

1 Performance of Wearable Pulse Oximetry During Controlled Hypoxia 2 Induction

3

4 Authors:

5 Yihang Jiang¹, Connor Spies³, Ali R. Roghanizad¹, PhD, Satasuk Joy Bhosai^{2,3}, MD, MPH,
6 Laurie Snyder^{2,3}, MD, MHS, Ashley Burke⁴, David MacLeod, MB BS, FRCA⁴, Jessilyn Dunn^{1,5}
7 PhD

8

9 * Corresponding Author

10 Email: jessilyn.dunn@duke.edu

11

12 1. Department of Biomedical Engineering, Duke University, Durham, NC, USA

13 2. Duke Clinical Research Institute, Durham, NC, USA

14 3. Department of Medicine, Duke University, Durham, NC, USA

15 4. Human Pharmacology & Physiology Lab, Duke University, Durham, NC, USA

16 5. Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA

17

18 **Abstract**

19 Oxygen saturation is a crucial metric used for monitoring patients with lung disease or other
20 conditions who are at risk of hypoxemia. Recently, consumer wearable devices began
21 incorporating oxygen saturation measurement tools, widening their use among individuals with
22 lung disease. However, their performance has not been well characterized, particularly during
23 blood oxygen desaturation, due to limitations in collecting comparison arterial blood gas
24 measurements during exercise-induced hypoxemia. To address this knowledge gap, we
25 performed a controlled oxygen desaturation study in nine healthy participants to mimic exercise-
26 induced hypoxemia, reducing the blood oxygen saturation levels from 100% to 60% using a gas
27 delivery system. We conducted a comparative analysis of the displayed oxygen saturation
28 readings of both the Masimo MightySat Rx finger pulse oximeter and Apple Watch Series 7 with
29 arterial blood oxygen saturation readings obtained from arterial blood gas samples. Both the
30 Masimo MightySat Rx pulse oximeter and the Apple Watch Series 7 tended to overestimate the
31 oxygen saturation. The Masimo pulse oximeter readings were more likely to fall within the
32 acceptable error range than the Apple Watch (49.03% compared to 32.14%). Notably, both
33 devices have limitations under low oxygen saturation levels (<88%) with an error of 1.96% and
34 4.99% for the Masimo and Apple Watch, respectively. Among the oxygen desaturation
35 measurements at a desaturation rate larger than two percent per minute, a rate which is clinically
36 correlated to sleep apnea, both devices increased their error slightly by 0.29% and 1.80%
37 respectively for the Masimo and Apple Watch. However, no statistically significant increase was
38 found.

39 **Author Summary**

40 Blood oxygen saturation (SpO₂) is an important measurement for monitoring patients with acute
41 and chronic conditions that are associated with hypoxemia, including chronic obstructive
42 pulmonary disease (COPD), heart failure, and sleep apnea. While smartwatches may provide a
43 novel method for continuous and unobtrusive SpO₂ monitoring, it is necessary to understand their
44 accuracy and limitations to ensure that they are used in a fit-for-purpose manner, especially under
45 conditions of low oxygen saturation.

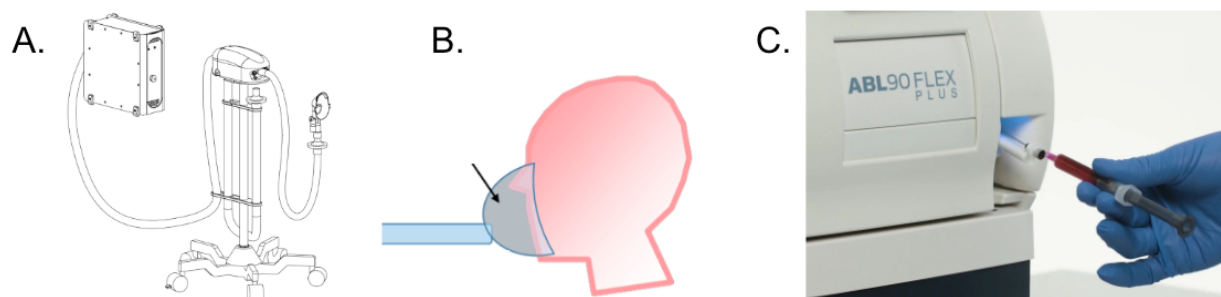
46 Introduction

47 Oxygen saturation represents how much hemoglobin is bound to oxygen. Measurement of oxygen
48 saturation is considered the fifth vital sign to gain insight into a person's health status alongside
49 body temperature, blood pressure, pulse rate and respiratory rate(1). Hypoxemia, often defined
50 as low oxygen saturation below 90%, is associated with symptoms such as shortness of breath,
51 rapid breathing, and fast heartbeat. Early identification of an abnormal oxygen saturation values
52 (2) is important in chronic obstructive pulmonary disease (COPD) patients where lung disease
53 can lead to hypoxemia symptoms and lower oxygen saturation during exercise, resting or sleep
54 (3–5). In COPD, oxygen desaturation during exercise can be caused by multiple factors such as
55 ventilation-perfusion mismatching, diffusion-type limitation, and shunting(6). In a study with 60
56 normoxic moderate-to-stable COPD patients, 33 patients (55%) desaturated during a 6-minute-
57 walk-test (6MWT), which is a standardized self-paced walking test that assesses the
58 cardiovascular response during daily activities(7). For COVID-19 patients with similar exercise-
59 induced desaturation who underwent a lung ultrasound scan, a significantly higher lung
60 ultrasound score was found as compared with patients who didn't have oxygen desaturation
61 larger than 4% indicating worsening odds of poor outcomes as a result of COVID-19 infection(8,9).
62

