# The impact of skin tone on the performance of pulse oximeters used by the National Health Service COVID oximetry @home scheme: a measurement and diagnostic accuracy study

Daniel S Martin (0000-0001-6220-8235), James C Doidge (0000-0002-3674-3100), Doug Gould (0000-0003-4148-3312), Tasnin Shahid (0000-0003-4320-6140), Alex Cowden (0009-0000-3776-8601), Walton N Charles (0000-0002-6012-5783), Amelia Francis Johnson (0009-0007-4515-4299), Roger Garrett, Catherine Mbema, Olusegun Olusanya (0000-0001-9667-658X), Eugene Healy (0000-0001-5591-6970), Kathryn Rowan (0000-0001-8217-5602), Paul Mouncey (0000-0002-8510-8517), David A Harrison (0000-0002-9002-9098)

On behalf of the EXAKT Study Investigators†


Daniel S Martin

Professor

Peninsula Medical School, University of Plymouth, John Bull Building, Plymouth Science Park, Research Way Plymouth PL6 8BT, UK

Intensive Care Unit, University Hospitals Plymouth NHS Trust, Plymouth, UK


James Doidge

Lead Statistician

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier House, 24 High Holborn, London WC1V 6AZ, UK


Doug Gould

Lead Researcher

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier House, 24 High Holborn, London WC1V 6AZ, UK

Accepted manuscript

Tasnin Shahid
Senior Trial Manager

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier
House, 24 High Holborn, London WC1V 6AZ, UK


Alex Cowden
Statistician

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier
House, 24 High Holborn, London WC1V 6AZ, UK


Walton N Charles
Research Associate

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier
House, 24 High Holborn, London WC1V 6AZ, UK


Amelia Francis Johnson
Trial Coordinator

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier
House, 24 High Holborn, London WC1V 6AZ, UK


Roger Garrett
Independent patient advisor

Independent patient advisor, Bristol, UK


Catherine Mbema
Director of Public Health

Lewisham Council, London, UK

Accepted manuscript

Olusegun Olusanya
Intensive Care Consultant

Department of Perioperative Medicine, St Bartholomew's Hospital, London, UK


Eugene Healy
Professor

Dermatopharmacology, University of Southampton, Southampton General Hospital, Southampton, UK

Dermatology, University Hospital Southampton NHS Foundation Trust, Southampton, UK


Kathryn Rowan
Scientific Advisor

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier House, 24 High Holborn, London WC1V 6AZ, UK


Paul Mouncey
Co-Director

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier House, 24 High Holborn, London WC1V 6AZ, UK


David A Harrison
Co-Director

Clinical Trials Unit, Intensive Care National Audit & Research Centre (ICNARC), Napier House, 24 High Holborn, London WC1V 6AZ, UK


†See the supplementary appendix for a full list of EXAKT Study Investigators


Correspondence to:

Professor Daniel Martin (0000-0001-6220-8235)
3

Peninsula Medical School

John Bull Building

Plymouth Science Park

Research Way

Plymouth

PL6 8BT

United Kingdom

daniel.martin@plymouth.ac.uk

† the EXAKT Study Investigators:

Katie Sweet, Daniel Hill, Linda Pipira, Oluwatosin Komolafe, Marcela P Vizcaychipi, Sanjeet Chana, Laura Martins, Ilhan Wardhere, Dhinesh Sundaran, Laszlo Hollos, Anna Williams, David Golden, Rebecca Seaman, Miriam Davey, Laura Kent, Nazrudeen Ali, Ibrahim Nasr, Dee Durrant, Bincy Kariyadil, Brendan Sloan, Sarah Buckley, Amy Major, Mathew Smith, Dumindu Karangoda, Estefania Treus, Georgina Luck, Shilpa Gurung, Mandeep K Phull, Nauman Hussain, Tatiana Pogreban, Aparna George; Nicholas Truman, Qasim T Ahmed, Victoria Cunliffe, Aayesha Kazi, Mark De Neef, Poh Choo Teoh, Glykeria Pakou, Amitaa Maharajh, Sarah Penkett, Muthukumaran Gourishankar, Paul Gill, Roxanne Gray, Thomas Billyard, Carl Hawkins, Brendan B Spooner, Geraldine Ward, Monica Popescu, Theodora Christodoulopoulou, Amrinder Sayan, Jamie Gonzales, Redmond P Tully, Joy Dearden, Michelle Mulcahy, Carrole Blessen, Andrew Achilleos, Rosaleeta Reece-Anthony, Christine B Manalo, Kay Ellen Spikes, Ahilanandan Dushianthan, Michael Carter, Karen Salmon, Rachel Burnish, Christopher Hebbes, Joanna Shak, Megha Mathews, Navneet Ghuman, Alex Mattin, Geraldine Hambrook, Tom O Lawton, Omar Jundi, Sharwend Supermanian, Louise T Akeroyd, Thaventhran Prabhahar, Dhanalakshmi Bakthavatsalam, Suhieb Alhourani, Leena

John, Mike Dean, Peter Mathews, Tabassum Khan; Lakshmi Aneesh, Dagan Lonsdale, Susannah Leaver, Sarah Farnell-Ward, Deborah Dawson, Nikki Yun, Maria Thanasi, Shreeja Dangol, Massimiliano Valcher, Angelo Rocha, Milafe Nimer, Frenita Dsouza, Rebecca Kanu, Maria Maiz Cordoba, Vincent Ventura, Edna Fernandes, Karen Lloyd, John Adam, Rohit Saha, Tom Williams, Kevin O'Reilly, Anna Broderick.

## Abstract

### Objectives

To assess the impact of skin tone on the measurement and diagnostic accuracy of five fingertip pulse oximeters used by patients in the NHS England COVID oximetry @home scheme.

### Design

Multi-centre diagnostic accuracy study (EXAKT).

### Setting

Twenty-four intensive care units in England between June 2022 and August 2024.

### Participants

903 critically ill adults on intensive care units screened for or enrolled into a trial evaluating different approaches to oxygen therapy.

### Interventions

Pulse oximetry-derived peripheral oxygen saturation ($SpO_2$) measurements were compared with paired arterial oxygen saturation ($SaO_2$) measurements from arterial blood analysed by co-oximetry (gold-standard). Skin tone (individual typology angle, ITA) was objectively measured using a hand-held spectrophotometer.

### Main outcome measures

Pulse oximeter measurement accuracy was assessed for bias, precision and overall accuracy. Diagnostic accuracy for identifying $SaO_2$ ≤92% was assessed by false negative and positive rates for $SpO_2$ using thresholds of ≤92% and ≤94%, and the area under the receiver operating characteristic curve, and by presence of occult hypoxaemia ($SaO_2$ <88% with $SpO_2$ >92%).

