



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

Ralf E Harskamp ,¹ Luuk Bekker,² Jelle C L Himmelreich,¹ Lukas De Clercq,¹ Evert P M Karregat,¹ Mengalvio E Sleswijk,² Wim A M Lucassen¹

To cite: Harskamp RE, Bekker L, Himmelreich JCL, *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 Res* 2021;**8**:e000939. doi:10.1136/bmjresp-2021-000939

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2021-000939>).

REH and LB are joint first authors.

Received 24 March 2021
Accepted 11 July 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of General Practice, Amsterdam UMC Locatie AMC, Amsterdam, North Holland, The Netherlands

²Flevoziekenhuis, Almere, Flevoland, The Netherlands

Correspondence to
Ralf E Harskamp;
r.e.harskamp@amsterdamumc.nl

ABSTRACT

Objectives To evaluate the performance of direct-to-consumer pulse oximeters under clinical conditions, with arterial blood gas measurement (SaO₂) as reference standard.

Design Cross-sectional, validation study.

Setting Intensive care.

Participants Adult patients requiring SaO₂-monitoring.

Interventions The studied oximeters are top-selling in Europe/USA (AFAC FS10D, AGPTEK FS10C, ANAPULSE ANP 100, Cocobear, Contec CMS50D1, HYLOGY MD-H37, Momed YM101, PRCMISEMED F4PRO, PULOX PO-200 and Zaccurate Pro Series 500 DL). Directly after collection of a SaO₂ blood sample, we obtained pulse oximeter readings (SpO₂). SpO₂-readings were performed in rotating order, blinded for SaO₂ and completed <10 min after blood sample collection.

Outcome measures Bias (SpO₂-SaO₂) mean, root mean square difference (A_{RMS}), mean absolute error (MAE) and accuracy in identifying hypoxaemia (SaO₂ ≤90%). As a clinical index test, we included a hospital-grade SpO₂-monitor (Philips).

Results In 35 consecutive patients, we obtained 2258 SpO₂-readings and 234 SaO₂-samples. Mean bias ranged from -0.6 to -4.8. None of the pulse oximeters met A_{RMS} ≤3%, the requirement set by International Organisation for Standardisation (ISO)-standards and required for Food and Drug Administration (FDA) 501(k)-clearance. The MAE ranged from 2.3 to 5.1, and five out of ten pulse oximeters met the requirement of ≤3%. For hypoxaemia, negative predictive values were 98%–99%. Positive predictive values ranged from 11% to 30%. Highest accuracy (95% CI) was found for Contec CMS50D1; 91% (86–94) and Zaccurate Pro Series 500 DL; 90% (85–94). The hospital-grade SpO₂-monitor had an A_{RMS} of 3.0% and MAE of 1.9, and an accuracy of 95% (91%–97%).

Conclusion Top-selling, direct-to-consumer pulse oximeters can accurately rule out hypoxaemia, but do not meet ISO-standards required for FDA-clearance

BACKGROUND

Pulse oximetry has become an indispensable, low-cost, non-invasive, diagnostic tool to assess a patients' oxygen saturation.

Key messages

What is the key question?

► To evaluate the performance of direct-to-consumer pulse oximeters in a clinical setting

What is the bottom line?

► The tested pulse oximeters do not meet the required standards, but can be used safely to rule out (but not rule in) hypoxaemia.

Why read on?

► Pulse oximeters are widely used for making treatment or referral decisions, thus understanding their limitations is essential.

Typically, this tool has a clip that can be put on a patient's finger to obtain information on the peripheral arterial oxygen saturation (SpO₂), which serves as a proxy for tissue oxygenation.¹ While most direct-to-consumer oximeters are not intended for use in clinical settings, pulse oximeters have evolved to play a pivotal role in routine medical care, and are an essential bedside tool in making treatment and/or referral decisions in community-based healthcare settings. In the current COVID-19 pandemic, the use of pulse oximeters has become even more indispensable, as hypoxaemia is a common diagnostic finding, with reports indicating that patients with COVID-19 may have few symptoms relative to the degree of hypoxaemia (termed 'silent hypoxaemia').² Given the risk of hypoxaemia, the WHO recommends home oximetry monitoring for patients with COVID-19 and risk factors for progression to severe disease.³ Given its importance in guiding medical decision-making, it is remarkable how little is known about the diagnostic accuracy of these devices when used under clinical conditions

in actual patients.⁴ As such, we aimed to evaluate whether popular direct-to-consumer fingertip pulse oximeters meet the standards for accuracy, as proposed by regulatory bodies under real-world conditions.

