# Accuracy of Multiple Pulse Oximeters in Stable Critically Ill Patients

Marie-Anne Blanchet, Gabriel Mercier, Antoine Delobel, Emi Nayet, Pierre-Alexandre Bouchard, Serge Simard, Erwan L'Her, Richard D Branson, and François Lellouche

**BACKGROUND: An accurate $S_{pO_2}$ value is critical in order to optimally titrate oxygen delivery to patients and to follow oxygenation guidelines. Limited prospective data exist on real-world performance of pulse oximeters in critically ill patients. The objective of this study was to assess accuracy and bias of the $S_{pO_2}$ values measured by several oximeters in hospitalized subjects. METHODS: We included stable adults in the ICU with an arterial catheter in place. Main exclusion criteria were poor $S_{pO_2}$ signal and $S_{pO_2}$ > 96%. In each subject, we simultaneously evaluated 4 oximeters: Nonin (Plymouth, Minnesota) embedded in the FreeO$_2$ device (OxyNov, Québec City, Québec, Canada), Masimo (Radical-7, Masimo, Irvine, California), Philips (FAST, Philips, Amsterdam, the Netherlands), and Nellcor (N-600, Medtronic, Minneapolis, Minnesota). Arterial blood gases were drawn and simultaneously each oximeters' $S_{pO_2}$ values were collected. $S_{pO_2}$ values were compared to the reference (arterial oxygen saturation [$S_{aO_2}$] value) to determine bias and accuracy. The ability for oximeters to detect hypoxemia and the impact of oximeters on oxygen titration were evaluated. RESULTS: We included 193 subjects (153 male, mean age 66 y) in whom 211 sets of measurements were performed. The skin pigmentation evaluated by Fitzpatrick scale showed 96.2% of subjects were light skin (types 1 and 2). One oximeter overestimated $S_{aO_2}$ (Philips, +0.9%), whereas the 3 others underestimated $S_{aO_2}$ (Nonin −3.1%, Nellcor −0.3%, Masimo −0.2%). $S_{aO_2}$ was underestimated with Nonin oximeter in 91.3% of the cases, whereas it was overestimated in 55.2% of the cases with Philips oximeter. Moderate hypoxemia ($S_{aO_2}$ 86–90% or $P_{aO_2}$ 55–60 mm Hg) was detected in 92, 33, 42, and 11% of the cases with Nonin, Nellcor, Masimo, and Philips, respectively. CONCLUSIONS: We found significant bias and moderate accuracy between the tested oximeters and the arterial blood gases in the studied population. These discrepancies may have important clinical impact on the detection of hypoxemia and management of oxygen therapy.** *Key words: oximeter; pulse oximetry; oxygen therapy monitoring; hypoxemia; hyperoxemia.* [Respir Care 2023;68(5):565–574. © 2023 Daedalus Enterprises]

## Introduction

Pulse oximetry is used daily by clinicians around the world to obtain a noninvasive measurement of $S_{pO_2}$, a reflection of arterial oxygen saturation ($S_{aO_2}$) that enables detection and amelioration of hypoxemia (main objective)[1-4] and hyperoxemia, both deleterious states.[5-12] Monitoring of pulse oximetry, often considered as the fifth vital sign, is recommended for patients with respiratory failure requiring oxygen therapy.[13] Several guidelines for oxygen therapy aim to reduce both the risk of hypoxemia and hyperoxemia.[13-15] In order to follow these guidelines, an $S_{pO_2}$ value that accurately represents the $S_{aO_2}$ value is required. It has been shown that small variations in $S_{pO_2}$ target greatly affect the oxygen flow required, which may have a relevant impact on clinical decisions.[16-19] Furthermore, $S_{pO_2}$ is frequently used to detect hypoxemia and assess the severity of patients with pneumonia. $S_{pO_2}$ and oxygen support also have an impact on clinical decision-making such as the initiation of corticosteroid treatment, intubation, or transfer to an ICU. For all these reasons, pulse oximeter accuracy is desirable. Although modern pulse oximetry was described 50 years ago, basics of this technology have not evolved significantly. Main evaluations of accuracy of these devices are carried out in healthy volunteers, with the exception of a few studies conducted in subjects.[20-24] For certification purpose, studies are conducted in healthy subjects,[25] whereas a number of common clinical conditions encountered in patients can reduce pulse oximetry accuracy. The conditions that may alter oximeters' precision include skin pigmentation,[20,26-30] skin temperature, skin thickness, nail polish, movement artifacts,[20,21,31-33] and poor peripheral

perfusion.[34,35] The impact of the oximeter brand used on measurement accuracy has been poorly evaluated in subjects with currently available devices.[21,22,26,27,31,36]

No recent study has simultaneously compared $S_{pO_2}$ values from multiple oximeters to the reference standard $S_{aO_2}$. As a result, the assessment of agreement between $S_{pO_2}$ and $S_{aO_2}$ with currently available oximeters is required. The objective of this study was to describe the accuracy and bias of the $S_{pO_2}$ value measured by several oximeters (Nonin, Nellcor, Masimo, and Philips) frequently used in clinical practice compared to the reference value, $S_{aO_2}$, measured by arterial blood gases in stable subjects admitted in an ICU.

SEE THE RELATED EDITORIAL ON PAGE 708

**QUICK LOOK**

**Current knowledge**

An accurate $S_{pO_2}$ value is critical in order to optimally titrate oxygen delivery to patients and to follow oxygenation guidelines. Oximeters' accuracy has not been well evaluated in subjects, although real-life conditions may reduce pulse oximetry accuracy in comparison with evaluations conducted in healthy volunteers with hands wrapped in a warming pad.

**What this paper contributes to our knowledge**

The present study shows that bias and accuracy of oximeters were variable in these medical instruments. We found significant bias and moderate accuracy between the tested oximeters and the arterial blood gases in the studied population. Oximeters are not all similar and provide different information. These discrepancies may have important impact on management of oxygen therapy and on clinical decision-making related to oxygenation.

## Methods

This study compared $S_{pO_2}$ values from 4 oximeters to concomitant $S_{aO_2}$ on 193 subjects in the ICU at Institut Universitaire de Cardiologie et de Pneumologie de Québec (IUCPQ), in Québec City, Canada, between March 24, 2021–September 17, 2021. The study was approved by the IUCPQ Research Ethics Committee with a waiver of consent.

To be eligible, adult subjects (age ≥ 18 y) were to be stable in terms of oxygen administration parameters (no change in $F_{IO_2}$ and $O_2$ flow for the last 30 min) and hospitalized in the ICU with an arterial catheter in place. Exclusion criteria were instability requiring frequent ventilator parameter modifications, poor $S_{pO_2}$ signal (based on the pulse oximetry waveform), recent ICU admission (considered unstable), infectious isolation, $S_{pO_2} > 96\%$, methemoglobinemia > 1.5%, and presence of dark nail polish.[33] Subjects could not be studied more than once per day.

