How Using Different Oximeters May Affect Clinical Decision-Making : A Method Comparison Study in Patients Receiving CPAP or Noninvasive Ventilation

Karege, Gatete; Egger, Bernard; Lawi, David; Bergeron, Anne; Janssens, Jean-Paul

How to cite

KAREGE, Gatete et al. How Using Different Oximeters May Affect Clinical Decision-Making : A Method Comparison Study in Patients Receiving CPAP or Noninvasive Ventilation. In: Chest, 2026, vol. 169, n° 1, p. 244–256. doi: 10.1016/j.chest.2025.08.018

This publication URL: https://archive-ouverte.unige.ch/unige:190463
Publication DOI: 10.1016/j.chest.2025.08.018

CHEST

How Using Different Oximeters May Affect Clinical Decision-Making

A Method Comparison Study in Patients Receiving CPAP or Noninvasive Ventilation

Gatete Karege, MD; Bernard Egger, MD; David Lawi, MD; Anne Bergeron, MD, PhD; and Jean-Paul Janssens, MD

BACKGROUND: Nocturnal pulsed oxygen saturation (Spo_2) monitoring is recommended for detecting residual sleep-disordered breathing, including nocturnal hypoventilation, in patients treated by noninvasive ventilation (NIV) or CPAP. It is a general assumption that different pulse oximetry devices will provide similar results. However, this may not be correct.

RESEARCH QUESTION: Does simultaneous nocturnal recording with different pulse oximetry devices lead to clinically meaningful differences?

STUDY DESIGN AND METHODS: In this prospective observational study, stable patients receiving CPAP or NIV with 3 different devices (a transcutaneous oxycapnograph with a probe on an earlobe and 2 pulse oximeters, 1 with a finger probe and the other with a ring probe) were monitored overnight. Results were compared using Bland-Altman graphs, Cohen's κ values, and interclass correlation coefficients (ICCs).

RESULTS: In all paired comparisons of devices, limits of agreement exceeded the predefined ± 2% limit for both mean and median Spo_2 and for oxygen desaturation index. The ICC values ranged between 0.77 and 0.84 for mean and median Spo_2 and between 0.33 and 0.75 for ODI. Cohen's κ values were moderate (0.46-0.73) for all paired comparisons.

INTERPRETATION: Use of different devices for monitoring stable patients receiving NIV or CPAP, all in current use in clinical practice and used by expert centers, may have an unpredictable impact on results and ensuing clinical decision-making.

CLINICAL TRIAL REGISTRATION: ClinicalTrials.gov; No.: NCT04474574; URL: www.clinicaltrials.gov

CHEST 2026; 169(1):244-256

KEY WORDS: CPAP; monitoring pulse oximetry; noninvasive ventilation; transcutaneous capnography

ABBREVIATIONS: IQR = interquartile range; NIV = noninvasive ventilation; ODI = oxygen desaturation index; $Ptcco_2$ = transcutaneous partial pressure of co_2; SDB = sleep-disordered breathing; Sao_2 = arterial oxygen saturation; Spo_2 = pulsed oxygen saturation

AFFILIATIONS: From the Division of Pulmonology, Department of Medicine (G. K., A. B., and J.-P. J.), Geneva University Hospitals, Geneva, Switzerland; Division of Pulmonology and Pulmonary Rehabilitation Center (B. E.), Rolle Hospital, Rolle, Vaud, Switzerland; Private practice (D. L.), Geneva, Switzerland; Faculty of Medicine, University of Geneva (A. B. and J.-P. J.), Geneva, Switzerland; Division of Pulmonology, La Tour Hospital, Meyrin (J.-P. J.), Geneva, Switzerland.

CORRESPONDENCE TO: Gatete Karege, MD; email: gatete.karege@hug.ch

DOI: https://doi.org/10.1016/j.chest.2025.08.018

Take-Home Points

Research Question: Does simultaneous overnight recording with 3 different pulse oximetry devices lead to clinically meaningful differences among stable patients treated by noninvasive ventilation (NIV) or CPAP?

Results: Paired comparisons of overnight mean and median pulsed oxygen saturation (Spo_2), oxygen desaturation indexes, and identification of significant hypoxemia showed substantial differences among devices.

Interpretation: Using different devices for monitoring Spo_2 in stable patients receiving NIV or CPAP may have an unpredictable impact on results and ensuing clinical decisions.

Pulse oximetry was introduced in clinical practice in 1974 and was accepted as an American Society of Anesthesiology standard of care tool in 1986.[1] Since then, it has been used extensively in clinical practice in anesthesiology, emergency departments, ambulances, hospital practice, and home care.[2] During the COVID-19 pandemic, pulsed oxygen saturation (Spo_2) of hemoglobin was referred to as the fifth vital sign in patients with respiratory distress as an adjunct to body temperature, heart rate, respiratory rate, and peak flow.[2] For patients treated by long-term noninvasive ventilation (NIV) or nasal CPAP, nocturnal monitoring is recommended for detecting residual sleep-disordered breathing (SDB), including nocturnal hypoventilation.[3-5] This involves performing nocturnal pulse oximetry or, whenever available, nocturnal capnography, which provides simultaneous assessment of transcutaneous partial pressure of CO_2 ($Ptcco_2$) and Spo_2.[6-9]

Pulse oximeters combine the principles of spectrophotometry and plethysmography: they measure the absorption of 2 wavelengths generated by 2 light-emitting diodes, which are specific for oxygenated or reduced hemoglobin (usually 660 and 940 nm) and filter the signals to integrate the pulsatile flow of arterial blood.[2] The algorithms of pulse oximeters allow discrimination between the fluctuations of Spo_2 related to the arterial flow (pulsatile component) and the continuous (nonpulsatile) venous or tissular component.[10-12] Signals treated by microprocessors provide Spo_2 values that are considered accurate within $\pm$ 2% when compared with measured arterial oxygen saturation (Sao_2) in the 80% to 99% range.[10,13] Calibration of the devices is performed in healthy volunteers subjected to a range of hypoxic conditions.[2] At lower values ($<$ 80%), reliability of Spo_2 decreases.[14-16] Both pulse oximeters and combined $Ptcco_2$ and Spo_2 probes (used in transcutaneous capnography) estimate Spo_2 using a similar technology (however, $Ptcco_2$ and Spo_2 probes contain a photosensor that measures light reflected by tissues, whereas conventional pulse oximeters measure absorption of light beams through the tissues via a photosensor).

