

# The effects of skin pigmentation on the accuracy of pulse oximetry in measuring oxygen saturation: a systematic review and meta-analysis

Chunhu Shi<sup>1,2\*</sup>, Mark Goodall<sup>3,4</sup>, Jo Dumville<sup>1,2</sup>, James Hill<sup>4,5</sup>, Gill Norman<sup>1,2</sup>, Oliver Hamer<sup>4,5</sup>, Andrew Clegg<sup>4,5</sup>, Caroline Leigh Watkins<sup>4,5</sup>, George Georgiou<sup>4,5</sup>, Alexander Hodgkinson<sup>6,7</sup>, Catherine Elizabeth Lightbody,<sup>8</sup> Paul Dark,<sup>9,10</sup> Nicky Cullum<sup>1,2</sup>

<sup>1</sup> School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

<sup>2</sup> NIHR Applied Research Collaboration Greater Manchester (ARC-GM), Manchester, UK

<sup>3</sup> Institute of Population Health, University of Liverpool, Liverpool, UK

<sup>4</sup> NIHR Applied Research Collaboration North West Coast (ARC-NWC), UK

<sup>5</sup> Applied Health Research Hub, University of Central Lancashire, Preston, UK

<sup>6</sup> NIHR School for Primary Care Research, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, UK

<sup>7</sup> NIHR Greater Manchester Patient Safety Translational Research Centre, Division of Population Health, Health Services Research & Primary Care, University of Manchester, UK

<sup>8</sup> Faculty of Health, University of Central Lancashire, Preston, UK

<sup>9</sup> NIHR Manchester Biomedical Research Centre, University of Manchester, Manchester, UK

<sup>10</sup> Northern Care Alliance NHS Foundation Trust, Salford Care Organisation, Greater Manchester, UK

\* Corresponding author

E-mail: [chunhu.shi@manchester.ac.uk](mailto:chunhu.shi@manchester.ac.uk)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## Abstract

### Background

Pulse oximetry was widely used in hospitals and at home to monitor blood oxygen during the COVID-19 pandemic. There have been concerns regarding potential bias in pulse oximetry measurements for people with dark skin. We aimed to assess the effects of skin pigmentation on the accuracy of oxygen saturation measurement by pulse oximetry ( $SpO_2$ ) compared with the gold standard  $SaO_2$  measured by CO-oximetry.

### Methods

We searched Ovid MEDLINE, Ovid Embase, and EBSCO CINAHL Plus (up to December 2021), as well as ClinicalTrials.gov and World Health Organization International Clinical Trials Registry Platform (up to August 2021). We identified studies comparing  $SpO_2$  values in any population, in any care setting, using any type of pulse oximeter, with  $SaO_2$  by standard CO-oximetry; and measuring the impact of skin pigmentation or ethnicity on pulse oximetry accuracy. We performed meta-analyses for mean bias (the primary outcome in this review) and its standard deviations (SDs) across studies included for each subgroup of level of skin pigmentation and ethnicity. We calculated accuracy root-mean-square ( $A_{rms}$ ) and 95% limits of agreement based on pooled mean bias and pooled SDs for each subgroup.

### Results

We included 32 studies (6505 participants); 27/32 (84.38%) in hospitals and none in people's homes. Findings of 14/32 studies (43.75%) were judged, via QUADAS-2, at high overall risk of bias. Fifteen studies measured skin pigmentation and 22 referred only to ethnicity. Compared with standard  $SaO_2$  measurement, pulse oximetry probably overestimates oxygen saturation in people with dark skin (pooled mean bias 1.11%; 95% confidence interval 0.29% to 1.93%) and people described as Black/African American (pooled mean bias 1.52%; 0.95% to 2.09%) (moderate- and low-certainty evidence). These results suggest that, for people with dark skin, pulse oximetry may overestimate blood oxygen saturation by around 1% on average compared with  $SaO_2$ . The bias of pulse oximetry measurements for people with other levels of skin pigmentation, or those from the White/Caucasian group is more uncertain. The data do not suggest overestimation in people from other ethnic groups such as those described as Asian, Hispanic, or mixed ethnicity (pooled mean bias 0.31%, 0.09% to 0.54%), but this evidence is low certainty. Whilst the extent of mean bias is small or negligible for all the subgroups of population evaluated, the associated imprecision is unacceptably large (with the pooled SDs > 1%). Nevertheless, when the extents of measurement bias and precision are considered jointly in  $A_{rms}$ , pulse oximetry measurements for all the subgroups appear acceptably accurate (with  $A_{rms}$  < 4%).

### Conclusions

Low-certainty evidence suggests that pulse oximetry may overestimate oxygen saturation in people with dark skin and people whose ethnicity is reported as Black/African American, compared with SaO<sub>2</sub>, although the overestimation may be quite small in hospital settings. The clinical importance of any overestimation will depend on the particular clinical circumstance. Pulse oximetry measurements appear accurate but imprecise for all levels of skin pigmentation. The evidence relates to clinician-measured oximetry in health care environments and may not be reflected in home pulse oximetry where other factors may also influence accuracy.

**Keywords:** pulse oximetry; arterial blood oxygen saturation; measurement bias; skin pigmentation; ethnicity; systematic review

## Author summary

### Why was this study done?

- Pulse oximetry was widely used in hospital and at home to measure blood oxygen levels during the COVID-19 pandemic.
- There was uncertainty as to whether skin pigmentation affects the accuracy of pulse oximetry measurements.

### What did the researchers do and find?

- We assessed, via systematic review, the effects of skin pigmentation on the accuracy of pulse oximetry measurement ( $SpO_2$ ) compared with  $SaO_2$  measured by standard CO-oximetry.
- In people with dark skin, oxygen saturation measured in hospital by pulse oximetry may be overestimated by an average of 1% compared with gold standard  $SaO_2$ , however, the evidence is of low certainty.
- The accuracy of pulse oximetry measured compared with standard  $SaO_2$  is quite uncertain for people with light or medium levels of skin pigmentation and for people from ethnic groups other than those described in papers as Black or African American.
- Pulse oximetry measurements appear to have acceptable overall accuracy (with  $A_{rms} < 4\%$ ) for all subgroups of population evaluated whilst the variation of oximetry readings appear unacceptably wide (with the pooled SDs  $> 1\%$ ).

### What do these findings mean?

- Hospital measured pulse oximetry may overestimate oxygen saturation in people with dark skin compared with  $SaO_2$  by approximately 1%.
- The implications of this finding in different clinical scenarios will vary but could be clinically important. Impacts are likely to be at the thresholds of being diagnosed as having hypoxaemia where even a small  $SpO_2$  overestimation could lead to clinically important hypoxaemia remaining undetected and untreated.
- How these findings extrapolate to community and home care settings is unclear.

## Introduction

Blood oxygen saturation levels require monitoring for health reasons in a wide range of circumstances. Low blood oxygen saturation, if identified to be hypoxemia, requires medical intervention and has been linked to an increased risk of death.[1] The gold standard measure of blood oxygen saturation levels (SaO<sub>2</sub>) requires a sample of arterial blood and measurement using CO-oximetry. Pulse oximetry, measuring SpO<sub>2</sub> as a proxy for SaO<sub>2</sub> using a non-invasive and inexpensive device, is frequently used to detect low blood oxygen levels. Pulse oximetry has been widely used during the COVID-19 pandemic, including in non-clinical settings, to detect hypoxemia and inform decisions to escalate care.[2]

The current WHO COVID-19 management guideline recommends the “*use of pulse oximetry monitoring at home as part of a package of care*” for symptomatic people with COVID-19.[3] Many countries have specific guidance or services for home pulse oximetry in line with this recommendation.[2,4] In the UK, the NHS England COVID Oximetry@home service provided oximetry devices to people with COVID-19 for their home self-monitoring of oxygen levels. Guidance recommended that those with pulse oximetry readings of 92% or less “*attend A&E as quickly as possible or call 999 immediately*”. [2]

The reporting of possible bias in pulse oximetry measurement, including due to skin pigmentation, raised a growing concern about the accuracy of oxygen self-monitoring.[5] Measurement bias could have serious clinical implications including the delay of urgent medical care.[6]. A recent US study analysed retrospective cohort data from more than 10,000 people, comparing where a diagnosis of *occult hypoxemia* (an SaO<sub>2</sub> of less than 88%) was missed by pulse oximetry.[7] Results showed people described as Black had ‘*nearly three times the frequency of occult hypoxemia that was not detected by pulse oximetry*’ as those described as White.[7] In November 2021, the UK Health Secretary ordered a review into racial bias in medical equipment, including pulse oximeters.

