

# Criterion validity of a newly developed Apple Watch app ('MVPA') compared to the native Apple Watch 'activity' app for measuring criterion moderate intensity physical activity

DIGITAL HEALTH  
Volume 11: 1–13  
© The Author(s) 2025  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076251326225  
journals.sagepub.com/home/dhj



Ashley Warner<sup>1</sup> , Natalie Vanicek<sup>2</sup> , Amanda Benson<sup>3</sup> ,  
Tony Myers<sup>4</sup>  and Grant Abt<sup>2</sup> 

## Abstract

**Introduction:** Many people fail to meet physical activity guidelines. One possible solution is wearable technology. Yet it is unclear if popular devices such as the Apple Watch can accurately measure intensity, which is a cornerstone of physical activity guidelines. We developed a bespoke Apple Watch app based on the use of relative intensity (%HRR), with the aim to determine if the bespoke app measures 'moderate' intensity more accurately compared to the native Apple Watch Activity app, using % oxygen consumption reserve (%VO<sub>2</sub>R) as the criterion.

**Methods:** Seventy-four participants (18–65 years) attended the laboratory twice. During Visit 1, they completed medical screening, and measurements of resting heart rate, maximal oxygen consumption, and maximal heart rate on a motorised treadmill. During Visit 2, participants completed 5-minute treadmill bouts starting at 3.5 km.h<sup>-1</sup>, increasing by 0.5 km.h<sup>-1</sup> until the bespoke app recorded 3 minutes at ≥40%HRR, and the native app recorded 5 minutes of 'exercise'. Oxygen consumption and heart rate were recorded. Bayesian posterior distributions were used to compare the two apps.

**Results:** At the walking speed that the native app recorded exercise, the mean (95% HDI) %VO<sub>2</sub>R was 33 (31–36)%, which is below moderate relative intensity (40%HRR). This is compared to a mean (95% HDI) of 43 (40–44)% for our bespoke app.

**Conclusion:** The bespoke app measured relative moderate intensity more accurately compared to the native app when compared to the %VO<sub>2</sub>R criterion. Exercise guidelines and wearable devices should incorporate relative measures of physical activity to better individualise monitoring and prescription.

## Keywords

Wearable technology, wearables, relative intensity, physical activity, exercise

Submission date: 29 October 2024; Acceptance date: 20 February 2025

## Introduction

There is a substantial body of evidence demonstrating that physical activity reduces morbidity and mortality risk.<sup>1–5</sup> This has led government organisations like the UK National Health Services (NHS) to advise the public on methods for engaging in physical activity.<sup>6</sup> In its guidance, the NHS suggests that participating in 'moderate' intensity (or above) physical activity leads to health benefits and advises people on how they can gauge 'moderate' intensity using subjective measures. The NHS suggests that moderate intensity "...will raise your heart rate, and make you

<sup>1</sup>School of Health and Life Sciences, University of the West of Scotland, Glasgow, UK

<sup>2</sup>School of Sport, Exercise and Rehabilitation Sciences, University of Hull, Hull, UK

<sup>3</sup>School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

<sup>4</sup>Department of Social Science, Sport and Business, Birmingham Newman University, Birmingham, UK

These authors contributed equally to this paper.

### Corresponding author:

Grant Abt, School of Sport, Exercise and Rehabilitation Sciences, University of Hull, Cottingham Rd, Hull HU6 7RX, UK.

Email: gabt@mac.com



breathe faster and feel warmer. One way to tell if you're working at a moderate intensity level is if you can still talk, but not sing".<sup>6</sup> Given these public facing recommendations it is important that they are examined for accuracy via scientific investigation.

As an alternative to these subjective methods of gauging 'moderate' intensity, the ACSM's guidelines for exercise testing and prescription<sup>7</sup> highlight a range of objective measures of exercise intensity, including both relative (% of heart rate reserve [%HRR], % of maximal heart rate [%HRmax], % of maximal oxygen consumption [%VO<sub>2max</sub>], % of oxygen consumption reserve [%VO<sub>2R</sub>]) and absolute (metabolic equivalents [METs]) methods. Relative intensity is described as the exercise intensity relative to the user's physiological capacity<sup>8</sup> and absolute intensity as the energy required to perform an activity.<sup>8</sup> However, there is a large disparity between relative and absolute measures of physical activity intensity.<sup>8-10</sup> For example, Warner et al.<sup>10</sup> conducted a systematic review with Bayesian meta-regression of secondary data to examine the agreement between relative and absolute intensity during walking and reported agreement in only 43% of trials. Moreover, a large variation in %HRR at a given MET was reported. For example, at 3 METs, which is considered the lower bound of moderate intensity, the model predicted that %HRR was 33%. This is lower than the 40% HRR considered as the lower bound of moderate intensity when using %HRR.<sup>7</sup> However, the 95% credible interval ranged from 18%HRR to 57%HRR, which spans both 'light' and 'moderate' intensity classifications and is very close to the lower bound of 'vigorous' intensity (60% HRR). As such, there is considered heterogeneity in relative intensity at a fixed value of absolute intensity.

Cadence, which is the number of steps taken per minute, is another measure that can be used to guide walking intensity. Although some researchers suggest that a walking cadence  $\geq 100$  steps.min<sup>-1</sup> is an appropriate heuristic value that will allow most people to reach absolute moderate intensity (3 METs),<sup>11-14</sup> recent evidence suggests this 'one-size-fits-all' simplistic approach may be misguided.<sup>15-17</sup> Recent research using relative intensity (%VO<sub>2R</sub>) has highlighted the impact both cardiorespiratory fitness and health status have on the cadence required to reach moderate and vigorous intensity.<sup>15</sup> Cardiorespiratory fitness plays an integral role in the required walking cadence to reach moderate intensity, which suggests that an individualised prescription of exercise intensity, would be a more appropriate measure of physical activity intensity compared to an absolute measure like METs.

As relative measures of physical activity use percentage thresholds of physiological parameters, they are likely to be more individualised and thus more accurate on an inter-individual basis compared to METs.<sup>10</sup> However, the main limitation of relative measures and prescription of individualised physical activity has been one of accessibility, or lack of, in a real-world setting. Relative intensity measures

routinely collected in laboratory settings, such as oxygen consumption or ventilatory threshold, are currently not feasible outside of a laboratory, mostly due to the inability of wearable devices to collect these measures.<sup>18</sup> However, heart rate measurement, and more specifically HRR, is now possible using a wearable device. The main advantage of using HRR is that it is highly associated with oxygen consumption reserve, which is a criterion measure of exercise intensity.<sup>19</sup> With the ever-developing wearables market,<sup>21</sup> the measurement of heart rate is rapidly becoming mainstream. Moreover, the accuracy of heart rate measurement at typical walking speeds is very good.<sup>22,23</sup> There are, however, some limitations that have been reported with using a wearable device to measure physical activity, with previous research indicating that wearable devices may overestimate physical activity intensity.<sup>24</sup> Yet this effect is usually due to the accelerometer in wearable devices using an external measure of absolute physical activity intensity, and therefore not individualised to the user. Abt et al.<sup>15</sup> reported that the exercise intensity at which the native Apple Watch activity app recognised that the user was 'exercising' was  $\sim 30\%$  VO<sub>2R</sub>, suggesting that this device was not accurate at guiding the user to exercising at criterion moderate intensity (40% VO<sub>2R</sub>). The underestimation of moderate relative intensity would lead to an overestimation of moderate-to-vigorous physical activity (MVPA) minutes throughout the day. This is important, because Apple suggest that meeting the default 30 minutes of 'exercise' per day (as measured by the Apple Watch) is associated with significant health benefits.<sup>25</sup>

Although the measurement of heart rate via a wrist worn wearable device now offers a viable option for measuring exercise intensity for millions of people,<sup>21</sup> it would be preferable to use a relative measure of intensity, such as HRR to do so. Heart rate reserve has been reported to be comparable to VO<sub>2R</sub>,<sup>19</sup> and therefore offers an accurate, yet feasible method for measuring relative intensity at a population level.<sup>19</sup> As the development of wearable devices grows, it is important for researchers, exercise professionals, and consumers to understand how relative, individualised measures of MVPA can be used in everyday exercise monitoring, prescription, and goal setting.

Therefore, the main aim of this study was to build on the results of Abt et al.<sup>26</sup> to examine whether a relative measure of exercise intensity, %HRR, can measure criterion exercise intensity more accurately compared to the native Apple Watch Activity app. We did this through the creation of a bespoke Apple Watch app (herein called the 'bespoke MVPA app') which measures %HRR during exercise.