### Results

11,018 paired $SpO_2$-$SaO_2$ measurements were analysed. All tested pulse oximeters overestimated at lower values and underestimated at higher values of $SaO_2$. $SpO_2$ readings were, on average, 0.6–1.5 percentage points higher for patients with darker (ITA −44°) than lighter (ITA 46°) skin tone. At both $SpO_2$ thresholds assessed, false negative rates increased with darker skin tones; the proportion of $SpO_2$ measurements >94% despite a paired $SaO_2$ ≤92% ranged from 5.3–35.3 percentage points higher for patients with darker than lighter

6

skin tone (7.6–62.2% vs 1.2–26.9%, rate ratio 2.3–7.1). Conversely, false positive rates decreased with darker skin tones.

## Conclusions

Five pulse oximeters, provided by the NHS England COVID oximetry @home scheme, yielded higher $SpO_2$ measurements for patients with darker, compared with lighter, skin tones that could translate into potentially clinically important differences in false negative and positive rates for detecting hypoxaemia.

## Trial registration

ClinicalTrials.gov NCT05481515

# Summary

**What is already known on this topic**

- Pulse oximeters have been associated with overestimation of arterial haemoglobin oxygen saturation in people with darker skin tones.
- Results from individual studies have been inconsistent and have often been limited by using ethnicity as a proxy for skin tone.
- Overestimation of arterial haemoglobin oxygen saturation could represent a significant clinical risk for patients with darker skin tones.

**What this study adds**

- Using spectrophotometry to generate an objective measure of skin tone, this study showed that all five pulse oximeters assessed gave higher $SpO_2$ readings for patients with darker skin tones than for patients with lighter skin tones.
- Current standards for performance of pulse oximeters based on accuracy root mean square may not be adequate to differentiate bias from precision.
- Even small variations in bias translated into substantial differences in false positive and false negative rates for diagnostic accuracy.

## Introduction

Arterial haemoglobin oxygen saturation ($SaO_2$) indicates the fraction of oxygenated haemoglobin, relative to total haemoglobin, in arterial blood and provides a useful indicator of hypoxaemia (low blood oxygen levels). The gold-standard measurement for $SaO_2$ is an arterial blood sample analysed by the co-oximeter of a blood gas machine. Pulse oximetry is a technology developed to estimate $SaO_2$ non-invasively by measuring the absorbance of two wavelengths of light (660 and 940 nm) transmitted across the vascular bed of a tissue. Differential light absorbance characteristics between oxygenated and deoxygenated haemoglobin permits the calculation of a ratio, allowing estimation of the $SaO_2$. Pulse oximetry is widely used throughout the world to monitor patients both in healthcare settings and, more recently, in the community.

Since pulse oximeters first became commercially available in the 1980s, advances in their design have led to the development of low-cost, battery-powered, portable devices that can be used at home by the public. During the COVID-19 pandemic, these fingertip pulse oximeters were supplied to people at home to enable remote monitoring of peripheral oxygen saturation ($SpO_2$) and support early detection of hypoxaemia.[1] In England, the National Health Service (NHS) introduced the COVID oximetry @home scheme. The effectiveness of such schemes has been difficult to evaluate retrospectively due to the complexity of the pandemic and multiple confounding factors.[2-4] Some low-cost pulse oximeters have been shown to be inaccurate, with their accuracy falling short of that required for regulatory approval.[5] Studies have also demonstrated that $SaO_2$ is overestimated by pulse oximetry in people with darker skin tones.[6] Overestimation of $SaO_2$ by pulse oximetry could lead to the late or absent diagnosis of hypoxaemia and subsequent delay or failure to initiate life-saving therapies.[7] The magnitude of any overestimation has been difficult to quantify due to the limitations of studies published to date. Many studies have used ethnicity or race as a proxy for skin tone, while others have assessed skin tone using the Fitzpatrick skin type or other colour comparison charts. Both methodological approaches have significant limitations that can be avoided by obtaining objective measurements of skin tone using spectrophotometry.[8,9]

The Exploring pulse oXimeter Accuracy across sKin Tones (EXAKT) study was commissioned and rapidly funded by the National Institute for Health and Care Research (NIHR) in response to concerns raised by the NHS in England about inaccuracies in pulse oximeters (HTA 21/608, 2022). The aim was to assess the measurement and diagnostic accuracy of five fingertip pulse oximeters provided by the NHS for use at home in the NHS

England COVID oximetry @home scheme and to assess the impact of skin tone on accuracy.

## Methods

### Sites and participants

The EXAKT study took advantage of the screening infrastructure for an ongoing, large multicentre randomised clinical trial, the UK-ROX trial.[10] Participants were adult patients (18 years and older) receiving mechanical ventilation and supplementary oxygen in 24 NHS adult intensive care units in England. Patients were recruited to the EXAKT study from those both enrolled into, and those screened but not enrolled into, the UK-ROX trial. Selection was weighted to ensure over-representation of patients with darker skin tones (see supplementary methods). Ethical approval was granted by the South Central Oxford C Research Ethics Committee (20/SC/0423) and the EXAKT protocol was prospectively registered with ClinicalTrials.gov (NCT05481515). Full details of eligibility criteria and consent procedures are provided in the supplementary methods. Patient data recorded for the EXAKT study were linked with routinely collected data from the ICNARC Case Mix Programme, the national clinical audit of adult critical care, to provide information on patient characteristics. No aspect of patients' clinical care was altered as a result of this study. Readings from the study pulse oximeters being evaluated were not used to base clinical decisions on. A separate pulse oximeter linked to a critical care monitor remained attached each patient throughout the study for clinical decision making. The critical care pulse oximeters were not evaluated for accuracy in this study.

### Pulse oximeters

Five battery-powered, fingertip pulse oximeters, provided by NHS England as those in use in the COVID oximetry @home scheme,[4] were evaluated: ChoiceMMed Oxywatch MD300C19 (FBC3776), ChoiceMMed Oxywatch MD300C13-CC12 (FBC3826), Creative Medical PC-60B1-BL (FBC3828), Biolight Meditech M70C (FBC 3831), Medlinket AM802 NHE (FBC3830). Two different pulse oximeter models were randomly allocated to each participating site. As the EXAKT study progressed, pulse oximeter use was regularly reviewed and oximeters moved around participating sites to ensure sufficient recruitment for each oximeter.

### Measurements of arterial oxygenation

The index test measure was pulse oximetry-derived $SpO_2$ and the reference standard was functional $SaO_2$ measured by the co-oximeter of a standard blood gas machine (gold-

standard). Contemporaneous, paired $SpO_2$-$SaO_2$ measurements were collected over a 24 hour period for each patient. Immediately prior to routine arterial blood gas (ABG) sampling from an indwelling arterial catheter, each of the two study pulse oximeters were placed on fingers of the same hand of the patient by a member of the clinical team. Details of placement of pulse oximeters are provided in the supplementary methods. After a period of stabilisation, the $SpO_2$ value from each device was recorded and the ABG sample was immediately taken. $SaO_2$ values from the co-oximeter were not corrected for patient temperature. A list of blood gas machines used in the study is provided in the supplementary methods.