The evaluated oximeters were Nonin (Plymouth, Minnesota) embedded in the $FreeO_2$ device (OxyNov, Québec City, Québec, Canada), Masimo (non–touch screen Radical-7, Masimo, Irvine, California), Philips (FAST, Amsterdam, the Netherlands), and Nellcor (N-600, Medtronic, Minneapolis, Minnesota). The slowest moving average time for $S_{pO_2}$ display was set on each of the oximeters (Nonin, 4 beats; Nellcor, 5–7 s; Masimo, 4 s; Philips, 5 s). Reusable digital sensors for all oximeters were used.

Before scheduled arterial blood gases, the oximeter sensors were placed on fingers in random order.[32] The fingers were randomized for the 4 tested oximeters (Nonin, Nellcor, Masimo, and Philips). The fingers used were the index, middle, and ring fingers of each hand.

Once all oximeters were in place, 2 min passed before the measurements were taken to ensure stability of the values. We used the pulse oximetry waveform to determine a signal quality for all oximeters. Four different technicians evaluated the pulse oximetry waveform for the included subjects. They all required mutual agreement on the quality of waveform. Then, each oximeter's $S_{pO_2}$ values were collected simultaneously to the arterial blood gas. $S_{aO_2}$ was used as the reference standard since it is commonly used as the reference in multiple previous studies (functional saturation was reported as recommended).[20-24] It was determined by multiwavelength oximetry (ABL800 FLEX, Radiometer, Copenhagen, Denmark).

Drs Blanchet, Delobel, Nayet, Bouchard, and Lellouche and Dr Mercier and Mr Simard are affiliated with Department of Medicine, Centre de recherche, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Québec, Canada. Dr L'Her is affiliated with Medical Intensive Care Unit, CHRU Brest, La Cavale Blanche, Brest, France. Mr Branson is affiliated with Department of Surgery, Division of Trauma and Critical Care, University Cincinnati, Cincinnati, Ohio.

Dr L'Her is a co-inventor of the $FreeO_2$ device used in this study and a co-founder of OxyNov Inc. Dr L'Her discloses relationships with GE Healthcare, Sedana Medical, Vygon, and Archeon. Mr Branson discloses relationships with Lungpacer, Engineered Medical Systems, Ventec Life Systems, and Vyaire. Mr Branson is Editor-in-Chief of RESPIRATORY CARE. Dr Lellouche is a co-inventor of the $FreeO_2$ device used in this study and a co-founder of OxyNov Inc. The remaining authors have disclosed no conflicts of interest.

Supplementary material related to this paper is available at http://www.rcjournal.com.

ClinicalTrials.gov registration NCT04772183.

Correspondence: François Lellouche MD PhD, Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec, 2725, Chemin Sainte-Foy, Québec, QC, Canada, G1V 4G5. E-mail: francois.lellouche@criucpq.ulaval.ca.

DOI: 10.4187/respcare.10582

The primary outcome measurements were $S_{pO_2}$ values for each evaluated oximeter and $S_{aO_2}$ values from the arterial blood gases. Subjects' vital signs (heart rate, breathing frequency, blood pressure, temperature, and oxygen saturation at inclusion) were collected, as well as cardiac index if available within the last 2 h for subjects after cardiac surgery. Skin color was determined using the Fitzpatrick scale;[37] comorbidities such as diabetes mellitus, presence of peripheral vascular disease, and left heart failure (defined by left ventricular ejection fraction < 40%) were all collected, as well as vasopressors or inotrope medication doses and the type of respiratory support. The main laboratory results collected were $S_{aO_2}$, pH, $HCO_3^-$, $P_{aCO_2}$, $P_{aO_2}$, lactate, methemoglobinemia, carboxyhemoglobinemia, hemoglobin, and glycosylated hemoglobin.[38]

### Statistical Analysis

From a previous unpublished experiment conducted in patients, with 42 measurements, the mean difference between 2 oximeters was 1.1 ± 1.2%. At a global significance level of 5% and applying the Bonferroni criteria for multiple comparisons, 210 records were required to detect a mean difference of 0.5% using a power of 95%. Using 35 randomization blocks, 210 participants could be enrolled. Continuous variables are reported as mean ± SD or median with interquartile range as appropriate. Nominal variables are reported as frequencies. The primary outcome evaluates the accuracy (random error) and bias (systematic error) of the $S_{pO_2}$ value measured by different oximeters frequently used in the clinical practice compared to the reference value of $S_{aO_2}$ measured by arterial gases in subjects managed in the ICU. The first approach was the use of Bland-Altman plot to analyze the agreement between values of $S_{aO_2}$ measured by arterial gases and values of $S_{pO_2}$ measured by oximeters. We reported mean bias, precision, and 95% CI. Mean bias was calculated as $S_{pO_2} - S_{aO_2}$ from each oximeter's value and the corresponding arterial blood value. Precision was calculated as SD of the bias. We reported root mean square error (RMSE), a parameter required by the Food and Drug Administration (FDA) for pulse oximeters' certification. RMSE is derived from calculations involving bias and precision.

Intra-class correlations were reported for assessing agreement between the reference values for $S_{aO_2}$ and pulse oximetry values from oximeters separately. Relationships between the reference values for $S_{aO_2}$ and pulse oximetry values from oximeters were analyzed using Pearson correlation coefficient. The linear mixed model was performed on a Latin-square design using an unrestricted covariance structure to compare oximeters (fixed factor). Blocks were linked to the random effect. The normality assumption was verified with the Shapiro-Wilk tests using residuals from the statistical model and transformed by the Cholesky metric. The Brown and Forsythe variation of Levene test statistic was used to verify the homogeneity of variances. Test was performed at 5% level.

Factors that modified the $S_{pO_2} - S_{aO_2}$ gap for each tested oximeter were analyzed using a linear regression model. Multivariable linear regressions were performed to determine independent predictors. Continuous variables were checked for the assumption of linearity. The final model was selected on the smallest Akaike criteria. The statistical analysis was not planned for multiplicity in test for secondary outcomes. All reports estimates are expressed using mean and 95% CI. Some baseline characteristics were used to investigate if any may explain the variability among oximeters by adding interaction terms between oximeters and variables. The hypoxemia detection was analyzed for each oximeter. The areas under the curve were compared using univariable models. Analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina) and R package (R Foundation for Statistical Computing, Vienna, Austria).

## Results

We collected 1,055 simultaneous measurements of oxygen saturation (one $S_{aO_2}$ and 4 $S_{pO_2}$ values for 211 inclusions in 193 subjects), 153 inclusions (73%) in males and 58 (27%) in females. Subjects could only be included once a day in the study. The mean age was 66 ± 11 y. The skin pigmentation, using the Fitzpatrick scale, was of type 1 or 2 in 96.2% of subjects. Sixty-four subjects had diabetes; 32 had peripheral vascular disease; 24 had left heart failure; 60 subjects received vasopressors or inotropes at time of measurement; 22 subjects received invasive ventilation (mean $F_{IO_2}$ was 0.37 ± 0.09); 15 received high-flow oxygen therapy; 3 received noninvasive ventilation; 87 received conventional oxygen therapy (mean $O_2$ flow was 2.9 ± 4.8 L/min), and 83 did not have respiratory support (Table 1 and study flow chart Figure S1, see related supplementary materials at http://www.rcjournal.com). Thirty-one subjects had tachycardia (heart rate > 100 beats/min); 15 were hypotensive (systolic blood pressure < 100 mm Hg), and none were hypothermic (core temperature < 35°C).

