Using Apple Watch to Estimate Cardio Fitness with VO₂ max

May 2021
# Contents

Overview ............................................................................................................................................3
Introduction ........................................................................................................................................3
   Definition .......................................................................................................................................3
   Measurement and Estimation ............................................................................................................3
   Utility ............................................................................................................................................3
   Genetics .......................................................................................................................................4
   Interventions .................................................................................................................................4
Cardio Fitness on Apple Watch ...........................................................................................................4
   Metric Description ........................................................................................................................4
   Development ................................................................................................................................7
      Study Design .............................................................................................................................7
      Statistical Methods ...................................................................................................................7
   Results ........................................................................................................................................8
   Discussion ...................................................................................................................................10
Conclusions ....................................................................................................................................12
References .....................................................................................................................................12
Overview

With watchOS 7, Apple Watch Series 3 and later use an updated algorithm to estimate a user’s cardio fitness level as measured by VO$_2$ max, the maximum volume of oxygen an individual can extract from inhaled air. This update extends VO$_2$ max estimates to lower ranges while also expanding availability of this metric. Additionally, with watchOS 7.2, users can view how their cardio fitness level is classified based on their age group and sex in the Health app on iPhone, and they can receive a notification if it falls within the low range. This paper provides a detailed understanding of the capabilities of these features, including testing and validation.

Introduction

Definition

VO$_2$ max is the maximum volume of oxygen an individual can extract from inhaled air and consume through cellular metabolism. As such, VO$_2$ max is a good overall indicator of cardiorespiratory fitness (CRF) because it incorporates multiple organ systems and is influenced by a number of factors at various points along the pathway, from respiration to end-organ oxygen consumption. VO$_2$ max values are typically normalized for body mass and reported as milliliters of oxygen per kilogram of body mass in one minute (ml/kg/min). They normally decline with increasing age and, at a population level, differ across biological sexes.

Measurement and Estimation

VO$_2$ max is measured during cardiopulmonary exercise testing (CPET) — a procedure in which an individual is asked to ride a stationary bike or ambulate on a treadmill at increasing intensity levels while wearing a face mask, which allows for the direct measurement of oxygen in inhaled and exhaled air. In most cases, the volume of oxygen that individuals consume during testing plateaus even though their work increases, and this plateau or VO$_2$ peak is assumed to be and referred to as the VO$_2$ max, despite a lack of certainty that a true maximum has been achieved.

In practice, VO$_2$ max or CRF is more commonly estimated from measurements taken during submaximal exertion because these tests are less expensive and less time intensive than maximal CPET, they require less exertion from and are more comfortable for the subject, and significant evidence exists to derive VO$_2$ max from submaximal exertion.

Utility

CRF, as measured by VO$_2$ max or the closely related metabolic equivalent (MET) — with 1 MET = ~3.5 ml/kg/min — has been repeatedly shown over the last 30 years to be a predictor of all-cause and cardiovascular mortality, as well as cardiovascular events for men and women. VO$_2$ max was independent of and more predictive than well-known risk factors for cardiovascular and all-cause mortality such as hypertension, obesity, and hypercholesterolemia.

Because of this prognostic utility, members of the medical and scientific community have advocated for inclusion of measures of CRF in routine medical practice as adjuncts to or even in place of traditional risk models such as Framingham. This predictive utility also applies outside the general population to disease-specific cohorts, such as individuals with heart failure, and in clinical decision-making related to specific events such as perioperative management and referral for cardiac rehabilitation. In response to these and other demonstrations of utility, the American Heart Association (AHA) in 2016 advocated for the assessment of CRF on a more routine basis, making a case for fitness as a vital sign.
Genetics
Genetics are strongly correlated with an individual's VO$_2$ max and changes in VO$_2$ max in response to exercise. At baseline, genetic factors are believed to determine approximately 50 to 70 percent of the VO$_2$ max differences observed between individuals$^{16,19}$ and approximately 20 to 60 percent of the variation in VO$_2$ max improvements in response to exercise training.$^5,20$