METHODS

We reported this diagnostic accuracy study in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 statement.⁵

STUDY DESIGN

We enrolled consecutive patients, who were at least 18 years of age, admitted to the intensive care unit (ICU) of a large community-based hospital (Flevoziekenhuis, Almere, the Netherlands). Eligible patients had a radial artery catheter for arterial blood oxygen saturation assessments (SaO₂) as part of routine medical care. Exclusion criteria were patients without a clinical indication for arterial access, those with known inherited forms of abnormal haemoglobin, and those who rapidly deteriorated due to acute haemodynamic compromise, in which the measurements required for this study could hinder medical interventions and thereby negatively affect patient safety.

Study procedures

Intensive Care personnel notified the site investigator (LB) of a potentially eligible patient for study enrolment. The site investigator enrolled each patient after verbal and written consent, either by the patient or his/her legal representative. During office hours, the site investigator was notified by the Intensive Care personnel when a blood gas sample was about to be performed. The site investigator positioned the pulse oximeters for readings directly after the arterial blood gas sample (SaO₂) was obtained. In order to reduce detection bias, the site investigator applied the pulse oximeters in a rotating fashion, maintaining a fixed device order but starting with a different device in each consecutive sample. As each SpO₂ measurement took 30s, the measurement time window involved at least 300s. The Intensive Care personnel (SaO₂) and site investigator (SpO₂-readings) reported their findings on separate digital forms, and were blinded for each other's findings.

Index test and devices

We evaluated 10 oximeters, which were selected from the top 10 of most purchased pulse oximeters on Amazon in at least two of the following countries: USA, UK, Germany, Italy or France. Amazon was chosen because of its dominance on the e-commerce market. In alphabetical order, these oximeters were as follows: AFAC FS10D, AGPTEK FS10C, ANAPULSE ANP 100, Cocobear, Contec CMS50D1, HYLOGY MD-H37, Mommmed YM101, PRCMISEMED F4PRO, PULOX PO-200 and Zacurate Pro Series 500DL. These pulse oximeters cost between

20€ and 50€ each, and all claimed to meet International Organisation for Standardisation (ISO) standards (see paragraph 'outcomes of interest' for specifics). As a clinical index test, we also included a hospital-grade pulse oximeter (Philips M1191BL sensor glove, Philips, The Netherlands), which was used as the clinical standard of care for continuous SpO₂ monitoring at the study site, and has met ISO standards and received 510(k) clearance of the Food and Drug Administration (FDA) (clearance number: K062455). We used the SpO₂ value that was shown on the pulse oximeter's display 30s after placement on a patient's fingertip. The same fingertip was used for each oximeter. When no result was displayed 30s after placement, this was documented as an invalid reading.

Reference standard

The reference standard was a point of care testing analyser (ABL90 Flex Plus, Radiometer Medical ApS, Brønshøj, Denmark, calibrated as per regulatory standards), which was used to perform blood gas analysis on arterial blood gas samples to obtain the arterial oxygen saturation (SaO₂) at the study site as part of routine Intensive Care.

Outcomes of interest

We formulated the following outcome measures: mean bias, root mean square difference (A_{RMS}), the mean absolute error (MAE) and diagnostic accuracy for hypoxaemia, defined as SaO₂ ≤90%. Mean bias is calculated as SpO₂–SaO₂, A_{RMS} and MAE are derived from calculations involving mean bias and precision (SD of bias). Because outliers have an excessive negative effect on results of the A_{RMS} parameter, we also assessed the MAE, a measurement that is more robust in the presence of outliers. The formulas to calculate these outcomes can be found in the supplement. We evaluated the diagnostic accuracy of the pulse oximeters according to the standards defined by the International Organisation for Standardisation in ISO 80601-2-61:2017, which supersedes the ISO 80601-2-61:2011 standard advised by the American FDA in their 510(k) Premarket Notification Submissions Guidance for pulse oximeters. This standard considers an A_{RMS} of ≤3% in the SaO₂ range of 70%–100% acceptable.^{6,7}

Data collection and sample size

We included data on SpO₂, SaO₂, heart rate, systolic blood pressure and diastolic blood pressure, sex, age, skin type (Fitzpatrick classification scale) as assessed by the site investigator, vasopressor use (ie, noradrenalin dose (mg/h)), body temperature (°C) and hand temperature to touch as assessed by the site investigator. We followed the FDA advice for sample size determination, which states that at least 200 measurements per pulse oximeter should be obtained from at least 10 subjects, of which at least two or 15% have a dark skin type (Fitzpatrick IV–VI).⁷