Mean $S_{aO_2}$ value was 93.6 ± 2.2%. There was positive bias of +0.9% for the Philips oximeter. There was negative bias for the oximeters Nonin, Nellcor, and Masimo, with bias of −3.1, −0.3, and −0.2%, respectively (Table 2). A Bland-Altman plot was drawn for each oximeter to illustrate agreement between $S_{pO_2}$ and $S_{aO_2}$. Mean bias ± precision and 95% CI were −3.1 ± 2.1 (−7.2 to 11.0) for Nonin, −0.3 ± 1.7 (−3.6 to 3.1) for Nellcor, −0.2 ± 1.9 (−4.0 to 3.6) for Masimo, and 0.9 ± 1.9 (−2.9 to 4.7) for Philips (Fig. 1). The Nonin oximeter underestimated $S_{aO_2}$ in 90.3% of $S_{pO_2}$ values, whereas Philips overestimated $S_{aO_2}$ in 55% of the cases (Table 2). Overall, the frequency of differences ≥ plus or

Table 1. Subject Characteristics and Clinical Features

| Characteristics and Clinical Features | ($n = 211$) |
|---|---|
| Age, y | 66 ± 11 |
| Men | 153 (73) |
| BMI, kg/m$^2$ | 29 ± 7 |
| Skin pigmentation Fitzpatrick scale | |
| Type 1 | 85 (40) |
| Type 2 | 118 (56) |
| Type 3 | 6 (3) |
| Type 4 | 1 (0) |
| Type 5 | 0 |
| Type 6 | 1 (0) |
| Location of the subjects | |
| Postoperative cardiac surgery ICU | 193 (91) |
| Respiratory/medical ICU | 18 (9) |
| Comorbidities | |
| Diabetes mellitus | 64 (31) |
| Peripheral vascular disease | 32 (15) |
| Left heart failure (LVEF < 40%) | 24 (11) |
| Vital signs | |
| $S_{pO_2}$ at inclusion, % | 93 ± 2 |
| Heart rate, beats/min | 85 ± 14 |
| Breathing frequency, breaths/min | 19 ± 5 |
| Temperature, °C | 36.7 ± 0.7 |
| Blood pressure | |
| Systolic, mm Hg | 126 ± 46 |
| Diastolic, mm Hg | 63 ± 14 |
| Respiratory support | |
| Mechanical ventilation | 22 (10) |
| $F_{IO_2}$ | 0.40 (0.30–0.40) |
| PEEP, cm $H_2O$ | 5 (5–8) |
| Noninvasive ventilation | 3 (1) |
| $F_{IO_2}$ | 0.35 (0.33–0.38) |
| PEEP, cm $H_2O$ | 5 (5–5) |
| Nasal high flow | 15 (7) |
| $F_{IO_2}$ | 0.35 (0.30–0.45) |
| Standard oxygen therapy | 87 (41) |
| Flow, L/min | 1.5 (1.0–2.0) |
| None | 83 (40) |
| Vasopressors and inotropes | 72 (34) |
| Noradrenaline | 18 (9) |
| Adrenaline | 34 (16) |
| Vasopressin | 5 (2) |
| Milrinone | 15 (7) |
| Dobutamine | 0 |
| Arterial blood gas analysis | |
| pH | 7.41 ± 0.05 |
| $P_{aCO_2}$, mm Hg | 40 ± 7 |
| $HCO_3^-$, mmol/L | 25 ± 3 |
| $P_{aO_2}$, mm Hg | 71 ± 9 |
| $S_{aO_2}$, % | 93.6 ± 2.2 |
| Lactates, mmol/L | 1.8 ± 1.8 |
| $F_{IO_2}$ | 0.23 (0.21–0.30) |
| Methemoglobinemia, % | 1.0 ± 0.3 |
| Carboxyhemoglobin, % | 1.5 ± 0.4 |
| Hemoglobin, g/L | 98 ± 17 |

Data are presented as *n* (%), median (interquartile range), or mean ± SD.
BMI = body mass index
LVEF = left ventricular ejection fraction
$S_{aO_2}$ = arterial oxygen saturation (as measured by blood gases)
$HCO_3^-$ = bicarbonate

minus 4% were present in 35.2, 2.4, 8.6, and 9.5% of the pairs with Nonin, Nellcor, Masimo, and Philips, respectively.[22] Bias ≥ plus or minus 10% was present in 1.5% of the pairs with the Nonin oximeter (Table 2). The impact of several clinical and biological parameters on the bias was assessed (peripheral arterial disease, heart failure, type of ventilation, vital signs, diabetes mellitus). There was only a minimal impact for the parameters tested. The impact of skin pigmentation could not be adequately assessed given the low number of subjects with dark-skin pigmentation (Table S1 and S2, see related supplementary materials at http://www.rcjournal.com).

In terms of accuracy, the tested oximeters all showed large random errors. The coefficient of determination ($r^2$) for each oximeter expressed moderate correlation between $S_{pO_2}$ values and $S_{aO_2}$ values (Figs. 1 and S2). The Nellcor oximeter had an $r^2$ coefficient that expressed a slightly greater accuracy than the other oximeters, with an $r^2$ value of 0.51. The 3 other oximeters $r^2$ coefficients were, respectively, 0.43, 0.43 and 0.44 for Nonin, Masimo, and Philips. The RMSE for each oximeter was of 3.73, 1.72, 1.95, and 2.13 for Nonin, Nellcor, Masimo, and Philips, respectively.

Oximeters frequently show diverging values for the same measurement. Nonin and Philips oximeters have shown the most significant bias, with mean bias of 4.0% (Table S3, see related supplementary materials at http://www.rcjournal.com). Masimo and Nellcor oximeters have shown concordant results, as the mean bias between these 2 oximeters' $S_{pO_2}$ values was of 0.1% (Table S3). The relationship between oximeters was represented by point clouds with poor determination coefficient ($r^2$), as shown on the Figure S5 (See related supplementary materials at http://www.rcjournal.com).

