Impact of Skin Pigmentation on Pulse Oximetry SpO2 and Wearable Pulse Rate Accuracy: A Meta-Analysis

Sanidhya Singh, Miles Bennett, Chen Chen, Sooyoon Shin, Hamid Ghanbari, Benjamin W. Nelson

Submitted to: Journal of Medical Internet Research
on: June 04, 2024

Table of Contents

Impact of Skin Pigmentation on Pulse Oximetry SpO2 and Wearable Pulse Rate Accuracy: A Meta-Analysis

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

[1]University of Michigan Medical School Ann Arbor US
[2]Verily Life Sciences LLC South San Francisco US
[3]Division of Digital Psychiatry Department of Psychiatry Harvard Medical School and Beth Israel Deaconess Medical Center Boston US

Corresponding Author:
Sanidhya Singh BSE
University of Michigan Medical School
1301 Catherine Street
Ann Arbor
US

Abstract

Background: Photoplethysmography (PPG) is a technology routinely used in clinical practice to assess blood oxygenation (SpO2) and pulse rate (PR). Skin pigmentation may influence accuracy, leading to health outcomes disparities.

Objective: This meta-analysis primarily aimed to evaluate the accuracy of PPG-derived SpO2 and PR by skin pigmentation. Secondarily, we aimed to evaluate statistical biases and the clinical relevance of PPG-derived SpO2 and PR according to skin pigmentation.

Methods: We identified 23 pulse oximetry studies (N=59,684; 197,353 paired SpO2-arterial blood observations) and 4 wearable PR studies (N=176; 140,771 paired photoplethysmography-electrocardiography observations). We evaluated accuracy according to skin pigmentation group by comparing SpO2 accuracy root-mean-square (Arms) values to the regulatory threshold of 3% and PR 95% limits of agreement (LoA) values to ±5 bpm, according to the standards of the American National Standards Institute, Advancing Safety in Medical Technology, and the International Electrotechnical Commission. We evaluated biases and clinical relevance using mean bias and 95% confidence intervals (CI).

Results: For SpO2, Arms were 3.96%, 4.71%, and 4.15% and pooled mean biases were 0.70% (95% CI: 0.17 to 1.22), 0.27% (95% CI: -0.64 to 1.19), and 1.27% (95% CI: 0.58 to 1.95) for light, medium, and dark pigmentation, respectively. For PR, 95% LoAs were -16.02 to 13.54, -18.62 to 16.84, and -33.69 to 32.54 and pooled mean biases were -1.24 bpm (95% CI: -5.31-2.83), -0.89 bpm (95% CI: -3.70-1.93), and -0.57 bpm (95% CI: -9.44-8.29) for light, medium, and dark pigmentation, respectively.

Conclusions: SpO2 and PR measurements may be inaccurate across all skin pigmentation groups, breaching FDA guidance and industry standards thresholds. Pulse oximeters significantly overestimate SpO2 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.

(JMIR Preprints 04/06/2024:62769)
DOI: https://doi.org/10.2196/preprints.62769

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Impact of Skin Pigmentation on Pulse Oximetry SpO2 and Wearable Pulse Rate Accuracy: A Meta-Analysis

Sanidhya Singh, B.S.E.[a], Miles Bennett, M.S.[b], Chen Chen, Ph.D[b], Sooyoon Shin, Ph.D.[b], Hamid Ghanbari, M.D.[a*], Benjamin W. Nelson, Ph.D.[b, c*]

[a]University of Michigan Medical School, Ann arbor, MI, USA [b]Verily Life Sciences LLC, South San Francisco, CA, USA. [c]Division of Digital Psychiatry, Department of Psychiatry, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA, USA.
* shared senior authorship

Corresponding Author:
Sanidhya Singh
University of Michigan Medical School
1301 Catherine Street
Ann Arbor, MI 48109
Phone: (734) 936-4000
Email: sansin@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 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 photoplethysmography-electrocardiography observations). We evaluated accuracy according to skin pigmentation group by comparing SpO_2 accuracy root-mean-square (A_{rms}) values to the regulatory threshold of 3% and PR 95% limits of agreement (LoA) values to ±5 bpm, according to the standards of the American National Standards Institute, Advancing Safety in Medical Technology, and the International Electrotechnical Commission. We evaluated biases and clinical relevance using mean bias and 95% confidence intervals (CI).
Results: For SpO_2, A_{rms} were 3.96%, 4.71%, and 4.15% and pooled mean biases were 0.70% (95% CI: 0.17 to 1.22), 0.27% (95% CI: -0.64 to 1.19), and 1.27% (95% CI: 0.58 to 1.95) for light, medium, and dark pigmentation, respectively. For PR, 95% LoAs were -16.02 to 13.54, -18.62 to 16.84, and -33.69 to 32.54 and pooled mean biases were -1.24 bpm (95% CI: -5.31-2.83), -0.89 bpm (95% CI: -3.70-1.93), and -0.57 bpm (95% CI: -9.44-8.29) for light, medium, and dark pigmentation, respectively.
Conclusions: SpO_2 and PR measurements may be inaccurate across all skin pigmentation groups, breaching FDA guidance and industry standards 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.

Keywords: health disparities, skin pigmentation; pulse oximetry; photoplethysmography

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 are 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 healthcare settings and as source of research endpoints [3,4].

Darker skin pigmentation may influence PPG-based SpO_2 and PR readings [5-8]. Patients with darker skin pigmentation are more likely than Caucasians to have overestimated SpO_2 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 [9-16].

Starting in 2013, a series of United States Food and Drug Administration (FDA) guidances, safety communications, and most recently attorney general statements have called to address darker skin pigmentation bias in pulse oximeters [17], yet these guidances lacked standards for assessing skin tone [18]. Recent pulse oximeter research has rekindled interest in skin pigmentation disparities in photoplethysmography (PPG) sensor accuracy, leading to increased media and regulatory attention [19-22]. In November 2023, over 24 attorney generals wrote a letter calling on the FDA to take urgent action to address pulse oximeter skin pigmentation disparities [23,24].

More recent research has also highlighted the potential presence of these biases in research-grade and consumer wearable PPG-based wearable devices [4,25-27], leading to calls to address inequity, bias, and discrimination in wearable health technology and clinical practice algorithms [28,29].

Accuracy guidance for pulse oximeters and heart rate/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 (standard deviation of bias) into a single metric [30]. FDA guidance has a threshold of $A_{rms} \leq 3\%$ [17] for transmittance devices and a threshold of $A_{rms} \leq 3.5\%$ [17] for ear clip and reflectance devices, while international thresholds are set at $A_{rms} \leq 4\%$ [30]. For heart rate/PR devices, overall accuracy can be assessed with 95% limits of agreement (LoAs). The American National Standards Institute (ANSI), Advancing Safety in Medical Technology (AAMI), and International Electrotechnical Commission (IEC) have set the recommendation that electrocardiography (ECG) devices have mean bias of ± 5 bpm or MAPE ≤ 10%, whichever is greater [31,32]. 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 in pulse oximeters limited to either consumer device PR [33] or pulse oximeter SpO_2 [34-36]. Only one meta-analysis focused on pulse oximetry for this topic is available [34], and an additional 7 studies have been published recently, providing an additional 50,980 subjects 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:

This meta-analysis provides open code and data (Materials–Open Code and Data in Multimedia Appendix 1).

Search Strategy and Selection Criteria:

We followed the PRISMA guidelines [37] and developed a MEDLINE systematic-review 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 [38]; 2) used arterial blood gas (SaO_2) or ECG as the reference device; 3) reported mean bias and standard deviation (SD), standard error (SE), 95% Limits of Agreement (LoA), or 95% confidence intervals (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) unretrievable full source texts; 4) remote photoplethysmography (rPPG); and 5) lack of SaO_2/ECG as 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 endpoints 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 (MB, HG) adjudicated any differences between the initial two reviewers and resolved any disagreements by checking manuscripts. First author verified the quality of manuscripts. Standardized data extraction was developed to extract the study characteristics (see Materials–Data Extraction in Multimedia Appendix 1). For one paper, we used participant-level data from supplementary materials to calculate mean bias and SE by skin pigmentation category [39]. This resulted in 27 studies in our final analysis (Figure 1, Table 1).

Figure 1. Study Selection Flow Chart.

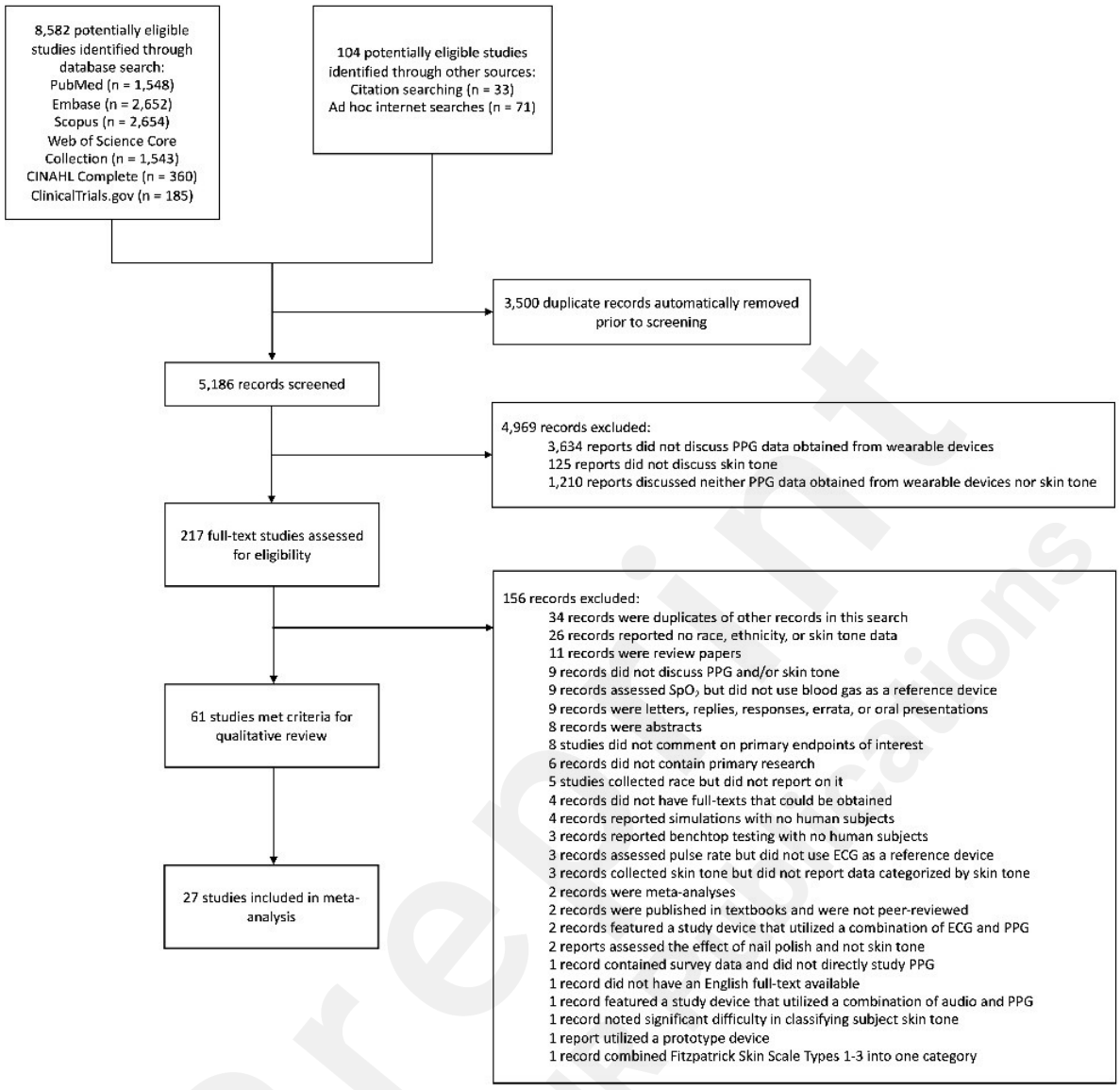


Table 1. Characteristics of Included Studies.