Statistical analysis

We assessed mean bias ($\text{SpO}_2 - \text{SaO}_2$) and SD, and subsequently calculated the A_{RMS} and MAE using the formulas as described in the supplemental data. We created Bland-Altman plots for each pulse oximeter to graphically display its bias ($\text{SpO}_2 - \text{SaO}_2$) to the reference standard (SaO_2). We added a zero-line and upper and lower limits of agreement ($\pm 1.96 \text{SD}$). To visualise the accuracy standards required by the regulatory bodies, $\pm 3\%$ lines are also displayed in the figures. As multiple observations were performed per individual, we calculated the SD by using the within-subject variance (σ^2) and the between-subject variance (σ_{μ}^2). Since A_{RMS} can be easily affected by the presence of outliers, we added MAE, which is more robust in the presence of outliers, as well as a sensitivity analysis restricting the sample to measurements within 1.96 SD of each pulse oximeter's mean and calculating A_{RMS} and MAE on this sample. This sensitivity analysis is analogous to discarding extreme readings as done in routine clinical care when a reading does not coincide with the patient's apparent clinical state. We also assessed the diagnostic performance (sensitivity, specificity, predictive values and accuracy) of pulse oximeters in detecting hypoxaemia, which we defined as $\text{SaO}_2 \leq 90\%$. Finally, we evaluated factors associated with poor performance of each pulse oximeter. We used bias as a continuous measurement for performance, and used logistic regression models in which we included relevant patient and pulse oximeter characteristics. We used SPSS (IBM SPSS, V.26.0, IBM, Armonk, USA), R software (R V.3.6.1, The R Foundation for statistical computing) and MedCalc Statistical Software V.18.5 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>; 2018) to conduct the analyses. We assessed statistical significance at the 0.05 level in all analyses.

RESULTS

Baseline characteristics

In July and August 2020, we enrolled 35 consecutive patients, with a median age of 69 years, and 40% female. [Table 1](#) displays the baseline characteristics. Patients were primarily admitted for respiratory failure due to COVID-19 or other pulmonary diseases (such as chronic obstructive pulmonary disease). Other diagnoses were septic shock, pulmonary embolism, and arterial thrombosis in extremities, severe hyperglycaemic and suicide attempt. Almost half of patients required vasopressors, and all patients required supplemental oxygen.

Blood gas samples and pulse oximeter readings

We obtained a total of 234 arterial blood gas samples; SaO_2 ranged between 85.6% and 99.8%, with the distribution illustrated in [figure 1](#). Of those 234 SaO_2 samples, 12 samples (5.1%) were classified as hypoxaemia. Moreover, each SaO_2 sample was followed by 10 SpO_2 measurements (one from each pulse oximeter), which resulted in a total of 2340 SpO_2 measurements. Eighty-two (3.5%)

Table 1 Patient and pulse oximeter characteristics

	Number/statistic
Patient characteristics (n=35)	
Age, years	69 (61–73)*
Female	14 (40%)
Fitzpatrick IV–VI skintype	5 (14.3%)
Body temperature, °C	37.2±0.9
Heart rate, beats per minute	84 (68–101)*
Systolic blood pressure (mm Hg)	125 (108–143)*
Diastolic blood pressure (mm Hg)	57 (50–66)*
Requiring vasopressor	16/35 (45.7%)
Noradrenalin dosage (mg/h)	0.4 (0.3–1.60)*
Requiring supplemental oxygen	35/35 (100%)
Pulse oximeter reading characteristics (n=2258)	
AFAC FS10D	232/234 (99.1%)
AGPTEK FS10C	233/234 (99.6%)
ANAPULSE ANP 100	202/234 (86.3%)
Cocobear	226/234 (96.6%)
Contec CMS50D1	228/234 (97.4%)
HYLOGY MD-H37	232/234 (99.1%)
Mommed YM101	229/234 (97.9%)
PRCISEMED F4 PRO	233/234 (99.6%)
PULOX-PO-200	231/234 (98.7%)
Zacurate Pro Series 500DL	212/234 (90.6%)
Total pulse oximeter readings	2258/2340 (96.5%)

Binary variables presented as number and percentage. Continuous variables were displayed as mean±SD when normally distributed, or median and 25th–75th percentile (*) when non-normally distributed.

SpO_2 measurements were invalid, as the pulse oximeter did not display a result after 30 s (also see [table 1](#)), leaving 2258 interpretable SpO_2 measurements, which were all obtained within 10 min (range 0–540 s) after acquiring the reference standard's blood sample.

Test results of pulse oximeters

[Table 2](#) shows the test characteristics of the pulse oximeters that are graphically displayed in [figure 2](#) and online supplemental figure S1. We found that most pulse oximeters overall had a negative mean bias, thus SpO_2 was on average lower than SaO_2 . Moreover, none of the tested pulse oximeters met the requirement of diagnostic accuracy, using the $A_{\text{RMS}} \leq 3\%$ threshold. When using MAE, which is less affected by outliers, five oximeters performed within a $\leq 3\%$ difference margin of the reference standard. When excluding extreme outliers ($>1.96 \text{SD}$), performance of Contec CMS50D1, Mommed YM101, Pulox-PO-200 and Zacurate Pro Series 500DL were within the $A_{\text{RMS}} \leq 3\%$ limit, and all but two oximeters (ANAPULSE ANP 100 and Cocobear) met the $\text{MAE} \leq 3\%$ thresholds.