$S_{aO_2}$ and $P_{aO_2}$ values allow detection of hypoxemia commonly defined as $S_{aO_2} < 90\%$ and $P_{aO_2} < 60$ mm Hg.[20] The ability to detect hypoxemia was variable between the tested oximeters. In the included subjects, 19 had a $P_{aO_2}$ value from 55–60 mm Hg, and 12 had an $S_{aO_2}$ value from 86–90%, suggesting moderate hypoxemia. The Nonin oximeter was the most sensitive of the 4 tested oximeters for detection of hypoxemia. It detected 100% of the states of

Table 2. Main Features of Evaluated Oximeters

| | $S_{aO_2}$ (Reference) | $S_{pO_2}$ | | | |
|---|---|---|---|---|---|
| | | Nonin | Nellcor | Masimo | Philips |
| $O_2$ saturation, % | 93.6 ± 2.2 | 90.5 ± 2.8 | 93.3 ± 2.3 | 93.4 ± 2.4 | 94.5 ± 2.4 |
| Mean bias, $S_{pO_2}-S_{aO_2}$, % | | −3.1* | −0.3† | −0.2 | 0.9* |
| Precision SD of the bias | | 2.1 | 1.7 | 1.9 | 1.9 |
| Bias 95% CI | | (−3.4 to −2.8) | (−0.5 to 0) | (−0.5 to 0.1) | (0.7−1.2) |
| Overestimation of $S_{aO_2}$(% measurements) | | 3.9 | 28.2 | 30.8 | 55.2 |
| Underestimation of $S_{aO_2}$(% measurements) | | 91.3 | 41.1 | 42.8 | 21.4 |
| Bias < −4 or ≥ +4% | | 73 (35) | 5 (2) | 18 (9) | 20 (10) |
| Bias < −10 or ≥ +10% | | 3 (1.5) | 0 | 0 | 0 |
| Detection of hypoxemia | | | | | |
| $S_{aO_2}$ < 90%‡ | | 11 (92) | 4 (33) | 5 (42) | 2 (17) |
| $P_{aO_2}$ < 60 mm Hg‡ | | 19 (100) | 5 (26) | 7 (37) | 2 (11) |

Data are presented as *n* (%) or mean ± SD unless otherwise noted.
$S_{aO_2}$ = arterial oxygen saturation (as measured by blood gases)
* $P < .001$.
† $P < .05$, comparison with $S_{aO_2}$ (paired *t* test).
‡Based on 12 subjects for $S_{aO_2} < 90\%$ and 19 subjects for $P_{aO_2} > 60$ mm Hg.
Overestimation was defined by a $S_{pO_2}$ value > $S_{aO_2}$ value.
Underestimation was defined by a $S_{pO_2}$ value < $S_{aO_2}$ value.

moderate hypoxemia. On the other hand, the Philips oximeter detected only 11% of hypoxemic events. Masimo and Nellcor detected 37% and 26% of hypoxemic events, respectively (Figs. 2, 3, and S4, see related supplementary materials at http://www.rcjournal.com).

Based on the cohort of the 211 concomitant measurements of oxygen saturation, we evaluated for each oximeter if oxygen (or $F_{IO_2}$) should be increased, decreased, or maintained constant to keep the subject within an $S_{pO_2}$ target of 92–96%. With this objective, it would require an increase oxygen in 60.6% versus 19.1% versus 18.7% versus 10.4% and to decrease oxygen in 1.0% versus 6.2% versus 9.1% versus 20.9% of the cases ($P < .001$) when using the Nonin, Nellcor, Masimo, or Philips oximeter, respectively.

## Discussion

Our study showed that in our population of stable subjects hospitalized in ICUs tested pulse oximeters had moderate accuracy (random error). In addition, each oximeter had significant differences for bias (systematic error) when compared to the reference standard, $S_{aO_2}$. The largest bias was shown with Nonin oximeter, with a mean bias of −3.1%, followed by Philips oximeter with a mean bias of +0.9%. Masimo and Nellcor oximeters did not show a significant bias, but, as for all tested oximeters, the correlation between $S_{pO_2}$ and $S_{aO_2}$ showed a moderate accuracy. Indeed, in about 14% of the subjects, the bias between pulse oximeters saturation and $S_{aO_2}$ was ≥ 4%; and in 0.4% of the measures, the bias reached ≥ 10%. Large bias was more frequent with the Nonin oximeter due to its systematic negative bias. These characteristics explain the clinical impact of the tested oximeters. Nonin oximeter was able to detect 100% of hypoxemia, whereas the Philips oximeter only detected 11%. In opposition, the oxygen weaning was much more frequent with the Philips oximeter in comparison with Nonin oximeter when usual $S_{pO_2}$ targets were utilized. The characteristics of the oximeters should be known by the users to adequately define the $S_{pO_2}$ target for patients based on the oximeter used. To our knowledge, this is the most recent study that compared several oximeters to concomitant $S_{aO_2}$ in ICU subjects.

The data presented in this study revealed that oximeters had a larger bias when tested critically ill subjects rather than healthy subjects. Previous studies conducted in healthy subjects found bias of −0.21% for Nonin, and −0.27% for Nellcor, on light-skinned healthy subjects (pigmentation similar to our population).[26] Another study conducted on healthy subjects found bias of about +1.0 to −1.5% for Nonin oximeter, of about −1 to +1% for Masimo, and of about −0.5% to +1% for Nellcor on light-skinned subjects.[27] In the present study, Nonin oximeter bias found in subjects was much greater than those found in previous studies conducted on healthy subjects, which might in part be related to the light-skin pigmentation (> 95% of the subjects included in the study). Pulse oximetry has been shown to overestimate $S_{aO_2}$ in dark-skin patients.[20,36] Therefore, with the Nonin oximeter, the bias might be lower (less underestimation) on dark-skinned patients, whereas for the Philips the bias would likely be greater.[36] Several other conditions that may affect arterial perfusion and oximeter accuracy were present in the subjects included in the