Study (Author, Year)	Sample Size	Test Device Evaluated	Reference Device	SaO_2 (%) Range	Participant Population	Research Setting	Skin Pigmentation Method
Pulse Oximetry							
Abrams 2003 [41]	200	Nellcor N-200	Radiometer ABL-520	Not reported	Medical- Adult patients with cirrhosis	Inpatient	Race- Black and White
Adler 1998 [42]	298	Nellcor D-25	4-wavelength spectro-photometer, or co-oximeter (Radiometer	50 to 99	Medical- Adult emergency room patients	Inpatient	Skin Tone- Munsell color system categorized into light, medium, or dark

			OSM3)				
Andrist 2022 [43]	1,061	Not Reported	Not Reported	Not Reported	Medical- Pediatric	Inpatient	Race
Barker 2023 [44]	75	Masimo SET® pulse oximeters with RD-SET® sensors	Radiometer ABL-835 Flex CO-Oximeter	70 to 100	Mixed- Healthy and mild systemic disease	Research-Lab	Race
Bickler 2005 [8]	23	-Nellcor N-595 with Nellcor OxiMax A finger probe -Novametrix 513s models (two types) -Nonin Onyx models (two types)	Radiometer OSM3	60 to 100	Healthy- non-smoking	Research-Lab	Ethnicity- light (Northern European) and dark (African American)
Bothma 1996 [45]	100	Simed S100e Nihon Koden Ohmeda 3740	IL482 Co-oximeter System	87·8 to 99·2	Medical- critically ill adult patients	Inpatient	Skin Tone- EEL reflectance spectrophotometer. All participants had dark skin tone
Burnett 2022 [46]	46,253	Unspecified Nellcor and Masimo devices	GEMStat Premier 3000	Not Reported	Medical- patients receiving anesthesia	Inpatient	Race and Ethnicity
Crooks 2022 [47]	2,997	Not reported	Not reported	Not reported	Medical- COVID-19	Inpatient	Race
Ebmeier 2018 [48]	394	Marquette Rac-4A monitors with Masimo sensors and Philips IntelliVue MP70 monitors with Philips Adult Reusable SpO_2 sensors	Radiometer ABL 800 FLEX arterial blood gas analyser	Not reported	Medical- multiple conditions	Inpatient	Ethnicity
Fawzy 2022 [15]	1,216	Not reported	ABL825, ABL827, or ABL90 blood gas analyzers	Not reported	Medical- COVID-19	Inpatient	Race and Ethnicity
Feiner 2007 [6]	36	Nellcor N-595 (OxiMax A adhesive probe) Nellcor N-595 (a clip-type probe) Masimo Radical (clip probe)	Radiometer OSM3 multiwavelength oximeter	60 to 100	Healthy- non-smoking	Research-Lab	Ethnicity- light (Caucasian), intermediate (Hispanic, Indian, Filipino, Vietnamese), and dark (African American) categories

		Masimo Radical (adhesive disposable probe) Nonin 9700 (clip-type probe) Nonin 9700 (disposable adhesive probe)					
Foglia 2017 [49]	36	Nellcor Oximax (Covidien) Masimo Rainbow SET Radical 7	Siemens Rapidlab 1265	60 to 92	Medical- Infants with cyanotic congenital heart disease and oxygen saturation <90%	Inpatient	Skin Tone- Munsell Color System
Jubran 1990 [50]	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 1996 [51]	8	Ohmeda 3700	IL 482 Co-oximeter	Not reported	Medical- Adults with stable condition with severe COPD	Research-Lab	Race- All White
Muñoz 2008 [52]	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 2020 [53]	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 to 100	Medical- Hospitalized adult patients	Inpatient and Outpatient	Skin Tone- Fitzpatrick Scale categorized into light, medium, dark

Ruppel 2023 [54]	774	Not reported	Not reported	Not reported	Medical- cardiac catheterization	Inpatient	Race
Sudat 2023 [55]	8,735	Not reported	Not reported	Not reported	Medical- multiple conditions	Inpatient	Race
Thrush 1994 [56]	25	Critikon Dinamap Plus Model 8700 Critikon Oxyshuttle Ohmeda 3700 Catalyst Research MiniOx IV	IL482 co-oximeter	80 to 100	Healthy- non-smoking adults	Research-Lab	Race- All White
Valbuena 2022 [10]	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 2021 [57]	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 2022 [58]	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 1991 [7]	33	Hewlett-Packard HP-47201A Ohmeda Biox IIA	IL282 co-oximeter	Not reported	Healthy- non-smoking volunteers	Research-Lab	Race- All Black
Pulse Rate							
Bent et al. (2020) [39]	53	Empatica E4 Apple Watch Fitbit Charge Garmin Vivosmart 3 Xiaomi Miband Biovotion	BittiumFaros 180, Bittium Inc.	NA	Healthy	Research-Lab	Skin Tone- Fitzpatrick Scale 1 (n = 7) 2 (n = 8) 3 (n = 10) 4 (n = 9) 5 (n = 9) 6 (n = 10)

		Everion					
Nelson et al. (2019) [59]	1	Apple Watch 3 Fitbit Charge 2	Vrije Universiteit Ambulatory Monitoring System	NA	Healthy	Research-Real World	Skin Tone- Fitzpatrick Scale 1 (n = 1)
Sanudo 2019 [60]	45	Apple Watch (version not reported)	Polar Chest Strap	NA	Healthy	Research-Lab	Skin Tone- Fitzpatrick Scale 2 (n = 15) 3 (n = 15) 4 (n = 15)
Chow 2020 [61]	40	Garmin Vivosmart HR+ Xiaomi Mi Band 2	Polar H7 Chest Strap	NA	Healthy	Research-Lab	Ethnicity and Skin Tone- East Asian, Fitzpatrick Scale 3 and 4

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 [40] (Figures S2 and S3 in Multimedia Appendix 1).

Statistical Analysis

Skin Tone Categorization

We mapped skin tone, race, and/or ethnicity into three primary skin pigmentation groups of light, medium, and dark, following published methodology [34], (see Table S7 in Multimedia Appendix 1) and examined biases for SpO_2 and PR by each skin pigmentation group (described below).

We elected to utilize this same skin pigmentation categorization schema as has previously been used with the goal of expanding upon the analysis of Shi et al. [34].

Statistical Analysis Methods

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 employed in prior meta-analyses within the discipline [34].

Evaluation endpoints and criteria

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

Accuracy: To evaluate accuracy of SpO_2, root mean square (A_{rms}) was used as an additional measure of accuracy for SpO_2 because of its commonality in the regulatory space [17]. It was calculated as $\sqrt{bias^2 + precision^2}$, where precision is the standard deviation of bias [41,42]. 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 measurements are closer to the true values whereas a higher A_{rms} value suggests lower accuracy, indicating that the measurements are further away from the true values.

For SpO_2, our study used the more strict A_{rms} >3% threshold to define inaccuracy as per FDA guidance [17]. We used this threshold as 22/23 (95.65%) 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% limits of agreement (LoA) of bias was used, and LoA was constructed as mean bias ± 1.96 * standard deviation (SD). For PR, a pooled 95% LoA not bounded by ± 5 bpm was considered inaccurate per ANSI/AAMI/IEC standards [31,32].

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% 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 [17]. 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 bpm it was considered clinically relevant as per ANSI/AAMI/IEC standards [31,32].

Table 2. Evaluation Endpoints and Criteria.[a]

Category	Objective	Evaluation Endpoint	Evaluation Criteria
SpO_2	Evaluate accuracy	A_{rms}	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 < -4% or lower bound of 95% CI > 4%
Pulse Rate	Evaluate accuracy	95% LoA	Accurate if 95% LoA is bounded by ± 5 bpm
	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, LoA = limits of agreement, CI = confidence interval.

Statistical meta-analysis methods

To obtain pooled results for above listed endpoints, we collected sample size, paired observations, mean bias, and standard deviation 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 effects (CHE) models were used to pool mean biases from various studies. Specifically, three-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 being compared using the same participants data, and these effect sizes are dependent. To account for the dependence, the CHE model utilized 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 [62]. 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 [34], and small ($\rho = 0.30$) and moderate ($\rho = 0.60$) correlation coefficients were also explored in sensitivity analyses. Additionally, one report [47] had substantially larger SE and we conducted a separate analysis with their results removed as a part of the sensitivity analysis. To evaluate 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. Forest plots were provided for visualization of 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, and we followed an analytical approach in prior meta-analyses within the discipline [34]. Specifically, SD of bias across studies were pooled using CHE models similar to those used to pool mean biases. And pooled mean bias and pooled SD were used to provide pooled estimates of 95% LoA and A_{rms}. Specifically,

- Overall accuracy $A_{rms} = \sqrt{the\ pooled\ mean\ bias^2 + the\ pooled\ SD^2}$;
- Overall 95% LoA = pooled mean bias ±1.96 * pooled SD.

Other methods

Descriptive statistics were provided for study characteristics, the device intended use, mean bias based on different skin tones, reference devices. All analyses were conducted using R (4.3.1). Statistical models were based on R packages *metafor* [40] and *clubSandwich* [63].

Results

Study Selection and Characteristics

Search strategy resulted in 8,582 records. We selected 27 studies for full review (pulse oximetry = 23, PR = 4; Figure 1, Table 1, 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, four PR studies with 176 participants and 140,771 paired PR-ECG observations were analyzed (Table 3).

Table 3. Summary Characteristics of Included Studies.[b]

Item	Sub Item	Pulse Oximetry N (Percent)	Pulse Rate N (Percent)
Participants/ participants		59,684	176
	Light	40,416 (67·72%)	31 (17·61%)
	Medium	9,967 (16·70%)	129 (73·30%)
	Dark	9,301 (15·58%)	16 (9·09%)
Paired Observations		197,353	140,771
	Light	131,008 (66·38%)	43,116 (30·60%)
	Medium	32,095 (16·26%)	90,733 (64·50%)
	Dark	34,250 (17·35%)	6,922 (4·92%)
Device Type Included in Study	Medical	23 (100%)	0 (0%)
	Non-Regulated	0 (0%)	4 (100%)*
Sensor Type	Transmittance	22 (95.65%)	0 (0%)
	Reflectance	1 (4.35%)	4 (100%)
Patient Population	Healthy	4 (17·39%)	3 (75%)
	Medical	18 (78·26%)	0 (0%)
	Healthy and Medical	1 (4·35%)	0 (0%)
	Not Reported	0 (0%)	1 (25%)
Research Setting	Medical Inpatient	15 (65·22%)	0 (0%)
	Medical Outpatient	1 (4·35%)	0 (0%)
	Medical Combined	1 (5·35%)	0 (0%)
	Research Lab	6 (26·09%)	3 (75%)
	Research Real World	0 (0%)	1 (25%)
Skin Pigmentation Method	Skin Tone	4 (17·39%)	4 (100%)
	Race	13 (56·52%)	0 (0%)
	Ethnicity	3 (13·04%)	0 (0%)
	Race and Ethnicity	3 (13·04%)	0 (0%)

[b] * = Medical device defined as devices that received regulatory clearance for either SpO_2 or PR.