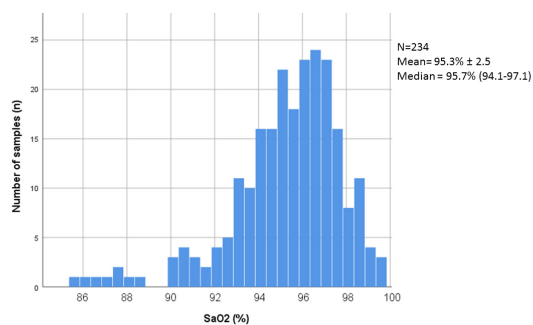


Figure 1 Distribution of arterial blood gas saturation (SaO₂) among obtained samples.

Detection of hypoxaemia

Table 3 summarises the accuracy of each pulse oximeter in detecting hypoxaemia. Overall, the oximeters with the highest specificity were Contec CMS50D1 (93%) and Zacurate Pro Series 500DL (91%). Oximeters with the highest sensitivity were Hylogy MD-H37 (92%) and Anapulse ANP 100 (91%). In terms of predictive values, all pulse oximeters performed well in ruling out hypoxaemia, with negative predictive values of 98%–99% in a population with ≈5% hypoxaemia measurements. Confirming the presence of hypoxaemia was poor, with positive predictive values of 11%–30%. For all reading, accuracy was highest for Contec CMS50D1 (91%) and Zacurate Pro Series 500DL (90%). As a sensitivity analysis, we also provided data on the performance of pulse oximeters at different abnormal oxygenation thresholds (SaO₂ ≤92% and SaO₂ ≤94), which can be found as online supplemental table S1).

Factors associated with poor pulse oximeter performance

Online supplemental table S2 displays factors that were associated with bias (SpO₂–SaO₂) of each pulse oximeter.

Overall, darker skin complexion and an inaccurate pulse rate measurement (difference between pulse oximeter pulse rate and ‘true’ pulse rate as captured by ICU monitor), were associated with poorer SpO₂ performance in five out of 10 pulse oximeters. Other factors that negatively affected results were systolic blood pressure and poor peripheral perfusion (cold hands to touch), in four and three pulse oximeters, respectively.

Test results and diagnostic accuracy of the SpO₂ continuous monitor (clinical index test)

The hospital-grade Philips sensor glove SpO₂ continuous monitor, our clinical index test, performed better with an A_{RMS} of 3.0 and MAE of 1.9 for all measurements (n=234), and A_{RMS} of 2.1 and MAE of 1.6 when excluding outliers (1.96 SD). For detecting hypoxaemia, the hospital-grade Philips sensor glove SpO₂ monitor had a sensitivity of 67% (35%–90%), specificity of 96% (93%–98%), positive and negative predictive values of 50% (31–69) and 98% (96%–99%), respectively, and an accuracy of 95% (91%–97%). The presence of cold hands/acra was associated with higher inaccuracy of the SpO₂ continuous monitor compared with SaO₂.

DISCUSSION

Our study found that the top selling low-cost, popular direct-to-consumer pulse oximeters do not meet the requirements set by the regulatory bodies (ISO/FDA), when compared with the gold standard of arterial oxygen saturation, as obtained by blood gas samples, in a population of intensive care patients. However, when extreme outliers were disregarded, four pulse oximeters would meet the A_{RMS} ≤3% requirements, and eight pulse oximeters would meet MAE ≤3% standards. The hospital-grade Philips sensor glove SpO₂ monitor performed slightly better. We found that the direct-to-consumer pulse

Table 2 Performance of pulse oximeters compared with arterial blood gas (using the metrics: mean bias, root mean square difference and mean absolute error)

	All measurements including outliers			Excluding outliers*	
	Mean bias (±SD)	A _{RMS} (%)	MAE (%)	A _{RMS} (%)	MAE (%)
AFAC FS10D	–2.6±4.4	5.1	3.1	3.3	2.5
AGPTEK FS10C	–2.2±4.0	4.6	2.9	3.0	2.3
ANAPULSE ANP 100	–4.8±5.9	7.5	5.1	5.4	4.4
Cocobear	–3.7±5.1	6.2	4.0	4.3	3.4
Contec CMS50D1	–0.6±4.5	4.5	2.3	2.7	2.0
HYLOGY MD-H37	–3.1±4.9	5.7	3.5	3.6	2.8
Mommed YM101	–1.7±3.6	3.9	2.4	2.6	1.9
PRCISEMED F4 PRO	–2.4±6.4	6.8	3.5	3.5	2.6
PULOX-PO-200	–1.6±3.8	4.1	2.6	2.8	2.2
Zacurate Pro Series 500DL	–1.4±4.5	4.7	2.5	2.5	1.9

*Only values within the limits of agreement (within 1.96 SD).

