Blood Oxygen app on Apple Watch

October 2022

MASIMO 2032
Apple v. Masimo
IPR2024-00294

Overview

Apple Watch — Series 6 or later, excluding Apple Watch SE[1] — is capable of measuring oxygen saturation of arterial hemoglobin (SpO_2) for fitness and wellness applications. The Apple Watch optical system uses a combination of light emitters and light sensors to take the blood oxygen measurement. A blood oxygen level represents the percentage of arterial hemoglobin in red blood cells that carry the oxygen from the lungs to the rest of the body. Blood oxygen is an established measure of overall wellness. This paper provides a detailed description of the Blood Oxygen feature on Apple Watch, including its testing and accuracy validation.

Introduction

The Blood Oxygen app on Apple Watch analyzes signals generated by sensors to provide estimates of the functional hemoglobin oxygen saturation of arterial blood using pulse oximetry technology. As public awareness of blood oxygen has increased, so too has demand for pulse oximeters, which has elevated the importance of describing device accuracy and how blood oxygen is determined. The purpose of this white paper is to share additional information about the Blood Oxygen feature on Apple Watch, its development, and its reading accuracy.

Wellness uses of the app include measurements while hiking and trekking at varying altitudes and awareness of one's natural or baseline measurements. Many parameters can affect blood oxygen, including oxygen concentration in the air that a user breathes. At higher altitude and lower barometric pressure, the fractional content of oxygen in inspired air decreases, which affects the blood oxygen saturation. Knowing the saturation can help users gauge activity and effort when traveling to or hiking at altitude. The ability to understand and trend blood oxygen levels can also be a helpful exercise metric for users. Normal physiologic response to exercise in a healthy person is maintaining or increasing SpO_2.

The Blood Oxygen app operates in two modes: on-demand spot checks that users initiate manually, and intermittent background measurements taken during low-movement conditions without requiring any user action. The app attempts to collect optical sensor data to generate SpO_2 readings when the user is stationary with the wrist in the desired posture for a short time period. Generating background spot check measurements when predefined conditions are met is referred to as "opportunistic" data acquisition. The Apple Watch irregular rhythm notification feature also uses this type of data acquisition method. It is important to note that the intermittent background measurements taken by Apple Watch are not the same as the continuous second-by-second measurement capability commonly available with bedside pulse oximeters.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

SpO_2 Technology

The Blood Oxygen feature on Apple Watch measures SpO_2 using conventional pulse oximetry methods: It shines red and near-infrared (IR) light into blood-perfused tissue, detects and processes the reemitted light photo-signals into respective photoplethysmograms (PPGs) that track the heartbeat-induced pulsations, determines the red-to-IR modulation ratio, and translates this into units of % SpO_2 through a predefined mapping relationship. Blood fully saturated with oxygen appears bright red and transitions to a darker brown color as oxygenation falls. The modulation ratio measured in pulse oximetry correlates with the color of the tissue's pulsing arterial blood and is thus used for determining SpO_2 — an estimate of the blood's true oxygenation as measured from an arterial blood draw, also known as the SaO_2 value.

The Apple Watch back crystal includes an array of light emitter and detector apertures configured as a "reflectance" sensor; emitted light scatters through the perfused tissues beneath Apple Watch, with a portion of that light reemerging and striking photodetectors along the same surface. The light sources used by the Blood Oxygen app and shared with other health features on Apple Watch comprise red, IR, and green LEDs operating at wavelengths of approximately 660, 850, and 525 nm, respectively.

For best results, Apple Watch should be worn snugly but comfortably. The back crystal should be approximately centered on the wrist, and it should be in complete contact with the soft tissue at the back of the wrist, farther up the arm than — and not touching — the ulnar styloid (wrist bone).

Pulse signals are smaller at the wrist than at the fingers and other conventional SpO_2 device probe sites due to differences in local vasculatures. Users should remain still and relaxed when manually taking readings, as with other pulse oximeters when pulses are weak. Apple Watch initiates intermittent spot checks automatically when it senses that the user is similarly still and that the wrist is in a proper position (the arm is generally horizontal and the palm is facing down).

