

Original Paper

# Performance of Wearable Pulse Oximetry During Controlled Hypoxia Induction: Instrument Validation Study

Yihang Jiang<sup>1</sup>, MS; Connor Spies<sup>2</sup>, BA; Ali R Roghanizad<sup>1</sup>, PhD; Will Ke Wang<sup>1</sup>, PhD; Satasuk Joy Bhosai<sup>2,3</sup>, MPH, MD; Laurie Snyder<sup>2,3</sup>, MD, MHS; Ashley Burke<sup>4</sup>, BA; David MacLeod<sup>4</sup>, MBBS; Jessilyn Dunn<sup>1,5</sup>, PhD

<sup>1</sup>Department of Biomedical Engineering, Duke University, Durham, NC, United States

<sup>2</sup>Department of Medicine, Duke University, Durham, NC, United States

<sup>3</sup>Department of Medicine, Duke Clinical Research Institute, Duke University, Durham, NC, United States

<sup>4</sup>Human Pharmacology and Physiology Lab, Duke University, Durham, NC, United States

<sup>5</sup>Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

## Corresponding Author:

Jessilyn Dunn, PhD  
Department of Biomedical Engineering  
Duke University  
534 Research Dr, # 448  
Durham, NC 27710  
United States  
Phone: 1 9542960640  
Email: [jessilyn.dunn@duke.edu](mailto:jessilyn.dunn@duke.edu)

## Abstract

**Background:** Oxygen saturation is a crucial metric used for monitoring patients with lung disease or respiratory illness who are at risk of hypoxemia (low blood oxygen saturation). Early and accurate identification of abnormal oxygen saturation is important for these patients, who may develop significant desaturation and hypoxemia symptoms during their daily activities.

**Objective:** This study aimed to evaluate the accuracy of Apple Watch Series 7 and a clinical-grade pulse oximeter, Masimo MightySat Rx, under hypoxemia and to assess whether measurement error is influenced by the oxygen desaturation rate (ODR).

**Methods:** We calculated the ODR of each measurement and conducted a comparative analysis of the displayed oxygen saturation readings from both the Masimo MightySat Rx finger pulse oximeter and Apple Watch Series 7 with arterial blood oxygen saturation (SaO<sub>2</sub>) readings obtained from a blood gas analyzer.

**Results:** Both the Masimo MightySat Rx pulse oximeter and the Apple Watch Series 7 tended to overestimate oxygen saturation. The pulse oximeter readings were more likely to fall within 2% of the acceptable (as specified by Masimo) peripheral oxygen saturation (SpO<sub>2</sub>) error range than the Apple Watch (49.03% vs 32.14%). Notably, both devices had limitations under low oxygen saturation levels (<88%), with an accuracy root mean square difference (A<sub>rms</sub>) of 3.52% (95% CI 3.18%-3.86%) and 5.82% (95% CI 5.32%-6.31%) for the Masimo MightySat Rx and Apple Watch Series 7, respectively. Among the blood oxygen measurements taken during a high ODR (ie, ≥2% SpO<sub>2</sub> per minute), which is a rate clinically correlated with sleep apnea, the A<sub>rms</sub> increased slightly by 0.75% for the Masimo MightySat Rx and decreased by 0.28% for the Apple Watch Series 7.

**Conclusions:** Both devices consistently overestimated SpO<sub>2</sub>, with accuracy declining notably during hypoxemia. The Apple Watch Series 7 mean bias suggests a likelihood of missing instances of hypoxemia, particularly at arterial oxygen saturation values below but close to 88%. Both the Apple Watch Series 7 and Masimo MightySat Rx exhibited A<sub>rms</sub> values exceeding the US Food and Drug Administration threshold under conditions of hypoxemia. While past studies have implicated high ODRs in increasing measurement error, we found no statistically significant relationship between ODR and measurement error for either device. Overall, our findings of SpO<sub>2</sub> overestimation and high A<sub>rms</sub> values underscore the need for caution when interpreting oxygen saturation values from these devices. The small sample size and limited diversity in skin tone and age restrict the generalizability of our findings. Future studies should include larger and more diverse populations to evaluate the performance of wearable-based pulse oximetry.

**Keywords:** pulse oximetry; oxygen saturation; smartwatches; hypoxemia; wearable electronic devices

## Introduction

Blood oxygen saturation, representing how much hemoglobin is bound with oxygen, is considered to be a useful vital sign for gaining insight into a person's health status alongside body temperature, blood pressure, heart rate, and respiratory rate [1]. Chronic lung diseases such as chronic obstructive pulmonary disease (COPD) and asthma can lead to significant desaturation and hypoxemia [2,3], and exercise-induced desaturation is common in these populations and associated with increased mortality [4,5]. Early and accurate identification of abnormal oxygen saturation [6] is important in patients who may develop hypoxemia symptoms and significant oxygen desaturation during exercise, rest, or sleep [7-9].

Measurement of oxygen saturation can be classified into 2 methods: invasive and noninvasive [10]. Arterial blood gas (ABG) analysis is an invasive measurement method that requires the collection of arterial blood samples and measures arterial oxygen saturation (SaO<sub>2</sub>) by a dedicated co-oximeter machine. While ABG analysis is considered the gold standard, arterial blood sample collection is painful and impractical for continuous monitoring [11,12]. Pulse oximetry is based on the principle that the optical absorption of light across specific wavelengths differs for oxyhemoglobin and deoxyhemoglobin, and this phenomenon allows for the noninvasive estimation of blood oxygen saturation (SpO<sub>2</sub>) [13]. SpO<sub>2</sub> measurement functionality was first introduced in popular consumer wearables in 2021, and roughly a decade later, in 2032, the global market value of wearable technology is projected to reach US \$191.58 billion [14]. Multiple studies comparing SpO<sub>2</sub> measurements from the Apple Watch Series 7 and finger-placed pulse oximeters found Pearson correlation coefficients ranging from 0.76 to 0.89 [15-18]. However, these studies did not evaluate the smartwatch's capability to detect and identify hypoxia specifically, which is when accuracy is most critical because incorrect measurements can put patients at risk of delayed recognition and treatment [19-21]. Occult hypoxemia, which occurs when a person's blood oxygen level appears normal when measured by pulse oximetry (SpO<sub>2</sub>>92%), but is actually low when measured by a more accurate test like ABG (SaO<sub>2</sub><88%), may be a clinically significant consequence of such unreliability [22]. In such a situation, hypoxemia is present but hidden or not detected by the pulse oximeter. Occult hypoxemia is concerning because it can delay the recognition and treatment of low oxygen levels, especially in critically ill patients or those with respiratory diseases such as COVID-19.

Additionally, while SpO<sub>2</sub> is often straightforward to measure when its level has reached a steady state, challenges arise in trying to accurately measure SpO<sub>2</sub> under dynamic change, particularly when those changes are rapid, such as during sleep apnea [23,24]. Blood oxygen levels typically decrease in people with obstructive sleep apnea (OSA) as

a result of gaps in their breathing [25]. The oxygen desaturation rate (ODR), defined as the change in SpO<sub>2</sub> per minute, is a metric used to quantify such changes in SpO<sub>2</sub> over time. During rapid desaturation, pulse oximeter accuracy has been shown to decrease due to averaging-window algorithms that smooth the signal over several seconds and include values before the desaturation event [26,27], causing the reported SpO<sub>2</sub> value to lag behind the patient's real-time oxygenation status and resulting in incorrect values. Because pulse oximeter accuracy may be mediated by the ODR, it is important to evaluate pulse oximeter performance under conditions of rapid oxygen desaturation. While such studies have been performed in the past, they typically do not use the gold standard SaO<sub>2</sub> method to measure blood oxygen saturation due to the burdensome nature of sample collection [28-30]. In this study, we evaluated whether the accuracies of SpO<sub>2</sub> measurements from a common pulse oximeter and a smartwatch were impacted by the ODR using gold standard SaO<sub>2</sub> measurements.

## Methods

### *Ethical Considerations*

This study was approved by the Duke University Health System Institutional Review Board (Pro00110458) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from participants prior to their participation in the study. All data were deidentified before analysis, and no personally identifiable information was collected or stored. No compensation was provided to participants. The manuscript and supplementary materials contain no images or information that could identify individual participants.

### *Study Population*

This study was conducted as an opportunistic investigation embedded within an existing controlled hypoxia study, which involved invasive arterial catheterization and prolonged exposure to reduced oxygen levels. As a result, the number of participants was constrained by the requirements of the parent study. Inclusion criteria included healthy male or female subjects between the ages of 18 and 45 years with a minimum body weight of 40 kg and a body mass index between 18 and 35 kg/m<sup>2</sup>. Individuals were excluded if they had peripheral vascular disease, Raynaud syndrome, cryoglobulinemia, or any collagen vascular disease affecting the fingers; a history of blood clots within the past 6 months; a history of sickle cell trait or thalassemia; a positive urine cotinine test; heparin allergy; essential tremor; gel nail polish or any other non-natural, nonremovable discoloration of the forefinger; and, based on venous blood sampling, abnormal hemoglobin levels, abnormal hemoglobin electrophoresis, carboxyhemoglobin levels greater than 3 g/dL, or methemoglobin levels greater than 2 g/dL.