A_{RMS}, root mean square difference; MAE, mean absolute error.

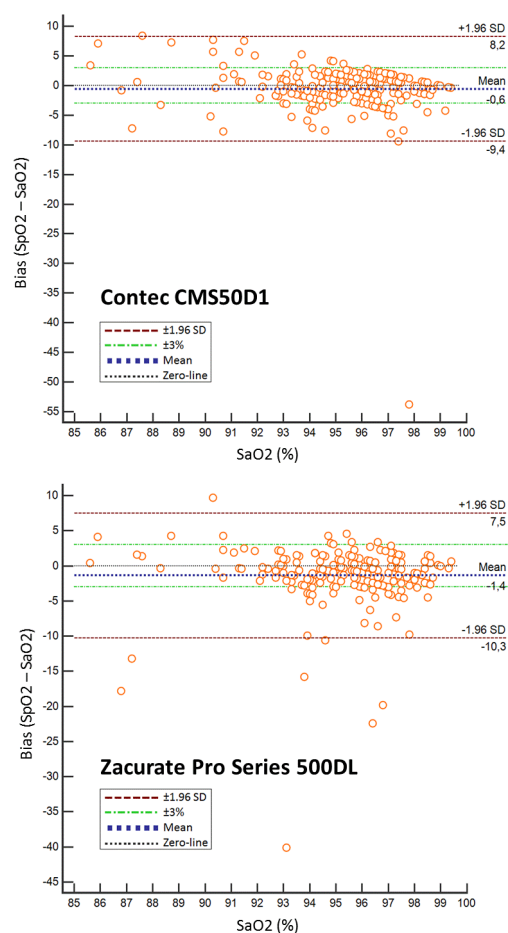


Figure 2 Bland-Altman plots of the bias compared with SaO₂ of two pulse oximeters with the lowest mean bias (Contec CMS50D1 -0.6 and Zacurate Pro series 500DL -1.4). *Bland-Altman plots for each pulse oximeter to graphically display its bias (SpO₂–SaO₂) to the reference standard (SaO₂). We added a zero-line and upper and lower limits of agreement (± 1.96 SD). To visualise the accuracy standards required by the regulatory bodies, $\pm 3\%$ lines are also displayed in the figures. *Bland-Altman plots of the other pulse oximeters can be found as a supplemental figure.

oximeters tested in this study performed well in ruling out hypoxaemia, but are not reliable in confirming the presence of hypoxaemia. Therefore, when such a pulse oximeter indicates a below normal SpO₂ (for instance, when used by a patient for home monitoring), confirmation with a medical-grade oximeter is required. Caution is warranted when factors are present that negatively affect the reliability of these pulse oximeters, such as an inaccurate pulse rate reading, darker skin pigmentation or cold extremities.

Strengths and limitations

Our study was performed under clinical conditions involving consecutive patients with direct access to arterial blood gas samples. Patients were diverse in terms of age, skin type and underlying clinical conditions. We used popular pulse oximeter devices and the number of samples was sufficient to draw accurate conclusions. Moreover, we used a hospital-grade SpO₂ device as a clinical index test. However, there are also a number of limitations that should be mentioned. First, the performance of fingertip pulse oximeters may have been affected by the poor health status and poorer peripheral perfusion of intensive care patients, when compared with community-based patients. Moreover, the distribution of SaO₂ is different in intensive care populations (ie, oxygenation and ventilation settings are set to strive for SaO₂ of 92%), when compared with community based populations, which would also affect performance. Furthermore, the use of sequential SpO₂ measurements meant that the time interval between the arterial blood gas draw (for SaO₂) and the SpO₂ measurement may have been up to several minutes. This is relevant, as minute-to-minute variation in oximetry readings may be present among critically ill patients, even in the relatively stable ICU patients that were enrolled in this study. We applied a rotating order of device usage in order to reduce bias from this limitation. Finally, we did not capture detailed ICU data (such as severity scores), or laboratory findings

Table 3 Accuracy of pulse oximeters in detecting hypoxaemia (prevalence of SaO₂ $\leq 90\%$ is 5.1%)