It is an important time to consider the current evidence base for the impact of skin pigmentation on the accuracy of pulse oximetry compared with the gold standard measure of SaO<sub>2</sub>. The only current relevant systematic review, published in 1995, included three studies that explicitly considered the impact of skin pigmentation on pulse oximetry accuracy.[8] The review concluded that pulse oximeters may overestimate blood oxygen saturation in people with high levels of skin pigmentation.[8] The recent rapid review by the NHS Race and Health Observatory came to similar conclusions but used a non-systematic review process, e.g., having no comprehensive searches, critical appraisal of included studies and meta-analysis.[6] Our objective was therefore to conduct a systematic review of relevant evidence for the influence of skin pigmentation on the accuracy of oxygen saturation measurement by pulse oximetry (SpO<sub>2</sub>) compared with SaO<sub>2</sub> measured by standard CO-oximetry.

## Methods

We used Cochrane methodologies to conduct this review,[9] and report it in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[10]

### Protocol and registration

We registered the protocol with the Open Science Framework (available here <https://osf.io/gm7ty>).[11]

### Eligibility criteria

We included any prospective or retrospective methods-comparison study, completed or ongoing, including full publications and conference abstracts.

We included studies that compared SpO<sub>2</sub> values in any population, in any care setting, measured using any type of commercially available pulse oximeter, with SaO<sub>2</sub> measured by standard CO-oximetry (in the same individuals).[12] We excluded studies that used: (1) prototype pulse oximetry devices; (2) pulse oximeters that require high-skilled specialists to operate (such as intra-partum pulse oximetry devices); and (3) pulse oximeters used for measuring venous blood oxygen saturation, as our focus was evaluating commercial pulse oximetry devices that could be potentially used at home for arterial oxygen saturation measurement.

We included studies investigating the accuracy of pulse oximetry based on both the level of skin pigmentation and ethnic group. We considered ethnicity as a sub-optimal but relevant proxy for level of skin pigmentation in relation to the accuracy of the biological reading of oxygen saturation.

For level of skin pigmentation measurement, we included studies that used a standardised measure of level of skin pigmentation such as the Fitzpatrick scale,[13] and studies that used an unstandardised or qualitative judgement of skin pigmentation levels such as so called 'light' or 'dark' level of skin pigmentation. The originally reported terms of skin pigmentation were mapped into 'low', 'medium' or 'high' pigmentation categories, without any further re-classification. For example, if a study classified skin as 'light' and 'dark', then the light skin group was mapped into the low pigmentation category, and the dark skin group into the high pigmentation category. For ethnicity, we included studies that grouped participants based on any ethnicity classification. We interpreted terms such as Black and White as describing ethnicity and only indirect indicators of skin pigmentation.[14]

Where study reports were inadequate to support inclusion decisions, we contacted authors for more information. We excluded studies where no response was received from contacting authors.

We focus here on reporting the comparative accuracy of pulse oximetry-produced SpO<sub>2</sub> in relation to SaO<sub>2</sub> measured by gold standard CO-oximetry. We therefore excluded studies that reported diagnostic test accuracy measures and those with inappropriate comparators, including: use of a reference pulse oximetry as the comparator, use of incorrect reference values of oxygen saturation e.g. arterial oxygen pressure (PaO<sub>2</sub>), calculated SaO<sub>2</sub>, fractional saturation (%O<sub>2</sub>Hb or FO<sub>2</sub>Hb).[12,15]

Following the British Standards Institution 2019 standards for pulse oximetry,[12] we included data on the overall accuracy (agreement between a test result and an accepted reference value reported as accuracy root-mean-square,  $A_{rms}$ ), mean bias and precision (and/or the limits of agreement for the SpO<sub>2</sub> and SaO<sub>2</sub> comparison) (Supporting Information Table S1). The overall accuracy ( $A_{rms}$ ) is a combination of mean bias and precision in a single measure.[12]. Mean bias is calculated as the mean difference between two measures (in this case SpO<sub>2</sub> – SaO<sub>2</sub>).[12] A larger mean difference indicates a larger bias value and values greater than 0 indicate overestimation with pulse oximetry.[12] Precision is commonly reported as the standard deviation of between-test mean difference with a larger standard deviation indicating less precision.[12]

The British Standards Institution standards for pulse oximetry gives  $A_{rms}$  primacy, as it details general accuracy and thus the suitability of the machine for its purpose. The value is a root mean square deviation calculation, and the relevance of this measure to clinical decision-making is not intuitive. Because of this we present mean bias as the review's primary outcome. This mean difference between 'true' blood oxygen saturation levels and those measured by pulse oximetry can more clearly indicate how clinical decisions referring to threshold values (e.g., admission to hospital with a pulse oximetry reading of 92% or lower) could be impacted by bias.

### **Information sources and literature searches**

In August 2021, we searched Ovid MEDLINE and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily (1946), Ovid Embase (1974), and EBSCO CINAHL Plus (Cumulative Index to Nursing and Allied Health Literature; 1937) to identify only English language reports of relevant studies given the limited time and resources available. Searches were updated in December 2021. There were no restrictions with respect to date of publication or study setting. See Supporting Information Box S1 for the search strategy used for Ovid MEDLINE.

In August 2021, we searched the ClinicalTrials.gov and World Health Organization International Clinical Trials Registry Platform for ongoing studies; and the reference lists of retrieved included studies, relevant systematic reviews, and guideline reports. We also contacted authors of key abstracts to request further information about their studies.

### **Study selection**

Two reviewers (CS and MG, or JH, OH) independently assessed titles and abstracts of the search results for relevance, with records stored and screened in the Abstrackr software.[16] Two reviewers (CS and MG or JH) independently inspected the full texts of all potentially eligible studies. The two reviewers resolved disagreements through discussion, involving a third reviewer where necessary.

### **Data extraction**

One reviewer (CS, or OH or JH) independently extracted data from included studies, which were checked by another reviewer (JH, MG, OH, GN). We resolved any disagreements through discussion. See Supporting Information Box 2 for data items we extracted.

We contacted study authors to clarify methods and data, where necessary. We transformed data into a format needed for analyses when required, e.g. from reported 95% limits of agreement or confidence interval (CI) to standard deviation (SD) using Wan's method.[17]

### **Risk of bias assessment**

There is no standard risk of bias tool for methods-comparison studies. We therefore chose QUADAS-2 for risk of bias assessment;[18] tailoring it as required, by adding and omitting signalling questions, see Supporting Information Box 3 for further information.

One reviewer (CS, or OH or JH) independently assessed the risk of bias for the included studies, checked by another reviewer (GN, MG, JH, OH) with any discrepancies resolved via discussion.

### **Data synthesis and analysis**

We summarised the included studies narratively and quantitatively; pooling evidence on  $A_{rms}$ , mean bias, precision (SD), and 95% limits of agreement. As specified in our protocol, we analysed data separately for studies which reported level of skin pigmentation and ethnicity. When pooling data for mean bias and its SD across studies, we used the correlated hierarchical effects model with small-sample corrections under the Robust Variance Estimation (RVE) framework (Supporting Information Box 4). The approach enabled us to include single-measure design study data, together with multiple dependent effect size estimates (i.e. mean bias and SD) of a repeated-measures design study in meta-analysis even when the dependence structure is unknown.[19-21] We used  $\tau^2$  (95% CI),  $I^2$  and the Q statistic and the related  $\chi^2$  test to fully assess heterogeneity in meta-analysis.