Descriptive Statistics on Race, Ethnicity, and Skin Tone

Pulse Oximetry. Skin pigmentation was classified by race in 13 (56.52%) studies, ethnicity in 3 (13.04%) studies, skin tone in 4 (17.39%) studies, and both race and ethnicity in 3 (13.04%) studies (Table 3). There were a total of 40,416 (67.72%) light skin pigmentation patients with 131,008 (66.38%) paired observations, 9,967 (16.70%) medium skin pigmentation patients with 32,095 (16.26%) paired observations, and 9,301 (15.58%) dark skin pigmentation patients with 34,250 (17.35%) paired observations.

PR. Skin pigmentation was classified by race or ethnicity in 0 (0.00%) studies, and skin tone in 4 (100%) studies (Table 3). There were a total of 31 (17.61%) light skin pigmentation participants with 43,116 (30.60%) paired observations, 129 (73.30%) medium skin pigmentation participants with 90,733 (64.50%) paired observations, and 16 (9.09%) dark skin pigmentation participants with 6,922 (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%, and 3.96%, 4.71%, and 4.15% for light, medium, and dark skin pigmentation, respectively (Table 4 and Figures 2-4). Of note, studies implementing multiple trial conditions or utilizing 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% to 1.35%) across all skin pigmentation groups using the CHE model. Between-study heterogeneity($I^2_{Level\,3}$) accounted for 14.53% of the total variation, while within-study heterogeneity ($I^2_{Level\,2}$) 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% to 1.22%) for light skin, 0.27% (95% CI: -0.64% to 1.19%) for medium skin, and 1.27% (95% CI: 0.58% to 1.95%) for dark skin. (Tables S1 and S3 in Multimedia Appendix 1).

Figure 2. Pulse oximetry forest plot for light skin pigmentation with multiple entries from the same study representing different devices examined [6, 8, 10, 15, 42-44, 46-58].

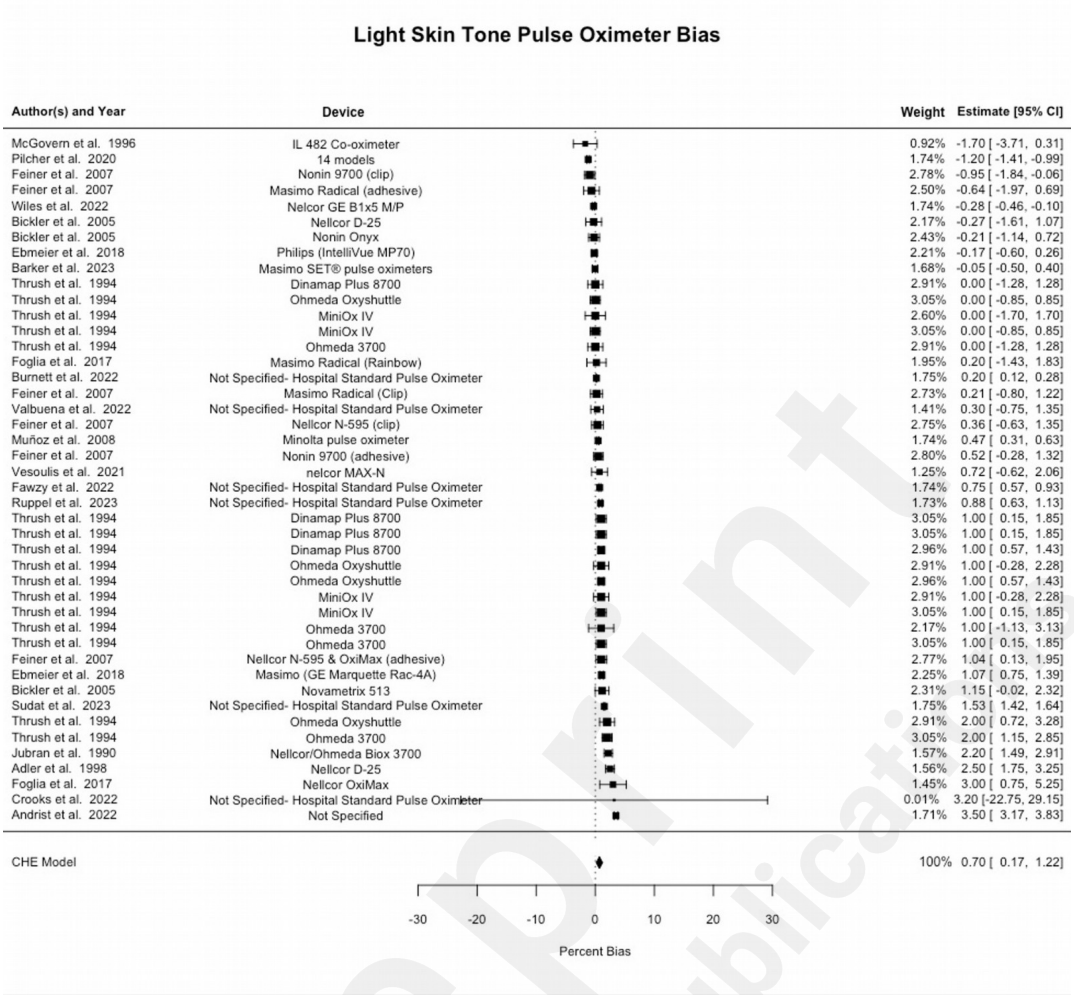


Figure 3. Pulse oximetry forest plot for medium skin pigmentation with multiple entries from the same study representing different devices examined [6, 10, 15, 42, 46-48, 53, 58].

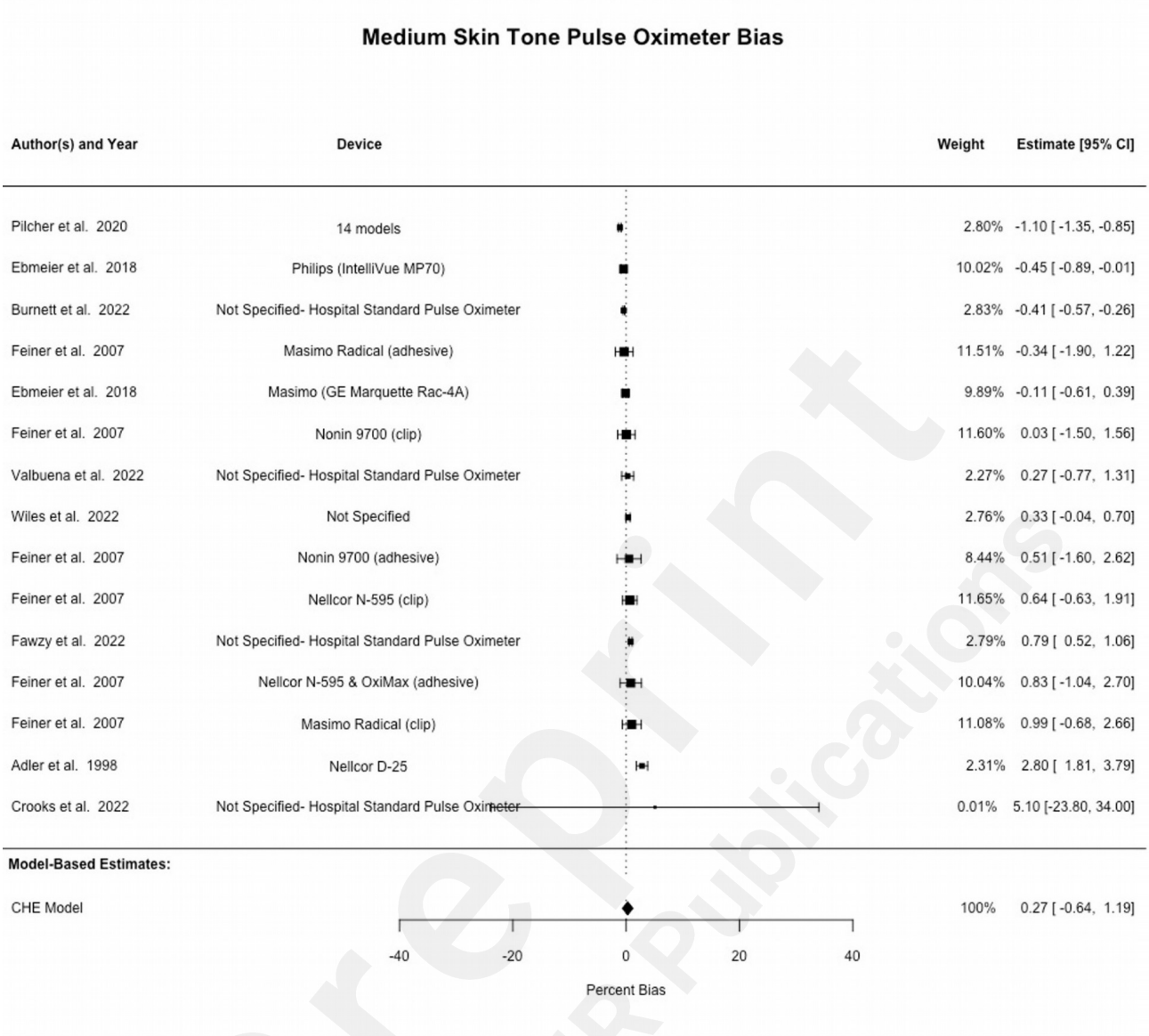


Figure 4. Pulse oximetry forest plot for dark skin pigmentation with multiple entries from the same study representing different devices examined [6-8, 10, 15, 41-50, 54, 55, 57, 58].