## Equipment

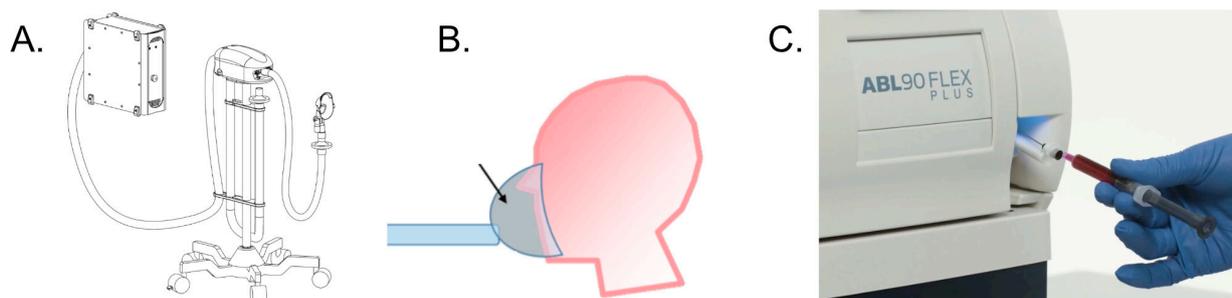
Equipment was used in this study to assess skin tone, control the desaturation process, and take the SpO<sub>2</sub> and SaO<sub>2</sub> measurements. To assess skin tone, the Delfin SkinColor-Catch colorimeter was used to report the individual typology angle (ITA), which classifies skin color types into 6 groups from very light to dark skin: very light (>55°), light (41° to <55°), intermediate (28° to <41°), tan (10° to <28°), brown (−30° to <10°), and dark (<−30°). In the desaturation process, we used the sequential gas delivery system RespirAct RA-MR to control the alveolar oxygen tension (PAO<sub>2</sub>) and partial pressure of end-tidal oxygen (PetO<sub>2</sub>) by changing inhaled oxygen and carbon dioxide levels [31]. The attained PAO<sub>2</sub> determined the partial pressure of oxygen in arterial blood (PaO<sub>2</sub>) at the alveolar-arterial interspace in the lung. The PaO<sub>2</sub> in turn determined the participants' SaO<sub>2</sub> prior to blood gas sampling. The Radiometer ABL90 Flex blood gas analyzer was used to analyze extracted ABG samples and obtain the functional SaO<sub>2</sub>. The Apple

Watch Series 7 smartwatch (reflectance pulse oximeter) and Masimo MightySat Rx finger pulse oximeter (transmissive pulse oximeter) were each used to measure SpO<sub>2</sub> separately. The Apple Watch Series 7 was chosen as it was previously identified to have the best performance among a selection of 4 commercial wearables in an analytic validation study [15,32,33], and the Masimo MightySat Rx was chosen because it is a US Food and Drug Administration (FDA) 510(k)-cleared pulse oximeter [34,35].

## Desaturation Protocol

In this study, we used a controlled gas delivery system to conduct the desaturation protocol and altered SaO<sub>2</sub> by adjusting end-tidal oxygen (PetO<sub>2</sub>) levels that participants inhaled (Figure 1). We then compared oxygen saturation (SpO<sub>2</sub>) measurements from Apple Watch Series 7 and Masimo MightySat Rx against the gold standard ABG measurements.

**Figure 1.** Workflow of the equipment used to adjust PetO<sub>2</sub> and measure SaO<sub>2</sub>. Adjusting the end-tidal PetO<sub>2</sub> setting in the RespirAct RA-MR gas delivery system (A) led the study participant to inhale gas with different oxygen, carbon dioxide, and nitrogen concentrations through a fitted mask (B). Blood samples taken once PetO<sub>2</sub> stabilized were processed by a blood gas analyzer (Radiometer ABL90 Flex) (C) to measure the resulting SaO<sub>2</sub> and PaO<sub>2</sub>. PaO<sub>2</sub>: partial pressure of oxygen in arterial blood; PetO<sub>2</sub>: partial pressure of end-tidal oxygen; SaO<sub>2</sub>: arterial oxygen saturation.



Before the desaturation procedure, participants lay in the supine position on a standard hospital stretcher with an arterial catheter placed in the forearm for blood sampling. The facemask that was connected to the sequential gas delivery system was donned by the participant. The Apple Watch Series 7 was placed comfortably but tightly on the right wrist, approximately 1 cm above the ulnar styloid process, to avoid movement during the experiment. The Masimo MightySat Rx was placed on the middle finger of the participant's right hand.

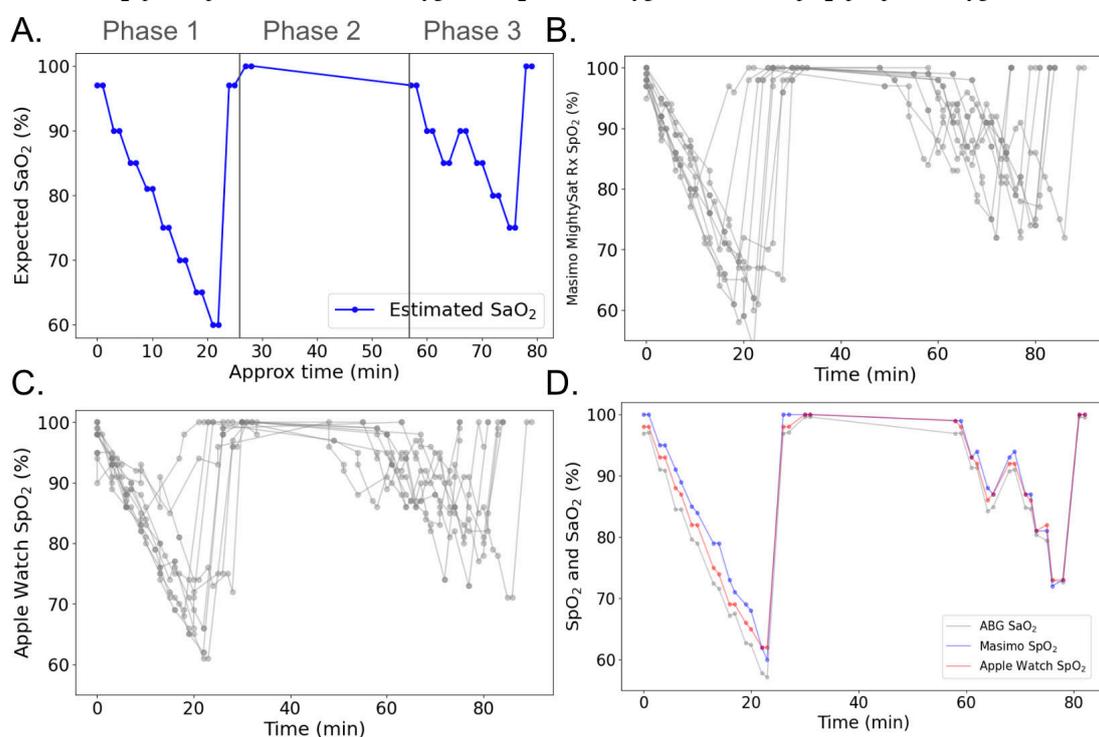
This study involved 3 phases: a sequential stepwise oxygen desaturation phase with monotonically decreasing oxygen saturation (Figure 2A, Phase 1), an oxygen saturation recovery phase (Figure 2A, Phase 2), and an interrupted oxygen desaturation phase, which included a brief increase in blood oxygen saturation during the overall desaturation sequence, enabling evaluation of pulse oximeters during directional changes in blood oxygen saturation (eg, desaturation to resaturation or resaturation to desaturation) (Figure 2A, Phase 3). Besides the oxygen saturation recovery phase, Phase 1 was designed to follow a standard step-down sequence [34], as recommended by the FDA, for assessing the performance of pulse oximeters during desaturation and hypoxemia, whereas Phase 3 was designed to evaluate device

performance when the desaturation was interrupted under hypoxemia. Blood oxygen levels (SaO<sub>2</sub>) were altered by controlling inhaled oxygen and carbon dioxide levels using the RespirAct gas delivery system, which correspondingly altered PaO<sub>2</sub> and PetO<sub>2</sub>, which in turn altered SaO<sub>2</sub> [36]. In the first sequence, the oxygen saturation (SaO<sub>2</sub>) was reduced in stepwise fashion from approximately 100% to approximately 60%, and then back to approximately 100% over the course of approximately 24 minutes. The second sequence involved holding participants' SaO<sub>2</sub> at approximately 100% for approximately 30 minutes. The third sequence involved decreasing participants' SaO<sub>2</sub> from approximately 100% to approximately 80%, then back to approximately 100% over the course of 21 minutes. The first oxygen desaturation sequence (Figure 2A, Phase 1) involved 8 steps of approximately 3 minutes each, with PetO<sub>2</sub> set to 90, 60, 45, 40, 37, 34, 32, and 250 mm Hg. Using the Severinghaus formula [37], SaO<sub>2</sub> levels at these stages were estimated to be 97%, 90%, 85%, 81%, 75%, 70%, 65%, 60%, and 100%, respectively. Following this sequence, PetO<sub>2</sub> was held at 250 mm Hg for approximately 30 minutes to allow participants to recover their maximum oxygen saturation (Figure 2A, Phase 2). The final oxygen desaturation sequence (Figure 2A, Phase 3) consisted of 7 steps of approximately 3 minutes

each, with  $PetO_2$  set to 90, 60, 50, 60, 50, 45, 40, and 250 mm Hg, corresponding to an estimated  $SaO_2$  of 97%, 90%, 85%, 90%, 85%, 80%, 75%, and 100%. During each step of the sequence,  $PetO_2$  was held at a steady state using the RespirAct RA-MR until 2 ABG samples were obtained: one collected once  $PetO_2$  had stabilized, and another collected 1 minute later, right before transitioning to the next target  $PetO_2$  level. Consecutive plateaus were separated by approximately 2 minutes to allow the gas delivery system to reach the next target  $PetO_2$  level. The only exception was

the recovery period between Phase 2 and Phase 3, during which participants were given a 30-minute rest interval. A participant who completed the full desaturation protocol was expected to have 36 ABG samples corresponding to 2 samples at each of the 18 steady-state plateaus, which themselves comprised 8 plateaus during Phase 1 (initial desaturation), 2 plateaus during Phase 2 (recovery at high  $PetO_2$ ), and 8 plateaus during Phase 3 (interrupted desaturation).

