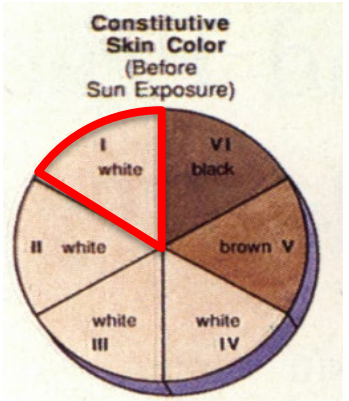
- FDA simulations indicate that increased melanin content may account for about 1-2% bias in SaO_2 = 85-90% range
- *Fawzy et al., MedRxiv 2023* clinical study found disparity using *ITA values alone*
- Melanin content may have important implications for the design of future clinical studies

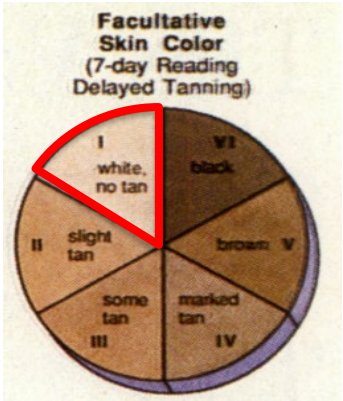


Bias estimates based on light-tissue interaction modeling (M_f = melanin volume fraction), from FDA-UMass collaboration

Subjective Skin Tone Scales

- Fitzpatrick Skin Phototype (FSP) scale
  - Originally developed to assess susceptibility to skin burn/tan
  - Not viable for high pigmentation subjects
- Semi-standardized
  - Modified or visual FSP (I-VI)
  - Massey-Martin
  - von Luschan
- Standardized, commercial color charts
  - Munsell
  - Pantone
- Standardized/traceable to color system (e.g., CIELab)
  - Monk Skin Tone (MST) scale





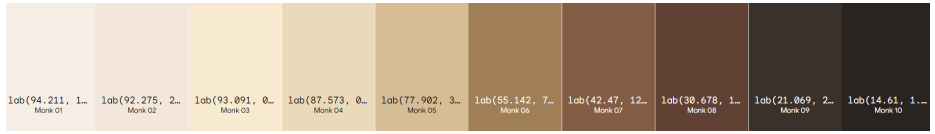
von Luschan Scale

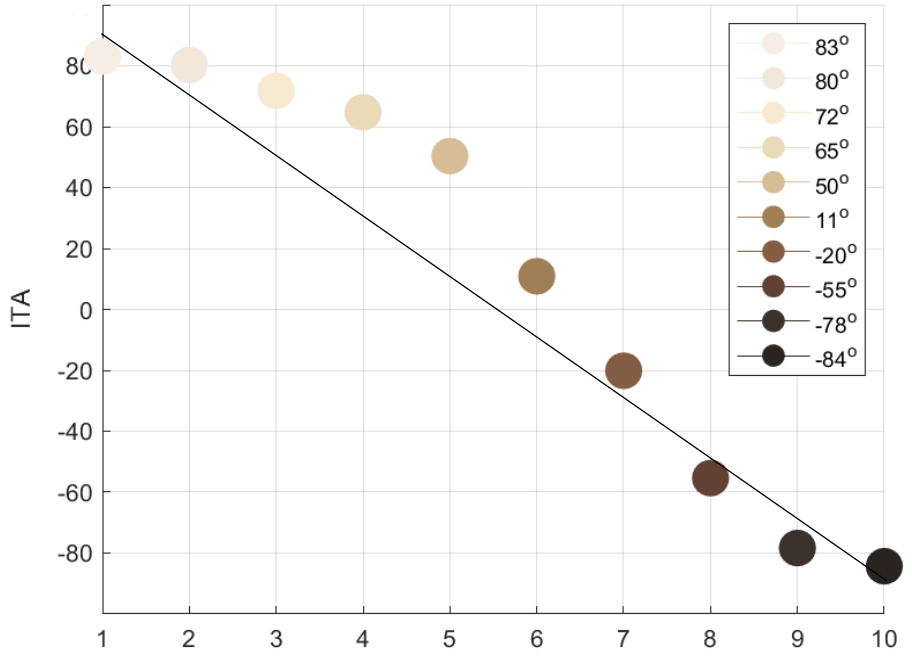
Fitzpatrick, Arch Dermatol 1988
Charlton et al., PLOS One 2020
Ash et al., Photon Lasers Med 2015
Varughese, P. M., et. al. *IJPD* 2018

Monk Skin Tone (MST) Scale

- Facial skin tone as an indicator linked to health [Monk et al., JHSB 2021;Heldreth et al., ACM JRC *in press*]
  - Perceived colorism, or human perception of skin color by oneself and/or others
- Standardized in color systems: CIELAB, RGB
- Better "perceived representativeness" than Fitzpatrick scale
- High level of inter-rater reliability
- Potential disadvantages
  - No commercially available color charts
  - Very low and high levels show minimal difference
  - Subjective methods prone to observer/rater errors, lighting, inter-observer bias

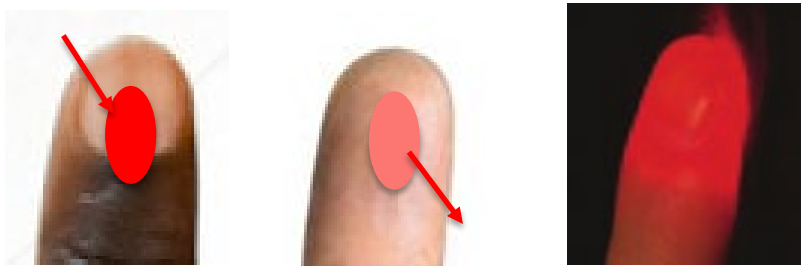
10-level scale



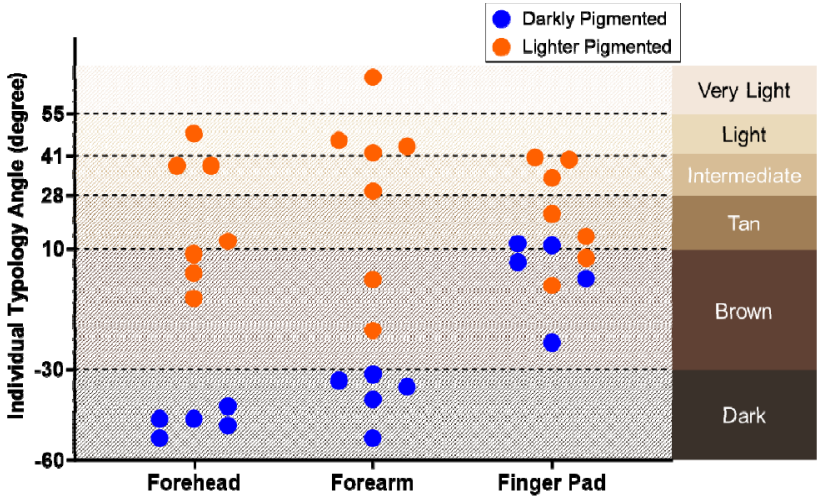


Pigmentation Measurement Sites

- Forehead is proposed site for MST evaluation
  - Provides wide range of pigmentation levels for relatively strong differentiation
- Finger is proposed site for ITA measurement
  - Fingernail, palmar finger contain low levels of melanin which can be difficult to differentiate
  - Fingertip pulse ox sensors detect light that interacts with adjacent dorsal skin pigmentation