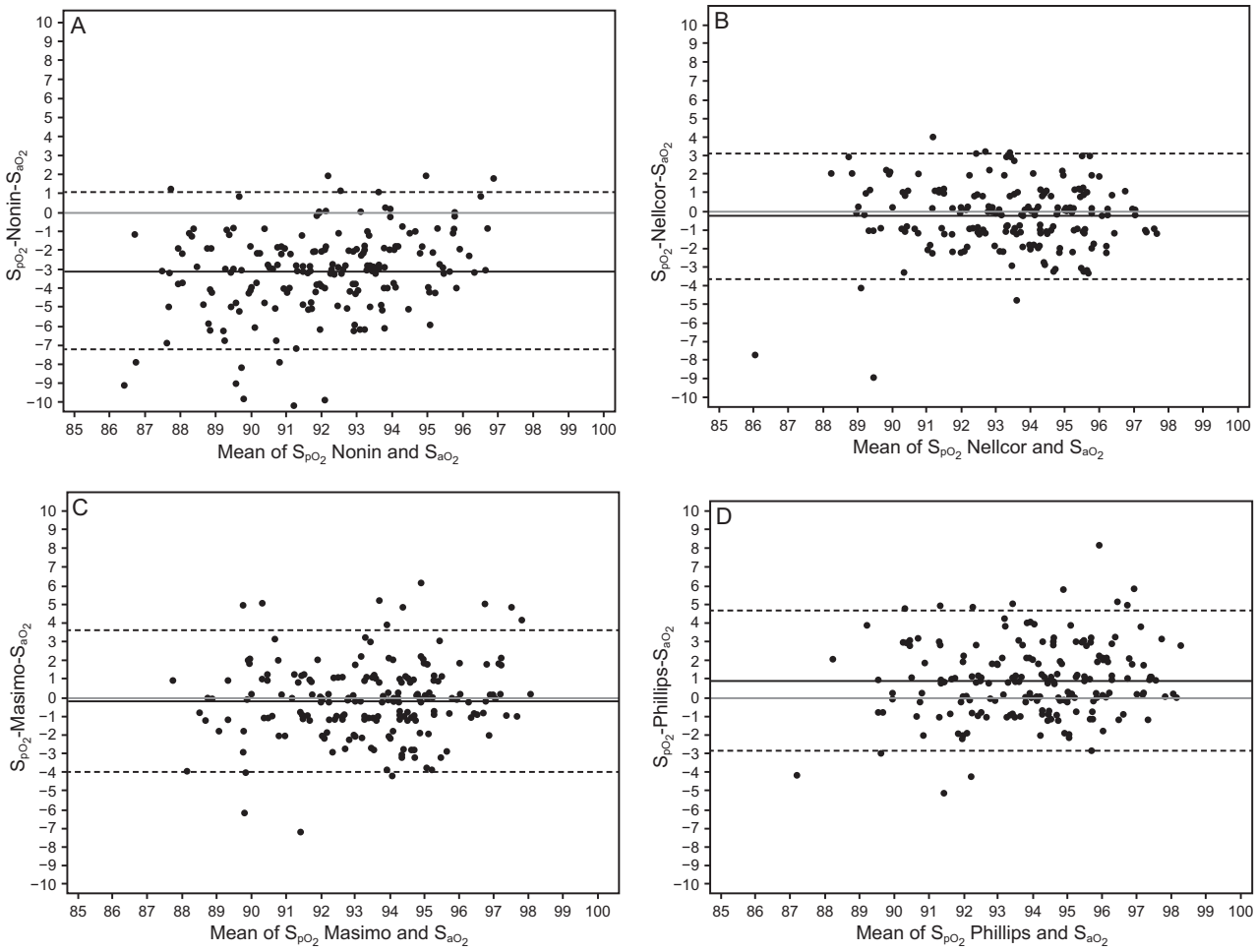


Fig. 1. Representation of the Bland-Altman plot for the comparison of the pulse oximeter saturation displayed by tested oximeters with the reference value (arterial oxygen saturation). A: Nonin, B: Nellcor, C: Masimo, and D: Philips oximeters. Black horizontal lines show the bias, and dashed lines denote 95% CI. $S_{aO_2}$ = arterial oxygen saturation.

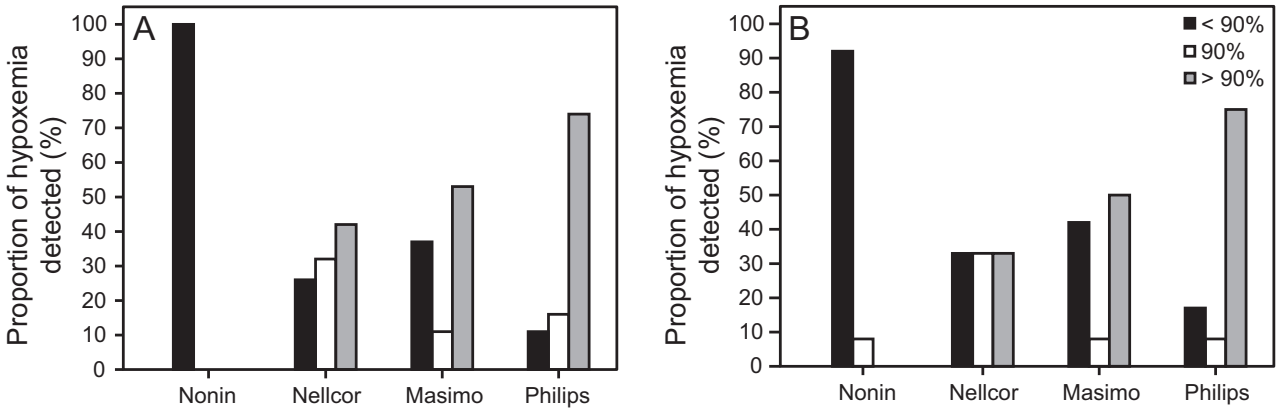


Fig. 2. Detection of arterial hypoxemia defined by $P_{aO_2} < 60$ mm Hg (A: in 19 subjects that met this definition) and $S_{aO_2} < 90\%$ (B: in 12 subjects) by the tested oximeters Nonin, Nellcor, Masimo, and Philips.

present study including peripheral vascular disease (15% of included subjects), left ventricular dysfunction (11%), utilization of vasopressors (34%), as well as the presence of cardiovascular conditions such as diabetes (30%). In addition, studies that were carried out in healthy subjects were conducted under somewhat artificial conditions with subjects sitting still with each hand wrapped in a warming pad to ensure a good circulation to the fingers.[26,27,31]