63 Measurement of oxygen saturation can be classified into two methods: invasive and non-
64 invasive(10). Arterial blood gas (ABG) analysis is an invasive oxygen saturation measurement
65 method that requires the collection of arterial blood samples from an artery and measurement of
66 arterial oxygen saturation (SaO_2) by a dedicated co-oximeter machine. ABG analysis is very
67 accurate and considered the 'gold standard', but ABG sample collection is painful and impractical
68 for continuous monitoring(11,12). Pulse oximetry is based on the principle that the optical
69 absorption of red and near-infrared light by oxyhemoglobin and deoxyhemoglobin differ and allow
70 the non-invasive estimate of oxygen saturation (SpO_2)(13). SpO_2 is considered a reliable
71 surrogate for SaO_2 , meeting the increasing need for home monitoring. Wearables devices have
72 integrated technology to measure and display oxygen saturation, SpO_2 , which helped the global
73 market value of smartwatches reach \$33.5 billion in 2022(14). However, there is concern that
74 there may be a limited accuracy of smartwatches to detect and identify low oxygen saturation,
75 and that such failures may a result of technical errors(15) or human factors(16). A lack of reliability
76 for low oxygen saturation detection would increase the risks for patients with hypoxemia (16–18).

77
78 Multiple research studies have used SaO_2 as the gold standard measure of oxygen saturation
79 and have focused on the bias of SaO_2 from pulse oximeters (SpO_2) or smartwatches without
80 investigating the accuracy of pulse oximeters during a continuous oxygen desaturation
81 process(19,20). Other research studies that used SpO_2 or SaO_2 as the reference standard during
82 a controlled desaturation process did not discuss the relationship between the accuracy and
83 desaturation rate(21,22). Our work(23) addresses this gap by analytically validating that pulse
84 oximeters can accurately respond to a drop in SaO_2 during a continuous oxygen desaturation
85 protocol. We hypothesized that the accuracy of pulse oximeters and wearables would differ under
86 hypoxemia and would also be affected by the oxygen desaturation rate. We employed a controlled
87 gas delivery system to conduct the controlled desaturation protocol and change the SaO_2 by
88 changing the setting of end-tidal oxygen (PetO_2) that participants breathe in (Fig 1). Oxygen

89 saturation readings from the Apple Watch Series 7, acknowledged for its superior best
90 performance among a selection of four commercial wearables(24,25), and Masimo MightySat Rx,
91 an FDA-cleared pulse oximeter, were compared against the gold standard of ABG analysis.



92
93 **Fig 1.** Flow chart depicting the desaturation protocol tools for adjusting $P_{et}O_2$ and measuring
94 SaO_2 . By adjusting the end-tidal $P_{et}O_2$ setting in the RespirAct RA-MR gas delivery system (A),
95 the study participant will inhale gas with different oxygen, carbon dioxide and nitrogen
96 concentrations through a fitted mask (B). Blood samples that were taken once the $P_{et}O_2$ was
97 stable were run through the Radiometer ABL90 Flex blood gas analyzer (C) to measure the
98 resulting SaO_2 and PaO_2 .
99

100 Results

101 Study Population

102 Nine individuals, 5 males and 4 females, provided written informed consent to Duke IRB protocol
103 Pro00105579 and completed the study. The median age was 25 years (range 19 - 28) and the
104 median Body Mass Index (BMI) was 24.4 kg/m² (range, 21.8-35.1 kg/m²). Participants ranged
105 from Type II to Type VI on the Fitzpatrick skin tone scale (Table 1). On the ITA skin tone scale
106 which defines ITA value larger than 50 degree as very light skin tone and ITA value smaller than
107 -50 degree as very dark skin tone(26), participants ranged from -27° to 54°, with a median value
108 of 7°.

109 **Table 1.** Patient demographics and characteristics

	Type II	Type III	Type IV	Type V	Type VI

Mean Age (Years)	22	19	24	25.25	27
Total Number (N)	2	1	1	4	1
Mean BMI	25.5	24	21.8	24.53	35.1
Gender (Counts)	Male(0)/ Female(2)	Male(1) / Female (0)	Male(0) / Female (1)	Male(3) / Female(1)	Male(1) / Female (0)
Mean ITA (degree)	+46.5	+23	+12	0	-27

110

111 Desaturation Protocol and ABG Analysis

112 The subjects underwent a controlled, stepwise descent of oxygen saturation from 100% to a target
113 of 60%. This was achieved by a dedicated gas delivery system, RespirAct (Thornhill Research,
114 Toronto, Canada) to target pre-specified end-tidal oxygen tensions, $P_{et}O_2$. In turn, the arterial
115 partial pressure of oxygen saturation ($P_{a}O_2$) determines the arterial oxygen saturation ($S_{a}O_2$),
116 which is measured by ABG and estimated non-invasively ($S_{p}O_2$) by the pulse oximeter and
117 smartwatch devices. Under stable $P_{et}O_2$ conditions, we obtained serial ABG samples to measure
118 $S_{a}O_2$.

119

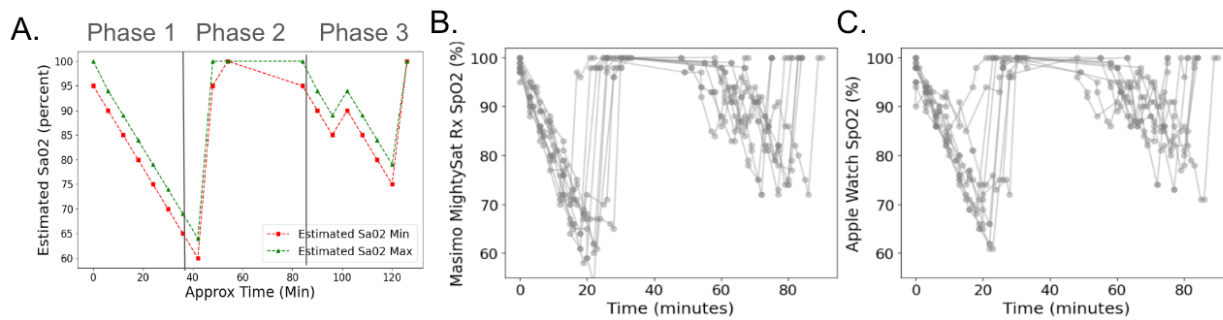
120 Measurements were taken across the 9 study participants at 18 targeted $P_{et}O_2$ levels ranging
121 from 32 to 250 mmHg, representing $S_{a}O_2$ from 60 to 100%. There were 308 $S_{a}O_2$ measurements
122 from the ABG analysis.