Dorsal Palmar/Ventral



Panel Question

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

Statistical Assessment of Pulse Oximeter Performance

GENE PENNELLO, PhD
Division of Imaging Diagnostics & Software Reliability
U.S. Food and Drug Administration
Center for Devices and Radiological Health
Office of Science and Engineering Laboratories

Outline

- Individual Agreement
  - Bias
  - Precision
- Descriptive Plots
  - Bland-Altman
  - Quantile-Quantile
- Co-Primary Analyses
  - Accuracy Root Mean Square error (ARMS)
  - Non-Disparate Performance Assurance
- Other Potential Objectives
  - Diagnostic Accuracy for $SaO_2 < 90\%$
  - Inverse Prediction Interval of Plausible SaO_2 values Given SpO_2 value

Individual Agreement

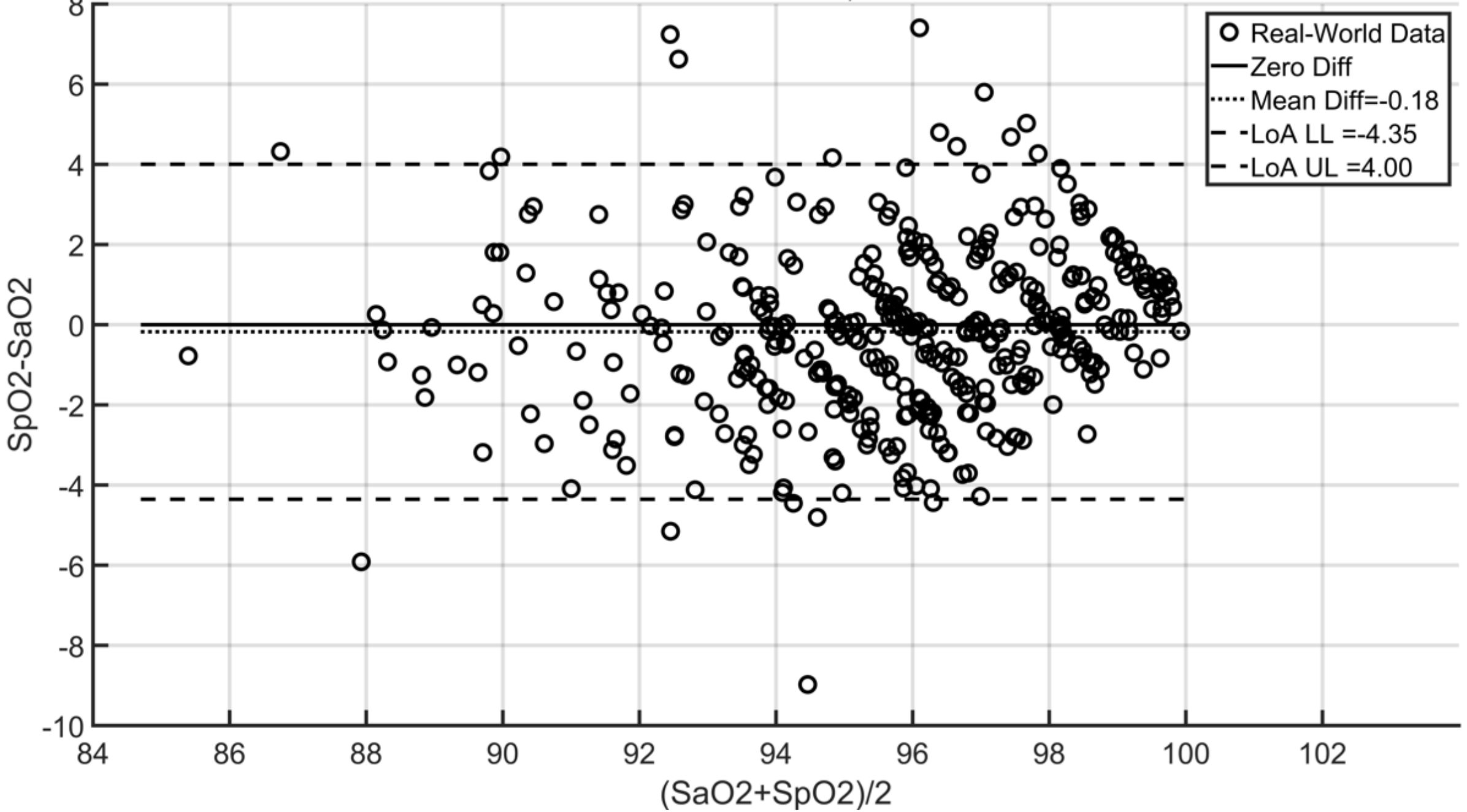
Individual Agreement

$\boldsymbol{SaO_2}$ = arterial oxygen saturation

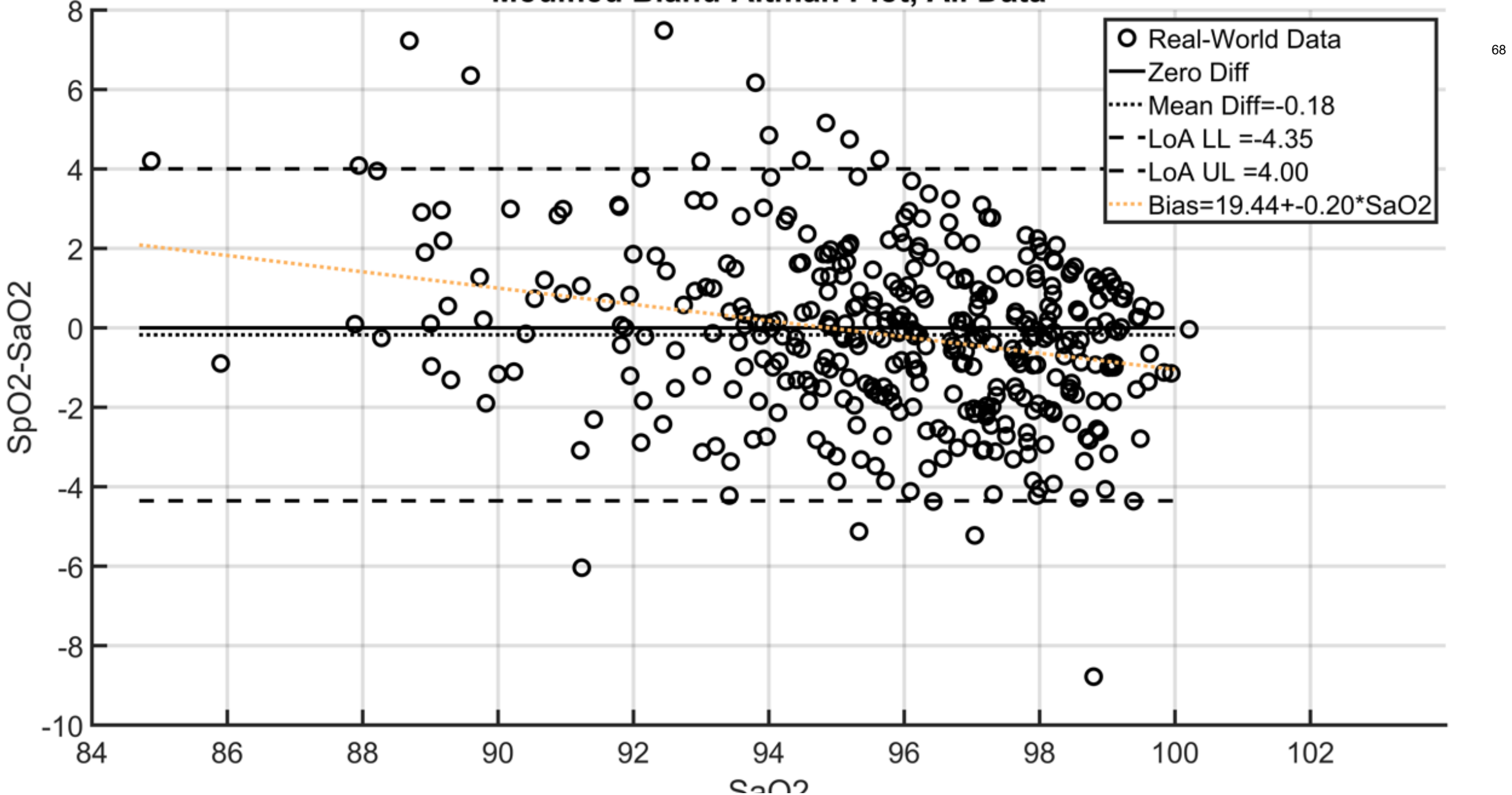
$\boldsymbol{SpO_2}$ = peripheral oxygen saturation

- **Error** $\boldsymbol{D = SpO_2 - SaO_2}$
- **Bias** $\boldsymbol{Bias}$ = average value of D in infinitely many paired repeated measures of (SaO_2, SpO_2)
- **Imprecision** $\boldsymbol{SD}$ = standard deviation of $\boldsymbol{D}$

Descriptive Plots

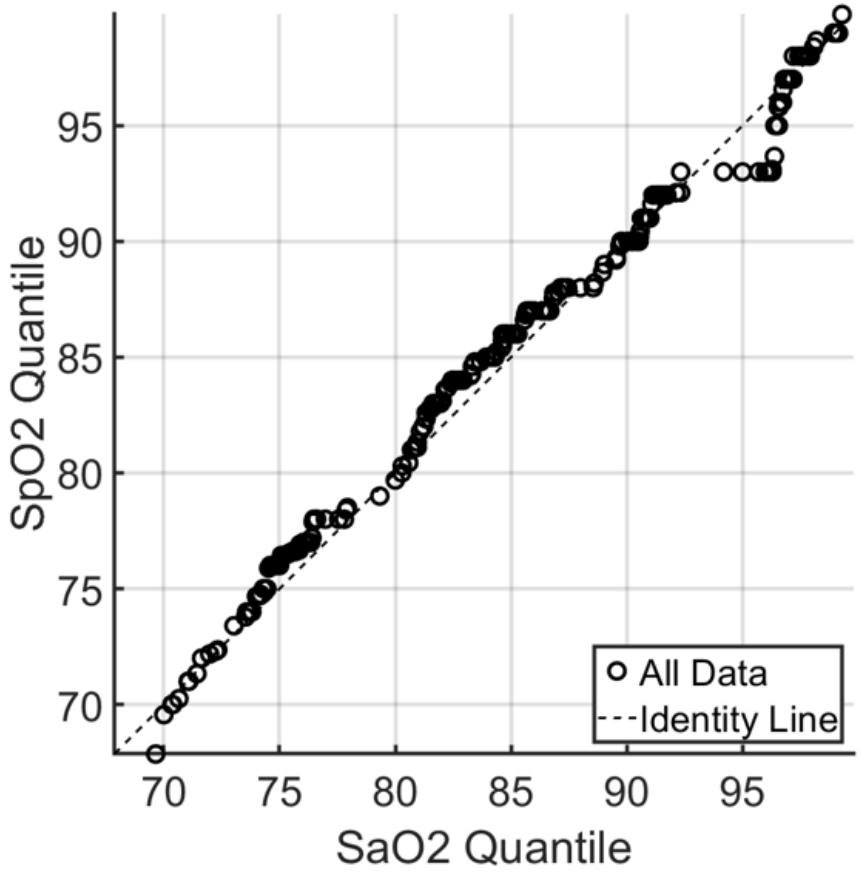


Bland-Altman Plot with zero difference line (solid), mean difference line (dotted), and 95% limits of agreement (LoA) (dashed).