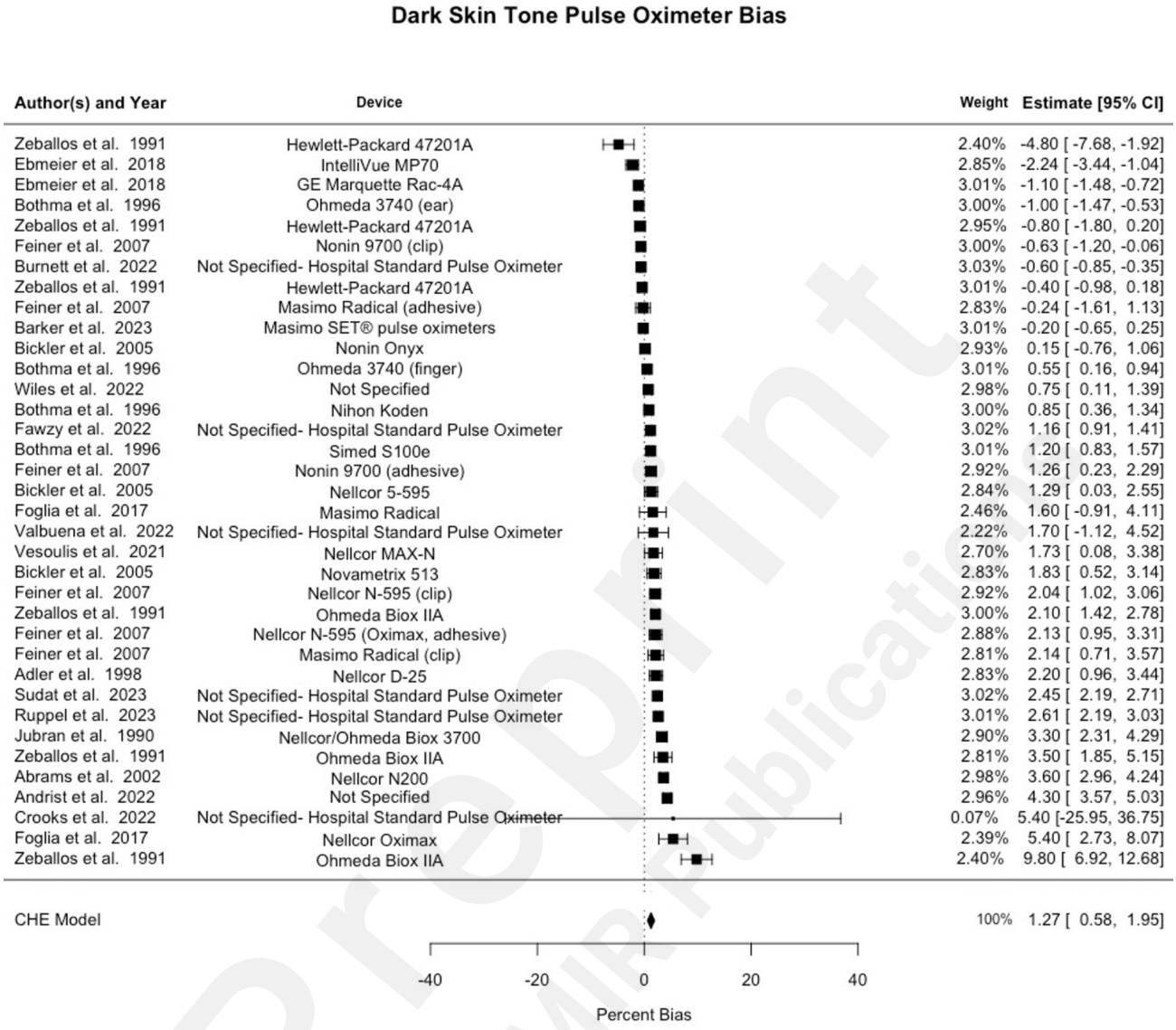


Table 4. Pulse Rate and Pulse Oximetry Bias by Skin Pigmentation.[c]

Device Type	Skin Pigment. Category	Number of Studies (evaluations)	Sample size (data pairs)	Unit	Pooled Mean Bias (95% CI)	Pooled SD (SE)	95% LoA	A_{rms} %	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)**	3·90 (1·36)	-6·94 to 8·34	3·96**	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**	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**	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**	98·55% (14·53% and 84·02%)
Pulse Rate	Light	3 (9)	31 (43,116)	bpm	-1·24 (-5·31 to 2·83)	7·54 (2·13)	-16·02* to 13·54*	–	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* to 16·84*	–	25·01% (0·00% and 25·01%)

Dark	1 (6)	16 (6,922)	bpm	-0·57 8·29)	(-9·44	to 16·89 (1·31)	-33·69* 32·54*	to –	13·70% (NA and 13·70%)
Combined	4 (24)	176 (140,771)	bpm	-0·29 3.29)	(-3·87	to 8·64 (1·67)	-17·23* 16·65*	to –	30·66% (26·76%) and 3·90%)

[c] ρ= 0·9 was used in CHE models to pool both mean bias and SD; Bold = statistically significant; * = exceeds ANSI Standards for Pulse Rate; ** = exceeds FDA Guidance for Pulse Oximetry. A_{rms} = accuracy root mean square, LoA = limits of agreement, NA = not applicable as there was only one PR study with dark skin pigmentation. CI = confidence interval, SD = standard deviation, SE = standard error.

PR. Analysis using the CHE model revealed limits of agreement (LoA) of -17.23 to 16.65 bpm and a mean bias of -0.29 bpm (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 bpm (95% CI: -5.31 to 2.83 bpm), -0.89 bpm (95% CI: -3.70 to 1.93 bpm), and -0.57 bpm (95% CI: -9.44 to 8.29 bpm) for the corresponding groups (Table 4 and Figures 5-7). Detailed results are provided in Tables S2 and S3 in Multimedia Appendix 1.

Figure 5. Pulse rate forest plot for light skin pigmentation [39, 59, 60].

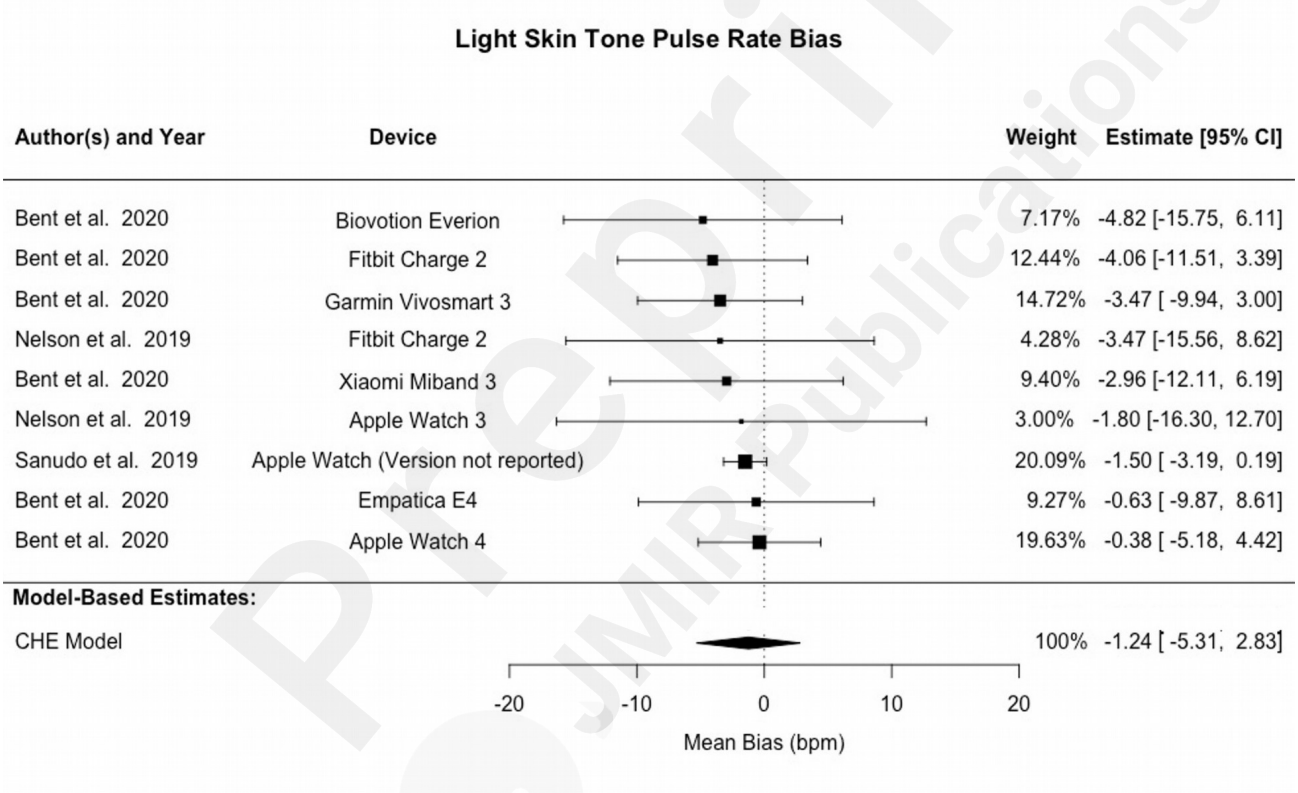


Figure 6. Pulse rate forest plot for medium skin pigmentation [39, 60, 61].

