Review

Impact of Skin Pigmentation on Pulse Oximetry Blood Oxygenation and Wearable Pulse Rate Accuracy: Systematic Review and Meta-Analysis

Sanidhya Singh[1], BSE; Miles Romney Bennett[2], MS; Chen Chen[2], PhD; Sooyoon Shin[2], PhD; Hamid Ghanbari[1*], MD, MPH; Benjamin W Nelson[2,3*], PhD

[1]University of Michigan Medical School, Ann Arbor, MI, United States

[2]Verily Life Sciences LLC, South San Francisco, CA, United States

[3]Division of Digital Psychiatry, Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Sanidhya Singh, BSE
University of Michigan Medical School
1301 Catherine Street
Ann Arbor, MI, 48109
United States
Phone: 1 7349364000
Email: sansin@med.umich.edu

Abstract

Background: Photoplethysmography (PPG) is a technology routinely used in clinical practice to assess blood oxygenation (SpO_2) and pulse rate (PR). Skin pigmentation may influence accuracy, leading to health outcomes disparities.

Objective: This systematic review and meta-analysis primarily aimed to evaluate the accuracy of PPG-derived SpO_2 and PR by skin pigmentation. Secondarily, we aimed to evaluate statistical biases and the clinical relevance of PPG-derived SpO_2 and PR according to skin pigmentation.

Methods: We identified 23 pulse oximetry studies (n=59,684; 197,353 paired SpO_2-arterial blood observations) and 4 wearable PR studies (n=176; 140,771 paired PPG-electrocardiography observations). We evaluated accuracy according to skin pigmentation group by comparing SpO_2 accuracy root-mean-square values to the regulatory threshold of 3% and PR 95% limits of agreement values to +5 or –5 beats per minute (bpm), according to the standards of the American National Standards Institute, Association for the Advancement of Medical Instrumentation, and the International Electrotechnical Commission. We evaluated biases and clinical relevance using mean bias and 95% CI.

Results: For SpO_2, accuracy root-mean-square values were 3.96%, 4.71%, and 4.15%, and pooled mean biases were 0.70% (95% CI 0.17%-1.22%), 0.27% (95% CI –0.64% to 1.19%), and 1.27% (95% CI 0.58%-1.95%) for light, medium, and dark pigmentation, respectively. For PR, 95% limits of agreement values were from –16.02 to 13.54, from –18.62 to 16.84, and from –33.69 to 32.54, and pooled mean biases were –1.24 (95% CI –5.31 to 2.83) bpm, –0.89 (95% CI –3.70 to 1.93) bpm, and –0.57 (95% CI –9.44 to 8.29) bpm for light, medium, and dark pigmentation, respectively.

Conclusions: SpO_2 and PR measurements may be inaccurate across all skin pigmentation groups, breaching U.S. Food and Drug Administration guidance and industry standard thresholds. Pulse oximeters significantly overestimate SpO_2 for both light and dark skin pigmentation, but this overestimation may not be clinically relevant. PRs obtained from wearables exhibit no statistically or clinically significant bias based on skin pigmentation.

(J Med Internet Res 2024;26:e62769) doi: 10.2196/62769

KEYWORDS

photoplethysmography; pulse oximetry; arterial blood gas; skin tone; skin pigmentation; bias; digital technology

XSL·FO

RenderX

Introduction

Photoplethysmography (PPG) technology has been used in medicine since the 1970s to assess pulse rate (PR) and blood oxygenation (SpO_2). The accuracy of PPG-based SpO_2 and PR is critical for medical practice, clinical decision-making, and patient outcomes [1].

Technological advancements have led to rapid expansion of this technology into consumer devices [2]. This has enabled consumers to continuously and unobtrusively track health status, generating information that is becoming commonly used in health care settings and as a source of research end points [3,4].

Researchers have identified several factors that influence the accuracy of PPG-based measurements, including skin temperature, sensor contact pressure, skin thickness, and hydration level [5,6]. Notably, darker skin pigmentation may influence PPG-based SpO_2 and PR readings [7-10]. Patients with darker skin pigmentation are more likely than patients with lighter skin color to have overestimated SpO_2 readings, leading to lower hospital admission; higher occult hypoxemia; and delayed or no access to dexamethasone, therapeutic oxygen, and COVID-19 therapies, resulting in increased hospital readmission, organ dysfunction, and mortality [11-18].

Starting in 2013, a series of U.S. Food and Drug Administration (FDA) guidances, safety communications, and more recently, statements from attorneys general have called to address darker skin pigmentation bias in pulse oximeters [19]; however, these guidances lacked standards for assessing skin tone [20]. Recent pulse oximeter research has rekindled interest in skin pigmentation disparities in PPG sensor accuracy, leading to increased media and regulatory attention [21-24]. In November 2023, more than 24 attorneys general wrote a letter calling on the FDA to take urgent action to address pulse oximeter skin pigmentation disparities [25,26].

More recent research has also highlighted the potential presence of these biases in research-grade and consumer PPG-based wearable devices [4,27-29], leading to calls to address inequity, bias, and discrimination in wearable health technology and clinical practice algorithms [30,31].

Accuracy guidance for pulse oximeters and heart rate or PR devices has been delineated by regulatory bodies, industry, and medical standards. Overall accuracy of pulse oximeters can be assessed with accuracy root-mean-square (A_{rms}), which combines mean bias and precision (SD of bias) into a single metric [32]. FDA guidance has a threshold of $A_{rms} \leq 3\%$ [19] for transmittance devices and a threshold of $A_{rms} \leq 3.5\%$ [19] for ear clip and reflectance devices, while international thresholds are set at $A_{rms} \leq 4\%$ [32]. For heart rate or PR devices, overall accuracy can be assessed with 95% limits of agreement (LoAs). The American National Standards Institute (ANSI), Association for the Advancement of Medical Instrumentation (AAMI), and International Electrotechnical Commission (IEC) have set the recommendation that electrocardiography (ECG) devices have mean bias of +5 or –5 beats per minute (bpm) or mean absolute percentage error $\leq 10\%$, whichever is greater [33,34]. Devices with accuracy measures breaching these accuracy thresholds can produce questionable results.

It is, therefore, critical to generate evidence to inform the design and calibration of these devices to reduce algorithmic bias and improve accuracy, so that PPG-based technology can generalize to all segments of the population, mitigating racial disparities in health outcomes. There have been systematic reviews on PPG skin pigmentation bias, but these have been limited to either consumer device PR [35] or pulse oximeter SpO_2 [36-38]. Only 1 meta-analysis focused on pulse oximetry for this topic is available [36], and an additional 7 studies have been published recently, providing an additional 50,980 participants and 182,369 paired observations. Therefore, performing a comprehensive meta-analysis to examine PPG accuracy, potential bias, and clinical relevance for SpO_2 and PR by skin pigmentation is timely.

Methods

Search Strategy and Selection Criteria

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [39] and developed a MEDLINE systematic search. Searches were performed between April 2022 and June 2023, and a final ad hoc search was conducted in June 2023. No additional studies were included after this date.

We included studies that (1) investigated PPG-derived SpO_2 or PR test devices per definition [40]; (2) used arterial blood gas (SaO_2) or ECG as the reference device; and (3) reported mean bias and SD, SE, 95% LoA, or 95% CI by race, ethnicity, or skin tone. Inclusion disagreements were resolved by consensus.
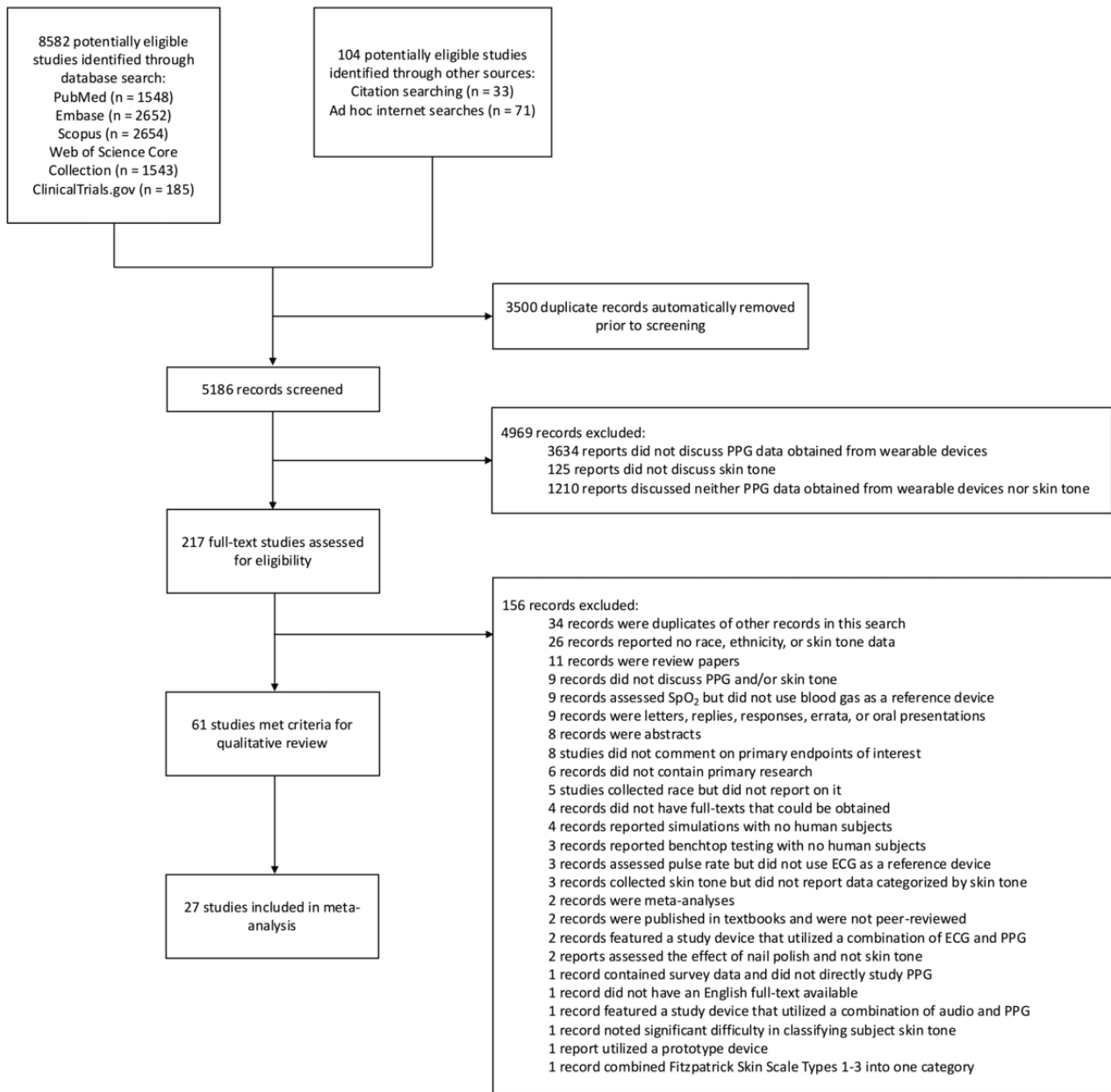
Exclusion criteria included (1) literature reviews, systematic reviews, commentary, and meta-analyses; (2) non-English manuscripts; (3) irretrievable full source texts; (4) studies on remote PPG; and (5) those that lack SaO_2 or ECG as the reference device. We chose to not exclude papers based on measurement hardware or underlying algorithms given that all measurement devices relied on contact-based PPG to measure the same end points of interest—SpO_2 and PR.

Data Analysis

Data Extraction

Data were independently extracted from published manuscripts (BWN and SS). A third and fourth reviewer (MRB and HG) adjudicated any differences between the initial 2 reviewers and resolved any disagreements by checking manuscripts. The first author verified the quality of manuscripts. Standardized data extraction was developed to extract the study characteristics (see the Materials—Data Extraction section in Multimedia Appendix 1). For 1 paper, we used participant-level data from the supplementary materials to calculate mean bias and SE by skin pigmentation category [41]. This resulted in 27 studies in our final analysis (Figure 1 and Table 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. ECG: electrocardiography; PPG: photoplethysmography; SpO$_2$: blood oxygenation.

