DOI:
10.3357/AMHP.6078.2023

Document Version
Peer reviewed version

Citation for published version (APA):
Hearn, E., Byford, J., Wolfe, C., Agyei, C., Hodkinson, P., Pollock, R., & Smith, T. (2023). Measuring Arterial Oxygen Saturation Using Wearable Devices Under Varying Conditions. *Aerospace medicine and human performance*, *94*(1), 42-47. https://doi.org/10.3357/AMHP.6078.2023

Aerospace Medicine and Human Performance

Measuring arterial oxygen saturation using wearable devices under varying conditions

--Manuscript Draft--

Manuscript Number:	AMHP6078R1
Full Title:	Measuring arterial oxygen saturation using wearable devices under varying conditions
Article Type:	Short Communication
Corresponding Author:	Thomas G. Smith, MBBS DAvMed DPhil FRCA FAsMA King's College London London, London UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	King's College London
Corresponding Author's Secondary Institution:	
First Author:	Eleanor L Hearn, MSc
First Author Secondary Information:	
Order of Authors:	Eleanor L Hearn, MSc
	Jack Byford, MSc
	Christopher Wolfe, MSc
	Cheryl Agyei, MSc
	Peter D Hodkinson, MRCP, PhD
	Ross D Pollock, PhD
	Thomas G Smith, DPhil, FRCA
Order of Authors Secondary Information:	

Measuring arterial oxygen saturation using wearable devices under varying conditions

Short title

Wearable devices and SpO_2

Authors and affiliations

Withheld as per journal instructions

Manuscript category

Short Communication

Manuscript metrics

Abstract word count:	248
Main text word count:	2,956
References:	13
Tables:	2
Figures:	2
Auxiliary information:	1 supplementary appendix

ABSTRACT

Introduction:

Recently developed wearable monitoring devices can provide arterial oxygen saturation (SpO_2) measurements, offering potential for use in aerospace operations. Pilots and passengers are already using these technologies, but their performance has not yet been established under conditions experienced in the flight environment such as environmental hypoxia and concurrent body motion.

Methods:

An initial evaluation was conducted in ten healthy subjects who were studied in a normobaric chamber during normoxia and at a simulated altitude of 15,000 ft (11.8% oxygen). SpO_2 was measured simultaneously using a standard pulse oximeter and four wearable devices: Apple Watch Series 6; Garmin Fēnix 6 watch; Cosinuss° Two in-ear sensor; and Oxitone 1000M wrist-worn pulse oximeter. Measurements were made while stationary at rest, during very slight body motion (induced by very low intensity cycling at 30W on an ergometer), and during moderate body motion (induced by moderate intensity cycling at 150W).

Results:

Missed readings, defined as failure to record an SpO_2 value within one minute, occurred commonly with all wearables. Even with only very slight body motion, most devices missed most readings (range of 12-82% missed readings), and the rate was higher with greater body motion (range 18-92%). One device tended to under-report SpO_2 while the other devices tended to over-report SpO_2. Performance decreased across the devices when oxygenation was reduced.

Discussion:

In this preliminary evaluation, the wearable devices studied did not perform to the same standard as a traditional pulse oximeter. These limitations may restrict their utility in-flight and require further investigation.

Keywords:

Pulse oximeter; pilot; hypoxemia; altitude; aviation; spaceflight

INTRODUCTION

With a global market size of approximately $40 billion in 2020, wearable technology is a growing industry with a broad impact that is likely to include the aerospace sector.[4] Wearable physiological monitoring devices, or ‘wearables’, are portable technologies that are intended to track physiological data such as calories burned, step count, heart rate and, more recently, arterial oxygen saturation (SpO_2). Owing to the accessibility and convenience of wearable technology, these devices have the potential to transform remote monitoring in patients at risk of hypoxemia, such as those with chronic obstructive pulmonary disease or COVID-19, and are marketed to consumers as a means of promoting health and wellbeing.

Aircrew are routinely exposed to mild-moderate hypoxia and, anecdotally, the use of wearables by pilots across general, commercial and military operations is increasing. Wearable measurements of in-flight SpO_2 are similarly appealing in other groups such as passengers, aeromedical patients and skydivers.[1] The ability to detect worsening hypoxemia during flight is highly desirable as it is dangerous and can develop for many reasons, such as reduced cabin pressure, unpressurised flight at high altitudes, pre-existing or acute illness, physical exertion (e.g. helicopter rear crew), high G acceleration, and failure of oxygen delivery and life-support systems. In recent years this has been particularly topical in the setting of military fast-jet operations due to the possible contribution of hypoxia to unexplained physiological events. However, it is important to establish the performance of new technologies prior to safety-critical use. With regards to isolated SpO_2 monitoring during flight, additional care is required as interpretation can be challenging or misleading even for accurate measurements, for example in the presence of hyperventilation.[2]

While the accuracy of heart rate data from wearables has been well-reported, the ability to measure SpO_2 is a newer feature and has not been comprehensively investigated.[8] Standard pulse oximeters used in medical practice utilise transmissive photoplethysmography (PPG), in which a light source and photodetector are located on opposite sides of a vascular bed (such as a finger or ear lobe) and the intensity of transmitted light of certain wavelengths is measured. The reliability of this technique is well established, but such devices tend to be somewhat obtrusive when used while performing other activities. In contrast, wearables are by their nature less obtrusive, but typically utilise the less established technique of reflective PPG, in which the light source and photodetector are positioned on the same side of a vascular bed, and the intensity of reflected light is measured.[12] Wearables are also designed for SpO_2 measurements to be made while completely stationary.