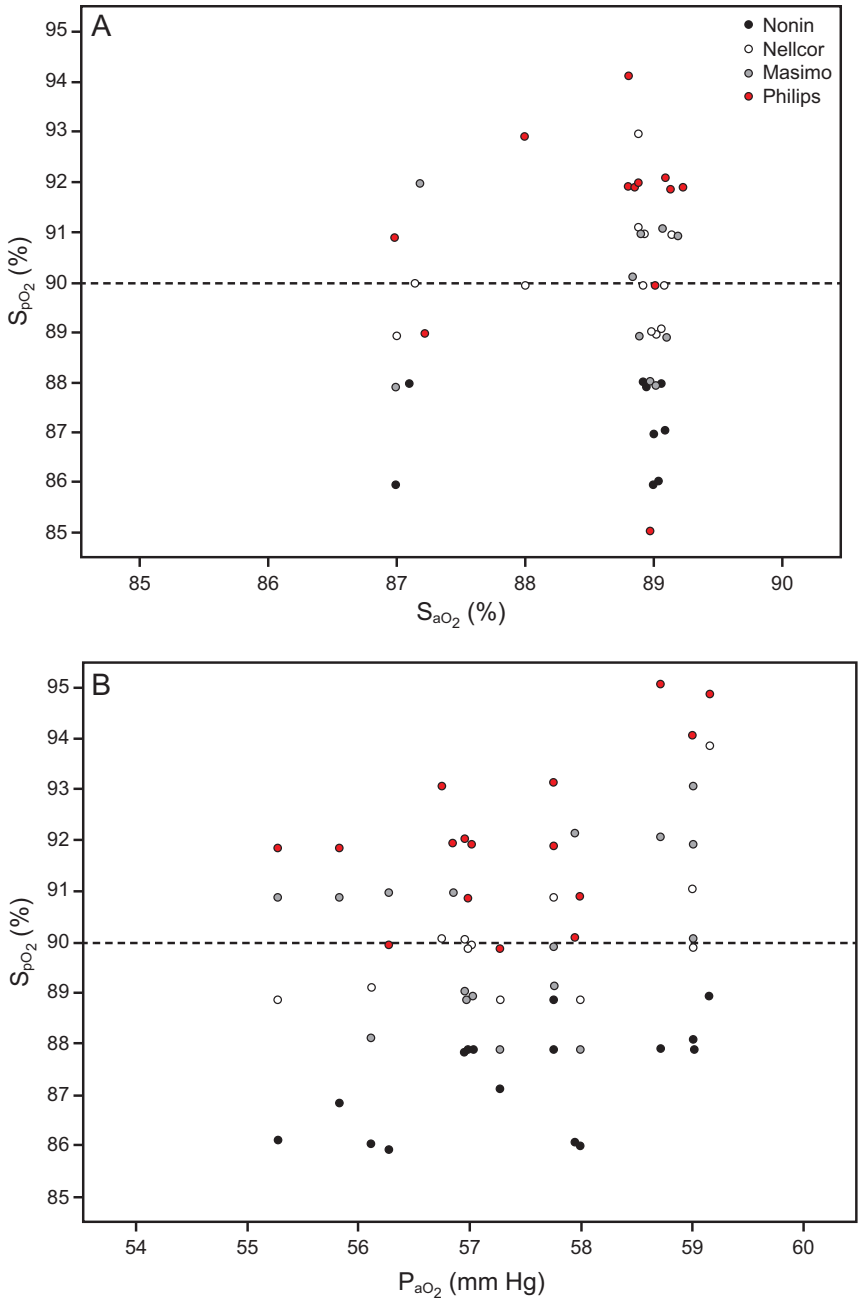


Fig. 3. $S_{pO_2}$ values of the 4 tested oximeters for $S_{aO_2}$ values < 90% (A) and for $P_{aO_2}$ values < 60 mm Hg (B). $S_{aO_2}$ = arterial oxygen saturation.

It has been recently highlighted that the bias between $S_{aO_2}$ and $S_{pO_2}$ related to skin pigmentation is more pronounced in studies conducted in subjects compared with studies conducted in healthy volunteers.[39] This is probably also true for the evaluation of oximeters, where the differences are greater in patients.

A limited number of studies comparing several oximeters have been recently conducted in subjects.[21,22,38] In those studies, none have made a systematic comparison of 4 oximeters with simultaneous arterial blood gases. A multi-center study in New Zealand/Australia collected 400 pairs of $S_{pO_2}$ and $S_{aO_2}$, but only one oximeter was evaluated for each subject.[22] Another study, conducted on mechanically ventilated subjects in the ICU in India, only compared 2 oximeters, Nonin and Philips.[21] Results from this study slightly diverge from our results, probably as a consequence of the skin pigment of the subjects. In comparison to our data, the bias found in this study was of similar amplitude between oximeters (+2.49 ± 2.99 for Philips and 0.46 ± 1.68 for Nonin) but shifted by +2–3%. The impact of

skin pigmentation on oximeters' bias has already been described.[20,26,36] Oximeters potentially overestimate $S_{aO_2}$ in patients with dark-skin pigmentation, which can explain the difference in the bias found in Singh et al[21] study and in our study. One recently published study evaluated oximeters' accuracy according to skin pigmentation but did not compare different oximeters and did not specify which oximeter was used.[36] However, it highlighted greater overestimation of $S_{aO_2}$ in dark-skinned subjects. Moreover, a recent study has shown that skin pigmentation was associated with the risk of hidden hypoxemia and increased mortality.[40]

Oximeters characteristics, such as the tendency to underestimate or overestimate, may have a clinical impact, and this should be taken into account in clinical practice. In the present study, the Nonin oximeter underestimated $S_{aO_2}$ in 91.3% of the measurements. For clinicians, this means that the $S_{aO_2}$ of the patient is probably above the displayed $S_{pO_2}$. With other tested oximeters, it is more difficult for the clinician to conclude if the $S_{aO_2}$ is above or below the displayed $S_{pO_2}$.

The Nonin oximeter has the largest bias but in return is more sensitive for the detection of hypoxemia. However, an oximeter that underestimates $S_{aO_2}$ can lead to exposing patients to hyperoxemia, which can also be deleterious.[5-12] Therefore, clinicians need to decrease $S_{pO_2}$ targets when using this type of oximeter to avoid hyperoxemia.

On the other hand, an oximeter that mainly overestimates $S_{aO_2}$, such as the Philips oximeter, lowers the risks of exposing patients to hyperoxemia. However, the main objective of oximeters (ie, to detect hypoxemia) may be missed. Therefore, clinicians may need to increase $S_{pO_2}$ targets when using this type of oximeter to avoid hypoxemia.

Concerning the oximeters that have low biases, as Nellcor and Masimo, $S_{pO_2}$ targets should not be adjusted, but clinicians must be aware of the weak accuracy of oximeters. This means that for a given $S_{pO_2}$ value displayed $S_{aO_2}$ may be as frequently both above or below expected.

There are potential clinical impacts that arise from a 2–3% bias between $S_{aO_2}$ and oximeters and from an even larger gap with several oximeters. Discrepancies related to skin pigmentation leading to occult hypoxemia in patients with dark skin have been described with potential impact on outcomes.[36,40-42] In addition, the clinical decision to titrate oxygen may differ when using different $S_{pO_2}$ target recommendations[13-15,43] and when using different oximeters. Our findings highlight the need for clinicians to be aware of the characteristics of the oximeter they use in their unit to avoid hypoxemia and hyperoxia and to adjust their $S_{pO_2}$ target. Current recommendations for $S_{pO_2}$ targets do not consider the variability related to oximeters used.[13-15] However, the present study tends to show that the variability between the oximeters and the difference between the current $S_{pO_2}$ recommendations have a synergistic effect on the oxygenation parameters. Indeed, the $O_2$ needs of a patient monitored with Philips oximeter following $S_{pO_2}$ target recommendations of 90–94%[15] would be greatly underestimated when compared with the $O_2$ needs of a patient monitored with Nonin oximeter following $S_{pO_2}$ target recommendations of 94–98%.[13] It has been shown that small variations of $S_{pO_2}$ targets may have a major impact on oxygen requirements.[16,18,19] A secondary analysis is currently underway to further analyze the impact of the oximeters' inaccuracy on $O_2$ management and on multiple clinical scores that use $S_{pO_2}$. In patients with ARDS, there is a potential risk of occult hypoxemia with an oximeter that overestimates $S_{aO_2}$ (Philips). This was shown to be potentially deleterious in severe subjects with shock.[44] On the contrary, there is a risk to use unnecessary high $F_{IO_2}$ with oximeters that underestimate $S_{aO_2}$ (Nonin), with potential risk of denitrogenation atelectasis and delayed oxygen weaning. If the $S_{pO_2}$ target is adjusted to the oximeter used, these risks may be reduced. Another clinical consequence is that $S_{pO_2}$ targets with automated oxygen titration are now decreased to 90% (and 88% in patients with COPD) in our institution as a Nonin oximeter is used with this device.[16]