8582 potentially eligible studies identified through database search:
PubMed (n = 1548)
Embase (n = 2652)
Scopus (n = 2654)
Web of Science Core Collection (n = 1543)
ClinicalTrials.gov (n = 185)

104 potentially eligible studies identified through other sources:
Citation searching (n = 33)
Ad hoc internet searches (n = 71)

3500 duplicate records automatically removed prior to screening

5186 records screened

4969 records excluded:
3634 reports did not discuss PPG data obtained from wearable devices
125 reports did not discuss skin tone
1210 reports discussed neither PPG data obtained from wearable devices nor skin tone

217 full-text studies assessed for eligibility

156 records excluded:
34 records were duplicates of other records in this search
26 records reported no race, ethnicity, or skin tone data
11 records were review papers
9 records did not discuss PPG and/or skin tone
9 records assessed SpO$_2$ but did not use blood gas as a reference device
9 records were letters, replies, responses, errata, or oral presentations
8 records were abstracts
8 studies did not comment on primary endpoints of interest
6 records did not contain primary research
5 studies collected race but did not report on it
4 records did not have full-texts that could be obtained
4 records reported simulations with no human subjects
3 records reported benchtop testing with no human subjects
3 records assessed pulse rate but did not use ECG as a reference device
3 records collected skin tone but did not report data categorized by skin tone
2 records were meta-analyses
2 records were published in textbooks and were not peer-reviewed
2 records featured a study device that utilized a combination of ECG and PPG
2 reports assessed the effect of nail polish and not skin tone
1 record contained survey data and did not directly study PPG
1 record did not have an English full-text available
1 record featured a study device that utilized a combination of audio and PPG
1 record noted significant difficulty in classifying subject skin tone
1 report utilized a prototype device
1 record combined Fitzpatrick Skin Scale Types 1-3 into one category

61 studies met criteria for qualitative review

27 studies included in meta-analysis

Table 1. Characteristics of included studies.

Device type and study (author, year)	Sample size, n	Test device evaluated	Reference device	SaO$_2$[a] (%) range	Participant population	Research setting	Skin pigmentation method
Pulse Oximetry							
Abrams et al [42], 2003	200	• Nellcor N-200	Radiometer ABL-520	Not reported	Medical—adult patients with cirrhosis	Inpatient	Race—Black and White
Adler et al [43], 1998	298	• Nellcor D-25	4-wavelength spectro-photometer, or co-oximeter (Radiometer OSM3)	50-99	Medical—adult emergency department patients	Inpatient	Skin tone—Munsell color system categorized into light, medium, or dark
Andrist et al [44], 2022	1061	• Not reported	Not reported	Not reported	Medical—pediatric	Inpatient	Race
Barker and Wilson [45], 2023	75	• Masimo SET[b] pulse oximeters with RD-SET sensors	Radiometer ABL-835 Flex CO-Oximeter	70-100	Mixed—healthy and mild systemic disease	Research—laboratory	Race
Bickler et al [10], 2005	23	• Nellcor N-595 with Nellcor OxiMax A finger probe • Novametrix 513s models (2 types) • Nonin Onyx models (2 types)	Radiometer OSM3	60-100	Healthy—non-smoking	Research—laboratory	Ethnicity—light (Northern European) and dark (African American)
Bothma et al [46], 1996	100	• Simed S100e • Nihon Koden • Ohmeda 3740	IL482 Co-oximeter System	87.8-99.2	Medical—critically ill adult patients	Inpatient	Skin tone—EEL[c] reflectance spectrophotometer; all participants had dark skin tone
Burnett et al [47], 2022	46,253	• Unspecified Nellcor and Masimo devices	GEMStat Premier 3000	Not reported	Medical—patients receiving anesthesia	Inpatient	Race and ethnicity
Crooks et al [48], 2022	2997	• Not reported	Not reported	Not reported	Medical—COVID-19	Inpatient	Race
Ebmeier et al [49], 2018	394	• Marquette Rac-4A monitors with Masimo sensors and Philips IntelliVue • MP70 monitors with Philips • Adult Reusable SpO$_2$[d] sensors	Radiometer ABL 800 FLEX SaO$_2$ analyzer	Not reported	Medical—multiple conditions	Inpatient	Ethnicity
Fawzy et al [17], 2022	1216	• Not reported	ABL825, ABL827, or ABL90 blood gas analyzers	Not reported	Medical—COVID-19	Inpatient	Race and ethnicity

XSL•FO
RenderX

Device type and study (author, year)	Sample size, n	Test device evaluated	Reference device	SaO_2[a] (%) range	Participant population	Research setting	Skin pigmentation method
Feiner et al [8], 2007	36	• Nellcor N-595 (Oxi-Max A adhesive probe) • Nellcor N-595 (a clip-type probe) • Masimo Radical (clip probe) • Masimo Radical (adhesive disposable probe) • Nonin 9700 (clip-type probe) • Nonin 9700 (disposable adhesive probe)	Radiometer OSM3 multi-wavelength oximeter	60-100	Healthy—non-smoking	Research—laboratory	Ethnicity—light (Caucasian), intermediate (Hispanic, Indian, Filipino, or Vietnamese), and dark (African American) categories
Foglia et al [50], 2017	36	• Nellcor Oximax (Covidien) • Masimo Rainbow SET Radical 7	Siemens Rapidlab 1265	60-92	Medical—infants with cyanotic congenital heart disease and oxygen saturation <90%	Inpatient	Skin tone—Munsell color system
Jubran and Tobin [51], 1990	54	• Nellcor pulse oximeter with disposable or reusable probes • Ohmeda-Biox3700 pulse oximeter with reusable probe	CO-oximetry	Not reported	Medical—critically ill, ventilator-dependent patients	Inpatient	Race—Black and White
McGovern et al [52], 1996	8	• Ohmeda 3700	IL 482 Co-oximeter	Not reported	Medical—adults with stable condition with severe COPD[f]	Research—laboratory	Race—all White
Muñoz et al [53], 2008	846	• Minolta Pulsox-7	IL 682 co-oximeter	Not reported	Medical—adults under assessment for long-term home oxygen therapy	Outpatient	Race—all White
Pilcher et al [54], 2020	400	• Carescape B450 monitor with Nellcor probe • GE Dash 3000 • Masimo Radical 7 • Masimo SET Quartz (unspecified) • Masimo SET Quartz Q400 • Nonin 2120 • Nonin 2140 • Nonin Avant (unspecified) • Nonin Avant 4000 • Nonin Avant 9700 • Nonin Lifesense Medair • Novametrix Model 512 • Ohmeda Biox 3700E with a GE TruSignal or Nellcor probe • Philips Intellivue MP70 with a GE TruSignal Nellcor or Philips probe • Welch Allyn with a Nellcor probe	Radiometer ABL800	72-100	Medical—hospitalized adult patients	Inpatient and outpatient	Skin Tone—Fitzpatrick Scale categorized into light, medium, and dark
	774	• Not reported	Not reported			Inpatient	Race

Device type and study (author, year)	Sample size, n	Test device evaluated	Reference device	SaO$_2$[a] (%) range	Participant population	Research setting	Skin pigmentation method
Ruppel et al [55], 2023				Not reported	Medical—cardiac catheterization		
Sudat et al [56], 2023	8735	• Not reported	Not reported	Not reported	Medical—multiple conditions	Inpatient	Race
Thrush and Hodges [57], 1994	25	• Critikon Dinamap Plus Model 8700 • Critikon Oxyshuttle Ohmeda 3700 Catalyst Research MiniOx IV	IL482 co-oximeter	80-100	Healthy—non-smoking adults	Research—laboratory	Race—all White
Valbuena et al [12], 2022	372	• Not reported	Not reported, blood gas analysis	Not reported	Medical—adult patients with respiratory failure or COVID-19	Inpatient	Race and ethnicity—White, Black, Hispanic, and Asian
Vesoulis et al [58], 2021	294	• Nellcor SpO$_2$ module with Neonatal-Adult MAX-N adhesive SpO$_2$ sensor (Covidien) (used with either Philips IntelliVue MP70 or MX800 monitors)	Radiometer ABL800 Flex	Not reported	Medical—preterm infants at neonatal intensive care unit	Inpatient	Race—White and Black
Wiles et al [59], 2022	194	• Nellcor reusable SpO$_2$ probes or Mindray disposable SpO$_2$ probes (GE Healthcare B1x5 M/P monitor)	RAPIDpoint 500 analyser (Siemens Healthcare GmbH)	Not reported	Medical—adult patients with COVID-19 pneumonitis	Inpatient	Race—Asian, Black, White, and other
Zeballos and Weisman [9], 1991	33	• Hewlett-Packard HP-47201A • Ohmeda Biox IIA	IL282 co-oximeter	Not reported	Healthy—non-smoking volunteers	Research—laboratory	Race—all Black

Pulse Rate

Device type and study (author, year)	Sample size, n	Test device evaluated	Reference device	SaO_2[a] (%) range	Participant population	Research setting	Skin pigmentation method
Bent et al [41], 2020	53	• Empatica E4 • Apple Watch • Fitbit Charge • Garmin Vivosmart 3 • Xiaomi Miband • Biovotion Everion	BittiumFaros 180, Bittium Inc.	N/A[g]	Healthy	Research —laboratory	Skin tone—Fitzpatrick Scale: • 1 (n=7) • 2 (n=8) • 3 (n=10) • 4 (n=9) • 5 (n=9) • 6 (n=10)
Nelson and Allen [60], 2019	1	• Apple Watch 3 • Fitbit Charge 2	Vrije Universiteit Ambulatory Monitoring System	N/A	Healthy	Research — real world	Skin tone—Fitzpatrick Scale: • 1 (n=1)
Sanudo et al [61], 2019	45	• Apple Watch (version not reported)	Polar Chest Strap	N/A	Healthy	Research —laboratory	Skin tone—Fitzpatrick Scale: • 2 (n=15) • 3 (n=15) • 4 (n=15)
Chow and Yang [62], 2020	40	• Garmin Vivosmart HR+ • Xiaomi Mi Band 2	Polar H7 Chest Strap	N/A	Healthy	Research —laboratory	Ethnicity and skin tone—East Asian, Fitzpatrick Scale 3 and 4

[a]SaO_2: arterial blood gas.

[b]SET: Signal Extraction Technology.

[c]EEL: electron energy loss.

[d]SpO_2: blood oxygenation.

[f]COPD: chronic obstructive pulmonary disease.

[g]N/A: not applicable.

Quality Assessment of the Overall Evidence

QUADAS-2 tool was used to evaluate risk of bias and applicability (Figure S1 in Multimedia Appendix 1). Funnel plots evaluated publication bias using the *metafor* package [63] (Figures S2 and S3 in Multimedia Appendix 1).

Statistical Analysis

Skin Tone Categorization

We mapped skin tone, race, and ethnicity into 3 primary skin pigmentation groups of light, medium, and dark, following published methodology [36] (see Table S1 in Multimedia Appendix 1) and examined biases for SpO_2 and PR by each skin pigmentation group (described below).

We elected to use this same skin pigmentation categorization schema as has previously been used, with the goal of expanding upon the analysis of Shi et al [36].

Statistical Analysis

The objective of the study was to assess whether the devices were accurate in estimating SpO_2 and PR when compared with a SaO_2 and ECG reference device, respectively, for each skin pigmentation group. If these measures were found to be inaccurate, biases were quantified and their clinical relevance was assessed. The analytical approach to execute these research objectives was formulated based on methodologies used in prior meta-analyses within the discipline [36].

Evaluation End Points and Criteria

A summary of the evaluation criteria used can be found in Table 2.