In patients receiving long-term NIV who have an Spo_2 of $<$ 90% for $>$ 10% of nocturnal recording time, several national guidelines recommend adding oxygen to ventilator tubing.[17-20] Indeed, in large observational studies of patients receiving long-term NIV, 36% to 40% of patients had supplemental oxygen on the ventilator.[21,22] For patients receiving CPAP, correcting oxygen desaturation index (ODI) and nocturnal Spo_2 are major end points: having prolonged or recurrent drops in Spo_2, or both, may lead to change in settings, a switch to NIV, or other adaptations of treatment, including providing supplemental oxygen in some cases.[23]

It is a general assumption that different pulse oximetry devices and transcutaneous capnography will provide similar values of Spo_2. However, this assumption may not be correct. Our main hypothesis is that different devices may not provide consistent results for nocturnal monitoring of Spo_2 in this setting, thus potentially impacting clinical decision-making. Therefore, the aims of this real-life method comparison study were: (1) to determine the agreement—or lack of agreement—between simultaneous recordings with 3 different devices of overnight mean and median Spo_2 values, receiving either CPAP or NIV; (2) to determine the agreement (or lack of) of the same devices for identifying patients with prolonged nocturnal hypoxemia using different Spo_2 thresholds; and (3) to determine if the same 3 devices provided similar results for ODI and for identifying patients with an ODI of $>$ 10 events/h vs $\leq$ 10 events/h as a marker of residual SDB with CPAP or NIV treatment.
The study was accepted by the Commission of the Canton of Geneva for Ethics of Research (Identifier: 2020-01813) and was registered at ClinicalTrials.gov (Identifier: NCT04474574).

Study Design and Methods

This study was performed at the Pulmonary Rehabilitation Centre of Rolle, Rolle Switzerland, a 40-bed specialized facility with an experienced team in the management of chronic respiratory failure and SDB and also in the use of monitoring devices for pulse oximetry and transcutaneous capnography.

Three devices currently used in our area for monitoring of patients receiving NIV or CPAP (all having received European Community Certification) were used simultaneously: device 1, the Sentec Digital Monitor with the OxiVent Sensor (transcutaneous $Ptcco_2$ and Spo_2 earlobe probe) and the SenTec Digital Monitor Software version SMB SW-V08.00 (Sentec); device 2, the ResMed complete oximetry kit for AirSense 10, AirCurve 10 CPAP, and Lumis bi-level positive pressure ventilators, which include an 8000 S finger sensor, an Xpod 3012 LP external wired pulse oximeter (Nonin Medical), and an Air10 oximeter adapter with Airview software (ResMed Schweiz GmBH); and device 3, the Viatom Checkme wrist oxygen device with a ring probe and oxygen Insight Pro version 1.87 software (Viatomtech).

Earlobe probe temperature for device 1 was set at 42° C, as recommended by the manufacturer for adults. This is well tolerated over an 8-hour period.[24,25] The averaging method can be adjusted from 2 to 32 seconds with external software, but was used at the default value in this study, that is, averaging data every 3 seconds (0.33 Hz). Reported accuracy for Spo_2 (root mean square error) for device 1 is ± 2.25%. An optional drift correction for $Ptcco_2$ exists, but this does not affect Spo_2.

The ResMed complete oximetry kit uses a reusable finger sensor, is compatible with S9 and A10 ResMed devices, and has a nonadjustable 1-Hz averaging rate. Data are accessed through Resmed Airview version 4.41.0-9.0 software or Resmed ResScan version 7.0.1.67. Accuracy is not reported by the manufacturer, to our knowledge. The Viatom Checkme wrist oxygen device has ring probe, a nonadjustable averaging rate (0.5 Hz), and a reported accuracy for Spo_2 (root mean square error) of ± 2.0% of between 80% and 99%.

Inclusion criteria for patients were: age older than 18 years, providing written informed consent for participation in the study, stable clinical condition, awake Spo_2 with room air of > 90%, long-term use of either NIV or CPAP ResMed devices compatible with use of device 2, no hemodynamic instability, no sclerodactyly, and no known abnormal hemoglobin readings. Nail polish was removed to avoid interference with pulse oximeter functioning.

Exclusion criteria were: long-term oxygen therapy, hemodynamic instability, a history of the Raynaud phenomenon or any type of vasculitis, any mechanical condition that could interfere with the quality of the signal (splint, bandage, or plaster), or acute exacerbation.[12,16,26]

Recordings lasting less than 4 hours were deleted from the analysis. For all patients included, the 3 devices were installed in the evening and uninstalled the next morning simultaneously by experienced nurses and left in place overnight.

The following parameters were recorded: age, sex, BMI, systolic and diastolic BP, medication, smoking history, indication for NIV or CPAP, and major comorbidities (predefined list). From the software of each device, the following variables were collected: mean, median, and minimal Spo_2; time spent with an Spo_2 of < 90%, < 88%, and < 80% (in minutes and as a percentage of total time of recording); total number of oxygen drops and ODI (≥ 3% drop)[27]; and mean and median heart rate (beats per minute). For device 1, mean and median $Ptcco_2$ values as well as time spent with a $Ptcco_2$ of > 6.7 kPa (50 mm Hg) also were collected. All data was recorded on Research Electronic Data Capture software (Vanderbilt University).

Statistical Analysis

Pulse oximetry data were described as mean (SD), median (interquartile range [IQR]), or count and percentage according to their distribution and nature. Bland-Altman plots were used to compare mean and median Spo_2 values and ODI values (> 3% drops) from the 3 devices, 2 by 2.[28] These plots show, for each patient, the difference between devices for Spo_2 on the y-axis (Spo_2 device A – Spo_2 device B), and the mean value between devices on the x-axis (Spo_2 device A + Spo_2 device B / 2). The graph also displays the bias (also referred to as accuracy: mean difference between the 2 measurements) and the upper and lower limits of agreement (± 1.96 × SD of bias). 95% of the values for (Spo_2 device A – Spo_2 device B) are within the limits of agreement.

Pulse oximetry data are expressed in percentages. Agreement between devices was considered acceptable if the limits of agreement did not exceed ± 2% (usually

reported limits of accuracy for Spo_2). Using the approximate formula of the SE of an SD ($\sqrt{3 \times SD^2 / n}$), it was anticipated that 50 participants should be recruited such that the width of the 95% CI of the upper limit of agreement did not exceed 1%. To analyze the reliability between results provided by the devices further, we reported interclass correlation coefficients using a 2-way mixed-effects model (ie, the interclass correlation coefficient in the Shrout and Fleiss[29] terminology).

Agreement as to impact on clinical decision-making (ie, patients fulfilling or not fulfilling criteria for nocturnal hypoxemia) was assessed by determining Cohen's κ value for predefined threshold values (≥ 10% of recording time with an Spo_2 of < 88% or < 90%). These threshold values were chosen because (1) they are provided by pulse oximeter software and (2) they are considered clinically relevant in this setting.[20,23] Significance level was set at: $\alpha = .05$ (2-sided) for all analyses.

