Thieme

Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise

Authors
Alaa Khushhal[1], Simon Nichols[2], Will Evans[3], Damien O. Gleadall-Siddall[1], Richard Page[1], Alasdair F. O'Doherty[4], Sean Carroll[1], Lee Ingle[1], Grant Abt[1]

Affiliations
1 The University of Hull, School of Life Sciences, Kingston upon Hull, United Kingdom of Great Britain and Northern Ireland
2 Sheffield Hallam University, Centre for Sport and Exercise Science, Sheffield, United Kingdom of Great Britain and Northern Ireland
3 Teesside University, School of Social Science, Humanities and Law, Middlesborough, United Kingdom of Great Britain and Northern Ireland
4 Northumbria University, Department of Sport, Exercise and Rehabilitation, Newcastle upon Tyne, United Kingdom of Great Britain and Northern Ireland

Key words
smartwatch, wearables, technology

received 07.03.2017
revised 24.07.2017
accepted 16.09.2017

Bibliography
DOI https://doi.org/10.1055/s-0043-120195
Sports Medicine International Open 2017; 1: E206–E211

ISSN 2367-1890

Correspondence
Dr. Grant Abt, PhD
The University of Hull, School of Life Sciences
Cottingham Road, Kingston upon Hull
HU6 7RX
United Kingdom of Great Britain and Northern Ireland
Tel.: +44/148/2463 397, Fax: +44/148/2465 149
g.abt@hull.ac.uk

ABSTRACT

We examined the validity and reliability of the Apple Watch heart rate sensor during and in recovery from exercise. Twenty-one males completed treadmill exercise while wearing two Apple Watches (left and right wrists) and a Polar S810i monitor (criterion). Exercise involved 5-min bouts of walking, jogging, and running at speeds of 4 km.h^{-1}, 7 km.h^{-1}, and 10 km.h^{-1}, followed by 11 min of rest between bouts. At all exercise intensities the mean bias was trivial. There were very good correlations with the criterion during walking (L: r = 0.97; R: r = 0.97), but good (L: r = 0.93; R: r = 0.92) and poor/good (L: r = 0.81; R: r = 0.86) correlations during jogging and running. Standardised typical error of the estimate was small, moderate, and moderate to large. There were good correlations following walking, but poor correlations following jogging and running. The percentage of heart rates recorded reduced with increasing intensity but increased over time. Intra-device standardised typical errors decreased with intensity. Inter-device standardised typical errors were small to moderate with very good to nearly perfect intraclass correlations. The Apple Watch heart rate sensor has very good validity during walking but validity decreases with increasing intensity.

Introduction

The measurement of heart rate (HR) during acute exercise is one of the most common and pragmatic methods for estimating exercise intensity and prescribing exercise training thresholds [2, 17], although some studies suggest that the linear relationship between HR and oxygen consumption is sometimes altered [5]. Similarly, heart rate recovery following acute maximal or sub-maximal exercise is a common method for characterising cardiorespiratory fitness and predicting mortality risk [4, 12]. Some studies though have reported that cardiac drift over time will result in a fixed HR recovery overestimating the time required to recover fully between exercise bouts [13]. Although the 12-lead electrocardiogram (ECG)

may be the 'gold standard' for measuring HR, the ECG monitoring equipment may be impractical or unrealistic for use outside of laboratory settings. Surrogate measures including HR monitors that connect wirelessly to an in-situ chest strap have been successfully validated against 12-lead ECG devices for measuring HR and heart rate variability at rest and during exercise [10, 14, 18].