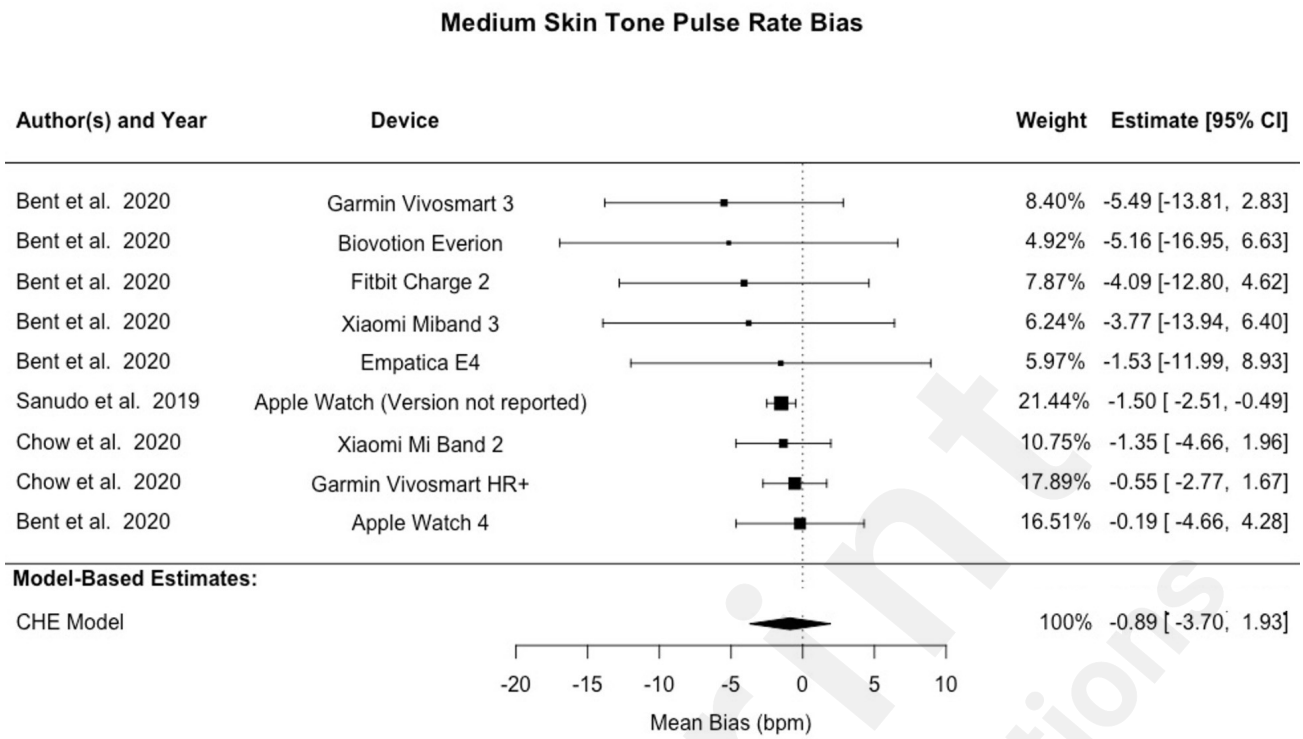
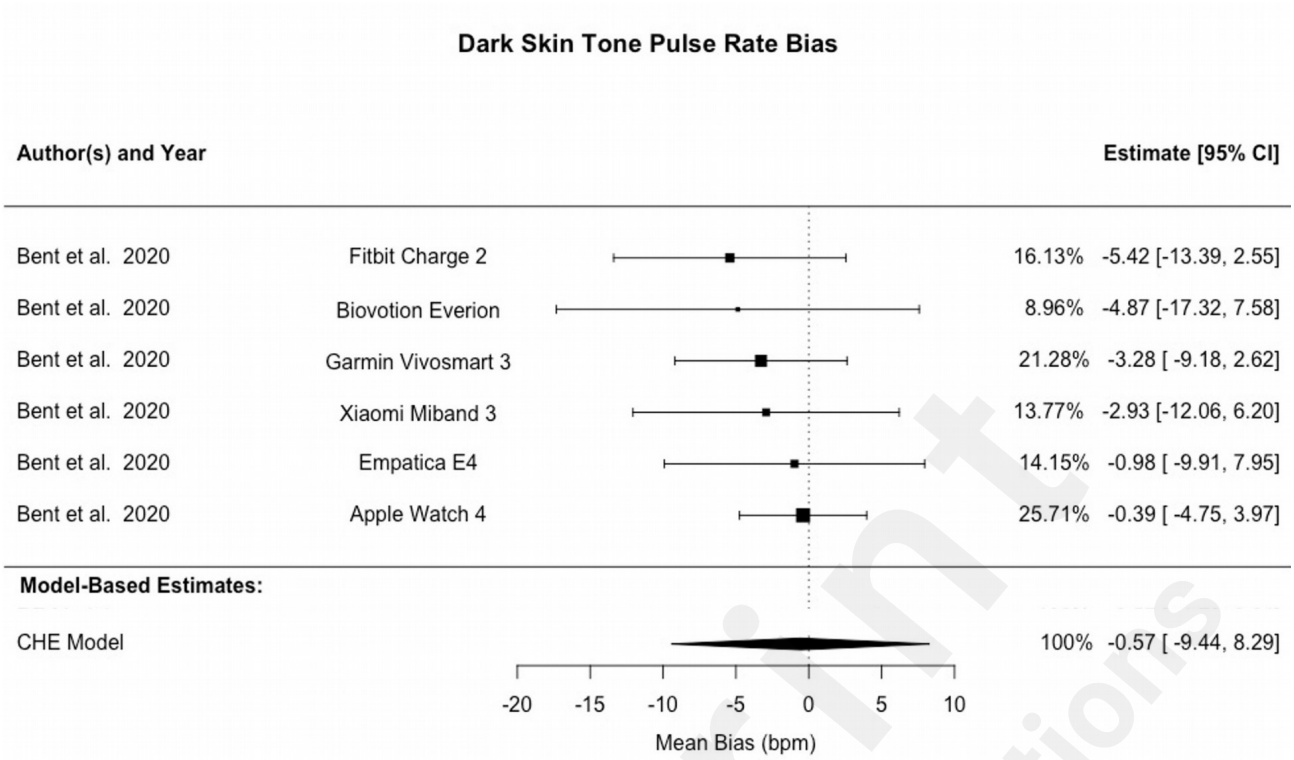


Figure 7. Pulse rate forest plot for dark skin pigmentation [39].



Sensitivity Analyses. Sensitivity analyses were performed for pulse oximetry SpO_2 by removing Crooks and colleagues [47] as an outlier due to substantially large SE (Tables S4 and S5 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 S4 in Multimedia Appendix 1.

Discussion

The study revealed a paucity of studies properly assessing skin pigmentation and 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. 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. And, in the sensitivity analyses without the outlier study of Crooks and colleagues [47], 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 [9-13,15], 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 [16]. Second, devices calibrated on individuals with medium

skin pigmentation (note that 48% of the US population is categorized as Fitzpatrick Skin Tone Scale (FSTS) III [64]) may lead to inaccurate readings for both lighter and darker skin pigmentation, since both may be suboptimally represented during algorithm training and testing.
Overall, these findings suggest that when pulse oximetry devices are deployed in their setting of intended use (i.e., 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 [64]. 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 and colleagues [34]. This method, however, classifies most Caucasians in the US as light, rather than medium-skin pigmentation [64]. Second, there were only four prior PR studies that collected participant skin tone that also reported on device accuracy and only one that had participants with dark pigmentation. Third, our study utilized 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 utilizing 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; future studies should delineate the effect of these and other moderating variables on SpO_2 and PR. Lastly, there are moderators that we didn't examine, such as medical versus healthy populations, degree of blood oxygenation, transmittance versus reflectance sensors, PPG light wavelength, or activity versus rest.

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 [17]. 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 [4,28,29,65-71]. 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 [4,28,29,65-71]:
Recommendation 1: It is vitally important for medical pulse oximetry devices as well as non-regulated research and consumer devices to incorporate the V3 framework [72] 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 [72], rather than relying on controlled laboratory settings where digits may be warmed prior to testing to maximize accuracy, 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 set 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 meta-analysis found 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.

Acknowledgements:

Author contributions:
Study concept and design: HG, SS, BWN
Data collection: HG, SS, MB, BWN
Data analysis and interpretation: BWN (data analysis and interpretation), CC (statistical method determination, code review)
Draft writing and review: BWN (abstract, introduction, methods, results, discussion, review), GH (methods, discussion, review), SS (methods, review), SS (review), MB (methods), CC (methods, review).
All authors have had full access to all the data in the study, final responsibility for the decision to submit for publication, and had direct access and verification of the underlying data reported in the manuscript.
Authors wish to acknowledge Shi et al. (2022) for providing open access to pulse oximetry bias for a subset of manuscripts included in this meta-analysis. Authors also wish to thank Whitney Townsend, University of Michigan research librarian, for helping develop and execute our literature search strategy.

Conflicts of interest

BWN, MB, SS, and CC report employment and equity ownership in Verily Life Sciences.

Abbreviations:

JMIR: Journal of Medical Internet Research

PPG: photoplethysmography
ECG: electrocardiography

References

1. Kyriacou PA, Allen J. *Photoplethysmography: Technology, Signal Analysis and Applications*. Academic Press; 2021. ISBN:9780128235256

2. Voice of the Consumer, Knowles M, Krasniansky A, et al. Consumer adoption of digital health in 2022: Moving at the speed of trust. Rock Health | We're powering the future of healthcare. Rock Health is a seed and early-stage venture fund that supports startups building the next generation of technologies transforming healthcare. Published February 21, 2023. Accessed March 1, 2023. https://rockhealth.com/insights/consumer-adoption-of-digital-health-in-2022-moving-at-the-speed-of-trust/

3. Al-Alusi MA, Khurshid S, Wang X, et al. Trends in Consumer Wearable Devices With Cardiac Sensors in a Primary Care Cohort. *Circ Cardiovasc Qual Outcomes*. 2022;15(7):e008833. doi: 10.1161/CIRCOUTCOMES.121.008833

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. doi: 10.1038/s41746-020-0297-4

5. Ries, A. L., Prewitt, L. M., & Johnson, J. J. Skin Color and Ear Oximetry. *Chest*. 1989;96(2):287-290. doi: 10.1378/chest.96.2.287

6. 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

7. 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

8. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005;102(4):715-719. doi: 10.1097/00000542-200504000-00004

9. 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. doi: 10.1056/NEJMc2029240

10. Valbuena VSM, Seelye S, Sjoding MW, 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. doi: 10.1136/bmj-2021-069775

11. 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. doi: 10.1001/jamainternmed.2022.2587

12. 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. doi: 10.4187/respcare.09795

13. Wong AKI, Charpignon M, Kim H, 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. doi: 10.1001/jamanetworkopen.2021.31674

14. Fawzy A, Wu TD, Wang K, 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. doi: 10.1001/jamanetworkopen.2023.30856

15. Fawzy A, Wu TD, Wang K, 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. doi: 10.1001/jamainternmed.2022.1906

16. 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

17. FDA. *Pulse Oximeters - Premarket Notification Submissions [510(k)s] Guidance for Industry and Food and Drug Administration Staff*. FDA; 2013. https://www.fda.gov/media/72470/download

18. FDA. *Review of Pulse Oximeters and Factors That Can Impact Their Accuracy*. FDA; 2022. https://www.fda.gov/media/162709/download

19. McFarling. FDA panel asks for improvements in pulse oximeters. Statnews. Published November 1, 2022. Accessed January 9, 2023. https://www.statnews.com/2022/11/01/fda-panel-asks-for-improvements-in-pulse-oximeters/

20. APA. WEARABLE DEVICES AS THERAPY TOOLS. Monitor in Psychology. Accessed January 10, 2023. https://www.apa.org/monitor/2021/2021-09-monitor.pdf

21. Center for Devices, Radiological Health. Pulse Oximeter Accuracy and Limitations: FDA Safety Communication. U.S. Food and Drug Administration. Accessed January 10, 2023. https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication

22. November 1, 2022: Anesthesiology and Respiratory Therapy Devices Panel. U.S. Food and Drug Administration. Published April 4, 2023. Accessed October 5, 2023. https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-1-2022-anesthesiology-and-respiratory-therapy-devices-panel-medical-devices-advisory

23. McFarling UL. Pulse oximeters' inaccuracies in darker-skinned people require urgent action, AGs tell FDA. STAT News. Published November 7, 2023. Accessed November 7, 2023. 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.

24. State of California Office of the Attorney General. Attorney General Letter to FDA. Published online November 1, 2023. https://oag.ca.gov/system/files/attachments/press-docs/23PR353%20Health%20Equity%20General%20Matter%20Multistate.pdf

25. Zinzuwadia A, Singh JP. Wearable devices—addressing bias and inequity. *The Lancet Digital Health*. 2022;4(12):e856-e857. doi:10.1016/s2589-7500(22)00194-7

26. Shachar C, Gerke S. Prevention of Bias and Discrimination in Clinical Practice Algorithms. *JAMA*. Published online January 5, 2023. doi:10.1001/jama.2022.23867

27. 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

28. Colvonen PJ, DeYoung PN, Bosompra NOA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. *Sleep*. 2020;43(10). doi:10.1093/sleep/zsaa159

29. Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms. State of California - Department of Justice - Office of the Attorney General. Published August 31, 2022. Accessed November 8, 2023. https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare

30. ANSI. *British Standards Institution (BSI) BS EN ISO 80601–2–61:2019: Medical Electrical Equipment – Part 2–61: Particular Requirements for Basic Safety and Essential Performance of Pulse Oximeter Equipment (ISO 80601–2–61:2017, Corrected Version 2018–02) London: British Standards Institution; 2019.*; 2019. Accessed October 30, 2023. https://webstore.ansi.org/standards/bsi/bseniso80601612019?gclid=Cj0KCQjwqP2pBhDMARIsAJQ0CzoRIm3aJK6juGmDJFoR0m6ykqn7mfbgNyMgSUgf3USfLNbfA9X0VukaAgoPEALw_wcB

31. ANSI/AAMI/IEC. *ANSI/AAMI/ IEC 60601- 2-27:2011 Medical Electrical Equipment — Part 2-27: Particular Requirements for the Basic Safety and Essential Performance of Electrocardiographic Monitoring Equipment*. ANSI/AAMI/IEC; 2011.