**Figure 2.** (A) Severinghaus-estimated  $SaO_2$  values. The solid blue dots represent the estimated  $SaO_2$  based on the  $PetO_2$  settings at each step of the protocol, and the solid lines show the overall trajectories. The 3 phases of the study protocol (1: oxygen desaturation; 2: oxygen saturation recovery; 3: interrupted oxygen desaturation) are separated using gray vertical lines. (B and C) Actual  $SpO_2$  measurements from the Masimo MightySat Rx (B) and the Apple Watch Series 7 (C) throughout the study protocol (N=9 study participants). (D) Actual  $SpO_2$  and  $SaO_2$  measurements from a representative participant resulting from the changing  $PetO_2$  in the inhaled gas mixture over the duration of the study. ABG: arterial blood gas; approx: approximate;  $PetO_2$ : partial pressure of end-tidal oxygen;  $SaO_2$ : arterial oxygen saturation;  $SpO_2$ : peripheral oxygen saturation.



Two  $SpO_2$  measurements, one each from the Apple Watch Series 7 and the Masimo MightySat Rx, were obtained simultaneously with each ABG sample at each steady-state plateau. Measurements on the smartwatch were manually triggered, and the participants were instructed to keep their arm still during the 15-second countdown while awaiting the  $SpO_2$  reading. The finger pulse oximeter reading was recorded once the smartwatch produced either a successful or unsuccessful measurement. If Masimo MightySat Rx or Apple Watch Series 7 failed to produce a measurement, we recorded a missing (invalid) measurement and made another attempt after readjusting the device. If both devices successfully produced readings, the smartwatch was manually triggered again to initiate the second measurement.

### Metrics and Statistical Analysis

To assess the performance of the MightySat Rx and Apple Watch Series 7, we compared each  $SpO_2$  measurement to the  $SaO_2$  measurements, determining whether values fell within,

above, or below the acceptable error range (2%) as defined by the Masimo MightySat Rx manual. We also recorded whether the  $SpO_2$  measurements were missing, indicating that the pulse oximetry device was unable to generate a measurement. The Bland-Altman method was used to assess the accuracy of  $SpO_2$  measurements across all readings, with separate analysis for readings where the  $SaO_2$  value was below 88%. The threshold of 88% was chosen in line with the British Thoracic Society guidelines [31] indicating the decision point for implementing intensive therapy to elevate oxygen saturation. The x-axes of the Bland-Altman plot are the mean of the  $SaO_2$  and  $SpO_2$  values and the y-axes are the difference between  $SaO_2$  and  $SpO_2$ . Metrics used to gauge each device's performance included mean directional error (MDE) (Equation 1), missingness (Equation 2), and average root mean square ( $A_{rms}$ ) difference (Equation 3), where  $p$  is the number of participants and  $v_i$  is the number of valid measurements for the  $i$ th participant. To evaluate device performance during oxygen desaturation, we excluded

measurements with SpO<sub>2</sub> or SaO<sub>2</sub> more than 99% (n=57), as well as measurements collected during the resaturation process (n=79), which was defined by an increase in SaO<sub>2</sub> relative to the preceding measurement (ie, between Phase 1 and Phase 2 of the protocol). For the remaining measurements (n=172) with ODR at least 2% per minute or less than 2% per minute. A 2-sided paired *t* test was used to examine the hypothesis that MDE differs between the 2 ODR scenarios. The choice of 2% per minute stems from the 4% oxygen desaturation index (ODI-4%), which is defined as the number of events per hour with a more than 4% decrease in SpO<sub>2</sub> within 2 minutes (ie, the equivalent of an average of a 2% change per minute) [24]. ODI-4% is a clinically important metric to help diagnose OSA [38]. However, because the sampling frequency of the SaO<sub>2</sub> measurements is less than 1 measurement every 2 minutes, this threshold was translated into an ODR of 2% per minute. Before statistical testing, the differences between each pair of SpO<sub>2</sub> and SaO<sub>2</sub> point measurements (see Table S1 in [Multimedia Appendix 1](#)) were mean-aggregated for each person at each of the 2 ODRs, resulting in 1 MDE per person per ODR scenario. This was done separately for the Apple Watch Series 7 and Masimo MightySat Rx SpO<sub>2</sub> measurements. These calculations resulted in 9 participant-level MDE values for each ODR scenario and for each device (see Table S1, column 7 in [Multimedia Appendix 1](#); Figure S1 for details of MDE values generation in [Multimedia Appendix 2](#); Figure S2 for Boxplots of MDE values in [Multimedia Appendix 3](#)). Two paired *t* tests were performed separately, 1 for each device,

on the 9 MDE values when ODR was at least 2% per minute and the 9 MDE values when ODR was less than 2% per minute to determine whether MDE values were statistically significantly different when ODR was above or below 2% per minute (see Figure S3 for normality check in [Multimedia Appendix 4](#)).

$$\text{MDE (Mean Directional Error)} = \frac{\sum_{i=1}^p \sum_{j=1}^{v_i} (\text{SpO}_{2i,j} - \text{SaO}_{2i,j})}{\sum_{i=1}^p v_i} \quad (1)$$

$$\text{Missingness (\%)} = 100 - \left( \frac{\sum_{i=1}^p v_i}{\text{Total measurements}} \right) \times 100 \quad (2)$$

$$A_{\text{rms}} (\text{Root Mean Square Difference}) = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^{v_i} (\text{SpO}_{2i,j} - \text{SaO}_{2i,j})^2}{\sum_{i=1}^p v_i}} \quad (3)$$

## Results

### Study Population

Nine individuals (5 males and 4 females) provided written informed consent under the Duke University Health System Institutional Review Board protocol (Pro00110458) and completed the study as described in the Methods. The median age was 25 (IQR 21-26) years, and the median body mass index was 24.4 (IQR 23.7-26.6) kg/m<sup>2</sup>. The median (IQR) value of ITA of all participants was 7° (-1° to 23°) ([Table 1](#)).

**Table 1.** Participant demographics and characteristics.

Participant number	Age	ITA <sup>a</sup> value	Reference ITA skin classification [39,40]	Gender
1	19	23	Tan	Male
2	25	54	Light	Female
3	27	-27	Brown	Male
4	26	-8	Brown	Male
5	24	12	Tan	Female
6	19	39	Intermediate	Female
7	21	2	Brown	Female
8	28	-1	Brown	Male
9	26	7	Brown	Male

<sup>a</sup>ITA: individual typology angle.

### Comparison Between ABG and Pulse Oximetry Measurements

Measurements were taken across 9 study participants at 18 set PetO<sub>2</sub> levels ranging from 32 to 250 mm Hg, which translates to an approximate SaO<sub>2</sub> values from 60% to 100% over the course of 3 phases of the study that involved monotonic oxygen desaturation, recovery, and interrupted oxygen desaturation ([Figure 2A](#); a representative participant's trajectory is shown in [Figure 2D](#)). There were 308 measurements of the "ground truth" ABG-based blood oxygen saturation (SaO<sub>2</sub>) which were compared with the SpO<sub>2</sub>

measurements generated by the Masimo MightySat Rx ([Figure 2B](#)) and the Apple Watch Series 7 ([Figure 2C](#)). All participants had maximum ABG-based SaO<sub>2</sub> measurements of 99% and maximum Masimo- and Apple watch-based SpO<sub>2</sub> measurements of 100% ([Table 2](#)). The minimum SaO<sub>2</sub> was less than 65% for 6/9 (67%) participants, and minimum SpO<sub>2</sub> was less than 65% for 7/9 (78%) participants using the Masimo MightySat Rx and for 2/9 (22%) participants using the Apple Watch Series 7. In general, the Apple Watch Series 7 measurements did not reach the same nadir values as the Masimo MightySat Rx measurements ([Figure 2B and C](#)).

**Table 2.** Maximum and minimum SpO<sub>2</sub> and SaO<sub>2</sub> values and interindividual ranges during the desaturation procedure.

Type of measurements	Maximum, % (range)	Minimum, % (range)
SpO <sub>2</sub> <sup>a</sup> measurements from Masimo MightySat Rx	100 (100-100)	53 (53-71)
SpO <sub>2</sub> measurements from Apple Watch Series 7	100 (100-100)	61 (61-85)
SaO <sub>2</sub> <sup>b</sup> measurements	99.8 (99.2-99.8)	56 (56-70)