Table 2. Evaluation end points and criteria for blood oxygenation (SpO_2) and pulse rate studies.

Category and objective	Evaluation end point	Evaluation criteria
SpO_2		
Evaluate accuracy	A_{rms}[a]	Accurate if $A_{rms} \leq 3\%$
Statistically significant bias	Mean bias (95% CI)	Statistically significant bias if 95% CI does not contain 0
Clinically relevant bias	Mean bias (95% CI)	Clinically relevant bias if either upper bound of 95% CI $\leq 4\%$ or lower bound of 95% CI $> 4\%$
Pulse rate		
Evaluate accuracy	95% LoA[b]	Accurate if 95% LoA is bounded by +5 or −5 bpm[c]
Statistically significant bias	Mean bias (95% CI)	Statistically significant bias if 95% CI does not contain 0
Clinically relevant bias	Mean bias (95% CI)	Clinically relevant bias if either upper bound of 95% CI ≤ 5 bpm or lower bound of 95% CI > 5 bpm

[a]A_{rms}: accuracy root-mean-square.

[b]LoA: limits of agreement.

[c]bpm: beats per minute.

Accuracy

To evaluate the accuracy of SpO_2, A_{rms} was used to measure accuracy, because of its commonality in the regulatory space [19]. It was calculated as $\sqrt{bias^2 + precision^2}$, where precision is the SD of bias [42,43]. A_{rms} is useful in clinical settings because it provides a single metric that accounts for both the systematic error (bias) and the random error (precision) in measurements and it provides a comprehensive assessment of how close the measurements are to the true values. A lower A_{rms} value indicates better accuracy, meaning that the measurement is closer to the true value, whereas a higher A_{rms} value suggests lower accuracy, indicating that the measurement is further away from the true value.

For SpO_2, our study used the stricter $A_{rms} > 3\%$ threshold to define inaccuracy as per FDA guidance [19]. We used this threshold as 22 (96%) out of 23 of pulse oximeter studies used a transmittance device and only one used an ear clip, which would have set a more liberal threshold of 3.5%. To evaluate accuracy of PR, the 95% LoA of bias was used, and LoA was constructed as mean bias (SD 1.96). For PR, a pooled 95% LoA not bounded by +5 or −5 bpm was considered inaccurate per ANSI/AAMI/IEC standards [33,34].

Bias

To evaluate bias of both SpO_2 and PR, mean bias and its 95% CI were used. For each data point that compares the test with the reference device, bias was constructed as *test device − reference device*. Based on the final results, a measure was considered to have statistically significant bias if the 95% CI of its mean bias did not contain 0, with the estimated mean bias over 0 indicating overestimation and under 0 indicating underestimation.

Clinical Relevance of Estimated Bias

To evaluate the clinical relevance of the estimated bias for SpO_2, if the 95% CI of mean bias was out of the +4% to −4% range

it was considered clinically relevant. This threshold is inferred from FDA guidance on *Pulse Oximeters—Premarket Notification Submissions [510(k)s]: Guidance for Industry and Food and Drug Administration Staff* [19]. To evaluate the clinical relevance of the estimated bias for PR, if the 95% CI of mean bias was out of the bound of +5 or −5 bpm, it was considered clinically relevant as per ANSI/AAMI/IEC standards [33,34].

Statistical Meta-Analysis

To obtain pooled results for above listed end points, we collected sample size, paired observations, mean bias, and SD from various studies. When not available, we either transformed relevant parameters, such as 95% LoA and 95% CI, into these statistics or obtained them by analyzing the raw data.

Methods to Pool Measures of Bias

Correlated and hierarchical effect (CHE) models were used to pool mean biases from various studies. Specifically, 3-level hierarchical models were constructed, with Level 1 representing individual data points collected in a study, Level 2 representing potentially multiple comparisons within a research study, and Level 3 representing various studies included in the meta-analysis. Within a study, there may be multiple devices compared using the same participants' data, and these effect sizes are dependent. To account for this dependence, the CHE model used the robust variance estimation (RVE) method, which allowed us to combine data from single-measure design studies with multiple dependent estimates of effect size, even when the dependence is unknown [64]. To execute the CHE model with RVE, we needed to plug in assumed correlation coefficients for these dependent effect sizes. We used correlation coefficient 0.9 following published methods [36], and small (=0.30) and moderate (=0.60) correlation coefficients were also explored in sensitivity analyses. Additionally, one report [48] had substantially larger SE, so we conducted a separate analysis excluding its results as part of the sensitivity analysis. To evaluate the heterogeneity of the studies, we reported the overall I^2 (percentage of variability in the effect sizes that is not caused

by sampling error) and its breakdown of within-study (I^2_{Level2}) and between-study (I^2_{Level3}) portions. We conducted subgroup analyses to see if there were statistically significant differences in bias between skin pigmentation categories (Table S2 from Multimedia Appendix 1). Forest plots were provided to visualize the mean bias (Figures S2 and S3 from Multimedia Appendix 1).

Methods to Pool Measures of Accuracy

When it comes to providing pooled 95% LoA and A_{rms}, no established methodologies exist, so we followed an analytical approach used in prior meta-analyses within the discipline [36]. Specifically, SD of bias across studies were pooled using CHE models similar to those used to pool mean biases. The pooled mean bias and pooled SD were then used to provide pooled estimates of 95% LoA and A_{rms}. Specifically,

- Overall accuracy $A_{rms} = \sqrt{\text{the pooled mean bias}^2 + \text{the pooled SD}^2}$
- Overall 95% LoA = pooled mean bias $\pm 1.96 \times$ pooled SD

Other Methods

Descriptive statistics were provided for study characteristics, the device intended use, mean bias based on different skin tones, and reference devices. All analyses were conducted using R (version 4.3.1; R Foundation for Statistical Computing) with statistical models based on R packages *metafor* [63] and *clubSandwich* [65].

Results

Study Selection and Characteristics

Search strategy resulted in 8582 records. We selected 27 studies for full review (pulse oximetry: n=23, PR: n=4; Figure 1, Table 1, and Table 2). A total of 23 pulse oximetry studies involving 59,684 participants with 197,353 paired SpO_2-SaO_2 observations were included. Additionally, 4 PR studies with 176 participants and 140,771 paired PR-ECG observations were analyzed (Table 3).

Table 3. Summary characteristics of included studies[a].

Item and subitem	Pulse oximetry (n=23), n (%)	Pulse rate (n=4), n (%)
Participants (pulse oximetry: n=59,684; pulse rate: n=176)		
Light skin pigmentation	40,416 (67.72)	31 (17.61)
Medium skin pigmentation	9967 (16.70)	129 (73.30)
Dark skin pigmentation	9301 (15.58)	16 (9.09)
Paired observations (pulse oximetry: n=197,353; pulse rate: n=140,771)		
Light skin pigmentation	131,008 (66.38)	43,116 (30.60)
Medium skin pigmentation	32,095 (16.26)	90,733 (64.50)
Dark skin pigmentation	34,250 (17.35)	6922 (4·92)
Device type included in study		
Medical	23 (100)	0 (0)
Nonregulated	0 (0)	4 (100)
Sensor type		
Transmittance	22 (96)	0 (0)
Reflectance	1 (4)	4 (100)
Patient population		
Healthy	4 (17)	3 (75)
Medical	18 (78)	0 (0)
Healthy and medical	1 (4)	0 (0)
Not reported	0 (0)	1 (25)
Research setting		
Medical inpatient	15 (65)	0 (0)
Medical outpatient	1 (4)	0 (0)
Medical combined	1 (5)	0 (0)
Research laboratory	6 (26)	3 (75)
Research in the real world	0 (0)	1 (25)
Skin pigmentation method		
Skin tone	4 (17)	4 (100)
Race	13 (57)	0 (0)
Ethnicity	3 (13)	0 (0)
Race and ethnicity	3 (13)	0 (0)

[a]Medical device defined as devices that received regulatory clearance for either blood oxygenation (SpO$_2$) or pulse rate (PR).

Descriptive Statistics on Race, Ethnicity, and Skin Tone

Pulse Oximetry

Skin pigmentation was classified by race in 13 (57%) out of 23 studies, ethnicity in 3 (13%) studies, skin tone in 4 (17%) studies, and both race and ethnicity in 3 (13%) studies (Table 3). Of the 59,684 patients with 197,353 paired observations, there were a total of 40,416 (67.72%) patients with light skin pigmentation with 131,008 (66.38%) paired observations, 9967 (16.70%) patients with medium skin pigmentation with 32,095 (16.26%) paired observations, and 9301 (15.58%) patients with dark skin pigmentation with 34,250 (17.35%) paired observations.

PR Results

Skin pigmentation was classified by race or ethnicity in 0 (0%) out of 4 studies and skin tone in 4 (100%) studies (Table 3). Of the 176 patients with 140,771 paired observations, there were a total of 31 (17.61%) participants with light skin pigmentation with 43,116 (30.60%) paired observations, 129 (73.30%) participants with medium skin pigmentation with 90,733 (64.50%) paired observations, and 16 (9.09%) participants with dark skin pigmentation with 6922 (4.92%) paired observations.

PPG Accuracy and Bias by Skin Pigmentation
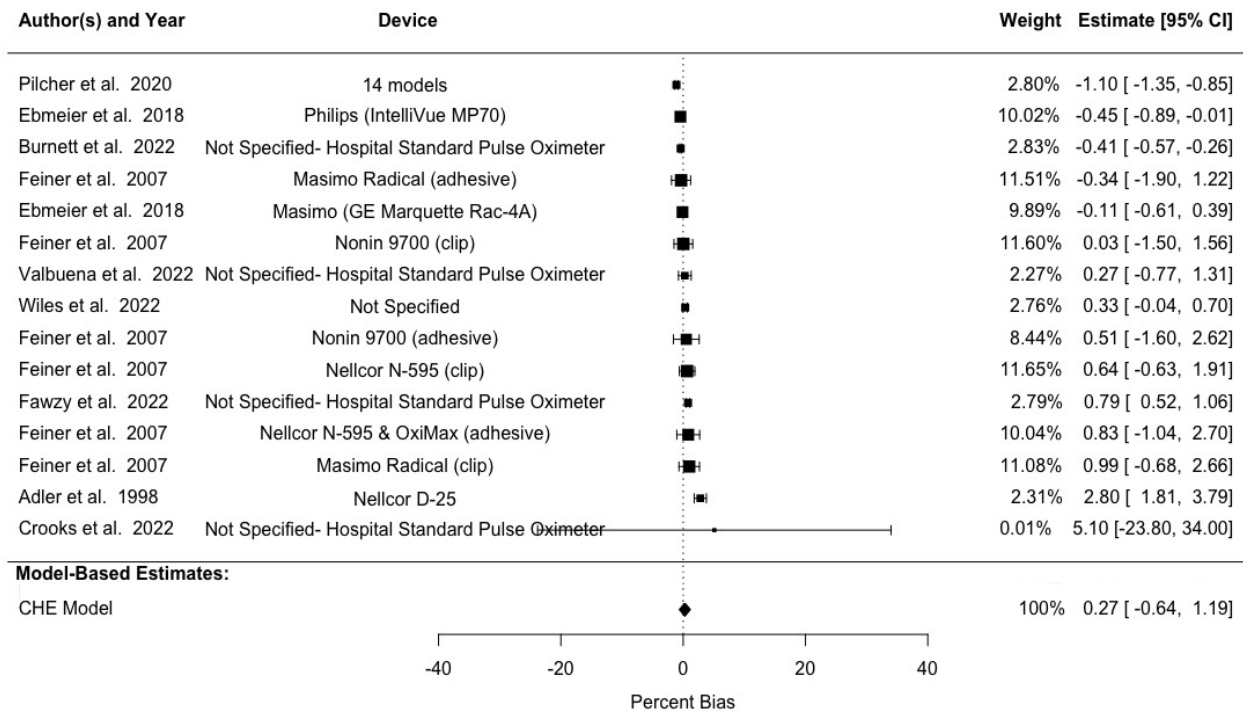
Pulse Oximetry