Development

During the development and evaluation of the Blood Oxygen feature, Apple collected data in multiple institutional review board (IRB)–approved studies involving many hundreds of participants who consented to the collection and use of their data for this purpose. These studies included controlled laboratory studies and supervised data collection sessions under a variety of user behaviors, cardiorespiratory conditions, and ambient environments, including real or simulated altitudes to span the 70–100% blood oxygen saturation range (based on conventional finger pulse oximetry or arterial blood sampling).

Subject pools included a wide range of skin types and tones to ensure that the sensor platform can accommodate the full range of users and maintain accuracy. At the wavelengths that Apple Watch uses, melanin is a strong light absorber — particularly in the green and red part of the spectrum — potentially making PPG measurements more difficult in users with darker skin tones. To account for this, the Apple Watch sensing platform senses the amount of detected light signals, and it automatically adjusts the LED current (and hence the light output), photodiode gain (sensitivity to light), and sampling rate to ensure adequate signal resolution across the range of human skin tones.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

Motion and low blood perfusion can obscure the underlying heartbeat-induced pulsatile signals, and both are well-known challenges to pulse oximeter reading accuracy and availability. Arm position can also impact SpO_2 readings because it can create a condition of "venous pulsation" where local arterial and venous blood compartments modulate with the cardiac cycle. Venous blood usually has substantially lower oxygen saturation, so its contribution could falsely lower SpO_2 measurements. To ensure accurate reporting, the Blood Oxygen app withholds readings when it determines that the PPG signals from Apple Watch are inadequate or that positional or movement conditions are unsuitable for reliable SpO_2 readings. If this occurs during a user-initiated session, a message displays indicating the potential reasons.

Performance Accuracy

Pulse oximeter accuracy is described in terms of the agreement between the device-reported SpO_2 and the true SaO_2 value, which is the gold standard reference for arterial blood oxygenation. Per the pulse oximeter International Standard[2] and FDA Guidance,[3] accuracy is computed as the root-mean-square of the pooled SpO_2–SaO_2 differences (A_{rms}) observed in a population of subjects spanning the full range of 70–100% SaO_2. Because measurements are statistically distributed, only about two-thirds of readings can be expected to fall within $\pm A_{rms}$ of the SaO_2 value. The standardized test methodology comprises a desaturation study conducted "under well-controlled, optimal laboratory conditions"[2] on at least 10 healthy adult subjects, as described in the ISO and FDA references noted above.

The Apple team finalized and prospectively validated the algorithm's Ratio-to-SpO_2 mapping function in a two-part study that included SaO_2 values from arterial blood sampling. Pooling and analyzing the paired observations from this development data, as described below, offers a perspective of the accuracy of the SpO_2 provided by Apple Watch when tested in the same manner as hospital-use pulse oximeters.

Data Collection

Apple contracted with a lab facility experienced in conducting similar studies under its IRB-approved protocol. Overall, 50 healthy adult subjects were enrolled, and each signed an informed consent form to allow the collection of their data for the purposes of this study. Subjects ranged from ages 19 to 40 (a mean of 26.6 years), split evenly by biological sex, and covering a wide range of skin tones (eight subjects had dark skin characterized visually as Fitzpatrick scale type V or VI). Each subject was fitted with a radial artery cannula, and periodic blood samples were drawn for analysis in a hemoximeter to determine SaO_2 spectroscopically.

On the other arm, each subject wore an Apple Watch Series 6 with an Apple Watch Sport Band near the center of the wrist, close but proximal of the styloid bone, with snug but comfortable band tightness. Sixteen watches — eight large (44mm case) and eight small (40mm case) — were distributed evenly across 48 of the subjects, independent of wrist size. (Two watches were cycled again for the remaining two subjects.) Hypoxia was induced in a stepwise manner by varying the subjects' inspired oxygen fraction, with blood samples taken during periods of stable saturation as indicated in real time by a pair of monitoring pulse oximeters using finger probes.