**Measurement of skin tone**

Skin tone was measured, for each patient, using a hand-held spectrophotometer (Konica Minolta CM-600d; see supplementary methods). In brief, four measurements were taken from normal skin on the dorsal surface of each patient's non-dominant hand, making sure to avoid any pigmented lesions on the skin. The International Commission on Illumination (CIE) L*, a*, and b* values for each reading were recorded and used to calculate the individual typology angle (ITA) as an objective continuous measure of skin tone, using the median L* and b* value for each patient.[11] Previously defined categories of skin tone based on ITA were used in descriptive results (very light, ITA ≥ 55°; light, 41° ≤ ITA < 55°; intermediate, 28° ≤ ITA < 41°; tan, 10° ≤ IT < 28°; brown, −30° ≤ITA < 10°; dark, ITA < −30°).[11,12] For simplicity when describing results, some comparisons are made between patients with median dark (the observed median ITA among patients with ITA < −30°) and median very light or light skin tone (median among patients with ITA ≥ 41°).

**Assessment of accuracy of pulse oximeters**

Measurement accuracy was assessed by bias (systematic error), precision (random error, or noise) and their combination for overall accuracy (assessed by accuracy root mean square, $A_{RMS}$) (see supplementary methods for definitions). Diagnostic accuracy was assessed by false negative rate (one minus sensitivity) and false positive rate (one minus specificity) of $SpO_2 \leq 92\%$ and $SpO_2 \leq 94\%$ for detecting $SaO_2 \leq 92\%$, and their combination as the area under the receiver operating characteristic curve (AUC) across all possible thresholds of $SpO_2$ for detecting $SaO_2 \leq 92\%$. $SpO_2$ thresholds were selected to correspond with the COVID oximetry @home guidance either to seek medical help (≤ 94%) or to attend the emergency department (≤ 92%).[13] 'Occult hypoxaemia' was assessed as the proportion of patients with $SaO_2 < 88\%$ who had $SpO_2 > 92\%$. The impact of skin tone (ITA) on performance was assessed for all seven metrics.

## Statistical analysis

Simulations informed by data from the UK-ROX trial and Case Mix Programme indicated that a sample of 900 patients yielding an anticipated 10,800 paired $SpO_2$-$SaO_2$ measurements would give at least 90% power to detect a linear variation in bias by skin tone corresponding to a 1.5 percentage point difference in bias between ITA values of 28° (lower boundary of the 'intermediate' category) and −30° (upper boundary of the 'dark' category).

For each accuracy metric, a single fractional probit regression model was fitted to the set of complete observations, adjusted for fixed effects of pulse oximeter model, ITA, $SaO_2$ and haemoglobin concentration, and interactions between pulse oximeter model and ITA, pulse oximeter model and $SaO_2$, and ITA and $SaO_2$. ITA, $SaO_2$, and haemoglobin concentration were modelled continuously using restricted cubic splines. To assess bias, this model was fitted with $SpO_2$ as the dependent variable. To assess precision, this model was fitted with the squared deviation between $SpO_2$ and the marginal predicted $SpO_2$ from the bias model as the dependent variable. To assess accuracy, this model was fitted with the squared difference between $SpO_2$ and paired $SaO_2$ as the dependent variable to estimate $A_{RMS}$. Models included fixed effects of site, and standard errors that account for repeated measures within patients.

To estimate false negative and false positive rates, multilevel logistic regression models were fitted with binary dependent variables of $SpO_2 \leq 92\%$ and $SpO_2 \leq 94\%$. Other than the addition of random effects for patients within sites, these models had identical structure to the models above. AUC was estimated using a probit ROC regression model with a dependent variable of $SaO_2 \leq 92\%$, independent variable of $SpO_2$ and adjusted for fixed effects of pulse oximeter model, ITA, and interactions between pulse oximeter model and ITA. The rate of occult hypoxaemia was estimated as the marginal predicted probability of $SpO_2 > 92\%$ averaged over the distribution of $SaO_2$ among patients with $SaO_2 < 88\%$.

See the supplementary methods for full details of statistical analyses and regression model outputs. Full analysis code can be viewed at https://github.com/ICNARC/EXAKT. Analyses were conducted using Stata/MP Version 18.0.

## Patient and public involvement

A patient advisor with lived experience was involved with the design of the study, the funding application (as a named applicant), the conduct of the study (as a member of the trial management group) and the editing of this manuscript (as a named author - RG). The study was designed following the NIHR INCLUDE Ethnicity Framework.[14] Given the nature of the research, we involved a public health professional (CM) with experience leading community

involvement work with Black, Asian and other minoritised communities during the COVID-19 pandemic. She helped our research team to form a community engagement partnership with Liberating Knowledge, who share experiences of belonging to groups racialised as minorities. Liberating Knowledge has enabled us to reach and communicate with communities who are likely to be most affected by the results of this study. We have worked with them to understand the current issues surrounding the area we studied and how to disseminate these new findings in a way that they are meaningful to those they may affect. Specifically, through Liberating Knowledge, we identified two public representatives to join our study management group meetings and provided input into the oversight of the study conduct. We hosted co-design workshops with members of voluntary, community and social enterprise organisations, research teams recruiting into the EXAKT study and members of the community. These co-design workshops aimed to explore the challenges faced in recruiting people from Black and South Asian communities and explore opportunities to raise awareness and identify approaches to increase the engagement of the wider community and identify the best ways to disseminate the results of the study. In addition to these activities, we produced regular blog posts focussed on the study and our engagement activities and presented at Health and Racial Justice Labs to raise awareness of the study more widely. Further to our work with Liberating Knowledge,  during the design phase of the study, we worked alongside the Centre for Ethnic Health Research, and hosted focus groups with public representatives to ensure we approached potential participants and their families appropriately and to review the information provided to patients recruited into EXAKT. As our study focused on patients with darker skin tones, we needed to understand how the patients and their relatives would feel about being approached to participate in a study related to skin tone.

## Results

### Participant characteristics

Between June 2022 and August 2024, 903 patients were recruited from 24 sites and included in the analysis (Supplementary figure 1; 1 patient was omitted from statistical models due to missing haemoglobin in all observations). Patient characteristics are summarised in table 1 and supplementary table 1. Of the 903 patients, 848 had linked routinely collected data available (94%; 566 male, 282 female; mean age 56 years); 232 (31%) were from a White ethnic group (lower than the UK population due to weighted sampling), 227 (31%) Asian, 177 (24%) Black, and 101 (14%) mixed or other ethnic group (ethnic group not available or not stated for 111 patients).