The pooled A_{rms} across different skin pigmentation groups was 3.6%, 3.96%, 4.71%, and 4.15% for combined, light, medium, and dark skin pigmentation, respectively (Table 4 and Figures 2-4 [8-11,17,42-51,54-56,58,59]). Of note, studies implementing multiple trial conditions or using multiple study devices were shown multiple times in Figures 2-4 to delineate different devices used within the same study. We observed a pooled mean percent bias of 0.82% (95% CI 0.29%-1.35%) across all skin pigmentation groups using the CHE model. Between-study heterogeneity (I^2_{Level3}) accounted for 14.53% of the total variation, while within-study heterogeneity (I^2_{Level2}) explained 84.02% of total variation. Delineating by skin pigmentation, the pooled mean percent bias from the CHE model was 0.70% (95% CI 0.17%-1.22%) for light skin, 0.27% (95% CI –0.64% to 1.19%) for medium skin, and 1.27% (95% CI 0.58%-1.95%) for dark skin (Tables S3 and S4 in Multimedia Appendix 1). Subgroup analyses found no statistically significant difference between light and medium skin pigmentation (estimate=0.113, SE 0.259; 95% CI –0.459 to 0.686), but there was a statistically significant difference between light and dark skin pigmentation (estimate=0.596, SE 0.240; 95% CI 0.069-1.123), such that pooled bias for those with darker skin was higher as compared to those with lighter skin (Table S2 in Multimedia Appendix 1).
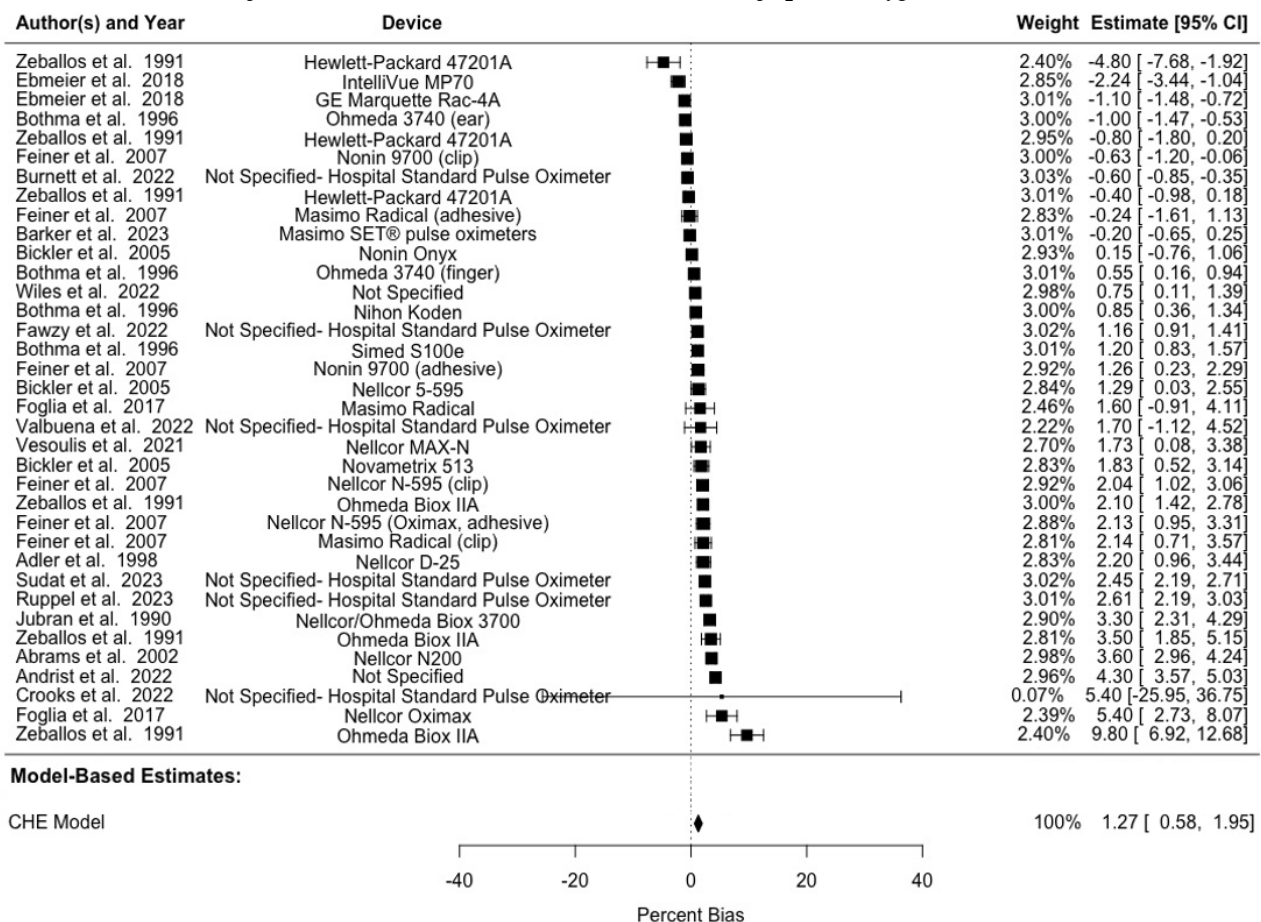
Figure 2. Forest plot showing bias in SpO$_2$ measurements in patients with light skin pigmentation. Multiple entries from the same study are included, each representing different devices evaluated. Squares denote study weight; center of squares denote observed study effect size; vertical lines denote study CIs; and diamond denotes pooled effect. CHE: correlated and hierarchical effect; SpO$_2$: blood oxygenation.
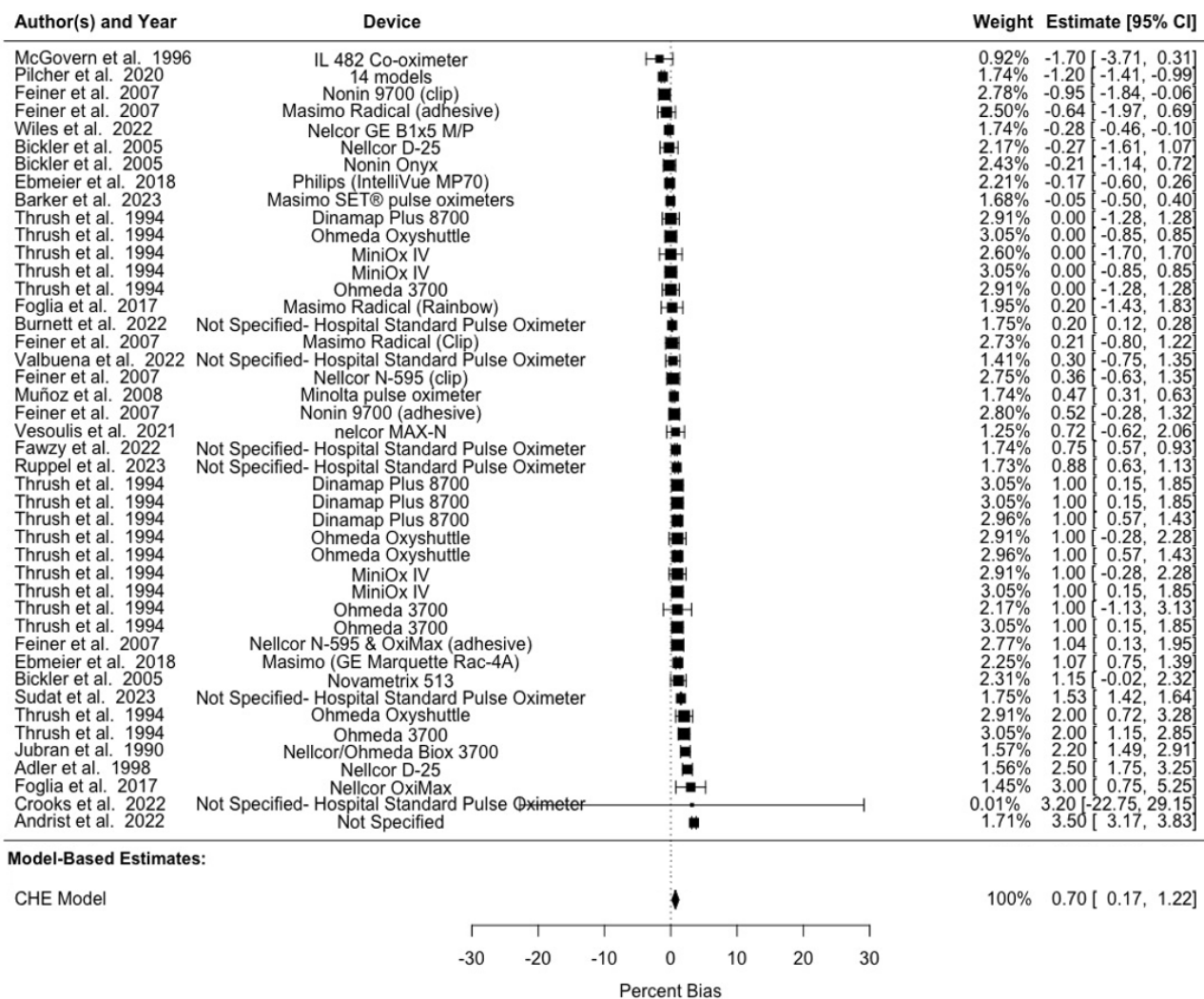
Figure 3. Forest plot showing bias in SpO_2 measurements in subjects with medium skin pigmentation. Multiple entries from the same study are included, each representing different devices evaluated. Note: squares denote study weight, center of squares denote observed study effect size, vertical lines denote study CIs, and diamond denotes pooled effect. CHE: correlated and hierarchical effect; SpO_2: blood oxygenation.

Author(s) and Year	Device	Weight	Estimate [95% CI]
Pilcher et al. 2020	14 models	2.80%	-1.10 [-1.35, -0.85]
Ebmeier et al. 2018	Philips (IntelliVue MP70)	10.02%	-0.45 [-0.89, -0.01]
Burnett et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	2.83%	-0.41 [-0.57, -0.26]
Feiner et al. 2007	Masimo Radical (adhesive)	11.51%	-0.34 [-1.90, 1.22]
Ebmeier et al. 2018	Masimo (GE Marquette Rac-4A)	9.89%	-0.11 [-0.61, 0.39]
Feiner et al. 2007	Nonin 9700 (clip)	11.60%	0.03 [-1.50, 1.56]
Valbuena et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	2.27%	0.27 [-0.77, 1.31]
Wiles et al. 2022	Not Specified	2.76%	0.33 [-0.04, 0.70]
Feiner et al. 2007	Nonin 9700 (adhesive)	8.44%	0.51 [-1.60, 2.62]
Feiner et al. 2007	Nellcor N-595 (clip)	11.65%	0.64 [-0.63, 1.91]
Fawzy et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	2.79%	0.79 [0.52, 1.06]
Feiner et al. 2007	Nellcor N-595 & OxiMax (adhesive)	10.04%	0.83 [-1.04, 2.70]
Feiner et al. 2007	Masimo Radical (clip)	11.08%	0.99 [-0.68, 2.66]
Adler et al. 1998	Nellcor D-25	2.31%	2.80 [1.81, 3.79]
Crooks et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	0.01%	5.10 [-23.80, 34.00]
Model-Based Estimates:			
CHE Model		100%	0.27 [-0.64, 1.19]

Figure 4. Forest plot showing bias in SpO_2 measurements in patients with dark skin pigmentation. Multiple entries from the same study are included, each representing different devices evaluated. Squares denote study weight; center of squares denote observed study effect size; vertical lines denote study CIs; and diamond denotes pooled effect. CHE: correlated and hierarchical effect; SpO_2: blood oxygenation.