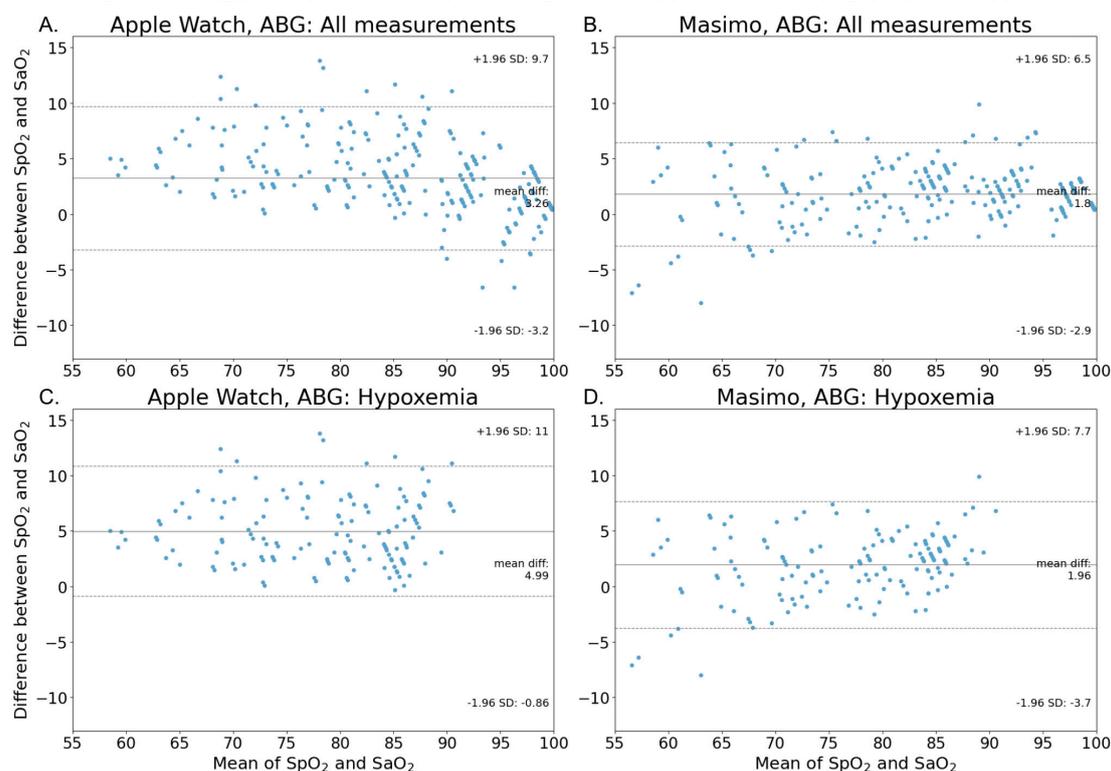
<sup>a</sup>SpO<sub>2</sub>: peripheral oxygen saturation.

<sup>b</sup>SaO<sub>2</sub>: arterial oxygen saturation.

Each measurement from the Masimo MightySat Rx and the Apple Watch Series 7 was compared with its concomitant ABG measurement to determine whether it fell within, above, or below 2% error, which was the acceptable range as defined by the Masimo MightySat Rx manual. The percentage of measurements falling into these different categories demonstrates whether there exists a trend of overestimation or underestimation of SpO<sub>2</sub> levels. Both the MightySat Rx and the Apple Watch Series 7 tended to overestimate SpO<sub>2</sub>, with

MDEs of 1.80% and 3.26%, respectively (Figure 3A, B). The Apple Watch Series 7 had a higher proportion of overestimated measurements (n=174, 56.49% of SpO<sub>2</sub> measurements were overestimated) as compared with the MightySat Rx (where n=137, 44.48% of measurements were overestimated). A higher percentage of Masimo MightySat Rx SpO<sub>2</sub> measurements (n=151, 49.03%) fell within the 2% error range of the reference SaO<sub>2</sub> values compared with Apple Watch Series 7 SpO<sub>2</sub> measurements (n=99, 32.14%).

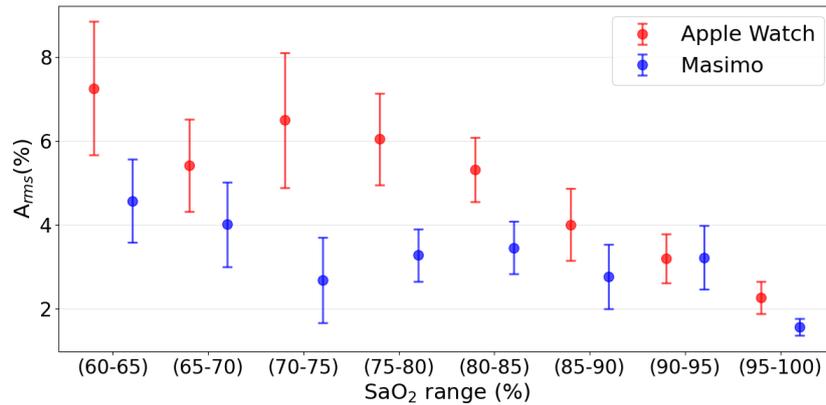
**Figure 3.** Bland-Altman plots demonstrate the differences between simultaneous ABG-based SaO<sub>2</sub> measurements and either all (A) or hypoxemia-only (SaO<sub>2</sub><88%) (C) SpO<sub>2</sub> measurements from the Apple Watch Series 7. The same comparisons are shown between SaO<sub>2</sub> and SpO<sub>2</sub> for the Masimo MightySat Rx (all, B; and hypoxemia-only, D). The solid line shows the mean difference of the measurements (MDE), and the dashed lines show the 95% limits of agreement. A<sub>rms</sub>: average root mean square; SaO<sub>2</sub>: arterial oxygen saturation; SpO<sub>2</sub>: peripheral oxygen saturation.



A<sub>rms</sub> across 5% SaO<sub>2</sub> bins demonstrated a decreasing trend (Figure 4) for both devices with increasing SaO<sub>2</sub>. The highest A<sub>rms</sub> values were observed in the lowest saturation ranges (60%-65%), reaching 7.26% for the Apple Watch Series 7 and 4.57% for the Masimo MightySat Rx. In contrast, the

lowest A<sub>rms</sub> values occurred at the highest saturation range (95%-100%), where A<sub>rms</sub> decreased to 2.27% for the Apple Watch Series 7 and 1.53% for the Masimo MightySat Rx. Overall, the Masimo MightySat Rx exhibited lower A<sub>rms</sub> values than the Apple Watch Series 7.

**Figure 4.** Average root mean square of SpO<sub>2</sub> measurements for Apple Watch Series 7 (red) and Masimo MightySat Rx (blue) binned by 5% oxygen saturation bins ranges from 60% to 100%. Points indicate the A<sub>rms</sub> within each SaO<sub>2</sub> range, and dashed lines represent the 95% CIs. ABG: arterial blood gas; SaO<sub>2</sub>: arterial oxygen saturation; SpO<sub>2</sub>: peripheral oxygen saturation.



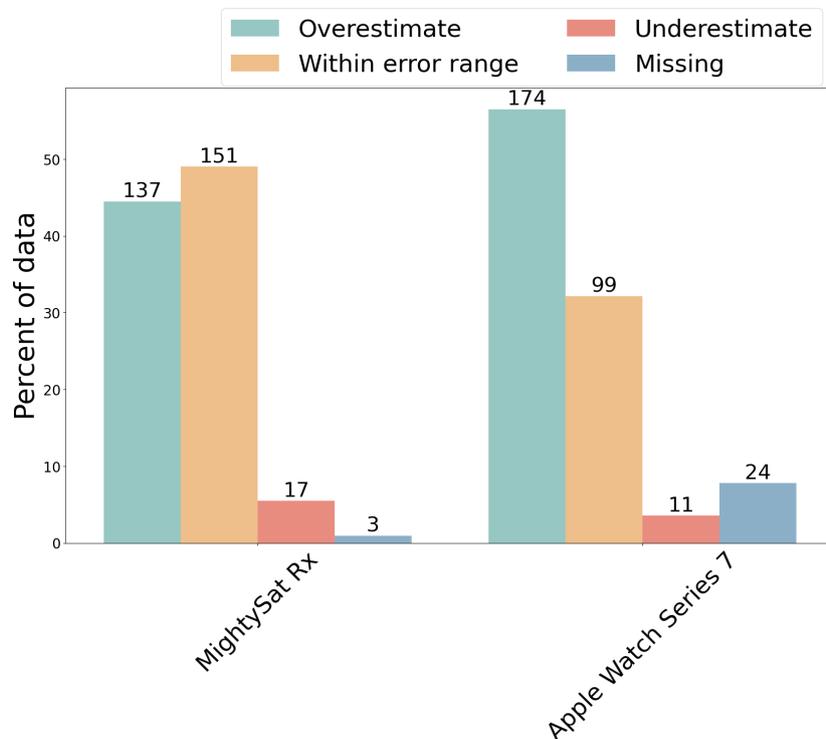
A Bland-Altman analysis demonstrated that SpO<sub>2</sub> measurements tended to be overestimated since the points in the figure mostly lie above the zero-bias line (Figure 3). Overall, the Masimo MightySat Rx showed higher accuracy against the SaO<sub>2</sub> ground truth measures than the Apple Watch Series 7, with an A<sub>rms</sub> difference in accuracy (95% CI) of 2.98% (2.72%-3.25%) for the Masimo MightySat Rx and 4.63% (4.26%-5%) for the Apple Watch Series 7.

**Missingness Evaluation**

Of attempted SpO<sub>2</sub> measurements, 305/308 (99%) and 284/308 (92%) were successfully reported by the Masimo

MightySat Rx and Apple Watch Series 7, respectively. In other words, when the device was prompted to make a measurement, the Masimo MightySat Rx was unable to produce 3/308 (0.97%) attempted measurements, whereas the Apple Watch Series 7 was unable to produce 24/308 (7.79%) attempted measurements (Figure 5, blue bar). Thus, in terms of missingness, the Apple Watch Series 7 had more missingness (24/308, 7.79%) compared with the MightySat Rx (3/308, 0.97%), indicating that both have high likelihood (>90%) for successfully obtaining a measurement when a measurement was attempted.

**Figure 5.** Relative percentages of data falling into the categories of overestimated measurements (green), measurements within acceptable error ranges (yellow), underestimated measurements (red), and missing measurements (blue) for the Masimo MightySat Rx and Apple Watch Series 7.