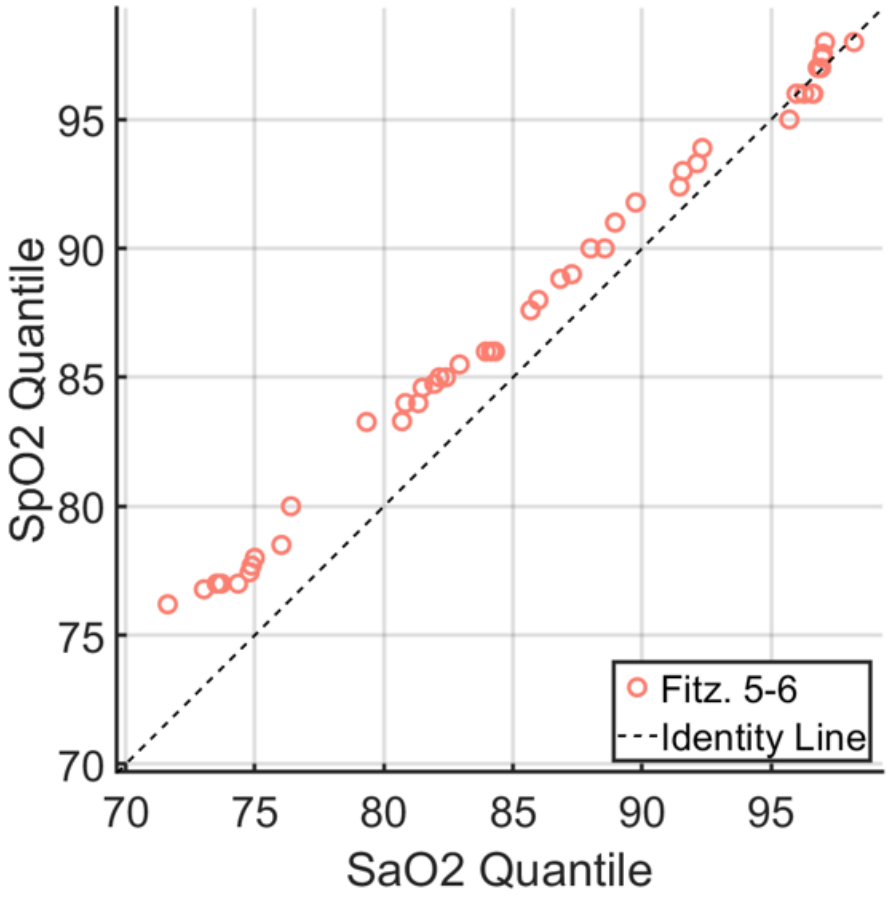


Modified Bland-Altman Plot with zero difference line (solid), mean difference line (dotted), and 95% limits of agreement (LoA) (dashed).

Quantile-Quantile (QQ) Plots



(a) very little SpO2 bias



(b) mostly positive SpO2 bias

QQ plots of SaO2 vs. SpO2 quantiles indicating (a) little SpO2 bias, and (b) positive SpO2 bias for most SaO2 values.

Accuracy Root Mean Square Error (Arms)

Accuracy Root Mean Square Error (A_{RMS})

- A_{RMS} is the root **mean** square of SpO_2 –SaO_2 .
  - A_{RMS} is an overall summary of the absolute error of SpO_2 for SaO_2.

- **Result:** $$A_{RMS}^2 = Bias^2 + SD^2$$

- **Performance Goal:**
  - Upper limit of 95% confidence interval on $A_{RMS} < 3\%$.

Sample Size for A_{RMS}

- **Null hypothesis** $H: \boldsymbol{A_{RMS} \geq 3\%}$
- **Alternative hypothesis** $A: \boldsymbol{A_{RMS} < 3\%}$
- For sample sizes of
  - 24 subjects, and
  - 20 repeated $(\boldsymbol{SaO_2, SpO_2})$ pairs per subject spanning $\boldsymbol{SaO_2 \in (70\%, 100\%)}$,
- if the true overall $\boldsymbol{A_{RMS} = 2.5\%}$, then
  - at the 2.5% 1-sided significance level, the power is $\boldsymbol{80\%}$ to reject H in favor of A,
  - i.e., to show $\boldsymbol{A_{RMS} < 3\%}$ with statistical significance,
  - i.e., for the 2-sided 95% confidence interval upper limit to be $< 3\%$.
- **Pre-caution**: Sample size providing 80% power may have to be $\boldsymbol{> 24}$ subjects if variation *between subjects* in $\boldsymbol{A_{RMS}}$ is substantial.

Non-Disparate Performance Assurance

Acceptable Maximum Difference in SpO_2 Bias between ITA or MST Levels

Proposed Acceptance Criteria: Non-Disparate Performance

- Across ITA *and* MST, the estimate of the maximum difference in SpO_2 bias* is

$< 1.5\%$ for $SaO_2 > 85\%$

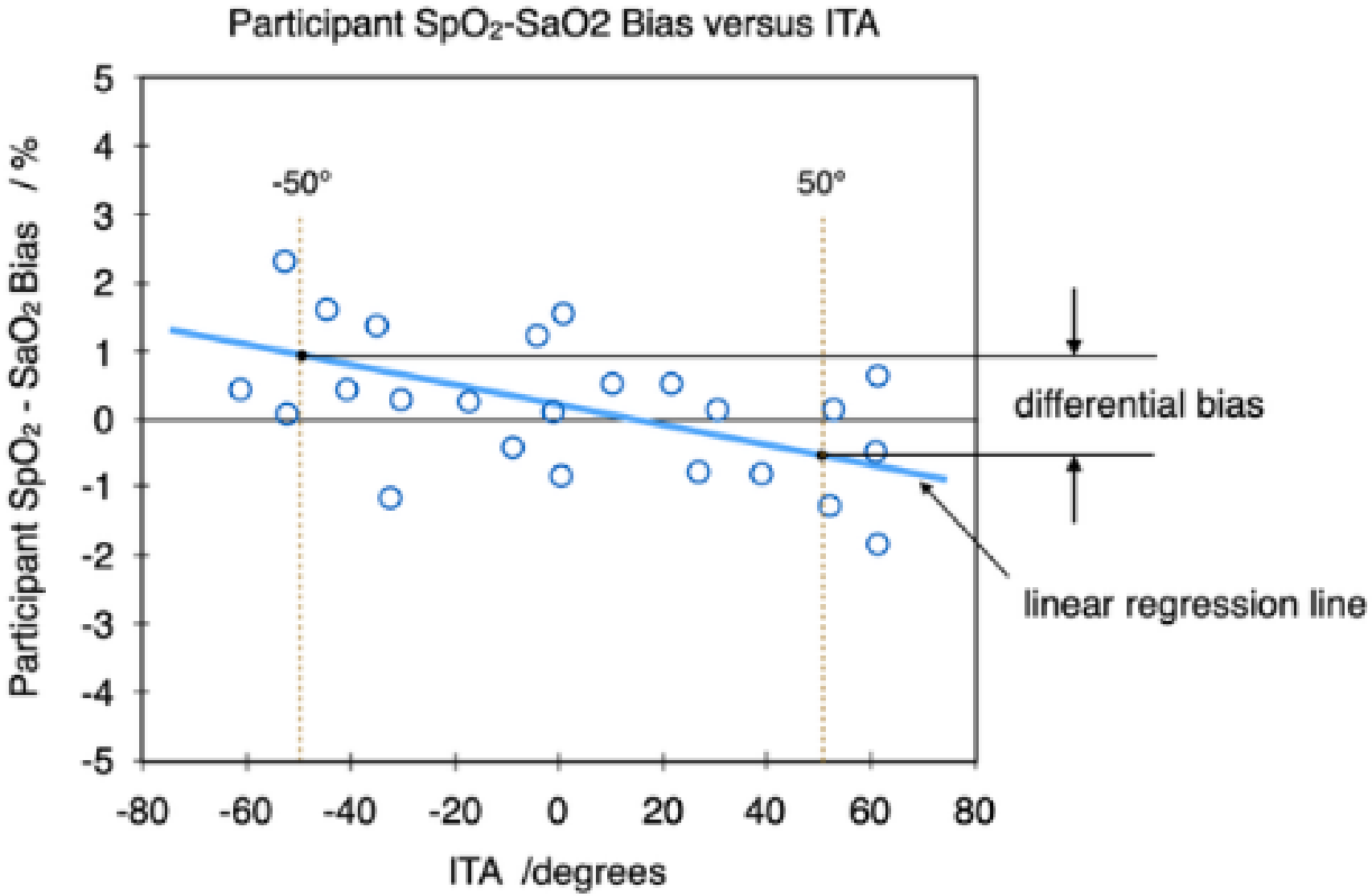
$< 3.5\%$ for $70\% \leq SaO_2 \leq 85\%$