## Methods

### Study preregistration

The study design, including a statistical analysis plan, was preregistered prior to data collection, and can be found at

**Table 1.** Proportion of female and male participants, together with the mean (SD) participant demographic data.

	Sex	Age (years)	Body mass (kg)	Stature (cm)	BMI (kg.m <sup>2</sup> )	Leg Length (cm)
Mean (SD)	43% F, 57% M	30.7 (10.3)	74.06 (14.8)	172.7 (10.1)	24.8 (4.4)	89.2 (6.2)
Range		40–58	45.4–131.8	153.0–192.7	16.44–39.14	72.6–103.6

<https://osf.io/j39fa>. STROBE guidelines were adhered to throughout the reporting of this observational study.<sup>27</sup> All raw data and coding script used to analyse the data can be found at <https://osf.io/3uejs/files/osfstorage>.

### Ethics

Ethics approval was obtained from the Faculty of Health Sciences Research Ethics Committee at the University of Hull, East Riding of Yorkshire, United Kingdom (REF – FHS230) prior to recruitment of participants.

### Study design

This was an observational study, with participants attending the laboratory on two occasions. The first visit included medical screening, obtaining written informed consent, resting measures, and measurement of maximal oxygen consumption and heart rate. The second visit was the main experimental trial.

### Participants

Participants were healthy adults aged between 18 and 65 years of age (see Table 1 for demographic data) and provided written informed consent to participate in the study. Participants were excluded if: 1) they were classified as moderate or high-risk according to the ACSM risk classification criteria,<sup>28</sup> 2) unable to walk on a motorised treadmill, 3) had a BMI  $\geq$  40), 4) were currently prescribed medication that altered their heart rate response to exercise (e.g., beta-blockers), or 5) had an injury or disability that altered their gait. Participants were recruited from the University and local community through written promotional material and personal communication.

### Sample size

As outlined in the study pre-registration, a minimum sample size of 50 participants was required, but with final sample size was determined by Bayesian sequential testing until a particular level of precision was attained or, if not achieved, by a stopping rule of 12 months from the beginning of data collection as time allocated for data collection was limited to this time period. Bayesian statistical methods are particularly suited to

sequential estimation where data come in a sample at a time and the posterior distribution is updated at each time point.<sup>29</sup> The precision required to stop data collection was when the width of the 95% HDI (high density interval)  $\leq$  standard deviation\*0.6. Using this criterion, a total of 74 participants completed all trials.

### Setting

Participants were recruited from May 2021 for a 12-month period. All laboratory trials were completed on campus at the University of Hull.

**Visit 1.** On initial visit to the laboratory, participants completed the following protocols in the sequence highlighted below.

### Preliminary measures

On entry to the laboratory, participants' nude body mass was measured to the nearest 0.1 kg using digital scales (WB-100MA Mark 3, Tanita Corporation, Tokyo, Japan). A wall-mounted stadiometer (Holtain Ltd, Dyfed, Wales, UK) was used to measure stretch stature<sup>30</sup> and leg length measurement taken from the anterior superior iliac spine to lateral malleolus.<sup>31</sup>

### Resting oxygen consumption and heart rate

Criterion resting VO<sub>2</sub> and heart rate were measured in a temperature-controlled laboratory. Participants lay in a supine position for 30 minutes with a cushion placed under their head. All other laboratory activity ceased, lights turned off, with participants were instructed to relax but not to go to sleep (closing of eyes was permitted). Breath-by-breath oxygen consumption (Cortex Metalyser 3B, GmbH, Germany), and heart rate were recorded continuously. Prior to the commencement of testing, the Cortex Metalyser was calibrated using a 3-point calibration: room air, a known gas concentration of oxygen and carbon dioxide, and volume calibrated using a 3-litre syringe in line with manufacturer guidelines.

The initial 10 minutes and final five minutes of oxygen consumption and heart rate data were discarded to allow for habituation, and expectation effects, respectively. The remaining 15 minutes of oxygen consumption and heart

rate data were used for the analysis of resting oxygen consumption and heart rate. During minutes 15–20, the bespoke MVPA app recorded a resting heart rate value using its built-in resting heart rate protocol.

### Maximal oxygen consumption

A graded exercise test completed on a motorised treadmill (h/p/cosmos, Pulsar, Nussdorf-Traunstein, Germany), using a step protocol commencing at 3 km.h<sup>-1</sup> and a 1% gradient, was used to measure maximal oxygen consumption. Treadmill speed was increased by 0.5 km.h<sup>-1</sup> every 30 seconds until running economy degraded to a point that the participant's posture and technique were affected, at this point, incline then increased every 30 seconds by 0.5% until volitional exhaustion. Oxygen consumption was measured continuously using the same system described for the resting oxygen consumption protocol. Maximal oxygen consumption criteria were: 1) a plateau in VO<sub>2</sub> despite an increase in workload, 2) an RER of 1.05, and 3) a minimum category ratio RPE of 9.<sup>32</sup>

*Visit 2.* Participants completed the experimental trial outlined below on their second visit to the laboratory.

### Experimental trial

Prior to completion of the main exercise trial, participants were instructed to avoid vigorous exercise for 24 hours, maintain their normal diet, and avoid caffeinated drinks for three hours. The main trial involved participants completing 5-minute exercise bouts on a motorised treadmill starting at a walking speed of 3.5 km.h<sup>-1</sup> and a gradient of 1% while wearing a Series 5 Apple Watch with the native Apple Watch activity app and bespoke app installed. Each subsequent bout increased speed by 0.5 km.h<sup>-1</sup> until three minutes of the 5-minute bout were recorded as ≥ 40% HRR by the bespoke MVPA app and all five minutes by the native Apple Watch Activity app. During each 5-minute exercise period, participants were told to maintain their normal walking gait and were not permitted to hold onto the handrails. Oxygen consumption (Cortex Metalysers 3B, GmbH, Germany) and heart rate, via ECG chest strap (Polar T31, Polar Electro, OY, Finland), were recorded. Between each walking bout, participants had five minutes of seated rest. Once the treadmill bout had terminated and the treadmill belt was stationary, a chair was placed on the treadmill and participants were instructed to sit, motionless, with their hands placed on the treadmill handrail in order to ensure no movement contributed to the measurement of 'exercise' on the native Apple Watch Activity app. Five minutes of seated rest were provided to ensure the native Apple Watch Activity app had adequate time to update the green 'exercise' ring in line with previously reported methods.<sup>26</sup> The green 'exercise

ring' is the Apple Watch's visual monitoring tool that displays daily minutes of completed exercise in the native Activity app.

### Development of the bespoke MVPA app

The bespoke MVPA Apple Watch app was designed to quantify MVPA and to overcome the barriers and limitations of the native Apple Watch Activity app. The bespoke MVPA app was built in Core app C and Unity language, while the Apple Watch connectivity to Apple iPhone was built in Swift and Objective C. Previous studies have indicated that the native Apple Watch Activity app underestimates the intensity required to achieve MVPA by ~25% when compared to directly measured %VO<sub>2</sub>R.<sup>22,26</sup> The bespoke MVPA app has been designed to overcome this limitation, using %HRR to quantify moderate intensity. The bespoke MVPA app uses %HRR because it has been reported to be similar to %VO<sub>2</sub>R,<sup>19</sup> and %HRR can be calculated easily by the Apple Watch using data already available from HealthKit (resting HR and estimated maximal heart rate). A fifth generation Apple Watch (Series 5) running watchOS 7.1.1, or later update, was used to determine moderate intensity exercise. The bespoke MVPA app used the ACSM definition of moderate intensity, defined as 40% to 59% of HRR or VO<sub>2</sub>R.<sup>28</sup> As VO<sub>2</sub>R data are not available through the Watch photoplethysmography (PPG) sensor, HRR was used to substitute for this, such that moderate intensity physical activity was defined as 40–59% HRR. The equation used to calculate moderate intensity within the bespoke MVPA app was:

$$\begin{aligned} \text{Target HRR} &= (\text{HR}_{\text{MAX}} - \text{HR}_{\text{REST}}) \\ &\times \text{exercise intensity (e.g., 0.4)} \\ &+ \text{HR}_{\text{REST}}. \end{aligned}$$

### MVPA activity log

Every second spent at an intensity at or above 40% HRR counted towards the accumulation of an MVPA minute. The accumulated total was displayed on the Apple Watch, via the bespoke MVPA app interface and on the companion MVPA iPhone app.