Recently developed wearables that can measure SpO_2 include consumer-grade products such as the Apple Watch 6 (Apple Inc, California, USA) and Garmin Fēnix 6 watch (Garmin Ltd, Kansas, USA), which are marketed for ‘general fitness and wellness purposes’ rather than for medical use. In contrast, the commercially available in-ear (‘hearable’) Cosinuss° Two (Cosinuss GmbH, Munich, Germany) has undergone testing in clinical settings, although comparative data has not been published and it is not currently classified as a medical device, while the wrist-worn Oxitone 1000M (Oxitone Medical, HaMerkaz, Israel) is an FDA-cleared medical monitor intended for clinical use. The Garmin and Apple watches and Cosinuss° Two use reflective PPG, while the Oxitone 1000M uses transmissive PPG. There is little published research reporting SpO_2 data from these devices. The Oxitone 1000M has been reported to provide accurate and precise SpO_2 values when measured in a stationary state,[5] while a recent study conducted in a respiratory outpatient clinic reported that the Apple Watch 6 appeared to be a reliable means of measuring SpO_2 in this controlled setting, although there were occasional outlying values.[10] An earlier Garmin watch model (the Fēnix 5X Plus) was found to over-estimate SpO_2 in volunteers studied in a normobaric chamber, especially at higher simulated altitudes, and it was noted that achieving a single measurement could take up to three minutes.[6] This highlights the potential for measurement failure to impact on performance – irrespective of its other qualities, a device that is unable to reliably achieve a timely reading is unlikely to be useful in the flight environment.

Although there is limited data and satisfactory performance cannot be assumed across the various technologies, these initial studies are generally encouraging with regards to use while stationary and under normoxic conditions. However, in-flight use does not necessarily allow such optimal conditions; achieving an absolutely motionless state can be challenging or impossible, and a lower range of SpO_2 may well be encountered. To our knowledge, no previous studies have investigated the potential combined effects of hypoxia and concurrent body motion of any degree. This initial study aimed to undertake a preliminary evaluation of four leading wearable devices in measuring SpO_2 under normoxic and hypoxic conditions while at rest and during relevant levels of body motion, including very minimal movement only marginally beyond a stationary state. The hypothesis was that their performance in measuring SpO_2 would be the same as that of a standard pulse oximeter. Our aim was to generate preliminary results and provide a basis for the definitive studies that are ultimately required.

METHODS

Subjects

This study was conducted in healthy volunteers and was approved by the King's College London Research Ethics Committee. It was conducted in accordance with the Declaration of Helsinki. All subjects provided written informed consent.

Equipment

The study was undertaken in a normobaric altitude chamber (Sporting Edge, UK) containing a cycle ergometer (Monark 818E, Monark Exercise, Vansbro, Sweden). Reference SpO_2 was measured continuously at the left index finger using a standard pulse oximeter (Pulse Oximeter 7840, Kontron Instruments Ltd, UK) recorded via PowerLab 8/35 and LabChart 8.0 (AD Instruments, Oxford, UK) and was compared with data from an Apple Watch 6 (at the left wrist), Garmin Fēnix 6 watch and Oxitone 1000M (at the right wrist) and a Cosinuss° Two (in the right ear). All wearables were attached and operated according to the manufacturer's instructions, and the Cosinuss° Two was fitted for size (small, medium or large). Simultaneous heart rate measurements were recorded from all monitors in parallel with SpO_2.

Procedure

Subjects attended the laboratory on two experimental days separated by a minimum of 24 hours. The protocol was identical on each occasion except that one day was conducted under normoxic conditions in room air (20.9% oxygen) and the other was conducted in hypoxic conditions at a simulated altitude of 15,000 ft (11.8% oxygen). This altitude was intended to extend nadir SpO_2 values into the 70-80% range. The order of normoxia and hypoxia was counter-balanced and subjects were blinded to each condition. Following instrumentation, subjects entered the hypoxia chamber and completed 10 minutes of seated rest. They then cycled on the ergometer for five-minute periods at very low intensity (30W) and at moderate intensity (150W) separated by five minutes of seated rest. These periods of cycling were intended as a reproducible means of inducing very slight body motion (30W) and moderate body motion (150W), with the added potential for exaggerating any hypoxemia.[13] Participants were instructed to remain otherwise still while cycling, and there was minimal associated motion of the arms and head, especially at 30W which requires only very gentle pedalling. A further five minutes of seated rest concluded testing. For each period of rest and cycling, measurements of SpO_2 and heart rate were recorded at three evenly-spaced time points. A maximum of one minute was allowed to obtain a reading from each device, after which a failed or ‘missed’ measurement was recorded.

Statistical Analysis

Data were normally distributed (Shapiro-Wilk test). The effect of hypoxia on SpO_2 and heart rate was analysed with paired t-tests (IBM SPSS Statistics v.26) using mean data for each period of rest or cycling (using SpO_2 and heart rate data obtained from the reference pulse oximeter). The accuracy and bias of measurements from the wearable devices were tested against the reference pulse oximeter using paired t-tests, Bland Altman analyses (GraphPad, Prism, v.26) and mean absolute percentage error (MAPE) score. MAPE was calculated using the following equation: ((actual value-forecast value)/actual value)*100. Statistical significance was assumed at $P < 0.05$ and data are presented as mean ± SD.