**Where SpO_2 bias is the mean of the difference D (where $D = SpO_2 - SaO_2$)*

Sample Size for Non-Disparate Performance Analysis

- For sample sizes of
  - 24 subjects, and
  - 20 repeated pairs $(\boldsymbol{SaO_2}, \boldsymbol{SpO_2})$ per subject spanning $\boldsymbol{SaO_2} \in (\boldsymbol{70\%}, \boldsymbol{100\%})$,
- Simulations suggest that for $\boldsymbol{70\% \leq SaO_2 < 85\%}$,
  - if true maximum difference in $\boldsymbol{SpO_2}$ bias $= \boldsymbol{2.0\%}$, then power is **80%** for **point estimate to be** $\boldsymbol{< 3.5\%}$.
- Simulations suggest that for $\boldsymbol{85\% \leq SaO_2 < 100\%}$,
  - if true maximum difference in $\boldsymbol{SpO_2}$ bias $= \boldsymbol{0.5\%}$, then power is **80%** for **point estimate to be** $\boldsymbol{< 1.5\%}$.

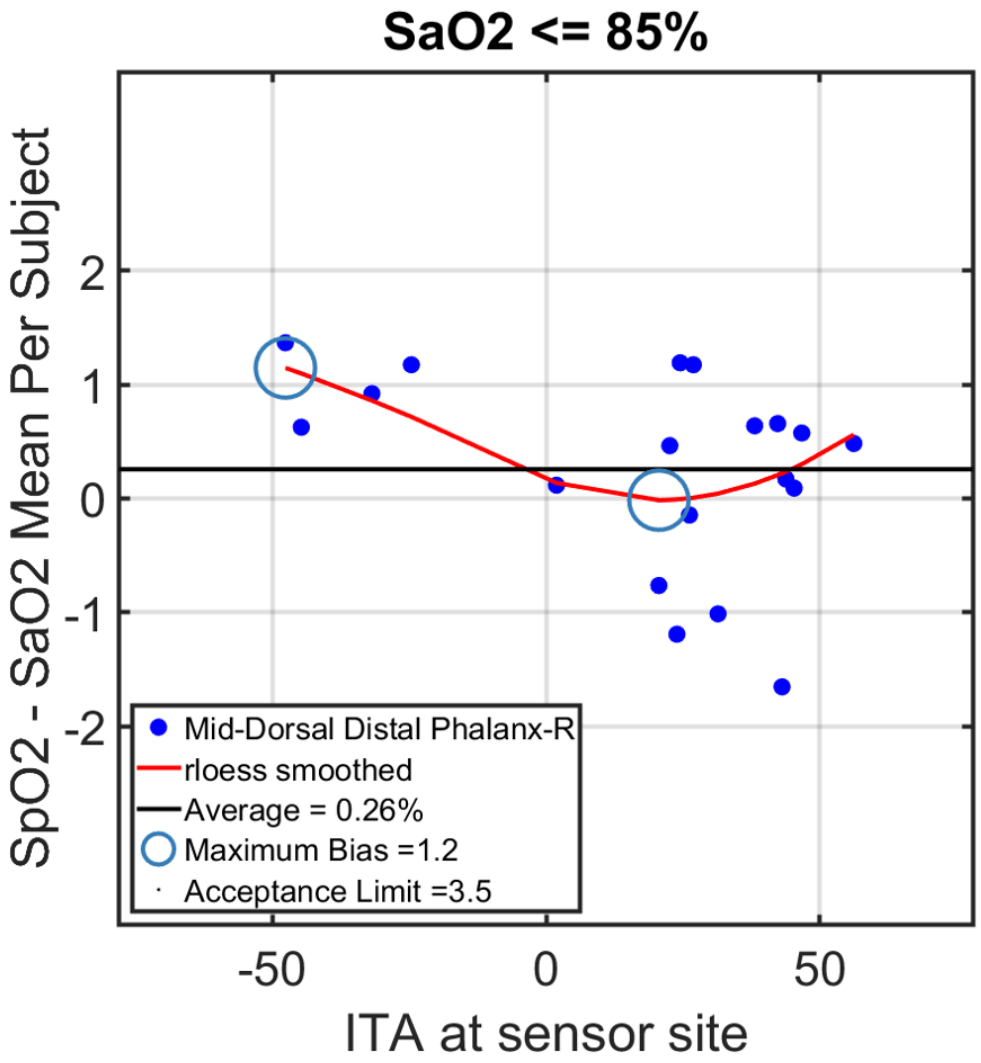
Data Consistent with SpO_2 Bias being Linear in ITA