## Statistical analysis

### Statistical models

Data from the native Apple Watch Activity app were compared to data from the bespoke MVPA app, and to the criterion measure of intensity, %VO<sub>2</sub>R. Data from the stage, when at least three minutes of the 5-minute bout met the moderate intensity criterion (40–59% HRR) measured by the MVPA app, were compared with data from the stage

when all five minutes of the 5-minute bout had been classified as ‘exercise’ by the native Apple Watch Activity app. %VO<sub>2</sub>R and %HRR were analysed separately, and different groups of models for these two dependent variables were produced. Posterior distributions were estimated for both the bespoke MVPA app and the native Apple Watch Activity app, and these distributions were compared to a 40% VO<sub>2</sub>R criterion value measured via ECG chest strap, as well as to each other. In the comparisons between apps and 40% HRR criterion, a Region of Practical Equivalence (ROPE) was used. This ROPE meant that, for practical purposes, values from 37% to 43% were considered equivalent to 40% HRR. Posterior distributions of the % HRR, estimated by the MVPA app, meant that the %HRR during all five minutes of the 5-minute bout had been classified as ‘exercise’ by the native Apple Watch Activity app. Together with the ECG chest strap, these data were modelled using a series of Bayesian regression models (see Models fitted subsection below for further details). In all %HRR estimates, the 40% HRR measured by the ECG chest strap was considered the response variable.

For %VO<sub>2</sub>R, posterior distributions were estimated for both the MVPA app and the native Apple Watch Activity app and these distributions were compared to a 40% VO<sub>2</sub>R criterion value, as well as to each other. In the comparisons between apps and metabolic cart, a ROPE was used. As with % HRR, the ROPE was set from 37% to 43%. The posterior distributions for the apps and metabolic cart were modelled using a series of Bayesian regression models (see Models fitted subsection below) and then compared to determine the best model in terms of out-of-sample prediction accuracy and data fit (see Comparison Methods subsection below). In all %VO<sub>2</sub>R estimates, the %VO<sub>2</sub>R measured by the metabolic cart was considered the response variable.

### Models fitted

The models estimating MVPA (40% HRR and 40% VO<sub>2</sub>R) were fitted with different response distributions (Gaussian, Skew Normal, and Student’s t-distribution) and different prior distributions: 1) general weakly informative prior, 2) a measurement constrained prior that distributes probability over possible values constrained by the measurement, and 3) strongly informative priors, informed by previous studies. Each of these models included: 1) fixed effect models, 2) random intercept models, where intercepts for each individual measure were allowed to vary, and 3) models where both intercepts and slopes for everyone were allowed to vary. This provided a pool of models for each analysis, which were compared (see Comparison methods subsection below). The basic structure of the fixed effects, random intercept and random slope models are detailed below. These were

adjusted to accommodate the different response distributions and priors outlined above.

Fixed effects model:

$$y_i \sim \text{Normal}(\mu_i, \sigma_e) \mu_i = \alpha + \beta x_i,$$

Random intercept model:

$$y_i \sim \text{Normal}(\mu_i, \sigma_e) \mu_i = \alpha_j[i] + \beta x_i \alpha_j \\ \sim \text{Normal}(\alpha, \sigma_\alpha);$$

Random intercept and slope model:

$$y_i \sim \text{Normal}(\mu_i, \sigma_e) \mu_i = \alpha_j[i] + \beta_j[i] x_i \alpha_j \\ \sim \text{Normal}(\alpha, \sigma_\alpha) \beta_j \sim \text{Normal}(\beta, \sigma_\beta).$$

In these models,  $\alpha_j[i]$  indicates that each participant is given a unique intercept, issued from a Normal/Gaussian distribution, centred on  $\alpha$ , the grand intercept, meaning that the model allowed for different mean values for each participant. The varying slopes were assigned a prior distribution centred on the grand slope  $\beta$ , and with standard deviation  $\sigma_\beta$ . X represents a categorical variable which includes the various methods of measurement to be compared. Given that the predictor categories from rating scales should not be assumed equidistant with respect to their relationship to the response variable, the distance between adjacent predictor categories is estimated from the data and may vary across categories.

### Comparison methods

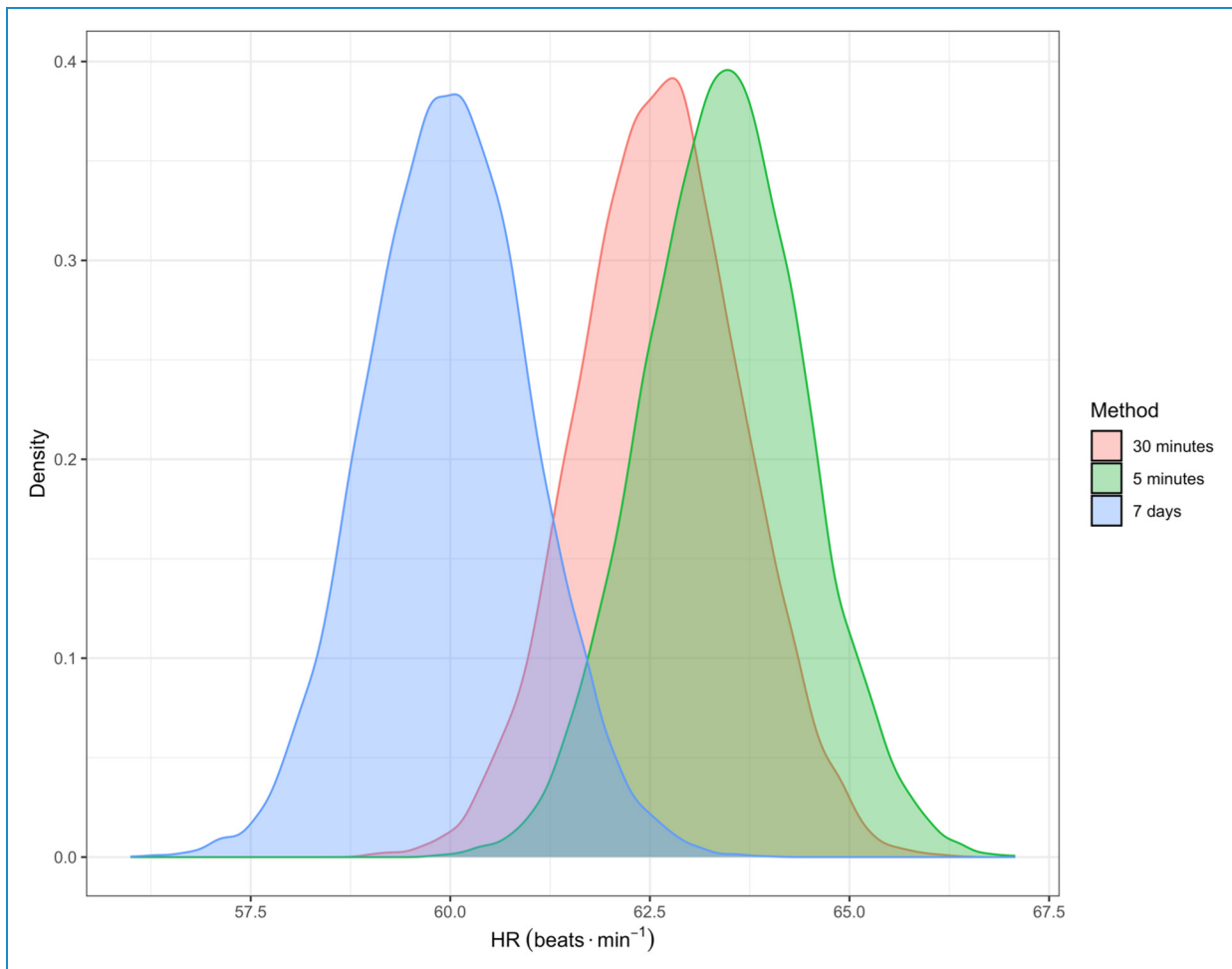
All the models produced were compared using three methods in the following order of priority: Leave-One-Out cross-validation (LOO), Bayes Factor, and Bayesian R<sup>2</sup>. Firstly, LOO information criterion (LOOIC) was used to determine the relative predictive performance of the models in terms of pointwise out-of-sample prediction accuracy using log-likelihoods from posterior simulations of parameter values.<sup>33</sup> A model was considered better if it produced a LOOIC difference greater than twice its corresponding standard error. Where a comparison of models using LOOIC does not achieve this level of difference, the second model comparison method was used where Bayes Factors (BF) quantified the support for one model over another. Where a particular model achieved a BF<sub>10</sub> greater than 10, relative to other models, it was considered the better model. To determine which model was better using LOOIC and then BF, the model with the highest Bayesian R<sup>2</sup> was selected. The Bayesian R<sup>2</sup> is seen as a data-based estimate of the proportion of variance explained for new data.<sup>34</sup>

### Results

Seventy-seven participants were initially recruited and consented to taking part in the study. Seventy-four participants (n=3 non-participation) completed all trials. Three

**Table 2.** Mean (SD) maximal oxygen consumption, resting oxygen consumption, resting heart rate (RHR) across three different methods, and maximal heart rate directly measured and estimated.