32. ANSI/AAMI. *ANSI/AAMI EC13-2002 Cardiac Monitors, Heart Rate Meters, And Alarms*. ANSI/AAMI; 2002. Accessed October 30, 2023. https://webstore.ansi.org/standards/aami/ansiaamiec132002

33. 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*. Published online November 14, 2022:1-9. doi: 10.1007/s40615-022-01446-9

34. Shi C, Goodall M, Dumville J, 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. doi: 10.1186/s12916-022-02452-8

35. 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* . 2022;22(9). doi:10.3390/s22093402

36. Al-Halawani R, Charlton PH, Qassem M, Kyriacou PA. A review of the effect of skin pigmentation on pulse oximeter accuracy. *Physiol Meas*. Published online May 12, 2023. doi:10.1088/1361-6579/acd51a

37. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi: 10.1136/bmj.n71

38. Center for Devices, Radiological Health. How to Determine if Your Product is a Medical Device. U.S. Food and Drug Administration. Published August 14, 2023. Accessed November 16, 2023. https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device

39. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med*. 2020;3:18. doi: 10.1038/s41746-020-0226-6

40. Viechtbauer W. *Metafor: A Meta-Analysis Package for R*. Github; 2023. doi:10.18637/jss.v036.i03

41. Abrams GA, 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. doi: 10.1053/jlts.2002.32252

42. Adler JN, Hughes LA, Vivilecchia R, Camargo CA Jr. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Acad Emerg Med*. 1998;5(10):965-970. doi: 10.1111/j.1553-2712.1998.tb02772.x

43. 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. doi: 10.1001/jamanetworkopen.2022.4584

44. Barker SJ, Wilson WC. Racial effects on Masimo pulse oximetry: a laboratory study. *J Clin Monit Comput*. 2023;37(2):567-574. doi: 10.1007/s10877-022-00927-w

45. Bothma PA, Joynt GM, Lipman J, et al. Accuracy of pulse oximetry in pigmented patients. *S Afr Med J*. 1996;86(5 Suppl):594-596.

46. Burnett GW, Stannard B, Wax DB, et al. Self-reported Race/Ethnicity and Intraoperative Occult Hypoxemia: A Retrospective Cohort Study. *Anesthesiology*. 2022;136(5):688-696. doi: 10.1097/ALN.0000000000004153

47. Crooks CJ, West J, Morling JR, et al. Pulse oximeter measurements vary across ethnic groups: an observational study in patients with COVID-19. *Eur Respir J*. 2022;59(4). doi:10.1183/13993003.03246-2021

48. Ebmeier SJ, Barker M, Bacon M, 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

49. Foglia EE, Whyte RK, Chaudhary A, et al. The Effect of Skin Pigmentation on the Accuracy of Pulse Oximetry in Infants with Hypoxemia. *J Pediatr*. 2017;182:375-377.e2. doi: 10.1016/j.jpeds.2016.11.043

50. 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

51. 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

52. 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. doi: 10.1183/09031936.00126507

53. Pilcher J, Ploen L, McKinstry S, 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. doi: 10.1186/s12890-019-1007-3

54. Ruppel H, Makeneni S, Faerber JA, et al. Evaluating the Accuracy of Pulse Oximetry in Children According to Race. *JAMA Pediatr*. 2023;177(5):540-543. doi: 10.1001/jamapediatrics.2023.0071

55. Sudat SEK, Wesson P, Rhoads KF, 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. doi: 10.1093/aje/kwac164

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

57. 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. doi: 10.1038/s41372-021-01230-3

58. Wiles MD, El-Nayal A, Elton G, 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. doi: 10.1111/anae.15581

59. 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. doi: 10.2196/10828

60. 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

61. Chow HW, Yang CC. 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. doi: 10.2196/14707

62. 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

63. Pustejovsky J. Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections [R package clubSandwich version 0.5.10]. Published 2023. Accessed May 29, 2024. https://CRAN.R-project.org/package=clubSandwich

64. 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

65. 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. doi: 10.1164/rccm.202209-1773ED

66. National Academies of Sciences, Engineering, and Medicine. 2022. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington, DC: The National Academies Press. https://doi.org/10. 17226/26479. *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*. National Academies of Sciences, Engineering, and Medicine; 2022. doi:10.17226/26479

67. Barton G, Lee NT, Resnick P. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. Published May 22, 2019. Accessed November 8, 2023. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

68. FDA. *Framework for the Use of Digital Health Technologies in Drug and Biological Product Development*. FDA; 2023. https://www.fda.gov/media/166396/download?attachment

69. Center for Drug Evaluation, Research. Digital Health Technologies for Remote Data Acquisition in Clinical Investigations. U.S. Food and Drug Administration. Published March 27, 2023. Accessed December 5, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations

70. Office of the Commissioner. Diversity Plans to Improve Enrollment of Participants From Underrepresented Racial and Ethnic Populations in Clinical Trials; Draft Guidance for Industry;

Availability. U.S. Food and Drug Administration. Published April 13, 2022. Accessed December 5, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations

71. Center for Drug Evaluation, Research. Enhancing the Diversity of Clinical Trial Populations — Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry. U.S. Food and Drug Administration. Published November 13, 2020. Accessed November 30, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial

72. Goldsack JC, Coravos A, Bakker JP, 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. doi: 10.1038/s41746-020-0260-4

73. Clark JM, Sanders S, Carter M, et al. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. *J Med Libr Assoc*. 2020;108(2):195-207. doi: 10.5195/jmla.2020.834

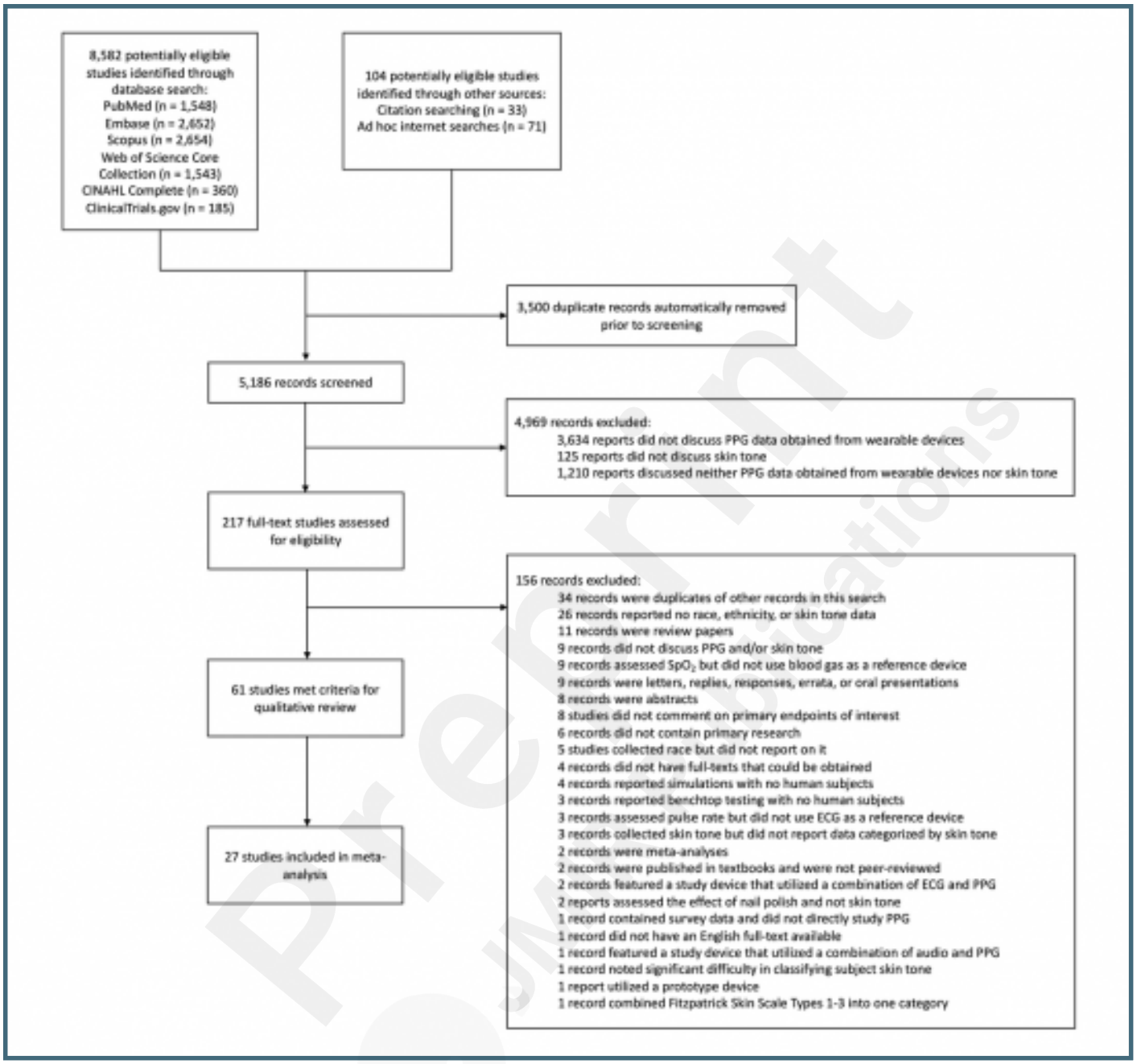
74. Higgins JPT, Deeks JJ. Selecting studies and collecting data. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd; 2008:151-185. ISBN: 9780470712184

75. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. The White House. Accessed November 16, 2023. https://obamawhitehouse.archives.gov/omb/fedreg_1997standards