123

124 Device Performance

125 Three hundred and five out of 308 (99%) and 284 out of 308 (92%) possible $S_{p}O_2$ measurements
126 were taken from Masimo MightySat Rx and Apple Watch Series 7, respectively. In other words,
127 only 3 measurements were unable to be taken by the Masimo device, but 24 measurements could

128 not be taken by the Apple Watch. The maximum SpO₂ measurement from Masimo MightySat Rx
129 was 100 percent, and all participants had SpO₂ measurements that reached 100 percent.
130 Likewise, the maximum SpO₂ measurements from Apple Watch Series 7 was 100 percent and all
131 participants had SpO₂ measurements that reached 100 percent. The minimum SpO₂
132 measurement from Masimo MightySat Rx was 53 percent, and minimum SpO₂ measurements
133 were below 65 percent for 7 out of 9 (78%) participants. However, the minimum SpO₂
134 measurement from Apple Watch Series 7 was 61 percent, and the minimum SpO₂ measurements
135 were below 65 percent for only 2 out of 9 (22%) participants. The estimated SaO₂ controlled by
136 the gas delivery system is shown in Fig 2A and the SpO₂ measurements of the two devices during
137 the desaturation protocols are shown in Fig 2B and Fig 2C, where it can be seen that the Apple
138 Watch measurements do not reach the same minimum values that the Masimo measurements
139 do.



140
141 **Fig 2. (A)** Estimated SaO₂ range using PetO₂ setting and Hemoglobin Dissociation Curve. The
142 green / red dots are the maximum / minimum estimated SaO₂. There are three phases
143 separated using solid black line: First oxygen desaturation sequence, oxygen saturation
144 recovery and second oxygen desaturation sequence. **(B and C)** SpO₂ measurements from the
145 Masimo MightySat Rx (B) and the Apple Watch Series 7 (C) during the oxygen desaturation
146 protocol (N=9 study participants).

147

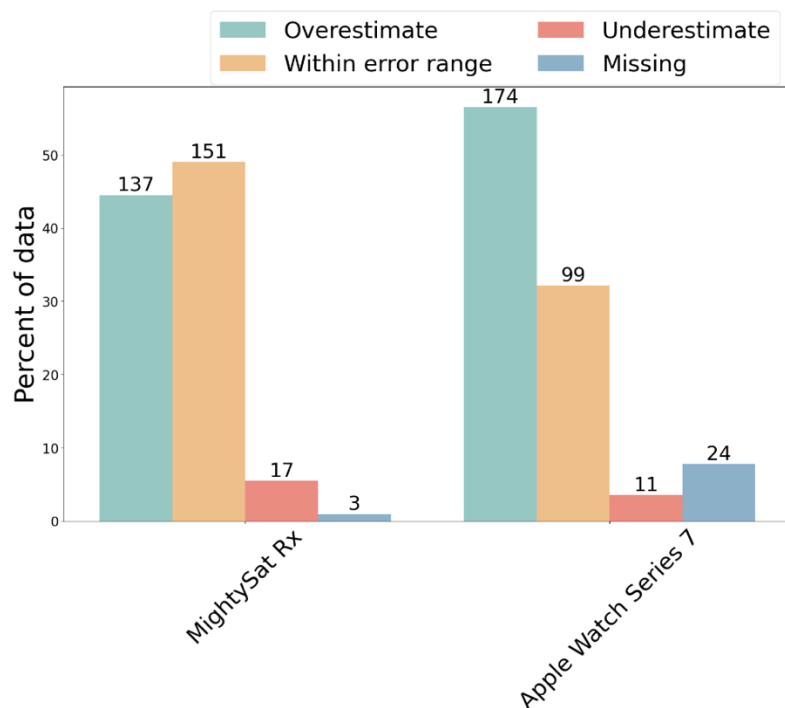
148

149 **Missingness and accuracy of measurements**

150

151 Each measurement from Masimo MightySat Rx and Apple Watch Series 7 was compared to the
152 arterial blood gas (ABG) measurement to determine whether it fell within, above, or below the

153 ABG measurement range. The percentage of measurements falling into these different categories
154 demonstrates whether there exists a trend of over- or underestimation of SpO₂ levels. Both the
155 MightySat Rx and the Apple Watch Series 7 tended to overestimate SpO₂ (Fig 3, Table 2). The
156 Apple Watch Series 7 had a higher percentage of overestimated measurements (56.49% of SpO₂
157 measurements were overestimated) as compared with the MightySat Rx (where 44.48% of
158 measurements were overestimated). The Masimo MightySat Rx had a higher percentage of
159 measurements (49.03%) falling within the accuracy range of the reference device compared to
160 Apple Watch Series 7 (32.14%). Ultimately, both devices have a problem of overestimating SpO₂,
161 and the percentage of measurements that were overestimated was even higher than the
162 percentage of measurements within the acceptable error range. In terms of missingness, the
163 Apple Watch Series 7 had more missingness (7.79%) compared to the MightySat Rx (0.97%),
164 indicating that both have high likelihood (>90%) for successfully obtaining a measurement when
165 a measurement was attempted. The Apple Watch Series 7's number of missing measurements
166 (24 missed out of 308) was higher than the Masimo MightySat Rx which only had 3 missing
167 measurements among 308 attempts.



168
169 **Fig 3.** Relative percentages of data falling into the categories of overestimated measurements
170 (red), measurements within acceptable error ranges (blue), underestimated measurements
171 (green), and missing measurements (grey) for the Masimo MightySat Rx and Apple Watch Series
172 7.