RESULTS

There were ten subjects (six men and four women) with mean age 27 ± 6 yr, weight 75 ± 15 kg, height 1.74 ± 0.11 m and body mass index 24 ± 3 kg/m^2. Fig 1 shows the effects of hypoxia and periods of cycling on the reference physiological data obtained using the standard pulse oximeter. SpO_2 was significantly lower during hypoxia at rest (82 ± 3% vs 98 ± 1%; t(29)=15.9, $P < 0.001$), during 30W cycling (76 ± 6% vs 98 ± 1%; t(9)=11.8, $P < 0.001$) and during 150W cycling (74 ± 7% vs 98 ± 1%; t(9)=12.2, $P < 0.001$). There was a small increase in heart rate during hypoxia compared with normoxia at rest (87 ± 14 bpm vs 75 ± 15 bpm; t(29)=6.4, $P < 0.001$) and similarly during 30W cycling (102 ± 13 bpm vs 91 ± 17 bpm; t(9)=3.4, $P = 0.008$) and 150W cycling (139 ± 14 bpm vs 127 ± 13 bpm; t(9)=2.7, $P = 0.026$).

[Figure 1 here]

Missed SpO_2 readings were common for all devices, with a progressive increase in the percentage of missed readings with increasing cycling intensity (Table I). At rest, the percentage of missed readings ranged between 2.5% and 20%, while during very low intensity cycling at 30W, when associated body motion was very minimal, most devices missed most readings (range 12-82%). During moderate intensity cycling at 150W, the percentage of missed readings ranged between 18% and 95%. Overall, the percentage of missed readings was lowest for the Cosinuss° Two and highest for the Oxitone 1000M. Mean absolute percentage error (MAPE) and percentage accuracy were calculated and are shown in Table I. With increasing cycling intensity, MAPE increased and percentage accuracy decreased. The Apple Watch 6 displayed the highest percentage accuracy independent of motion status, whilst the Garmin Fēnix 6 showed the lowest percentage accuracy. Equivalent data for heart rate is shown in Table II. Missed heart rate readings were generally less frequent, while overall, from rest to 150W cycling, MAPE increased and percentage accuracy decreased.

[Table I and Table II here]

Figure 2 shows all recorded SpO_2 data (at rest and while cycling) for each of the respective devices during normoxia and hypoxia. Under normoxic conditions, when values were successfully obtained the SpO_2 data from the Apple Watch 6 (t(4)=0.5898, P = 0.6) and Oxitone 1000M (t(4)=1.215, P = 0.3) were not significantly different from reference data obtained from the traditional pulse oximeter. However, SpO_2 readings from the Garmin Fēnix 6 (t(4)=4.867, P = 0.008) and Cosinuss° Two (t(4)=3.964, P = 0.017) were significantly different from the corresponding reference data. During hypoxia, the Cosinuss° Two (t(4)=0.3653, P = 0.7) was the only device to provide SpO_2 measurements that were not significantly different from the reference data; the Apple Watch 6 (t(4)=8.025, P = 0.001), Garmin Fēnix 6 (t(4)=4.094, P = 0.015) and Oxitone 1000M (t(4)=3.812, P = 0.019) data were significantly different from the reference data. Equivalent data for heart rate is shown in the supplementary online appendix (Figure A1).

[Figure 2 here]

Overall, when normoxic and hypoxic measurements were combined, the Apple Watch 6, Garmin Fēnix 6 and Oxitone 1000M all tended to over-report SpO_2 both at rest and while cycling, whilst the Cosinuss°Two tended to under-report SpO_2 (Figure A2 in the supplementary online appendix). Compared with the reference SpO_2 data, the Apple Watch 6 had the smallest mean bias (rest: 1.7 ± 2.1%; 30W cycling: 1.2 ± 3.4%; 150W cycling: 1.9 ± 2.3%), while the Cosinuss° Two had the largest mean bias (rest: -2.9 ± 3.0%; 30W: -1.5 ± 3.7%; 150W: -6.5 ± 5.2%). The Oxitone 1000M over-reported SpO_2 with a higher mean bias (rest: 2.0 ± 1.8%; 30W: 3.4 ± 3.8%; 150W: 5.3 ± 6.5%) during cycling compared with at rest

(Figure A2). Equivalent data for heart rate is shown in the supplementary online appendix (Figure A3).

DISCUSSION

This preliminary study of four wearable devices indicates that, across a range of SpO_2 values and levels of body motion, the ability of each of the respective devices to measure SpO_2 diverged substantially from that of a traditional pulse oximeter. A high proportion of readings were recorded as 'missed' when the device failed to provide a measurement within one minute, which would be considered a potentially critical operational failure in many aviation contexts. Missed measurements were common even at rest for most devices, and none were able to reliably provide SpO_2 measurements during cycling at moderate or even low intensity, when associated movement of the rest of the body was very minimal. The Apple Watch 6 had the highest accuracy with a potentially acceptable bias when SpO_2 values were achieved, but the device missed the majority of readings in the presence of very slight body motion, and missed nearly all readings when body motion was at a moderate level. These wearable devices are designed for SpO_2 measurements to be taken in a stationary state, but this is likely to be difficult or impossible to achieve during flight operations. Measurements were frequently missed even when there was only the slightest body motion, and it is therefore questionable whether these devices would be able to obtain measurements reliably in many real-world settings including aerospace environments.