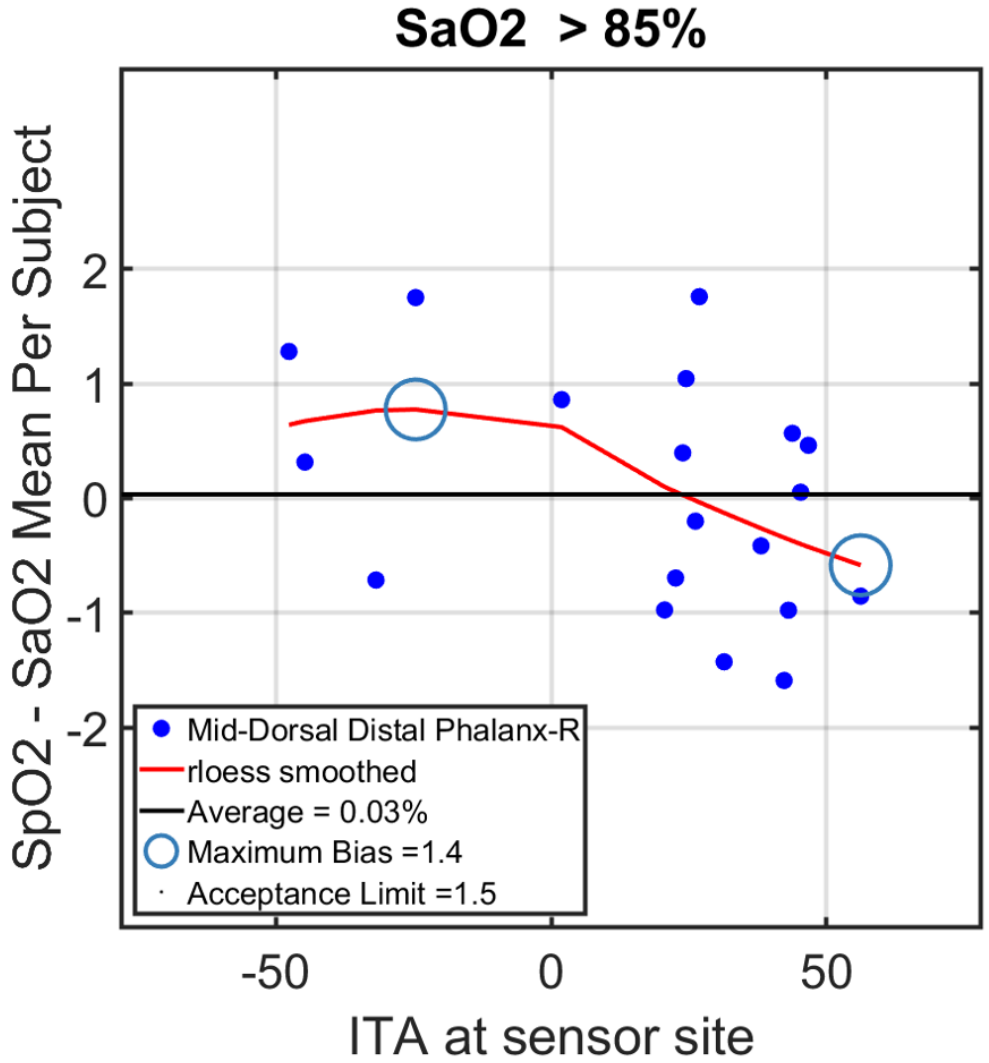


Linear Mixed Effects Model

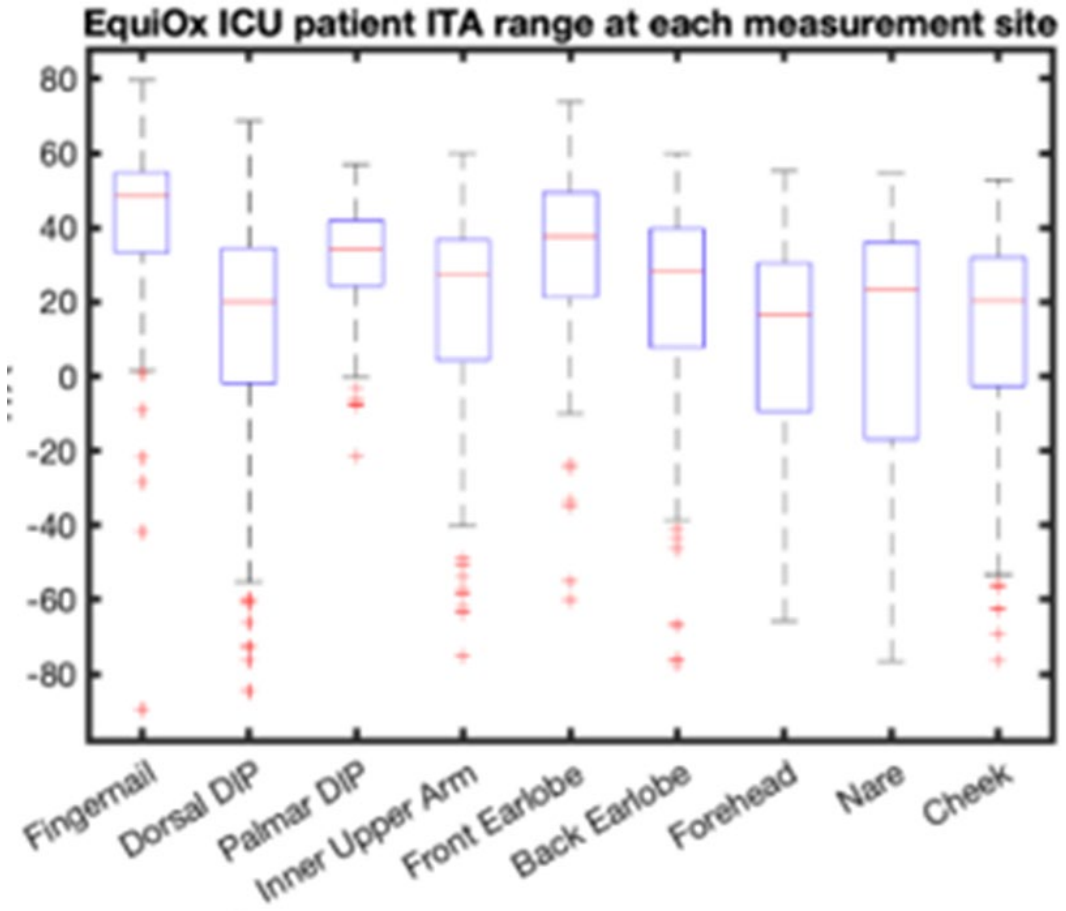
- Response variable is $\boldsymbol{D = SpO_2 - SaO_2}$.
- $\boldsymbol{SpO_2}$ **bias** is the expected value (i.e., mean) of $\boldsymbol{D}$.
- $\boldsymbol{SpO_2}$ **bias** is assumed to be **linear** in
  - $\boldsymbol{SaO_2}$,
  - $\boldsymbol{ITA}$, and
  - their **interaction**.
- The model should be checked for goodness-of-fit to the data.
  - In particular, the assumption that $\boldsymbol{SpO_2}$ **bias is linear in** $\boldsymbol{ITA}$ should be checked.