173
174 **Table 2.** Percentage and number of measurements falling in four categories

	No. (%) of overestimated SpO ₂ values	No. (%) of within error range	No. (%) of underestimated SpO ₂ values	No. (%) of missed measurements
Apple Watch Series 7	137 (44)	151 (49)	17 (6)	3 (1)
MightySat Rx	174 (56)	99 (32)	11 (4)	24 (8)

177
178
179 A Bland–Altman analysis demonstrated the difference between the SpO₂ measurements from the
180 Apple Watch and Masimo MightySat Rx and reference SaO₂ measurements (Fig 4A, Fig 4B). The
181 MDE of SpO₂ for the Masimo MightySat Rx and the Apple Watch was 1.80% and 3.26%,
182 respectively. This analysis demonstrates again the trend of overestimation for both devices and

183 that the severity of overestimation by the Apple Watch Series 7 was higher than the
184 overestimation by the Masimo MightySat Rx.

185
186

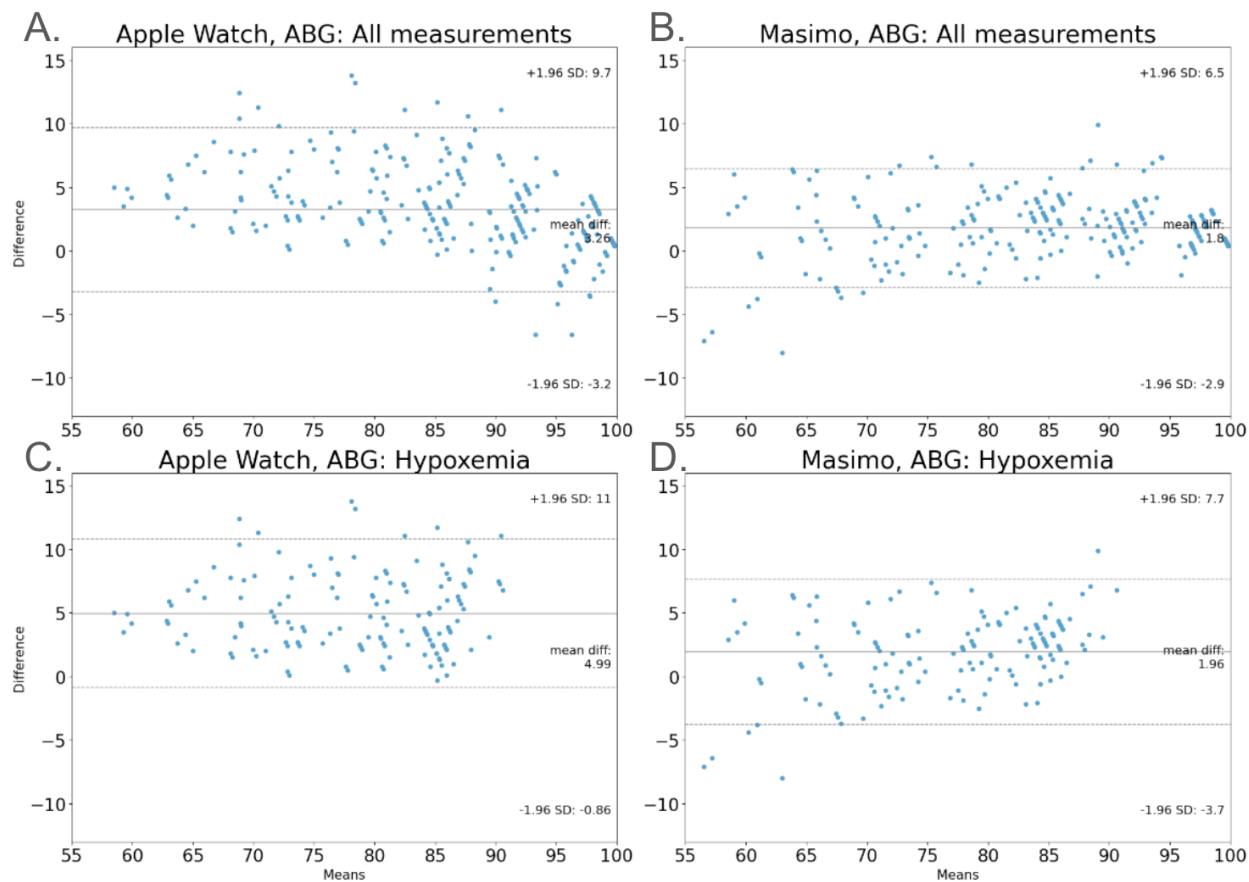
187 **Accuracy of measurements under hypoxemia**

188

189 Among the 308 measurements in the ABG analysis, 165 SaO₂ measurements (54%) are below
190 88 percent, which is the clinical threshold for hypoxemia(27). Among these 165 valid SaO₂
191 measurements, 164 (99%) and 140 (85%) SpO₂ measurements were collected simultaneously
192 from the Masimo MightySat Rx and Apple Watch Series 7 respectively.

193

194 A Bland–Altman analysis demonstrated the difference between the SpO₂ measurements from
195 each the Apple Watch and the Masimo MightySat Rx against ABG SaO₂ measurements under
196 hypoxemia. (Fig 4C, Fig 4D) The MDE of SpO₂ for the Masimo MightySat Rx and the Apple Watch
197 were 1.96% and 4.99%, respectively. This analysis demonstrates that the severity of the
198 overestimation increases under hypoxemia. Also, it demonstrates that the device has lower
199 accuracy during hypoxemia compared to normal blood oxygen conditions.



200

201 **Fig 4.** Bland-Altman plots demonstrate the differences between all simultaneous (A) SpO₂
202 measurements of the smartwatch (Apple Watch Series 7), (B) oximeter (Masimo MightySat Rx)
203 and SaO₂ measurements from ABG analysis. Under hypoxemia (SaO₂ < 88 %), it also
204 demonstrates simultaneous SpO₂ measurements of the (C) smartwatch (Apple Watch Series 7),
205 (D) oximeter (Masimo MightySat Rx) and SaO₂ measurements from ABG analysis. The solid line
206 is the mean difference of the measurements (MDE). Dashed lines are the 95% limits of agreement.