Author(s) and Year	Device	Weight	Estimate [95% CI]
Zeballos et al. 1991	Hewlett-Packard 47201A	2.40%	-4.80 [-7.68, -1.92]
Ebmeier et al. 2018	IntelliVue MP70	2.85%	-2.24 [-3.44, -1.04]
Ebmeier et al. 2018	GE Marquette Rac-4A	3.01%	-1.10 [-1.48, -0.72]
Bothma et al. 1996	Ohmeda 3740 (ear)	3.00%	-1.00 [-1.47, -0.53]
Zeballos et al. 1991	Hewlett-Packard 47201A	2.95%	-0.80 [-1.80, 0.20]
Feiner et al. 2007	Nonin 9700 (clip)	3.00%	-0.63 [-1.20, -0.06]
Burnett et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	3.03%	-0.60 [-0.85, -0.35]
Zeballos et al. 1991	Hewlett-Packard 47201A	3.01%	-0.40 [-0.98, 0.18]
Feiner et al. 2007	Masimo Radical (adhesive)	2.83%	-0.24 [-1.61, 1.13]
Barker et al. 2023	Masimo SET® pulse oximeters	3.01%	-0.20 [-0.65, 0.25]
Bickler et al. 2005	Nonin Onyx	2.93%	0.15 [-0.76, 1.06]
Bothma et al. 1996	Ohmeda 3740 (finger)	3.01%	0.55 [0.16, 0.94]
Wiles et al. 2022	Not Specified	2.98%	0.75 [0.11, 1.39]
Bothma et al. 1996	Nihon Koden	3.00%	0.85 [0.36, 1.34]
Fawzy et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	3.02%	1.16 [0.91, 1.41]
Bothma et al. 1996	Simed S100e	3.01%	1.20 [0.83, 1.57]
Feiner et al. 2007	Nonin 9700 (adhesive)	2.92%	1.26 [0.23, 2.29]
Bickler et al. 2005	Nellcor 5-595	2.84%	1.29 [0.03, 2.55]
Foglia et al. 2017	Masimo Radical	2.46%	1.60 [-0.91, 4.11]
Valbuena et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	2.22%	1.70 [-1.12, 4.52]
Vesoulis et al. 2021	Nellcor MAX-N	2.70%	1.73 [0.08, 3.38]
Bickler et al. 2005	Novametrix 513	2.83%	1.83 [0.52, 3.14]
Feiner et al. 2007	Nellcor N-595 (clip)	2.92%	2.04 [1.02, 3.06]
Zeballos et al. 1991	Ohmeda Biox IIA	3.00%	2.10 [1.42, 2.78]
Feiner et al. 2007	Nellcor N-595 (Oximax, adhesive)	2.88%	2.13 [0.95, 3.31]
Feiner et al. 2007	Masimo Radical (clip)	2.81%	2.14 [0.71, 3.57]
Adler et al. 1998	Nellcor D-25	2.83%	2.20 [0.96, 3.44]
Sudat et al. 2023	Not Specified- Hospital Standard Pulse Oximeter	3.02%	2.45 [2.19, 2.71]
Ruppel et al. 2023	Not Specified- Hospital Standard Pulse Oximeter	3.01%	2.61 [2.19, 3.03]
Jubran et al. 1990	Nellcor/Ohmeda Biox 3700	2.90%	3.30 [2.31, 4.29]
Zeballos et al. 1991	Ohmeda Biox IIA	2.81%	3.50 [1.85, 5.15]
Abrams et al. 2002	Nellcor N200	2.98%	3.60 [2.96, 4.24]
Andrist et al. 2022	Not Specified	2.96%	4.30 [3.57, 5.03]
Crooks et al. 2022	Not Specified- Hospital Standard Pulse Oximeter	0.07%	5.40 [-25.95, 36.75]
Foglia et al. 2017	Nellcor Oximax	2.39%	5.40 [2.73, 8.07]
Zeballos et al. 1991	Ohmeda Biox IIA	2.40%	9.80 [6.92, 12.68]
Model-Based Estimates:			
CHE Model		100%	1.27 [0.58, 1.95]

XSL•FO
RenderX

Table 4. Pulse rate and pulse oximetry bias by skin pigmentation[a].

Device type and skin pigment category	Studies (evaluations), n	Sample size (data pairs), n	Unit	Pooled mean bias (95% CI)	Pooled SD (SE)	95% LoA[b]	A_{rms}[c] (%)	Overall I^2 (between- and within-study heterogeneity)
Pulse oximetry								
Light	20 (44)	40,416 (131,008)	Percent	*0.70 (0.17 to 1.22)* [d]	3.90 (1.36)	–6.94 to 8.34	3.96[e]	97.45% (46.44% and 51.01%)
Medium	9 (15)	9,967 (32,095)	Percent	0.27 (–0.64 to 1.19)	4.71 (1.71)	–8.95 to 9.50	4.71[e]	95.31% (81.16% and 14.14%)
Dark	19 (36)	9,301 (34,250)	Percent	*1.27 (0.58 to 1.95)*	3.96 (1.30)	–6.49 to 9.02	4.15[e]	98.46% (0.00% and 98.46%)
Combined	23 (95)	59,684 (197,353)	Percent	*0.82 (0.29 to 1.35)*	3.50 (1.28)	–6.04 to 7.68	3.60[e]	98.55% (14.53% and 84.02%)
Pulse rate								
Light	3 (9)	31 (43,116)	bpm[f]	–1.24 (–5.31 to 2.83)	7.54 (2.13)	–16.02[g] to 13.54[g]	—[h]	10.99% (0.00% and 10.99%)
Medium	3 (9)	129 (90,733)	bpm	–0.89 (–3.70 to 1.93)	9.05 (1.75)	–18.62[g] to 16.84[g]	—	25.01% (0.00% and 25.01%)
Dark	1 (6)	16 (6,922)	bpm	–0.57 (–9.44 to 8.29)	16.89 (1.31)	–33.69[g] to 32·54[g]	—	13.70% (N/A[i] and 13.70%)
Combined	4 (24)	176 (140,771)	bpm	–0.29 (–3.87 to 3.29)	8.64 (1.67)	–17.23[g] to 16.65[g]	—	30.66% (26.76% and 3.90%)

[a]ρ=0.9 was used in correlated and hierarchical effect models to pool both mean bias and SD.

[b]LoA: limits of agreement.

[c]A_{rms}: accuracy root-mean-square

[d]Italicization denotes statistical significance.

[e]Exceeds U.S. Food and Drug Administration guidance for pulse oximetry.

[f]bpm: beats per minute.

[g]Exceeds American National Standards Institute standards for pulse rate.

[h]Not applicable.

[i]N/A: not available.

PR Results

Analysis using the CHE model revealed LoA values of –17.23 to 16.65 bpm and a mean bias of –0.29 (95% CI –3.87 to 3.29) bpm across all studies. Heterogeneity analysis demonstrated that 26.76% of the variation in bias (I^2_{Level3}) stemmed from between-study differences, while 73.34% (I^2_{Level2}) originated from within-study variation. Our analysis revealed 95% LoA of –16.02 to 13.54 bpm, –18.62 to 16.84 bpm, and –33.69 to 32.54 bpm for light, medium, and dark skin pigmentation groups, respectively. Mean biases were –1.24 (95% CI –5.31 to 2.83) bpm, –0.89 (95% CI –3.70 to 1.93) bpm, and –0.57 (95% CI –9.44 to 8.29) bpm for the corresponding groups (Table 4 and Figures 5-7 [41,60-62]). Detailed results are provided in Tables S4 and S5 in Multimedia Appendix 1. Subgroup analyses found no statistically significant difference between light and medium or between light and dark skin pigmentation pooled bias (*P*≥.05; Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 5. Forest plot showing pulse rate measurement bias in patients with light skin pigmentation. Squares denote study weight; center of squares denotes observed study effect size; vertical lines denote study CIs; and diamond denotes pooled effect. CHE: correlated and hierarchical effect.

Author(s) and Year	Device		Weight	Estimate [95% CI]
Bent et al. 2020	Biovotion Everion		7.17%	-4.82 [-15.75, 6.11]
Bent et al. 2020	Fitbit Charge 2		12.44%	-4.06 [-11.51, 3.39]
Bent et al. 2020	Garmin Vivosmart 3		14.72%	-3.47 [-9.94, 3.00]
Nelson et al. 2019	Fitbit Charge 2		4.28%	-3.47 [-15.56, 8.62]
Bent et al. 2020	Xiaomi Miband 3		9.40%	-2.96 [-12.11, 6.19]
Nelson et al. 2019	Apple Watch 3		3.00%	-1.80 [-16.30, 12.70]
Sanudo et al. 2019	Apple Watch (Version not reported)		20.09%	-1.50 [-3.19, 0.19]
Bent et al. 2020	Empatica E4		9.27%	-0.63 [-9.87, 8.61]
Bent et al. 2020	Apple Watch 4		19.63%	-0.38 [-5.18, 4.42]
Model-Based Estimates:				
CHE Model			100%	-1.24 [-5.31, 2.83]

Mean Bias (bpm)
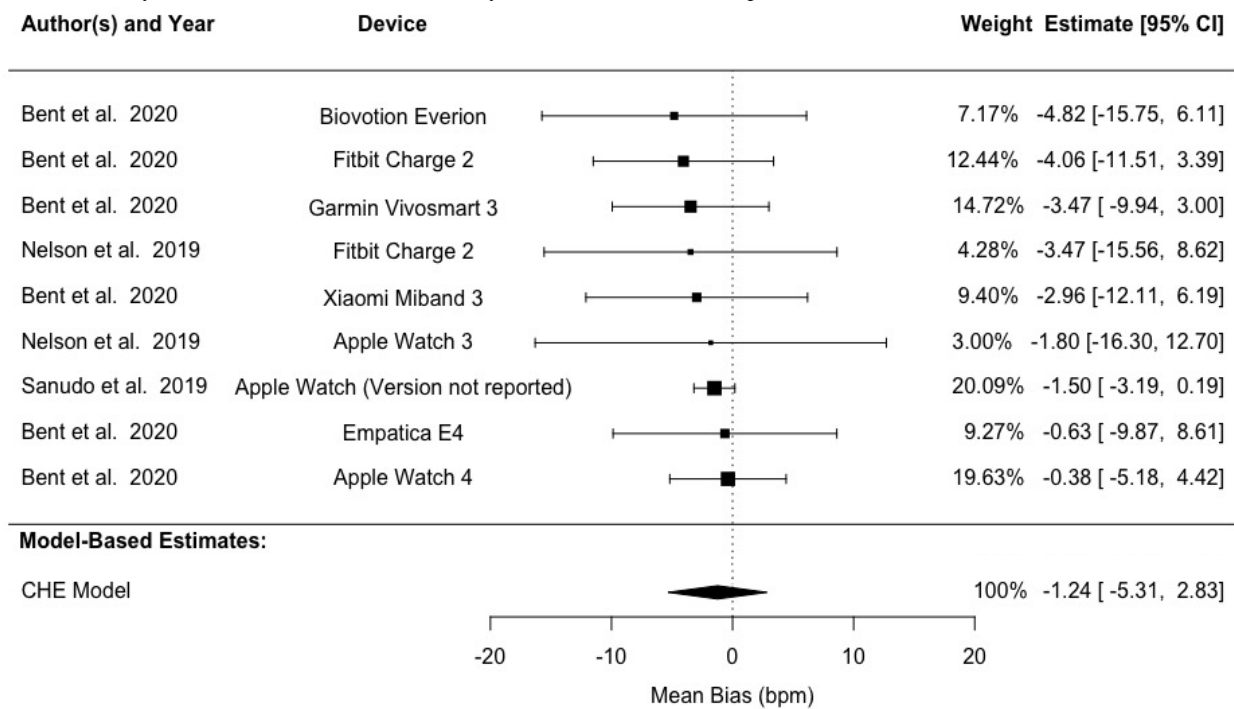
Figure 6. Forest plot showing pulse rate measurement bias in patients with medium skin pigmentation. Squares denote study weight; center of squares denotes observed study effect size; vertical lines denote study CIs; and diamond denotes pooled effect. CHE: correlated and hierarchical effect.