Non-Linear Regression of SpO_2 Bias on ITA





ITA Range by Location, ICU Patients, RW Data



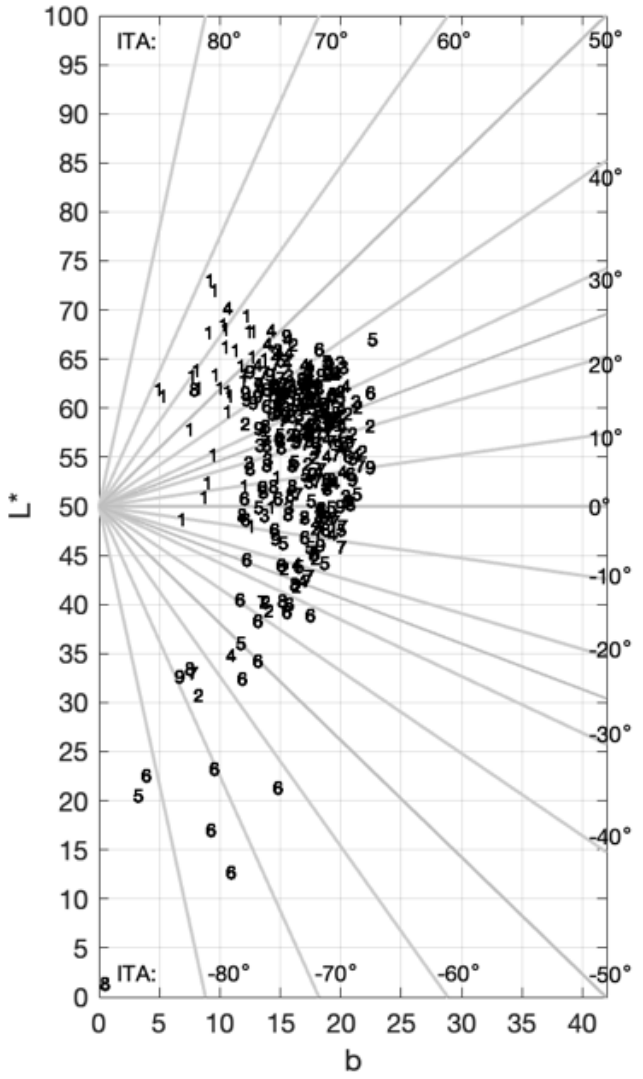
Plot 1. ITA ranges in ICU patients

Location	Interquartile Range
Fingernail	$(35,55)$
Dorsal DIP	
Palmar DIP	
Inner Upper Arm	
Front Earlobe	
Back Earlobe	
Forehead	
Nare	
Cheek	

- **smallest** 25th percentile of ITA is −15°
- **largest** 75th percentile of ITA is 55°

ITA Range by Location, Desaturation Controlled Study, N=34

Location	ITA min,max
1	−11.5, 68.3
2	−67.0, 46.9
3	−3.8, 47.1
4	−54.5, 62.2
5	−83.7, 44.8
6	−81.9, 41.3
7	−65.6, 44.8
8	−89.4, 56.7
9	−69.0, 48.2



Diagnostic Accuracy

ROC Analysis

Diagnostic Accuracy

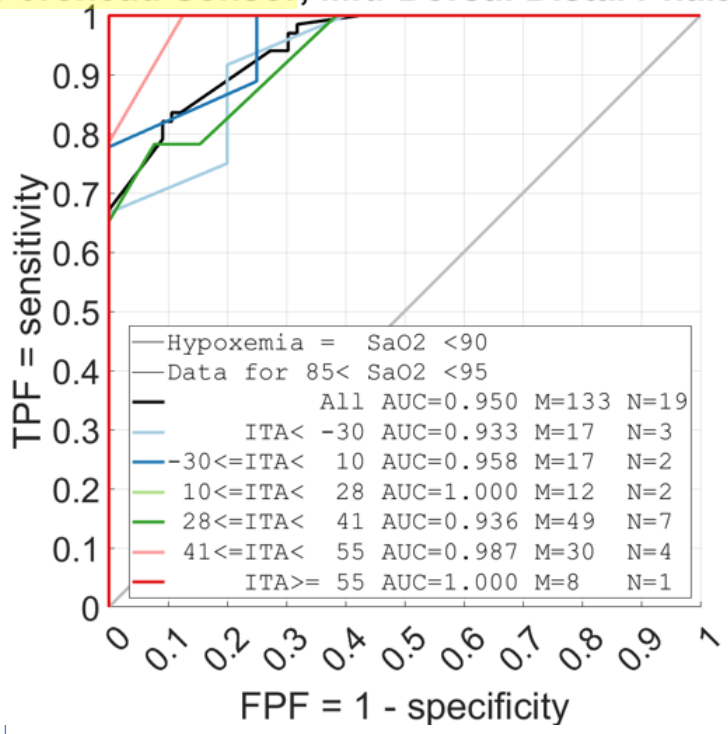
- **Target Condition**: $SaO_2 < 90\%$.
- **Test Positive Result:** $SpO_2 <$ some cut-off value.
  - e.g., $90\%, 92\%, 94\%$, etc.
- **True Positive Fraction (TPF)**, or **Sensitivity**
  - Probability that a subject with the condition tests positive for it.
- **False Positive Fraction (FPF)**, or **1 – Specificity**
  - Probability that a subject without the condition tests positive for it.

Receiver Operating Characteristic (ROC) Curve

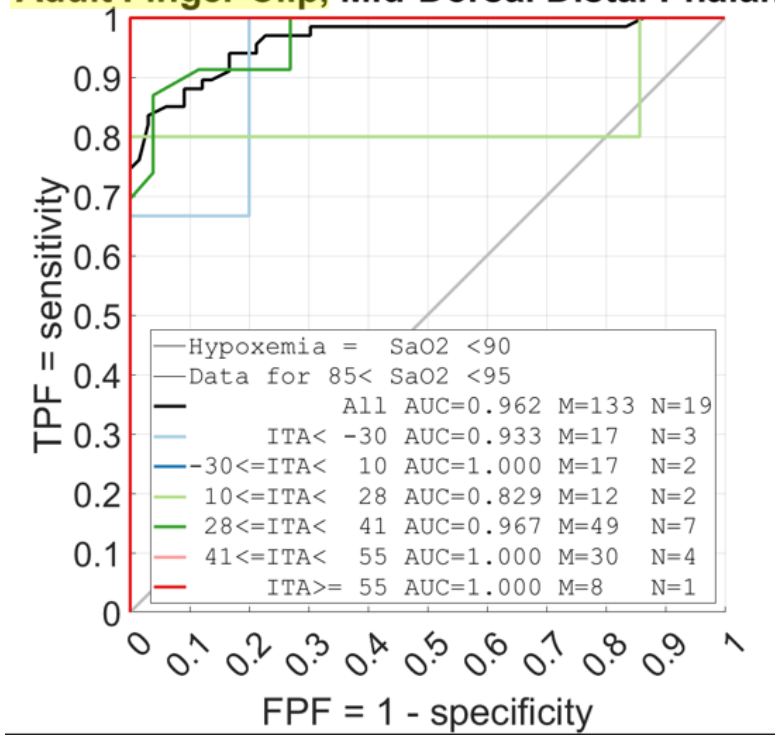
- **Target Condition**: $SaO_2 < 90\%$.
- **Test Positive Result:** $SpO_2 <$ some cut-off value.
- Receiver operating characteristic (**ROC**) curve
  - Plot of TPF vs FPF for every possible cut-off in SpO_2 that could be used to define a test positive result for the condition $SaO_2 < 90\%$.
- **AUC** is the area under the ROC curve.