Recent advances in technology have led to the integration of photoplethysmography (PPG) into wrist-worn devices for the purpose of estimating HR. The PPG technique is a simple non-invasive optical method that detects beat-to-beat pulsatile changes in blood flow [1, 11]. The Apple Watch, commercially released in 2015, is one such device that uses PPG to measure HR. Although the Apple Watch has become the world's highest-selling smartwatch with almost 12 million in sales in 2016 [3], there are very few studies that have examined its validity or reliability for measuring HR. Wallen et al. [15] recently examined the validity of the Apple Watch in 22 healthy males and females during a 1-h protocol of low exercise intensity, supine and seated rest, walking, and running on a treadmill and cycling on an ergometer. These authors reported a mean (SD) difference of –1.3 (4.4) beats.min^{-1} and limits of agreement of –9.9 to 7.3 beats.min^{-1} between the Apple Watch and an ECG. However, HR was recorded manually and the process of how HR data were extracted is not clearly explained. Wang et al. [16] examined the validity of the Apple Watch HR compared to an ECG and a Polar chest strap in 50 males and females. Participants exercised on a motorised treadmill at 3.2 km.h^{-1}, 4.8 km.h^{-1}, 6.4 km.h^{-1}, 8 km.h^{-1}, and 9.6 km.h^{-1}, for 3 min at each stage while wearing two of four wrist-worn devices (Fitbit Charge HR, Apple Watch, Mio Alpha, and Basis Peak). There was a correlation of r = 0.91 (95 % CI: 0.88 to 0.93) between the Apple Watch and the ECG. The limits of agreement ranged from –27 to + 29 beats.min^{-1} compared to the ECG. However, HR was taken only once manually at the end of each 3-min stage, which is a serious limitation and questions how well each data point represents the mean HR. There was also no indication on which wrist the Apple Watch was worn.

To our knowledge, no study has investigated the validity of the Apple Watch HR sensor during controlled walking, jogging and running, during recovery from controlled exercise, or the intra- and inter-device reliability. It is important to examine the validity and reliability of modern wearable devices because it is well established that a dose-response relationship exists between exercise intensity and health outcomes, which places emphasis on the accurate monitoring of exercise intensity. Therefore, the aim of the study was to investigate the validity and intra- and inter-device reliability of the Apple Watch HR sensor during walking, jogging, and running activities and during recovery from each of these activities.

Materials and Methods

Study population

Our study was approved by the institutional ethics committee and meets the ethical standards of the journal [6]. Twenty-nine healthy male participants were recruited and provided written informed consent. However, eight did not complete the study; one participant withdrew due to an unrelated injury, and given the heterogeneity in participant fitness seven others were excluded because they were unable to complete all three bouts of exercise (walking, jogging, and running). Participant cardiorespiratory fitness was not assessed because the relative physiological response was not a primary measure of interest. Twenty-one healthy male participants (mean [SD]; age 31.4 [7.2] y; BMI 26.1 [2.9] kg.m^{-2}) completed the study, and of these, 20 were right-hand dominant. Eleven participants were British (white skin) and 10 were Asian (brown skin), with no participant having black skin. All participants were recreationally active and involved in a wide range of activities including walking, running, resistance training and soccer. Ten participants described their fitness status as highly fit, nine as moderately fit, and one as unfit. We lost heart rate data from one participant from the right Apple Watch during running only in trial 1 due to a data recording error. The inclusion criteria were that participants be free from known disease, not taking any form of medication, and aged > 18 years. Participants with a diagnosis of cardio-metabolic disease were excluded.

Experimental design

Participants visited the exercise testing laboratory on three separate occasions. The first visit was used to screen participants for eligibility and to familiarise them with the exercise protocol. The second visit was the first testing session and included walking, jogging, and running on a treadmill (GE T2100 treadmill) at 1 % inclination for 5 min at 4 km.h^{-1}, 7 km.h^{-1}, and 10 km.h^{-1}, respectively. These speeds were selected based on pre-study pilot testing and were mean values representing walking, jogging, and running speeds. Each bout of exercise was followed by approximately 11 min of rest. Based on data from pilot testing, 11 min was sufficient time to allow data from the Apple Watches to be transferred to the paired iPhone, together with allowing the HR to return to baseline in order to avoid any carry-over effects between intensity stages. The final visit replicated the testing protocol conducted in the second visit, with the laboratory conditions maintained between trials. The mean (SD) days between the second and third sessions was 7 (4). All testing visits were scheduled at the same time of the day. All participants were advised not to eat a large meal or consume caffeine for at least three hours before testing and to avoid moderate to vigorous physical activity in the 24 h before testing.