Results

Patients were included prospectively among patients admitted for elective evaluation of a positive pressure device at Rolle Hospital between July 1, 2023, and October 30, 2023. Fifty-one patients were included. One patient was omitted from analysis because of a recording time of < 4 hours. The remaining 50 patients (20 female [40%], all White) were 68 ± 10 years of age, were normotensive (systolic BP, 135 ± 18 mm Hg; diastolic BP, 72 ± 11 mm Hg), and had an average BMI of 28.4 ± 9.4 kg/m^2; 21 patients (42%) demonstrated obesity. Thirty-one patients were receiving long-term NIV (62%) and 19 patients (38%) were receiving CPAP. Details as to medication, comorbidities, and indication for NIV or CPAP are provided in e-Table 1. Table 1 summarizes the results provided by the software for the 3 devices tested. Only device 1 provided $Ptcco_2$. For 2 patients, data for device 2 were not available because of technical problems. Total median recording time was 444 minutes (IQR, 413-488 minutes), and the minimal value was 324 minutes. Boxplots for mean and median Spo_2 and ODI are shown in Figure 1. Percentage of total recording with an Spo_2 of < 90% and < 88% was reported for each device (Table 1). For all devices and patients, median time spent with Spo_2 of < 80% was 0 minutes (IQR, 0-0 minutes).

Figures 2, 3, and 4 show Bland-Altman plots for paired comparisons of mean and median Spo_2 and ODI (≥ 3%) among devices. For all paired comparisons, limits of agreement exceeded the predefined threshold of ± 2%. Bias was > 2% when comparing mean and median Spo_2 between devices 1 and 3: the Sentec Digital Monitor tended to overestimate Spo_2 when compared with the Viatom Checkme wrist oxygen device. Values for bias and upper and lower limits of agreement are summarized in Table 2. For ODI values, limits of agreement among devices were high for all paired comparisons, and bias between devices exceeded 5 events/h when comparing device 3 with device 1 or 2.

Table 3 provides Cohen's κ value for agreement between devices for identifying patients with nocturnal hypoxemia (2 different thresholds) or residual SDB according to ODI. For all paired analyses except 1, κ values indicate only moderate agreement between devices for identifying patients with > 10% of recording time of < 90%, of < 88%, or with an ODI < 10 events/h (e-Tables 2).

Discussion

This is, to our knowledge, the first study to compare simultaneous overnight recordings of Spo_2 with 3 different devices in hemodynamically and clinically stable adults treated with NIV or CPAP. Spo_2 was measured using either a probe put on the earlobe or on a finger or a ring probe. Results showed poor agreement among devices for mean and median Spo_2 and ODI, with wide limits of agreement. In all paired comparisons, limits of agreement between devices exceeded the predefined ± 2% threshold. One of the devices (Sentec Digital Monitor) overestimated mean and median Spo_2 with a bias of > 2% when compared with the Viatom Checkme wrist oxygen device. Whatever threshold chosen for Spo_2 (> 10% of total recording time of < 90% or < 88%), agreement between devices was only moderate for identifying patients with significant hypoxemia. This was also the case for the ODI (≥ 3% drops). These differences among commonly used European Community Certification-approved devices may impact clinical decision-making.

The first important finding of this study is that using different Spo_2 devices impacts identifying whether a patient receiving NIV or CPAP is hypoxemic. According to the device used, paired analysis showed inconsistent results in 12% to 28% of patients (e-Table 2).

TABLE 1] Results Obtained With the 3 Devices Tested (2 Missing Values for Device 2 for Technical Reasons)

Device	No.	Median	IQR	Range
Device 1[a]				
Mean Spo_2	50	92.0	91.0-95.0	87-97
Median Spo_2	50	93.0	91.0-94.8	87-97
Time spent with Spo_2 < 90%, %	50	3.0	0.0-33.5	0-93
Time spent with Spo_2 < 88%, %	50	0.0	0.0-10.5	0-62
ODI, events/h[b]	50	5.0	1.0-12.0	0-34
Mean $Ptcco_2$, kPa	50	5.5	5.1-5.9	4.1-7.2
Median $Ptcco_2$, kPa	50	5.5	5.1-5.8	4.1-7.2
Mean HR, beats/min	50	73.0	65.3-77.8	47-98
Device 2[c]				
Mean Spo_2	48	91.5	89.0-93.3	83-96
Median Spo_2	48	91.5	89.0-93.3	84-96
Time spent with Spo_2 < 90%, %	48	15.5	2-50.5	0-98
Time spent with Spo_2 < 88%, %	48	3.0	0-30.8	0-83
ODI, events/h[b]	48	6.0	3.0-8.3	0-29
Mean HR, beats/min	48	73.0	66.0-79.5	45-127
Device 3[d]				
Mean Spo_2	50	90.0	88-92	84-97
Median Spo_2	50	90	88-92	84-97
Time spent with Spo_2 < 90%, %	50	17.8	3.4-61.9	0-100
Time spent with Spo_2 < 88%, %	48	5.6	1.1-35.6	0-97.4
ODI, events/h[b]	46	11.0	4.3-17.8	0-53
Mean HR, beats/min	50	70.0	64.3-77.8	46-97

Results obtained with the 3 devices tested (2 missing values for device 2 for technical reasons). For all devices and patients, time spent with Spo_2 < 80% (%) was 0 (IQR, 0-0). HR = heart rate; IQR = interquartile range; ODI = oxygen desaturation index; $Ptcco_2$ = transcutaneous partial pressure of co_2; Spo_2 = pulsed oxygen saturation.
[a]Sentec Digital Monitor with OxiVent Sensor (transcutaneous $Ptcco_2$ and Spo_2 probe).
[b]≥ 3% drop in Spo_2.
[c]ResMed oximetry kit with 8000 S Sensor and Xpod 3012 pulse oximeter.
[d]Viatom Checkme wrist oxygen device.

The conditions of the present study aimed to minimize technical artefacts that could impact Spo_2 values. Factors known to affect accuracy of Spo_2 readings are well known: altitude, hypoperfusion or arrhythmia, motion artifacts, ambient light interference, skin pigmentation (skin color bias), tattoos or skin dyes, hemoglobin variants, nail polish or artificial nails, site of probe, high skin temperature, and hypercapnia.[30-32] In this study, patients received no vasopressor treatment and showed normal systemic pressures, motion artefacts were minimal (sleep), ambient lighting was low or absent, positioning of the devices was performed by an experienced staff, and patients had no nail polish on the probe finger.[11] All patients were of White race, excluding any impact of skin pigmentation. None had any known hemoglobin variants. High values of $Paco_2$ also have been associated with lower performance of pulse oximeters: in this study, average $Ptcco_2$ values were within normal limits (median, 5.5 kPa; IQR, 5.1-5.8 kPa).[33]

Among potential causes of altered Spo_2 signals, 2 factors may have contributed to the discrepancies found between devices. The first is skin temperature: the Sentec $Ptcco_2$ and Spo_2 probe heats the skin at 42° C to enhance transcutaneous diffusion of oxygen and co_2. This increases local perfusion, and thus may increase Spo_2 values. The second is the probe site: regarding arterial blood flow, the earlobe is more central than the fingertips, which may explain a trend for slightly higher values obtained at the earlobe. Using the earlobe as the site for the Sentec probe is the default procedure in adults, although some centers use probes on the forehead or on the thorax. A systematic meta-analysis including 7,021 paired samples from 2,817 patients confirmed that the earlobe was the probe site that