**Summary of key measures**

A total of 5,656 measurements of $SaO_2$ were collected, yielding 11,018 observation points when paired with synchronous $SpO_2$ measurements from up to two pulse oximeters per patient. Median $SaO_2$ was 96% (interquartile range 94 to 97%) and 891 (16%) measurements were of $SaO_2 \leq 92\%$ (table 1). The range of participant skin tones is shown in figure 1 and Supplementary figure 2, described by ITA and skin tone category. The median within-patient standard error of ITA estimation was 3 degrees.

[Table 1 – Patient characteristics and arterial blood gas observations]

[Figure 1 – Distribution of skin tone in the EXAKT study cohort, by pulse oximeter]

The relationship between $SpO_2$ and $SaO_2$ is illustrated in figure 2 and Supplementary figures 3A–E. All pulse oximeters tended to overestimate at lower values of $SaO_2$ and underestimate at higher values of $SaO_2$ (as indicated by the solid orange regression lines position above or below the dashed green lines of exact agreement), but with substantial imprecision at all levels of $SaO_2$ (as indicated by the scattered observation points). There was variation across the pulse oximeters in the degree of over- and underestimation, and the tipping point at which overestimation switched to underestimation (where the regression line crosses the line of exact agreement).

[Figure 2 – Relationship between $SpO_2$ and $SaO_2$, by pulse oximeter]

**The effect of skin tone on pulse oximeter accuracy**

All pulse oximeters tended to give higher $SpO_2$ readings at darker skin tones (figure 3, Supplementary tables 2–4 and Supplementary figures 4–7). For any given $SaO_2$ across the range studied, $SpO_2$ readings were on average 0.6–1.5 percentage points higher for a patient with median 'dark' skin tone (ITA −44°) than for a patient with median 'light' skin tone (ITA 46°) (Supplementary table 2). However, because all pulse oximeters tended to

underestimate high saturations and overestimate low saturations, the higher readings for darker skin tones could translate into either increasing positive bias or decreasing negative bias, depending on the $SaO_2$ (worse bias for patients with low saturation but potentially less bias for patients with higher saturations; see Supplementary figure 4). Consequently, there were no clear and consistent patterns in accuracy, as assessed by $A_{RMS}$, across skin tones (Supplementary figures 6).

[Figure 3 – Variation in measurement accuracy at $SaO_2$ 92% across a range of skin tones]

[Figure 4 – Variation in diagnostic accuracy to detect $SaO_2 \leq$ 92% across a range of skin tones]

**The effect of skin tone on false positive and false negative rates for detecting hypoxaemia**

For all pulse oximeters and at both thresholds assessed ($SpO_2 \leq$ 92% and $SpO_2 \leq$ 94%), the false negative rate for identifying $SaO_2 \leq$ 92% increased with darker skin tone (figure 4, Supplementary tables 5–8 and Supplementary figure 8). For example, across the five pulse oximeters, the false negative rate for $SpO_2 \leq$ 94% (i.e. the proportion of $SpO_2$ readings > 94% despite an associated $SaO_2 \leq$ 92%) ranged from 5.3 to 35.3 percentage points higher for a patient with median 'dark' skin tone than for a patient with median 'light' skin tone (7.6–62.2% vs 1.2–26.9%, rate ratio 2.3–7.1). Conversely, the false positive rate for identifying $SaO_2 \leq$ 92% decreased with darker skin tone. Absolute values of the false negative and false positive rates were driven by the overall bias of the pulse oximeter model, with pulse oximeters that tended to underestimate $SaO_2$ having lower false negative rates and higher false positive rates, and pulse oximeters that tended to overestimate $SaO_2$ having higher false negative rates and lower false positive rates. The overall discriminatory power of $SpO_2$, to predict an $SaO_2 \leq$ 92%, assessed as the area under the ROC curve, tended to decrease with darker skin tone (figure 4). Estimated rates of occult hypoxaemia were imprecise due to the small number of observations with $SaO_2 <$ 88% but tended to increase with darker skin tone (figure 4).

Regression model outputs are shown in detail in the supplementary materials.

# Discussion

## Principal findings

We found that all five fingertip pulse oximeters supplied for home use in the NHS England COVID oximetry @home scheme returned higher $SpO_2$ values for patients with darker skin tones than patients with lighter skin tones but the same $SaO_2$. The nature and magnitude of this bias varied between pulse oximeters and was dependent on both an individual's skin tone and the underlying $SaO_2$. At high saturations, all pulse oximeters underestimate $SaO_2$ and higher readings for darker skin tones counterintuitively mean less bias. However, each pulse oximeter exhibiting a tipping point below which $SaO_2$ was overestimated and darker skin increased the bias.

Overall measurement accuracy ($A_{RMS}$) was primarily driven by precision and was not sensitive to the relatively small variations in bias by skin tone. However, these small variations in bias led to marked differences in diagnostic accuracy, with much higher false negative rates (failure to identify true hypoxaemia, or occult hypoxaemia) but correspondingly lower false positive rates (false alarms) in patients with darker skin tones. While the relationships between skin tone and accuracy were non-linear, the clinically important relationships with bias, false negative rates and false positive rates were generally monotonic (always increasing or decreasing with skin tone, within the bounds of the confidence intervals illustrated).

## Comparison with other studies

Our findings align with laboratory investigations of healthy volunteers, smaller clinical studies and retrospective database analyses conducted over the last four decades, that have reported relationships between skin tone and diagnostic accuracy.[6] However, our findings undermine the simplistic view that oximeter function is worse in patients with darker skin tones. Of note, two of the largest prospective clinical accuracy studies to date reported conflicted findings. A study of 394 critically ill patients on ICU reported that bias was greater in patients with darker rather than lighter skin tones,[15] whilst the other recruited 400 hospital in- and out-patients but found that skin tone did not affect pulse oximeter accuracy.[16] Both studies differed from ours in that they employed subjective measures of skin tone. The degree to which pulse oximeters overestimate $SaO_2$ in people with darker skin tones has been difficult to quantify as it varies between devices, is exacerbated by low $SaO_2$ and many studies have used ethnicity as a proxy measure for skin tone. Our study evaluated five low-cost pulse oximeters but did not compare measurements to those from devices commonly used in a hospital setting. The devices were supplied to the study team through the