Pulse oximeter	TP/FN	FP/TN	Sensitivity	Specificity	PPV	NPV	Accuracy
AFAC FS10D	9/3	46/174	75 (43–95)	79 (73–84)	16 (11–23)	98 (96–99)	79 (73–84)
AGPTEK FS10C	10/2	39/182	83 (52–98)	82 (77–87)	20 (15–27)	99 (96–100)	82 (77–87)
ANAPULSE ANP 100	10/1	79/112	91 (59–100)	59 (51–66)	11 (9–14)	99 (95–100)	60 (53–67)
Cocobear	10/2	63/151	83 (52–98)	71 (64–77)	14 (10–18)	99 (96–100)	71 (65–77)
Contec CMS50D1	7/5	16/200	58 (28–85)	93 (88–96)	30 (18–46)	98 (95–99)	91 (86–94)
HYLOGY MD-H37	11/1	51/169	92 (62–100)	77 (71–82)	18 (14–22)	99 (96–100)	78 (72–83)
Mommed YM101	9/3	31/186	75 (43–95)	86 (80–90)	23 (15–32)	98 (96–99)	85 (80–89)
PRCISEMED F4 PRO	8/4	39/182	67 (35–90)	82 (77–87)	17 (11–25)	98 (95–99)	82 (76–86)
PULOX-PO-200	9/3	30/189	75 (43–95)	86 (81–91)	23 (16–32)	98 (96–99)	86 (81–90)
Zacurate Pro Series 500DL	8/2	19/183	80 (44–97)	91 (86–94)	30 (20–42)	99 (96–100)	90 (85–94)

The values for PPV, NPV and accuracy are dependent on disease prevalence.

FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive.



on our patients, such as haemoglobin or acid–base status. These variables are of relevance as SpO₂-readings are based on photoplethysmographic measurements using infrared wavelengths through (de)oxyhaemoglobin. As such, a decrease in haemoglobin blood concentration may affect capturing a sufficient signal. Furthermore, a change in carbon dioxide concentration and/or pH shift could in turn alter the oxyhaemoglobin dissociation curve (Bohr effect), resulting in inaccurate SpO₂ as well as SaO₂ measurements.

Prior studies

Many commonly used direct-to-consumer pulse oximeters do not undergo rigorous in vivo testing, and thus, little is known about the accuracy of these devices. An important study which did perform in vivo testing of inexpensive pulse oximeters was published by Lipnick *et al* in 2016.⁴ In this study, six finger pulse oximeters (not cleared by the FDA) were evaluated in 22 healthy subjects, in which stable SaO₂ plateaus between 70% and 100% were achieved under controlled conditions via a partial rebreathing circuit. The study found that two pulse oximeters tested (Contec CMS50DL and Beijing Choice C20) demonstrated an A_{RMS} of ≤3%, hereby meeting the ISO criteria for accuracy. Of the tested oximeters, the Contec CMS50DL may be comparable with the Contec CMS50D1 model that we tested. In our study, which was performed under clinical conditions, the Contec device did not fare as well, although it was still one of the better pulse oximeters that we tested. Prior studies found that oximeters performed worse in hypoxic conditions, with mean bias increasing at lower oxygen saturations compared with arterial blood gas⁴ or a conventional bedside pulse oximeter.^{8,9} In our study, we found a similar observation in some, but not all oximeters. Still, from a clinical perspective, it is particularly important to have minimal bias in the range of 90%–95%, as this is where a hypoxaemic state should be differentiated from a non-hypoxaemic state.

Implications for practice

Due to the current COVID-19 pandemic, pulse oximeters have become more popular than ever before. Despite their limitations, these devices present a welcome tool for remote monitoring of patients and for ruling out hypoxaemia, particularly in a population where the prevalence of hypoxaemia is low. In our population, in which approximately 5% was hypoxaemic, a selection of top selling low-cost devices were able to safely rule out hypoxaemia in virtually all cases (98%–99%). This percentage would even further approach 100% in low prevalence settings. Still, our findings, as those of other studies^{4,8,9} illustrate that in patients with other symptoms suggestive of hypoxaemia, physicians should remain alert, particularly in high-risk patients with preexisting pulmonary disease. In these scenarios, devices that are FDA-cleared should be used instead, as they show a smaller degree of

increasing bias during lower SaO₂ conditions.^{10–12} Finally, irrespective of the device used, it is important for clinicians to realise that there are a number of factors that negatively impact the reliability of pulse oximetry. These factors include poor peripheral perfusion, inaccurate pulse rate measurement, motion, anaemia, nail polish, and dark skin pigmentation, as shown in our study as well as by others.^{13–16}

CONCLUSION

Direct-to-consumer pulse oximeters are widely available and in high demand. Most of these low cost pulse oximeters have not been rigorously tested. In this study, we tested 10 popular pulse oximeters in ICU patients with direct access to arterial blood gas. Overall, we found that the tested pulse oximeters would not meet strict ISO requirements used by the FDA in their 510(k) premarket notification clearance process. However, most devices can safely rule out hypoxaemia in the vast majority of patients, which is particularly relevant in community-based populations with a low a priori hypoxaemia risk. Future studies are warranted to further assess the accuracy of pulse oximeters in community-based patients, and to gain insight how to further improve this non-invasive, low-cost, and potentially life-saving technology.

Acknowledgements We would like to thank all participating patients, their families, as well as the personnel of the Intensive Care of FlevoZiekenhuis.