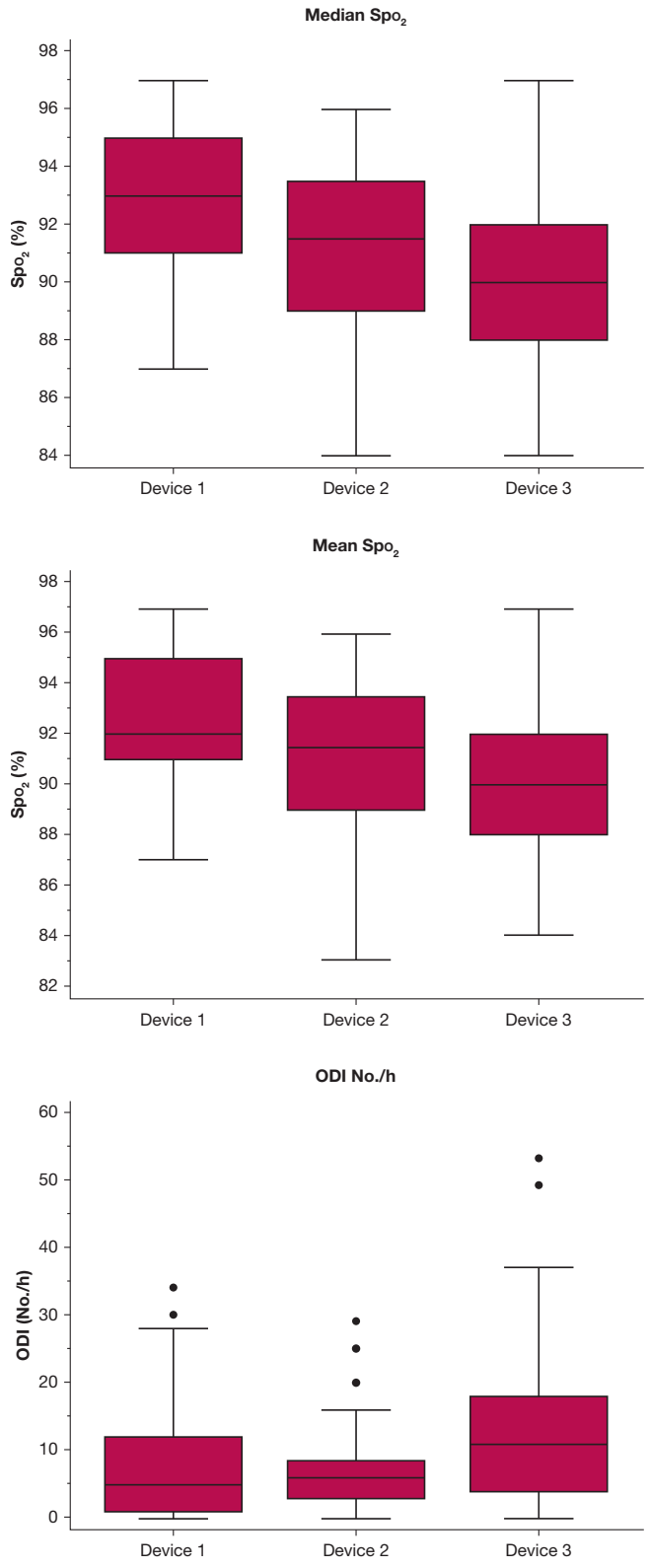

Figure 1 – *A-C, Box plots graphs for median Spo_2 (A), mean Spo_2 (B), and ODI (C) for the 3 devices tested in 50 patients (2 missing values for device 2; see text for explanation). The box shows the interquartile range [IQR (25th-75th percentile)], the horizontal line marks the median, whiskers extend to 1.5 × IQR, and circles indicate points beyond the whiskers. Device 1 is a Sentec Digital Monitor with OxiVent Sensor (transcutaneous partial pressure of Co_2 and Spo_2 probe). Device 2 is a ResMed oximetry kit with 8000 S Sensor and an Xpod 3012 pulse oximeter. Device 3 is an Viatom Checkme wrist oxygen device. ODI = oxygen desaturation index; Spo_2 = pulsed oxygen saturation.*

