Accuracy of the Apple Watch Oxygen Saturation Measurement in Adults: A Systematic Review

Paul Windisch 1, Christina Schröder 1, Robert Förster 1, Nikola Cihoric 2, Daniel R. Zwahlen 1

1. Department of Radiation Oncology, Kantonsspital Winterthur, Winterthur, CHE 2. Department of Radiation Oncology, Inselspital, University Hospital of Bern, Bern, CHE

Corresponding author: Paul Windisch, paulwindisch93@gmail.com

Abstract

The purpose of this review is to summarize the research on the accuracy of oxygen saturation (spO2) measurements using the Apple Watch (Apple Inc., Cupertino, California). The Medline and Google Scholar databases were searched for papers evaluating the spO2 measurements of the Apple Watch vs. any kind of ground truth and records were analyzed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

The five publications with 973 total patients that met the inclusion criteria all used the Apple Watch Series 6 and described 95% limits of agreement of +/- 2.7 to 5.9% spO2. However, outliers of up to 15% spO2 were reported. Only one study had patient-level data uploaded to a public repository. The Apple Watch Series 6 does not show a strong systematic bias compared to conventional, medical-grade pulse oximeters. However, outliers do occur and should not cause concern in otherwise healthy individuals. The impact of race on measurement accuracy should be investigated.

Introduction And Background

Wearables are blurring the lines between lifestyle devices and medical products. In the case of the Apple Watch (Apple Inc., Cupertino, California), some features, like the electrocardiogram (ECG) function have been evaluated in large studies and received clearance by the United States Food and Drug Administration (FDA) while others have not and are framed as wellness trackers [1].

Despite this distinction, researchers have started investigating the potential use of these non-approved features for remote monitoring settings in the future [2]. An already existing issue is that many users who just use the watch as a lifestyle device are likely unaware of the distinction and might approach their providers with questions regarding conspicuous values measured by the watch. The ability of the watch to provide oxygen saturation (spO2) measurements is one of those features that has previously been available only with medical-grade pulse oximeters.

However, unlike medical-grade pulse oximeters, the Apple Watch can’t use the common transmissive pulse oximetry where light is passed through a thin part of the body, and instead has to rely on reflectance pulse oximetry where light is passed into the wrist and only the reflected light is measured at the photodiode. This approach is considered more challenging as changes in spO2 tend to produce smaller changes in the signals which also appear to be less stable [3].

With this technology now being available to a large population, the purpose of this systematic review is therefore to summarize the research that has been published to date to provide information on the accuracy of the measurements and guide further research.

Review

The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [4]. Original articles that evaluated the spO2 measurement feature of any Apple Watch series against any kind of ground truth were included. No constraints regarding the language of the publication were applied. Articles had to be published no earlier than 2015.

The Medical Literature Analysis and Retrieval System Online (MEDLINE) database were searched on October 18th, 2022, via the PubMed interface. The query was designed to include studies with either the words “Apple Watch” or “Apple Smartwatch” in the title or abstract as well as at least one word indicating the investigation of oxygen saturation. The complete search query that was used for PubMed was therefore: (Apple Watch|title/abstract) OR (Apple Smartwatch|title/abstract) AND (spO2|title/abstract) OR
For Google Scholar, two separate queries were used, and the filtering by year of publication was done through the user interface: ‘allintitle: SpO2 OR Oxygen OR Oximetry ‘Apple SmartWatch’, and ‘allintitle: SpO2 OR Oxygen OR Oximetry ‘Apple Watch’.

After the exclusion of duplicates, the titles, as well as abstracts, were screened, and only relevant publications proceeded to full-text screening. The decision as to whether a study met the inclusion criteria of the review was made by two authors (PW, CS) without using automated tools. A third author (DRZ) acted as a referee in case of a potential disagreement between the two authors responsible for screening. The review had not been registered beforehand, and no protocol had been published. Two authors (PW and CS) independently extracted data and discussed any discrepancies. Data were extracted with regards to 1) Study parameters (title, authors, year of publication, number of patients); 2) clinical parameters (conditions, patient age, spO2 measurements); 3) technical parameters (Apple Watch series, wristband, ground truth); and 4) study parameters (data availability, conflict of interest, funding). The inclusion workflow is depicted in Figure 1.