Our findings suggest that targets used for the titration of $O_2$ should be adapted to the oximeter used. Moreover, oxygen-free days has recently been proposed as a possible outcome marker in clinical trials,[45] and this parameter may be influenced significantly by the type of oximeter used.[46]

Our study has a number of limitations. The number of subjects included was relatively low compared to large databases.[40] However, our method assured that the $S_{pO_2}$ values and the $S_{aO_2}$ values were taken exactly concomitantly, in subjects who had been stable for several minutes, whereas in studies using large databases the $S_{aO_2}$ value was matched with the $S_{pO_2}$ value within 5–10 min.[36,40-42] Knowing that oxygenation may vary within seconds after activity the only rigorous comparison of $S_{aO_2}$ and $S_{pO_2}$ requires concomitant measurements of $S_{pO_2}$ and $S_{aO_2}$ as we designed the study. Moreover, in the present study, the sampling made it possible to answer the main hypothesis: to demonstrate a difference of 0.5% between $S_{aO_2}$ and oximeters $S_{pO_2}$.

Our study's main limitation is the low diversity of skin pigmentation of our subjects. More than 96% of the subjects had a Fitzpatrick score of 1 or 2, which reflects the local population but limits the applicability of our results to this specific population. Similar comparisons are required on patients with darker-skin pigmentation given the significant potential impact.[36,40] In terms of detection of hypoxemia, we had a limited number of subjects with $S_{aO_2} < 90\%$, and these results must be confirmed. The absence of perfusion index value can also be raised as a limitation. However, our subjects were stable, thus not in shock, and a member of the research team made sure each patient had good perfusion before the measurements were taken. The RMSE was higher for the Nonin oximeter in the present study. However, even if it is required by the FDA, RMSE

is not the best measure of oximeter accuracy and is unintelligible for most bedside clinicians. It is likely that the oximeters' bias are shifted in patients with dark pigmentation. It was around +0.5 with Nonin and +2.5% with Philips oximeter, as shown in Singh study conducted in India, leading to higher RMSE for Philips. Consequently, the RMSE value is probably more representative when included subjects have balanced skin pigmentations, which is not the case here. It should be reported for each subjects' subgroups of skin pigmentation as recently proposed.[47]

## Conclusions

The present study shows that bias and accuracy of oximeters are variable. Each oximeter behaves differently; some overestimate the oxygenation (Philips FAST); some underestimate the oxygenation (Nonin embedded in the $FreeO_2$), and others have mean $S_{pO_2}$ values close to the mean $S_{aO_2}$ value (Nellcor N-600 and Masimo Radical-7) but with significant inaccuracy. With the latter, it is thus difficult for clinicians to determine whether these overestimate or underestimate oxygenation. All 4 of the tested oximeters showed a moderate correlation with $S_{aO_2}$. Therefore, oximeters should be used knowing their limits; and when $S_{pO_2}$ is used by clinicians to target the oxygen flow or $F_{IO_2}$, the oxygenation targets should probably be adapted based on the oximeter and skin pigmentation. For example, with guidelines that recommend a target $S_{pO_2}$ between 92–96%, on a population with the same characteristics as studied in this paper, this target might be increased to 93–97% with the Philips oximeter tested in the study and decreased to 89–93% with the Nonin oximeter tested in the study. More values could be required to assess these guidelines modification suggestions.

These marked differences may partly explain the major variations in recent recommendations for oxygen therapy: from 90–94%[15] to 92–96%[14] to 94–98%.[13,43,48] Despite the fact that the first pulse oximeters were described 50 years ago, and as these devices are now used daily by clinicians around the world, their characteristics should be better known to optimize their utilization for titrating and monitoring oxygen therapy.

### REFERENCES


1. Galatius-Jensen S, Hansen J, Rasmussen V, Bildsoe J, Therboe M, Rosenberg J. Nocturnal hypoxemia after myocardial infarction: association with nocturnal myocardial ischemia and arrhythmias. Br Heart J 1994;72(1):23-30.
2. Gill NP, Wright B, Reilly CS. Relationship between hypoxemic and cardiac ischemic events in the perioperative period. Br J Anaesth 1992;68(5):471-473.
3. Jubran A. Pulse oximetry. Crit Care 2015;19(1):272.
4. Sun Z, Sessler DI, Dalton JE, Devereaux PJ, Shahinyan A, Naylor AJ, et al. Postoperative hypoxemia is common and persistent: a prospective blinded observational study. Anesth Analg 2015;121(3):709-715.
5. Aubier M, Murciano D, Fournier M, Milic-Emili J, Pariente R, Derenne JP. Central respiratory drive in acute respiratory failure of patients with chronic obstructive pulmonary disease. Am Rev Respir Dis 1980;122(2):191-199.
6. Austin MA, Wills KE, Blizzard L, Walters EH, Wood-Baker R. Effect of high-flow oxygen on mortality in chronic obstructive pulmonary disease patients in prehospital setting: randomized controlled trial. BMJ 2010;341:c5462.
7. Cameron L, Pilcher J, Weatherall M, Beasley R, Perrin K. The risk of serious adverse outcomes associated with hypoxemia and hyperoxemia in acute exacerbations of COPD. Postgrad Med J 2012;88 (1046):684-689.
8. Chu DK, Kim LH, Young PJ, Zamiri N, Almenawer SA, Jaeschke R, et al. Mortality and morbidity in acutely ill adults treated with liberal versus conservative oxygen therapy (IOTA): a systematic review and meta-analysis. Lancet 2018;391(10131):1693-1705.
9. Farquhar H, Weatherall M, Wijesinghe M, Perrin K, Ranchord A, Simmonds M, et al. Systematic review of studies of the effect of hyperoxia on coronary blood flow. Am Heart J 2009;158(3):371-377.
10. McNulty PH, King N, Scott S, Hartman G, McCann J, Kozak M, et al. Effects of supplemental oxygen administration on coronary blood flow in patients undergoing cardiac catheterization. Am J Physiol Heart Circ Physiol 2005;288(3):H1057-1062.
11. Floyd TF, Ratcliffe SJ, Detre JA, Woo YJ, Acker MA, Bavaria JE, et al. Integrity of the cerebral blood-flow response to hyperoxia after cardiopulmonary bypass. J Cardiothorac Vasc Anesth 2007;21(2):212-217.
12. Stub D, Smith K, Bernard S, Nehme Z, Stephenson M, Bray JE, et al; AVOID Investigators. Air versus oxygen in ST-segment elevation myocardial infarction. Circulation 2015;131(24):2143-2150.
13. O'Driscoll BR, Howard LS, Earis J, Mak V; British Thoracic Society Emergency Oxygen Guideline Group; Group BTSEOGD Group. BTS guideline for oxygen use in adults in health care and emergency settings. Thorax 2017;72(Suppl 1):ii1-ii90.
14. Beasley R, Chien J, Douglas J, Eastlake L, Farah C, King G, et al. Thoracic Society of Australia and New Zealand oxygen guidelines for acute oxygen use in adults: "Swimming between the flags." Respirology 2015;20(8):1182-1191.
15. Siemieniuk RAC, Chu DK, Kim LH, Guell-Rous MR, Alhazzani W, Soccal PM, et al. Oxygen therapy for acutely ill medical patients: a clinical practice guideline. BMJ 2018;363:k4169.
16. Lellouche F, L'Her E. Automated oxygen flow titration to maintain constant oxygenation. Respir Care 2012;57(8):1254-1262.
17. Lellouche F, Bouchard PA, Roberge M, Simard S, L'Her E, Maltais F, et al. Automated oxygen titration and weaning with $FreeO_2$ in patients with acute exacerbation of COPD: a pilot randomized trial. Int J Chron Obstruct Pulmon Dis 2016;11:1983-1990.
18. L'Her E, Dias P, Gouillou M, Riou A, Souquiere L, Paleiron N, et al. Automatic versus manual oxygen administration in the emergency department. Eur Respir J 2017;50(1).
19. L'Her E, Jaber S, Verzilli D, Jacob C, Huiban B, Futier E, et al. Automated closed-loop versus standard manual oxygen administration after major abdominal or thoracic surgery: an international multi-center randomized controlled study. Eur Respir J 2021;57(1).
20. Jubran A, Tobin MJ. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. Chest 1990; 97(6):1420-1425.
21. Singh AK, Sahi MS, Mahawar B, Rajpurohit S. Comparative evaluation of accuracy of pulse oximeters and factors affecting their performance in a tertiary intensive care unit. J Clin Diagn Res 2017;11(6): OC05-OC08.
22. Pilcher J, Ploen L, McKinstry S, Bardsley G, Chien J, Howard L, et al. A multi-center prospective observational study comparing arterial blood gas values to those obtained by pulse oximeters used in adult