207

208

209

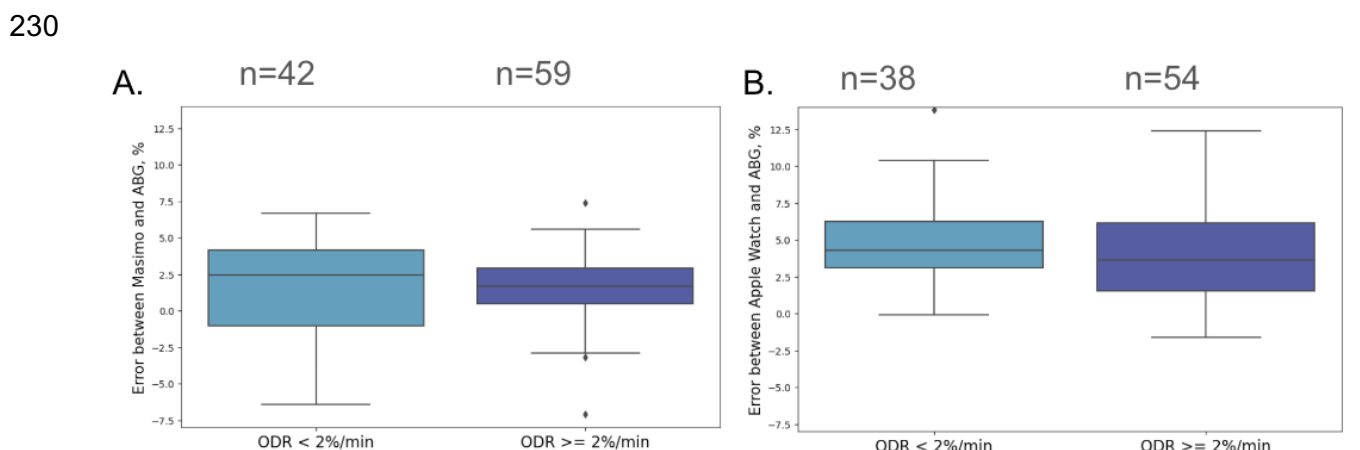
210 Accuracy of measurements during desaturation

211

212 Among the 308 measurements in the ABG analysis, 102 SaO₂ measurements are collected during
213 the first oxygen desaturation sequence (Fig 2A, Phase 1) which consisted of 8 stable desaturation
214 steps that were evenly spaced in time (see methods). Among these 102 valid SaO₂
215 measurements, 101 and 92 SpO₂ measurements were collected simultaneously from Masimo
216 MightySat Rx and Apple Watch Series 7 respectively. Among the 101 and 92 SpO₂ measurements,

217 42 out of 101 and 38 out of 92 measurements are taken when the oxygen desaturation rate (ODR)
218 is larger than 2 percent per minute.

219
220 Welch's t-test models was used to explore whether a statistically significant difference exists
221 between measurements with an oxygen desaturation rate above or below two percent per
222 minute—equivalent to the 4% oxygen desaturation index (ODI-4%) which is defined as the number
223 of events with a more than 4% decrease in SpO₂ within 120 seconds—has been found related to
224 sleep apnea in multiple studies (28,29). When the ODR is larger than 2 percent per minute, the
225 MDE for the Masimo MightySat Rx increases from 1.60% to 1.89% and for the Apple Watch
226 increases from 2.14% to 3.94%. (Fig 5) No significant correlation between ODR and the error of
227 the Masimo MightySat Rx and Apple Watch Series 7 were found (p-values = 0.64 and 0.17,
228 respectively), indicating that, while there is a trend of error increasing with ODR, this trend is not
229 statistically significant.



232 Fig 5. Boxplots demonstrate the different errors of pulse oximetry between measurements with
233 oxygen desaturation rate < 2%/min and >= 2%/min (A) SpO₂ measurements of the Masimo
234 MightySat Rx (B) SpO₂ measurements of the smartwatch (Apple Watch Series 7)

235

236 **Materials and Methods**

237 **Study Population**

238 This study was conducted under the approval of the Duke University Health System (DUHS)
239 Institutional Review Board and all participants signed informed consent before participating
240 (Pro00105579 and Pro00110458). The criteria for study inclusion were healthy volunteers at least
241 18 years old. Exclusion criteria included individuals with peripheral vascular disease, Raynaud's
242 syndrome, cryoglobulinemia or any collagen vascular disease affecting the fingers, a history of
243 blood clots in the last 6 months, an essential tremor, or gel nail polish or any other non-natural,
244 non-removable discoloration of the forefinger.

245

246 **Devices and Data Collection**

247 This study involved two sequential, stepwise oxygen desaturation sequences. In the first
248 sequence the oxygen saturation (SaO_2) was reduced in stepwise fashion from ~100% to ~60%,
249 and then back to ~100% over the course of approximately 60 minutes. The second sequence
250 involved decreasing participants' SaO_2 from ~100% to ~80%, and then back to ~100% over the
251 course of 48 minutes. During the oxygen desaturation sequences, the displayed pulse oximeter
252 saturation (SpO_2) of two devices (the Apple Watch Series 7 and the Masimo MightySat Rx) were
253 recorded at the time of arterial blood draws. Prior to beginning the desaturation protocol, skin tone
254 of participants was measured using an objective colorimeter device (the Delfin SkinColorCatch)
255 as well as visually by the study team using the Fitzpatrick scale (FP). The RespirAct RA-MR and
256 the Radiometer ABL90 Flex were used respectively to alter and measure the SaO_2 . The RespirAct
257 RA-MR is a sequential gas delivery system used to control oxygen saturation and provide steady-
258 state plateaus prior to the blood gas sampling(30). The blood gas samples were then analyzed
259 by the Radiometer ABL90 Flex co-oximeter to yield the arterial oxygen saturation (SaO_2),

260 considered to be the 'gold standard' measurement. Once a steady state of PetO₂ and PetCO₂ was
261 achieved by the gas delivery system, arterial samples were drawn and analyzed using the
262 Radiometer ABL90 Flex.