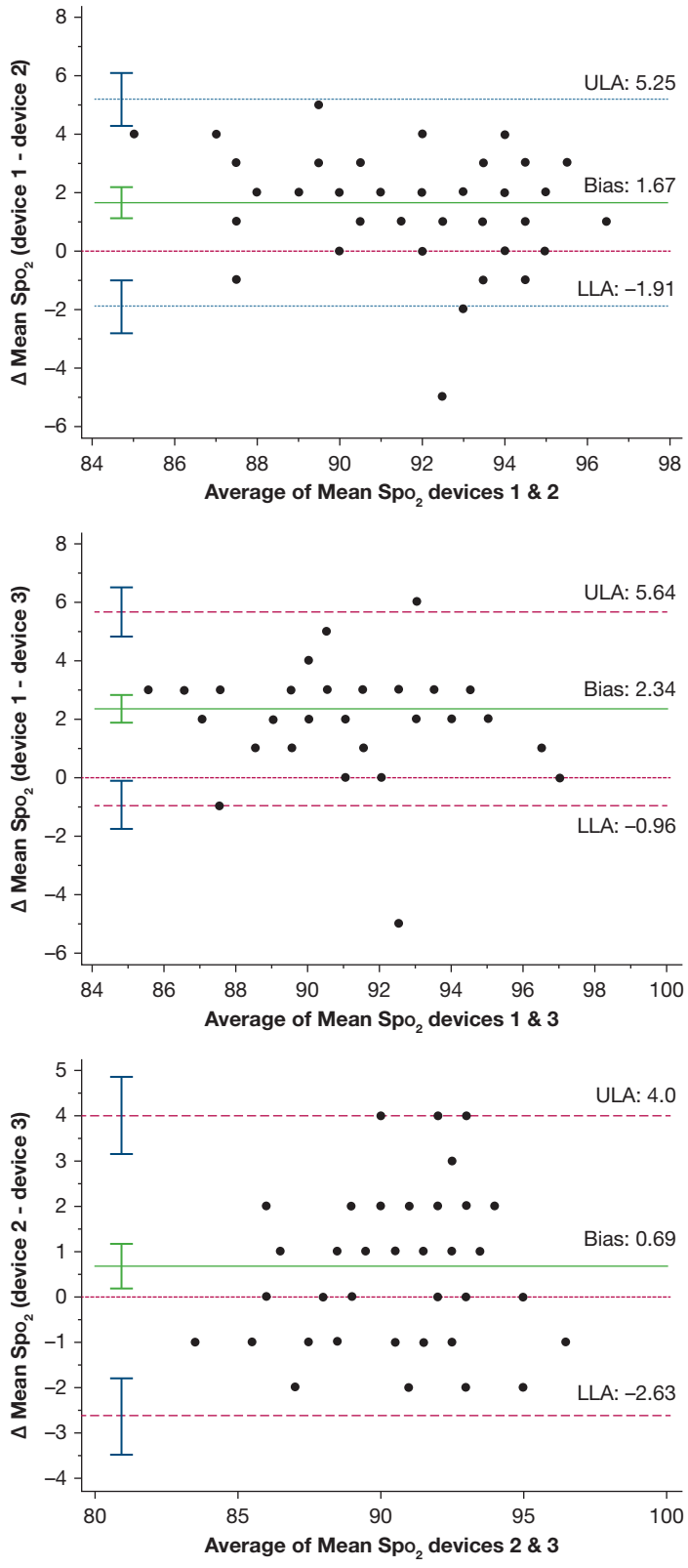

Figure 2 – *A-C, Bland-Altman plots showing paired comparisons of mean Spo_2 values for the 3 devices tested in 50 patients (paired comparisons with device 2 computed with 48 patients): devices 1 and 2 (A), devices 1 and 3 (B), and devices 2 and 3 (C). Each circle on the graph represents a patient. Device 1 is a Sentec Digital Monitor with OxiVent Sensor (transcutaneous partial pressure of co_2 and Spo_2 probe). Device 2 is a ResMed oximetry kit with 8000 S Sensor and Xpod 3012 pulse oximeter. Device 3 is a Viatom Checkme wrist oxygen device. LLA = lower limit of agreement (bias – 1.96 × SD); Spo_2 = pulsed oxygen saturation; ULA = upper limit of agreement (bias + 1.96 × SD).*

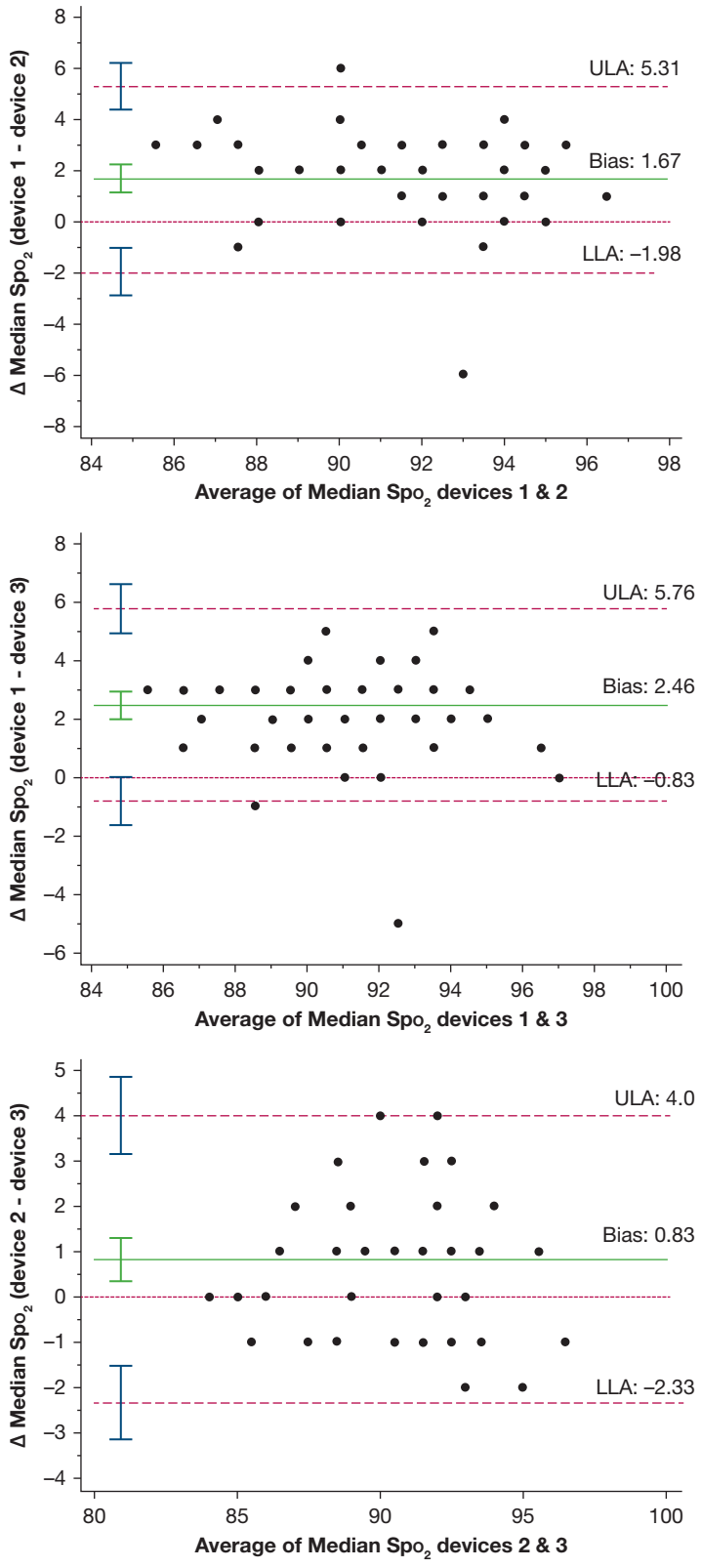

Figure 3 – *A-C, Bland-Altman plots showing median Spo_2 values in 50 patients (paired comparisons with device 2 computed with 48 patients) of the 3 devices tested: devices 1 and 2 (A), devices 1 and 3 (B), and devices 2 and 3 (C). Each circle on the graph represents a patient. LLA = lower limit of agreement (bias – 1.96 × SD); Spo_2 = pulsed oxygen saturation; ULA = upper limit of agreement (bias + 1.96 × SD).*