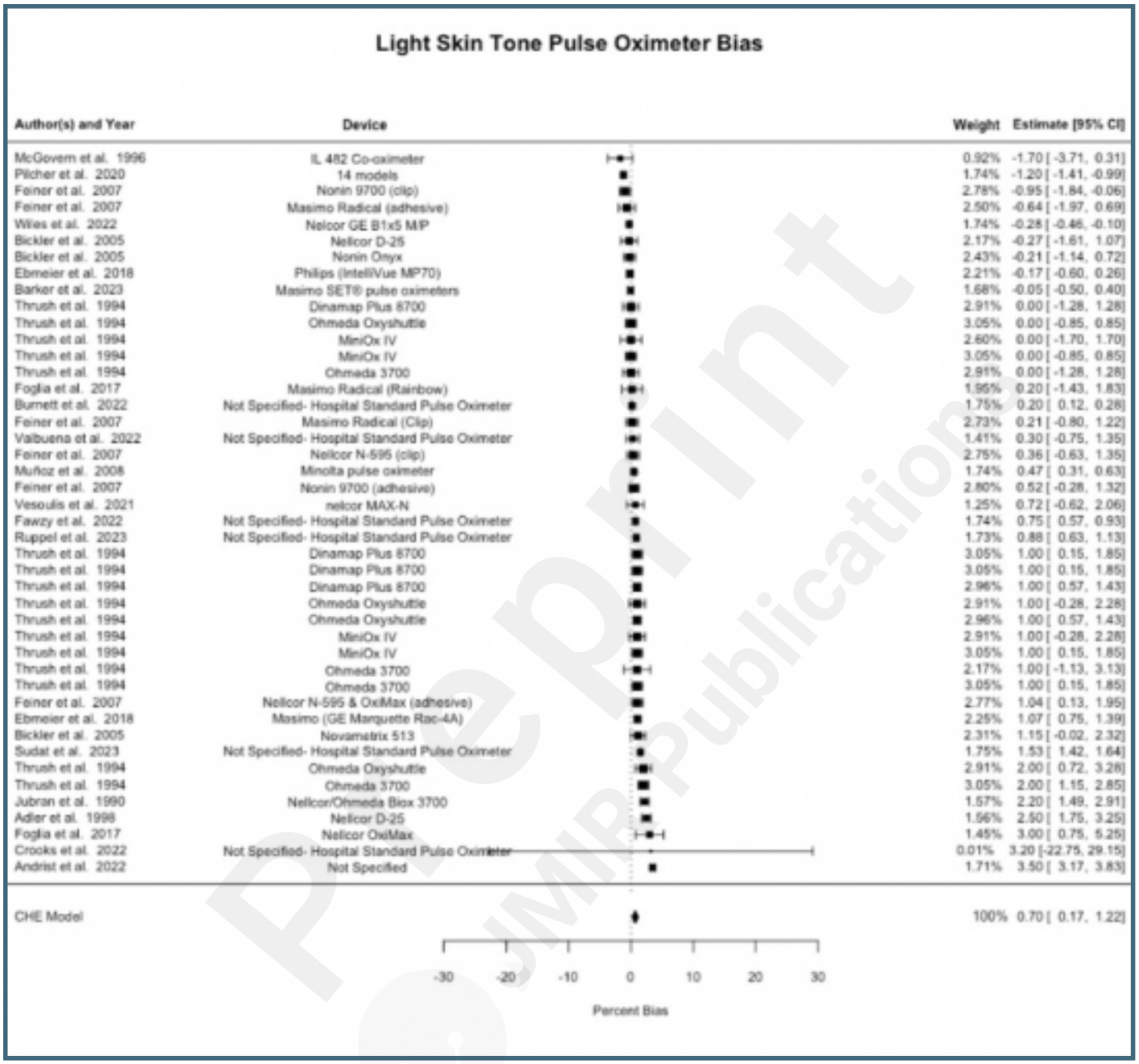
Supplementary Files

Figures

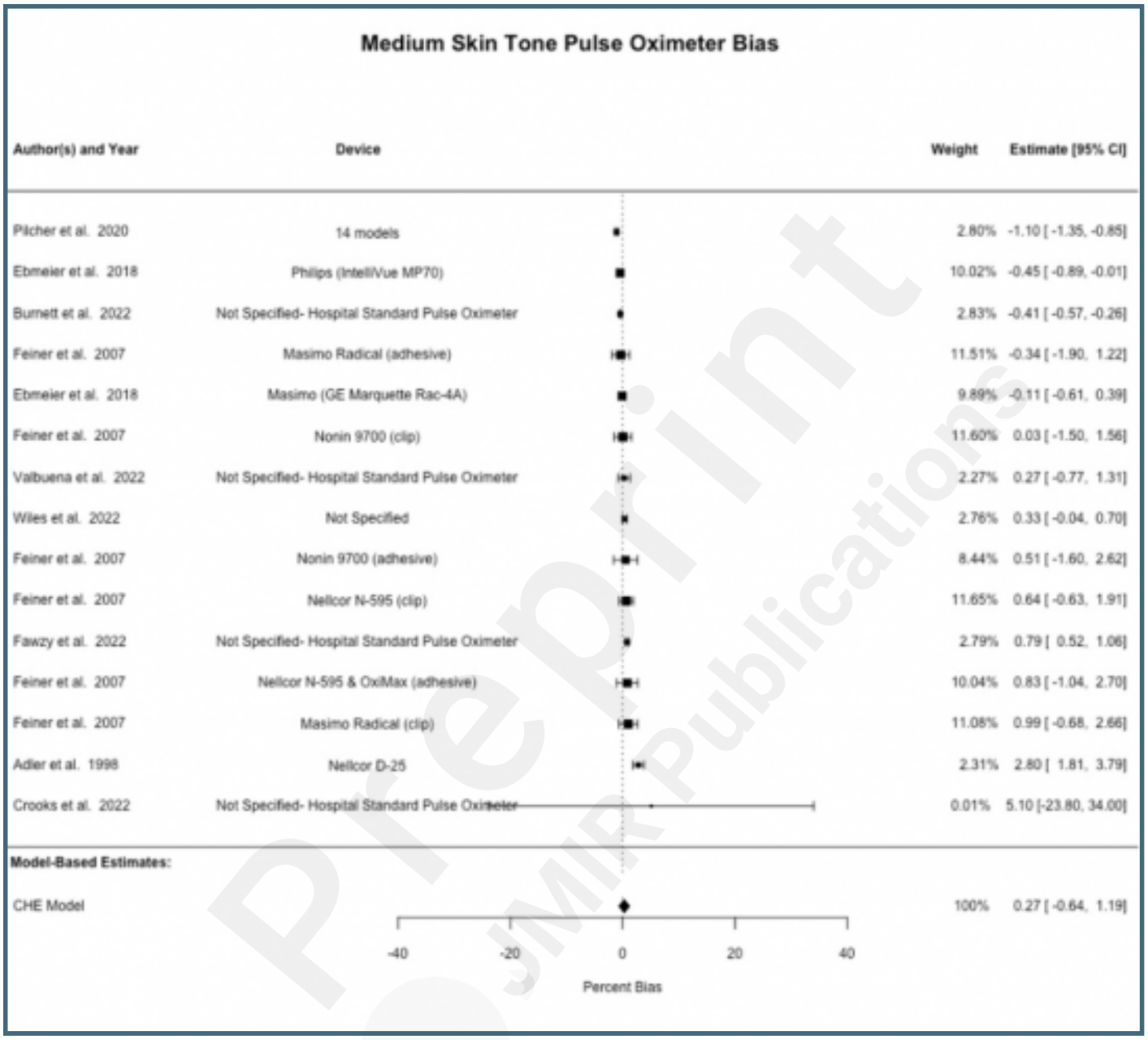
Study selection flow chart.



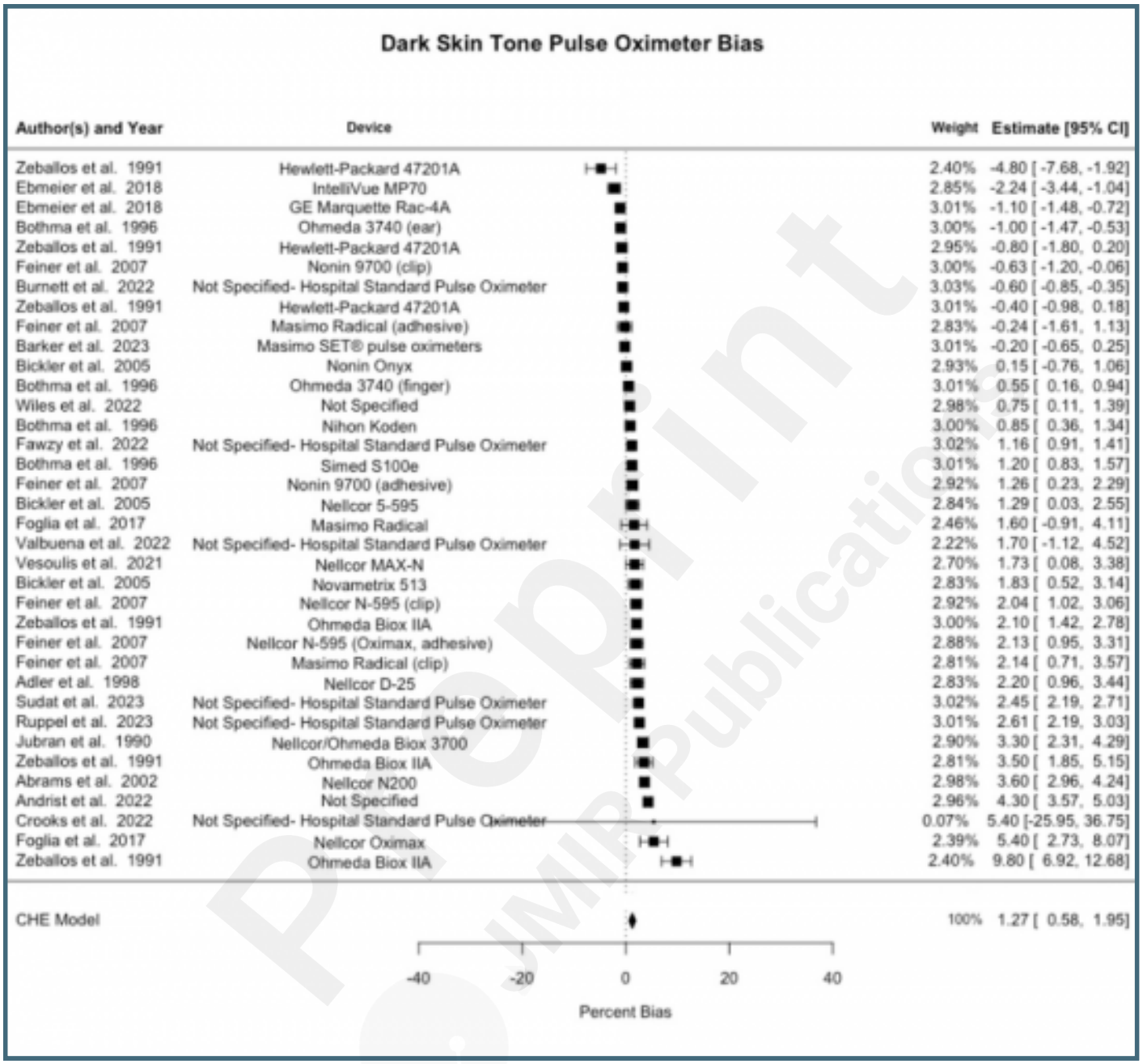
Pulse oximetry forest plot for light skin pigmentation with multiple entries from the same study representing different devices examined [6, 8, 10, 15, 42-44, 46-58].