## Accuracy of Measurements Under Hypoxemia

The clinical threshold for hypoxemia is blood oxygen saturation less than 88% [41]. As noted in the Introduction, occult hypoxemia is a situation where a patient actually has blood oxygen saturation less than 88%, but their blood oxygen saturation measurements read as higher (ie, above the threshold that would lead to a clinical intervention). Among the 308 ABG-based SaO<sub>2</sub> measurements, 165 (54%) were less than 88%. Among those, the Masimo MightySat Rx and Apple Watch Series 7 produced 164 (99%) and 150 (91%) simultaneous SpO<sub>2</sub> measurements, respectively; the remaining were attempted but the devices did not generate values and thus those values were considered missing. Among those SpO<sub>2</sub> measurements where SaO<sub>2</sub> was less than 88%, 86/164 (52.44%) of the Masimo MightySat Rx measurements and 128/150 (85.33%) of the Apple Watch Series 7 measurements were overestimated. Importantly, 3 (1.8%) of the Masimo MightySat Rx measurements and 6 (4%) of the Apple Watch Series 7 measurements were actually reported to be above 92%. These instances of occult hypoxemia occurred in participants 2 and 7 using the Masimo MightySat Rx, and in participants 2, 8, and 9 using the Apple Watch Series 7. Notably, only one measurement from a single participant (participant #2) demonstrated an instance of occult hypoxemia that was detected simultaneously on both devices.

Looking specifically under conditions of hypoxemia, the MDE of SpO<sub>2</sub> for the Masimo MightySat Rx and the Apple Watch Series 7 was 1.96% and 4.99%, respectively; higher than the overall MDE of 1.80% and 3.26% reported

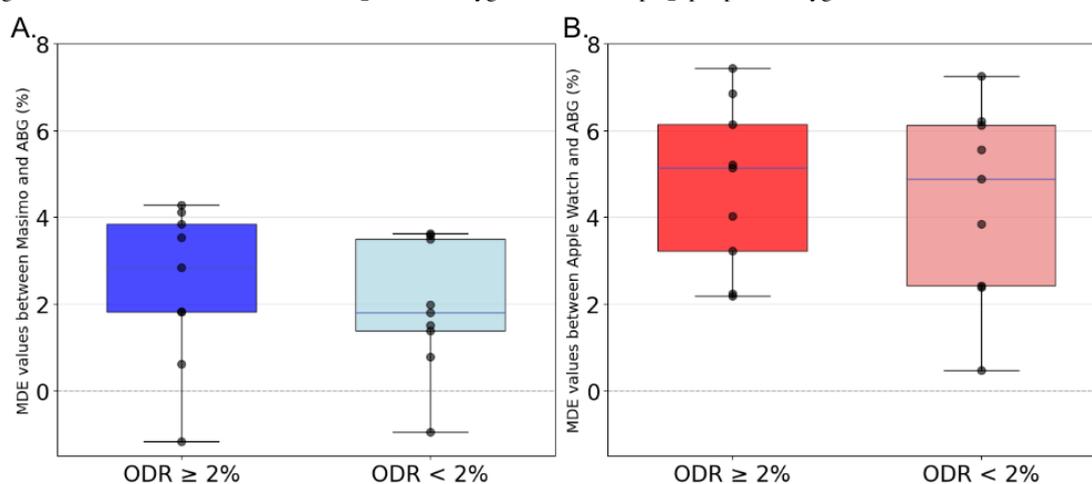
above (Figure 3C, D; Table S1 in Multimedia Appendix 1, which represents participant-level MDE of SpO<sub>2</sub> for Masimo MightySat Rx and the Apple Watch Series 7). The A<sub>rms</sub> (95% CI) during hypoxemia was 3.52% (3.18%-3.86%) and 5.82% (5.32%-6.31%) SpO<sub>2</sub> for the Masimo MightySat Rx and the Apple Watch Series 7, respectively, which are also both higher than the overall A<sub>rms</sub>.

## Relationship Between Device Performance and ODR

Before comparing device performance under ODR at least 2% per minute and ODR less than 2% per minute conditions, we excluded 57 fully saturated measurements (SpO<sub>2</sub> or SaO<sub>2</sub>>99%) and 79 resaturation measurements. Among the remaining 172 desaturating SaO<sub>2</sub> measurements (Table S2 in Multimedia Appendix 5) out of the total 308 measurements, 169 and 157 SpO<sub>2</sub> measurements were collected from the Masimo MightySat Rx and the Apple Watch Series 7, respectively. Among those, 83/169 Masimo MightySat Rx and 78/157 Apple Watch Series 7 measurements occurred when the ODR was greater than two percent per minute.

When the ODR increased above 2% per minute, the Masimo MightySat Rx A<sub>rms</sub> increased from 3.04% to 3.79%, and the MDE increased from 1.86% to 2.38%. Also at ODR greater than 2% per minute, the Apple Watch Series 7 MDE increased from 4.16% to 4.41% (Figure S2 in Multimedia Appendix 3) but A<sub>rms</sub> decreased from 5.50% to 5.22%. No significant correlation was found (Figure 6) between the ODR and MDE for the Masimo MightySat Rx nor the Apple Watch (2-sided paired *t* test: *P*=.12 and .22, respectively).

**Figure 6.** Boxplots demonstrate the participant-level mean directional error of (A) Masimo MightySat Rx and (B) Apple Watch Series 7 between SpO<sub>2</sub> and arterial oxygen saturation measurements when the oxygen desaturation rate is at least 2% per minute and less than 2% per minute. ABG: arterial blood gas; MDE: mean directional error; SaO<sub>2</sub>: arterial oxygen saturation; SpO<sub>2</sub>: peripheral oxygen saturation.



## Discussion

### Principal Findings

We designed and implemented this study to evaluate the performance of optical blood oxygen saturation measurements (SpO<sub>2</sub>) from smartwatches and finger pulse oximeters under conditions of hypoxemia that were induced by

a controlled oxygen desaturation protocol. There are three main findings of our study: (1) both the Masimo MightySat Rx and the Apple Watch Series 7 often overestimate true blood oxygen saturation values, with 44% and 56% of their measurements in this study, respectively, found to be higher than the ground truth SaO<sub>2</sub> measures; (2) both devices have lower accuracy during hypoxemia than at normal blood oxygen saturation; and (3) both devices showed a trend

toward lower accuracy when the ODR is larger than 2% per minute, but the difference is not statistically significant.

Our previous study [32], which used the Masimo MightySat Rx as the reference measurement for oxygen saturation to analytically validate four commercial wearables in 49 patients with low blood oxygen saturation, many of whom had chronic pulmonary disease, found that the Apple Watch Series 7 overestimated blood oxygen saturation measurements 17% of the time. In this current study, where we used ABG measurements of SaO<sub>2</sub> as the reference standard “ground truth” measure, as well as a controlled oxygen desaturation study protocol in healthy individuals to achieve low blood oxygen saturation, the Apple Watch Series 7 overestimated blood oxygen saturation 56% of the time. In this study, we also observed that the Masimo MightySat Rx, the reference standard used in the previous study, also overestimated oxygen saturation levels 44% of the time. One likely reason for the discrepancy in the Apple Watch Series 7 overestimation rate between the 2 studies is the choice of the reference standard measurement—the ABG measurements used in the current study are a more trustworthy ground truth measurement. Another likely factor is the difference in the distribution of oxygen saturation values between the 2 studies. Specifically, our previous study did not include a desaturation protocol, and the SpO<sub>2</sub> measurements obtained from the Masimo MightySat Rx ranged only from 82% to 100%, which may have contributed to a lower observed rate of overestimation. Blood oxygen saturation overestimation is particularly dangerous under conditions of occult hypoxemia (where the actual SaO<sub>2</sub> is <88% but the measured SpO<sub>2</sub> is >92%), when a patient’s blood oxygen saturation is thought to be normal and care is delayed [42]. In our study, although the Apple Watch Series 7 had a high rate of overestimation (56% of measurements), only 6/150 (4%) of the Apple Watch Series 7 measurements and 3/164 (1.8%) of the Masimo MightySat Rx measurements actually fell into the category of occult hypoxemia. It is likely that, had there been more measurements concentrated only slightly below 88%, there would have been more occurrences of occult hypoxemia observed in our study. This is because the mean bias of the Apple Watch Series 7 measurements under hypoxemia conditions (4.99%) just slightly exceeds the minimum difference (4%) between SpO<sub>2</sub> and SaO<sub>2</sub> that defines occult hypoxemia (92% vs 88%).

Our study also demonstrates that the performance of both the finger pulse oximeter and the smartwatch suffers during hypoxemia as compared with conditions of normal blood oxygen saturation. The overestimation rate increased from 44.48% to 52.44% for the Masimo MightySat Rx and from 56.49% to 85.33% for the Apple Watch Series 7, while the A<sub>rms</sub> increased from 2.98% to 3.52% and 4.63% to 5.82% during hypoxemia for the 2 devices, respectively. For 510(k) clearance, the FDA currently recommends an A<sub>rms</sub> of less than 3% for transmission pulse oximeters (eg, Masimo MightySat Rx) and less than 3.5% for reflectance oximeters (eg, Apple Watch) [34]. According to the FDA’s 2025 draft guidance, the A<sub>rms</sub> specification for both sensor types will be standardized to less than 3% [43]. Both

devices used in our study exhibited A<sub>rms</sub> values exceeding the FDA threshold under conditions of hypoxemia. A previous study that recruited 50 healthy adults and conducted oxygen desaturation cycles from approximately 100% to 70% SaO<sub>2</sub> evaluated the accuracy of Apple Watch SpO<sub>2</sub> measurements using ABG as the gold standard; this study also reported a trend of overestimation in the Apple Watch Series 6, with an A<sub>rms</sub> of 2.18% (95% CI 1.55%-2.84%), meeting current or upcoming FDA A<sub>rms</sub> requirements. The missingness in their study was 5.29% (966 of 1020 attempted measurements gave readings) as compared with 7.79% (284 of 308 attempted measurements gave readings) in our study, likely due to more measurements in our study attempted at lower oxygen saturation [44]. Another study involving 167 participants with acute exacerbation of COPD reported an MDE of 0.46% between Apple Watch Series 6 measurements and ABG-based SaO<sub>2</sub> measurements [30].