The reduction in the performance of wearables in the presence of any movement of the body is attributable to motion artefact. As technology advances and becomes progressively miniaturised, this more readily exposes the PPG signal to noise such as motion artefact, and movement of the PPG sensor that alters the direction in which the light signal is emitted. This is particularly pertinent when the motion artefact frequency corresponds with that of the PPG signal (0.5-5.0Hz). Typically, motion artefact noise relates to a frequency of 0.01-10Hz, thus regularly overlapping with the PPG band.[7]

A further factor that should be considered is the potential for variation in peripheral circulation to affect SpO_2 measurements. Poor perfusion can cause a decrease in the ratio of arterial to venous blood at the sensor location, reduced venous saturation through a larger oxygen extraction ratio, and lower pulse amplitude. In addition, motion artefact can have a more profound impact when pulse amplitude is suppressed as it exerts a greater influence on the PPG signal.[9] Poor perfusion could conceivably have lowered the SpO_2 readings of the wrist-worn wearables in this study if a redistribution of blood flow to the exercising muscles in the lower limbs occurred. However, this seems unlikely as any such effect would also have applied to the reference pulse oximeter, and we note that the Cosinuss° Two (situated in the ear) was the only device to consistently under-report SpO_2.

The performance of wearables in measuring SpO_2 has only been investigated in a small number of studies, in which data was obtained at rest.[5,10,6] A perfectly motionless state provides optimal conditions and may explain the more favourable comparative data obtained with the Apple Watch 6,[10] Oxitone 1000M[5] and the predecessor Garmin Fēnix 5X Plus watch.[6] The latter study also explored the effect of reducing inspired oxygen concentration and demonstrated a larger bias at a simulated altitude of 12,000 ft compared with lower altitudes.[6] In the current study we observed a decrease in the performance of SpO_2 measurements under hypoxic conditions compared with during normoxia in all four wearable devices. Pulse oximeter performance is known to be reduced at lower SpO_2 values,[11] and in this context, the possibility that wearables may be additionally unreliable when oxygenation is lower, such as at altitude, warrants particular caution regarding their use in aerospace operations.

This study had several limitations. The sample size was intended to allow an initial preliminary evaluation of multiple wearables across varying conditions. The results are preliminary in nature and are intended to serve as the basis for more definitive research. Subjects were young and healthy and were primarily from a white ethnic background, precluding any analysis of the effect of skin pigmentation.[3] Cycling does not replicate actual in-flight conditions and was used as a reproducible surrogate for relevant levels of body motion, as this is the aspect of pedalling that has the potential to impair readings from wearable devices. The protocol did not target associated metabolic activity, which is not directly related to function of wearable monitors. It should be noted that hardware and software for these technologies remain under continuing development and improvement. Furthermore, consumer grade products such as the Apple Watch 6 and Garmin Fēnix 6 carry disclaimers that SpO_2 readings are not intended for medical use, and associated product information acknowledges that various factors may affect measurements including a user's individual anatomy, the fit of the device and ambient light conditions.

Wearable technology is rapidly advancing, and with further development the ability to measure SpO_2 unobtrusively offers great potential to be useful in a multitude of settings, including as a means of early detection of hypoxemia in clinical populations. This could encompass ambulatory and outpatient settings as well as ward-based, perioperative and critical care medicine. Ultimately, wearable-derived SpO_2 data may likewise offer benefits as in-flight tools, whether for pilots, passengers, aeromedical patients, rear crew or skydivers.

Based on this preliminary study, we suggest that further research and development is required before this can be generally recommended. Future investigations may consider ways to minimise movement-associated noise infiltrating reflective PPG signals, and should encompass relevant populations and environmental conditions including actual in-flight measurements.

In summary, while wearable devices offer great promise, in this preliminary study the four wearable devices investigated did not perform to the same standard as a traditional pulse oximeter for SpO_2 measurements. Limitations associated with varying conditions, including minimal body motion, may well apply in real-world settings including aviation and spaceflight, and further research into the use of wearables in these domains is required.

ACKNOWLEDGEMENTS

The authors thank Dr James Clark for developing the hypoxia chamber facility, and Lindsey Marjoram for expert technical assistance.

Financial Disclosure Statement: This work was supported by King's College London. The authors have no competing interests to declare.

REFERENCES

1. Bradke BS, Everman BR. Mild hypoxia of a skydiver making repeated, medium-altitude aircraft exits. Aerosp Med Hum Perform. 2020; 91(2): 110–115.

2. Ernsting J. Limitations of pulse oximetry in aviation medicine. In: 53rd International Congress of Aviation and Space Medicine. Warsaw, Poland: 2005.

3. Feiner J, Severinghaus J, Bickler P. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg. 2007; 105: S18-S23.

4. Grand View Research. Wearable Technology Market Size, Share & Trends Analysis Report By Product (Wrist-Wear, Eye-Wear & Head-Wear, Foot-Wear, Neck-Wear, Body-wear), By Application, By Region, And Segment Forecasts, 2021 - 2028 [Internet]. 2021 [cited 23 December 2021]. p. 1. Available from: https://www.grandviewresearch.com/industry-analysis/wearable-technology-market.