Interventions
Improving or maintaining VO$_2$ max over time is strongly associated with decreased mortality. In a study of over 500 men followed for 11 years, Laukkanen et al. found that for every 1 ml/kg/min increase in VO$_2$ max, risk of death decreased by 9 percent.$^{21}$ At a study level, high-intensity interval training yields the greatest improvements in VO$_2$ max.$^{22,23,24}$ Over the course of programs that ranged from 6 to 12 weeks, improvements in VO$_2$ max were generally approximately 5 to 10 percent (in ml/kg/min). It's important to note that decreases in VO$_2$ max with decreased activity or inactivity have been reported to be of similar or greater magnitude (up to a 27 percent decline) over far shorter time periods (2 to 3 weeks).$^{25,26}$ Increasing physical activity in the absence of improvements to VO$_2$ max doesn't appear to confer the same survival benefit in those whose VO$_2$ max increases compared with those whose VO$_2$ max doesn't.$^{27}$

Cardio Fitness on Apple Watch
This paper describes the development and validation of the cardio fitness metric, an estimate of VO$_2$ max using Apple Watch. The intended audience of this paper is researchers, healthcare providers, and developers who may be interested in using this estimation in their work, as well as customers who would like to know more about VO$_2$ max and how it's measured and validated using Apple Watch. Additional information on how to set up and view VO$_2$ max estimates for Apple Watch can be found at support.apple.com/en-us/HT211856.

Metric Description
Cardio fitness on Apple Watch is an estimation of a user's VO$_2$ max in ml/kg/min, made based on measuring a user's heart rate response to physical activity. Updates to the algorithm used to estimate VO$_2$ max in watchOS 7 extend estimates to lower ranges of cardio fitness (14 to 60 ml/kg/min) for users with Apple Watch Series 3 or later. A view of VO$_2$ max in the Health app in iOS 14 under Cardio Fitness is shown in figure 1. A VO$_2$ max value may be generated after walking, running, or hiking outdoors on relatively flat ground (that is, a grade of less than 5 percent incline or decline) with adequate GPS, heart rate signal quality, and exertion (an approximate increase of 30 percent of the range from resting heart rate to max). A user's first such workout won't generate an estimate, and a user needs to have worn Apple Watch for one day before a first estimate can be generated.
These estimates of VO\textsubscript{2} max are based on submaximal predictions of VO\textsubscript{2} max rather than peak VO\textsubscript{2}. As such, users don't need to achieve peak heart rate to receive an estimate; however, a notion of peak heart rate is needed. Because of this, users who take medications that may reduce their peak heart rate can indicate in Health Details in the Health app that they're taking this medication to enable more accurate VO\textsubscript{2} max estimates (see figure 2).

With a feature introduced in iOS 14.3, users ages 20 and above have the option of being notified if their cardio fitness level, as measured by estimated VO\textsubscript{2} max, is consistently and confidently low enough to suggest risk of long-term health issues or current limitation in daily activities. For users ages 20 to 59, this threshold for notification is the lowest quintile for sex and age by decade as determined by the Fitness Registry and Importance of Exercise National Database;\textsuperscript{2} for users ages 60 and above, absolute VO\textsubscript{2} max thresholds of 18 and 15 ml/kg/min are used for males and females, respectively, based on data suggesting these are thresholds for independent living at the extremes of age in both sexes.\textsuperscript{28} Users who want to receive Low Cardio Fitness notifications must opt in, which requires completing an onboarding experience in the Health app that describes the feature; collects information like age, sex, and relevant medications needed to provide an accurate alert; outlines factors that can lower your cardio fitness; and provides optional education content describing the importance of VO\textsubscript{2} max and potential causes of a notification (see figure 3).
Using Apple Watch to Estimate Cardio Fitness with VO2 max.  

May 2021

Figure 2: Medications that affect heart rate can be noted in Health Details in the Health app in iOS 14

Figure 3: Onboarding for Low Cardio Fitness notifications in the Health app in iOS 14
Development

Study Design
Apple collected data for the design and validation of the VO2 max metric across multiple studies, approved by an institutional review board (IRB), involving participants who consented to the collection and use of their data for this purpose.

Study participants completed VO2 max tests, VO2 submax tests, or both — referred to as cardiopulmonary exercise testing, or CPET — while wearing Apple Watch Series 4. A variety of CPET test protocols were used, including both treadmill and cycle ergometers. Each participant completed up to six CPETs over the course of the studies, with at least 10 days between consecutive tests to ensure participants had sufficient time to rest between tests and adequate periods for data collection before and after individual CPETs. CPET data was used to verify that the protocol was completed correctly and that participants reached at least 60 percent of the predicted maximum heart rate. Tests for which there were a gas exchange abnormality, low-quality heart rate signal, detected arrhythmia, reported pain, or biomechanical inefficiency were excluded from all further analyses. Tests that passed the verification steps were used in algorithm development. To obtain the reference VO2 max of each participant, linear projections were made using heart rate and VO2 in the submax range to determine VO2 max based on age-predicted maximum heart rate. Predicted maximum heart rate was lowered for users on heart rate–limiting medications, such as beta blockers, according to published predictions.

