

The Accuracy of Wrist-worn Heart Rate Monitors across a Range of Exercise Intensities

Paul Hough*, Mark Glaister, Adam Pledger

School of Sport, Health & Applied Science, St Mary's University, Twickenham, London, TW1 4SX, United Kingdom

*Corresponding author: Paul.hough@stmarys.ac.uk

Abstract Commercially available wrist-worn heart rate (HR) monitors have become increasingly popular. However, the accuracy of the devices across a range of exercise intensities is largely unknown. This study investigated the accuracy of four wrist-worn HR monitors (Apple Watch Series 1, Fitbit Charge, TomTom Touch, and Mio Fuse). Eighteen adults completed three trials on a cycle ergometer wearing a chest-worn HR monitor (Polar). Trial 1 established the HR-power output relationship, and resting and maximum HR. In trials 2 and 3, participants were fitted to an electrocardiogram (ECG) and completed a step test consisting of 5 x 3 minute stages at 40 - 80% of HR reserve (determined in trial 1) whilst wearing two wrist-worn HR monitors. Relative to ECG, there were no differences in HR between the devices during exercise ($p = 0.239$), and no device \times exercise intensity interaction ($p = 0.370$). There were no instances where ECG and Polar data differed by $\geq 5 \text{ b}\cdot\text{min}^{-1}$. Conversely, there were two instances (2.2%) with the Apple, four (4.4%) with the Mio, 10 (11.1%) with the TomTom, and 19 (21.1%) with the Fitbit. A chest-worn HR monitor offers greater accuracy compared to wrist-worn devices.

Keywords: Heart-rate, accuracy, fitness, watch, technology

Cite This Article: Paul Hough, Mark Glaister, and Adam Pledger, "The Accuracy of Wrist-worn Heart Rate Monitors across a Range of Exercise Intensities." *Journal of Physical Activity Research*, vol. 2, no. 2 (2017): 112-116. doi: 10.12691/jpar-2-2-8.

1. Introduction

The measurement of heart rate (HR) is widely used to monitor the intensity of physical activity (PA), devise exercise programmes based on HR training zones [1,2] and estimate cardiorespiratory fitness (CRF). [3] Technological advances since the early 1980s have led to the development of commercially available HR monitors, which consist of a chest strap (transmitter) that communicates with a watch (receiver) via telemetry, providing a continuous reading of HR. [4] These devices are reported to produce valid and reliable HR data when compared against an electrocardiogram (ECG) at rest and during various modes and intensities of exercise. [4,5,6]

Although HR chest belt systems are highly accurate, wearing a belt across the chest can be inconvenient for some users and may cause minor discomfort if worn for extended periods. To address these issues, wrist-worn devices have recently been developed for commercial use. These devices have implemented photoplethysmography (PPG) techniques to measure HR without the need for a chest belt. PPG is a non-invasive method for the detection of HR using a probe, typically light emitting diodes, to shine directly onto the skin and detect changes in the blood volume to determine HR. The PPG technique is based on the principle that blood flow through the artery is inversely related to the amount of light refracted. [7] Recent evidence suggests that the PPG method has acceptable validity in the measurement of HR during

walking and running activities. [8] However, some studies [9,10] have reported that wrist-worn devices become less accurate in measuring HR as exercise intensity increases, possibly due to increases in upper body movement during faster running speeds, rather than an error in the device per se. In contrast, Stahl and colleagues [8] reported an increase in HR precision at raster running speeds, possibly due to an enhancement in blood perfusion.

Studies that have indicated a reduction in the accuracy of wrist-worn HR devices as exercise intensity increases have tested the devices using fixed, rather than relative, intensities of exercise. [8,9,10] As exercise intensities for each participant were not prescribed relative to the level of CRF, this could influence the precision of the HR response due to participants performing exercise within different intensity domains. For example, running at a particular speed, could be within the moderate intensity domain for one individual; whereas the same speed could be vigorous, potentially non steady-state, exercise for an individual with a lower level of CRF.