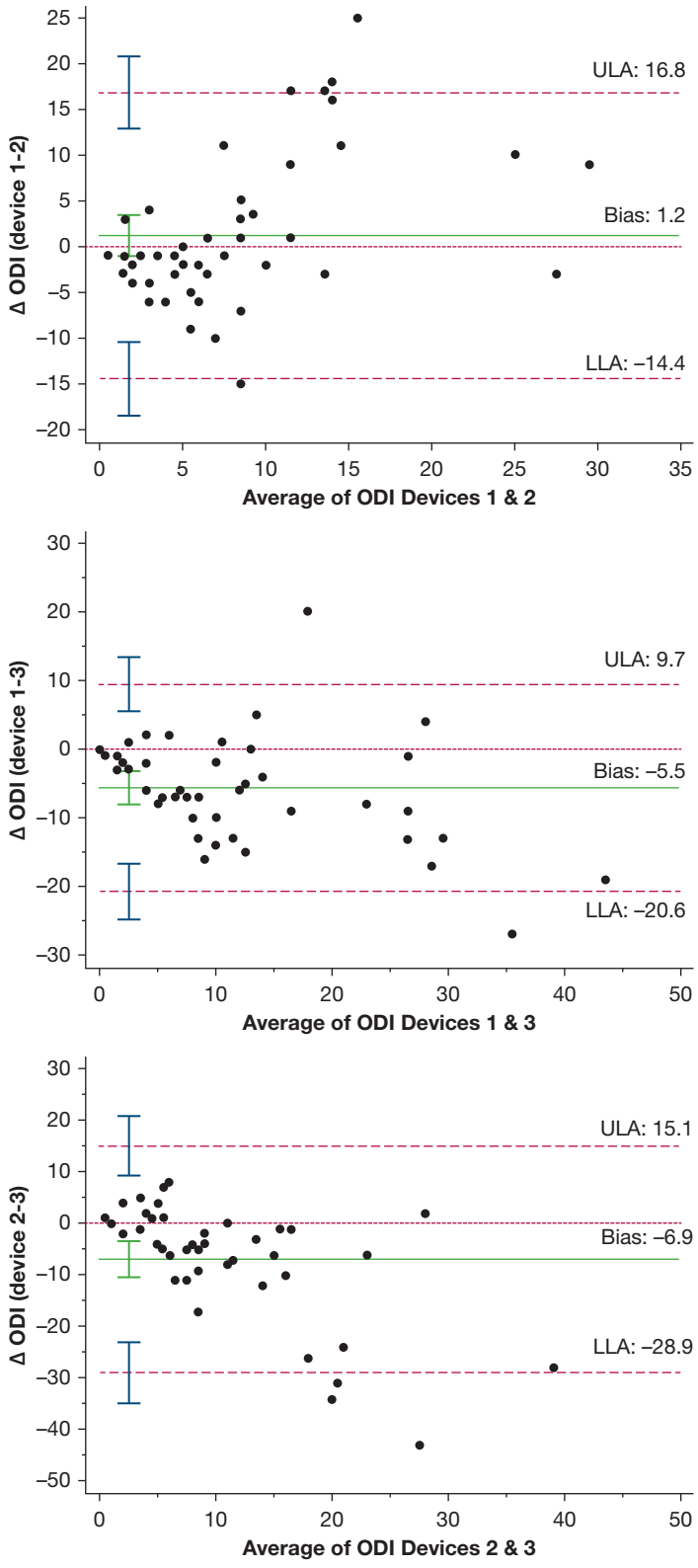

Figure 4 – *A-C, Bland-Altman showing oxygen desaturation index (ODI) values (events/h; ≥ 3% drops) in 50 patients (paired comparisons with device 2 computed with 48 patients) of the 3 devices tested: devices 1 and 2 (A), devices 1 and 3 (B), and devices 2 and 3 (C). Each circle on the graph represents a patient. LLA = lower limit of agreement (bias – 1.96 × SD); Spo_2 = pulsed oxygen saturation; ULA = upper limit of agreement (bias + 1.96 × SD).*

TABLE 2] Bland-Altman Plot Findings

Variable	LLA	Bias	ULA	ICC	95% CI
Mean Spo_2					
Device 1 vs 2	−1.91	1.67	5.25	0.78	0.65-0.87
Device 1 vs 3	−0.96	2.34	5.64	0.81	0.69-0.89
Device 2 vs 3	−2.63	0.69	4.0	0.83	0.71-0.90
Median Spo_2					
Device 1 vs 2	−1.98	1.67	5.31	0.77	0.63-0.87
Device 1 vs 3	−0.83	2.46	5.75	0.81	0.69-0.89
Device 2 vs 3	−2.33	0.83	4.0	0.84	0.74-0.91
ODI					
Device 1 vs 2	−14.43	1.2	16.83	0.47	0.21-0.66
Device 1 vs 3	−20.61	−5.52	9.68	0.75	0.60-0.86
Device 2 vs 3	−28.98	−6.9	15.07	0.33	0.07-0.58

ICCs and 95% CIs estimated using a 2-way mixed-effects model. Device 1 is a Sentec Digital Monitor with OxiVent Sensor (transcutaneous partial pressure of co_2 and Spo_2 probe). Device 2 is a ResMed oximetry kit with 8000 S Sensor and Xpod 3012 pulse oximeter. Device 3 is a Viatom Checkme wrist oxygen device. Tested in 50 patients; comparisons with device 2 computed with 48 patients. ICC = intraclass correlation coefficient; LLA = lower limit of agreement (bias − 1.96 × SD of bias); ODI = oxygen desaturation index; Spo_2 = pulsed oxygen saturation; ULA = upper limit of agreement (bias + 1.96 × SD of bias).

provided the best clinical accuracy when compared with the chest or forearm.[34]

A few other specificities of these devices must be mentioned. First, minimal differences exist in default averaging time for Spo_2 among the devices studied (0.33-1 Hz). Actual sampling in pulse oximeters occurs approximately 25 times/s, but the display is averaged to minimize errors and to stabilize the signal.[2] We did not modify this parameter because: (1) it is seldom if ever done in clinical practice, (2) 2 of the devices had a non-adjustable averaging time and different software was required for the third device, and (3) differences were minimal and it is unlikely that they would be relevant for averaged overnight values. Second, the technology used for estimating Spo_2 is slightly different because both conventional (finger probe) pulse oximeters measure absorption of light beams through the nail bed, whereas the $Ptcco_2$ and Spo_2 earlobe probe and Viatom Checkme ring probe measure light reflected by the soft tissues. To our knowledge, this is not known to affect Spo_2 values. Third, it is possible that calibration procedures vary according to device and manufacturer. European Community Certification requires that pulse oximeters have been tested and certified to be accurate to root mean square error of $< 4\%$ at an Sao_2 of between 70% and 100%. Calibration is performed in healthy individuals exposed to changes in Fio_2 to alter the Sao_2 and compared with Sao_2 values.[35] Finally and importantly, microprocessors and algorithms used to treat the light signals and filter the nonpulsatile vs the pulsatile component of blood flow may differ substantially among devices.[2]

It is surprising that so few studies address the issue of the comparability of Spo_2 overnight recordings using

TABLE 3] κ Values for Agreement Among Devices as to Identifying Patients With Hypoxemia at Different Thresholds and Increased ODI

Variable	Lower Estimate	Cohen's κ	Upper Estimate
No. of patents with $Spo_2 < 90\% > 10\%$ of time			
Device 1 vs 2	0.25	0.48	0.7
Device 2 vs 3	0.28	0.53	0.77
Device 1 vs 3	0.24	0.46	0.68
No. of patients with $Spo_2 < 88\% > 10\%$ of time			
Device 1 vs 2	0.22	0.47	0.72
Device 2 vs 3	0.53	0.73	0.93
Device 1 vs 3	0.38	0.6	0.82
ODI ≥ 10 events/h			
Device 1 vs 2	0.25	0.48	0.7
Device 2 vs 3	0.28	0.53	0.77
Device 1 vs 3	0.24	0.46	0.68

κ values between 0.21 and 0.4 indicate fair agreement, between 0.41 and 0.6 indicate moderate agreement, between 0.61 and 0.8 indicated substantial agreement, and between 0.81 and 1.0 indicate almost perfect agreement. Device 1 is a Sentec Digital Monitor with OxiVent Sensor (transcutaneous $Ptcco_2$ and Spo_2 probe). Device 2 is a ResMed oximetry kit with 8000 S Sensor and Xpod 3012 pulse oximeter. Device 3 is a Viatom Checkme wrist oxygen device. Tested in 50 patients; comparisons with device 2 computed with 48 patients. ODI = oxygen desaturation index; Spo_2 = pulsed oxygen saturation.