In addition to wearing Apple Watch during proctored CPETs, participants wore Apple Watch and carried iPhone during their day-to-day activities throughout the course of the studies. These activities included workouts logged by participants. Data from a variety of Apple Watch sensors (photoplethysmograph, accelerometer, gyroscope, barometer, and GPS) was collected during this period and used in VO2 max algorithm design.

A subset of study participants was withheld from all algorithm design data to verify algorithm accuracy and guard against overfitting. Algorithm performance was calculated by comparing the last valid VO2 max estimate on Apple Watch with the average of submaximal projections from all curated CPETs for each participant unless otherwise specified.

Statistical Methods
The validity of VO2 max on Apple Watch was computed as the mean and standard deviation of errors between the last valid mean VO2 max estimate on Apple Watch and the mean submaximal VO2 max projection from all curated CPETs for each participant. Reliability, reported as intraclass correlation coefficient (ICC), was evaluated by calculating the absolute agreement per participant between the last valid VO2 max estimate on Apple Watch and a previous VO2 max on Apple Watch estimated at least 28 days prior. The consistency of VO2 max on Apple Watch is expressed as the median and 90th percentile standard deviation per participant of all VO2 max estimates on Apple Watch for participants who had at least five estimates. Finally, the availability of VO2 max on Apple Watch is computed in two ways: as the percentage of all outdoor pedestrian workouts longer than 5.75 minutes from all participants yielding a VO2 max estimate on Apple Watch, and as the percentage of participants who completed at least 10 outdoor pedestrian workouts longer than 5.75 minutes who received at least one VO2 max estimate on Apple Watch after 10 workouts.
Results

Baseline characteristics of the participants whose data was used for design and validation are summarized in table 1.

Table 1. Participant characteristics

<table>
<thead>
<tr>
<th></th>
<th>Design (N = 534)</th>
<th>Validation (N = 221)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender — number (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>191 (36)</td>
<td>94 (43)</td>
</tr>
<tr>
<td>Male</td>
<td>343 (64)</td>
<td>127 (57)</td>
</tr>
<tr>
<td><em><em>Age — years</em> (mean ± SD)</em>*</td>
<td>53 ± 18</td>
<td>55 ± 17</td>
</tr>
<tr>
<td><strong>Age distribution — number (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45 years</td>
<td>207 (39)</td>
<td>74 (33)</td>
</tr>
<tr>
<td>45–54 years</td>
<td>67 (13)</td>
<td>26 (12)</td>
</tr>
<tr>
<td>55–65 years</td>
<td>57 (11)</td>
<td>36 (16)</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td>203 (38)</td>
<td>85 (38)</td>
</tr>
<tr>
<td><strong>Reference VO\textsubscript{2} max — ml/kg/min (mean ± SD)</strong></td>
<td>31.7 ± 10.6</td>
<td>29.7 ± 10.5</td>
</tr>
<tr>
<td><strong>Length of observation — days (mean ± SD)</strong></td>
<td>441 ± 137</td>
<td>390 ± 138</td>
</tr>
<tr>
<td><strong>Comorbidities — number (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td>51 (10)</td>
<td>17 (8)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>38 (7)</td>
<td>23 (10)</td>
</tr>
<tr>
<td>History of stroke</td>
<td>9 (2)</td>
<td>5 (2)</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>41 (8)</td>
<td>24 (11)</td>
</tr>
<tr>
<td>History of myocardial infarction</td>
<td>34 (6)</td>
<td>16 (7)</td>
</tr>
<tr>
<td>COPD</td>
<td>4 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>10 (2)</td>
<td>5 (2)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>121 (22)</td>
<td>47 (21)</td>
</tr>
<tr>
<td><strong>Smoking status (cigarettes) — number (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>5 (1)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Former smoker</td>
<td>63 (12)</td>
<td>37 (17)</td>
</tr>
<tr>
<td>Never smoked</td>
<td>300 (56)</td>
<td>129 (58)</td>
</tr>
<tr>
<td>Smoking status unknown</td>
<td>166 (31)</td>
<td>54 (24)</td>
</tr>
<tr>
<td><strong>BMI category — number (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight (BMI &lt; 18.5)</td>
<td>1 (&lt;1)</td>
<td>2 (&lt;1)</td>
</tr>
<tr>
<td>Normal weight (18.5 ≤ BMI &lt; 25.0)</td>
<td>215 (40)</td>
<td>99 (45)</td>
</tr>
<tr>
<td>Overweight (25.0 ≤ BMI &lt; 30.0)</td>
<td>220 (41)</td>
<td>77 (35)</td>
</tr>
<tr>
<td>Obese (BMI ≥ 30.0)</td>
<td>98 (18)</td>
<td>43 (19)</td>
</tr>
</tbody>
</table>