16

procurement route used to provide oximeters for the NHS England COVID oximetry @home scheme and represent the devices in use in the programme at that time. Other similar devices are available and may be in use both within the scheme and elsewhere. The findings of this study may not be generalisable to other models of pulse oximeter in different clinical settings. However, in a recent, much smaller study using a similar design to ours, investigators reported the accuracy of an FDA-approved critical care pulse oximeter using data from 350 paired $SpO_2$-$SaO_2$ measurements from 12 critically ill adults, five with dark and seven with lighter skin tone.[17] Using the same spectrophotometer as our in study to calculate ITA, the investigators reported significant overestimation of $SaO_2$ by the pulse oximeter for patients with darker skin tone. In another study, spectrophotometry was used in a laboratory setting to measure the ITA of healthy volunteers exposed to hypoxic gas mixtures whilst evaluating the accuracy of 11 different pulse oximeters.[8] The investigators reported that five of the 11 devices had an $A_{RMS}$ of >3% and that nine performed worse for participants with the darkest skin tones. Comparison of our study to most of the other published literature in this field is difficult as it either did not measure skin tone directly, or did so using colour charts, a method know to be flawed.[18,19]

**Strengths and limitations of this study**

The strengths of our study include the large number of participants and simultaneous prospective measurements, the wide distribution of skin tones, and the objective measurement of skin tone using spectrophotometry. Sophisticated statistical modelling approaches were used to assess the performance of the pulse oximeters across multiple domains of both measurement and diagnostic accuracy. We mitigated the impact of noise in skin tone measurement by deriving ITA from four L* and b* measurements for each patient, with little residual noise in the resulting average. The study does, however, have some limitations. To enable a gold-standard measurement for comparison and to explore accuracy in those with low $SaO_2$, the study was conducted in critically ill patients. Other elements of critical illness that we were unable to account for in analyses may have affected the results. In particular, critically ill patients may be prone to poor perfusion of their digits due to hypotension, hypovolaemia or the use of vasopressor therapy. Poor perfusion has been shown to exacerbate the skin tone related bias of pulse oximetry.[20] This could reduce the generalisability of the findings to a community based setting, where these pulse oximeters are designed to be used. However, in the CIE L*a*b* colour space; L* represents dark through to light (correlating with skin colour); a* represents cutaneous erythema and is impacted by melanin composition and cutaneous blood flow; and b* reflects an individual's constitutional pigmentation and their ability to tan.[21] As the calculation of ITA does not include a* values, it is unlikely that skin tone values in this study were affected by perfusion.

17

It is possible that differences between ITA measurements taken from the dorsal surface of the hand could differ slightly from those at the fingertip, where pulse oximeters were placed. Finally, while the range of oxygen saturations observed covered the key decision thresholds in current treatment guidelines for home use, only 16% of $SaO_2$ values were 92% or lower and only 3% were below 88%.

## Implications for clinical practice

Healthcare providers need to remember that pulse oximetry is a way of estimating $SaO_2$ that has several limitations and that performance can vary considerably between models. It is their responsibility to understand the performance characteristics of the device they are using. Manufacturers should ensure robust testing of devices prior to marketing them and continue to monitor performance post-marketing. All performance data should be made available to end-users.

The findings of our study and work that preceded it need to be considered when determining the clinical relevance of $SpO_2$ readings and steps taken to mitigate any potential harm that could occur by failing to appreciate the effect of skin tone on pulse oximetry accuracy. $SpO_2$ readings should be interpreted in the context of other clinical information and trends in $SpO_2$ values given greater importance than single readings, particularly in patients with darker skin tones. False negative readings translate to a failure to detect hypoxaemia, which could have important clinical ramifications; whilst false positive readings may result in the unnecessary administration of oxygen, and other therapies. Healthcare systems should develop guidance to inform and aid practitioners, patients and the public, particularly in settings where additional clinical readings from other medical measurement devices would not be available. As the accuracy of pulse oximetry varies between devices, across skin tones and is dependent on $SaO_2$, it is not possible to generate unified clinical guidance from the findings of our study.

## Implications for standards and regulation

The current International Organization for Standardization (ISO) standard on basic safety and essential performance of pulse oximeters requires observations from only 10 individuals, with no requirements related to skin tone.[22] It relies on a single performance metric, $A_{RMS}$, which our study has demonstrated is insensitive to potentially clinically important variations in performance by skin tone (and would be similarly insensitive to other biases of a comparable magnitude). This ISO standard is currently in the process of being reviewed and revised. In the UK, the Medicines and Healthcare products Regulatory Agency (MHRA) does not approve medical devices directly. Market approval is expected to include compliance with the ISO standard, but this is not mandatory. The US Food and Drug Administration

18

(FDA) has recently published draft guidance aimed at improving the accuracy of pulse oximeters used in healthcare settings.[23] The draft guidance recommends pulse oximeters are evaluated in 150 or more participants with a diverse range of skin tones, assessed by both Monk Skin Tone (MST) scale and ITA, and 20–24 paired measurements of $SpO_2$ and $SaO_2$ per participant. In addition to $A_{RMS}$, the draft guidance includes recommended success criteria in terms of variation in the bias across MST categories and by ITA. While the draft FDA guidance goes some way to addressing the issues identified in our study, a 1.5 percentage point variation in bias across a 100° range of ITA would be considered 'non-disparate performance' for $SaO_2$ values in the range we studied. This is at the upper end of the variation in bias seen in our study, which translated into substantial variation in diagnostic accuracy. As $SpO_2$ thresholds may be used to guide clinical practice or healthcare seeking behaviour, more stringent criteria may need to be considered.

**Implications for further research**

This study was designed to evaluate the accuracy of fingertip pulse oximeters designed to be used at home at oxygen saturations in the range relevant to clinical decision-making in the community. Therefore, the findings are not directly generalisable to high fidelity pulse oximeters used in secondary care settings, or to lower oxygen saturations relevant to clinical decision-making in these settings. Future research should investigate whether the signals detected in these hospital-based retrospective studies are replicated using objective measures of skin tone. Our study has highlighted the complexity of the effect skin tone on pulse oximeter accuracy and should serve as a framework to generate further research questions aiming to reduce bias in pulse oximetry. There remains no gold standard method for measuring skin tone.[9] We used spectrophotometry to generate ITA values to avoid the well described limitations of colour charts and the Fitzpatrick skin type,[24] and took measurements from the dorsal surface of the hand as this is close to the fingertip. ITA values vary according to where on the body the measurements are taken,[8] and further work needs to be undertaken to determine which site provides the most relevant values to optimise pulse oximetry. Also, in settings where spectrophotometry is not possible, reliable alternatives need to be evaluated.

**Conclusions**

For any given $SaO_2$ across the range studied, the five fingertip pulse oximeters evaluated gave $SpO_2$ readings on average 0.6–1.5 percentage points higher for a patient with median 'dark' skin tone than for a patient with median 'light' skin tone. These small variations in bias translated into substantial differences in false positive and false negative rates for detecting hypoxaemia.