	$VO_{2max}$ ( $mL.kg^{-1}.min^{-1}$ )	$VO_{2rest}$ ( $mL.kg^{-1}.min^{-1}$ )	Chest strap RHR ( $beats.min^{-1}$ )	7 - day RHR ( $beats.min^{-1}$ )	MVPA app RHR ( $beats.min^{-1}$ )	$HR_{max}$ recorded ( $beats.min^{-1}$ )	Gellish $HR_{max}$ ( $beats.min^{-1}$ )
Mean (SD)	40 (10)	3.6 (0.6)	63 (9)	60 (7)	64 (9)	184 (11)	186 (7)
Range	20.1- 63	2-4.8	43.8-82.0	45.7-73.3	45.2-87.9	148-211	166.4-194.4



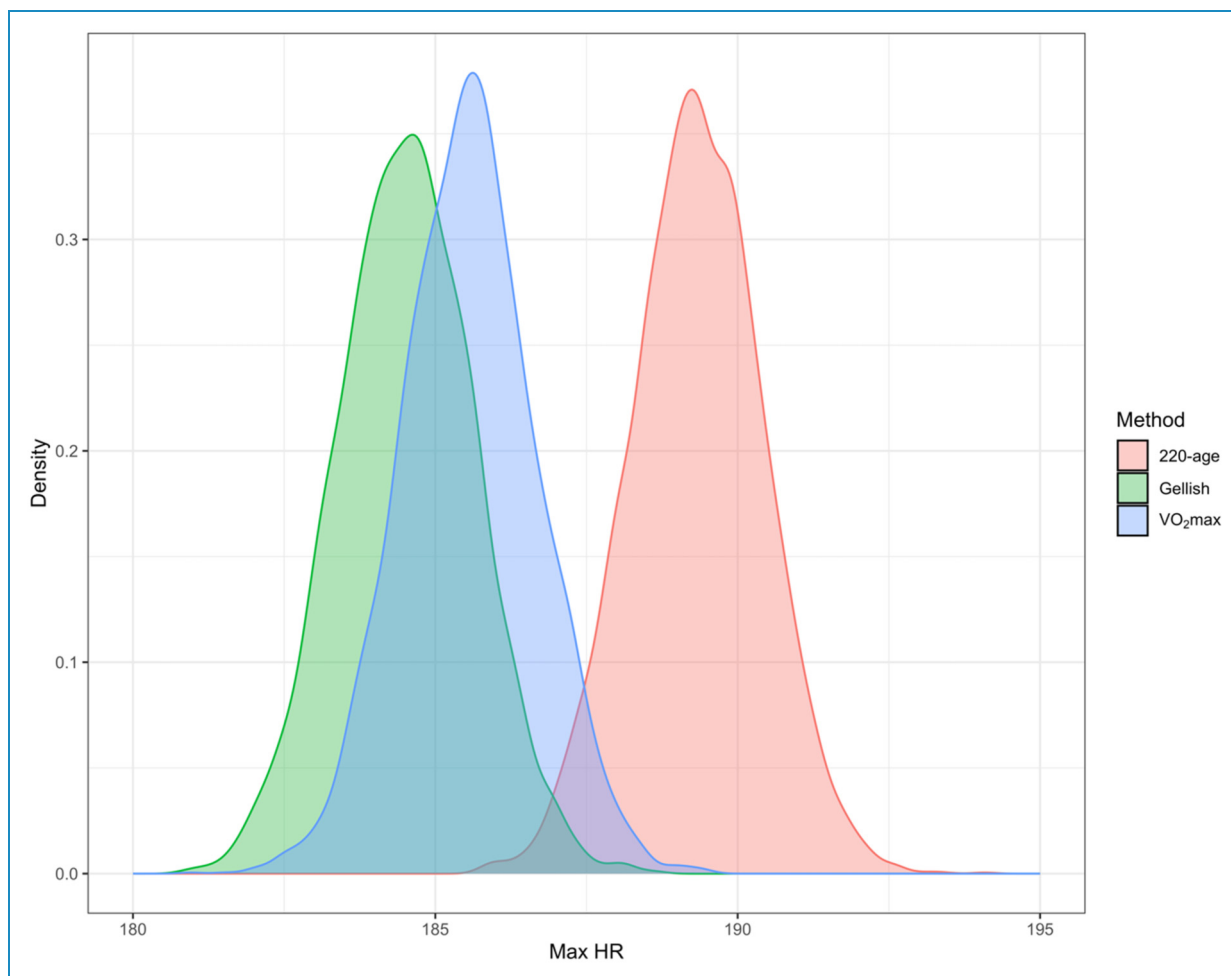
**Figure 1.** Posterior distributions for the measurement of resting heart rate using 30-minute, 5-minute supine laboratory measurements, and 7-day mean from the Apple watch.

participants did not complete all the trials because of lack of time ( $n = 2$ ) and family bereavement ( $n = 1$ ).

Table 2 shows the mean (SD) for maximal oxygen consumption and resting oxygen consumption as measured using an online gas analysis system, chest strap ECG resting heart rate, 7-day RHR recorded via the Apple Watch, 5-min MVPA app reading taken during the resting

$VO_2$  protocol, recorded  $HR_{max}$  taken from the  $VO_{2max}$  protocol, and age-predicted  $HR_{max}$  using the Gellish formula.<sup>35</sup>

Figure 1 illustrates the posterior distributions for resting heart rate from the three protocols used to record resting heart rate. Specifically, the mean effect (ME) and credible interval (CI) for each posterior distribution were: 7-day protocol ME: 60 CI: 58 to 62  $beats.min^{-1}$ , 5-min protocol



**Figure 2.** Posterior distributions for measured HR<sub>max</sub> and predicted HR<sub>max</sub> from the Gellish (206.9-[0.67\*age]) and traditional age-prediction (220-age) formulae.

ME: 63 CI: 61 to 65 beats.min<sup>-1</sup>, and 30-minute protocol ME: 63 CI: 61 to 65 beats.min<sup>-1</sup>.

Figure 2 illustrates the maximum heart rate protocols from the two predictive methods, Gelliish et al.<sup>35</sup> and 220 – age, compared to criterion HR<sub>max</sub> measured during the maximal oxygen consumption test. Specifically, the mean effect (ME) and credible interval (CI) for each posterior distribution were: Gellish ME: 185 CI: 182 to 187 beats.min<sup>-1</sup>, 220 – age ME: 189 CI: 187 to 191 beats.min<sup>-1</sup> and maximal oxygen consumption test ME: 186 CI: 183 to 188 beats.min<sup>-1</sup>.

Mean (SD) %HRR, %VO<sub>2</sub>R and treadmill speed required to advance the Apple Watch green ‘exercise’ ring by five minutes and record three minutes of MPA on the MVPA app (reported as the inflection point) are presented in Table 3.

Mean (SD) %HRR, %VO<sub>2</sub>R, treadmill speed and cadence required to meet 3 METs (absolute moderate intensity) are presented in Table 4 (METs calculated using the hybrid method, where measured oxygen consumption is divided by 3.5 mL.kg<sup>-1</sup>.min<sup>-1</sup>).

**Table 3.** Mean (SD) %VO<sub>2</sub>R, %HRR and treadmill speed at the inflection point for the native Apple watch activity app and MVPA app.

Mean (SD)	%VO <sub>2</sub> R	%HRR (ECG)	%HRR (watch)	Speed (km.h <sup>-1</sup> )
Native Activity app	33 (10)	33 (10)	34 (10)	5.36 (1)
Range	14.1-68.3	18.9-58.2	16.7-64.0	4.0-7.0
MVPA app	43 (9)	43 (6)	44 (3)	6.00 (1)
Range	14.2-59.9	27.8-57.7	37.8-51.6	3.5-8.0

Table 5 displays the posterior distribution mean effect and credible intervals for the 3 MET inflection point for %HRR, %VO<sub>2</sub>R, treadmill speed and cadence.