FIGURE 1: Workflow of the literature search according to PRISMA guidelines

Source: Page et al. [4]

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

The query returned 17 articles (12 from PubMed, and five from Google Scholar), three of which were duplicates. Eight articles were original articles that were excluded due to not comparing Apple Watch spO2...
measurements to a ground truth [5-12]. One article was excluded due to being an editorial [13]. All excluded articles and the respective reasons for exclusion are provided in Table 1.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
<th>Year</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yamagami et al.</td>
<td>Early Detection of Symptom Exacerbation in Patients With SARS-CoV-2 Infection Using the Fitbit Charge 3 (DEXTERITY): Pilot Evaluation</td>
<td>2021</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>O'Neill et al.</td>
<td>2-Hydroxybenzylamine (2-HOBA) to Prevent Early Recurrence of Atrial Fibrillation After Catheter Ablation: Protocol for a Randomized Controlled Trial Including Detection of AF Using a Wearable Device</td>
<td>2021</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>Chandrasekaran et al.</td>
<td>Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights From a National Survey</td>
<td>2020</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>LaMunion et al.</td>
<td>Use of Consumer Monitors for Estimating Energy Expenditure in Youth</td>
<td>2020</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>Abt et al. [9]</td>
<td>Walking Cadence Required to Elicit Criterion Moderate-Intensity Physical Activity Is Moderated by Fitness Status</td>
<td>2019</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>Xie et al. [11]</td>
<td>Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study</td>
<td>2018</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>Abt et al. [12]</td>
<td>Measuring Moderate-Intensity Exercise with the Apple Watch: Validation Study</td>
<td>2018</td>
<td>Original article but out of scope</td>
</tr>
<tr>
<td>Abt et al. [10]</td>
<td>The Validity and Inter-device Variability of the Apple Watch™ for Measuring Maximal Heart Rate</td>
<td>2018</td>
<td>Original article but out of scope</td>
</tr>
</tbody>
</table>

**TABLE 1: Excluded studies and reason for exclusion**

The remaining five articles were included [2,14-17]. A summary of extracted characteristics is available in Table 2.
<table>
<thead>
<tr>
<th>Publication</th>
<th>Device</th>
<th>Wrist band</th>
<th>Ground truth</th>
<th>Number of patients &amp; condition</th>
<th>Age (range) [years]</th>
<th>SpO₂ according to ground truth [range] [%]</th>
<th>Pearson correlation coefficient</th>
<th>Race</th>
<th>Data availability</th>
<th>Conflict of interest</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Littell et al. (2022) [17]</td>
<td>Series 6 (not mentioned)</td>
<td>Adjustable elastic wristband made of rubber</td>
<td>GE Healthcare Nellcor Carescape Dinamap V100 pulse oximeter</td>
<td>508 patients (208 adults, 270 children) with congenital heart disease</td>
<td>Adults median: 32 (18 - 74); Children median: 10 (0 - 1 - 17)</td>
<td>Adult median: 97 (78 - 103); Children median: 97 (73 - 100)</td>
<td>0.813</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>None</td>
<td>Czech Technical University</td>
</tr>
<tr>
<td>Spaccarotella et al. (2022) [18]</td>
<td>Series 6 (not mentioned)</td>
<td>Not mentioned</td>
<td>Covidien Nellcor Portable SpO₂ Patient Monitoring System</td>
<td>64 patients &lt;3 years with weight &lt;13 kg ordered as part of their medical care</td>
<td>Mean: 7.2 (5.1 - 18)</td>
<td>Median: 98 (78 - 100)</td>
<td>0.76</td>
<td>Not mentioned</td>
<td>Uploaded to public repository</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Pipek et al. (2021) [19]</td>
<td>Series 6 (44 mm)</td>
<td>Not mentioned</td>
<td>Medtronic POD-2 Finger Oximeter, Multiblog Ox-06 Oximeter</td>
<td>100 adult patients (23 with COPD, 61 with interstitial lung disease, 16 healthy controls)</td>
<td>Mean: 94.4% (not mentioned)</td>
<td>Mean: 94.4% (not mentioned)</td>
<td>0.81</td>
<td>Assessed with Fitzpatrick skin type, and no information on distribution</td>
<td>Not mentioned</td>
<td>None</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>

**TABLE 2: Summary of extracted study parameters**

All publications declared that no conflict of interest was present. Two publications declared that no funding for the project had been received [16], and one publication declared a grant from the university as a funding source [17], while the remaining two publications did not mention funding [14,15].