There is no established approach to pooling data for  $A_{rms}$  and 95% limits of agreement across studies directly. In this review, we used the pooled mean bias and the pooled SDs produced

by related meta-analyses, and followed the British Standard Institute methods to calculate the  $A_{rms}$  [12] and Tipton and Shuster's methods to calculate the population 95% limits of agreement.[20,22]

Within the R open-source software environment (version 4.1.2; R Foundation for Statistical Computing),[23] we performed RVE meta-analyses using functions available in packages of *clubSandwich*,[24] and *metafor*,[25] and used the *forestplot* package to produce forest plots that graphically present pooled – or originally reported when data pooling was impossible – results of  $A_{rms}$ , mean bias (SD), and 95% limits of agreement.[26] Supporting Information Box 4 presents generic R codes used for meta-analysed. All codes and analyses were checked by our statistician co-author (AH).

When meta-analysis was not appropriate, we synthesised relevant evidence following the Synthesis Without Meta-analysis in systematic reviews (SWiM) reporting guideline.[27]

One reviewer (CS) assessed the certainty of evidence on mean bias by referring to the GRADE approach developed for the test accuracy topic, checked by another reviewer (GN).[28,29] Using this approach the certainty of mean bias findings could be assessed as at high, moderate, low or very low certainty. In interpreting review findings, we used the British Standards Institution-recommended thresholds to judge the accuracy of pulse oximetry, viz. for pulse oximetry to be accurate by British Standards Institution standards, the overall accuracy  $A_{rms}$  should be within 4% over the range of 70% to 100% SaO<sub>2</sub>. For the mean bias (and precision), SpO<sub>2</sub> measures should be within +/- 2% of CO-oximetry measures and the variation for repeated SpO<sub>2</sub> measures should be within one SD of the mean bias.[12] With the mean bias as the primary outcome, any pooled mean bias of > 0% would suggest a risk of hypoxemia. Given pulse oximeter devices commonly present integers in percentage, we rounded pooled estimates to be integers when interpreting the related findings such as rounding mean bias values within +/-0.50% to 0%.

We analysed data on pulse oximeters of different brands/manufacturers separately where possible, or presented originally reported data where meta-analysis was not appropriate.

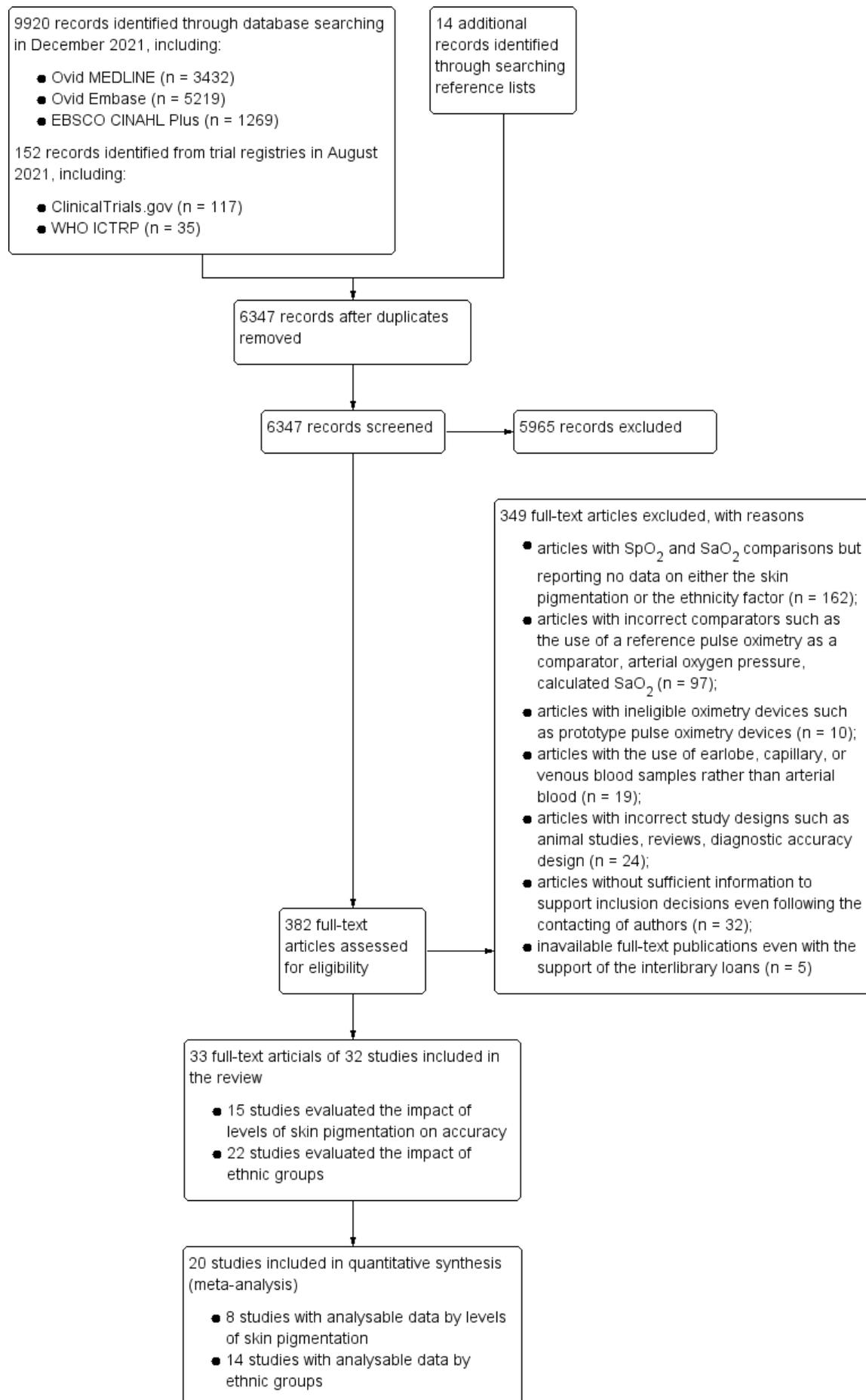
We undertook pre-planned sensitivity analyses to assess the robustness of meta-analysis, (1) excluding studies in which all participants had the same level of skin pigmentation or were from the same ethnicity, (2) excluding studies that have no data directly available for a meta-analysis unless data are transformed, and (3) excluding studies at high overall risk of bias. We also undertook post-hoc sensitivity analysis to assess the impact of excluding studies that used descriptors of ethnicity to indicate levels of skin pigmentation. This is because those study data might be systematically different from the others that used standardised skin pigmentation measurement methods such as the Fitzpatrick scale.

We assessed publication bias following a qualitative approach without using funnel plots or Egger's tests, given existing methods were not considered appropriate for the case of this review.[30]

## Results

### Search results

Following a search in August 2021 and an update search in December 2021, we assessed titles and abstracts of 9920 records identified from electronic databases and 152 from trial registries. We also assessed 14 records identified by screening the reference lists of included studies and existing reviews. After assessing full texts of 382 potentially relevant records, we identified 33 publications of 32 studies – published between 1985 and 2021 – as eligible for inclusion in this review (Figure 1).[31-63] We also identified one ongoing study from electronic searches.[64] By contacting authors we received raw or study-level summary data for two studies.[39,51]



## Figure 1. The study selection flowchart

### Characteristics of included studies

Table 1 summarises included studies (Supporting Information Table S2 presents a more detailed overview).