Figure 3 illustrates the oxygen consumption reserve (%VO<sub>2</sub>R) at the inflection point for the native Apple Watch

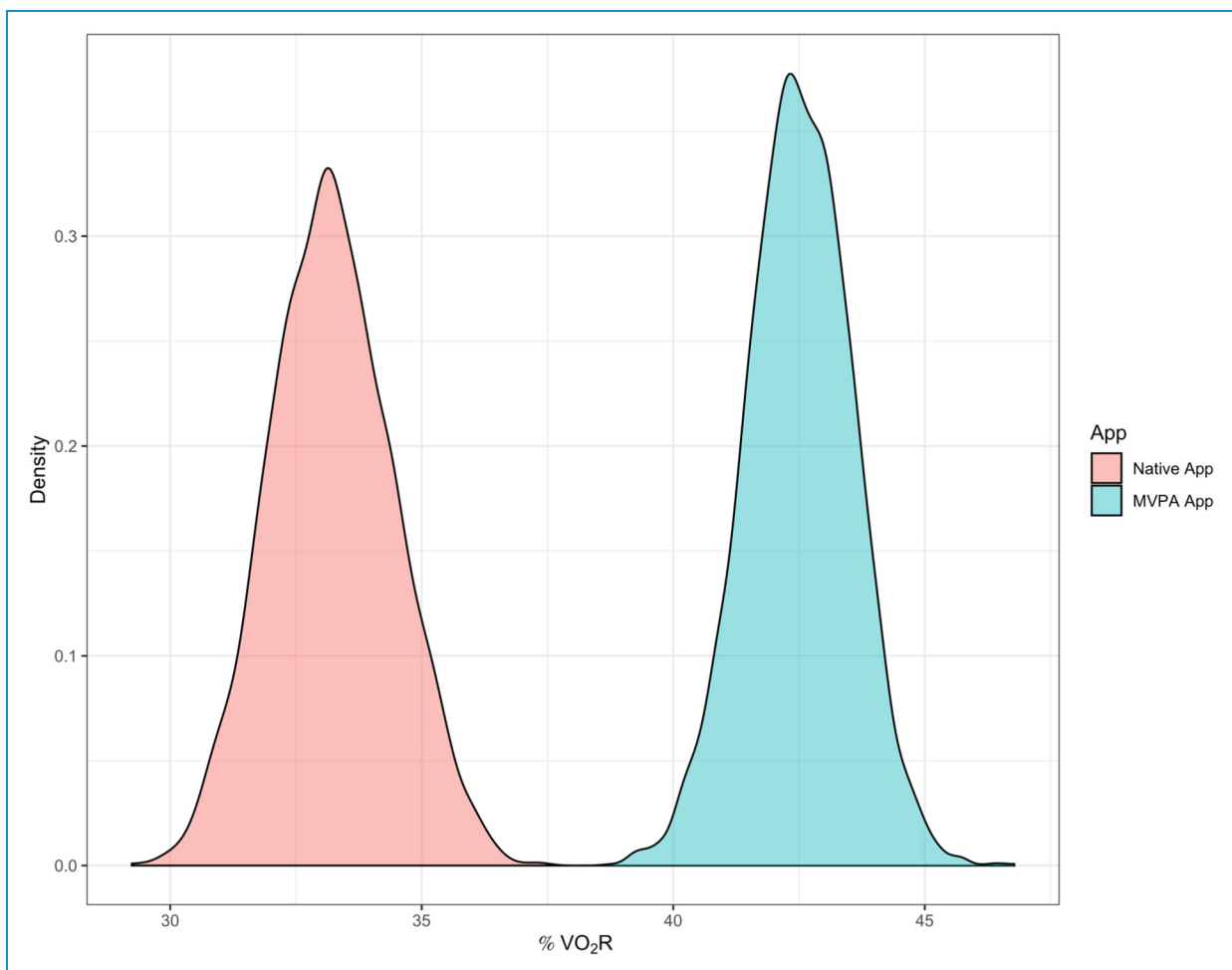
**Table 4.** Mean (SD) %VO<sub>2</sub>R, %HRR, treadmill speed and cadence at 3 METs.

Mean (SD)	%VO <sub>2</sub> R	%HRR (ECG)	Cadence (steps.min <sup>-1</sup> )	Speed (km.h <sup>-1</sup> )
3 METs	23 (8)	25 (8)	104 (9)	3.8 (0.4)
Range	11.1-49.0	11-53.2	90-136	3.5-5.0

**Table 5.** Mean effect (95% CI) for the 3 MET inflection point for %HRR, %VO<sub>2</sub>R, treadmill speed and cadence.

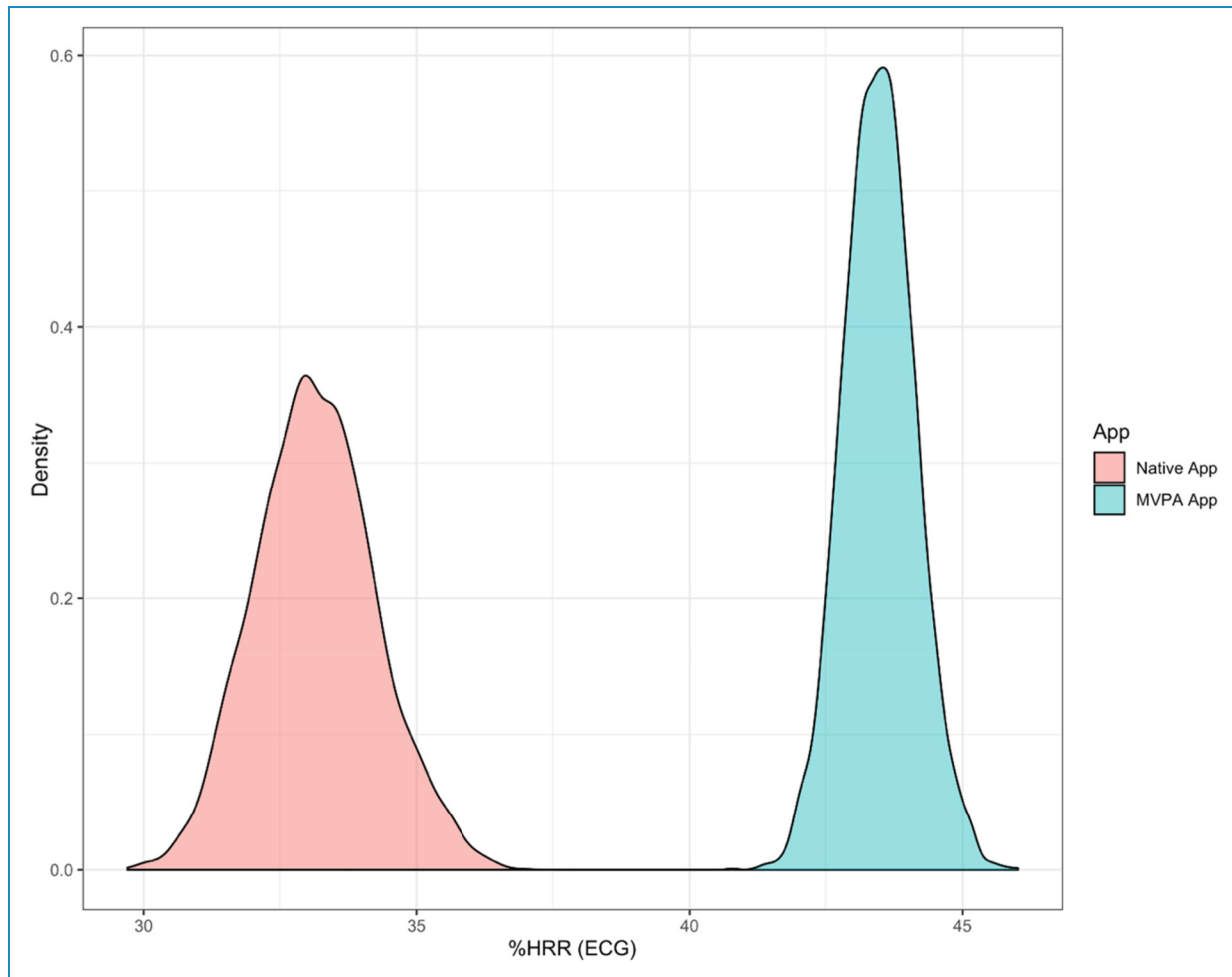
Mean effect (95% CI)	%VO <sub>2</sub> R	%HRR (ECG)	Cadence (steps.min <sup>-1</sup> )
3 METs	22 (20.7-23.14)	23.7 (22.6 - 25)	102 (100-103)

Activity app whereby all five minutes had been recorded as ‘exercise’ and the bespoke MVPA app had recorded a minimum of three minutes of the 5-minute bout as criterion moderate intensity. Specifically, the mean effect (ME) and credible interval (CI) for the native Apple Watch activity app were: ME: 33% CI: 31% to 36% VO<sub>2</sub>R, and for the bespoke MVPA app: ME: 43% CI: 40% to 44% VO<sub>2</sub>R. The Bayesian R<sup>2</sup> value for the bespoke MVPA app as a data-based estimate of the proportion of variance at 40% VO<sub>2</sub>R was 0.75. For the native Apple Watch Activity app Bayesian R<sup>2</sup> was 0.066. The native Apple Watch Activity app recorded 0% of participants above 40% VO<sub>2</sub>R when the native Apple Watch activity app’s green exercise ring recorded minutes of ‘exercise’, compared to 95% of participants being above 40% VO<sub>2</sub>R when using our ‘MVPA’ app Figure 4 illustrates the %HRR at the inflection point for the native Apple Watch Activity app whereby all five minutes had been recorded



**Figure 3.** Posterior distributions for oxygen consumption reserve (%VO<sub>2</sub>R) at the inflection point for the native Apple watch activity app and bespoke MVPA app.