Instrumentation and data acquisition

During each trial participants wore a Polar HR monitor chest strap (T13, Polar Electro, OY, Finland) with the corresponding watch (Polar S810i, Polar Electro, OY, Finland), placed over the handrail of the treadmill and two Apple Watch Sport devices (Series 0, watchOS 2.0.1, Apple Inc., California, USA) – one on the left wrist and another on the right wrist. Both Apple Watches connected wirelessly via Bluetooth to two iPhone 5S smartphones (Apple Inc., California, USA). The sampling time for the Polar S810i HR monitor was set at 5 s intervals. Following exercise the HR data were transferred from the Polar S810i HR monitor to the Polar Pro Trainer 5 software. To measure HR on each Apple Watch, we used the 'Workout' app. The 'Workout' app nominally records HR at 5-s intervals. On cessation of each trial the HR data were synced automatically to the 'Health' database on its paired iPhone. To retrieve the raw

HR and sampling time data from the 'Health' database, a bespoke iPhone app was written. The bespoke app was written in Xcode 7.2.1 using the language Swift 2.1 and using the methods provided by the HealthKit framework (Apple Inc., California, USA).

Data analysis

Data were log-transformed prior to analysis to avoid bias resulting from non-uniformity of error. All data were analysed using custom-designed Microsoft Excel spreadsheets [7]. The mean and standard deviation (SD) for each exercise and recovery period were used to report descriptive data. We report the standardised typical error of the estimate, standardised mean bias, and Pearson product moment correlation coefficients to assess validity, together with the 95 % limits of agreement to aid comparisons with other studies. Standardised typical error and intraclass correlation were used to measure inter- and intra-device reliability. Uncertainties in these estimates are reported as 90 % confidence intervals. The following definitions were used to interpret the strength of the Pearson correlation coefficients used to assess the validity of the HR data and the intraclass correlation coefficients used to assess the inter- and intra-device reliability of the HR data: very poor (r = 0.45 to 0.69), poor (r = 0.70 to 0.84), good (r = 0.85 to 0.94), very good (r = 0.95 to 0.994) and excellent (r ≥ 0.995) [8]. The following definitions were used to interpret the validity of the HR data using the standardised typical error of the estimate: trivial, < 0.1; small, 0.1 to 0.29; moderate 0.3 to 0.59; large ≥ 0.6 [7]. Standardised typical error was doubled prior to interpretation using the following scale: trivial, < 0.2; small, 0.2 to 0.59; moderate, 0.6 to 1.19; large, 1.2 to 1.99; very large, 2.0 to 3.99; extremely large, ≥ 4.0 [7].

Results

Validity of Apple Watch HR during walking, jogging and running

The standardised mean bias showed there was no obvious under- or overestimation of the mean HR at any of the exercise intensities (▶ **Table 1**). There were very good correlations between the left and right Apple Watches and the criterion during walking, and good correlations during jogging. For running, there was a poor correlation for the left watch, but a good correlation for the right watch. Standardised typical error of the estimate increased as the exercise intensity increased, being small, moderate, and moderate/large, for walking, jogging, and running, respectively (▶ **Table 1**). The 95 % limits of agreement are displayed in ▶ **Table 1**. Although the Apple Watch nominally measures HR every 5 s, we were able to test this by examining the exact time that each HR was recorded. The mean (SD) percent of all possible HRs recorded by the Apple Watch reduced with increasing exercise intensity but increased over time (▶ **Fig. 1**).

Validity of Apple Watch HR in recovery from walking, jogging and running

The standardised mean bias showed there were small overestimations of the mean HR after walking, jogging and running, and standardised typical error of the estimate increased as the exercise intensity increased (▶ **Table 1**). There were good correlations between the left and right Apple Watches and the criterion after walking. There were poor correlations following jogging and running (▶ **Table 1**).

Reliability of the Apple Watch HR during walking, jogging and running

The intra-device reliability (Trial 1 vs Trial 2) increased with exercise intensity such that the ICCs increased and the standardised typical errors decreased (▶ **Table 2**). The inter-device reliability between the left and right Apple Watches during Trial 2 showed very good to nearly perfect ICCs and small to moderate standardised typical errors.