Raw unprocessed watch signals were recorded continuously throughout the sessions using proprietary data collection software, independent of the blood sampling. At the end of each session, recorded signals were downloaded from Apple Watch for offline processing using the Blood Oxygen app algorithm (with watchOS 8). SpO_2 values were computed using individual 15-second segments of watch signals; for direct comparisons with SaO_2, segment start times were aligned with the beginning of each respective blood draw. Study data was collected in two parts — the first 26 subjects' signals contributed to the algorithm's Ratio-to-SpO_2 mapping function (calibration), and the final 24 subjects' data was used to validate those results and characterize the system's A_{rms} accuracy against a blood reference. The processing and analysis of this collected data were conducted solely by Apple.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

Performance characteristics were evaluated across the 70–100% SaO_2 span, as well as a narrower 85–95% SaO_2 span that encompasses the region generally associated with the onset of less-than-normal oxygen saturation. Five subject groupings are presented: light skin (Fitzpatrick I–IV), dark skin (Fitzpatrick V–VI), male, female, and overall. The measures comprise A_{rms} and its 95 percent confidence interval (CI) computed by bootstrapping among the subjects, mean SpO_2–SaO_2 difference and bootstrapped 95 percent CI, upper and lower limits of agreement (LOA) computed per the Bland-Altman method accounting for repeated measures,[4] and SpO_2 reading availability.

Results

Figure 1 below illustrates a study session for one of the subjects. Each subject was exposed to two desaturation cycles from ~100% to ~70% SaO_2, with a recovery and rest in the middle.

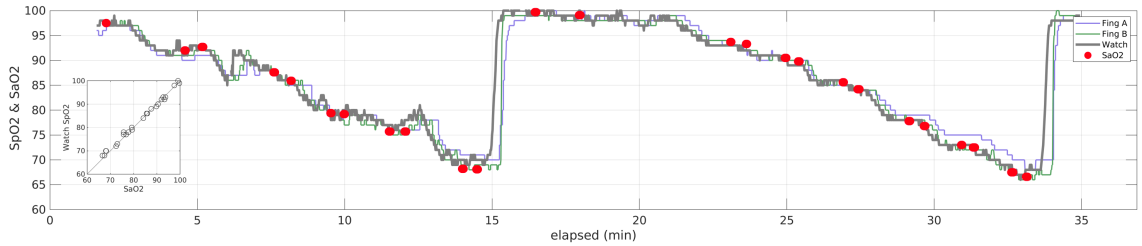


Figure 1. This trend plot represents one of the sessions from the validation data set. SpO_2 trends for the monitored fingers — Fing A and Fing B — and for Apple Watch are shown by the blue, green, and gray lines, respectively. Sampled blood SaO_2 values are indicated by the red dots. The inset to the left plots Apple Watch SpO_2 compared with SaO_2 during this session.

All 50 subject sessions provided data. Only valid data, as assessed before the analysis, was included: when blood draws were obtained while saturation was stable, when SaO_2 values were available and uncorrupted, and when the wrist was still and oriented properly. Overall, there were 1,020 such observation periods with sampled $SaO_2 \geq 70\%$. Simultaneous SpO_2 values were available in 966 of these periods (514 and 452 from the first and second sets, respectively, with 54 instances of the algorithm deeming PPG signals to be inadequate), resulting in an overall reading availability of 94.7 percent. The median number of paired observations per subject session was 21 (range of 6–25), with four contributing fewer than 10 pairs. Performance results for the two data sets are shown in table 1 and figures 2 and 3.

The 70–100% A_{rms} was 1.77% SpO_2 in the first 26 subjects and 2.18% SpO_2 in the final 24-subject validation set. Linear regression is y = 0.961x + 3.382 for the calibration data set and y = 0.959x + 3.906 for the validation set; confidence intervals for these two regression lines overlap across the entire span. With comparable regressions in the two data sets, and with overlap in respective A_{rms} and mean difference confidence intervals, performance comparisons across gender and skin tone are presented for the pooled data to achieve higher statistical power within each subgroup, as shown in table 2 and figures 4 and 5.