ROC Plots, Controlled Desaturation Study Data

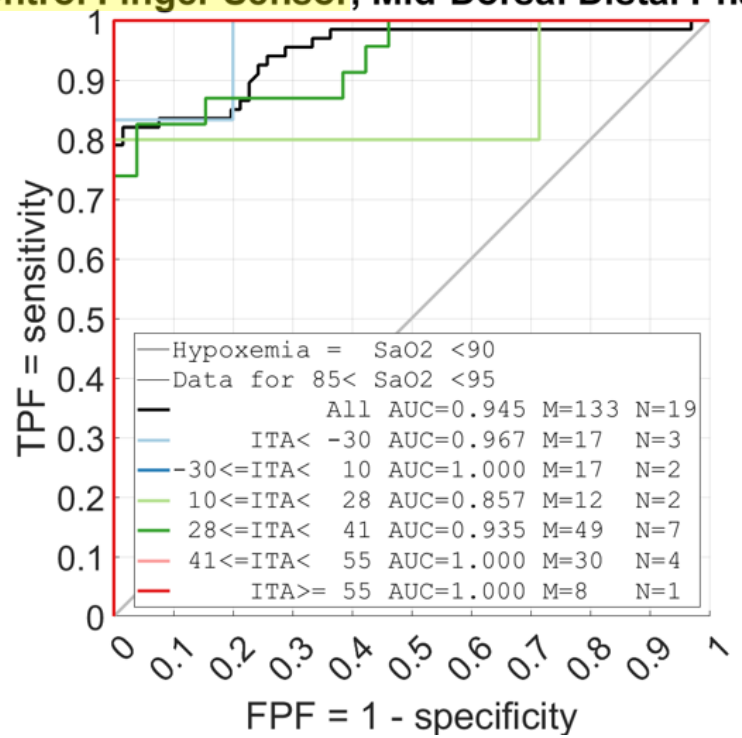
Forehead Sensor, Mid-Dorsal Distal Phalanx-L



Adult Finger Clip, Mid-Dorsal Distal Phalan

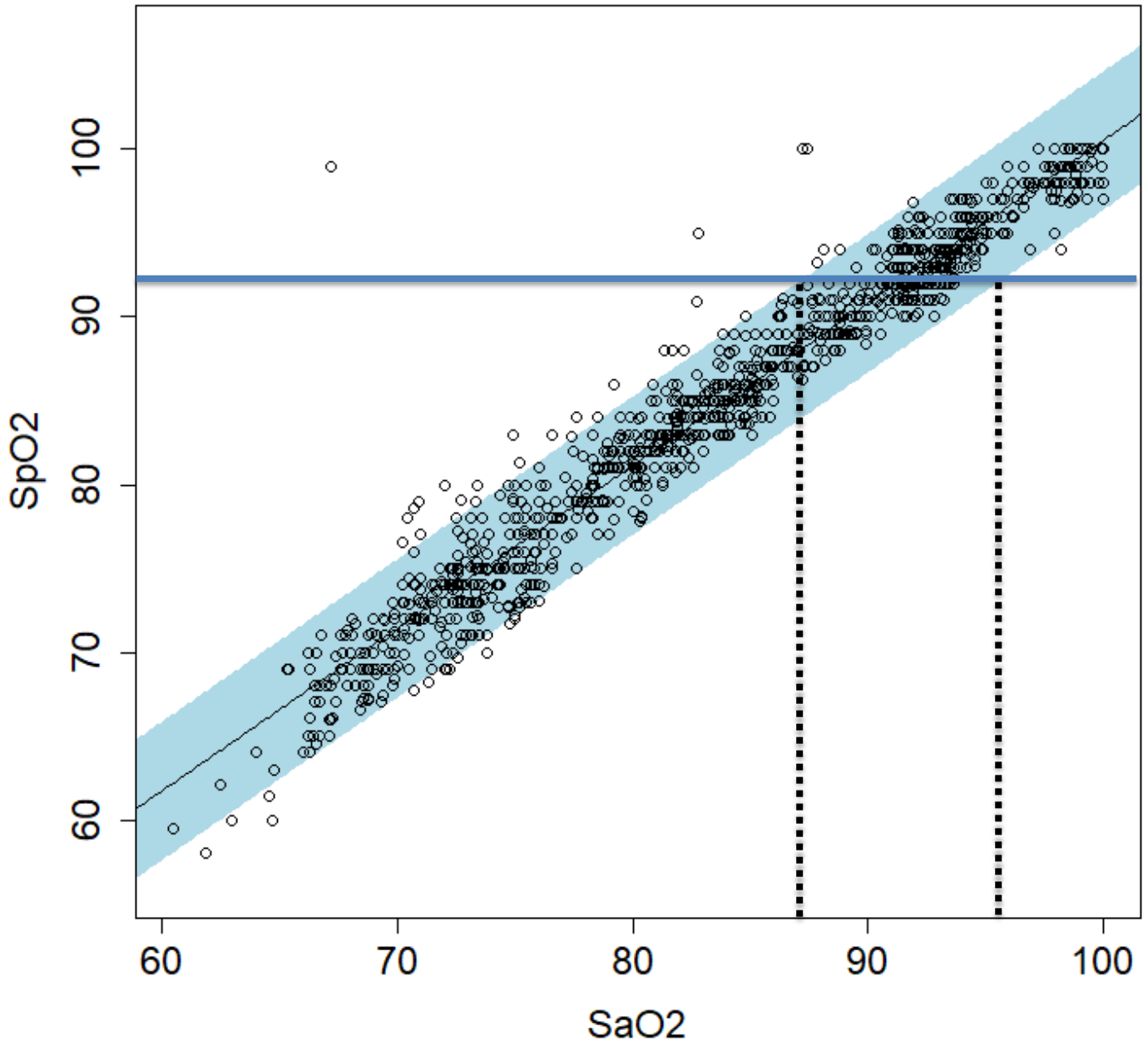


Control Finger Sensor, Mid-Dorsal Distal Phalanx-L



Target Condition: $SaO_2 < 90\%$

Inverse Prediction



95% pointwise inverse prediction band of actual SaO2 value given the SpO2 value.

Summary

Summary

- FDA proposes sample sizes of
  - $N = 24$ subjects,
  - $M = 480$ total number of simultaneous measurements of $(\boldsymbol{SaO_2}, \boldsymbol{SpO_2})$.
- FDA proposes co-primary analyses of
  - $\boldsymbol{A_{RMS}}$
  - Disparity in $\boldsymbol{SpO_2}$ **bias** across ITA levels and MST levels
    - **Question**: For the ITA analysis, what ITA interval should be used for evaluating the maximum difference in $\boldsymbol{SpO_2}$ bias between ITA values?
    - **Note**: MST analysis was not covered in this presentation, but is similar to ITA analysis.
- Other analyses may be worth considering, e.g.,
  - Diagnostic accuracy of $\boldsymbol{SpO_2}$ for classifying a target condition, e.g., $\boldsymbol{SaO_2 < 90\%}$
  - Prediction interval of plausible $\boldsymbol{SaO_2}$ values given the $\boldsymbol{SpO_2}$ value
  - Repeatability of the $\boldsymbol{SpO_2}$ value, e.g., standard deviation or coefficient of variation of $\boldsymbol{SpO_2}$ among repeated measures under the same or different conditions of measurement.
  - Association of Race / Ethnicity with $\boldsymbol{SpO_2}$ bias.