Author(s) and Year	Device		Weight	Estimate [95% CI]
Bent et al. 2020	Garmin Vivosmart 3		8.40%	-5.49 [-13.81, 2.83]
Bent et al. 2020	Biovotion Everion		4.92%	-5.16 [-16.95, 6.63]
Bent et al. 2020	Fitbit Charge 2		7.87%	-4.09 [-12.80, 4.62]
Bent et al. 2020	Xiaomi Miband 3		6.24%	-3.77 [-13.94, 6.40]
Bent et al. 2020	Empatica E4		5.97%	-1.53 [-11.99, 8.93]
Sanudo et al. 2019	Apple Watch (Version not reported)		21.44%	-1.50 [-2.51, -0.49]
Chow et al. 2020	Xiaomi Mi Band 2		10.75%	-1.35 [-4.66, 1.96]
Chow et al. 2020	Garmin Vivosmart HR+		17.89%	-0.55 [-2.77, 1.67]
Bent et al. 2020	Apple Watch 4		16.51%	-0.19 [-4.66, 4.28]
Model-Based Estimates:				
CHE Model			100%	-0.89 [-3.70, 1.93]

Mean Bias (bpm)
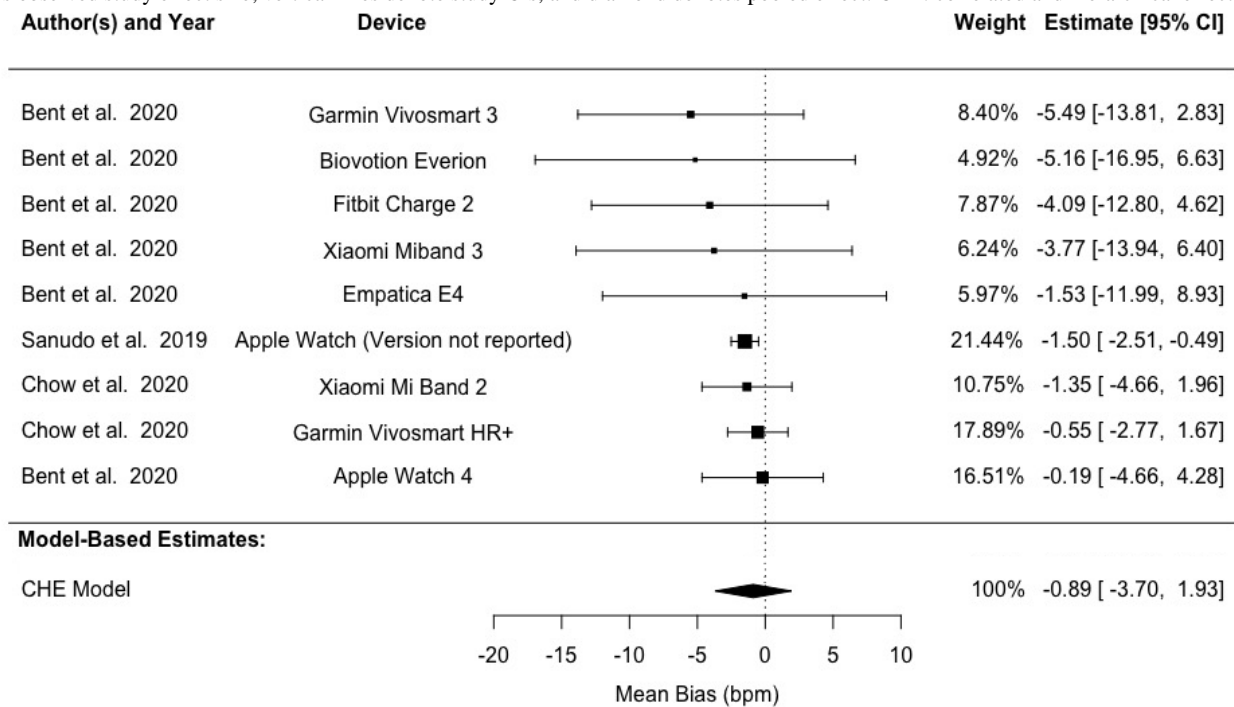
Figure 7. Forest plot showing pulse rate measurement bias in patients with dark skin pigmentation. Squares denote study weight; center of squares denotes observed study effect size; vertical lines denote study CIs; and diamond denotes pooled effect. CHE: correlated and hierarchical effect.

Author(s) and Year	Device		Weight	Estimate [95% CI]
Bent et al. 2020	Fitbit Charge 2		16.13%	-5.42 [-13.39, 2.55]
Bent et al. 2020	Biovotion Everion		8.96%	-4.87 [-17.32, 7.58]
Bent et al. 2020	Garmin Vivosmart 3		21.28%	-3.28 [-9.18, 2.62]
Bent et al. 2020	Xiaomi Miband 3		13.77%	-2.93 [-12.06, 6.20]
Bent et al. 2020	Empatica E4		14.15%	-0.98 [-9.91, 7.95]
Bent et al. 2020	Apple Watch 4		25.71%	-0.39 [-4.75, 3.97]

Model-Based Estimates:

CHE Model			100%	-0.57 [-9.44, 8.29]

Mean Bias (bpm): -20 -15 -10 -5 0 5 10
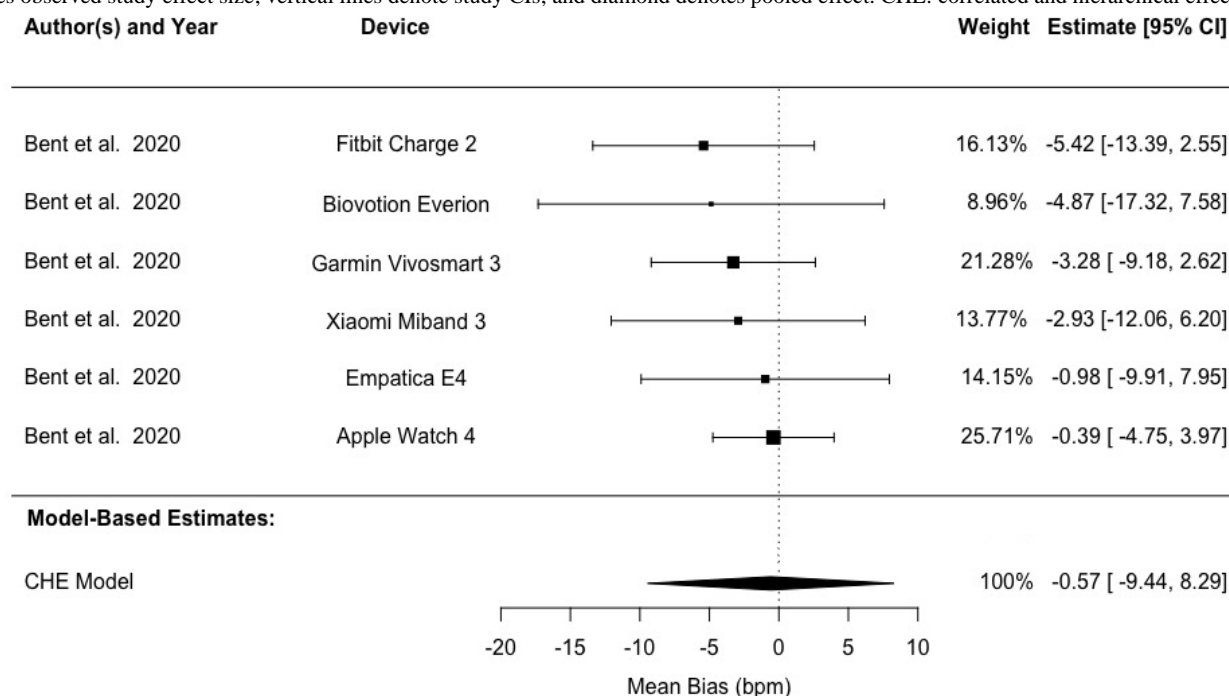
Sensitivity Analyses

Sensitivity analyses were performed for pulse oximetry SpO_2 by removing Crooks et al [48] as an outlier due to substantially large SE (Tables S6 and S7 in Multimedia Appendix 1). The pooled mean percent biases and 95% LoA were essentially unchanged, while the pooled A_{rms} across all skin pigmentation groups was 3.00%, 2.91%, and 3.40% for light, medium, and dark skin pigmentation, respectively. Sensitivity analyses using CHE models with a range of correlation coefficients (including small and moderate values) yielded similar conclusions to those obtained with a high correlation coefficient (ρ=0.9), except for the PR analysis in the light skin pigmentation group. This analysis revealed a statistically significant bias when a low correlation coefficient (ρ=0.3) was used. Details are provided in Table S6 in Multimedia Appendix 1.

Discussion

Principal Findings

The study revealed a paucity of studies properly assessing skin pigmentation, a tendency for uneven distribution across different skin pigmentation groups (overrepresentation of light skin for SpO_2 and a single study with representation of dark skin for PR), and lack of consumer-device reporting despite growing use in clinical setting. Results suggest inaccurate SpO_2 and PR measurements across all skin pigmentation groups as they breach FDA and ANSI/AAMI/IEC standards, respectively, with wearable accuracy varying considerably depending on the model, which may be due to date of model production or algorithm development. Pulse oximeters may also overestimate SpO_2 significantly for light and dark skin pigmentation, but without clinically relevant bias. We did not find statistically significant or clinically relevant bias in wearable PR devices.

Despite not meeting FDA guidance across all groups, pulse oximeter SpO_2 was inaccurate only across medium and dark skin pigmentation groups when compared to the more liberal international thresholds. Additionally, in the sensitivity analyses without the outlier study of Crooks et al [48], all pooled A_{rms} values dropped, resulting in inaccurate pulse oximeter SpO_2 only for dark skin pigmentation and no group exceeding international thresholds.

The results showing pulse oximeters significantly overestimating SpO_2 were expected for dark pigmentation and supported by findings on patient outcomes [11-15,17], but overestimated values for light pigmentation were unexpected. Two possible reasons come to mind. First, less melanin in lighter skin could distort the PPG signal [18]. Second, devices calibrated on individuals with medium skin pigmentation (note that 48% of the US population is categorized as Fitzpatrick Skin Tone Scale III [66]) may lead to inaccurate readings for both lighter and darker skin pigmentation, since both may be suboptimally represented during algorithm training and testing.

Lastly, it should be noted that both SpO_2 and PR studies had lower between-study heterogeneity and higher within-study heterogeneity, indicating that the studies that were included in these analyses were largely consistent with one another and that most of the variation came from within studies potentially due to higher variability across devices used or participants enrolled within each study.

Overall, these findings suggest that when pulse oximetry devices are deployed in their setting of intended use (ie, uncontrolled settings, such real-world medical settings and home environments with diverse patient populations), the performance observed in analytical validation studies may not generalize.