Current evidence suggests that the accuracy of wrist-worn HR devices is dependent on the device used and the intensity of exercise. However, no study to date has tested the accuracy of devices using relative exercise intensities. Additionally, it is unclear if the accuracy of HR readings is decreased when exercise intensity increases. Therefore, the aim of this study was to determine the accuracy of four popular wrist-worn devices (Apple Watch Series 1, Fitbit Charge 2, Mio Fuse and TomTom Touch) in measuring HR during incremental cycling at fixed, relative exercise intensities.

2. Methods

2.1. Participants

Eighteen healthy adults (16 males, 2 females) volunteered to participate in the study which was approved by St Mary's University's Ethics Committee. Mean \pm s for age, height, body mass, and body mass index were: 30 ± 8 yrs, 1.76 ± 0.08 m, 75.3 ± 12.8 kg, and 24.1 ± 2.8 kg·m⁻², respectively. Exclusion criteria included: (1) current use of any medications that influence the cardiovascular system; (2) presence of any known cardiorespiratory, neurological, musculoskeletal, or metabolic diseases or abnormalities; (3) being unaccustomed to regular (at least 75 minutes per week) vigorous exercise. Participants completed a medical history form and provided written informed consent before commencing the experimental procedures.

2.2. Experimental Design

All participants made three visits to the laboratory for data collection. Trial 1 consisted of a baseline assessment to establish resting HR, the relationship between HR and power output, and maximum HR (HR_{max}) for each participant. During trials 2 and 3, wrist-worn HR monitors were used alongside an electrocardiograph (ECG) system (Vyntus ECG, JAEGER, Würzburg, Germany) and chest-worn HR monitor (Polar S610i; Polar Electro Oy, Kempele, Finland). Participants were instructed not to eat or drink anything other than water in the three hours preceding each visit and to refrain from any strenuous PA and caffeine consumption for 24 hours before each visit.

2.3. Baseline Assessment

Upon arriving at the laboratory, each participant completed a medical history questionnaire and consent form. Height and body mass were measured, and a coded strap (Polar T31; Polar Electro Oy, Kempele, Finland) was fitted around the chest of each participant at the level of the xiphoid process. Following five minutes of seated, undisturbed rest, resting HR was recorded using the HR receiver. A validated prediction equation [11] was used to calculate predicted HR_{max} and 85% HRR for use in the comparative trials.

The cycle ergometer (SRM, SRM International, Jülich, Germany) was adjusted for each participant at the start of the baseline assessment, with measurements noted for replication in subsequent trials. Participants then completed a submaximal step protocol consisting of six to nine 3-minute stages, with power output increasing by 20 W at the start of each stage. The starting power output was selected by the investigator based on the reported training status of each participant, to enable the completion of a minimum of five and a maximum of nine stages before 85% HR_{max} was achieved. All participants completed a minimum of five stages. Mean HR during the final 60 s of each stage was obtained using HR analysis software (Polar Precision Performance Software, Polar Electro, Oy, Finland). Following a five minute rest period, the

participant performed a maximal ramp protocol on the cycle ergometer. The initial power output of the ramp test was the final power of sub-maximal test minus 30%, and was increased by 5 W at 15 s intervals. The test was terminated when the participant reached volitional exhaustion or cadence fell below 70 rpm. The HR_{max} achieved during the test was obtained using the HR analysis software.