Littell and colleagues were the only author group that uploaded data to a public repository [16]. Two of the remaining publications mentioned the data being available from the corresponding author upon reasonable request [17], while the others did not mention data availability [14,15].

Pipek et al. evaluated the performance of an Apple Watch Series 6 in 100 adult patients [15]. Measurements were conducted with the Apple Watch worn according to the manufacturer’s instructions and one pulse oximeter clip on the index and middle finger of the same hand. All values were noted at the moment the Apple Watch finished its measurement. Of the 100 patients, 61 had a history of interstitial lung disease, 23 had a history of COPD, and 16 were healthy controls. The mean SpO₂ measured with the conventional pulse oximeters was 94.4%. The mean SpO₂ measured with the Apple Watch was 95.9%. The Pearson correlation coefficient between watch and conventional oximeter measurements was 0.81, with limits of agreement of -2.7% and +4.1%. The Pearson correlation coefficient between the two conventional oximeter measurements was 0.995. The authors concluded that the Apple Watch Series 6 was a reliable way to obtain SpO₂ measurements in patients with lung diseases under controlled conditions.

Spaccarotella et al. evaluated the performance of an Apple Watch Series 6 in 257 adult patients [2]. Measurements were conducted according to the manufacturer’s instructions on the same arm within one minute of each other. All measurements were performed twice and then averaged. Of the 257 patients, 141 had a history of cardiovascular disease, 60 had a history of lung disease, and 56 were healthy controls. The mean age was 64 years. While the study did not mention the median SpO₂, the Pearson correlation coefficient was 0.89, with limits of agreement of -5.5% and +5.0%. The authors conclude that the watch could be used to assess SpO₂ in healthy patients, as well as in those with cardiovascular or lung disease.

Littell et al. evaluated the performance of an Apple Watch series 6 in 84 patients with an age of <23 years but weight >5 kg who had an ECG ordered as part of their routine clinical care at a pediatric cardiology unit [16]. For the SpO₂ measurement, patients were instructed to remain still for 15 seconds while the watch was placed on their wrists. In small children <10 kg, the watch was placed on other areas of the body with a
measurements should always be discussed and viewed in the context of the respective clinical scenario. However, the question if the Apple Watch is appropriate for spO2 measurements only one that allowed for the watch to be placed on other regions of the body besides the wrist which also poses the question of how much the accuracy of the measurements is affected [14].

Rafl et al. evaluated the performance of an Apple Watch Series 6 in 24 healthy adults that were exposed to hypoxic conditions [17]. The watch measurements were taken from the left wrist according to the manufacturer’s instructions with the conventional pulse oximeter being attached to the middle finger of the same hand. Two measurements were taken during a two-minute stabilization phase where participants breathed ambient air. In the following five minutes, the desaturation phase, participants breathed hypoxic air (12% O2), and spO2 was measured every 30 seconds. In the following stabilization phase, participants breathed ambient air again, and spO2 was measured every 30 seconds until it returned to normal values. Each participant underwent the procedure twice, with at least one hour in between, resulting in 642 spO2 measurements for the whole study. The mean age of participants was 24 years. The spO2 readings in the study range from 76 – 100% though it is unclear on which device this range is based. The authors report 95% limits of agreement of -5.8% and +5.9%. In patients and conditions where the conventional pulse oximeter measured >90% spO2, the Apple Watch measured on average 1% higher than the conventional device. The authors concluded that the spO2 measuring was sufficiently advanced for indicative measurement outside of the clinic.

While we found several studies that evaluated the spO2 measurements of the Apple Watch compared to a conventional, medical-grade pulse oximeter as the ground truth, differences in the testing procedures and the presentation of the resulting data made it difficult to compare between publications and prevented us from attempting a quantitative synthesis.