Contributors RH and LB designed the study protocol. LB was involved in the recruitment of patients for participation in data collection, supervised by MES. LB performed the statistical analysis and interpreted all data with statistical guidance from JCLH, LDC, EPMK, WL and RH. RH and LB drafted the manuscript and all authors contributed to its revision. All authors read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the local ethics committee and the board of directors of the Flevoziekenhuis. We obtained informed written consent from each participant. When a patient was unable to provide informed consent himself/herself, for example, when sedated or unconscious, a legal representative provided informed consent on the patient's behalf.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

 Ralf E Harskamp <http://orcid.org/0000-0001-9041-0350>

REFERENCES

- 1 Jubran A. Pulse oximetry. *Crit Care* 2015;19:272.
- 2 Tobin MJ, Laghi F, Jubran A. Why COVID-19 silent hypoxemia is baffling to physicians. *Am J Respir Crit Care Med* 2020;202:356–60.
- 3 World Health Organization. COVID-19 clinical management. Available: <https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1> [Accessed 25 Jan 2021].
- 4 Lipnick MS, Feiner JR, Au P, et al. 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.
- 5 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- 6 International Organization for Standardization. *Medical electrical equipment — part 2-61: particular requirements for basic safety and essential performance of pulse oximeter equipment*, 2017.
- 7 U.S.D.o.H.a.H. Services. *Pulse oximeters - premarket notification submissions [510(k)s] guidance for industry and food and drug administration staff*. Rockville: U.S. Department of Health and Human Services, 2013.
- 8 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;109:154–8.
- 9 Hudson AJ, Benjamin J, Jardeleza T, et al. Clinical interpretation of peripheral pulse oximeters labeled “not for medical use”. *Ann Fam Med* 2018;16:552–4.
- 10 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.
- 11 da Costa JC, Faustino P, Lima R, et al. Research: comparison of the accuracy of a pocket versus standard pulse oximeter. *Biomed Instrum Technol* 2016;50:190–3.
- 12 Levrat Q, Petitpas F, Bouche G, et al. Usefulness of pulse oximetry using the set technology in critically ill adult patients. *Ann Fr Anesth Reanim* 2009;28:640–4.
- 13 Adler JN, Hughes LA, Vivilecchia R, et al. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Acad Emerg Med* 1998;5:965–70.
- 14 Coté CJ, Goldstein EA, Fuchsman WH, et al. The effect of nail polish on pulse oximetry. *Anesth Analg* 1988;67:683–6.
- 15 Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 2005;102:715–9.
- 16 Severinghaus JW, Koh SO. Effect of anemia on pulse oximeter accuracy at low saturation. *J Clin Monit* 1990;6:85–8.

Supplemental data

Definition of Root mean square difference

$$A_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

A_{RMS} : root mean square difference;

n : number of samples;

y_j : Pulse oximeter measurement;

\hat{y}_j : Reference standard measurement

We squared the difference (either positive, negative or 0) between a pulse oximeter measurement and the reference standard measurement. Then we calculated the mean of the sum of squares, and determined the root of that mean. As the difference between the pulse oximeter and the reference standard is squared, the A_{RMS} is always positive.

Definition of Mean absolute error

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

MAE : mean absolute error;

n : number of samples;

y_j : Pulse oximeter measurement;

\hat{y}_j : Reference standard measurement

We calculated the absolute difference (either positive or 0) between a pulse oximeter measurement and the reference standard measurement. We then determined the mean of these absolute differences. Similarly to the A_{RMS} , the MAE is always positive, as it concerns an absolute difference.

References

Kampakis, S. Performance measures RMSE MAE. 2020 [cited 2020 01-10-2020]; Available from: <https://thedata scientist.com/performance-measures-rmse-mae/>.

Lipnick, M.S., The Accuracy of 6 Inexpensive Pulse Oximeters Not Cleared by the Food and Drug Administration: The Possible Global Public Health Implications. *Anesthesia and Analgesia*, 2016. 123(2): p. 339-344.

Table S1. Accuracy of pulse oximeters in detecting abnormal oxygenation status, using either $\text{SaO}_2 \leq 92\%$ or $\text{SaO}_2 \leq 94\%$ as threshold