2.4. Comparative Trials

During trials 2 and 3, each participant was fitted with two of the four wrist-worn devices (one on each wrist) using a randomised and counterbalanced method. The devices tested were the Apple Watch Series One (Apple Inc., California, United States), Fitbit Charge HR (Fitbit Inc., San Francisco, United States), TomTom Touch (TomTom International B.V., Amsterdam, The Netherlands), and Mio Fuse (Mio Global, Canada). The wrist-worn devices were fitted according to the instructions of each manufacturer. Prior to ECG electrode placement, the skin was prepared with an alcohol preparation pad and, where applicable, shaved with a disposable razor. Five silver-chloride electrodes (Ambu Ltd., Cambridgeshire, UK) were then placed on the anterior torso in a standard five-lead configuration and connected to the ECG system. Each participant then performed a submaximal, step protocol on the cycle ergometer, which consisted of five, 3-minute stages at 40, 50, 60, 70, and 80% HRR (determined using linear regression from the submaximal HR responses in Trial 1). For consistency, and given that some of the wrist-worn devices did not have a facility to record HR continuously, HR data were recorded manually 10 s from the end of each stage. At the end of the final stage the participant rested passively on the cycle ergometer and recovery HR was recorded after 2 minutes.

2.5. Statistical Analyses

Data were analysed using the Statistical Package for the Social Sciences (IBM SPSS Statistics; Armonk, New York, USA). Measures of central tendency and spread are presented as means \pm standard deviation. Relationships between ECG measures of HR and those of each of the alternative devices were determined using Pearson correlation coefficients with associated 95% confidence limits. The effects of exercise intensity on HR responses across the various HR devices were evaluated using a two-way (device \times exercise intensity) analysis of variance (ANOVA), with differences in HR between each device and the ECG response as the dependent variable. Differences in recovery HR between ECG and the various devices were evaluated using a one-way ANOVA, with differences in HR relative to ECG again used as the dependent variable. In line with previous research [5] comparing HR responses between ECG and a traditional HR monitor (Polar RS400; Polar Electro Oy, Kempele, Finland) differences in HR between ECG and the various devices ≥ 5 b·min⁻¹ were considered to be outside of the normal range of variability. α was set at 0.05 for all analyses.

3. Results

3.1. Heart Rates during Exercise

There were no instances, across any of the exercise intensities, where ECG and Polar data differed by ≥ 5 $\text{b}\cdot\text{min}^{-1}$. In contrast there were two instances (2.2%) with the Apple, four (4.4%) with the Mio Fuse, 19 (21.1%) with the Fitbit, and 10 (11.1%) with the TomTom. As such, 18 participants achieved a full data-set of HR responses that were < 5 $\text{b}\cdot\text{min}^{-1}$ of the ECG responses for the Polar, 16 for the Apple, 14 for the Mio Fuse, 10 for the Fitbit, and 12 for the TomTom. Only four participants achieved a full data set from all devices without any discrepancies ≥ 5 $\text{b}\cdot\text{min}^{-1}$ relative to ECG. Heart rate responses during exercise determined from all devices are presented in Table 1, with corresponding correlation coefficients between ECG and each device presented in Table 2. Relative to ECG, there were no significant differences in HR responses between the devices during

exercise ($F = 1.499$; $p = 0.239$), and there was no device \times exercise intensity interaction effect ($F = 1.027$; $p = 0.370$). There was, however, an effect of exercise intensity on HR responses relative to ECG ($F = 3.637$; $p = 0.032$), but *post hoc* tests were unable to identify where those differences were.

3.2. Heart Rates during Recovery

Heart rate responses at the end of the 2-minute recovery period from all devices are presented in Table 3, along with correlation coefficients between ECG and each device. There were no instances, at the end of the recovery period, where ECG and Polar data differed by ≥ 5 $\text{b}\cdot\text{min}^{-1}$. In contrast, there were five instances (27.8%) with the Apple, two (11.1%) with the Mio, two (11.1%) with the Fitbit, and three (16.7%) with the TomTom. Relative to ECG there were no significant differences between the devices in HR at the end of the recovery period ($F = 2.847$; $p = 0.068$).