5. Guber A, Epstein Shochet G, Kohn S, Shitrit D. Wrist-sensor pulse oximeter enables prolonged patient monitoring in chronic lung diseases. J Med Systems. 2019;43: 230.

6. Lauterbach C, Romano P, Greisler L, Brindle R, Ford K, Kuennen M. Accuracy and reliability of commercial wrist-worn pulse oximeter during normobaric hypoxia exposure under resting conditions. Res Q Exerc Sport. 2020; 92: 549-558.

7. Lee J, Kim M, Park H, Kim I. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. Sensors. 2020; 20: 1493.

8. Nelson B, Low C, Jacobson N, Areán P, Torous J, Allen N. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med. 2020; 3: 90.

9. Petterson M, Begnoche V, Graybeal J. The effect of motion on pulse oximetry and its clinical significance. Anesth Analg. 2007; 105: S78-S84.

10. Pipek L, Nascimento R, Acencio M, Teixeira L. Comparison of SpO_2 and heart rate values on Apple Watch and conventional commercial oximeters devices in patients with lung disease. Sci Rep. 2021; 11: 18901.

11. Pulse Oximeter Accuracy and Limitations [Internet]. U.S. Food and Drug Administration. 2021 [cited 23 December 2021]. Available from: https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication.

12. Tamura T. Current progress of photoplethysmography and SPO_2 for health monitoring. Biomed Eng Lett. 2019; 9: 21-36.

13. Wiseman, Kelly, Swanny, McNamara, Beckert. Hypoxemia in healthy subjects at moderate altitude. Aviat Space Environ Med. 2013 84: 22-26

TABLES

Table I.

		Apple Watch 6	Garmin Fēnix 6	Cosinuss° Two	Oxitone 1000M
Number of data points					
	Rest	160	160	160	160
	30W cycling	60	60	60	60
	150W cycling	60	60	60	60
Missed readings (% of total)					
	Rest	2.5%	20%	11%	14%
	30W cycling	65%	65%	12%	82%
	150W cycling	95%	83%	18%	92%
Mean absolute percentage error					
	Rest	-2.26	-2.19	2.66	-2.39
	30W cycling	-0.80	-3.92	2.06	-3.44
	150W cycling	-4.21	-9.89	3.33	-6.69
Accuracy (%)					
	Rest	97.7	97.8	97.3	97.6
	30W cycling	99.2	96.1	97.9	96.6
	150W cycling	95.8	90.1	96.7	93.3

Table I. SpO_2 measurements: number of data points, percentage of missed readings, mean absolute percentage error and percentage accuracy for each device measuring SpO_2 at rest and during cycling at 30W and 150W.

Table II.

		Apple Watch 6	Garmin Fēnix 6	Cosinuss° Two	Oxitone 1000M
Number of data points					
	Rest	160	160	160	160
	30W cycling	60	60	60	60
	150W cycling	60	60	60	60
Missed readings (% of total)					
	Rest	0%	2%	7%	5%
	30W cycling	0%	2%	12%	67%
	150W cycling	0%	0%	20%	77%
Mean absolute percentage error					
	Rest	1.05	0.8	7.64	2.56
	30W cycling	-7.51	7.91	0.51	9.71
	150W cycling	-2.33	29.41	45.14	33.32
Accuracy (%)					
	Rest	98.95	99.2	92.36	97.44
	30W cycling	92.49	92.09	99.49	90.29
	150W cycling	97.67	70.59	54.86	66.68

Table II. Heart rate measurements: number of data points, percentage of missed readings, mean absolute percentage error (MAPE) and percentage accuracy for each device measuring heart rate at rest and during cycling at 30W and 150W.

FIGURE LEGENDS

Figure 1. Mean arterial oxygen saturation and heart rate at rest and cycling at 30W and 150W under normoxic (20.9% oxygen) and hypoxic (11.8% oxygen) conditions. Solid red lines and circles denote normoxia. Dashed blue lines and squares denote hypoxia. Asterisks denote a statistically significant effect of hypoxia ($P < 0.05$). Data are mean ± SD.

Figure 2. Arterial oxygen saturation measured by the reference pulse oximeter and wearable devices during normoxia (red boxes) and hypoxia (blue boxes). Data are from all conditions combined (rest and cycling). The mean, interquartile range (boxes) and maximum and minimum values (bars) are shown. Asterisks denote a statistically significant difference ($P < 0.05$) between reference data obtained from the traditional pulse oximeter and data from the respective wearable devices.

Figure 1 (revised)