Reliability of the Apple Watch HR in recovery from walking, jogging and running

The intra-device reliability (Trial 1 vs Trial 2) in recovery from exercise increased with exercise intensity such that the ICCs increased and the standardised typical errors decreased (▶ **Table 2**). The inter-device reliability between the left and right Watches during Trial 2 showed nearly perfect ICCs and small to moderate standardised typical errors.

Discussion

This is the first study to examine the validity and intra- and inter-device reliability of the Apple Watch for measuring HR during and in recovery from controlled walking, jogging, and running. We observed that the Apple Watch has very good validity for measuring HR during walking and good validity in recovery from walking. However, the validity of the Apple Watch for measuring HR during exercise decreases with increasing intensity and the proportion of HR values recorded by the watch decreases with increasing exercise intensity. The intra-device reliability is good during walking and in recovery from walking and improved with the higher exercise intensity associated with jogging and running. The inter-device reliability is very good with low standardised typical errors and good to very good ICCs.

Our findings are largely in agreement with Wallen et al. [15] who reported HR from the Apple Watch during rest, cycling and walking at three speeds: 2.7, 4.0 and 5.5 $km.h^{-1}$. They reported a trivial underestimation of the mean HR from the Apple Watch compared to an ECG (1 $beat.min^{-1}$) with the 95 % LoA being –10 to 7 $beats.min^{-1}$. The Apple Watch was the most accurate among the four devices at low exercise intensity. Our results revealed a trivial mean bias during walking at 4 $km.h^{-1}$, with the mean bias and 95 % LoA being 0 (–6 to 6) and 0 (–8 to 8) for the left and right Apple Watches, respectively. The difference between our results and Wallen and colleagues [15] may be because they measured the mean HR across different exercise modes but we measured HR during treadmill exercise only.

Wang and colleagues [16] examined the accuracy of wrist-worn watches in 50 participants during walking, jogging, and running on a treadmill for three minutes at 3.2, 4.8, 6.4, 8 and 9.6 $km.h^{-1}$, respectively. However, only 25 of the 50 participants wore an Apple Watch. These authors reported that the accuracy of the four devices, including the Apple Watch, decreased with increasing exercise intensity and our findings are in agreement with this. Although they reported a correlation of r = 0.91 (95 % CI: 0.88 to 0.93) between

▶ **Table 1** Validity of measuring HR with the Apple Watch during and in recovery from walking, jogging, and running.