patients attending Australian and New Zealand hospitals. BMC Pulm Med 2020;20(1):7.
23. Van de Louw A, Cracco C, Cerf C, Harf A, Duvaldestin P, Lemaire F, et al. Accuracy of pulse oximetry in the intensive care unit. Intensive Care Med 2001;27(10):1606-1613.
24. Clayton DG, Webb RK, Ralston AC, Duthie D, Runciman WB. A comparison of the performance of 20 pulse oximeters under conditions of poor perfusion. Anaesthesia 1991;46(1):3-10.
25. ISO. 80601-2- 61:2011, Medical electrical equipment — Part 2–61: Particular requirements for basic safety and essential performance of pulse oximeter equipment.
26. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology 2005; 102(4):715-719.
27. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg 2007;105(6 Suppl):S18-s23.
28. Henry NR, Hanson AC, Schulte PJ, Warner NS, Manento MN, Weister TJ, Warner MA. Disparities in hypoxemia detection by pulse oximetry across self-identified racial groups and associations with clinical outcomes. Crit Care Med 2022;50(2):205-211.
29. Crooks CJ, West J, Morling JR, Simmonds M, Juurlink I, Briggs S, et al. Pulse oximeter measurements vary across ethnic groups: an observational study in patients with COVID-19. Eur Respir J 2022;59(4).
30. Tobin MJ, Jubran A. Pulse oximetry, racial bias, and statistical bias. Ann Intensive Care 2022;12(1):2.
31. Louie A, Feiner JR, Bickler PE, Rhodes L, Bernstein M, Lucero J. Four types of pulse oximeters accurately detect hypoxia during low perfusion and motion. Anesthesiology 2018;128(3):520-530.
32. Basaranoglu G, Bakan M, Umutoglu T, Zengin SU, Idin K, Salihoglu Z. Comparison of SpO2 values from different fingers of the hands. Springerplus 2015;4:561.
33. Hinkelbein J, Koehler H, Genzwuerker HV, Fiedler F. Artificial acrylic finger nails may alter pulse oximetry measurement. Resuscitation 2007;74(1): 75-82.
34. Nesseler N, Frénel JV, Launey Y, Morcet J, Mallédant Y, Seguin P. Pulse oximetry and high-dose vasopressors: a comparison between forehead reflectance and finger transmission sensors. Intensive Care Med 2012;38(10):1718-1722.
35. Wilson BJ, Cowan HJ, Lord JA, Zuege DJ, Zygun DA. The accuracy of pulse oximetry in emergency department patients with severe sepsis and septic shock: a retrospective cohort study. BMC Emerg Med 2010;10:9.
36. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020;383(25): 2477-2478.
37. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol 1988;124(6):869-871.
38. Pu LJ, Shen Y, Lu L, Zhang RY, Zhang Q, Shen WF. Increased blood glycohemoglobin A1c levels lead to overestimation of arterial oxygen saturation by pulse oximetry in patients with type 2 diabetes. Cardiovasc Diabetol 2012;11:110.
39. Okunlola O, Lipnick MS, Batchelder P, Bernstein M, Feiner J, Bickler P. Pulse oximeter performance, racial inequity, and the work ahead. Respir Care 2022;67(2):252-257.
40. Wong AI, Charpignon M, Kim H, Josef C, de Hond AAH, Fojas JJ, et al. Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality. JAMA Netw Open 2021;4(11): e2131674.
41. Valbuena VSM, Seelye S, Sjoding MW, Valley TS, Dickson RP, Gay SE, et al. Racial bias and reproducibility in pulse oximetry among medical and surgical inpatients in general care in the Veterans Health Administration 2013–19: multi-center, retrospective cohort study. BMJ 2022;378:e069775).
42. Fawzy A, Wu TD, Wang K, Robinson ML, Farha J, Bradke A, et al. Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19. JAMA Intern Med 2022;182(7):730-738.
43. Piraino T, Madden M, K JR, Lamberti J, Ginier E, S LS. Management of adult patients with oxygen in the acute care setting. Respir Care 2022;67(1):115-128.
44. Barrot L, Asfar P, Mauny F, Winiszewski H, Montini F, Badie J, et al; LOCO2 Investigators and REVA Research Network. Liberal or conservative oxygen therapy for acute respiratory distress syndrome. N Engl J Med 2020;382(11):999-1008.
45. Moskowitz A, Shotwell MS, Gibbs KW, Harkins M, Rosenberg Y, Troendle J, et al. Oxygen-free days as an outcome measure in clinical trials of therapies for COVID-19 and other causes of new-onset hypoxemia. Chest 2022;162(4):804-814.
46. Lellouche F, Blanchet M, Branson R. Oxygen-free days and the confounders of clinical practice. Chest Dec 2022;162(6):e331-e332.
47. Sjoding MW, Iwashyna TJ, Valley TS. Change the framework for pulse oximeter regulation to ensure clinicians can give patients the oxygen they need. Am J Respir Crit Care Med 2023;207(6):661-664.
48. O'Driscoll BR, Howard LS, Earis J, Mak V. British Thoracic Society Guideline for oxygen use in adults in health care and emergency settings. BMJ Open Respir Res 2017;4(1):e000170.