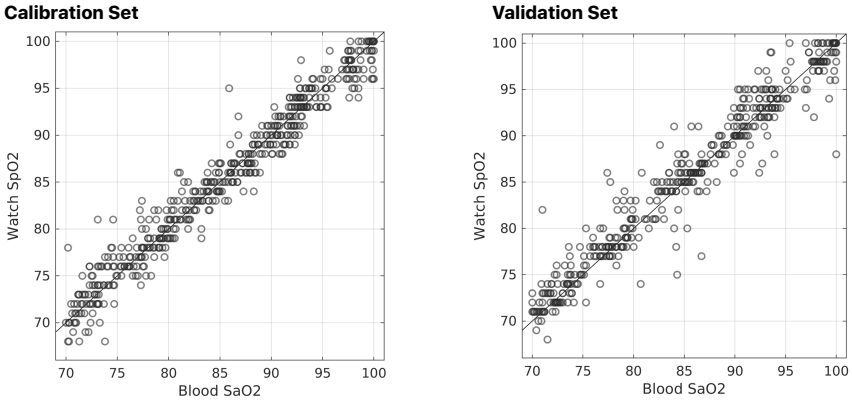


Figure 2. These scatterplots compare paired Apple Watch SpO_2 with SaO_2 for the calibration (N=26 subjects) and validation (N=24 subjects) data sets.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

Table 1. Summary Statistics over the 70–100% and 85–95% SaO_2 Spans

SaO_2 Span	Data Set	#Subjects / #Pairs / #Tries	A_{rms} [95% CI]	Mean Difference [95% CI]	95% LOA
70–100%	Calibration	26 / 514 / 551	1.77 [1.46–2.11]	+0.06 [-0.36–0.52]	-3.39–3.51
	Validation	24 / 452 / 469	2.18 [1.55–2.84]	+0.37 [-0.28–0.97]	-3.85–4.59
85–95%	Calibration	26 / 194 / 203	1.67 [1.37–2.00]	-0.14 [-0.61–0.32]	-3.44–3.17
	Validation	24 / 175 / 177	2.05 [1.41–2.63]	+0.24 [-0.34–0.80]	-3.79–4.26

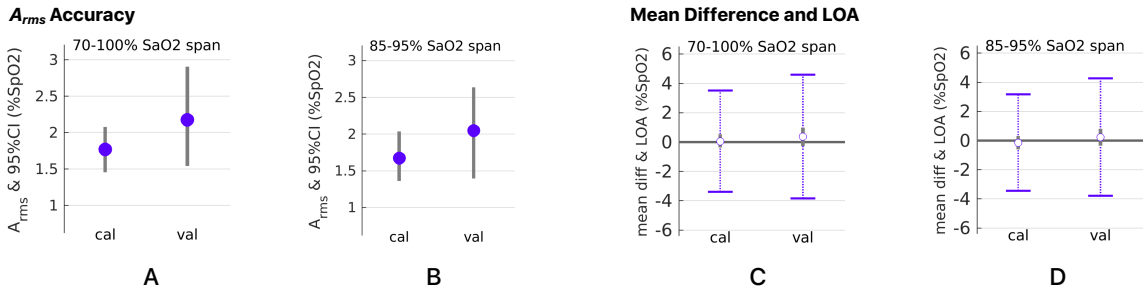


Figure 3. Charts A and B show A_{rms} and the associated 95 percent confidence intervals for the calibration and validation data sets over the two SaO_2 spans. Charts C and D show the mean SpO_2–SaO_2 differences indicated by the open circles for the two data sets and spans, with gray bars indicating the 95 percent confidence intervals for the means; 95 percent LOA for the individual observed differences are indicated by dotted lines and upper and lower blue lines.

Below, figure 4 provides plots for paired data broken out by subgroups of male, female, light skin tones, and dark skin tones. Figure 5 and table 2 summarize the performance characteristics over the 70–100% and 85–95% SaO_2 spans across each subgroup and overall.