different devices in adults. Most studies either have compared point readings with Sa_{O_2} measured by arterial blood gas samples or have compared point values among devices.[14,32,35-37] It is beyond the scope of this article to review all data regarding Sp_{O_2} reliability. Bias (in this case, mean difference of readings between measured Sp_{O_2} and Sa_{O_2}, measured by arterial sampling), and limits of agreement (bias ± 1.96 SD of bias) tend to vary considerably from one device to another. A large difference between upper and lower limits of agreement reflects a high range in potential differences of readings between Sp_{O_2} and Sa_{O_2}: 95% of these differences are expected to lie between the upper and lower limits of agreement. Reliability of the $Ptcc_{O_2}$ and Sp_{O_2} probes (as in device 1) mostly has been studied in comparison with Pac_{O_2} values.[7] However, point values comparing Sp_{O_2} and Sa_{O_2} values have been published and suggest that agreement between $Ptcc_{O_2}$ plus Sp_{O_2} devices and Sa_{O_2} is good.[24,38-41] In summary, overnight recordings of Sp_{O_2} using different pulse oximeters in optimal conditions can result in a clinically significant difference in estimation of oxygen requirements without any obvious explanation.

The second important finding was a major difference in overnight ODIs between devices and thus in the identification of residual SDB receiving CPAP or NIV. Limits of agreement for ODI in paired comparisons of devices were high. Cohen's κ values were moderate (0.46-0.53). Patients were classified differently as having or not having residual SDB in 30% to 43% of cases (e-Table 2C). We used an arbitrary threshold value for ODI of ≥ 10 events/h referring to Georges et al,[42] who showed that an apnea-hypopnea index threshold of 10 events/h was sensitive and specific for identifying patients with residual SDB receiving NIV.

The fact that choice of oximeter affects ODI was reported previously.[43] Zafar et al[43] compared overnight recordings of Sp_{O_2} by 4 different pulse oximeters in candidates for CPAP because of OSA syndrome. Three devices showed a low bias for ODI (0.3 ± 1.7 events/h), but 1 device showed a bias of 2 ± 7 events/h when compared with the 3 other devices. In 35 of 113 patients, using a cutoff of 15 events/h, results affected eligibility for CPAP according to Medicare guidelines. In our study, performed in patients receiving CPAP or NIV, inconsistency between devices was substantial. We did not compare ODI values with data from ventilator software because of the lack of clear definition of respiratory events quantified by ventilator software receiving NIV.

We must point out a few limitations of this study. First, it was not the aim of this study to determine which was the best device for assessing Sp_{O_2} or ODI, and this study does not provide this information. Indeed, the gold standard for Sp_{O_2} is Sa_{O_2}. This would have required invasive procedures to assess arterial blood gases (repeated arterial sampling or a radial catheter). Our aim was to compare frequently used devices in a usual clinical setting and to assess (in)consistency of results and how this affected the indication for supplemental oxygen. The same is true for ODI: a formal evaluation of the relevance of ODI receiving positive pressure (NIV or CPAP) would require either polygraphy or polysomnography because of the complexity of the respiratory events that may occur.[3,44,45] Our aim was pragmatic: the question was, did the 3 devices classify, in a similar manner, patients for whom further testing was necessary?

We acknowledge that, although overnight Sp_{O_2} values are important for monitoring these patients, they are not sufficient per se to choose a clinical option. The clinical decision will depend on algorithms that integrate items such as indication for NIV or CPAP, type of device, information provided by ventilator software (ie, adherence, leaks, apnea-hypopnea index), detection of residual hypoventilation, or CPAP or ventilator settings.[3,5]

Interpretation

This study showed that using different devices for monitoring stable patients receiving NIV or CPAP, all in current use in clinical practice and used by expert centers, may affect both results and clinical decision-making unpredictably. This raises the question of formal independent validation studies of devices marketed for clinical use and of the possibility of periodic (re)calibration of these devices. Slightly higher values obtained with the $Ptcc_{O_2}$ and Sp_{O_2} probe require further studies for confirmation.

Funding/Support

This study received financial support from the Pulmonary League of Geneva (Ligue Pulmonaire Genevoise; www.lpge.ch), a nonprofit organization involved in home care of patients with SDB and chronic respiratory failure.

Financial/Nonfinancial Disclosures

None declared.

Acknowledgments

Author contributions: G. K., B. E., D. L., A. B., and J.-P. J. take responsibility for the integrity of the data, the accuracy of the data analysis; all contributed to planning and design of study and to data interpretation and analysis. G. K., B. E., D. L., and J. -P. J. contributed substantially in data collection. G. K., B. E., D. L., A. B., and J.-P. J. contributed to writing and revision of the article.

Role of sponsors: The Pulmonary League of Geneva was not involved in study design, analysis or preparation of manuscript.

Other contributions: The authors thank Antoine Poncet, the Center for Clinical Research and Division of Clinical Epidemiology, Department of Health and Community Medicine, Geneva University Hospitals, for his support for the statistical and methodologic aspects of this study.

Additional information: The e-Tables are available online under "Supplementary Data."