263

264 **Desaturation Protocol**

265 Two measurements of SpO₂ from each of the Apple Watch and Masimo MightySat were obtained
266 at each of the oxygenation plateaus (where PetO₂ was held at a steady state by the RespirAct
267 RA-MR) to coincide with the timing of the arterial blood gas samples. During this procedure,
268 participants were lying in the supine position on a standard hospital stretcher with an arterial
269 catheter placed in their forearms. The first oxygen desaturation sequence (Figure 2A, Phase 1)
270 involved 8 steps of approximately 6 minutes each with PetO₂ set to 90, 60, 50, 60, 50, 37, 34, and
271 32 mmHg and estimated maximum SaO₂ from Hemoglobin Dissociation Curve to be 100, 94, 89,
272 84, 79, 74, 69, and 64 percent. Following this sequence, PetO₂ will be set to 250 mmHg for
273 approximately 30 minutes to allow participants to recover their maximum oxygen saturation. The
274 second oxygen desaturation sequence consisted of 8 steps of approximately 6 minutes each with
275 PetO₂ set to 90, 60, 50, 60, 50, 45, 40 and 250 mmHg, and estimated maximum SaO₂ to be 100,
276 94, 89, 94, 89, 84, 79 and 100. The smartwatch was placed tightly on the right wrist, approximately
277 one centimeter above the ulnar styloid process. Measurements on the watch were manually
278 triggered when PetO₂ reached steady state, and if the watch failed to produce a measurement, it
279 was repositioned on the participant's wrist and another measurement was attempted. If the device
280 still failed to produce a measurement, no value was recorded for this specific time point and it was
281 labeled as a missing (unobtainable) observation. The Masimo MightySat Rx was placed on the
282 middle finger of the participant's right hand, and recorded continuously at the rate of one reading
283 per second during the study procedure.

284 **Metrics and statistical analysis**

285 To assess the effectiveness of the pulse oximetry devices, we compared each SpO₂
286 measurement from the Apple Watch and the Masimo MightySat Rx to the SaO₂ measurements,
287 determining if they fell within, above, or below the reference range (2%) which is considered the
288 usual error range of most finger tip oximeters(1), or if they were missing. The Bland-Altman
289 method was employed to assess the accuracy of SpO₂ measurements across all readings, with
290 separate analysis for readings where SaO₂ value was below 88%. The threshold of 88% was
291 chosen in line with the British Thoracic Society guidelines(27) indicating the point for intensive
292 therapy to elevate oxygen saturation. Mean directional error (MDE) (Eq 1) and missingness (Eq
293 2) were utilized as metrics to gauge device performance. To evaluate device performance during
294 oxygen desaturation, we focused on observations in the initial oxygen desaturation phase, where
295 SaO₂ settings transitioned uniformly from 100 percent to 60 percent. We utilized a linear mixed
296 effects model to examine the impact of fixed effects from oxygen desaturation rate and random
297 effects from individual participants on the difference between SpO₂ and SaO₂ measurements.

$$\text{MDE (Mean directional error)} = \frac{\sum(\text{Reference SpO}_2 - \text{Detected SpO}_2)}{\text{Number of valid measurements}} \quad (1)$$

$$\text{Missingness(\%)} = 100 - \left(\frac{\text{Valid measurements}}{\text{Total measurements}} \right) * 100 \quad (2)$$

298

299

300

301 **Discussion**

302 Wearables may be used for continuous measurements of oxygen saturation (SpO₂) for COPD
303 patients at home. To evaluate the performance of these devices under hypoxemia, we applied
304 two oxygen desaturation sequences on nine subjects with controlled, stepwise descent of

305 oxygen saturation from 100% to a target of 60%. We used the clinically important threshold of
306 88%, a decision point for administering supplemental oxygen, to evaluate MDE using Bland-
307 Altman analysis. We compared the readings from the Masimo pulse oximeter and the Apple
308 Watch Series 7 with 'gold standard' ABG readings under both low and normal oxygen saturation
309 levels.

310
311 How best to evaluate the performance of pulse oximeters under hypoxemia remains an
312 unsolved problem. In the US FDA's guidance on pulse oximetry FDA assessment(31), it is
313 stated that FDA-cleared pulse oximeters will have the lowest accuracy below 80%, but there is
314 no standard to evaluate the performance during hypoxemia with rapid oxygen saturation value
315 changes. Calculating the average bias between the 'gold standard' SaO₂ and device-measured
316 SpO₂ over all values of blood oxygen saturation is insufficient to understand a tool's
317 performance as that performance may change at lower values or during rapid change(17).

318 Thus, evaluating pulse oximeters during hypoxemia requires a particularly challenging form of
319 clinical validation (e.g., as defined in the V3 framework), involving observing decreases in
320 oxygen saturation over both longer timescales (e.g., on the order of hours or days, which is
321 more common in this field), as well as during rapid change (e.g. on the order of minutes). An
322 example of a longer timescale clinical validation study occurred in COPD patients for one week;
323 however, SpO₂ was used as the reference standard instead of SaO₂ because the study was
324 performed in an ambulatory setting, which introduced potential inaccuracies in the reference
325 standard(32). We recommend that, in the evaluation of new tools for measuring blood oxygen
326 saturation, it is important to 1) carefully evaluate the performance of the device under
327 hypoxemic conditions at a variety of levels of low blood oxygenation, 2) specifically explore
328 performance during hypoxemia through oxygen desaturation sequences over shorter timescales
329 to determine how well the device can capture rapid changes in blood oxygen saturation, 3) use
330 arterial blood gas measurements as the gold standard measurement against which to determine

331 performance of the device, and 4) explore the accuracy of pulse oximeters during hypoxemia on
332 people with different skin tones since it is reported black patients had nearly three times the
333 frequency of occult hypoxemia (low oxygen levels in the blood) as detected by blood gas
334 measurements but not detected by pulse oximetry(33).