Strengths and Limitations

There were a few limitations, mostly from limitations inherent in prior studies, that should be noted. First, our skin pigmentation categorization approach has strengths but also limitations. It was an effort to overcome the reporting heterogeneity and the tendency in the published literature to conflate data collection on race, ethnicity, and skin tone. This is particularly problematic as skin tone is a physiological concept (determined by the melanin amount in the basal layer of the epidermis), while ethnicity and race are largely social constructs, with high underlying physiological heterogeneity [66]. To reduce heterogeneity, this meta-analysis reclassified race, ethnicity, and skin tone into a universal schema for skin pigmentation based on the system used by Shi et al [36]. This method, however, classifies most White people in the United States as light rather than medium skin pigmentation [66]. Second, there were only 4 prior PR studies that collected participant skin tone that also reported on device accuracy and only 1 that had participants with dark pigmentation. Third, our study used stated FDA and ANSI/AAMI/IEC standards as set thresholds to gauge device performance. It is possible that guidelines and thresholds cited in our study may change in the future, potentially limiting the applicability of the conclusions drawn in this paper. Fourth, our study chose to group papers using a variety of patient populations and testing methodologies with the goal of aggregating the largest pool of data possible on which to draw conclusions across multiple contexts. As described above, low between-study heterogeneity indicates that the studies used in this meta-analysis were largely consistent with one another. This reduces the likelihood of a moderating variable having a significant effect between studies. Instead, it suggests that most of the variation originated within studies, possibly due to higher variability across devices used or participants enrolled within each study. Despite this, we conducted subgroup analyses to examine the potential moderating variables of medical versus healthy populations for pulse oximeters (see Table S8 in Multimedia Appendix 1). We did not conduct these analyses for pulse rate because most studies did not specify whether subjects were healthy or part of medical populations. We did not examine the potential moderating effects of sensor type, transmittance versus reflectance sensors, wavelength of light used by sensors, and location of sensor placement given the core sensor technology studied was PPG across all included studies. The level of SpO_2 was not examined given this study's goal of gathering the largest sample upon which to study the effect of skin tone. Participant activity level was not evaluated as a moderator variable because many included studies did not delineate participant activity level explicitly. Future studies should evaluate the effect of these and other moderating variables on SpO_2 and PR.

Evidence Generation Guidelines for Future Analytical Validation Studies

The 2013 FDA guidance may be insufficient to ensure accuracy in pulse oximeters across all skin pigmentation and settings of intended use [19]. But there are now multiple FDA guidances for digital health tools requiring fit-for-purpose evidence as well as growing concern/guidance on clinical research diversity [4,30,31,67-73]. The FDA currently categorizes consumer devices as low-risk wellness products, exempting them from stringent regulatory oversight. However, as these devices are integrated into clinical decision-making and used as tools in clinical research [3,4], it becomes crucial to understand and communicate their advantages and limitations. Increased reliance on consumer devices increases the demand for accurate devices whose performance features and potential impact on health outcomes are known with transparency. To generate fit-for-purpose evidence applicable to diverse population, here, we propose 5 recommendations based on FDA guidance and literature (Textbox 1) [4,30,31,67-73]:

Textbox 1. Recommendations for future analytical validation studies.

Recommendation 1

It is vitally important for medical pulse oximetry devices as well as nonregulated research and consumer devices to incorporate the V3 framework [74] for sensor verification and analytical validation of derived SpO_2 and PR values for regulatory submission before these metrics can be responsibly deployed in medical, consumer, and research settings.

Recommendation 2

When possible, analytically validate devices in settings of intended use [74], rather than relying on controlled laboratory settings where digits may be warmed prior to testing, and confirm device accuracy in all subgroups (sex, race, skin pigmentation, healthy vs medical populations).

Recommendation 3

Use objective measures of skin pigmentation, rather than relying on race and ethnicity, as this will reduce heterogeneity in studies and allow for a more accurate understanding of how skin pigmentation impacts device performance.

Recommendation 4

Industry should set a priori maximum allowable difference thresholds using FDA and ANSI guidelines, properly power each subgroup, and require that 95% LoA fit within these standards for each subgroup and in the setting of intended use before receiving regulatory approval, production, and deployment.

Recommendation 5

Future studies should report device and firmware versions, as firmware updates may include changes in underlying algorithms influencing accuracy of metric generation, as described previously in the literature [4].

Conclusions

PPG has been applied in clinical practice for decades, and its accuracy in patients with different skin pigmentation has long been in question. Whether this technology contributes to diagnostic biases and by how much is only more pressing for clinicians and patients with the advent of consumer wearable PPG sensors and the growing interest and incorporation of these devices into clinical practice and in clinical research. This systematic review and meta-analysis found that pulse oximeter SpO_2 and wearable PR were inaccurate across all skin pigmentation groups as the resulting accuracy values breached FDA guidance and ANSI/AAMI/IEC standard thresholds, respectively, although pulse oximeter SpO_2 was only found to be inaccurate for dark skin pigmentation in sensitivity analyses. In addition, despite not exceeding clinically relevant bias thresholds, pulse oximeters were found to significantly overestimate SpO_2 for light and dark skin pigmentation. No systematic or clinically relevant bias was found in estimation of PR. The recommendations in this paper can help advise patients, study participants, care providers, device manufacturers, application developers, researchers, and legislators on best practices going forward.

Data Availability

Open code and data are available on Open Science Framework [75] (see the Materials—Open Code and Data section in Multimedia Appendix 1).

Authors' Contributions

SS contributed to study concept and design, data collection, and writing portions of the *Methods* section of the manuscript. HG contributed to study concept and design, data collection, and writing portions of the *Methods* and *Discussion* sections of the manuscript. BWN contributed to study concept and design; data collection; data analysis and interpretation; and writing portions of the *Abstract*, *Introduction*, *Methods*, *Results*, and *Discussion* sections of the manuscript. MRB contributed to data collection and writing portions of the *Methods* section of the manuscript. CC contributed to statistical analysis method determination, code review, and writing portions of the *Methods* section of the manuscript. All authors (SS, MRB, CC, SS, HG, and BWN) reviewed the final manuscript. HG and BWN contributed to this project as shared senior authors.

Conflicts of Interest

BWN, MRB, SS, and CC report past or current employment and/or equity ownership in Verily Life Sciences. HG serves or has served as a consultant for Verily Life Sciences, Boston Scientific, Huxley Medical, and Happitech.

Multimedia Appendix 1

Supplementary information.
[DOCX File , 619 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 89 KB-Multimedia Appendix 2]

References

1. Kyriacou PA, Allen J. Photoplethysmography: Technology, Signal Analysis and Applications. Cambridge, MA. Academic Press; 2021.
2. Knowles M, Krasniansky A, Nagappan A. Consumer adoption of digital health in 2022: moving at the speed of trust. Rock Health. Feb 21, 2023. URL: https://rockhealth.com/insights/consumer-adoption-of-digital-health-in-2022-moving-at-the-speed-of-trust/ [accessed 2023-03-01]
3. Al-Alusi MA, Khurshid S, Wang X, Venn RA, Pipilas D, Ashburner JM, et al. Trends in consumer wearable devices with cardiac sensors in a primary care cohort. Circ Cardiovasc Qual Outcomes. 2022;15(7):e008833. [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.121.008833] [Medline: 35758032]
4. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med. 2020;3:90. [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
5. Khan M, Pretty CG, Amies AC, Elliott R, Shaw GM, Chase JG. Investigating the effects of temperature on photoplethysmography. IFAC-PapersOnLine. 2015;48(20):360-365. [doi: 10.1016/j.ifacol.2015.10.166]
6. Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel). 2021;11(4):126. [FREE Full text] [doi: 10.3390/bios11040126] [Medline: 33923469]
7. Ries AL, Prewitt LM, Johnson JJ. Skin color and ear oximetry. Chest. 1989;96(2):287-290. [doi: 10.1378/chest.96.2.287] [Medline: 2752811]

8. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg. 2007;105(6 Suppl):S18-S23. [doi: 10.1213/01.ane.0000285988.35174.d9] [Medline: 18048893]

9. Zeballos RJ, Weisman IM. Reliability of noninvasive oximetry in black subjects during exercise and hypoxia. Am Rev Respir Dis. 1991;144(6):1240-1244. [doi: 10.1164/ajrccm/144.6.1240] [Medline: 1741533]

10. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology. 2005;102(4):715-719. [FREE Full text] [doi: 10.1097/00000542-200504000-00004] [Medline: 15791098]

11. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med. 2020;383(25):2477-2478. [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]

12. Valbuena VSM, Seelye S, Sjoding MW, Valley TS, Dickson RP, Gay SE, et al. Racial bias and reproducibility in pulse oximetry among medical and surgical inpatients in general care in the veterans health administration 2013-19: multicenter, retrospective cohort study. BMJ. 2022;378:e069775. [FREE Full text] [doi: 10.1136/bmj-2021-069775] [Medline: 35793817]

13. Gottlieb ER, Ziegler J, Morley K, Rush B, Celi LA. Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit. JAMA Intern Med. 2022;182(8):849-858. [FREE Full text] [doi: 10.1001/jamainternmed.2022.2587] [Medline: 35816344]

14. Okunlola OE, Lipnick MS, Batchelder PB, Bernstein M, Feiner JR, Bickler PE. Pulse oximeter performance, racial inequity, and the work ahead. Respir Care. 2022;67(2):252-257. [FREE Full text] [doi: 10.4187/respcare.09795] [Medline: 34772785]

15. Wong AI, Charpignon M, Kim H, Josef C, de Hond AAH, Fojas JJ, et al. Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality. JAMA Netw Open. 2021;4(11):e2131674. [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.31674] [Medline: 34730820]

16. Fawzy A, Wu TD, Wang K, Sands KE, Fisher AM, Arnold Egloff SA, et al. Clinical outcomes associated with overestimation of oxygen saturation by pulse oximetry in patients hospitalized with COVID-19. JAMA Netw Open. 2023;6(8):e2330856. [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.30856] [Medline: 37615985]

17. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. JAMA Intern Med. 2022;182(7):730-738. [FREE Full text] [doi: 10.1001/jamainternmed.2022.1906] [Medline: 35639368]

18. Keller MD, Harrison-Smith B, Patil C, Arefin MS. Skin colour affects the accuracy of medical oxygen sensors. Nature. 2022;610(7932):449-451. [doi: 10.1038/d41586-022-03161-1] [Medline: 36261563]

19. Pulse oximeters - premarket notification submissions [510(k)s] guidance for industry and Food and Drug Administration staff. U.S. Food and Drug Administration. Mar 04, 2013. URL: https://www.fda.gov/media/72470/download [accessed 2024-08-24]

20. Review of pulse oximeters and factors that can impact their accuracy. U.S. Food and Drug Administration. 2022. URL: https://www.fda.gov/media/162709/download [accessed 2024-08-24]

21. McFarling UL. FDA panel asks for improvements in pulse oximeters. STAT News. 2022. URL: https://www.statnews.com/2022/11/01/fda-panel-asks-for-improvements-in-pulse-oximeters/ [accessed 2023-01-09]

22. American Psychiatric Association. Wearable Devices as Therapy Tools. Monitor on Psychology. Sep 2021. URL: https://www.apa.org/monitor/2021/2021-09-monitor.pdf [accessed 2023-01-10]

23. Center for Devices and Radiological Health. Pulse oximeter accuracy and limitations: FDA safety communication. U.S. Food and Drug Administration. URL: https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication [accessed 2023-01-10]

24. November 1, 2022: anesthesiology and respiratory therapy devices panel of the medical devices advisory committee meeting announcement. U.S. Food and Drug Administration. Apr 4, 2023. URL: https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-1-2022-anesthesiology-and-respiratory-therapy-devices-panel-medical-devices-advisory [accessed 2023-10-05]

25. McFarling UL. Pulse oximeters' inaccuracies in darker-skinned people require urgent action, AGs tell FDA. STAT News. Nov 07, 2023. URL: https://www.statnews.com/2023/11/07/pulse-oximeters-attorneys-general-urge-fda-action/#:~:text=Pulse%20oximeters?%20overestimation%20of%20oxygen,for%20severe%20Covid%2D19%20infections [accessed 2023-11-07]

26. Attorney General letter to FDA. State of California Office of the Attorney General. Nov 01, 2023. URL: https://oag.ca.gov/system/files/attachments/press-docs/23PR353%20Health%20Equity%20General%20Matter%20Multistate.pdf [accessed 2023-11-01]

27. Zinzuwadia A, Singh JP. Wearable devices-addressing bias and inequity. Lancet Digit Health. 2022;4(12):e856-e857. [FREE Full text] [doi: 10.1016/S2589-7500(22)00194-7] [Medline: 36335031]