**Table 1. Summary characteristics of the included studies**

Items	Summary statistics (n (%))
Study designs (32 studies)	
<i>Prospective design</i>	29 (90.62%)
<i>Retrospective design</i>	3 (9.38%)
Repeated-measures design (32 studies)	
<i>Yes (that is, more than one SpO<sub>2</sub>-SaO<sub>2</sub> data pair collected per person)</i>	21 (65.62%)
<i>No</i>	11 (34.38%)
Care settings (32 studies)	
<i>Hospital care settings</i>	27 (84.38%)
<i>Laboratory setting*</i>	5 (15.62%)
Types of participants (32 studies)	
<i>Children</i>	7 studies (21.88%), with 1608 participants involving: <ul style="list-style-type: none"> <li>• a current critical illness,[33,36,61]</li> <li>• hypoxemic conditions and/or cyanotic congenital heart disease,[40,42,43,55]</li> </ul>
<i>Adults</i>	25 (78.12%), with 4897 participants involving a variety of health conditions: <ul style="list-style-type: none"> <li>• healthy volunteers,[34,39,57,59,63]</li> <li>• critical illnesses or conditions needing intensive care unit admission and/or mechanical ventilation,[35,37,44-48,56]</li> <li>• pulmonary/respiratory conditions including COVID-19,[38,49,53,54,60,62]</li> <li>• cirrhosis,[31]</li> <li>• chronic rheumatic heart disease,[58]</li> <li>• postoperative hypothermia,[41]</li> <li>• hospitalised patients in general,[32,51]</li> <li>• adults under the need of a long-term home oxygen therapy[50]</li> </ul>
Sample sizes (32 studies)	Median 50 (range: 6 to 1562)

Age (32 studies)	
Mean or median specified (n = 23)	Median 56.40 years (range: 4 days to 69 years)
Factors related to skin pigmentation (32 studies)	
<i>Levels of skin pigmentation</i>	15 studies (46.88%), with 1800 participants
<i>Descriptors of ethnicity</i>	22 studies (68.88%), with 4910 participants

Note: \* The studies at the laboratory setting (n = 5) only recruited healthy volunteers.

The 32 studies (6505 participants) reported SpO<sub>2</sub>-SaO<sub>2</sub> comparison evaluations of 54 different pulse oximeters from 26 manufacturers. As the gold standard comparator measure, SaO<sub>2</sub> was directly measured using either one of six generic (or 14 specific) types of CO-oximetry methods, or one of two combinations of CO-oximetry methods (Supporting Information Table S3).

#### **Assessment of risk of bias and applicability**

Using QUADAS-2, we considered 14/32 studies (43.75%) to be at unclear risk of bias for all four domains or high risk of bias for at least one domain,[34,38,44,47,48,53,54,56-61,63] and the remaining 18 (56.25%) to be at low risk of bias for at least one of the four domains (Figure 2).

Key limitations in downgrading risk of bias were: (1) for the patient selection domain, where specific sub-populations were inappropriately excluded from a study[61] or there was no explicit exclusion or even inclusion criteria (19 studies); (2) for index test and reference standard domains, where there was no blinding information for either pulse oximetry SpO<sub>2</sub> measurements (20 studies) or CO-oximeter SaO<sub>2</sub> readings (30 studies); and (3) for the flow and timing domain, where there were sub-optimally long time intervals between SpO<sub>2</sub> readings and the arterial blood sampling for SaO<sub>2</sub> measurement,[44] or the exclusion of participants (3/25, 12.00%) from study data analysis without rationale.[59]

We judged the applicability concern as high for one study,[60] moderate for 13 studies,[33,36,37,40,42,43,47,48,51,56,57,59,62] and low in terms of all three applicability considerations for the remaining 18 studies. Applicability concerns largely resulted from the lack of specifications on pulse oximeters evaluated, CO-oximeter devices used, and/or arterial blood sampling procedures, thus SpO<sub>2</sub> or SaO<sub>2</sub> could not be reproduced.

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Abrams 2002	?	+	?	+	+	+	+
Adler 1998	+	+	?	?	+	+	+
Avant 1997	+	?	?	?	+	?	?
Bickler 2005	?	?	?	?	+	+	+
Bothma 1996	?	+	?	?	+	+	+
Brooks 2020	+	?	?	?	+	+	?
Ebmeier 2018	+	+	?	?	+	+	?
Escourrou 1990	?	?	?	?	+	+	+
Feiner 2007	+	?	?	+	+	+	+
Foglia 2017	?	?	+	?	+	?	+
Gabrielczyk 1988	?	+	?	+	+	+	+
Harris 2016	+	+	?	+	+	+	?
Harris 2019	+	+	?	+	+	+	?
Harskamp 2021	+	+	+	⊖	+	+	+
Hinkelbein 2006	+	+	?	+	+	+	+
Hinkelbein 2007	+	+	?	+	+	+	+
Jubran 1990	?	?	?	?	+	?	+
Lee 1993	?	?	?	?	?	?	?
McGovern 1996	?	?	?	+	+	+	+
Munoz 2008	?	?	?	+	+	+	+
Pilcher 2020	?	+	?	+	+	?	?
Ries 1985	?	?	?	?	+	+	+
Ries 1989	?	?	?	?	+	+	+
Ross 2014	+	+	?	+	+	+	+
Schallom 2018	?	?	?	?	+	+	?
Smyth 1986	?	?	?	?	+	+	?
Stewart 1991	?	?	?	?	+	+	+
Thrush 1994	?	?	?	⊖	?	+	+
Valbuena 2021	?	?	?	?	+	⊖	?
Vesoulis 2021	⊖	?	?	?	+	+	+
Wiles 2021	+	?	?	?	+	+	?
Zeballos 1991	?	?	?	?	+	+	+

⊖ High      ? Unclear      + Low

Figure 2. Risk of bias assessment results

### **Association between level of skin pigmentation and pulse oximetry accuracy**

Fifteen of the 32 studies (1800 participants) reported by level of skin pigmentation.[32,34,35,37-44,51-54,63] Eight of these studies (1297 participants) had available data and were included in the meta-analyses:[32,34,35,37,39,40,51,52,63] Supporting Information Table S4 presents the mapping of originally reported terms of skin pigmentation into 'low', 'medium' or 'high' pigmentation categories. Table 2 presents pooled association data (further data detail along with GRADE assessment details are found in Supporting Information Figures S1-S3 and Table S5).

**Table 2. Result summaries of meta-analysis for levels of skin pigmentation and ethnic groups**

Subgroup categories	No. of studies (evaluations)	Sample size (data pairs)	Calculated $A_{rms}$ , %	Pooled mean bias (95% CI), %	Pooled SD (95% CI), %	Calculated 95% limits of agreement, %	Overall $I^2$ (between-studies and within-study heterogeneity) in mean bias data pooling
<b>High (dark) skin pigmentation</b>							
Primary analysis	8 (24)	221 (3270)	1.88	1.11 (0.29 to 1.93)	1.52 (1.30 to 1.79)	-1.87 to 4.09	98.03% (0% and 98.03%)
Sensitivity analysis, excluding studies with high overall risk of bias	6 (15)	177 (2691)	1.75	0.87 (-0.46 to 2.19)	1.52 (1.20 to 1.93)	-2.11 to 3.84	–
Sensitivity analysis, excluding studies with no use of standardised scales for measuring skin pigmentation	5 (9)	160 (474)	1.79	0.89 (-1.37 to 3.14)	1.55 (1.12 to 2.15)	-2.16 to 3.93	–
Sensitivity analysis, excluding studies with participants of high skin pigmentation alone	6 (14)	88 (2738)	1.95	1.12 (-0.16 to 2.39)	1.60 (1.26 to 2.03)	-2.02 to 4.26	–
<b>Medium pigmentation</b>							
Primary analysis	4 (10)	406 (1323)	1.58	-0.58 (-2.25 to 1.09)	1.47 (1.08 to 2.00)	-3.46 to 2.30	92.65% (82.39% and 10.25%)
Sensitivity analysis, excluding studies with no use of standardised scales for measuring skin pigmentation	3 (4)	399 (399)	1.66	-0.79 (-3.03 to 1.45)	1.46 (1.00 to 2.14)	-3.66 to 2.08	–
<b>Low (light) skin pigmentation</b>							
Primary analysis	6 (15)	670 (2865)	1.53	-0.35 (-1.36 to 0.67)	1.49 (1.23 to 1.81)	-3.27 to 2.58	92.73% (22.42% and 70.31%)