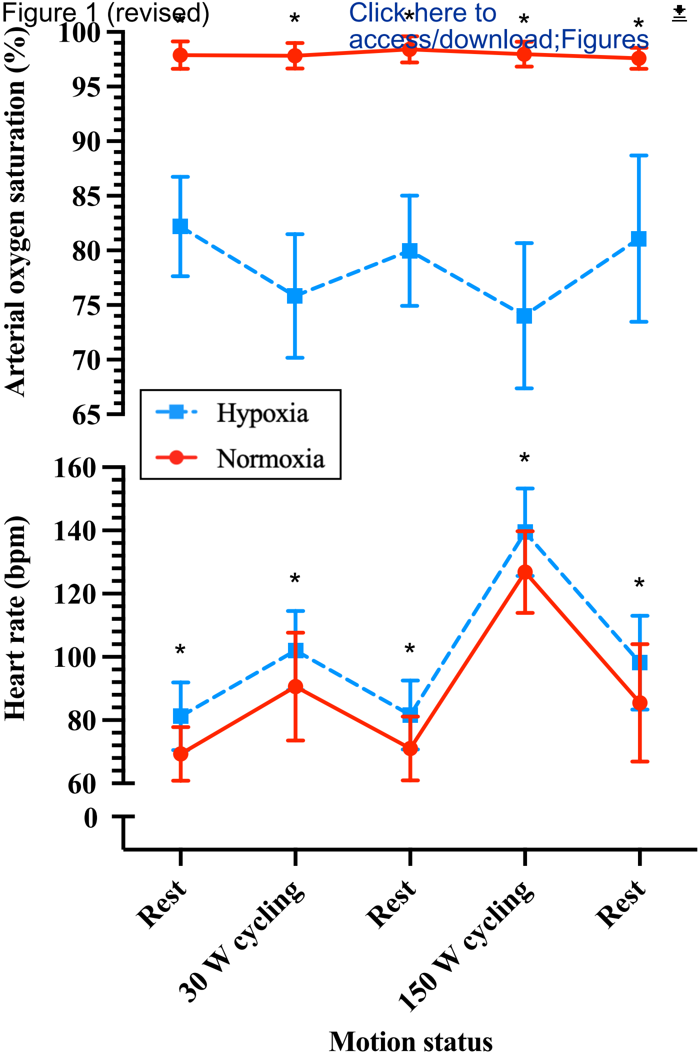
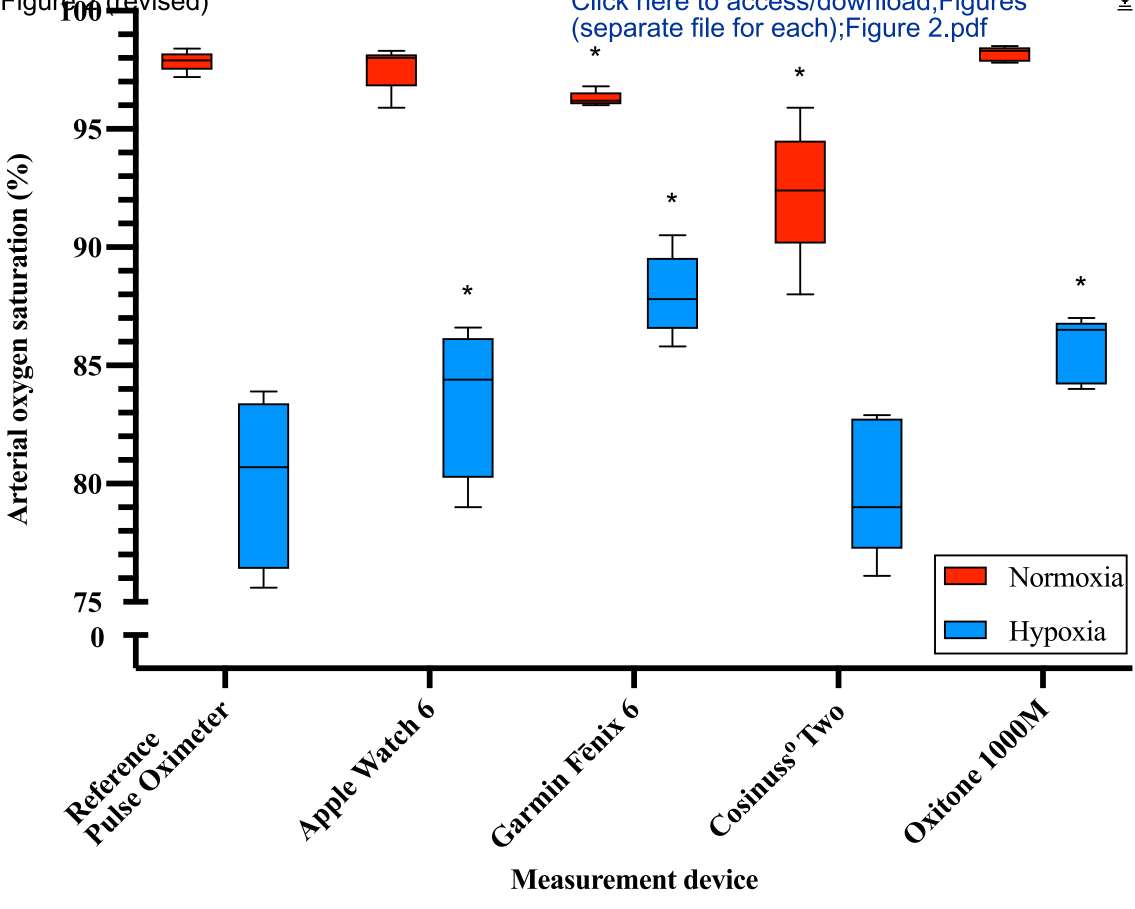


Figure 2 (revised)



Auxiliary Material: Supplementary Online Appendix

Supplement to:
Measuring arterial oxygen saturation using wearable devices under varying conditions

SUPPLEMENTARY RESULTS

Figure A1: Heart rate measurements obtained from the reference pulse oximeter and wearable devices during normoxia and hypoxia.