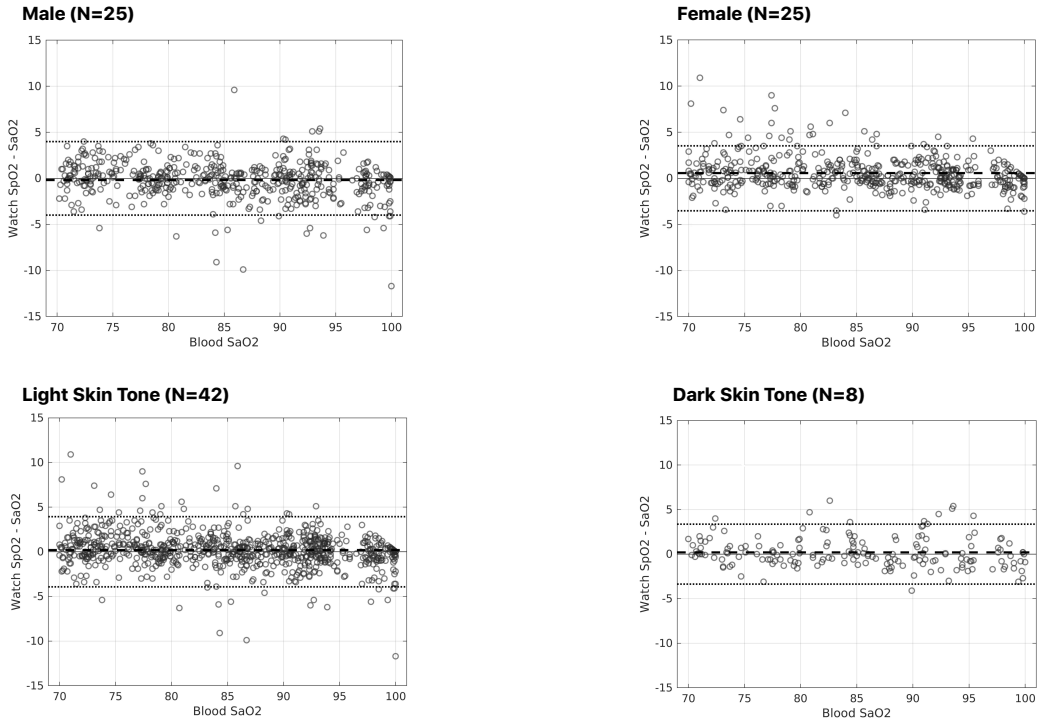


Figure 4. Modified Bland-Altman plots illustrate the paired Apple Watch SpO_2 and SaO_2 data in each of the four subgroups over the 70–100% span. The mean difference is shown by the dashed lines, and 95 percent LOA is shown by the dotted lines. These plots use decimal precision SpO_2 to better distinguish overlapping data.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