Table 1. Heart rates across a range of exercise intensities as determined by electrocardiogram (2 trials) and several other devices. Intensities are based on percentages of heart rate reserve. Values are means \pm standard deviation

Device	40% HRR	50% HRR	60% HRR	70% HRR	80% HRR
ECG 1 ($\text{b}\cdot\text{min}^{-1}$)	111 \pm 12	123 \pm 16	134 \pm 16	146 \pm 16	156 \pm 18
Polar ($\text{b}\cdot\text{min}^{-1}$)*	111 \pm 12	123 \pm 16	135 \pm 16	146 \pm 16	157 \pm 18
Apple ($\text{b}\cdot\text{min}^{-1}$)*	111 \pm 13	123 \pm 16 ¹	134 \pm 14 ¹	146 \pm 16	157 \pm 17
Mio Fuse ($\text{b}\cdot\text{min}^{-1}$)*	111 \pm 13 ²	122 \pm 16 ¹	134 \pm 16 ¹	146 \pm 16	156 \pm 18
ECG 2 ($\text{b}\cdot\text{min}^{-1}$)	111 \pm 11	122 \pm 14	134 \pm 16	146 \pm 15	158 \pm 16
Fitbit ($\text{b}\cdot\text{min}^{-1}$)†	106 \pm 17 ⁶	114 \pm 19 ⁶	132 \pm 20 ³	144 \pm 18 ²	158 \pm 15 ²
TomTom ($\text{b}\cdot\text{min}^{-1}$)†	109 \pm 12 ⁴	118 \pm 17 ³	135 \pm 16 ¹	146 \pm 15 ²	157 \pm 15

Note: HRR = heart rate reserve; ECG = electrocardiogram; * measured at same time as ECG 1; † measured at same time as ECG 2. Superscripted numbers indicate the number of instances where responses differed by ≥ 5 $\text{b}\cdot\text{min}^{-1}$ relative to ECG.

Table 2. Correlation coefficients between heart rate responses measured by electrocardiogram versus several other devices across a range of exercises intensities based on percentages of heart rate reserve. Unless otherwise indicated, values are based on a sample size of 18 and include all data

Device	ECG				
	40% HRR	50% HRR	60% HRR	70% HRR	80% HRR
Polar	0.998 [0.995 – 0.999]	1.000 [1.000 – 1.000]	0.999 [0.997 – 1.000]	0.999 [0.997 – 1.000]	0.999 [0.997 – 1.000]
Apple	0.995 [0.986 – 0.998]	0.984 [0.957 – 0.994]	0.958 [0.889 – 0.985]	0.995 [0.986 – 0.998]	0.995 [0.986 – 0.998]
Mio	0.973 [0.927 – 0.990]	0.954 [0.878 – 0.983]	0.993 [0.981 – 0.997]	0.997 [0.992 – 0.999]	0.999 [0.997 – 1.000]
Fitbit	0.895 [0.735 – 0.961]	0.620 [0.216 – 0.843]	0.931 [0.821 – 0.974]	0.904 [0.756 – 0.964]	0.987 [0.965 – 0.995]
TomTom	0.963 [0.898 – 0.987]*	0.712 [0.352 – 0.889]*	0.997 [0.991 – 0.999]*	0.990 [0.972 – 0.996]*	0.997 [0.992 – 0.999]

Note: ECG = electrocardiogram; HRR = heart rate reserve; * $n = 17$ (no data received from the device for one participant over the first four exercise intensities). Values in square brackets represent 95% confidence limits.

Table 3. Heart rates after 2 minutes of recovery from a step incremental test as determined by electrocardiogram (2 trials) and several other devices, along with corresponding correlation coefficients between ECG and those same devices. Heart rate values are means \pm standard deviation. $n = 18$ for all measures

	ECG 1	Polar*	Apple*	Mio Fuse*	ECG 2	Fitbit†	TomTom†
Heart rate ($\text{b}\cdot\text{min}^{-1}$)	100 \pm 20	100 \pm 20	103 \pm 20	100 \pm 20	100 \pm 18	101 \pm 18	99 \pm 19
Correlation coefficient	-	0.998 [0.995 – 0.999]	0.960 [0.894 – 0.985]	0.953 [0.876 – 0.983]	-	0.990 [0.973 – 0.996]	0.938 [0.838 – 0.977]

Note: ECG = electrocardiogram; * heart rate measured at same time as ECG 1; † heart rate measured at same time as ECG 2. Values in square brackets represent 95% confidence limits.