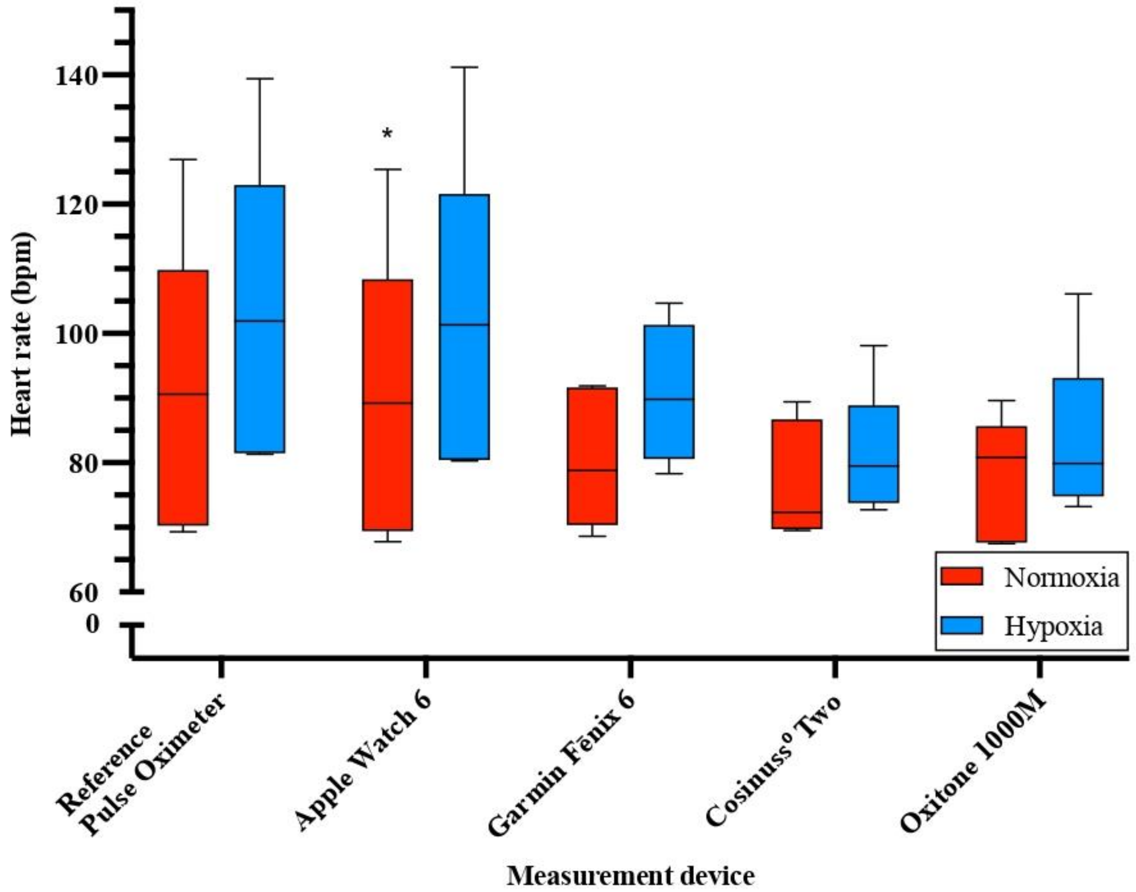


Figure A1. Heart rate measured by the reference pulse oximeter and wearable monitors during normoxia (red boxes) and normoxia (blue boxes). Data are from all conditions (rest and cycling combined). The mean, interquartile range (boxes) and maximum and minimum values (bars) are shown. Asterisks denote a statistically significant difference (*P* <0.05) between reference data obtained from the traditional pulse oximeter and data from the respective wearable devices. During normoxia, the Apple Watch 6 was the only device to show a significant difference in heart rate data compared with the reference pulse oximeter (t(4)=4.762, *P* = 0.009). During hypoxia, there was no significant difference between the heart rate data obtained from any of the wearables and the reference pulse oximeter.

Figure A2: Bland-Altman analyses of SpO_2 data obtained from wearable devices.