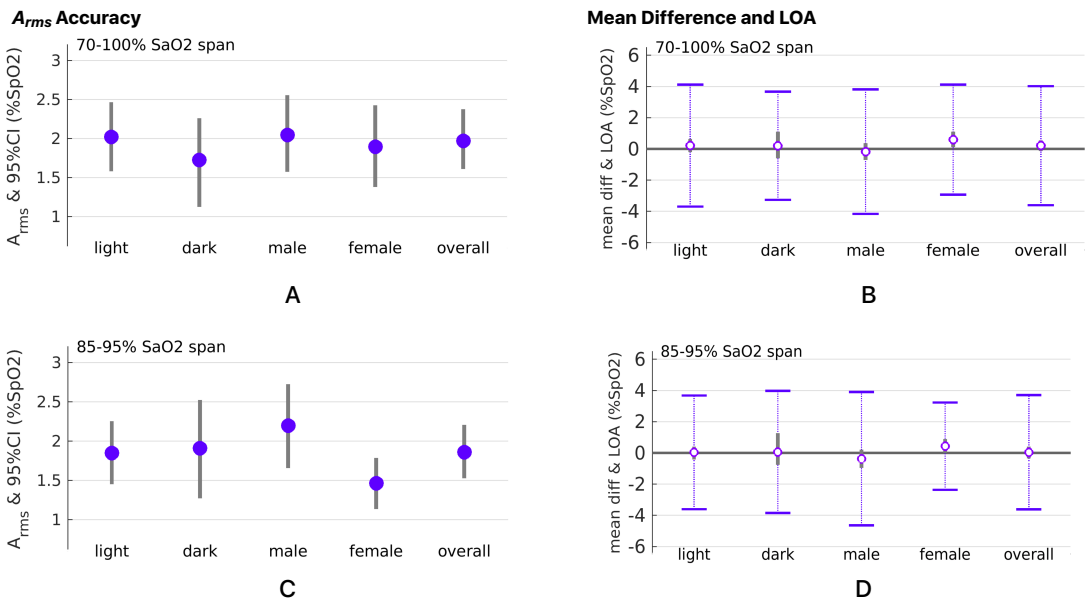


Figure 5. These charts compare performance for the data in the 70–100% and 85–95% SaO_2 across the five groups: light skin tone subjects, dark skin tone subjects, male subjects, female subjects, and overall. A_{rms} is shown in charts A and C, with the gray error bars indicating 95 percent CI. The mean SpO_2–SaO_2 differences are shown in charts B and D for the two spans by the open circles, with the gray bars indicating 95 percent CI; 95 percent LOA in the observed differences are indicated by the upper and lower blue lines.

Table 2. Performance Observations Overall and in Subgroups

SaO_2 Span	Subgroup	#Subjects / #Pairs / #Tries	A_{rms} [95% CI]	Mean Difference [95% CI]	95% LOA
70–100%	**Overall**	**50 / 966 / 1020**	**1.97 [1.61–2.38]**	**0.21 [-0.13–0.55]**	**-3.61–4.02**
	Light skin	42 / 794 / 843	2.02 [1.58–2.47]	0.21 [-0.21–0.63]	-3.70–4.12
	Dark skin	8 / 172 / 177	1.73 [1.12–2.26]	0.20 [-0.61–1.11]	-3.26–3.67
	Male	25 / 486 / 515	2.05 [1.57–2.56]	-0.17 [-0.71–0.36]	-4.16–3.81
	Female	25 / 480 / 505	1.90 [1.38–2.43]	0.59 [0.12–1.11]	-2.93–4.11
85–95%	**Overall**	**50 / 369 / 380**	**1.86 [1.53–2.21]**	**0.04 [-0.35–0.40]**	**-3.62–3.70**
	Light skin	42 / 309 / 320	1.85 [1.45–2.25]	0.04 [-0.42–0.40]	-3.60–3.68
	Dark skin	8 / 60 / 60	1.91 [1.27–2.53]	0.06 [-0.77–1.27]	-3.85–3.97
	Male	25 / 180 / 185	2.20 [1.66–2.73]	-0.37 [-0.97–0.21]	-4.64–3.90
	Female	25 / 189 / 195	1.47 [1.14–1.79]	0.43 [0.01–0.89]	-2.36–3.23

MASIMO 2032
Apple v. Masimo
IPR2024-00294

Discussion

The Blood Oxygen app on Apple Watch provides accurate and validated on-demand and background measurements of SpO_2. The observed 50-subject A_{rms} accuracy of 1.97% SpO_2 is within the typical specification limits defined in the U.S. FDA Guidance document (≤3.0% or ≤3.5% SpO_2, depending on sensor type)[3] and the ISO standard limit (≤4% SpO_2)[2] when tested according to the methods described above. This A_{rms} value is also similar to those of medical-grade devices used in hospitals when tested in the same manner.

Accuracy, LOA, and mean SpO_2–SaO_2 difference (bias) were comparable across the four subgroups and did not differ statistically from one another in either of the SaO_2 spans. Recent literature has raised concerns of significant SpO_2 reading bias and degraded accuracy in Black patients. In the subjects included in our controlled lab study, we did not observe a skin-tone dependence in A_{rms} or mean reading differences when compared with blood.

Equivalent to conventional pulse oximetry, performance can be affected if the sensing optics do not make complete contact with the skin or are worn very tightly. These suboptimal conditions commonly result in unavailable readings, but they can also affect accuracy — creating SpO_2 readings that may overestimate or underestimate SaO_2. Much of the outlier scatter seen in figure 2 resulted from three subject sessions — one in the first data set and two in the second — in which the watch was not worn with recommended snugness. One of these sessions (a male subject with well-perfused light skin) is highlighted in figure 6, with the observations overlaid on the remaining data. The other two noted sessions (both female subjects with well-perfused light skin) account for 10 of the high-reading outliers seen with SaO_2 < 85%. In the absence of these three data sets, the overall pooled 70–100% A_{rms} improves from 1.97% to 1.67% SpO_2.