4. Discussion

The aim of this study was to investigate the accuracy of four commercially available wrist-worn HR monitors. The results indicate that none of the wrist-worn HR monitors demonstrated the same level of accuracy as an ECG or chest-worn HR monitor. Therefore, in accordance with previous studies, a chest-worn HR device offers the best accuracy for the measurement of HR during exercise and recovery. [5,6,12] Although, relative to ECG, there were no significant differences in HR responses between any of the devices during exercise, the number of instances where the wrist-worn devices differed by $\geq 5 \text{ b}\cdot\text{min}^{-1}$ from ECG measurements raises some concerns over their level of precision. Overall, the Apple and Mio Fuse devices offered the greatest level of accuracy across the five exercise intensities (see Table 2).

Previous wrist-worn HR monitor studies have suggested either an improvement [8], or a reduction [9,10] in precision as exercise intensity increases; the former attributed to increases in blood perfusion, and the latter to increased hand/wrist movement at faster running speeds. The present study is the first to investigate the precision of wrist-worn HR monitors during incremental exercise at various relative, rather than fixed, exercise intensities. Moreover, the use of cycling as the mode of exercise controlled for the risk of upper body movement influencing the measurements. Although, relative to ECG, there was a significant effect of exercise intensity on HR responses during exercise, that effect appears to be mostly attributable to a preponderance of errors (differences from ECG $\geq 5 \text{ b}\cdot\text{min}^{-1}$) with the Fitbit and TomTom devices at the lower exercise intensities (see Table 1 & Table 2). A reduction in the number of errors at the higher exercise intensities supports the possibility that device precision improves as blood perfusion increases. However, the fact that post hoc tests were unable to find significant differences between devices at each intensity, coupled with the absence of similar error responses with the Apple and Mio Fuse devices suggests that any such effect may be device-specific.

Despite strong correlations with criterion measurements and the absence of statistically significant differences, the failure of the wrist-worn devices to consistently produce accurate HR readings is consistent with reports on similar devices [8,9,10]. Previous research has attributed inconsistencies in HR readings, in part, to artefacts from excessive hand/wrist movement during running [6,9]. However, the present study used cycling based exercise to avoid hand/wrist movements throughout the protocol, yet there were still several instances where HR values from the wrist-worn devices differed from ECG by $\geq 5 \text{ b}\cdot\text{min}^{-1}$. While it is not possible to determine whether the responses would have resulted in more errors if running had been chosen as the mode of exercise, research into the effects of exercise mode on the precision of wrist-worn devices is certainly warranted.

Recovery HR is often measured during athletic training programmes [13], sometimes as a means of determining recovery duration during interval training [14]. Recovery HR is also measured in clinical practice as a prognostic marker [15,16]. Therefore, it is important that wrist-worn

devices are sensitive to a rapid flux in HR. In the present study, the failure of the wrist-worn devices to consistently produce accurate HR readings in recovery, despite, as during exercise, strong correlations with ECG readings and the absence of significant differences, raises some concerns. Indeed, the frequency of errors in HR from the wrist-worn devices at the end of the recovery period provides credence to the hypothesis that the precision of wrist-worn devices may be influenced by blood perfusion.

5. Conclusions

Heart rate is commonly used to monitor and prescribe cardiovascular-based exercise for athletic [2] and clinical populations [17] and is also used to estimate CRF [3]. The results of the present study indicate that wrist-worn HR monitors offer an acceptable level of accuracy in measuring HR during cardiovascular exercise. However, it is recommended that a chest-worn HR monitor is used when accuracy is paramount across a range of exercise intensities. The present study indicated instances of random error in HR readings with wrist-worn devices which, in part, might be due to individual factors, such as blood perfusion. Therefore, individuals wishing to purchase a wrist-worn HR device should, ideally, compare the device with a chest-worn HR monitor to determine if the device offers acceptable precision during the PA/exercise that will routinely be performed. If a comparison between devices is not feasible then, based on the results of this study, the Apple Watch and Mio Fuse devices appear to provide the most valid measures of HR during cycling; with the latter providing the best all-round response.