**Figure 4.** Posterior distributions for %HRR at the inflection point for the native Apple watch activity app and bespoke MVPA app. Heart rate is from the ECG chest strap.

as ‘exercise’ and the bespoke MVPA app had recorded a minimum of three minutes of the 5-minute bout as criterion moderate intensity. Specifically, the mean effect (ME) and credible interval (CI) for the native Apple Watch activity app were: ME: 33% CI: 31% to 35% HRR, and the bespoke MVPA app: ME: 44% CI: 43% to 45% HRR. The Bayesian  $R^2$  value for the bespoke MVPA app as a data-based estimate of the proportion of variance at 40% HRR was 0.64. For the native Apple Watch activity app Bayesian  $R^2$  was 0.015

## Discussion

The aim of this study was to examine whether %HRR more accurately measures criterion ‘moderate’ exercise intensity (40-59% HRR) compared to the native Apple Watch ‘Activity’ app when compared to a criterion measure, % oxygen consumption reserve (%VO<sub>2</sub>R). The main finding of this study is that using %HRR more accurately measures moderate intensity than the Apple Watch Activity app

(green ‘exercise’ ring) when compared to the %VO<sub>2</sub>R criterion. The mean (95%CI) %HRR for the bespoke MVPA app and native Apple Watch Activity app were 43% (40% to 44%) and 33% (31% to 36%), respectively. These data show that at the point both apps (bespoke MVPA and native Activity) indicated that participants were exercising at moderate intensity, only the bespoke MVPA app achieved this criterion, with the native Apple Watch Activity app intensity being well below the 40% HRR threshold. Moreover, the entire posterior distribution for the native Activity app was outside the pre-registered ROPE, and the 95% credible interval for the bespoke MVPA app of 40% to 44% suggests that our estimations of these population effects are very precise. We can also describe the size of effect by the fact that when using the native ‘Activity’ app, 0% of participants were above 40% VO<sub>2</sub>R when the Activity app’s green ‘exercise’ ring advanced, compared to 95% of participants being above 40% VO<sub>2</sub>R when using the bespoke MVPA app.

Given these results, %HRR may be the most accurate and accessible individualised measure of exercise intensity

that can be incorporated into a wrist-worn wearable device. The popularity of wearable devices,<sup>21</sup> and the ability to scale individualised measurement of moderate (and vigorous) exercise intensity at a population level, should not be underestimated. The posterior distributions comparing the bespoke MVPA app with the Apple Watch native Activity app highlight that using %HRR is likely to measure physical activity intensity of moderate or above (40%  $\text{VO}_2\text{R}$ ) more accurately compared to the Apple Watch native Activity app, which was unlikely to measure intensity accurately compared to criterion intensity % $\text{VO}_2\text{R}$ . These findings are in line with previous investigations,<sup>15,17,26</sup> highlighting the need for individualisation of exercise monitoring using relative intensity, as physical activity intensity is influenced by cardiorespiratory fitness status. For example, in the current study the minimum treadmill speed required to elicit moderate physical activity (MPA) for an individual (40% HRR) by the bespoke MVPA app was  $3.5 \text{ km}\cdot\text{h}^{-1}$  while the fastest treadmill speed required to reach MPA for an individual was  $8.0 \text{ km}\cdot\text{h}^{-1}$ . This highlights the individualised nature of the intensity spectrum and how participants' cardiorespiratory fitness plays an integral role in the speed required to elicit MPA. Wearable technology manufacturers may wish to consider incorporating relative measures of physical activity intensity into their exercise monitoring.

Accurate measurement of RHR is important if relative measures of intensity are to be used, and trusted, in wearable devices. Although a 12-lead ECG measurement of RHR is seen as the gold standard,<sup>36</sup> there are obvious practical implications for using this method outside of laboratory conditions. The resting heart rate data measured in the current study demonstrate that, in comparison to the criterion 30-minute supine recorded RHR, the 5-minute protocol embedded in the bespoke MVPA app overestimates mean RHR by  $\sim 1 \text{ beats}\cdot\text{min}^{-1}$ , while the mean 7-day Apple Watch estimate seems to underestimate RHR by  $\sim 3 \text{ beats}\cdot\text{min}^{-1}$ . These results have implications for real world app useability and subsequent physical activity monitoring via wearable technology. First, it is unlikely that wearable device users would be willing to maintain a supine position for a period (either 30 minutes or 5 minutes) while heart rate is recorded. Second, although the underestimation of RHR using the 7-day Apple Watch data may result in slightly different %HRR measurements, this effect might be trivial in practice. For example, the data displayed in Figure 4 showing the posterior distribution of %HRR for the bespoke MVPA app are based on the use of the 7-day RHR data from the Apple Watch, and yet this distribution matches quite closely to that displayed for % $\text{VO}_2\text{R}$  (Figure 3). Moreover, the difference in HRR at 40% between a RHR of 60 and 63 for a person with a HRmax of 200 is  $2 \text{ beats}\cdot\text{min}^{-1}$ . So, the use of background RHR as recorded automatically by the Apple Watch might represent a 'sweet spot' balance between convenience and accuracy.

While to the best of the authors' knowledge there are no current studies investigating the validity of recording RHR via the Apple Watch compared to ECG at rest, a recent investigation<sup>37</sup> recorded resting heart rate via an Apple Watch and ECG during 3 minutes of seated rest prior to a maximal oxygen consumption protocol. The mean absolute percentage error between Apple Watch and ECG was 1.7%. The Apple Watch's useability in measuring health and physical activity may have a wider impact than the prescription of exercise intensity alone. For example, Greiwe and Nyenhuis<sup>38</sup> suggest that telemonitoring of health measures (such as RHR) may transform the way health is measured and diagnosed. Never-the-less, additional research into the validity and accuracy of resting heart rate monitoring with wearable devices might be required, especially if wearable device manufacturers start to incorporate relative measures of exercise intensity that are dependent on these values.

The posterior distributions of  $\text{HR}_{\text{max}}$  (Figure 2) indicate that the Gellish formula<sup>35</sup> was the most accurate at predicting  $\text{HR}_{\text{max}}$  ( $\sim 1 \text{ beats}\cdot\text{min}^{-1}$  difference) while the 220 - age calculation overestimated  $\text{HR}_{\text{max}}$  ( $\sim 4 \text{ beats}\cdot\text{min}^{-1}$  difference) compared to the criterion measured during a maximal oxygen consumption test. These findings are in line with previous studies<sup>39-41</sup> reporting that 220 - age did not accurately predict  $\text{HR}_{\text{max}}$ . In order to prescribe and monitor exercise intensity more accurately, an upper ( $\text{HR}_{\text{max}}$ ) and lower (RHR) bound of heart rate is required to quantify intensity thresholds using %HRR.<sup>42,43</sup> These findings have important implications in a real-world setting. The likelihood of individuals performing maximal oxygen consumption testing is very low, and therefore predictive measures are often used to calculate maximal values.<sup>44</sup> Based on the findings of this study, those predictions should use the Gellish et al.<sup>35</sup> formula as it aligns more closely with the  $\text{HR}_{\text{max}}$  recorded during the maximal oxygen consumption test performed in this study. Using other predictions may lead to inaccurate exercise intensity prescription.