335

336 Limitations

337 The main limitation of this study is the small sample size, leading to insufficient representation
338 across skin tone categories. Three out of the five Fitzpatrick skin tone categories only have one
339 participant each, making it impossible to generalize our results to people with similar skin tones.
340 Our study used wide coverage of oxygen saturation values since the target SpO₂ ranged from
341 60% to 100%. However, this protocol cannot demonstrate SpO₂ measurement performance
342 over long periods since each participant only spent 60 minutes in the first oxygen desaturation
343 sequence and 48 minutes in the second one. It is necessary to consistently monitor the oxygen
344 saturation value overnight because nocturnal hypoxemia commonly occurs among COPD
345 patients(34,35). It would be infeasible to induce oxygen desaturation for such a long period in
346 healthy adults, which is why it is necessary to perform clinical validation among COPD patients
347 with nocturnal desaturation, defined as having a SpO₂ value below 90% for more than 30% of
348 the time in bed during one or more nights(36,37). Future studies should recruit more subjects
349 with at least 15% of darkly pigmented subjects guided by FDA(31) and prolong the monitoring
350 process for the nocturnal desaturation caused by COPD or sleep apnea to verify the accuracy
351 and variability of wearables.

352

353 Acknowledgements

354 The content is solely the responsibility of the authors and does not necessarily represent the
355 official views of AstraZeneca or the National Institutes of Health.

356
357

358 References

- 359 1. Pulse Oximetry > Fact Sheets > Yale Medicine [Internet]. [cited 2023 Dec 12]. Available
360 from: <https://www.yalemedicine.org/conditions/pulse-oximetry>
- 361 2. Vold ML, Aasebø U, Wilsgaard T, Melbye H. Low oxygen saturation and mortality in an adult
362 cohort: the Tromsø study. *BMC Pulm Med*. 2015 Feb 12;15:9.
- 363 3. Davis J. WebMD. [cited 2023 Dec 12]. What Is COPD Hypoxia? Available from:
364 <https://www.webmd.com/lung/copd/copd-hypoxia>
- 365 4. Knowler MT, Dunagan DP, Adair NE, Chin J Robert. Baseline Oxygen Saturation Predicts
366 Exercise Desaturation Below Prescription Threshold in Patients With Chronic Obstructive
367 Pulmonary Disease. *Archives of Internal Medicine*. 2001 Mar 12;161(5):732–6.
- 368 5. Little SA, Elkholy MM, Chalmers GW, Farouk A, Patel KR, Thomson NC. Predictors of
369 nocturnal oxygen desaturation in patients with COPD. *Respiratory Medicine*. 1999 Mar
370 1;93(3):202–7.
- 371 6. Full article: Exertional Desaturation in Patients with Chronic Obstructive Pulmonary Disease
372 [Internet]. [cited 2024 May 16]. Available from:
373 <https://www.tandfonline.com/doi/full/10.3109/15412550903341497#2b85d6ca-6520-4a3d-8e4a-aa9f2ee3f33d-b6de7b7c-de82-45a5-9538-313dd15c6659>
374
- 375 7. Dogra AC, Gupta U, Sarkar M, Padam A. Exercise-induced desaturation in patients with
376 chronic obstructive pulmonary disease on six-minute walk test. *Lung India*. 2015;32(4):320–
377 5.
- 378 8. Carlucci A, Paneroni M, Carotenuto M, Bertella E, Cirio S, Gandolfo A, et al. Prevalence of
379 exercise-induced oxygen desaturation after recovery from SARS-CoV-2 pneumonia and use
380 of lung ultrasound to predict need for pulmonary rehabilitation. *Pulmonol*. 2023 Dec
381 1;29:S4–8.
- 382 9. Perrone T, Soldati G, Padovini L, Fiengo A, Lettieri G, Sabatini U, et al. A New Lung
383 Ultrasound Protocol Able to Predict Worsening in Patients Affected by Severe Acute
384 Respiratory Syndrome Coronavirus 2 Pneumonia. *Journal of Ultrasound in Medicine*. 2021
385 Aug;40(8):1627–35.