28. Shachar C, Gerke S. Prevention of bias and discrimination in clinical practice algorithms. JAMA. 2023;329(4):283-284. [doi: 10.1001/jama.2022.23867] [Medline: 36602791]

29. Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. JAMA. 2023;329(4):285-286. [doi: 10.1001/jama.2022.23870] [Medline: 36602795]

30. Colvonen P, DeYoung P, Bosompra N, Owens R. Limiting racial disparities and bias for wearable devices in health science research. Sleep. 2020;43(10):zsaa159. [FREE Full text] [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]

31. State of California Department of Justice. Attorney General Bonta launches inquiry into racial and ethnic bias in healthcare algorithms. State of California Office of the Attorney General. Aug 31, 2022. URL: https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare [accessed 2023-11-08]

32. Medical electrical equipment particular requirements for basic safety and essential performance of pulse oximeter equipment (British standard). ANSI. URL: https://tinyurl.com/ayfrdpau [accessed 2023-10-23]

33. ANSI/AAMI/IEC. IEC 60601-2-27 particular requirements for the basic safety and essential performance of electro cardiographic monitoring. ITC India. URL: https://www.itcindia.org/iec-60601-2-27-particular-requirements-for-the-basic-safety-and-essential-performance-of-electro-cardiographic-monitoring/#:~:text=What%20is%20IEC%2060601%2D2,processing%2C%20alarms%2C%20and%20displays [accessed 2024-08-24]

34. ANSI/AAMI EC13-2002 cardiac monitors, heart rate meters, and alarms. ANSI. 2002. URL: https://webstore.ansi.org/standards/aami/ansiaamiec132002 [accessed 2023-10-30]

35. Koerber D, Khan S, Shamsheri T, Kirubarajan A, Mehta S. Accuracy of heart rate measurement with wrist-worn wearable devices in various skin tones: a systematic review. J Racial Ethn Health Disparities. 2023;10(6):2676-2684. [FREE Full text] [doi: 10.1007/s40615-022-01446-9] [Medline: 36376641]

36. Shi C, Goodall M, Dumville J, Hill J, Norman G, Hamer O, et al. The accuracy of pulse oximetry in measuring oxygen saturation by levels of skin pigmentation: a systematic review and meta-analysis. BMC Med. 2022;20(1):267. [FREE Full text] [doi: 10.1186/s12916-022-02452-8] [Medline: 35971142]

37. Cabanas AM, Fuentes-Guajardo M, Latorre K, León D, Martín-Escudero P. Skin pigmentation influence on pulse oximetry accuracy: a systematic review and bibliometric analysis. Sensors (Basel). 2022;22(9):3402. [FREE Full text] [doi: 10.3390/s22093402] [Medline: 35591092]

38. Al-Halawani R, Charlton PH, Qassem M, Kyriacou PA. A review of the effect of skin pigmentation on pulse oximeter accuracy. Physiol Meas. 2023;44(5):05TR01. [FREE Full text] [doi: 10.1088/1361-6579/acd51a] [Medline: 37172609]

39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]

40. Center for Devices and Radiological Health. How to determine if your product is a medical device. U.S. Food and Drug Administration. Sep 29, 2022. URL: https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device [accessed 2023-11-16]

41. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med. 2020;3:18. [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]

42. Abrams G, Sanders MK, Fallon MB. Utility of pulse oximetry in the detection of arterial hypoxemia in liver transplant candidates. Liver Transpl. 2002;8(4):391-396. [FREE Full text] [doi: 10.1053/jlts.2002.32252] [Medline: 11965585]

43. Adler JN, Hughes LA, Vivilecchia R, Camargo CA. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. Acad Emerg Med. 1998;5(10):965-970. [FREE Full text] [doi: 10.1111/j.1553-2712.1998.tb02772.x] [Medline: 9862586]

44. Andrist E, Nuppnau M, Barbaro RP, Valley TS, Sjoding MW. Association of race with pulse oximetry accuracy in hospitalized children. JAMA Netw Open. 2022;5(3):e224584. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.4584] [Medline: 35357460]

45. Barker SJ, Wilson WC. Racial effects on Masimo pulse oximetry: a laboratory study. J Clin Monit Comput. 2023;37(2):567-574. [FREE Full text] [doi: 10.1007/s10877-022-00927-w] [Medline: 36370242]

46. Bothma PA, Joynt GM, Lipman J, Hon H, Mathala B, Scribante J, et al. Accuracy of pulse oximetry in pigmented patients. S Afr Med J. 1996;86(5 Suppl):594-596. [Medline: 8914569]

47. Burnett G, Stannard B, Wax D, Lin H-M, Pyram-Vincent C, DeMaria S, et al. Self-reported race/ethnicity and intraoperative occult hypoxemia: a retrospective cohort study. Anesthesiology. 2022;136(5):688-696. [FREE Full text] [doi: 10.1097/ALN.0000000000004153] [Medline: 35231085]

48. Crooks CJ, West J, Morling JR, Simmonds M, Juurlink I, Briggs S, et al. Pulse oximeter measurements vary across ethnic groups: an observational study in patients with COVID-19. Eur Respir J. 2022;59(4):2103246. [FREE Full text] [doi: 10.1183/13993003.03246-2021] [Medline: 35086839]

49. Ebmeier SJ, Barker M, Bacon M, Beasley RC, Bellomo R, Knee Chong C, et al. A two centre observational study of simultaneous pulse oximetry and arterial oxygen saturation recordings in intensive care unit patients. Anaesth Intensive Care. 2018;46(3):297-303. [doi: 10.1177/0310057X1804600307] [Medline: 29716488]

50. Foglia EE, Whyte RK, Chaudhary A, Mott A, Chen J, Propert KJ, et al. The effect of skin pigmentation on the accuracy of pulse oximetry in infants with hypoxemia. J Pediatr. 2017;182:375-377.e2. [FREE Full text] [doi: 10.1016/j.jpeds.2016.11.043] [Medline: 27939107]

51. Jubran A, Tobin MJ. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. Chest. 1990;97(6):1420-1425. [doi: 10.1378/chest.97.6.1420] [Medline: 2347228]

52. McGovern JP, Sasse SA, Stansbury DW, Causing LA, Light RW. Comparison of oxygen saturation by pulse oximetry and co-oximetry during exercise testing in patients with COPD. Chest. 1996;109(5):1151-1155. [doi: 10.1378/chest.109.5.1151] [Medline: 8625659]

53. Muñoz X, Torres F, Sampol G, Rios J, Martí S, Escrich E. Accuracy and reliability of pulse oximetry at different arterial carbon dioxide pressure levels. Eur Respir J. 2008;32(4):1053-1059. [FREE Full text] [doi: 10.1183/09031936.00126507] [Medline: 18480106]

54. Pilcher J, Ploen L, McKinstry S, Bardsley G, Chien J, Howard L, et al. A multicentre prospective observational study comparing arterial blood gas values to those obtained by pulse oximeters used in adult patients attending Australian and New Zealand hospitals. BMC Pulm Med. 2020;20(1):7. [FREE Full text] [doi: 10.1186/s12890-019-1007-3] [Medline: 31918697]

55. Ruppel H, Makeneni S, Faerber JA, Lane-Fall MB, Foglia EE, O'Byrne ML, et al. Evaluating the accuracy of pulse oximetry in children according to race. JAMA Pediatr. 2023;177(5):540-543. [FREE Full text] [doi: 10.1001/jamapediatrics.2023.0071] [Medline: 36939727]

56. Sudat SEK, Wesson P, Rhoads K, Brown S, Aboelata N, Pressman AR, et al. Racial disparities in pulse oximeter device inaccuracy and estimated clinical impact on COVID-19 treatment course. Am J Epidemiol. 2023;192(5):703-713. [FREE Full text] [doi: 10.1093/aje/kwac164] [Medline: 36173743]

57. Thrush D, Hodges MR. Accuracy of pulse oximetry during hypoxemia. South Med J. 1994;87(4):518-521. [doi: 10.1097/00007611-199404000-00019] [Medline: 8153783]

58. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. J Perinatol. 2022;42(1):79-85. [FREE Full text] [doi: 10.1038/s41372-021-01230-3] [Medline: 34642469]

59. Wiles MD, El-Nayal A, Elton G, Malaj M, Winterbottom J, Gillies C, et al. The effect of patient ethnicity on the accuracy of peripheral pulse oximetry in patients with COVID-19 pneumonitis: a single-centre, retrospective analysis. Anaesthesia. 2022;77(2):143-152. [FREE Full text] [doi: 10.1111/anae.15581] [Medline: 34542168]

60. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth. 2019;7(3):e10828. [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]

61. Sañudo B, de Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the apple watch. J Med Syst. 2019;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]

62. Chow H, Yang C. Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. JMIR Mhealth Uhealth. 2020;8(4):e14707. [FREE Full text] [doi: 10.2196/14707] [Medline: 32343255]

63. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft. 2010;36(3):1-48.

64. Pustejovsky JE, Tipton E. Meta-analysis with robust variance estimation: expanding the range of working models. Prev Sci. 2022;23(3):425-438. [doi: 10.1007/s11121-021-01246-3] [Medline: 33961175]

65. Pustejovsky J. clubSandwich: cluster-robust (sandwich) variance estimators with small-sample corrections. R Project. URL: https://CRAN.R-project.org/package=clubSandwich [accessed 2024-05-29]

66. Keiser E, Linos E, Kanzler M, Lee W, Sainani KL, Tang JY. Reliability and prevalence of digital image skin types in the United States: results from national health and nutrition examination survey 2003-2004. J Am Acad Dermatol. 2012;66(1):163-165. [doi: 10.1016/j.jaad.2011.02.044] [Medline: 22177642]

67. Sjoding MW, Iwashyna TJ, Valley TS. Change the framework for pulse oximeter regulation to ensure clinicians can give patients the oxygen they need. Am J Respir Crit Care Med. 2023;207(6):661-664. [FREE Full text] [doi: 10.1164/rccm.202209-1773ED] [Medline: 36260769]

68. National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Committee on Women in Science, Engineering, and Medicine, Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research. Bibbins-Domingo K, Helman A, editors. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington, DC. National Academies Press; Jul 13, 2022.

69. Lee NT, Resnick P, Barton G. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. May 22, 2019. URL: https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ [accessed 2023-11-08]

70. Framework for the use of digital health technologies in drug and biological product development. U.S. Food and Drug Administration. Mar 2023. URL: https://www.fda.gov/media/166396/download?attachment [accessed 2024-08-24]

71. Center for Drug Evaluation and Research. Digital health technologies for remote data acquisition in clinical investigations. U.S. Food and Drug Administration. Dec 2023. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations [accessed 2023-12-05]

72. Office of the Commissioner. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials; draft guidance for industry. U.S. Food and Drug Administration. URL: https://www.fda.gov/

regulatory-information/search-fda-guidance-documents/
diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations [accessed 2023-12-05]

73. Center for Drug Evaluation and Research. Enhancing the diversity of clinical trial populations — eligibility criteria, enrollment practices, and trial designs guidance for industry. U.S. Food and Drug Administration. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial [accessed 2023-11-30]

74. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med. 2020;3:55. [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]

75. Singh S, Bennett M, Chen C, Ghanbari H, Nelson B. Photoplethysmography pulse oximetry and pulse rate accuracy by skin tone: a meta-analysis. Open Science Framework. URL: http://osf.io/qngmz/ [accessed 2024-09-30]

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
ANSI: American National Standards Institute
A_{rms}: accuracy root-mean-square
bpm: beats per minute
CHE: correlated and hierarchical effect
ECG: electrocardiography
FDA: Food and Drug Administration
IEC: International Electrotechnical Commission
LoA: limits of agreement
PPG: photoplethysmography
PR: pulse rate
PRIMSA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RVE: robust variance estimation
SaO_2: arterial blood gas
SpO_2: blood oxygenation

XSL•FO
RenderX