Cardiorespiratory fitness plays an integral role in the walking cadence required to meet a relative measure of moderate intensity.<sup>15</sup> When physical activity is monitored relatively via a physiological measure, such as %HRR, and thus individualised via intensity thresholds, the need for a walking cadence prescription is only useful when walking is performed in short bursts. In such a circumstance, the delay in heart rate response does not allow for physiological mechanisms to be measured quickly enough. Although previous studies and public health guidelines have monitored/promoted physical activity using absolute measures such as step count (typically 10,000 steps per day, or 3,000 steps in 30 minutes),<sup>12,45,46</sup> using a relative measure of physical activity intensity as shown in the current study is a more accurate measure of physical activity intensity. The growing popularity of wearable devices now makes this individualised approach a reality, and at scale. Ultimately,

the use of individualised relative intensity measures for physical activity monitoring and prescription has potential implications for mortality and morbidity risks, improving people's quality of life and subsequently reducing pressures on health services. As such, a move toward greater emphasis on individualised physical activity intensity monitoring and prescription should be incorporated into public health guidelines.

Measuring physical activity accurately on a population scale is complex. However, based on the findings from the current study, the Apple Watch would overestimate MVPA when compared to measuring MVPA using the criterion  $\%VO_2R$  (or  $\%HRR$ ). Mean treadmill speed at which the Apple Watch registered physical activity on the green 'exercise' ring was  $0.8 \text{ km}\cdot\text{h}^{-1}$  ( $5.4 \text{ km}\cdot\text{h}^{-1}$  compared to  $6.2 \text{ km}\cdot\text{h}^{-1}$ ) slower than the bespoke MVPA app. Currently, the Apple Watch records physical activity via an accumulation of accelerometry data and intermittent heart rate data.<sup>47</sup> This is likely configured to prolong battery life of the wearable device, as using the PPG sensor continually requires additional battery power. Therefore, as a relative measure of intensity is not used regularly, the Apple Watch is likely overestimating physical activity, because it measures physical activity mostly via an absolute accelerometry measurement. Additionally, the point at which 3 METs was reached (in this case METs were calculated via oxygen consumption divided by a resting oxygen consumption value of  $3.5 \text{ mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ ), also indicates a substantial overestimation of physical activity intensity when compared to criterion  $\%VO_2R$  (or  $\%HRR$ ). At 3 METs,  $\%HRR$  mean (SD) was 25 (8)% and  $\%VO_2R$  was 23 (8)% substantially under the 40% moderate intensity boundary. Additionally, cadence was  $104$  (9)  $\text{steps}\cdot\text{min}^{-1}$  which may indicate why  $100 \text{ steps}\cdot\text{min}^{-1}$  has been advocated as the walking cadence required to elicit moderate intensity.<sup>11,48</sup> However, the findings of the current study support that using  $\%HRR$ , which takes into consideration two measures of physiological capacity (resting HR and maximal HR), is likely to measure physical activity intensity more accurately on an individual basis. Given the potentially large overestimation of MVPA throughout the day when using the Apple Watch Activity app (green 'exercise' ring), a greater degree of personalisation is required. The need for personalisation of exercise intensity thresholds requires either a more continual use of background heart rate measurements by the Apple Watch during everyday ambulation, or a greater personalisation of data (feedback) given to users to ensure walking cadence meets relative moderate intensity thresholds. Currently the Apple Watch has heart rate zones for physical activity displayed in the Workout app<sup>49</sup> that are based on  $\%HRR$ . The default zone 1 begins at 60% HRR, or vigorous intensity as defined by the ACSM.<sup>7</sup> These zones, and the monitoring of exercise via  $\%HRR$ , are not used to quantify exercise on the native Apple Watch Activity app during everyday ambulation, or when an

exercise mode is selected in the Workout app. Although heart rate zones can be modified manually, it is unlikely that Apple Watch users will change the zones to replicate ACSM guidelines.<sup>7</sup> Starting the first heart rate zone at vigorous intensity is unlikely to increase adherence to exercise.<sup>50</sup> The default Apple Watch heart rate zones would benefit from being aligned to the moderate and vigorous exercise intensity guidelines recommended by the ACSM,<sup>7</sup> by setting zone 1 (moderate intensity) as 40-59% HRR and zone 2 (vigorous intensity) as 60-90% HRR.

## Study limitations

This study has focussed on using  $\%HRR$  that can be measured by wearable devices and is therefore accessible to the general population. Nevertheless, there are limitations in using this method that are worth noting. While  $\%HRR$  is individualised to an extent, the method uses fixed percentage boundaries to define intensity categories (e.g., moderate intensity is defined at 40–59% HRR). Physiological markers such as lactate or ventilatory thresholds may offer a more individualised and subsequent better method of monitoring physical activity intensity as they reflect a real physiological threshold for each individual.<sup>51</sup> Additionally, ventilatory thresholds have been reported to better anchor physical activity intensity as the metabolic stimulus is better normalised across people with varying fitness levels.<sup>52</sup> However, at the present time wearable devices do not have the ability to measure such physiological thresholds, and future research in this field may be dictated by wearable device manufacturers making these measurements available. Therefore,  $\%HRR$  currently offers the most accurate relative measure available for use at a population level. Never-the-less, future research may wish to identify how accurate the prescription of training zones offered on wearable devices such as the Apple Watch correspond to physiological phenomena like ventilatory thresholds.

A potential limitation of the current study is that the Apple Watches used were Series 5, with the latest Watches now available being Series 10. Although there is no public information available on whether, or how, the Apple Watch has changed how it measures 'exercise' via the green ring, it is possible that some changes have been made to improve the accuracy of the Watch for measuring MVPA. One of the problems that all research studies examining consumer technology suffer from is the rapid advancement of technology, and the inevitable difference between the technology used in the study for data collection and that available to the consumer when the study is published.

## Conclusion

The bespoke MVPA app measured relative moderate intensity more accurately compared to the native Apple Watch

Activity app (green ‘exercise’ ring) when assessed against the criterion %VO<sub>2</sub>R measure. Exercise guidelines and wearable devices like the Apple Watch should incorporate relative measures of physical activity to individualise physical activity monitoring and prescription more accurately, and with a concomitant move away from arbitrary absolute values that apply a single exercise goal to the mass population, such as 100 steps·min<sup>-1</sup> walking cadence. A shift towards individualised physical activity intensity prescription is available via wearable technology, and incorporating individualised prescription has the potential to improve population health.

**Author contributions:** Conception (GA, AW), acquisition of data (AW), data extraction, data analyses (TM, GA, AW), interpretation of data (all authors), writing the manuscript (all authors), final approval of the submitted version (all authors). The corresponding author attests that all listed authors meet the criteria for authorship and that no others meeting the criteria have been omitted.

**Consent to participate:** Written consent was obtained from all participants prior to participation.

**Consent for publication:** Not applicable

**Data availability:** Anonymised data at <https://osf.io/3uejs/>


**Declaration of conflicting interest:** Apple Inc. provided the Apple Watch devices for the research. Apple was not involved in the design of the research, nor was it involved in the collection, analysis, or interpretation of the research data, or the content of this or any related publication.


**Ethical considerations:** Institutional ethics approval was obtained from the Faculty of Health Sciences Research Ethics Committee at the University of Hull (REF – FHS230).