Accepted manuscript

**Summary box**

| **What is already known on this topic** |
| --- |
| • Pulse oximeters measure peripheral haemoglobin oxygen saturation ($SpO_2$) and were developed to identify hypoxaemia (low blood oxygen levels) but their accuracy is known to be affected by several factors. |
| • Combining evidence from previous studies indicates that darker skin tones, often assessed using ethnicity as a proxy for skin tone, are associated with overestimation of arterial haemoglobin oxygen saturation ($SaO_2$). |
| • Heterogeneity in study designs including a lack of objective skin tone measurement has prevented quantification of the impact of skin tone on the measurement and diagnostic accuracy of pulse oximeters. |
| What this study adds |
| • The five pulse oximeters assessed all gave higher $SpO_2$ readings for patients with darker skin tones than for patients with lighter skin tones. |
| • While absolute differences in readings were small, they can result in substantially higher rates of false negatives and lower rates of false positives in the diagnosis of hypoxaemia. |
| • Neither bias nor accuracy root mean square (the measure used to define manufacturing standards for pulse oximeter accuracy) adequately reflected differences in diagnostic accuracy across skin tones. |

Word count: 4439

**Contributors**

DSM, DG, JD, CM, OO, EH, KR, PM, and DAH conceptualised and designed the study; DSM, DG, CM, OO, EH, KR, PM, and DAH obtained funding for the study; DSM, JD, DG, TS, AC, WNC, AFJ, RG, CM, OO, EH, KR, PM, DAH conducted the study; TS and AFJ were responsible for project administration; JD, AC, WNC, and DAH were responsible for data curation; JD, AC, WNC, and DAH were responsible for study software; JD, AC, WNC, and DAH undertook formal analysis of the data; DSM, JD, DG, AC, WNC, EH, KR, PM, and DAH

20

validated the findings; JD, AC, WNC, DAH created the visualisations; DSM, JD, DG, KR,

PM, and DAH wrote the original draft of the manuscript; DSM, JD, DG, TS, AC, WNC, AFJ,

RG, CM, OO, EH, KR, PM, and DAH reviewed and edited versions of the manuscript. JD,

AC and DAH directly accessed and verified the underlying data reported in the manuscript.

DSM is the guarantor. The corresponding author attests that all listed authors meet

authorship criteria and that no others meeting the criteria have been omitted. All authors

revised the manuscript and approved the final manuscript as submitted.

Competing interests

All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/disclosure-of-interest/](www.icmje.org/disclosure-of-interest/) and declare: DSM, JD, DG, TS, AC, WNC, AFJ, CM, OO, EH, KR, PM and DAH
report funding from the National Institute for Health and Care Research (NIHR) Health
Technology Assessment Programme to their institutions for the work presented in the
manuscript. The authors report no financial relationships with any organisations that might
have an interest in the submitted work in the previous three years; no other relationships or
activities that could appear to have influenced the submitted work. DSM is Chief Investigator
for the UK-ROX trial and a member of the Medicines and Healthcare products Regulatory
Agency Interim Devices Working Group. DSM and EH contributed to the Equity in medical
devices: independent review (2024). EH is a trustee of the British Skin Foundation and a
member of the British Association of Dermatologists (both organisations have expressed
interests in "skin of colour") and has been a member of the European Society for Pigment
Cell Research Board. KR is Director of the NIHR Health and Social Care Delivery Research
Programme.

**Ethical approval**

Ethical approval was granted by the South Central - Oxford C Research Ethics Committee (20/SC/0423).

**Data sharing**

Deidentified participant data and associated documentation will be made available to researchers for any reasonable purpose, subject to a data use agreement, following completion and publication of all pre-planned analyses. To request access to data, please visit: https://www.icnarc.org/data-services/access-our-data/. Full analysis code can be viewed at https://github.com/ICNARC/EXAKT

**Transparency**

The lead author (the manuscript's guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

**Dissemination statement**

We will initially disseminate the findings of our study to the public via our public website, using plain language. We are working with community groups who have assisted us throughout the study to produce infographics, animations and plain language summaries specifically for members of the public who may be particularly affected by the findings. These will form the next phase of our public dissemination programme. We have already begun to engage with key healthcare and policy stakeholders in the United Kingdom to communicate our findings to them. We will continue to work with these stakeholders after the publication of this manuscript to facilitate future research and clinical guidance.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

This article is published Open Access and a CC BY licence is needed.

Accepted manuscript

**References**

1. Greenhalgh T, Knight M, Inda-Kim M, Fulop NJ, Leach J, Vindrola-Padros C. Remote management of covid-19 using home pulse oximetry and virtual ward support. *BMJ* 2021; **372**: n677.

2. Alboksmaty A, Beaney T, Elkin S, et al. Effectiveness and safety of pulse oximetry in remote patient monitoring of patients with COVID-19: a systematic review. *Lancet Digit Health* 2022; **4**: e279-e89.

3. Rubinstein F, Williams R, Dumville J, et al. Patient use of pulse oximetry to support management of COVID-19 in Greater Manchester: A non-randomised evaluation using a target trial approach. *PLoS One* 2024; **19**: e0310822.

4. Sherlaw-Johnson C, Georghiou T, Morris S, et al. The impact of remote home monitoring of people with COVID-19 using pulse oximetry: A national population and observational study. *EClinicalMedicine* 2022; **45**: 101318.

5. Lipnick MS, Feiner JR, Au P, Bernstein M, Bickler PE. The Accuracy of 6 Inexpensive Pulse Oximeters Not Cleared by the Food and Drug Administration: The Possible Global Public Health Implications. *Anesth Analg* 2016; **123**: 338-45.

6. Martin D, Johns C, Sorrell L, et al. Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: a systematic review. *Br J Anaesth* 2024; **132**: 945-56.

7. 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 Network Open* 2023; **6**: e2330856-e.

8. Leeb G, Auchus I, Law T, et al. The performance of 11 fingertip pulse oximeters during hypoxemia in healthy human participants with varied, quantified skin pigment. *EBioMedicine* 2024; **102**: 105051.

9. Weir VR, Dempsey K, Gichoya JW, Rotemberg V, Wong AI. A survey of skin tone assessment in prospective research. *NPJ Digit Med* 2024; **7**: 191.

10.    Martin DS, Gould DW, Shahid T, et al. Conservative Oxygen Therapy in Mechanically Ventilated Critically Ill Adult Patients: The UK-ROX Randomized Clinical Trial. *JAMA* 2025; **334**: 398-408.

11.    Del Bino S, Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br J Dermatol* 2013; **169 Suppl 3**: 33-40.

12.    Chardon A, Cretois I, Hourseau C. Skin colour typology and suntanning pathways. *Int J Cosmet Sci* 1991; **13**: 191-208.

13.    NHS England. Suspected coronavirus (COVID-19): Important information to keep you safe while at home. Version 8. 21 April 2022. https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/12/C1634-suspected-Coronavirus-important-information-to-keep-you-safe-at-home-V8_210422.pdf (accessed 15 April 2025).