*Based on year of birth.
Algorithm performance for the design and validation data sets is reported in table 2. A plot of reference (average CPET-derived VO$_2$ max per user compared with final Apple Watch–estimated VO$_2$ max) for the design and validation participants is shown in figure 4. Algorithm performance was assessed on data collected during workouts. For a subset of participants (132 design and 62 validation), VO$_2$ max was also estimated outside of workouts during periods of outdoor walking to assess the ability to estimate VO$_2$ max when a workout hasn’t been initiated on Apple Watch. In these users, nonworkout estimates were an average of 0.32 ml/kg/min higher than workout estimates in the design group. No significant difference was detected between workout and nonworkout estimates in the validation group.

![Figure 4: Reference compared with estimated VO$_2$ max (ml/kg/min) for participants in the design and validation sets](image-url)
Assessment of CRF using VO₂ max has received increasing attention as a means of risk stratification, with some advocating for its consideration as a vital sign. In practice and despite demonstrated benefits in application, objective measurement of CRF using CPET remains infrequent in part due to expense, participant burden, and limited acceptance as standard of care across multiple specialties. Accurate estimation of a user’s VO₂ max through wearable technology could expand CRF screening to a large segment of the population at lower cost, and it may allow for remote monitoring of patients between clinical visits in programs such as cardiac rehabilitation. Such accurate and available estimates of VO₂ max may also be used to guide risk stratification and response to programs designed to reduce risk, such as preoperative assessment and rehabilitation.

The improved algorithm for VO₂ max estimation on Apple Watch described here was designed for and validated in a population with reference VO₂ max covering a wide range of cardio fitness levels, as shown in figure 4. Nearly half the study participants were over the age of 55, and approximately 10 percent had known coronary artery disease. Racial and ethnic diversity in the study population didn’t approximate the U.S. population; however, heart rate — a key input to VO₂ max estimates on Apple Watch — has been shown to be consistently accurate across multiple skin tones in both internal and external studies.

Extension of VO₂ max estimates to lower ranges in watchOS 7 combined with estimation outside of workouts increases the availability of this metric for individuals with low cardio fitness. Over 90 percent of participants with at least one outdoor walk, outdoor run, or hiking workout longer than three minutes and tracked using the Workout app received at least one estimate of VO₂ max on Apple Watch. Increasing the number of outdoor pedestrian workouts will increase the likelihood and accuracy of a VO₂ max estimate on Apple Watch.

### Discussion

**Table 2. VO₂ max performance**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Design (N = 534)</th>
<th>Validation (N = 221)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validity</strong></td>
<td>Error (mean estimated VO₂ max - mean projected submax reference) — ml/kg/min (mean ± SD)</td>
<td>1.2 ± 4.4</td>
<td>1.4 ± 4.7</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>ICC A-1 comparison of last VO₂ max estimate using data and metadata from only that session to a VO₂ max estimate more than 28 days previous using only data and metadata from that session — ICC [confidence interval]</td>
<td>0.89 [0.86, 0.91]</td>
<td>0.86 [0.80, 0.90]</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td>SD of pVO₂ max per user — ml/kg/min (% of pVO₂ max)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>1.2 (3.7%)</td>
<td>1.2 (3.4%)</td>
</tr>
<tr>
<td></td>
<td>90th percentile</td>
<td>2.6 (7.6%)</td>
<td>2.6 (7.2%)</td>
</tr>
<tr>
<td><strong>Availability</strong></td>
<td>Percentage of outdoor pedestrian workouts longer than 5.75 minutes receiving an estimate</td>
<td>79%</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>Percentage of participants who completed at least 10 outdoor pedestrian workouts longer than 5.75 minutes and received at least one estimate in their first 10 workouts</td>
<td>93%</td>
<td>93%</td>
</tr>
</tbody>
</table>
VO₂ max estimation by Apple Watch is accurate and reliable relative to commonly used methods of measuring VO₂ max, with an average error of less than 1 MET and an ICC of more than 0.85. The accuracy of VO₂ max on Apple Watch approaches the accuracy of the reference; submaximal exercise testing protocols have previously been measured to have approximately zero mean error and a standard error of 1 MET. In terms of test-retest reliability, VO₂ max on Apple Watch has an ICC of 0.87 in the validation data compared with 0.75 in submaximal treadmill testing.