SaO₂ ≤ 92%	Pulse oximeter	TP/FN	FP/TN	Sensitivity	Specificity	PPV	NPV	Accuracy
	AFAC FS10D	19/6	66/141	76 (55-91)	68 (61-74)	22 (18-28)	96 (92-98)	69 (63-75)
	AGPTEK FS10C	18/7	64/144	72 (51-88)	69 (63-75)	22 (17-28)	95 (92-97)	70 (63-75)
	ANAPULSE ANP 100	18/5	76/103	78 (56-93)	58 (50-65)	19 (15-24)	95 (90-98)	60 (53-67)
	Cocobear	22/1	90/113	96 (78-100)	56 (49-63)	20 (17-23)	99 (94-100)	60 (53-66)
	Contec CMS50D1	13/12	34/169	52 (31-72)	83 (77-88)	28 (19-38)	93 (90-96)	80 (74-85)
	HYLOGY MD-H37	21/4	79/128	84 (64-95)	62 (55-68)	21 (17-25)	97 (93-99)	64 (58-70)
	Mommed YM101	18/7	52/152	72 (51-88)	75 (68-80)	26 (20-33)	96 (92-98)	74 (68-80)
	PRCMISEMED F4 PRO	15/9	64/145	63 (41-81)	69 (63-76)	19 (14-25)	94 (91-96)	69 (62-75)
	PULOX-PO-200	20/5	52/154	80 (59-93)	75 (68-81)	28 (22-34)	97 (93-99)	75 (69-81)
	Zacurate Pro Series 500DL	15/7	37/153	68 (45-86)	81 (74-86)	29 (21-38)	96 (92-98)	79 (73-85)
SaO₂ ≤ 94%	AFAC FS10D	64/6	76/86	91 (82-97)	53 (45-61)	46 (41-50)	93 (87-97)	65 (58-71)
	AGPTEK FS10C	62/8	71/92	89 (79-95)	56 (48-64)	47 (42-51)	92 (86-96)	66 (60-72)
	ANAPULSE ANP 100	60/5	102/35	92 (83-97)	26 (18-34)	37 (34-40)	88 (74-94)	47 (40-54)
	Cocobear	71/2	88/71	97 (90-100)	45 (37-53)	45 (41-48)	97 (90-99)	61 (55-68)
	Contec CMS50D1	55/15	41/117	79 (67-87)	74 (66-81)	57 (50-64)	89 (83-93)	75 (69-81)
	HYLOGY MD-H37	67/3	86/76	96 (88-99)	47 (39-55)	44 (40-48)	96 (89-99)	62 (55-68)
	Mommed YM101	60/9	56/104	87 (77-94)	65 (57-72)	52 (46-57)	92 (86-96)	72 (65-77)
	PRCMISEMED F4 PRO	57/12	75/89	83 (72-91)	54 (46-62)	43 (38-48)	88 (81-93)	63 (56-69)
	PULOX-PO-200	62/7	70/92	90 (80-96)	57 (49-65)	47 (42-52)	93 (87-96)	67 (60-73)
	Zacurate Pro Series 500DL	55/9	45/103	86 (75-93)	70 (62-77)	55 (48-61)	92 (86-95)	75 (68-80)

The values for PPV, NPV and accuracy are dependent on disease prevalence

TP= true positive, FP= false positive, FN= false negative, TN= true negative, PPV= positive predictive value, NPV= negative predictive value.

Table S2. Factors associated with poor performance of each pulse oximeter

Pulse oximeter	Associated factor(s)	Beta	Std Error	P-value
AFAC FS10D	Heart rate bias	0.04	0.01	0.002
	Systolic blood pressure	0.04	0.01	0.005
AGPTEK FS10C	Heart rate bias	0.04	0.01	0.002
	Fitzpatrick scale IV-VI	1.96	0.93	0.04
ANAPULSE ANP 100	None	-	-	-
Cocobear	Systolic blood pressure	0.03	0.02	0.03
	Fitzpatrick scale IV-VI	2.30	1.21	0.05
Contec CMS50D1	Cold hands to touch	1.59	0.77	0.04
HYLOGY MD-H37	Fitzpatrick scale IV-VI	3.07	1.13	0.007
Momed YM101	Heart rate bias	0.04	0.01	<0.001
	Systolic blood pressure	0.03	0.01	0.006
	Fitzpatrick scale IV-VI	2.40	0.82	0.004
PRCMISEMED F4 PRO	Cold hands to touch	2.69	1.03	0.01
	Heart rate bias	0.06	0.02	0.002
PULOX-PO-200	Cold hands to touch	1.60	0.62	0.01
	Systolic blood pressure	0.03	0.01	0.006
Zacurate Pro Series 500DL	Heart rate bias	0.06	0.02	0.002
	Fitzpatrick scale IV-VI	2.34	1.1	0.038

Performance of each pulse oximeter was based on bias ($SpO_2 - SaO_2$), a continuous metric. Included variables in logistic models were: age (years), sex, Fitzpatrick scale (I-III vs IV-VI), heart rate bias (difference between pulse oximeter heart rate and true heart rate as captured by ICU monitor), body temperature (degrees C), cold hands to touch (yes/no), systolic blood pressure (mmHg), and use of vasopressor drugs (yes/no).

Figure S1. Bland-Altman plots of performance of pulse oximeters (bias compared with SaO₂)