14.    Treweek S, Banister K, Bower P, et al. Developing the INCLUDE Ethnicity Framework—a tool to help trialists design trials that better reflect the communities they serve. *Trials* 2021; **22**: 337.

15.    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**: 297-303.

16.    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 Pulmonary Medicine* 2020; **20**: 7.

17.    Fawzy A, Ali H, Dziedzic PH, et al. Skin Pigmentation and Pulse Oximeter Accuracy in the Intensive Care Unit: A Pilot Prospective Study. *Am J Respir Crit Care Med* 2024; **210**: 355-8.

18.    Eilers S, Bach DQ, Gaber R, et al. Accuracy of self-report in assessing Fitzpatrick skin phototypes I through VI. *JAMA Dermatol* 2013; **149**: 1289-94.

19.    Hannon L, DeFina R. Just Skin Deep? The Impact of Interviewer Race on the Assessment of African American Respondent Skin Tone. *Race Soc Probl* 2014; **6**: 356-64.

20.    Gudelunas MK, Lipnick M, Hendrickson C, et al. Low Perfusion and Missed Diagnosis of Hypoxemia by Pulse Oximetry in Darkly Pigmented Skin: A Prospective Study. *Anesth Analg* 2024; **138**: 552-61.

21.    Ly BCK, Dyer EB, Feig JL, Chien AL, Del Bino S. Research Techniques Made Simple: Cutaneous Colorimetry: A Reliable Technique for Objective Skin Color Measurement. *J Invest Dermatol* 2020; **140**: 3-12.e1.

22.    The British Standards Institution. BS EN ISO 80607-2-61:2019: 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.

23. US Department of Health and Human Services Food and Drug Administration. Pulse Oximeters for Medical Purposes - Non-Clinical and Clinical Performance Testing, Labeling, and Premarket Submission Recommendations: Draft Guidance for Industry and Food and Drug Administration Staff. 7 January 2025. https://www.fda.gov/media/184896/download (accessed 15 April 2025).

24. Ware OR, Dawson JE, Shinohara MM, Taylor SC. Racial limitations of fitzpatrick skin type. *Cutis* 2020; **105**: 77-80.

**Figure legends**

**Figure 1  Distribution of skin tone in the EXAKT study cohort, by pulse oximeter**

Each dot represents one patient in the EXAKT study cohort, with each value (lightness and yellow components) averaged across four spectrophotometer readings of the dorsal aspect of the hand. Colours displayed are approximately indicative of patient's skin tone and may be affected by display settings or print quality. Individual typology angle (ITA) is indicated by the angle from above or below a horizontal line from ($L^* = 50$, $b^* = 0$), with the boundaries between skin tone categories visualised as dashed lines at 55°, 41°, 28°, 10° and −30°. Pulse oximeters: A, ChoiceMMed Oxywatch MD300C19; B, ChoiceMMed Oxywatch MD300C13-CC12; C, Creative Medical PC-60B1-BL; D, Biolight Meditech M70C; E, Medlinket AM802 NHE.

**Figure 2  Relationship between $SpO_2$ and $SaO_2$, by pulse oximeter**

Each dot represents one observation, illustrating the relationship between $SpO_2$ and $SaO_2$, across the range of $SaO_2$ from 80% to 100%, with the dashed line indicating exact agreement. In the top row, the data are presented with $SpO_2$ on the vertical axis, the same data are presented with the difference between $SpO_2$ and $SaO_2$ (error or bias) on the vertical axis. Solid lines illustrate the adjusted mean error (bias) with shaded 95% confidence intervals. Pulse oximeters: A, ChoiceMMed Oxywatch MD300C19; B, ChoiceMMed Oxywatch MD300C13-CC12; C, Creative Medical PC-60B1-BL; D, Biolight Meditech M70C; E, Medlinket AM802 NHE.

**Figure 3  Variation in measurement accuracy at $SaO_2$ 92%, across a range of skin tones**

Pulse oximeters: A, ChoiceMMed Oxywatch MD300C19; B, ChoiceMMed Oxywatch MD300C13-CC12; C, Creative Medical PC-60B1-BL; D, Biolight Meditech M70C; E, Medlinket AM802 NHE. Solid lines indicate modelled estimates, with shading of 95% confidence limits. Dashed lines indicate exact agreement or perfect prediction. Bias: mean error (systematic error); Precision: root mean squared deviation (random error). Accuracy: accuracy root mean square (total error). See supplementary methods for full definitions, details of statistical modelling, and estimates at different levels of $SaO_2$.

**Figure 4  Variation in diagnostic accuracy to detect SaO$_2$ ≤ 92%, across a range of skin tones**

Pulse oximeters: A, ChoiceMMed Oxywatch MD300C19; B, ChoiceMMed Oxywatch MD300C13-CC12; C, Creative Medical PC-60B1-BL; D, Biolight Meditech M70C; E, Medlinket AM802 NHE. Solid lines indicate modelled estimates, with shading of 95% confidence limits. Dashed lines indicate exact agreement or perfect prediction. False negative rate: failure of pulse oximeter to correctly identify hypoxaemia; False positive rate: pulse oximeter indicating hypoxaemia when no hypoxaemia is present; Area under receiver operative characteristic (ROC) curve: likelihood that a pulse oximeter will indicate a lower oxygen saturation for someone with hypoxemia than for someone without hypoxemia; Occult hypoxaemia: failure of pulse oximeter to identify severe hypoxaemia. See supplementary methods for full definitions, details of statistical modelling, and estimates at different levels of SpO$_2$.