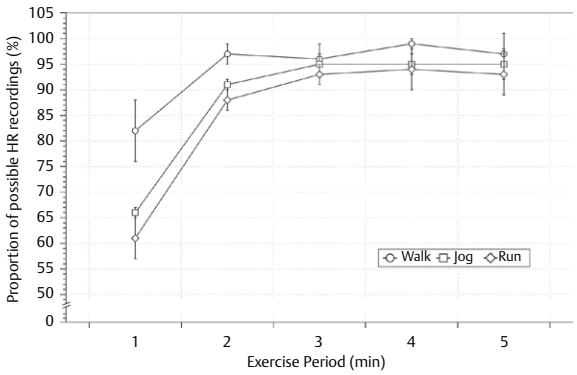
Exercise Mode	Left Wrist	Right Wrist
Walking		
Mean (SD) of criterion (Polar) (beats.min^{-1})	95 (14)	95 (14)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	94 (13)	95 (14)
Standardised mean bias (90 % CI)	− 0.03 (− 0.11 to 0.06)	0.01 (− 0.09 to 0.11)
Standardised typical error of the estimate (90 % CI)	0.23 (0.18 to 0.32)	0.26 (0.21 to 0.36)
Correlation coefficient (90 % CI) vs criterion (Polar)	0.97 (0.94 to 0.99)	0.97 (0.93 to 0.98)
Mean bias (95 % limits of agreement) (beats.min^{-1})	0 (− 6 to 6)	0 (− 8 to 8)
In recovery from walking		
Mean (SD) of criterion (Polar) (beats.min^{-1})	83 (14)	83 (14)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	88 (14)	89 (13)
Standardised mean bias (90 %CI)	0.32 (0.15 to 0.49)	0.37 (0.22 to 0.52)
Standardised typical error of the estimate (90 % CI)	0.46 (0.36 to 0.63)	0.41 (0.32 to 0.56)
Correlation coefficient (90 % CI) vs criterion (Polar)	0.89 (0.78 to 0.95)	0.92 (0.83 to 0.96)
Mean bias (95 % limits of agreement) (beats.min^{-1})	5 (− 8 to 18)	5 (− 7 to 17)
Jogging		
Mean (SD) of criterion (Polar) (beats.min^{-1})	133 (15)	133 (15)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	132 (16)	133 (16)
Standardised mean bias (90 % CI)	− 0.03 (− 0.18 to 0.11)	0.01 (− 0.15 to 0.16)
Standardised typical error of the estimate (90 % CI)	0.37 (0.29 to 0.50)	0.40 (0.32 to 0.55)
Correlation coefficient (90 % CI) vs criterion (Polar)	0.93 (0.86 to 0.97)	0.92 (0.84 to 0.96)
Mean bias (95 % limits of agreement) (beats.min^{-1})	1 (− 19 to 21)	1 (− 19 to 21)
In recovery from jogging		
Mean (SD) of criterion (Polar) (beats.min^{-1})	118 (19)	118 (19)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	131 (19)	130 (21)
Standardised mean bias (90 % CI)	0.59 (0.32 to 0.86)	0.53 (0.30 to 0.77)
Standardised typical error of the estimate (90 % CI)	0.72 (0.57 to 0.98)	0.61 (0.49 to 0.84)
Correlation coefficient (90 % CI) vs criterion (Polar)	0.71 (0.47 to 0.86)	0.80 (0.61 to 0.90)
Mean bias (95 % limits of agreement) (beats.min^{-1})	12 (− 18 to 42)	11 (− 16 to 38)
Running		
Mean (SD) of criterion (Polar) (beats.min^{-1})	155 (17)	155 (17)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	157 (18)	157 (18)
Standardised mean bias (90 % CI)	0.11 (− 0.12 to 0.35)	0.11 (− 0.09 to 0.31)
Standardised typical error of the estimate (90 % CI)	0.61 (0.48 to 0.83)	0.53 (0.42 to 0.72)
Correlation coefficient (90 % CI) vs criterion (Polar)	0.81 (0.62 to 0.91)	0.86 (0.72 to 0.93)
Mean bias (95 % limits of agreement) (beats.min^{-1})	2 (− 21 to 25)	2 (− 18 to 22)
In recovery from running		
Mean (SD) of criterion (Polar) (beats.min^{-1})	144 (22)	144 (22)
Mean (SD) of practical (Apple Watch) (beats.min^{-1})	156 (19)	155 (20)
Standardised mean bias (90 % CI)	0.50 (0.25 to 0.75)	0.44 (0.19 to 0.69)
Standardised typical error of the estimate (90 % CI)	0.67 (0.54 to 0.92)	0.67 (0.53 to 0.92)
Correlation coefficient (90 %CI) vs criterion (Polar)	0.76 (0.54 to 0.88)	0.75 (0.53 to 0.88)
Mean bias (95 % limits of agreement) (beats.min^{-1})	12 (− 18 to 42)	10 (− 20 to 40)

CI: Confidence interval.

the Apple Watch and an ECG, the limits of agreement ranged from –27 to + 29 beats.min^{-1}. Moreover, 7 of their 50 participants were African American, yet a previous study reported that the correlation between Apple Watch HR and an ECG was different between those with darker and lighter skin [15], which may have affected their results. However, our study did not include any participants with dark skin colour. Wang and colleagues [16] also recorded the HR manually at the end of each 3-min stage, which questions how well each data point represents the mean HR.

We have extended the findings of Wallen et al. [15] and Wang et al. [16] by measuring HR continuously (but nominally) every 5 s (rather than every 3 min, or manually) thereby substantially increasing the validity of the measured mean and standard deviation. The development of our bespoke in-house software allowed us to con-

tinuously record HR from the wrist watches and facilitated the collection of more frequent measurements which is previously unreported in the published literature. Although the Apple Watch nominally measures HR every 5 s, our data shows that the proportion of HR values actually measured by the Apple Watch decreases with increasing exercise intensity (▶ **Fig. 1**), which is most likely contributing to the decreased validity of the Apple Watch for measuring HR at higher exercise intensities during running. This finding might suggest that the more rapid arm movement at higher exercise intensities is increasing the movement artefact, thereby affecting the ability of the Apple Watch to measure HR. Although logic would suggest that blood flow to the wrist would be increased at higher exercise intensities, the increased movement artefact might disproportionately counteract this, leading to a degraded frequency in the HR measurement. Although we do not have direct access to the algorithms used to calculate HR from the PPG data, we can speculate that the missing HR values from the Apple Watch could result from the software purposefully not reporting HR values determined to be physiologically implausible.