**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Apple, (grant number Investigator Support Programme).

**ORCID iDs:** Ashley Warner  <https://orcid.org/0000-0002-5173-9716>

Natalie Vanicek  <https://orcid.org/0000-0002-9602-3172>

Amanda Benson  <https://orcid.org/0000-0003-2959-8969>

Tony Myers  <https://orcid.org/0000-0003-4516-4829>

Grant Abt  <https://orcid.org/0000-0002-4079-9270>

## References

- Barengo NC, Hu G, Lakka TA, et al. Low physical activity as a predictor for total and cardiovascular disease mortality in middle-aged men and women in Finland. *Eur Heart J* 2004; 25: 2204–2211.
- Barengo NC, Antikainen R, Borodulin K, et al. Leisure-Time physical activity reduces total and cardiovascular mortality and cardiovascular disease incidence in older adults. *J Am Geriatr Soc* 2017; 65: 504–510.
- Nocon M, Hiemann T, Müller-Riemenschneider F, et al. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. *Eur J Cardiovasc Prev Rehabil* 2008; 15: 239–246.
- Freitas PD, Silva AG, Ferreira PG, et al. Exercise improves physical activity and comorbidities in obese adults with asthma. *Med Sci Sports Exerc* 2018; 50: 1367–1376.
- Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. *Lancet Public Health* 2022; 7: e219–e228.
- nhs.uk. Physical activity guidelines for adults aged 19 to 64. *nhs.uk*, <https://www.nhs.uk/live-well/exercise/exercise-guidelines/physical-activity-guidelines-for-adults-aged-19-to-64/> (2022, accessed 9 May 2022).
- American College of Sports Medicine, Liguori G, Feito Y, et al. *ACSM's guidelines for exercise testing and prescription*. 11th ed. Philadelphia, PA: Wolters Kluwer Health, 2022.
- Siddique J, Aaby D, Montag SE, et al. Individualized relative-intensity physical activity accelerometer cut points. *Med Sci Sports Exercise* 2020; 52: 398–407.
- Gil-Rey E, Maldonado-Martin S, Palacios-Samper N, et al. Objectively measured absolute and relative physical activity intensity levels in postmenopausal women. *Eur J Sport Sci* 2019; 19: 539–548.
- Warner A, Vanicek N, Benson A, et al. Agreement and relationship between measures of absolute and relative intensity during walking: a systematic review with meta-regression. *PLoS One* 2022; 17: e0277031.
- Tudor-Locke C, Han H, Aguiar EJ, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *Br J Sports Med* 2018; 52: 776–788.
- Marshall SJ, Levy SS, Tudor-Locke CE, et al. Translating physical activity recommendations into a pedometer-based step goal: 3000 steps in 30 minutes. *Am J Prev Med* 2009; 36: 410–415.
- Nielson R, Vehrs PR, Fellingham GW, et al. Step counts and energy expenditure as estimated by pedometry during treadmill walking at different stride frequencies. *J Phys Act Health* 2011; 8: 1004–1013.
- Abel M, Hannon J, Mullineaux D, et al. Determination of step rate thresholds corresponding to physical activity intensity classifications in adults. *J Phys Act Health* 2011; 8: 45–51.
- Abt G, Bray J, Myers T, et al. Walking cadence required to elicit criterion moderate-intensity physical activity is moderated by fitness status. *J Sports Sci* 2019; 17: 1989–1995.
- O'Brien MW, Kivell MJ, Wojcik WR, et al. Step rate thresholds associated with moderate and vigorous physical activity in adults. *Int J Environ Res Public Health* 2018; 15: 2545.
- Serrano F, Slaght J, Sénéchal M, et al. Identification and prediction of the walking cadence required to reach moderate intensity using individually-determined relative moderate intensity in older adults. *J Aging Phys Act* 2017; 25: 205–211.
- Lehtonen E, Gagnon D, Eklund D, et al. Hierarchical framework to improve individualised exercise prescription in adults: a critical review. *BMJ Open Sport & Exercise Medicine* 2022; 8: e001339.

19. Swain DP and Leutholtz BC. Heart rate reserve is equivalent to %VO<sub>2</sub> reserve, not to %VO<sub>2</sub>max. *Med Sci Sports Exercise* 1997; 29: 410–414.
20. Kite-Powell J. *Polar: the original fitness tracker and heart rate monitor*. Forbes, 2016; <https://www.forbes.com/sites/jenniferhicks/2016/02/28/polar-the-original-fitness-tracker-and-heart-rate-monitor/> (accessed 18 November 2022).
21. Statista. Global connected wearable devices 2016–2022. *Statista* 2023. <https://www.statista.com/statistics/487291/global-connected-wearable-devices> (accessed 13 February 2023).
22. Khushhal A, Nichols S, Evans W, et al. Validity and reliability of the apple watch for measuring heart rate during exercise. *Sports Medicine International Open* 2017; 1: E206–E211.
23. Zhang Y, Weaver RG, Armstrong B, et al. Validity of wrist-worn photoplethysmography devices to measure heart rate: a systematic review and meta-analysis. *J Sports Sci* 2020; 38: 2021–2034.
24. Thompson D, Batterham AM, Peacock OJ, et al. Feedback from physical activity monitors is not compatible with current recommendations: a recalibration study. *Prev Med* 2016; 91: 389–394.
25. Apple UK. Apple Watch - Close Your Rings. *Apple (United Kingdom)*, <https://www.apple.com/uk/watch/close-your-rings/> (2023, accessed 5 June 2023).
26. Abt G, Bray J and Benson AC. Measuring moderate-intensity exercise with the Apple watch: validation study. *JMIR Cardio* 2018; 2: e6.
27. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.
28. Riebe D, Ehrman JK and Liguori G, et al. *ACSM guidelines for exercise testing and prescription*. 10th ed. Philadelphia, PA: Wolters Kluwer, 2018.
29. Marsman M and Wagenmakers E-J. Bayesian Benefits with JASP. *European Journal of Developmental Psychology* 2017; 14: 545–555.
30. Norton K, Norton L and Sadgrove D. Position statement on physical activity and exercise intensity terminology. *J Sci Med Sport* 2010; 13: 496–502.
31. Terry MA, Winell JJ, Green DW, et al. Measurement variance in limb length discrepancy: clinical and radiographic assessment of interobserver and intraobserver variability. *J Pediatr Orthop* 2005; 25: 197–201.
32. Edvardsen E, Hem E and Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PLoS One* 2014; 9: e85276.
33. Vehtari A, Gelman A and Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 2017; 27: 1413–1432.
34. Gelman A, Goodrich B, Gabry J, et al. R-squared for Bayesian regression models. *Am Stat* 2019; 73: 307–309.
35. Gellish RL, Goslin BR, Olson RE, et al. Longitudinal modeling of the relationship between age and maximal heart rate. *Med Sci Sports Exercise* 2007; 39: 822–829.
36. Bayés de Luna A, Bayes-Genis A, Brugada R, et al. *Clinical electrocardiography: a textbook*. 4th ed. Wiley, 2012, pp.23–33. DOI: 10.1002/9781118392041
37. Ho W-T, Yang Y-J and Li T-C. Accuracy of wrist-worn wearable devices for determining exercise intensity. *DIGITAL HEALTH* 2022; 8: 20552076221124393.
38. Greiwe J and Nyenhuis SM. Wearable technology and how this can be implemented into clinical practice. *Curr Allergy Asthma Rep* 2020; 20: 36.
39. Machado FA and Denadai BS. Validity of maximum heart rate prediction equations for children and adolescents. *Arq Bras Cardiol* 2011; 97: 136–140.
40. Nes BM, Janszky I, Wisløff U, et al. Age-predicted maximal heart rate in healthy subjects: the HUNT fitness study. *Scand J Med Sci Sports* 2013; 23: 697–704.
41. Sarzynski MA, Rankinen T, Earnest CP, et al. Measured maximal heart rates compared to commonly used age-based prediction equations in the heritage family study. *Am J Hum Biol* 2013; 25: 695–701.
42. Dalleck LC and Kravitz L. Relationship between %heart rate reserve and %VO<sub>2</sub> reserve during elliptical crosstrainer exercise. *J Sports Sci Med* 2006; 5: 662–671.
43. Lounana J, Champion F, Noakes TD, et al. Relationship between %HRmax, %HR reserve, %VO<sub>2</sub>max, and %VO<sub>2</sub> reserve in elite cyclists. *Med Sci Sports Exercise* 2007; 39: 350–357.
44. Sartor F, Vernillo G, de Morree HM, et al. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. *Sports Med* 2013; 43: 865–873.
45. Le-Masurier GC, Sidman CL and Corbin CB. Accumulating 10,000 steps: does this meet current physical activity guidelines? *Res Q Exerc Sport* 2003; 74: 389–394.
46. Tudor-Locke C, Sisson SB, Collova T, et al. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Can J Appl Physiol* 2005; 30: 666–676.
47. Apple. Monitor your heart rate with apple watch. *Apple Support* 2022. <https://support.apple.com/en-us/HT204666> (accessed 7 June 2023).
48. Aguiar EJ, Gould ZR, Ducharme SW, et al. Cadence-based classification of minimally moderate intensity during over-ground walking in 21- to 40-year-old adults. *J Phys Act Health* 2019; 16: 1092–1097.
49. Apple. Monitor your heart rate with Apple Watch - Apple Support, <https://support.apple.com/en-us/HT204666>.
50. Ekkekakis P and Biddle SJH. Extraordinary claims in the literature on high-intensity interval training (HIIT): iV. Is HIIT associated with higher long-term exercise adherence?. *Psychol Sport Exerc* 2023; 64: 102295.
51. Iannetta D, Inglis EC, Mattu AT, et al. A critical evaluation of current methods for exercise prescription in women and men. *Med Sci Sports Exercise* 2020; 52: 466–473.
52. Weatherwax R, Harris N, Kilding AE, et al. Incidence of  $\dot{V}O_{2\max}$  responders to personalized versus standardized exercise prescription. *Med Sci Sports Exercise* 2019; 51: 681–691.