With the new algorithm, VO₂ max estimates for users who take heart rate–limiting medications — such as beta blockers and calcium channel blockers — and report this in the Health app on iPhone paired to Apple Watch should be more accurate relative to estimates made in previous versions of iOS and watchOS. The handling of these medications doesn’t differentiate dosage, cardioselectivity, or intrinsic sympathomimetic activity of some beta blockers, all of which could potentially be meaningful inputs but have been omitted in favor of usability. With this approach, the estimated error for users in the validation cohort who take beta blockers and calcium channel blockers decreased from 11.8 ± 4.0 ml/kg/min to 1.6 ± 3.1 ml/kg/min when the Health Details settings appropriately reflected their medication usage. Users who take heart rate–limiting medications and don’t enter this information will receive higher-than-actual estimates; those who take low or as-needed doses of these medications that don’t consistently reduce max heart rate (for example, propranolol for performance anxiety) will likely receive more accurate estimates if they don’t enter this information. Given the highly prevalent usage of these medications, accounting for them appropriately is critical to accurately estimating VO₂ max, especially for older users.

In some conditions, a user’s VO₂ max estimate may be inaccurate. Users with an incorrect age, sex, or weight entered in the Health app may have consistently inaccurate VO₂ max estimates. Normal physiological changes associated with pregnancy may lead to inaccurate estimates. Individual estimates may be low if sensor data is recorded during behaviors that increase the user’s work in ways that Apple Watch can’t accurately detect. Common instances of such behaviors include carrying significant weight beyond body weight, such as a heavy backpack or child, and walking or running on ground that increases the user’s work, such as sand. Similarly, using an assistive device or pushing a stroller may decrease availability or accuracy of VO₂ max estimates on Apple Watch. Factors that increase heart rate, such as dehydration, caffeine intake, extreme heat, or recent transition to high altitudes may also lead to underestimates. VO₂ max accuracy on Apple Watch can be increased by performing frequent outdoor pedestrian workouts, by achieving higher exertion during workouts, and by wearing Apple Watch consistently throughout the day beyond typical workout sessions.

Users with chronotropic incompetence — a condition where heart rate doesn’t appropriately increase to compensate for demand — may receive overestimates of VO₂ max. Chronotropic incompetence is primarily associated with heart failure, which occurs in approximately 30 to 80 percent (depending on diagnostic criteria) of patients with the condition. It has also been linked to significant proportions of patients with chronic obstructive pulmonary disease (COPD), lupus, and other autoimmune conditions.

In addition to chronotropic incompetence, other medical conditions can also decrease the accuracy of VO₂ max estimates on Apple Watch. These include medical conditions or devices that decouple heart rate from movement or exercise (for example, pain, arrhythmias, pacemakers, or cardiac-assist devices); medical conditions that severely limit exercise tolerance, preventing patients from reaching heart rates close to their predicted maximum heart rate (for example, peripheral arterial disease); and medical conditions that significantly increase the difficulty of ambulation, such as skeletal or neuromuscular conditions causing gait inefficiency (for example, multiple sclerosis or cerebral palsy).
Conclusions

With watchOS 7 on Apple Watch Series 3 and later, VO\textsubscript{2} max estimates have been expanded to lower ranges of cardio fitness while providing users with the option to receive a notification if their cardio fitness level is low for their age and sex. This extended range, along with increased estimate availability and the option for users taking heart rate–limiting medications to receive more accurate estimates than were previously available, may increase researchers’ and clinicians’ ability to use this metric for tracking fitness in older adults and in the presence of comorbid health conditions.

References


© 2021 Apple Inc. All rights reserved. Apple, the Apple logo, Apple Watch, iPhone, and watchOS are trademarks of Apple Inc., registered in the U.S. and other countries. iOS is a trademark or registered trademark of Cisco in the U.S. and other countries and is used under license. Other product and company names mentioned herein may be trademarks of their respective companies.