References

1. Van Meter A, Williams U, Zavala A, et al. Beat to beat: a measured look at the history of pulse oximetry. *J Anesth Hist*. 2017;3(1):24-26.
2. Tekin K, Karadogan M, Gunaydin S, Kismet K. Everything about pulse oximetry-part 1: history, principles, advantages, limitations, inaccuracies, cost analysis, the level of knowledge about pulse oximeter among clinicians, and pulse oximetry versus tissue oximetry. *J Intensive Care Med*. 2023;38(9):775-784.
3. Arnal JM, Oranger M, Gonzalez-Bermejo J. Monitoring systems in home ventilation. *J Clin Med*. 2023;12(6):2163.
4. Janssens JP, Borel JC, Pepin JL, Somno NIVG. Nocturnal monitoring of home non-invasive ventilation: the contribution of simple tools such as pulse oximetry, capnography, built-in ventilator software and autonomic markers of sleep fragmentation. *Thorax*. 2011;66(5):438-445.
5. Janssens JP, Cantero C, Pasquina P, Georges M, Rabec C. Monitoring long term noninvasive ventilation: benefits, caveats and perspectives. *Front Med (Lausanne)*. 2022;9:874523.
6. Eberhard P, Gisiger PA, Gardaz JP, Spahn DR. Combining transcutaneous blood gas measurement and pulse oximetry. *Anesth Analg*. 2002;94(1 suppl): S76-S80.
7. Aarrestad S, Tollefsen E, Kleiven AL, Qvarfort M, Janssens JP, Skjonsberg OH. Validity of transcutaneous PCO2 in monitoring chronic hypoventilation treated with non-invasive ventilation. *Respir Med*. 2016;112:112-118.
8. Dullenkopf A, Bernardo SD, Berger F, Fasnacht M, Gerber AC, Weiss M. Evaluation of a new combined SpO2/PtcCO2 sensor in anaesthetized paediatric patients. *Paediatr Anaesth*. 2003;13(9):777-784.
9. Duiverman ML. Tricks and tips for home mechanical ventilation. Home mechanical ventilation: set-up and monitoring protocols. *Pulmonology*. 2021;27(2):144-150.
10. Milner QJ, Mathews GR. An assessment of the accuracy of pulse oximeters. *Anaesthesia*. 2012;67(4):396-401.
11. Chan ED, Chan MM, Chan MM. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respir Med*. 2013;107(6): 789-799.
12. Pretto JJ, Roebuck T, Beckert L, Hamilton G. Clinical use of pulse oximetry: official guidelines from the Thoracic Society of Australia and New Zealand. *Respirology*. 2014;19(1):38-46.
13. Jensen LA, Onyskiw JE, Prasad NG. Meta-analysis of arterial oxygen saturation monitoring by pulse oximetry in adults. *Heart Lung*. 1998;27 (6):387-408.
14. Severinghaus JW, Naifeh KH. Accuracy of response of six pulse oximeters to profound hypoxia. *Anesthesiology*. 1987;67(4):551-558.
15. Trivedi NS, Ghouri AF, Lai E, Shah NK, Barker SJ. Pulse oximeter performance during desaturation and resaturation: a comparison of seven models. *J Clin Anesth*. 1997;9(3):184-188.
16. Jubran A. Pulse oximetry. *Crit Care*. 2015;19(1):272.
17. McKim DA, Road J, Avendano M, et al. Home mechanical ventilation: a Canadian Thoracic Society clinical practice guideline. *Can Respir J*. 2011;18(4): 197-215.
18. Windisch W, Geiseler J, Simon K, Walterspacher S, Dreher M; on behalf of the Guideline Commission. German national guideline for treating chronic respiratory failure with invasive and non-invasive ventilation: revised edition 2017—part 1. *Respiration*. 2018;96(1): 66-97.
19. Windisch W, Geiseler J, Simon K, Walterspacher S, Dreher M; on behalf of the Guideline Commission. German national guideline for treating chronic respiratory failure with invasive and non-invasive ventilation: revised edition 2017—part 2. *Respiration*. 2018;96(2): 171-203.
20. Janssens JP, Michel F, Schwarz EI, et al. Long-term mechanical ventilation: recommendations of the swiss society of pulmonology. *Respiration*. 2020:1-36.
21. Patout M, Lhuillier E, Kaltsakas G, et al. Long-term survival following initiation of home non-invasive ventilation: a European study. *Thorax*. 2020;75(11): 965-973.
22. Cantero C, Adler D, Pasquina P, et al. Long term noninvasive ventilation in the Geneva Lake area: indications, prevalence and modalities. *Chest*. 2020;158(1): 279-291.
23. Patil SP, Collop NA, Chediak AD, Olson EJ, Vohra KP, Panel OTE. Optimal NIV Medicare access promotion: patients with OSA: a technical expert panel report from the American College of Chest Physicians, the American Association for Respiratory Care, the American Academy of Sleep Medicine, and the American Thoracic Society. *Chest*. 2021;160(5): e409-e417.
24. Hazenberg A, Zijlstra JG, Kerstjens HA, Wijkstra PJ. Validation of a transcutaneous CO(2) monitor in adult patients with chronic respiratory failure. *Respiration*. 2011;81(3):242-246.
25. Janssens JP, Perrin E, Bennani I, de Muralt B, Titelion V, Picaud C. Is continuous transcutaneous monitoring of PCO2 (TcPCO2) over 8 h reliable in adults? *Respir Med*. 2001;95(5):331-335.
26. Lellouche F, Branson R. U.S. Food and Drug Administration strategy to evaluate pulse oximeters: the same cause will produce the same effects. *Am J Respir Crit Care Med*. 2024;209(11):1301-1303.
27. Berry RB, Budhiraja R, Gottlieb DJ, et al. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2012;8(5):597-619.
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.
29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2): 420-428.
30. Leon-Valladares D, Barrio-Mateu LA, Cortes-Carmona N, et al. Determining factors of pulse oximetry accuracy: a literature review. *Rev Clin Esp (Barc)*. 2024;224(5):314-330.
31. Giuliano KK, Bilkovski RN, Beard J, Lamminmaki S. Comparative analysis of signal accuracy of three SpO(2) monitors during motion and low perfusion conditions. *J Clin Monit Comput*. 2023;37 (6):1451-1461.
32. Weber U, Tomschik E, Resch I, et al. Comparison of two new generation pulse oximeters during emergency ambulance transportation. *Eur J Anaesthesiol*. 2011;28(2):92-96.
33. Munoz X, Torres F, Sampol G, Rios J, Marti S, Escrich E. Accuracy and reliability of pulse oximetry at different arterial carbon dioxide pressure levels. *Eur Respir J*. 2008;32(4): 1053-1059.
34. Conway A, Tipton E, Liu WH, et al. Accuracy and precision of transcutaneous carbon dioxide monitoring: a systematic review and meta-analysis. *Thorax*. 2019;74(2):157-163.
35. 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(2):338-345.

36. Stell D, Noble JJ, Kay RH, et al. Exploring the impact of pulse oximeter selection within the COVID-19 home-use pulse oximetry pathways. *BMJ Open Respir Res*. 2022;9(1):e001159.

37. 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(3):154-158.

38. Kocher S, Rohling R, Tschupp A. Performance of a digital PCO2/SPO2 ear sensor. *J Clin Monit Comput*. 2004;18(2):75-79.

39. Ekkernkamp E, Welte L, Schmoor C, et al. Spot check analysis of gas exchange: invasive versus noninvasive methods. *Respiration*. 2015;89(4):294-303.

40. Chhajed PN, Miedinger D, Baty F, et al. Comparison of combined oximetry and cutaneous capnography using a digital sensor with arterial blood gas analysis. *Scand J Clin Lab Invest*. 2010;70(1):60-64.

41. Torok SS, Leuppi JD, Baty F, Tamm M, Chhajed PN. Combined oximetry-cutaneous capnography in patients assessed for long-term oxygen therapy. *Chest*. 2008;133(6):1421-1425.

42. Georges M, Adler D, Contal O, et al. Reliability of apnea-hypopnea index measured by a home bi-level pressure support ventilator versus a polysomnographic assessment. *Respir Care*. 2015;60(7):1051-1056.

43. Zafar S, Ayappa I, Norman RG, Krieger AC, Walsleben JA, Rapoport DM. Choice of oximeter affects apnea-hypopnea index. *Chest*. 2005;127(1):80-88.

44. Gonzalez-Bermejo J, Perrin C, Janssens JP, et al. Proposal for a systematic analysis of polygraphy or polysomnography for identifying and scoring abnormal events occurring during non-invasive ventilation. *Thorax*. 2012;67(6):546-552.

45. Gonzalez-Bermejo J, Janssens JP, Rabec C, et al. Framework for patient-ventilator asynchrony during long-term non-invasive ventilation. *Thorax*. 2019;74:715-717.