Sensitivity analysis, excluding studies with high overall risk of bias	5 (12)	660 (2245)	1.57	-0.47 (-1.77 to 0.83)	1.50 (1.21 to 1.87)	-3.42 to 2.47	–
Sensitivity analysis (scale)	4 (6)	648 (667)	1.62	-0.54 (-2.52 to 1.43)	1.53 (1.18 to 1.98)	-3.53 to 2.45	–
<b>Participants described as Black/African American</b>							
Primary analysis	9 (22)	459 (5753)	2.27	1.52 (0.95 to 2.09)	1.68 (1.32 to 2.14)	-1.78 to 4.82	96.39% (0% and 96.39%)
Sensitivity analysis, excluding studies with high overall risk of bias	4 (10)	67 (2892)	1.99	1.47 (-0.21 to 3.16)	1.35 (1.08 to 1.69)	-1.17 to 4.11	–
Sensitivity analysis, excluding studies with data transformation	7 (20)	316 (3110)	2.26	1.55 (0.85 to 2.25)	1.64 (1.28 to 2.11)	-1.67 to 4.77	–
Sensitivity analysis, excluding studies with Black/African American populations alone	8 (16)	426 (5621)	2.35	1.63 (0.77 to 2.49)	1.69 (1.25 to 2.28)	-1.68 to 4.94	–
<b>Participants from other ethnic groups, described as Asian, Hispanic or mixed ethnicity</b>							
Primary analysis	3 (9)	522 (2646)	1.58	0.31 (0.09 to 0.54)	1.55 (0.53 to 4.53)	-2.72 to 3.35	47.95% (0% and 47.95%)
Sensitivity analysis, excluding studies with high overall risk of bias	2 (7)	41 (2165)	1.23	0.36 (-0.24 to 0.95)	1.18 (0.14 to 9.78)	-1.95 to 2.66	–
Sensitivity analysis, excluding studies with data transformation	2 (8)	488 (1405)	2.07	0.30 (-0.80 to 1.40)	2.04 (0.32 to 12.81)	-3.70 to 4.31	–
<b>White/Caucasian</b>							
Primary analysis	13 (48)	2195 (12870)	1.64	0.55 (-0.21 to 1.31)	1.55 (1.31 to 1.82)	-2.48 to 3.58	94.39% (69.92% and 24.47%)

Sensitivity analysis, excluding studies with high overall risk of bias	8 (26)	1293 (7770)	1.42	0.36 (-0.88 to 1.61)	1.38 (1.19 to 1.59)	-2.33 to 3.06	–
Sensitivity analysis, excluding studies with data transformation	10 (45)	1044 (5485)	1.67	0.63 (-0.38 to 1.63)	1.54 (1.32 to 1.80)	-2.40 to 3.65	–
Sensitivity analysis, excluding studies with White/Caucasian populations alone	8 (16)	1246 (9791)	1.84	0.98 (-0.08 to 2.05)	1.56 (1.16 to 2.10)	-2.07 to 4.04	–

Notes. Calculated accuracy root-mean-square ( $A_{rms}$ ) was derived based on the pooled mean bias and the pooled SD from related meta-analyses (Supporting Information Box S4). According to the British Standards Institution standards,[12] pulse oximetry with an  $A_{rms}$  larger than 4% is considered inaccurate.

The pooled mean bias was produced by meta-analysis of data on mean SpO<sub>2</sub>–SaO<sub>2</sub> bias. Similarly, the pooled SD was from meta-analysis of all SD data reported in included studies. According to a commonly used criteria (mean bias of +/- 2% (one SD)), [12] an accurate pulse oximetry should produce SpO<sub>2</sub> measures that are within +/- 2% of CO-oximetry measures and the variation of bias values for repeated SpO<sub>2</sub> measures should be within one SD.

The 95% limits of agreement was calculated using the method described by Bland and Altman,[22] based on the pooled mean bias, and the pooled SD (Supporting Information Box S4).

Full results of heterogeneity tests are in Supporting Information Figures S1-S3 and S5-S7.

People with high levels of pigmentation probably have overestimated oxygen saturation readings from hospital-based pulse oximetry compared with standard SaO<sub>2</sub> readings (8 studies, 24 comparisons, 3270 SpO<sub>2</sub>-SaO<sub>2</sub> pairs from 221 participants): pooled mean bias 1.11% (95% CI 0.29% to 1.93%), moderate-certainty evidence. This means that, on average, pulse oximetry probably overestimates blood oxygen saturation by approximately 1% in this group, but the overestimation may be as low as 0.29% or almost 2%. The evidence for people with medium skin pigmentation is uncertain, meaning further research is likely to alter findings. The evidence for people with light skin pigmentation does not suggest clinically important systematic bias (pooled mean bias -0.35, 95% CI -1.36 to 0.67), but it is of low certainty.

For all the levels of skin pigmentation, the  $A_{rms}$  values are around 2% or lower (95% CI non-estimable), and the pooled SD values are around 1.50% on average (Table 2). These data mean that, for people with any level of skin pigmentation, about 68% of their pulse oximetry readings would be within  $\pm 2\%$  of the CO-oximetry readings, with one SD indicating a variation around the mean bias of minus 1.50% to plus 1.50%.

We tested the sensitivity of the findings:  $A_{rms}$  and SD values were generally consistent but there was increased uncertainty for mean bias findings. Supporting Information Figure S4 presents evidence for different types of pulse oximeters evaluated: overall, most devices slightly overestimated oxygen saturation in people with high levels of skin pigmentation, with imprecision around estimates.

Seven of the 15 studies (with 503 participants) were not included in meta-analysis (Supporting Information Table S6):[38,41-44,53,54] six compared differences in mean bias between pigmentation levels, rather than reporting mean bias by levels of skin pigmentation. Where relevant and available we present data by device type in Supporting Information Table S6, there may be some variation in the amount of mean bias based on device used but data are too limited to draw further conclusions.

### **Association between ethnicity and pulse oximetry accuracy**

Twenty-two of the 32 studies (4910 participants) described participants by ethnicity rather than level of skin pigmentation[31,33,34,36,38,39,41,45-50,55-63] and we included 14 studies (3510 participants) in meta-analyses.[31,33,34,39,45-47,49,50,59,60-63]. Pooled data are shown in Table 2 (further data are reported in Supporting Information Figures S5-S7 and Table S5).

Oxygen saturation measured for people described in study reports as Black or African American may be overestimated using hospital pulse oximetry compared with standard SaO<sub>2</sub> readings: mean bias 1.52% (95% CI 0.95% to 2.09%), low-certainty evidence. The 95% confidence interval of this estimate ranges between an overestimation of 1% to 2%. The

evidence for people described in studies as Asian, Hispanic, or of mixed ethnicity does not indicate a clinically important systematic bias (mean bias 0.31%, 0.09% to 0.54%), but it is of low certainty. The evidence is uncertain for groups described in papers as White/Caucasian, meaning further research is likely to alter findings (very low certainty evidence).

In the studies which referred to ethnicity rather than skin pigmentation, the  $A_{rms}$  values are around 2% or lower (95% CI non-estimable) for all groups, and the pooled SD values are around 1.50% on average (Table 2). This means that for people from the ethnic groups studied, about 68% of their pulse oximetry readings would be within  $\pm 2\%$  of the CO-oximetry readings, with one SD indicating a variation around the mean bias of minus 1.50% to plus 1.50%.

We tested the sensitivity of the findings:  $A_{rms}$  and SD values were generally consistent but there was increased uncertainty for mean bias findings. Supporting Information Figure S8 presents evidence for each type of pulse oximeter evaluated: overall, most devices overestimated oxygen saturation in people described as Black or African American.

Eight of the 22 studies (1400 participants) were not included in meta-analysis (Supporting Information Table S7):[36,38,41,48,550-58] all had no accuracy data reported by ethnicity groups.

## Discussion

### Summary of findings

This review summarises the evidence from 32 studies which measure the impact of skin pigmentation (15 studies) or ethnicity (22 studies) on the accuracy of pulse oximetry. With mean bias as the primary outcome, the evidence suggests that for people with darkly pigmented skin and people described in study reports as Black or African American, oxygen saturation may be overestimated by pulse oximetry in hospital compared with gold standard SaO<sub>2</sub>. Whilst the existing evidence suggests that the degree of measurement bias (when rounded integers) is small at around the recommended criterion of 2% or less, this evidence is of low certainty and further confirmatory primary research is needed. It should also be emphasised that these results are for clinician-measured oximetry in controlled clinical environments and do not necessarily reflect the measurement bias of home pulse oximetry by patients or carers. The evidence suggests that pulse oximetry for people with other levels of skin pigmentation is less likely to be overestimated but the evidence is uncertain. The low certainty for much of the data presented means that further research could overturn these conclusions. For all the subgroups of populations evaluated, whilst the degree of mean bias is small or negligible, pulse oximetry readings appear unacceptably imprecise (pooled SDs are greater than the recommended criterion of 1%). Nevertheless, when the extents of measurement bias and precision are considered jointly in  $A_{rms}$ , pulse oximetry measurements

for all the subgroups appear acceptably accurate (with  $A_{rms} <$  the British Standards Institution recommended threshold of 4%).