**Table 1    Patient characteristics and arterial blood gas observations**

| | Pulse oximeter | | | | | Overall |
|---|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | **E** | |
| *Patient characteristics* | | | | | | |
| N | 369 | 351 | 351 | 346 | 388 | 903 |
| Age [1], mean (SD) [N] | 56.1 (16.1) [348] | 55.4 (15.8) [329] | 55.7 (15.4) [337] | 55.6 (16.2) [312] | 55.0 (16.8) [369] | 55.6 (16.0) [848] |
| Male [1], n (%) [N] | 234 (67.2%) [348] | 217 (66.0%) [329] | 234 (69.4%) [337] | 193 (61.9%) [312] | 253 (68.6%) [369] | 566 (66.7%) [848] |
| Ethnic group [2], n (%) | [N = 312 (84.6%)] | [N = 277 (78.9%)] | [N = 296 (84.3%)] | [N = 274 (79.2%)] | [N = 314 (80.9%)] | [N = 737 (81.6%)] |
|   White | 104 (33.3%) | 99 (35.7%) | 75 (25.3%) | 88 (32.1%) | 98 (31.2%) | 232 (31.5%) |
|   Asian | 104 (33.3%) | 83 (30.0%) | 111 (37.5%) | 66 (24.1%) | 90 (28.7%) | 227 (30.8%) |
|   Black | 72 (23.1%) | 52 (18.8%) | 72 (24.3%) | 84 (30.7%) | 73 (23.2%) | 177 (24.0%) |
|   Mixed or Other | 32 (10.3%) | 43 (15.5%) | 38 (12.8%) | 36 (13.1%) | 53 (16.9%) | 101 (13.7%) |
| Skin tone group, n (%) | | | | | | |
|   Very light or light [3] | 46 (12.5%) | 54 (15.4%) | 29 (8.3%) | 46 (13.3%) | 47 (12.1%) | 111 (12.3%) |
|   Intermediate | 60 (16.3%) | 53 (15.1%) | 57 (16.2%) | 47 (13.6%) | 61 (15.7%) | 139 (15.4%) |
|   Tan | 77 (20.9%) | 79 (22.5%) | 78 (22.2%) | 74 (21.4%) | 88 (22.7%) | 198 (21.9%) |
|   Brown | 122 (33.1%) | 115 (32.8%) | 112 (31.9%) | 124 (35.8%) | 137 (35.3%) | 305 (33.8%) |
|   Dark | 64 (17.3%) | 50 (14.2%) | 75 (21.4%) | 55 (15.9%) | 55 (14.2%) | 150 (16.6%) |
| Chronic respiratory disease [4], n (%) [N] | 24 (7.6%) [317] | 28 (9.3%) [300] | 13 (4.9%) [267] | 18 (6.9%) [259] | 15 (4.4%) [344] | 49 (6.6%) [744] |
| Reason for admission to critical care (body system) [1], n (%) | [N = 348 (94.3%)] | [N = 329 (93.7%)] | [N = 337 (96.0%)] | [N = 312 (90.2%)] | [N = 369 (95.1%)] | [N = 848 (93.9%)] |
|   Respiratory | 105 (30.2%) | 101 (30.7%) | 100 (29.7%) | 95 (30.4%) | 99 (26.8%) | 250 (29.5%) |
|   Cardiovascular | 70 (20.1%) | 57 (17.3%) | 59 (17.5%) | 70 (22.4%) | 82 (22.2%) | 169 (19.9%) |
| *Arterial blood gas observations* | | | | | | |
| N | 2177 | 2215 | 2010 | 2355 | 2251 | 5656 |
| $SaO_2$, %, median (IQR) | 96 (94, 98) | 95 (93, 97) | 96 (94, 98) | 96 (94, 97) | 96 (94, 98) | 96 (94, 97) |
| $SaO_2$ range [5], n (%) | | | | | | |
|   < 88% | 30 (1.4%) | 36 (1.6%) | 28 (1.4%) | 35 (1.5%) | 44 (2.0%) | 95 (1.7%) |
|   88% – 92% | 306 (14.1%) | 391 (17.7%) | 245 (12.2%) | 322 (13.7%) | 284 (12.6%) | 796 (14.1%) |
|   93% – 94% | 357 (16.4%) | 414 (18.7%) | 322 (16.0%) | 400 (17.0%) | 339 (15.1%) | 940 (16.6%) |
|   > 94% | 1484 (68.2%) | 1374 (62.0%) | 1415 (70.4%) | 1598 (67.9%) | 1584 (70.4%) | 3825 (67.6%) |
| Haemoglobin, g/L, median (IQR) | 101 (86, 116) | 104 (90, 123) | 99 (85, 113) | 100 (86, 117) | 99 (85, 115) | 101 (86, 117) |
| Carboxyhaemoglobin [6], %, median (IQR) | 1.1 (0.7, 1.5) | 1.2 (0.8, 1.5) | 1.2 (0.8, 1.5) | 1.2 (0.8, 1.5) | 1.2 (0.9, 1.5) | 1.2 (0.8, 1.5) |
| Methaemoglobin [7], %, median (IQR) | 0.7 (0.5, 1.0) | 0.8 (0.5, 1.1) | 0.7 (0.3, 1.0) | 0.6 (0.4, 0.9) | 0.8 (0.5, 1.1) | 0.7 (0.4, 1.0) |

Pulse oximeters: A, ChoiceMMed Oxywatch MD300C19; B, ChoiceMMed Oxywatch MD300C13-CC12; C, Creative Medical PC-60B1-BL; D, Biolight Meditech M70C; E, Medlinket AM802 NHE.

[1] Based on linked routine data. N=55 (6.1%) not linked.

[2] Based on linked routine data. N=55 (6.1%) not linked, N=111 (12.3%) recorded as "Not stated".

[3] Due to small numbers, "very light" (N=13) is combined with "light" in tables and figures.

[4] Shortness of breath with light activity due to pulmonary disease and evident within the six months prior to admission. Based on linked routine data. N=55 (6.1%) not linked, N=104 (11.5%) missing.

[5] Rounded to nearest percentage point, for consistency with $SpO_2$.

[6] N=36 (0.6%) missing.

[7] N=71 (1.3%) missing.

**Print abstract**

**Study question.** What is the impact of skin tone on the measurement and diagnostic accuracy of five fingertip pulse oximeters provided by the NHS for use at home in the NHS England COVID oximetry @home scheme?

**Methods.** 903 critically ill adults on intensive care units, screened for or enrolled into a trial evaluating oxygen therapy, were enrolled into this study. Pulse oximetry-derived peripheral oxygen saturation ($SpO_2$) measurements were compared with paired arterial oxygen saturation ($SaO_2$) measurements from arterial blood analysed by co-oximetry. Skin tone was objectively measured using a hand-held spectrophotometer and presented as individual typology angle (ITA). Diagnostic accuracy for identifying $SaO_2$ ≤92% was assessed by false negative and positive rates for $SpO_2$ using thresholds of ≤92% and ≤94%, and the area under the receiver operating characteristic curve, and by presence of occult hypoxaemia ($SaO_2$ <88% with $SpO_2$ >92%).

**Study answer and limitations.** $SpO_2$ readings were, on average, 0.6–1.5 percentage points higher for patients with darker (ITA −44°) than lighter (ITA 46°) skin tone. At both $SpO_2$ thresholds assessed, false negative rates increased with darker skin tones; the proportion of $SpO_2$ measurements >94% despite a paired $SaO_2$ ≤92% ranged from 5.3–35.3 percentage points higher for patients with darker than lighter skin tone (7.6–62.2% vs 1.2–26.9%, rate ratio 2.3–7.1). Conversely, false positive rates decreased with darker skin tones. A key limitaion was that the study was conducted in critically ill patients, which may limit the generalisability of the findings.

**What this study adds.** Using an objective measure of skin tone to prospectively evaluate pulse oximeter accuracy in a large cohort of patients, small variations in bias translated into substantial differences in false positive and false negative rates for detecting hypoxaemia.

Accepted manuscript

ClinicalTrials.gov NCT05481515