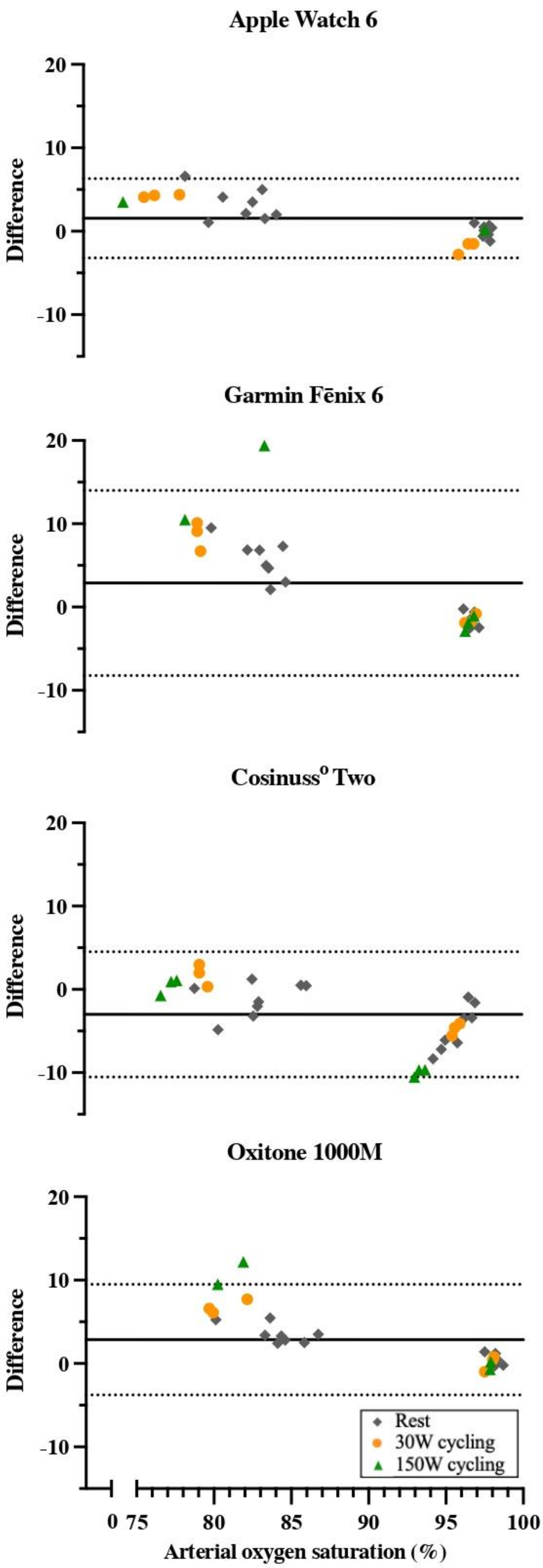


Figure A2. Bland-Altman plots of SpO_2 (%) for each device compared with the reference data. Solid lines represent mean bias and dashed lines represent 95% limits of agreement. Grey diamonds show data obtained at rest, orange circles show data obtained during 30W cycling and green triangles show data obtained during 150W cycling.

Figure A3: Bland-Altman analyses of heart rate data obtained from wearable devices.

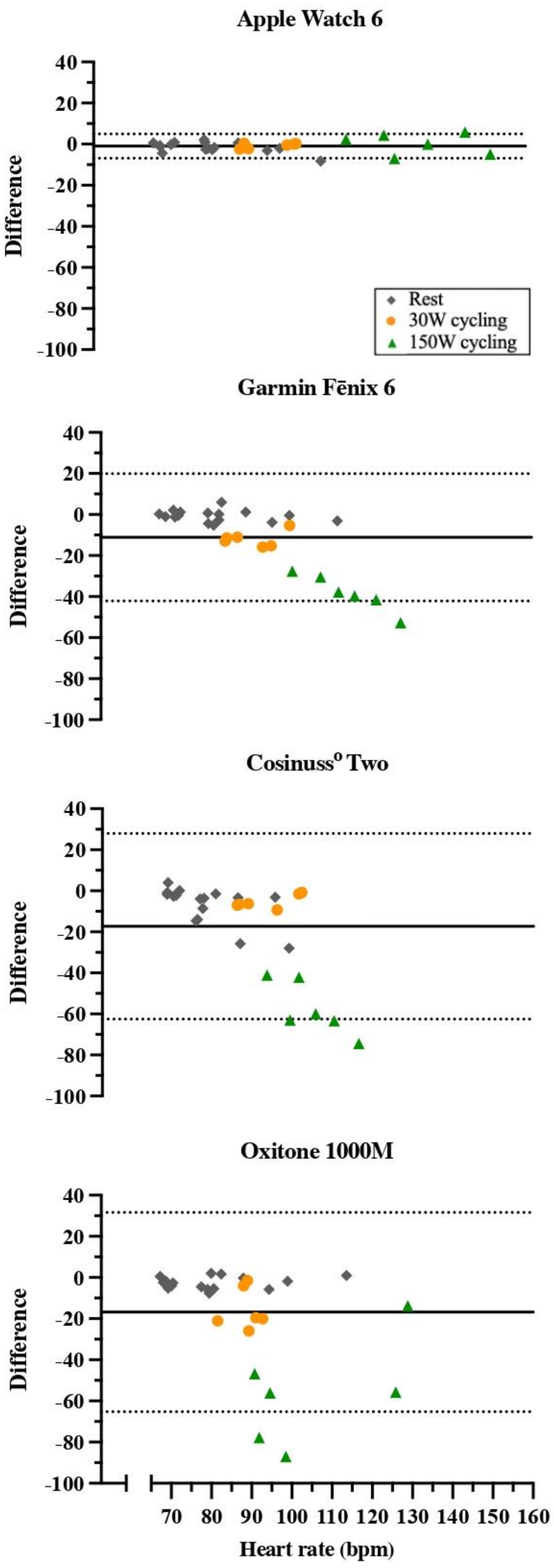


Figure A3. Bland-Altman plots of heart rate (bpm) for each device compared with the reference data. Solid lines represent mean bias and dashed lines represent 95% limits of agreement. Grey diamonds show data obtained at rest, orange circles show data obtained during 30W cycling and green triangles show data obtained during 150W cycling. Overall, all devices under-reported heart rate. Mean bias and variability of bias generally increased from rest to 150W cycling across the devices. The Garmin Fēnix 6 had the smallest mean bias at rest (-0.7 ± 2.8 bpm), whilst the Apple Watch had the smallest mean bias during exercise (30 W: -0.7 ± 1.3 bpm; 150 W: +0.1 ± 5.1 bpm).