### **Overall completeness and applicability of evidence**

Our search was comprehensive and up to date as of 14 December 2021 however, there are limitations in the completeness and applicability of the evidence identified.

Pulse oximetry is widely used in clinical practice and promoted for home use during the COVID-19 pandemic.[2] We are aware of many factors (such as oxygen saturation ranges at baseline, types of pulse oximeter probe, comorbidities, movement, age of the patient) which could theoretically affect pulse oximetry accuracy in the real world.[8] However, most of the included studies (27/32, 84.38%) in this review were based in hospital settings and this review only addresses skin pigmentation and ethnicity. Most included studies had limited information as to whether the pulse oximeters evaluated were appropriate for home self-monitoring. Therefore, little is known for the case of pulse oximetry undertaken by untrained people at home where other factors such as movement need to be considered.

Pulse oximeters have been developed and upgraded since 1970s,[65] and old devices might have been withdrawn. The included studies were published between 1985 and 2021 and some of the older studies may have used discontinued devices. Nevertheless, the overestimation of oxygen saturation for darker skin appears to be consistent in general across most devices evaluated. To keep the completeness of evidence in this review, we included study data for all pulse oximeter devices included.

We regarded mean bias as the primary outcome in this review given it is intuitive. We also present the results for the other outcomes recommended by British Standards Institution (overall accuracy and precision) for completeness.

### **Comparison with other studies**

To our knowledge, there is one existing systematic review (published in 1995) and a rapid review in this area.[6,8] The Jensen review used reference measures of  $SaO_2$  ( $PaO_2$ , calculated  $SaO_2$ , and  $\%O_2Hb$ ) that are now considered incorrect or outdated.[12,15] The review considered skin pigmentation as a factor affecting pulse oximetry accuracy, but did not address ethnicity-related evidence. With no estimation of the size of the bias, the review concluded that pulse oximeters may overestimate oxygen saturation for people with darker skin. This conclusion however is only based on one study with data available for synthesis regarding skin pigmentation.

The rapid review by NHS Race and Health Observatory came to similar conclusions regarding pulse oximetry accuracy, but its methods were not clear. Of all nine studies included in the rapid review, two had no data on  $SpO_2$ - $SaO_2$  comparisons: one with a diagnostic accuracy design,[7] and one comparing two types of pulse oximetry with each other[66]. This rapid

review did not differentiate between evidence for ethnicity and skin pigmentation, which is considered inappropriate.[14]

By comparison we identified 15 eligible studies exploring the influence of skin pigmentation and added the evidence of 22 studies which examined the influence of ethnicity. Importantly, we considered internationally recognised SaO<sub>2</sub> measured by the gold standard CO-oximetry as the comparator for pulse oximetry. Our review, with more studies, adds well-produced evidence to the NHS Race and Health Observatory guidance and other national guidelines with further insights into the current size and precision around current accuracy estimates as noted below.

The UK British Standards Institution 2019, originating from the 2017 standards of the International Organization for Standardization (ISO),[67] specify performance requirements of pulse oximeter devices:[12] they are  $A_{rms}$  within 4% (the US Food and Drug Administration recommend a more conservative threshold of 3%)[68] and the mean bias (precision) value of +/- 2% (one SD) for acceptable pulse oximetry. In relation to these, the review results suggest an acceptable level of  $A_{rms}$  and bias but unacceptable imprecision. However, there is a need to consider the uncertainty in the mean bias estimates.

Whilst this is a relatively small amount of mean bias its impact on clinical decision making could be significant at threshold values for diagnosis of hypoxaemia, leading to clinically important hypoxaemia remaining undetected and untreated. Underestimated SpO<sub>2</sub> readings also have the potential to be harmful, resulting in unnecessary treatment with oxygen (and the risk of hyperoxaemia) and wider impacts such as delayed hospital discharge. Two recent diagnostic studies provide evidence on clinical implications resulting from the bias in pulse oximetry for blood oxygen saturation levels[7,69]. In these studies people described as Black had a higher risk of '*occult hypoxemia that was not detected by pulse oximetry*' compared with those described as White.[7] This may suggest that even small amounts of mean bias, when at the margins of diagnostic thresholds, could have an impact on diagnostic accuracy. Further understanding of these impacts could be explored via evidence synthesis of diagnostic accuracy (classification) studies to assess the clinical implications of the findings in relation to clinical decision-making thresholds.

The bias of pulse oximetry in specific sub-populations might have further implications. A recent decision analytical Markov modelling study evaluated the cost-utility of remote pulse oximetry monitoring over a 3-week time horizon from a US health sector perspective.[70] It found that remote monitoring dominates current standard care, by saving USD \$11472 and gaining QALYs of 0.013. Besides, people with access to remote monitoring had 87% fewer hospitalisations and 77% fewer deaths. However, this study did not consider the pulse oximetry bias in specific sub-populations and future research is needed.

## Strengths and limitations of the review

We followed prespecified methods to summarise all relevant evidence to minimise the risk of bias in the review process.[11] We ran comprehensive electronic searches with the support of an information specialist, searched trial registries, and checked references of included studies and relevant systematic reviews identified. We used British Standards Institution recommended gold standard CO-oximetry as the comparator for pulse oximetry measures. We contacted study authors to clarify details and request study data. We developed a correlated hierarchical effects model and used RVE approaches to meta-analyse data.[21] This approach deals with correlations of multiple effect size estimates within a study, rather than treating them as independent (this treatment is common but inappropriate if conventional inverse variance analytical approach is chosen).[19] As a result of using RVE approaches, we included not only independent data (of 11 studies) in analyses but also more data from studies (n = 21) with repeated-measures design.

This review has some limitations. Firstly, seven studies considering the impact of skin pigmentation and eight studies considering ethnicity compared SpO<sub>2</sub>-SaO<sub>2</sub> bias data between different subgroups and presented only tests of significance results, rather than SpO<sub>2</sub> and SaO<sub>2</sub> data per se at each subgroup level. Also, at least two studies used diagnostic accuracy design that only presented proportions of participants with specific ranges of SpO<sub>2</sub> in relation to specific SaO<sub>2</sub> values, again rather than SpO<sub>2</sub> and SaO<sub>2</sub> data per se [7,69]. We contacted authors of these studies to request relevant data and received data for two studies.[39,51] If more data were received, then the results of this review could change.

Secondly, we are aware of the potential difference between the concepts of race and ethnicity and the recent call to distinguish them from each other.[71] For simplicity, we chose to use the term of 'ethnicity' throughout this review given race and ethnicity are context/country-specific concepts and there is no globally accepted racial/ethnic classification approaches to clearly distinguish them.[14] If we had treated race and ethnicity data separately, the evidence base would change slightly, however, we would not expect the overall conclusion to change. We also acknowledge that measurement of level of skin pigmentation is limited by the use of scales like the Fitzpatrick scale.[72] Such scales are criticised as being too blunt a measure of skin pigmentation, an issue that impacts on the findings of this review and which needs to be considered in research going forward.

Thirdly, in meta-analysis, we grouped data from studies where people were grouped by ethnicity and referred to as 'Asian', 'Hispanic' or mixed ethnicity into a single subgroup; mainly because the included studies mixed these ethnic groups together in this way. We were not able to produce evidence for each of these ethnic groups separately, although we acknowledge potential differences in measurement bias between them.

Fourthly, we did not consider the differences between specific pulse oximeter devices (54 models included), the differences between children and adults and their health conditions, or the difference in skin pigmentation measurement methods between studies (Table 1 and Supporting Information Table S4). In terms of pulse oximeters evaluated, there may be differences between devices for the use of health professionals in hospitals and those for home self-monitoring. Because of these, meta-analyses in this review demonstrated between-studies heterogeneity inevitably (Table 2). However, we found, across devices evaluated and types of participants, included studies were largely consistent in suggesting oxygen saturation overestimation of using pulse oximetry. We therefore chose to pool study data, without undertaking further subgroups for these differences.