References

- [1] American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise, *Medicine & Science in Sports & Exercise*, 43. 1334-1359. 2011.
- [2] Mann, T., Lamberts R.P. and Lambert M.I. Methods of prescribing relative exercise intensity: physiological and practical considerations, *Sports Medicine*, 43. 613-625. 2013.
- [3] Sykes K. and Roberts, A. The Chester step test-a simple yet effective tool for the prediction of aerobic capacity. *Physiotherapy*, 90. 183-188. 2004.
- [4] Kinnunen, H and Heikkilä, I. The timing accuracy of the Polar Vantage NV heart rate monitor, *Journal of Sports Science*, 16. S107-S110. 1998.
- [5] Engström, E., Ottosson, E., Wohlfart, B., Grundström, N. and Wisén, A. Comparison of heart rate measured by Polar RS400 and ECG, validity and repeatability. *Advances in Physiotherapy*. 14. 115-22. 2012.
- [6] Terbizan, D.J., Dolezal, B.A. and Albano, C. Validity of seven commercially available heart rate monitors. *Measurement in Physical Education & Exercise Science*, 6. 243-247. 2002.
- [7] Spierer, D.K., Rosen, Z., Litman, L.L. and Fujii, K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *Journal of Medical Engineering & Technology*, 39. 264-271. 2015.
- [8] Stahl, S.E., An, H.S., Dinkel, D.M., Noble, J.M. and Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & exercise Medicine*, 2:e000106. 2016.

- [9] Lee, M. and Gorelick, M. Validity of the Smarthealth watch to measure heart rate during rest and exercise. *Measurement in Physical Education and Exercise Science*, 15. 18-25. 2011.
- [10] Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P. and Gillinov, M. *JAMA Cardiology*, 2. 104-106. 2017.
- [11] Tanaka, H, Monahan, K.G. and Seals, D.S. Age - predicted maximal heart rate revisited. *Journal of American College of Cardiology*, 37. 153-156. 2001.
- [12] Wajciechowski, J., Gayle, R., Andrews, R. and Dintiman, G. The accuracy of radio telemetry heart rate monitor during exercise. *Clinical Kinesiology*, 45. 9-12. 1991.
- [13] Darr, K.C., Bassett, D.R., Morgan, B.J. and Thomas, D.P. Effects of age and training status on heart rate recovery after peak exercise. *American Journal of Physiology*, 254. H340-H343. 1988.
- [14] Tocco, F., Sanna, I., Mulliri, G., Magnani, S., Todde, F., Mura, R., Ghiani, G., Concu, A., Melis, F. and Crisafulli A. Heart rate unreliability during interval training recovery in middle distance runners. *Journal of Sports Science & Medicine*, 14. 466-472. 2015.
- [15] Cole, C.R, Blackstone, E.H., Pashkow, F.J., Snader, C.E. and Lauer, M.S. Heart rate recovery immediately after exercise as a predictor of mortality. *New England Journal of Medicine*, 341. 1351-1357. 1999.
- [16] Nanas, S., Anastasiou-Nana, M., Dimopoulos, S., Sakellariou, D., Alexopoulos, G., Kapsimalakou, S., Papazoglou, P., Tsolakis, E., Papazachou, O., Roussos, C. and Nanas J. Early heart rate recovery after exercise predicts mortality in patients with chronic heart failure. *International Journal of Cardiology*, 110. 393-400. 2006.
- [17] Beale, L., Silberbauer, J., Carter, H., Doust, J. and Brickley G. Exercise heart rate guidelines overestimate recommended intensity for chronic heart failure patients. *British Journal of Cardiology*, 17. 133-137. 2010.