When the ODR exceeded 2% per minute, A<sub>rms</sub> increased from 3.04% to 3.79% for the Masimo MightySat Rx and decreased from 5.50% to 5.22% for the Apple Watch Series 7. Even though the difference between SaO<sub>2</sub> and SpO<sub>2</sub> was not found to be statistically significant during desaturation, the fact that A<sub>rms</sub> exceeded 3% for both devices indicates that they should be used with caution, particularly under rapid desaturation conditions. Accurate performance during oxygen desaturation is crucial not only for monitoring oxygen levels but also for clinical decision-making. This finding is especially relevant for home monitoring in patients with OSA, where the number of desaturation episodes (ie, a drop in mean oxygen saturation of ≥4% over the course of 120 s) is linked to apnea and hypopnea [24]. Inaccuracies and missingness during desaturation could result in misclassification of these episodes, leading to inappropriate assessment and treatment plans for patients.

Although current FDA guidance recommends validation of pulse oximeter accuracy within the SpO<sub>2</sub> range of 70% to 100%, our results highlight the importance of evaluating device performance at lower oxygen saturation levels. Device accuracy declined substantially as SaO<sub>2</sub> decreased from 100% to 60%, with A<sub>rms</sub> increasing from 2.27% to 7.26% for the Apple Watch Series 7 and from 1.53% to 4.57% for the Masimo MightySat Rx. These results suggest that limiting validation to SpO<sub>2</sub> values of at least 70% may overlook performance limitations in more severe hypoxemia, where accurate measurement is critical for timely clinical interventions. Expanding regulatory validation requirements to include lower saturation ranges may therefore provide a more comprehensive assessment of pulse oximeter performance.

## Limitations

The main limitation of this study is the small sample size, which resulted in insufficient representation across skin tone categories and age ranges. According to the FDA guidance for pulse oximeters, 10 or more healthy participants that vary in age and gender should be included, and a minimum of 2 subjects, or 15% of the study participant pool, need to be darkly pigmented [34]. However, in our study, only four

(light, intermediate, tan, brown) out of the seven possible ITA skin types (very light, light, intermediate, tan, brown, dark, and very dark) were represented. A benefit of our study, however, was the use of objective skin tone measures using a device rather than the more common subjective human assessment using a comparison scale like Fitzpatrick or Monk (Figure S4 in [Multimedia Appendix 6](#)). Moreover, only the brown and tan skin types were represented by more than one participant, and we did not have any representation from the dark, very dark, and very light categories. This lack of diversity across skin tones limits the generalizability of our findings to individuals with lighter or darker skin tones. We recognize this as a limitation in our study as a result of the limited time and resources available to perform the study more comprehensively.

While our study captured a wide range of oxygen saturation values from 60% to 100%, covering the FDA-recommended SpO<sub>2</sub> accuracy validation range of 70% to 100%, we did not measure the response time of the finger pulse oximeter, and the Apple Watch Series 7 reported SpO<sub>2</sub> readings after a fixed 15-second countdown. Previous studies have shown that finger pulse oximeters may have a delay in displaying the blood oxygen saturation values during the onset of hypoxia [45]. Several factors can contribute to this delay, including the measurement site and poor peripheral perfusion [45]. Other research suggests that fingertip oximeters have the slowest response in desaturation changes from normoxemia to hypoxemia [46]. Future studies could consider evaluating the response time during desaturation events to ensure that pulse oximeters provide not only accurate but also timely measurements.

Although our experiment simulated the process of oxygen desaturation in healthy participants, there are additional factors to consider that may prevent these results from generalizing to patients of older age and/or who may experience oxygen desaturation as a result of a health condition. The age range of our study participants was between 19 and 28 years; COPD is most prevalent in people aged 65 to 84 [47]. Arterial stiffness of older people results in changes in the propagation of the pulse to the periphery, thereby influencing the peripheral pulse timing and shape characteristics which can affect the performance of pulse oximeters [48].

## Funding

This study was supported in part by funds from AstraZeneca (LS and SJB), by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (R01DK133531 to JD), and by NSF CAREER #2339669. The content is solely the responsibility of the authors and does not necessarily represent the official views of AstraZeneca or the National Institutes of Health.

## Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

JD sits on the Google Consumer Health Advisory Board and is a consultant to Samsung Research America and Jones Day. All other authors have no conflicts to declare.

## Multimedia Appendix 1

Finally, this protocol cannot demonstrate SpO<sub>2</sub> measurement performance over long periods since each participant only spent 60 minutes in the first oxygen desaturation sequence and 48 minutes in the second one. While it would be infeasible to induce oxygen desaturation for such a long period in healthy adults, clinically, it is necessary to be able to consistently monitor oxygen saturation values overnight, for example, in COPD patients who often experience nocturnal hypoxemia [2,49]. It is therefore necessary to perform clinical validation among COPD patients with nocturnal desaturation, defined as having a SpO<sub>2</sub> value below 90% for more than 30% of the time in bed during one or more nights [50, 51]. Future clinical validation studies should thus perform real-world monitoring over longer periods, and particularly during sleep, in COPD and sleep apnea patients to capture nocturnal desaturation caused by those conditions [52].

## Conclusion

In this controlled hypoxemia study, we evaluated how well the Masimo MightySat Rx and the Apple Watch Series 7 oxygen saturation measurements (SpO<sub>2</sub>) correspond to ABG measurements (SaO<sub>2</sub>) across a range of arterial oxygen saturation levels, ranging from approximately 60% to 100%. Both devices consistently overestimated SpO<sub>2</sub>, with accuracy declining notably during hypoxemia. The Apple Watch Series 7 mean bias suggests a likelihood for missing instances of hypoxemia, particularly at SaO<sub>2</sub> values below, but close to, 88%. Both the Apple Watch Series 7 and Masimo MightySat Rx exhibited A<sub>rms</sub> values exceeding the FDA threshold under conditions of hypoxemia. While past studies have implicated high ODRs in increasing measurement error, we found no statistically significant relationship between ODR and measurement error for either device. Overall, our findings of SpO<sub>2</sub> overestimation and high A<sub>rms</sub> values underscore the need for caution when interpreting oxygen saturation values from these devices. The small sample size and limited diversity in skin tone and age restrict the generalizability of our findings. Future studies should include larger and more diverse populations to evaluate the performance of wearable-based pulse oximetry.

Participant-level mean directional error (MDE) and variability of SpO<sub>2</sub> measurement error for Apple Watch Series 7 and Masimo MightySat Rx.

[\[DOCX File \(Microsoft Word File\), 14 KB-Multimedia Appendix 1\]](#)

### Multimedia Appendix 2

Workflow for generating participant-level mean directional error (MDE) values for the Masimo MightySat Rx and the Apple Watch Series 7 prior to performing paired *t* tests.

[\[PNG File \(Portable Network Graphics File\), 180 KB-Multimedia Appendix 2\]](#)

### Multimedia Appendix 3

Boxplots demonstrate the different errors of pulse oximetry between measurements with oxygen desaturation rate  $\geq 2\%$ /min and  $< 2\%$ /min (A) SpO<sub>2</sub> measurements of the Masimo MightySat Rx (B) SpO<sub>2</sub> measurements of the smartwatch (Apple Watch Series 7).

[\[PNG File \(Portable Network Graphics File\), 47 KB-Multimedia Appendix 3\]](#)

### Multimedia Appendix 4

Quantile-quantile (QQ) plots of participant-level mean directional error (MDE) for Masimo MightySat Rx and Apple Watch Series 7.

[\[PNG File \(Portable Network Graphics File\), 66 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

SpO<sub>2</sub> and SaO<sub>2</sub> measurements when oxygen desaturation rate (ODR)  $\geq 2\%$  or ODR  $< 2\%$ .

[\[DOCX File \(Microsoft Word File\), 59 KB-Multimedia Appendix 5\]](#)

### Multimedia Appendix 6

SpO<sub>2</sub> measurement bias by sex (left panel) and skin tone (right panel) for the Masimo MightySat Rx (top row) and Apple Watch Series 7 (bottom row).

[\[PNG File \(Portable Network Graphics File\), 132 KB-Multimedia Appendix 6\]](#)