Fifthly, we only searched for English language publications. However, there are probably no major differences between summary treatment effects in English-language restricted meta-analyses and other language-inclusive meta-analyses.[73] We did not search for literature in the format of preprints or resources other than trial registries given the limited time and resources available. We considered the possible publication bias in assessing the certainty of evidence using GRADE approach.

Finally, there is no existing approach to risk of bias and GRADE assessment in the case of this review. We used the QUADAS-2 and GRADE approaches developed for the test accuracy topic: both are relevant but not specific to this review. However, we were only able to assess the certainty of evidence for mean bias and found the existing GRADE approach inapplicable for precision,  $A_{rms}$  and limits of agreement.

## **Conclusions**

Pulse oximetry may overestimate blood oxygen saturation levels for people with dark skin compared with gold standard  $SaO_2$  measures. The extent of measurement bias appears relatively small, but pulse oximetry measurements appear accurate for all levels of skin pigmentation when the small or negligible bias and imprecision identified are considered jointly in the overall accuracy.

The evidence for the measurement bias identified however is uncertain. A small bias may be crucial for some patients: particularly at the threshold that informs clinical decision making. Health professionals should consider the potential bias of pulse oximetry readings for people with dark skin, and should also be aware that multiple potential factors could affect the accuracy of pulse oximeter readings.

## **Ethical approval**

Not required.

## **Data Availability:**

All relevant data are within the manuscript and its Supporting Information files. No additional data available.

### **Acknowledgments**

We thank the information specialist Catherine Harris for helping design search strategies and run electronic database searches in August 2021. We also thank Dr John Feiner and Dr Paul Young for sharing their study data and others who responded to our queries for clarifying study methods in deciding the eligibility of their studies.

### **Author Contributions:**

**Conceptualization:** Chunhu Shi, Mark Goodall, Jo Dumville, Andrew Clegg, Caroline Leigh Watkins, George Georgiou, Catherine Elizabeth Lightbody, Nicky Cullum

**Data Curation:** Chunhu Shi

**Investigation:** Chunhu Shi, Mark Goodall, James Hill, Gill Norman, Oliver Hamer

**Formal analysis:** Chunhu Shi, Jo Dumville, Alexander Hodkinson

**Funding acquisition:** Jo Dumville, Caroline Leigh Watkins, Nicky Cullum

**Methodology:** Chunhu Shi, Jo Dumville, James Hill, Gill Norman, Andrew Clegg, Alexander Hodkinson

**Project administration:** Chunhu Shi, Jo Dumville, James Hill

**Resources:** Jo Dumville, Andrew Clegg, George Georgiou

**Software:** Chunhu Shi

**Supervision:** Jo Dumville, Andrew Clegg, Caroline Leigh Watkins, Nicky Cullum

**Validation:** Alexander Hodkinson, Jo Dumville, Gill Norman

**Visualization:** Chunhu Shi

**Writing – original draft:** Chunhu Shi, Jo Dumville, Nicky Cullum

**Writing – review & editing:** Chunhu Shi, Mark Goodall, Jo Dumville, James Hill, Gill Norman, Oliver Hamer, Andrew Clegg, Alexander Hodkinson, Caroline Leigh Watkins, George Georgiou, Catherine Elizabeth Lightbody, Paul Dark, Nicky Cullum

### **Funding:**

The review is funded by the National Institute for Health Research Applied Research Collaboration (NIHR ARC) Greater Manchester and NIHR ARC North West Coast (ARC NWC). The views expressed in this article are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

### **Competing interests:**

The authors have declared that no competing interests exist.

## References

1. Swigris JJ, Zhou X, Wamboldt FS, Du Bois R, Keith R, Fischer A, et al. Exercise peripheral oxygen saturation (SpO<sub>2</sub>) accurately reflects arterial oxygen saturation (SaO<sub>2</sub>) and predicts mortality in systemic sclerosis. *Thorax*. 2009 Jul 1;64(7):626-30. Pmid: 19359269
2. NHS England. COVID Oximetry@home. <https://www.england.nhs.uk/nhs-at-home/covid-oximetry-at-home/>
3. World Health Organization (WHO). Covid-19 clinical management: living guideline (updated 25.1.21). 2021. <https://apps.who.int/iris/handle/10665/338882>
4. The Royal Australian College of General Practitioners. Managing COVID-19 at home with assistance from your general practice: A guide, action plan and symptom diary for patients. East Melbourne, Vic: RACGP, 2021.
5. Luks AM, Swenson ER. Pulse oximetry for monitoring patients with covid-19 at home: potential pitfalls and practical guidance. *Ann Am Thorac Soc*. 2020;17:1040-6. pmid: 32521167
6. NHS Race and Health Observatory. Pulse oximetry and racial bias: Recommendations for national healthcare, regulatory and research bodies. March 2021. <https://www.nhsrho.org/wp-content/uploads/2021/03/Pulse-oximetry-racial-bias-report.pdf>
7. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *N Engl J Med*. 2020 Dec 17;383(25):2477-8. pmid: 33326721
8. Jensen LA, Onyskiw JE, Prasad NG. Meta-analysis of arterial oxygen saturation monitoring by pulse oximetry in adults. *Heart Lung*. 1998 Nov 1;27(6):387-408. pmid:9835670
9. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2 (updated February 2021). Cochrane, 2021. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).
10. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021 Mar 29;372. pmid: 33781993
11. Shi C. The effects of skin pigmentation and race/ethnicity on the accuracy of pulse oximetry in measuring oxygen saturation: a systematic review and meta-analysis. 2021. <https://osf.io/gm7ty>
12. 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.

13. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988 Jun 1;124(6):869-71. pmid:3377516
14. Silver SE. Skin color is not the same thing as race. Arch Dermatol. 2004 Mar 1;140(3):361. pmid:15023784
15. Toffaletti J, Zijlstra WG. Misconceptions in reporting oxygen saturation. Anesth Analg. 2007 Dec 1;105(6):S5-9. pmid:18048899
16. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr. Proceedings of the 2<sup>nd</sup> ACM SIGHIT International Health Informatics Symposium. Miami, FL: 28–30 Jan 2012.
17. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014 Dec;14(1):1-3. pmid:25524443
18. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. pmid: 22007046.
19. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. 2010 Jan;1(1):39-65. pmid: 26056092
20. Tipton E, Shuster J. A framework for the meta-analysis of Bland–Altman studies based on a limits of agreement approach. Stat Med. 2017 Oct 15;36(23):3621-35. pmid:28664537
21. Pustejovsky JE, Tipton E. Meta-analysis with Robust Variance Estimation: Expanding the range of working models. Prev Sci. 2021;7:1-4. pmid:33961175
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i:307–310 pmid:2868172
23. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
24. Pustejovsky J. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.5.3. 2021. <https://CRAN.R-project.org/package=clubSandwich>
25. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 2010, 36(3), 1-48. <https://doi.org/10.18637/jss.v036.i03>
26. Gordon M, Lumley T. forestplot: Advanced Forest Plot Using ‘grid’ Graphics. R package version 2.0.1. 2021. <https://CRAN.R-project.org/package=forestplot>
27. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SwiM) in systematic reviews: reporting guideline. BMJ. 2020 Jan 16;368. pmid:31948937

28. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. *J Clin Epidemiol*. 2020 Jun 1;122:129-41. pmid:32060007
29. Schünemann HJ, Mustafa RA, Brozek J, Steingart KR, Leeflang M, Murad MH, et al. GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2020 Jun 1;122:142-52. pmid:32058069
30. Wolfgang (<https://stats.stackexchange.com/users/1934/wolfgang>), Metafor package: bias and sensitivity diagnostics. 7 June 2015 [cited 05 January 2022]. In: Stack Exchange [Internet]. Available from: <https://stats.stackexchange.com/q/155875>.
31. Abrams GA, Sanders MK, Fallon MB. Utility of pulse oximetry in the detection of arterial hypoxemia in liver transplant candidates. *Liver Transpl*. 2002 Apr 1;8(4):391-6. pmid:11965585
32. Adler JN, Hughes LA, Vtvilecchia R, Jr CA. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Acad Emerg Med*. 1998 Oct;5(10):965-70. pmid:9862586
33. Avant MG, Lowe N, Torres Jr A. Comparison of accuracy and signal consistency of two reusable pulse oximeter probes in critically ill children. *Respiratory Care*. 1997;42(7):698-704.
34. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology*. 2005 Apr 1;102(4):715-9. pmid:15791098
35. Bothma PA, Joynt GM, Upman J, Hon H, Mathala B, Scribante J, et al. Accuracy of pulse oximetry in pigmented patients. *S Afr Med J*. 1996;86(5). 8914569
36. Brooks JC, Raman S, Gibbons K, Williams T, Dunster KR, Pegg D, et al. Transcutaneous Oxygen Saturation Accuracy in Critically Ill Children. *Research Square*; 2020. DOI: 10.21203/rs.2.21938/v1.
37. Ebmeier SJ, Barker M, Bacon M, Beasley RC, Bellomo R, Chong CK, 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 May;46(3):297-303. pmid:29716488
38. Escourrou PJ, Delaperche MF, Visseaux A. Reliability of pulse oximetry during exercise in pulmonary patients. *Chest*. 1990 Mar 1;97(3):635-8. pmid:2106411
39. 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 Dec 1;105(6):S18-23. pmid:18048893

40. Foglia E, Whyte R, Chaudhary A, Mott A, Chen J, Propert K, et al. Accuracy and precision of pulse oximetry in hypoxemic infants. *J Pediatr* 2017;182:375-7
41. Gabrielczyk MR, Buist RJ. Pulse oximetry and postoperative hypothermia: An evaluation of the Nellcor N-100 in a cardiac surgical intensive care unit. *Anaesthesia*. 1988 May;43(5):402-4. pmid:3400853
42. Harris BU, Char DS, Feinstein JA, Verma A, Shiboski SC, Ramamoorthy C. Accuracy of pulse oximeters intended for hypoxemic pediatric patients. *Pediatr Crit Care Med* 2016; 17:315–320. pmid:26914626
43. Harris BU, Stewart S, Verma A, Hoen H, Stein ML, Wright G, et al. Accuracy of a portable pulse oximeter in monitoring hypoxemic infants with cyanotic heart disease. *Cardiol Young*. 2019 Aug;29(8):1025-9. pmid:31304897
44. Harskamp R, Bekker L, Himmelreich J, De Clercq L, Karregat E, Sleeswijk M, et al. Performance of popular pulse oximeters compared with simultaneous arterial oxygen saturation or clinical-grade pulse oximetry: a cross-sectional validation study in intensive care patients. *BMJ Open Respir Res*. 2021 Sep;8(1):e000939. pmid: 34489238
45. Hinkelbein J, Genzwuerker HV, Sogl R, Fiedler F. Effect of nail polish on oxygen saturation determined by pulse oximetry in critically ill patients. *Resuscitation*. 2007;72(1):82-91. pmid:17098347
46. Hinkelbein J, Koehler H, Genzwuerker HV, Fiedler F. Artificial acrylic finger nails may alter pulse oximetry measurement. *Resuscitation*. 2007;74(1):75-82. pmid:17353079
47. Jubran A, Tobin MJ. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. *Chest* 1990; 97: 1420–5. pmid:2347228
48. Lee KH, Hui KP, Tan WC, Lim TK. Factors influencing pulse oximetry as compared to functional arterial saturation in multi-ethnic Singapore. *Singapore Med J* 1993; Vol 34: 385-7. pmid:8153680
49. 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 May 1;109(5):1151-5. pmid:8625659
50. 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 Oct 1;32(4):1053-9. pmid:18480106
51. 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 Dec;20(1):1-9. pmid:31918697

52. Ploen L, Pilcher J, Beckert L, Swanney M, Beasley R. An investigation into the bias of pulse oximeters. *Respirology*. 2016, 21 (Suppl 2), 6.
53. Ries AL, Farrow JT, Clausen JL. Accuracy of two ear oximeters at rest and during exercise in pulmonary patients. *Am Rev Respir Dis*. 1985 Sep;132(3):685-9. pmid:4037542
54. Ries AL, Prewitt LM, Johnson JJ. Skin color and ear oximetry. *Chest*. 1989 Aug 1;96(2):287-90. pmid:2752811
55. Ross PA, Newth CJ, Khemani RG. Accuracy of pulse oximetry in children. *Pediatrics*. 2014 Jan 1;133(1):22-9. pmid:24344108
56. Schallom M, Prentice D, Sona C, Arroyo C, Mazuski J. Comparison of nasal and forehead oximetry accuracy and pressure injury in critically ill patients. *Heart Lung*. 2018 Mar 1;47(2):93-9. pmid:29402444
57. Smyth RJ, D'Urzo AD, Slutsky AS, Galko BM, Rebuck AS. Ear oximetry during combined hypoxia and exercise. *J Appl Physiol*. 1986 Feb 1;60(2):716-9. pmid:3949671
58. Stewart KG, Rowbottom SJ. Inaccuracy of pulse oximetry in patients with severe tricuspid regurgitation. *Anaesthesia*. 1991 Aug;46(8):668-70. pmid:1887977
59. Thrush D, Hodges MR. Accuracy of pulse oximetry during hypoxemia. *South Med J*. 1994 Apr 1;87(4):518-21. pmid:8153783
60. Valbuena VS, Barbaro RP, Claar D, Valley TS, Dickson RP, Gay SE, et al. Racial bias in pulse oximetry measurement among patients about to undergo ECMO in 2019-2020, a retrospective cohort study. *Chest*. 2021;S0012-3692(21)04065-4. pmid:34592317
61. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J Perinatol*. 2021;42(1):79-85 . pmid:34642469
62. 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*. 2021;77(2):143-152. pmid:34542168
63. Zeballos RJ, Weisman IM. Reliability of noninvasive oximetry in black subjects during exercise and hypoxia. *Am Rev Respir Dis*. 1991;144:1240-4. pmid:1741533
64. Blanchet MA, Mercier G, Bouchard PA, Rousseau E, Lellouche F. Accuracy of pulse oximetry (SpO<sub>2</sub>) with different oximeters. *Oxygap study*. *Intensive Care Medicine Experimental* 2021, 9(Suppl 1):51.
65. Severinghaus JW, Honda Y. History of blood gas analysis. VII. Pulse oximetry. *J Clin Monit*. 1987 Apr;3(2):135-8. pmid:3295125

66. Smith RN, Hofmeyr R. Perioperative comparison of the agreement between a portable fingertip pulse oximeter v. a conventional bedside pulse oximeter in adult patients (COMFORT trial). *S Afr Med J*. 2019 Mar 1;109(3):154-8. pmid:30834870
67. International Organization for Standardization. Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment (ISO 80601-2-61:2017). <https://committee.iso.org/standard/67963.html>
68. Food and Drug Administration. Pulse Oximeters – Premarket Notification Submissions: Guidance for Industry and Food and Drug Administration Staff. [https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM081352pdf\\_2013](https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM081352pdf_2013). Accessed 10 February 2022.
69. Wong AK, Charpignon M, Kim H, Josef C, de Hond AA, 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. pmid:34730820
70. Padula WV, Miano MA, Kelley MA, Crawford SA, Choy BH, Hughes RM, et al. A cost-utility analysis of remote pulse-oximetry monitoring of patients with COVID-2019. *Value in Health*. 2021 Oct 23. <https://doi.org/10.1016/j.jval.2021.09.008>
71. Flanagan A, Frey T, Christiansen SL, Bauchner H. The reporting of race and ethnicity in medical and science journals: comments invited. *JAMA*. 2021 Mar 16;325(11):1049-52. pmid:33616604
72. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial limitations of Fitzpatrick skin type. *Cutis*. 2020 Feb;105(2):77-80. pmid: 32186531.
73. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care*. 2012 Apr;28(2):138-44. pmid:22559755