- 386 10. Collins JA, Rudenski A, Gibson J, Howard L, O'Driscoll R. Relating oxygen partial pressure,
387 saturation and content: the haemoglobin–oxygen dissociation curve. *Breathe* (Sheff). 2015
388 Sep;11(3):194–201.
- 389 11. Diagnostic accuracy of venous blood gases compared to arterial blood gases | European
390 Respiratory Society [Internet]. [cited 2023 Dec 12]. Available from:
391 https://erj.ersjournals.com/content/42/Suppl_57/P3987
- 392 12. Plüddemann A, Thompson M, Heneghan C, Price C. Pulse oximetry in primary care:
393 primary care diagnostic technology update. *Br J Gen Pract*. 2011 May;61(586):358–9.
- 394 13. Chan ED, Chan MM, Chan MM. Pulse oximetry: Understanding its basic principles
395 facilitates appreciation of its limitations. *Respiratory Medicine*. 2013 Jun;107(6):789–99.
- 396 14. Global Market Insights Inc. [Internet]. [cited 2023 Dec 12]. Smartwatch Market Size & Share
397 – Analysis Report, 2032. Available from: [https://www.gminsights.com/industry-](https://www.gminsights.com/industry-analysis/smartwatch-market)
398 [analysis/smartwatch-market](https://www.gminsights.com/industry-analysis/smartwatch-market)
- 399 15. An assessment of the accuracy of pulse oximeters - Milner - 2012 - Anaesthesia - Wiley
400 Online Library [Internet]. [cited 2023 Dec 12]. Available from:
401 [https://associationofanaesthetists-](https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2011.07021.x)
402 [publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2011.07021.x](https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2011.07021.x)
- 403 16. Wilson BJ, Cowan HJ, Lord JA, Zuege DJ, Zygun DA. The accuracy of pulse oximetry in
404 emergency department patients with severe sepsis and septic shock: a retrospective cohort
405 study. *BMC Emerg Med*. 2010 May 5;10:9.
- 406 17. Accuracy of Pulse Oximeters in Detecting Hypoxemia in Patients with Chronic
407 Thromboembolic Pulmonary Hypertension - PMC [Internet]. [cited 2023 Dec 12]. Available
408 from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4433184/>
- 409 18. Louie A, Feiner JR, Bickler PE, Rhodes L, Bernstein M, Lucero J. Four Types of Pulse
410 Oximeters Accurately Detect Hypoxia during Low Perfusion and Motion. *Anesthesiology*.
411 2018 Mar 1;128(3):520–30.
- 412 19. Rafi J, Bachman TE, Rafi-Huttova V, Walzel S, Rozanek M. Commercial smartwatch with
413 pulse oximeter detects short-time hypoxemia as well as standard medical-grade device:
414 Validation study. *Digit Health*. 2022 Oct 11;8:20552076221132127.
- 415 20. Santos M, Vollam S, Pimentel MA, Areia C, Young L, Roman C, et al. The Use of Wearable
416 Pulse Oximeters in the Prompt Detection of Hypoxemia and During Movement: Diagnostic
417 Accuracy Study. *Journal of Medical Internet Research*. 2022 Feb 15;24(2):e28890.
- 418 21. The desaturation response time of finger pulse oximeters during mild hypothermia* -
419 MacLeod - 2005 - Anaesthesia - Wiley Online Library [Internet]. [cited 2024 Jun 9]. Available
420 from: [https://associationofanaesthetists-](https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2004.04033.x)
421 [publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2004.04033.x](https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2044.2004.04033.x)
- 422 22. Pulse Oximeter Performance during Rapid Desaturation - PMC [Internet]. [cited 2024 Jun 9].
423 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9185462/>

- 424 23. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification,
425 analytical validation, and clinical validation (V3): the foundation of determining fit-for-
426 purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digit Med*. 2020 Dec;3(1):55.
- 427 24. Jiang Y, Spies C, Magin J, Bhosai SJ, Snyder L, Dunn J. Investigating the accuracy of blood
428 oxygen saturation measurements in common consumer smartwatches. *PLOS Digit Health*.
429 2023 Jul 12;2(7):e0000296.
- 430 25. Windisch P, Schröder C, Förster R, Cihoric N, Zwahlen DR. Accuracy of the Apple Watch
431 Oxygen Saturation Measurement in Adults: A Systematic Review. *Cureus*. 15(2):e35355.
- 432 26. Agarwal M, Anand S, Patro M, Gothi D. Early versus non-early desaturation during 6MWT in
433 COPD patients: A follow-up study. *Lung India*. 2023;40(3):235–41.
- 434 27. BTS guideline for emergency oxygen use in adult patients | Thorax [Internet]. [cited 2023
435 Dec 12]. Available from: https://thorax.bmj.com/content/63/Suppl_6/vi1
- 436 28. Jung H, Kim D, Lee W, Seo H, Seo J, Choi J, et al. Performance evaluation of a wrist-worn
437 reflectance pulse oximeter during sleep. *Sleep Health*. 2022 Oct 1;8(5):420–8.
- 438 29. Temirbekov D, Güneş S, Yazıcı ZM, Sayın İ. The Ignored Parameter in the Diagnosis of
439 Obstructive Sleep Apnea Syndrome: The Oxygen Desaturation Index. *Turk Arch*
440 *Otorhinolaryngol*. 2018 Mar;56(1):1–6.
- 441 30. Ikeda K, MacLeod DB, Grocott HP, Moretti EW, Ames W, Vacchiano C. The Accuracy of a
442 Near-Infrared Spectroscopy Cerebral Oximetry Device and Its Potential Value for Estimating
443 Jugular Venous Oxygen Saturation. *Anesth Analg*. 2014 Dec;119(6):1381–92.
- 444 31. Health C for D and R. Pulse Oximeter Accuracy and Limitations: FDA Safety
445 Communication. FDA [Internet]. 2023 Nov 17 [cited 2023 Dec 12]; Available from:
446 [https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-](https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication)
447 [limitations-fda-safety-communication](https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication)
- 448 32. Buekers J, Theunis J, De Boever P, Vaes AW, Koopman M, Janssen EV, et al. Wearable
449 Finger Pulse Oximetry for Continuous Oxygen Saturation Measurements During Daily Home
450 Routines of Patients With Chronic Obstructive Pulmonary Disease (COPD) Over One Week:
451 Observational Study. *JMIR Mhealth Uhealth*. 2019 06;7(6):e12866.
- 452 33. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry
453 Measurement. *N Engl J Med*. 2020 Dec 17;383(25):2477–8.
- 454 34. Dewan NA, Nieto FJ, Somers VK. Intermittent Hypoxemia and OSA. *Chest*. 2015
455 Jan;147(1):266–74.
- 456 35. Oxygen saturation during daily activities in chronic obstructive pulmonary disease |
457 European Respiratory Society [Internet]. [cited 2023 Dec 20]. Available from:
458 <https://erj.ersjournals.com/content/9/12/2584.long>
- 459 36. Iliaz S, Cagatay T, Bingol Z, Okumus G, Iliaz R, Kuran G, et al. Does the 6-minute walk test
460 predict nocturnal oxygen desaturation in patients with moderate to severe COPD? *Chron*
461 *Respir Dis*. 2015 Feb;12(1):61–8.

- 462 37. Chaouat A, Weitzenblum E, Kessler R, Charpentier C, Ehrhart M, Levi-Valensi P, et al.
463 Sleep-related O₂ desaturation and daytime pulmonary haemodynamics in COPD patients
464 with mild hypoxaemia. *Eur Respir J.* 1997 Aug;10(8):1730–5.

465