## References

1. Pulse oximetry. Yale Medicine. URL: <https://www.yalemedicine.org/conditions/pulse-oximetry> [Accessed 2023-12-12]
2. Soguel Schenkel N, Burdet L, de Muralt B, Fitting JW. Oxygen saturation during daily activities in chronic obstructive pulmonary disease. *Eur Respir J*. Dec 1996;9(12):2584-2589. [doi: [10.1183/09031936.96.09122584](https://doi.org/10.1183/09031936.96.09122584)] [Medline: [8980973](https://pubmed.ncbi.nlm.nih.gov/8980973/)]
3. Inwald D, Roland M, Kuitert L, McKenzie SA, Petros A. Oxygen treatment for acute severe asthma. *BMJ*. Jul 14, 2001;323(7304):98-100. [doi: [10.1136/bmj.323.7304.98](https://doi.org/10.1136/bmj.323.7304.98)] [Medline: [11451788](https://pubmed.ncbi.nlm.nih.gov/11451788/)]
4. Panos RJ, Eschenbacher W. Exertional desaturation in patients with chronic obstructive pulmonary disease. *COPD*. Dec 2009;6(6):478-487. [doi: [10.3109/15412550903341497](https://doi.org/10.3109/15412550903341497)] [Medline: [19938972](https://pubmed.ncbi.nlm.nih.gov/19938972/)]
5. Dogra AC, Gupta U, Sarkar M, Padam A. Exercise-induced desaturation in patients with chronic obstructive pulmonary disease on six-minute walk test. *Lung India*. 2015;32(4):320-325. [doi: [10.4103/0970-2113.159550](https://doi.org/10.4103/0970-2113.159550)] [Medline: [26180379](https://pubmed.ncbi.nlm.nih.gov/26180379/)]
6. Vold ML, Aasebø U, Wilsgaard T, Melbye H. Low oxygen saturation and mortality in an adult cohort: the Tromsø study. *BMC Pulm Med*. Feb 12, 2015;15:9. [doi: [10.1186/s12890-015-0003-5](https://doi.org/10.1186/s12890-015-0003-5)] [Medline: [25885261](https://pubmed.ncbi.nlm.nih.gov/25885261/)]
7. Davis J. What is COPD hypoxia? WebMD. URL: <https://www.webmd.com/lung/copd/copd-hypoxia> [Accessed 2023-12-12]
8. Knowler MT, Dunagan DP, Adair NE, Chin R Jr. Baseline oxygen saturation predicts exercise desaturation below prescription threshold in patients with chronic obstructive pulmonary disease. *Arch Intern Med*. Mar 12, 2001;161(5):732-736. [doi: [10.1001/archinte.161.5.732](https://doi.org/10.1001/archinte.161.5.732)] [Medline: [11231707](https://pubmed.ncbi.nlm.nih.gov/11231707/)]
9. Little SA, Elkholy MM, Chalmers GW, Farouk A, Patel KR, Thomson NC. Predictors of nocturnal oxygen desaturation in patients with COPD. *Respir Med*. Mar 1999;93(3):202-207. [doi: [10.1016/s0954-6111\(99\)90009-4](https://doi.org/10.1016/s0954-6111(99)90009-4)] [Medline: [10464879](https://pubmed.ncbi.nlm.nih.gov/10464879/)]
10. Collins JA, Rudenski A, Gibson J, Howard L, O'Driscoll R. Relating oxygen partial pressure, saturation and content: the haemoglobin-oxygen dissociation curve. *Breathe (Sheff)*. Sep 2015;11(3):194-201. [doi: [10.1183/20734735.001415](https://doi.org/10.1183/20734735.001415)] [Medline: [26632351](https://pubmed.ncbi.nlm.nih.gov/26632351/)]
11. Moran O, Brown L, Murray H, Rang L. Diagnostic accuracy of venous blood gases compared to arterial blood gases. *Eur Respir J*. Sep 2013;42(Suppl 57):3987. URL: <https://publications.ersnet.org/content/erj/42/suppl57/p3987?implicit-login=true%26983> [Accessed 2023-12-12] [doi: [10.1183/13993003/erj.42.Suppl\\_57.P3987](https://doi.org/10.1183/13993003/erj.42.Suppl_57.P3987)]

12. Plüddemann A, Thompson M, Heneghan C, Price C. Pulse oximetry in primary care: primary care diagnostic technology update. *Br J Gen Pract.* May 2011;61(586):358-359. [doi: [10.3399/bjgp11X572553](https://doi.org/10.3399/bjgp11X572553)] [Medline: [21619771](https://pubmed.ncbi.nlm.nih.gov/21619771/)]
13. Chan ED, Chan MM, Chan MM. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respir Med.* Jun 2013;107(6):789-799. [doi: [10.1016/j.rmed.2013.02.004](https://doi.org/10.1016/j.rmed.2013.02.004)] [Medline: [23490227](https://pubmed.ncbi.nlm.nih.gov/23490227/)]
14. Wearable technology market size, share & industry analysis, by type (passive and active), by technology (iot based, AR and VR, and others), by end-use (health & fitness, BFSI, gaming & entertainment, fashion, travel, education, and logistics & warehouse), and regional forecast, 2026–2034. *Fortune Business Insights.* 2026. URL: <https://www.fortunebusinessinsights.com/wearable-technology-market-106000> [Accessed 2026-03-08]
15. Windisch P, Schröder C, Förster R, Cihoric N, Zwahlen DR. Accuracy of the Apple Watch oxygen saturation measurement in adults: a systematic review. *Cureus.* Feb 2023;15(2):e35355. [doi: [10.7759/cureus.35355](https://doi.org/10.7759/cureus.35355)] [Medline: [36974257](https://pubmed.ncbi.nlm.nih.gov/36974257/)]
16. Spaccarotella C, Polimeni A, Mancuso C, Pelaia G, Esposito G, Indolfi C. Assessment of non-invasive measurements of oxygen saturation and heart rate with an Apple smartwatch: comparison with a standard pulse oximeter. *J Clin Med.* Mar 8, 2022;11(6):1467. [doi: [10.3390/jcm11061467](https://doi.org/10.3390/jcm11061467)] [Medline: [35329793](https://pubmed.ncbi.nlm.nih.gov/35329793/)]
17. Littell L, Roelle L, Dalal A, et al. Assessment of Apple Watch Series 6 pulse oximetry and electrocardiograms in a pediatric population. *PLOS Digit Health.* Aug 2022;1(8):e0000051. [doi: [10.1371/journal.pdig.0000051](https://doi.org/10.1371/journal.pdig.0000051)] [Medline: [36812630](https://pubmed.ncbi.nlm.nih.gov/36812630/)]
18. Pätz C, Michaelis A, Markel F, et al. Accuracy of the Apple Watch oxygen saturation measurement in adults and children with congenital heart disease. *Pediatr Cardiol.* Feb 2023;44(2):333-343. [doi: [10.1007/s00246-022-02987-w](https://doi.org/10.1007/s00246-022-02987-w)] [Medline: [35995951](https://pubmed.ncbi.nlm.nih.gov/35995951/)]
19. Milner QJW, Mathews GR. An assessment of the accuracy of pulse oximeters. *Anaesthesia.* Apr 2012;67(4):396-401. [doi: [10.1111/j.1365-2044.2011.07021.x](https://doi.org/10.1111/j.1365-2044.2011.07021.x)] [Medline: [22324874](https://pubmed.ncbi.nlm.nih.gov/22324874/)]
20. 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.* May 5, 2010;10:9. [doi: [10.1186/1471-227X-10-9](https://doi.org/10.1186/1471-227X-10-9)] [Medline: [20444248](https://pubmed.ncbi.nlm.nih.gov/20444248/)]
21. Bhutta BS, Alghoula F, Berim I. Hypoxia. In: *StatPearls.* StatPearls Publishing; 2025. URL: <http://www.ncbi.nlm.nih.gov/books/NBK482316> [Accessed 2026-01-26]
22. Valbuena VSM, Seelye S, Sjoding MW, et al. Racial bias and reproducibility in pulse oximetry among medical and surgical inpatients in general care in the Veterans Health Administration 2013-19: multicenter, retrospective cohort study. *BMJ.* Jul 6, 2022;378:e069775. [doi: [10.1136/bmj-2021-069775](https://doi.org/10.1136/bmj-2021-069775)] [Medline: [35793817](https://pubmed.ncbi.nlm.nih.gov/35793817/)]
23. Jung H, Kim D, Lee W, et al. Performance evaluation of a wrist-worn reflectance pulse oximeter during sleep. *Sleep Health.* Oct 2022;8(5):420-428. [doi: [10.1016/j.sleh.2022.04.003](https://doi.org/10.1016/j.sleh.2022.04.003)] [Medline: [35817700](https://pubmed.ncbi.nlm.nih.gov/35817700/)]
24. Temirbekov D, Güneş S, Yazıcı ZM, Sayın İ. The ignored parameter in the diagnosis of obstructive sleep apnea syndrome: the oxygen desaturation index. *Turk Arch Otorhinolaryngol.* Mar 2018;56(1):1-6. [doi: [10.5152/tao.2018.3025](https://doi.org/10.5152/tao.2018.3025)] [Medline: [29988275](https://pubmed.ncbi.nlm.nih.gov/29988275/)]
25. Landsberg R, Friedman M, Ascher-Landsberg J. Treatment of hypoxemia in obstructive sleep apnea. *Am J Rhinol.* 2001;15(5):311-313. [Medline: [11732817](https://pubmed.ncbi.nlm.nih.gov/11732817/)]
26. Horakova L, Roubik K. Performance of different pulse oximeters can affect the duration of field breathing experiments. Presented at: 2019 E-Health and Bioengineering Conference (EHB); Nov 21-23, 2019:1-4; Iasi, Romania. 2019.[doi: [10.1109/EHB47216.2019.8969917](https://doi.org/10.1109/EHB47216.2019.8969917)]
27. DeMeulenaere S. Pulse oximetry: uses and limitations. *J Nurse Pract.* May 2007;3(5):312-317. [doi: [10.1016/j.nurpra.2007.02.021](https://doi.org/10.1016/j.nurpra.2007.02.021)]
28. Rafl J, Bachman TE, Rafl-Huttova V, Walzel S, Rozanek M. Commercial smartwatch with pulse oximeter detects short-time hypoxemia as well as standard medical-grade device: validation study. *Digit Health.* 2022;8:20552076221132127. [doi: [10.1177/20552076221132127](https://doi.org/10.1177/20552076221132127)] [Medline: [36249475](https://pubmed.ncbi.nlm.nih.gov/36249475/)]
29. Santos M, Vollam S, Pimentel MA, et al. The use of wearable pulse oximeters in the prompt detection of hypoxemia and during movement: diagnostic accuracy study. *J Med Internet Res.* Feb 15, 2022;24(2):e28890. [doi: [10.2196/28890](https://doi.org/10.2196/28890)] [Medline: [35166690](https://pubmed.ncbi.nlm.nih.gov/35166690/)]
30. Arslan B, Sener K, Guven R, et al. Accuracy of the Apple Watch in measuring oxygen saturation: comparison with pulse oximetry and ABG. *Ir J Med Sci.* Feb 2024;193(1):477-483. [doi: [10.1007/s11845-023-03456-w](https://doi.org/10.1007/s11845-023-03456-w)] [Medline: [37440093](https://pubmed.ncbi.nlm.nih.gov/37440093/)]
31. Bengtsson J, Bake B, Johansson A, Bengtson JP. End-tidal to arterial oxygen tension difference as an oxygenation index. *Acta Anaesthesiol Scand.* Mar 2001;45(3):357-363. [doi: [10.1034/j.1399-6576.2001.045003357.x](https://doi.org/10.1034/j.1399-6576.2001.045003357.x)] [Medline: [11207474](https://pubmed.ncbi.nlm.nih.gov/11207474/)]
32. Jiang Y, Spies C, Magin J, Bhosai SJ, Snyder L, Dunn J. Investigating the accuracy of blood oxygen saturation measurements in common consumer smartwatches. *PLoS Digit Health.* Jul 2023;2(7):e0000296. [doi: [10.1371/journal.pdig.0000296](https://doi.org/10.1371/journal.pdig.0000296)] [Medline: [37437005](https://pubmed.ncbi.nlm.nih.gov/37437005/)]

33. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med*. 2020;3:55. [doi: [10.1038/s41746-020-0260-4](https://doi.org/10.1038/s41746-020-0260-4)] [Medline: [32337371](https://pubmed.ncbi.nlm.nih.gov/32337371/)]
34. Pulse oximeters - premarket notification submissions [510(k)s]: guidance for industry and food and drug administration staff. U.S. Food and Drug Administration (FDA), Center for Devices and Radiological Health (CDRH); 2013. URL: <https://www.fda.gov/media/72470/download> [Accessed 2024-05-30]
35. 510(k) premarket notification. US Food and Drug Administration. URL: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K201456> [Accessed 2022-03-03]
36. Benni PB, MacLeod D, Ikeda K, Lin HM. A validation method for near-infrared spectroscopy based tissue oximeters for cerebral and somatic tissue oxygen saturation measurements. *J Clin Monit Comput*. Apr 2018;32(2):269-284. [doi: [10.1007/s10877-017-0015-1](https://doi.org/10.1007/s10877-017-0015-1)] [Medline: [28374103](https://pubmed.ncbi.nlm.nih.gov/28374103/)]
37. Severinghaus JW. Simple, accurate equations for human blood O<sub>2</sub> dissociation computations. *J Appl Physiol Respir Environ Exerc Physiol*. Mar 1979;46(3):599-602. [doi: [10.1152/jappl.1979.46.3.599](https://doi.org/10.1152/jappl.1979.46.3.599)] [Medline: [35496](https://pubmed.ncbi.nlm.nih.gov/35496/)]
38. Whenn CB, Wilson DL, Ruehland WR, Churchward TJ, Worsnop C, Tolson J. The impact of study type and sleep measurement on oxygen desaturation index calculation. *J Clin Sleep Med*. May 1, 2024;20(5):709-717. [doi: [10.5664/jcsm.10982](https://doi.org/10.5664/jcsm.10982)] [Medline: [38169424](https://pubmed.ncbi.nlm.nih.gov/38169424/)]
39. Chardon A, Cretois I, Hourseau C. Skin colour typology and suntanning pathways. *Int J Cosmet Sci*. Aug 1991;13(4):191-208. [doi: [10.1111/j.1467-2494.1991.tb00561.x](https://doi.org/10.1111/j.1467-2494.1991.tb00561.x)] [Medline: [19291061](https://pubmed.ncbi.nlm.nih.gov/19291061/)]
40. Del Bino S, Bernerd F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br J Dermatol*. Oct 2013;169 Suppl 3:33-40. [doi: [10.1111/bjd.12529](https://doi.org/10.1111/bjd.12529)] [Medline: [24098899](https://pubmed.ncbi.nlm.nih.gov/24098899/)]
41. O'Driscoll BR, Howard LS, Davison AG, British Thoracic Society. BTS guideline for emergency oxygen use in adult patients. *Thorax*. Oct 2008;63 Suppl 6:vi1-vi68. [doi: [10.1136/thx.2008.102947](https://doi.org/10.1136/thx.2008.102947)] [Medline: [18838559](https://pubmed.ncbi.nlm.nih.gov/18838559/)]
42. Henry NR, Hanson AC, Schulte PJ, et al. Disparities in hypoxemia detection by pulse oximetry across self-identified racial groups and associations with clinical outcomes. *Crit Care Med*. Feb 1, 2022;50(2):204-211. [doi: [10.1097/CCM.0000000000005394](https://doi.org/10.1097/CCM.0000000000005394)] [Medline: [35100193](https://pubmed.ncbi.nlm.nih.gov/35100193/)]
43. Pulse oximeters for medical purposes - non-clinical and clinical performance testing, labeling, and premarket submission recommendations. U.S. Department of Health and Human Services; U.S. Food and Drug Administration; Center for Devices and Radiological Health; 2025. URL: <https://www.fda.gov/media/184896/download> [Accessed 2025-06-05]
44. Blood oxygen app on Apple Watch. Apple Inc; 2022. URL: [https://www.apple.com/healthcare/docs/site/Blood\\_Oxygen\\_app\\_on\\_Apple\\_Watch\\_October\\_2022.pdf](https://www.apple.com/healthcare/docs/site/Blood_Oxygen_app_on_Apple_Watch_October_2022.pdf) [Accessed 2026-03-02]
45. MacLeod DB, Cortinez LI, Keifer JC, et al. The desaturation response time of finger pulse oximeters during mild hypothermia. *Anaesthesia*. Jan 2005;60(1):65-71. [doi: [10.1111/j.1365-2044.2004.04033.x](https://doi.org/10.1111/j.1365-2044.2004.04033.x)] [Medline: [15601275](https://pubmed.ncbi.nlm.nih.gov/15601275/)]
46. Khairunnisa S, Soesanti I, Listyarifah D. Classification of finger pulse oximeter based on their response time using quantitative analysis. Presented at: The 5th International Conference on Bioinformatics, Biotechnology, and Biomedical Engineering (BioMIC 2023); Sep 6-7, 2023; Yogyakarta, Indonesia. 2023.[doi: [10.1051/bioconf/20237502002](https://doi.org/10.1051/bioconf/20237502002)]
47. Chronic obstructive pulmonary disease among adults aged 18 and over in the United States, 1998–2009. National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC); 2011. URL: <https://www.cdc.gov/nchs/data/databriefs/db63.pdf> [Accessed 2025-01-26]
48. Allen J, Murray A. Age-related changes in the characteristics of the photoplethysmographic pulse shape at various body sites. *Physiol Meas*. May 2003;24(2):297-307. [doi: [10.1088/0967-3334/24/2/306](https://doi.org/10.1088/0967-3334/24/2/306)] [Medline: [12812416](https://pubmed.ncbi.nlm.nih.gov/12812416/)]
49. Dewan NA, Nieto FJ, Somers VK. Intermittent hypoxemia and OSA: implications for comorbidities. *Chest*. Jan 2015;147(1):266-274. [doi: [10.1378/chest.14-0500](https://doi.org/10.1378/chest.14-0500)] [Medline: [25560865](https://pubmed.ncbi.nlm.nih.gov/25560865/)]
50. Iliaz S, Cagatay T, Bingol Z, et al. Does the 6-minute walk test predict nocturnal oxygen desaturation in patients with moderate to severe COPD? *Chron Respir Dis*. Feb 2015;12(1):61-68. [doi: [10.1177/1479972314562406](https://doi.org/10.1177/1479972314562406)] [Medline: [25480424](https://pubmed.ncbi.nlm.nih.gov/25480424/)]
51. Chaouat A, Weitzenblum E, Kessler R, et al. Sleep-related O<sub>2</sub> desaturation and daytime pulmonary haemodynamics in COPD patients with mild hypoxaemia. *Eur Respir J*. Aug 1997;10(8):1730-1735. [doi: [10.1183/09031936.97.10081730](https://doi.org/10.1183/09031936.97.10081730)] [Medline: [9272911](https://pubmed.ncbi.nlm.nih.gov/9272911/)]
52. Khondakar KR, Kaushik A. Role of wearable sensing technology to manage long COVID. *Biosensors (Basel)*. Dec 31, 2022;13(1):62. [doi: [10.3390/bios13010062](https://doi.org/10.3390/bios13010062)] [Medline: [36671900](https://pubmed.ncbi.nlm.nih.gov/36671900/)]

## Abbreviations

**ABG:** arterial blood gas

**A<sub>rms</sub>:** average root mean square

**COPD:** chronic obstructive pulmonary disease

**FDA:** US Food and Drug Administration

**ITA:** individual typology angle  
**MDE:** mean directional error  
**ODR:** oxygen desaturation rate  
**OSA:** obstructive sleep apnea  
**PAO<sub>2</sub>:** alveolar oxygen tension  
**PaO<sub>2</sub>:** partial pressure of oxygen in arterial blood  
**PetO<sub>2</sub>:** partial pressure of end-tidal oxygen  
**SaO<sub>2</sub>:** arterial oxygen saturation  
**SpO<sub>2</sub>:** peripheral oxygen saturation

*Edited by Amy Schwartz, Matthew Balcarras; peer-reviewed by Benjamin Smarr, Veronika Rafl-Huttova, William B Orr; submitted 04.Oct.2025; final revised version received 02.Jan.2026; accepted 13.Jan.2026; published 27.Mar.2026*

*Please cite as:*

*Jiang Y, Spies C, Roghanizad AR, Wang WK, Bhosai SJ, Snyder L, Burke A, MacLeod D, Dunn J  
Performance of Wearable Pulse Oximetry During Controlled Hypoxia Induction: Instrument Validation Study  
JMIR Form Res 2026;10:e85253  
URL: <https://formative.jmir.org/2026/1/e85253>  
doi: [10.2196/85253](https://doi.org/10.2196/85253)*

© Yihang Jiang, Connor Spies, Ali R Roghanizad, Will Ke Wang, Satasuk Joy Bhosai, Laurie Snyder, Ashley Burke, David MacLeod, Jessilyn Dunn. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 27.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.