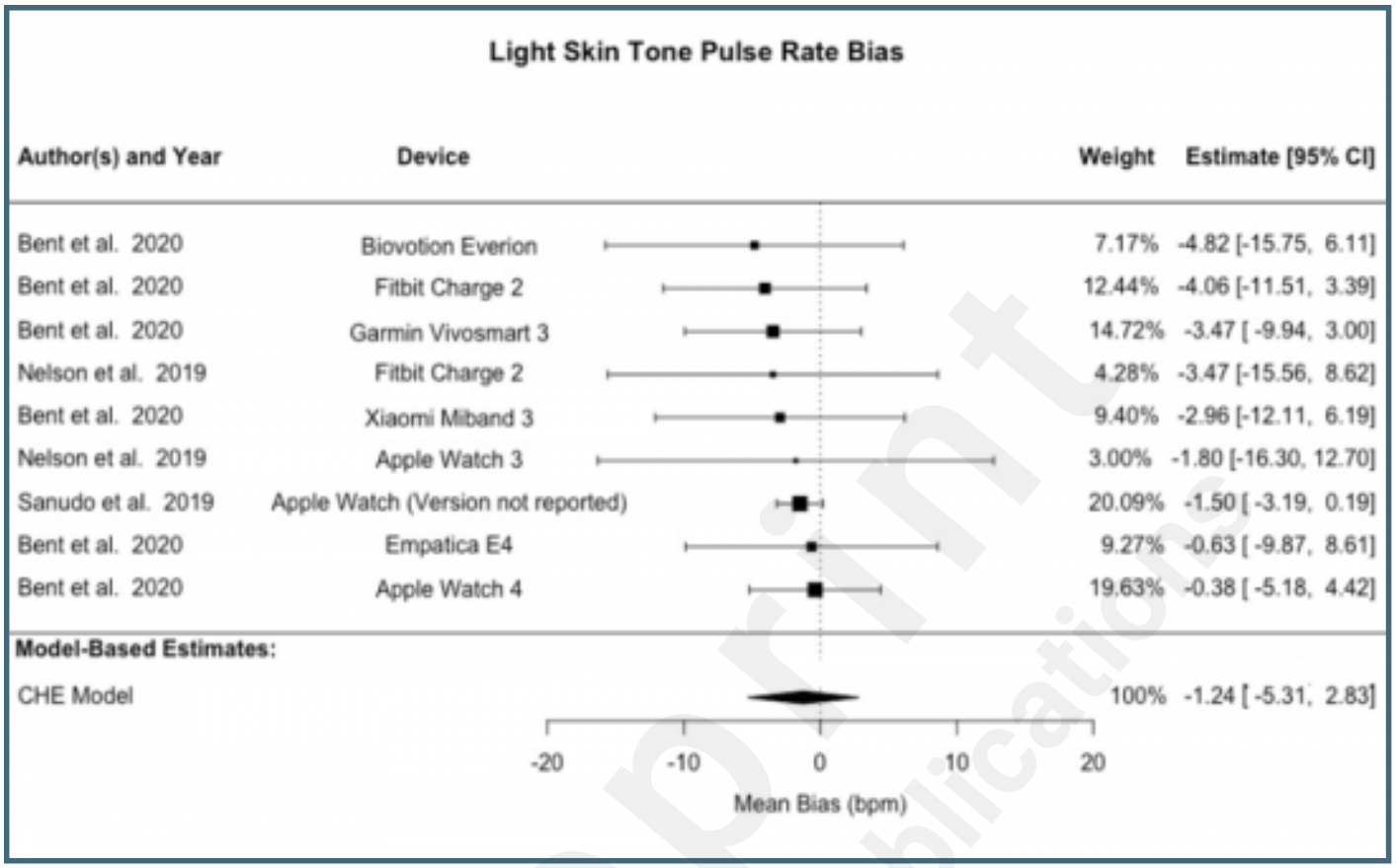
Pulse oximetry forest plot for medium skin pigmentation with multiple entries from the same study representing different devices examined [6, 10, 15, 42, 46-48, 53, 58].



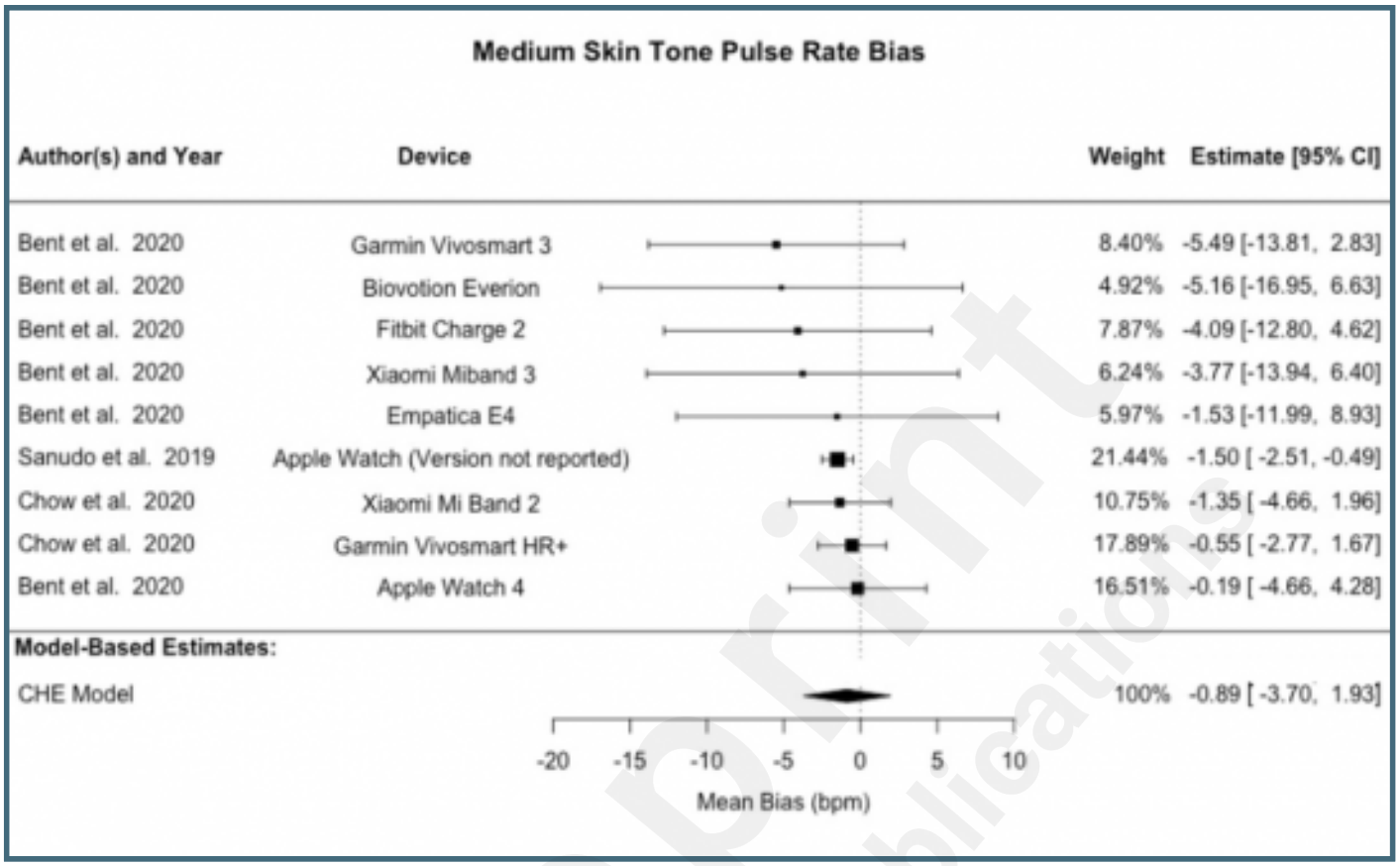
Pulse oximetry forest plot for dark skin pigmentation with multiple entries from the same study representing different devices examined [6-8, 10, 15, 41-50, 54, 55, 57, 58].



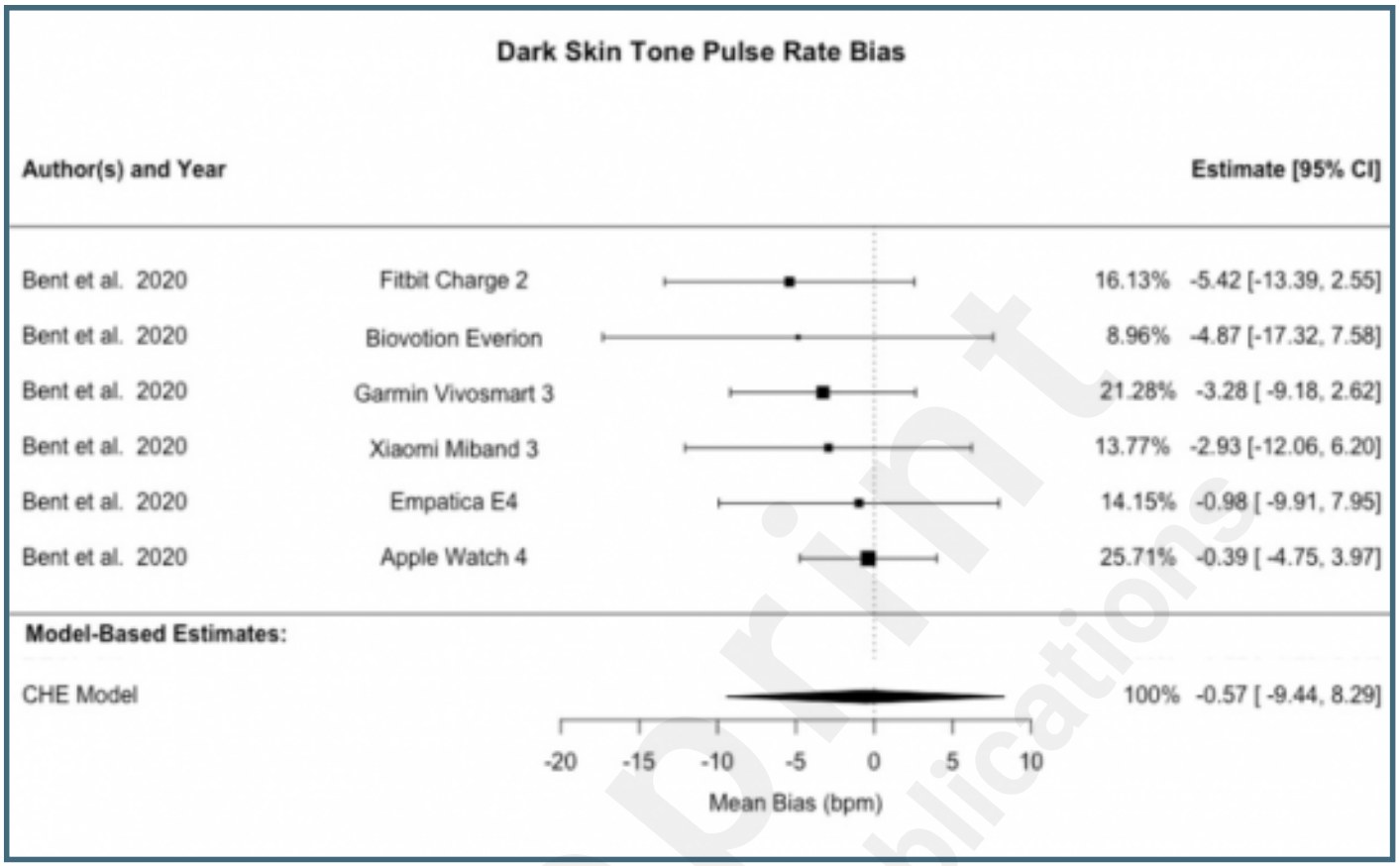
Pulse rate forest plot for light skin pigmentation [39, 59, 60].



Pulse rate forest plot for medium skin pigmentation [39, 60, 61].



Pulse rate forest plot for dark skin pigmentation [39].



Multimedia Appendixes

Supplementary information.
URL: http://asset.jmir.pub/assets/84686b6479e0d3adb5869174c731710c.docx

CONSORT (or other) checklists

PRISMA 2020 Checklist.
URL: http://asset.jmir.pub/assets/55caa205656f9fd62ecc212b8ec483f9.pdf