This applies in particular to the way the watch was worn during testing. The publication by Pätz et al. demonstrates that using an appropriate wristband has an impact on the percentage of failed measurements which also poses the question of how much the accuracy of the measurements is affected [14]. This question also applies to the study by Littell et al., which was the other study that included pediatric patients and the only one that allowed for the watch to be placed on other regions of the body besides the wrist [16]. It should be noted that Apple makes the spO2 measurements only available to users who have entered an age ≥ 18 years [18].

Four of the five studies drew positive conclusions regarding the usability of the spO2 measurements in their respective populations. This was mainly based on the fact that no study demonstrated a strong systematic bias of the watch measurements. However, the question if the Apple Watch is appropriate for spO2 measurements should always be discussed and viewed in the context of the respective clinical scenario.

The most common scenario today is a supposedly healthy individual using the Apple Watch as a wellness
tracker. The only way the watch could provide a benefit in this scenario is by alerting those people who are actually not completely healthy but have an underlying issue that causes a decreased oxygen saturation. The question remains as to whether the watch will be the thing that makes someone see a physician earlier than actual symptoms such as shortness of breath during physical activity.

For the vast majority of individuals using the watch as a wellness tracker, it will likely not provide a meaningful benefit in terms of health guidance and might even cause psychological distress to those who worry about outliers that the watch occasionally produces. Since the normal range for oxygen saturation is 95 - 100%, and the limits of agreement of the reviewed studies are in the +/- 2.7 to 5.9% range, it is to be expected that users will eventually be shown measurements outside of the normal range which, if occurring in isolation, should be no cause for concern.

Whether the watch is suited for remote monitoring in patients with established conditions that affect their oxygen saturation depends on the goal of that monitoring. Since the Apple Watch does not conduct permanent measurements but rather conducts measurements several times during the day, it is not suitable to detect sudden drops in spO\textsubscript{2} since those might happen outside of measurements. However, if the goal is to see how the oxygen saturation for a patient changes on average, especially over longer periods of time, the watch could be a good and convenient option.

A question that remains to be addressed is if the accuracy of the measurements depends on skin color, which has also been investigated for conventional pulse oximeters where differences in the risk of occult hypoxemia have been reported [19,20]. Only the study by Pipek and colleagues investigated this question. The other publications either described their populations as homogeneously caucasian (Rafl et al.) or white (Spaccarotella et al.) or didn’t cover the subject. While Pipek and colleagues did not find a significant difference in the measurements due to skin color assessed using the Fitzpatrick type, the distribution of skin color among their participants is unclear [15].

Many of the uncertainties mentioned previously could be addressed and mitigated by articles publishing patient-level data like the publication by Littell and colleagues [16]. Ideally, this data would not only contain spO\textsubscript{2} measurements but also additional information such as Fitzpatrick skin type, skin abnormalities, temperature, wrist circumference, etc.

Possible limitations at the review level include the fact that only articles with ‘Apple Watch’ or ‘Apple Smartwatch’ in their title or abstract were retrieved by the query, which could have led to articles using only terms like ‘wearable’ being missed. However, it seems appropriate to assume that any research using the Apple Watch is likely to use one of the terms somewhere in either title or the abstract. In addition, we also did not notice any papers in the discussion sections of the included manuscripts that were not returned by our query.

Only two databases were queried, but this limitation is mitigated by the fact that the majority of publications in the field of digital health appear in PubMed or Google Scholar-indexed journals.

Limitations at the study level include the heterogeneous measurement and reporting processes. In addition, all studies used conventional, medical-grade oximeters as the ground truth instead of arterial blood samples which represent the FDA’s standard when it comes to evaluating new spO\textsubscript{2} measuring devices [21]. Lastly, all studies used manually triggered measurements when the person was in an appropriate position at rest. Therefore, the performance of the Apple Watch when it does automatic measurements as a person is going about their day could not be assessed.

**Conclusions**

Our review suggests that the Apple Watch Series 6 does not show a strong systematic bias compared to conventional, medical-grade pulse oximeters. However, outliers appear to occur fairly often even though we could not determine a definitive frequency and should not cause concern in otherwise healthy individuals. The impact of skin color on measurement accuracy should be investigated.

**Additional Information**

**Disclosures**

**Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors declare that no financial support was received from any organization for the submitted work. **Financial relationships:** Paul Windisch declare(s) personal fees from dacadoo AG. PW reports consulting for dacadoo AG, a technology company that has business activity related to Apple products. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

**References**