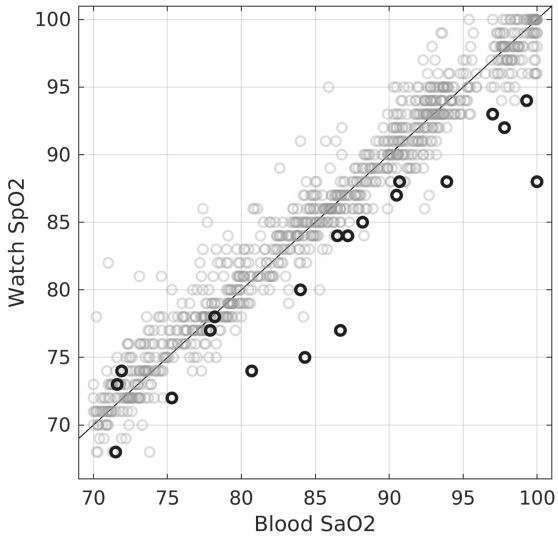


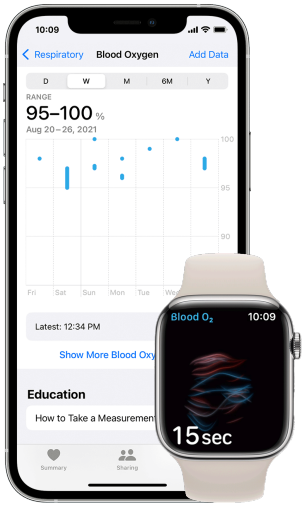
Figure 6. Observations from one subject session in which the watch was not worn with recommended snugness are shown by the dark points overlaid on the overall data shown in light gray.

Factors like interference from the wrist bone, a loose or overly tight watch band, and low skin perfusion may make it difficult to obtain readings from Apple Watch. Performance may be improved by moving Apple Watch farther away from the wrist bone, ensuring that the band is snug, and avoiding cold wrists and hands. When taking on-demand measurements, it is also important for users to be still and relaxed to obtain an SpO_2 reading.

MASIMO 2032
Apple v. Masimo
IPR2024-00294

HealthKit

HealthKit provides a central repository for health and fitness data on iPhone and Apple Watch. A user's background and on-demand SpO_2 values are displayed in the Health app and can be viewed by day, week, month, or year. Values taken at barometric pressures generally found at altitudes above approximately 5000 feet are annotated "high elevation environment." Values taken during sleep are also labeled. Understanding individual level trends allows the user to see the variability in their average values, as well as highs and lows and when those values occurred (for example, during air travel or sleep).



Conclusion

Apple Watch includes a range of features that focus on health, fitness, safety, and staying connected. The Blood Oxygen app on Apple Watch is an accurate wrist-based pulse oximeter capable of both on-demand and background measurements. The data is available in the Health app on iPhone, allowing users to track values and trends. Features that track activity, heart rate, cardio fitness, and SpO_2 make Apple Watch a powerful wellness device for all users. Its blood oxygen measurements are accurate, as described in this white paper, and can be helpful in assessing general wellness.

[1]Based on availability as of October 2022. [2]ISO. 2017. "ISO 80601-2-61: Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment," second edition. Corrected February 2018. www.iso.org/standard/67963.html. [3]FDA. "Pulse Oximeters - Premarket Notification Submissions [510(k)s] - Guidance for Industry and Food and Drug Administration Staff." Regulatory Information. Guidance document issued March 4, 2013. fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug. [4]Bland, J. Martin, Douglas G. Altman. 2007. "Agreement between methods of measurement with multiple observations per individual." *Journal of Biopharmaceutical Statistics* 17, no. 4: 571–82. doi.org/10.1080/10543400701329422.

© 2022 Apple Inc. All rights reserved. Apple, the Apple logo, Apple Watch, HealthKit, iPhone, and watchOS are trademarks of Apple Inc., registered in the U.S. and other countries.

MASIMO 2032
Apple v. Masimo
IPR2024-00294