▶ **Fig. 1** The mean (SD) of all possible heart rate recordings actually measured by the Apple Watch during each minute of the 5-min exercise period for walking, jogging, and running.

It is also clear from ▶ **Fig. 1** that during the first minute of exercise (particularly at higher exercise intensities), the Apple Watch is not recording between approximately 20 % and 40 % of HRs. Although we cannot say for certain, we suspect this is also related to a combination of blood-flow and motion artefact issues previously mentioned. Given that the PPG sensor estimates HR by measuring changes in blood flow, the limited blood flow to the wrist at the initiation of exercise [9] might lower the confidence of the predictive algorithms to accurately measure HR. As suggested before when related to the effect of exercise intensity on the proportion of HRs recorded, the Apple Watch software may discard all measured HRs until the algorithm is confident that it is recording a physiologically plausible value. Given these issues and based on our data, we would urge caution when analysing Apple Watch HR data of less than three minutes in duration.

Our study is also the first to examine the reliability of HR measured by the Apple Watch. Although we did not measure gait parameters during trials, there is clearly variation in the arm movement of participants within and between trials leading to variation in the amount of movement artefacts. However, our data show that the reliability between watches (inter-device) is higher than within watches (intra-

▶ **Table 2** Intra-device and inter-device reliability of HR as measured by the Apple Watch during and in recovery from walking, jogging, and running.

Intra-device					
Mode	**N**	**Time**	**Wrist**	**ICC (90 % CI)**	**STE (90 % CI)**
Walk	21	T1 vs T2	Left	0.84 (0.68 to 0.92)	0.92 (0.74 to 1.62)
Walk	21	T1 vs T2	Right	0.74 (0.51 to 0.87)	1.24 (0.98 to 1.70)
Jog	21	T1 vs T2	Left	0.82 (0.64 to 0.91)	1.00 (0.80 to 1.38)
Jog	21	T1 vs T2	Right	0.95 (0.88 to 0.97)	0.26 (0.20 to 0.35)
Run	21	T1 vs T2	Left	0.91 (0.82 to 0.96)	0.64 (0.52 to 0.88)
Run	20	T1 vs T2	Right	0.92 (0.84 to 0.97)	0.60 (0.48 to 0.84)
Walk – Rec	21	T1 vs T2	Left	0.86 (0.73 to 0.93)	0.84 (0.66 to 1.16)
Walk – Rec	21	T1 vs T2	Right	0.86 (0.72 to 0.93)	0.84 (0.68 to 1.16)
Jog – Rec	21	T1 vs T2	Left	0.86 (0.72 to 0.93)	0.84 (0.68 to 1.16)
Jog – Rec	21	T1 vs T2	Right	0.91 (0.82 to 0.96)	0.66 (0.52 to 0.90)
Run – Rec	20	T1 vs T2	Left	0.96 (0.91 to 0.98)	0.44 (0.36 to 0.62)
Run – Rec	20	T1 vs T2	Right	0.90 (0.79 to 0.95)	0.70 (0.56 to 0.98)
Inter-device					
Wrist	**N**	**Time**	**Mode**	**ICC (90 % CI)**	**STE (90 % CI)**
L vs R	21	T2	Walk	0.97 (0.94 to 0.99)	0.36 (0.30 to 0.50)
L vs R	21	T2	Jog	0.91 (0.82 to 0.96)	0.66 (0.52 to 0.90)
L vs R	21	T2	Run	0.99 (0.97 to 0.99)	0.26 (0.20 to 0.34)
L vs R – Rec	21	T2	Walk	0.98 (0.95 to 0.99)	0.32 (0.26 to 0.44)
L vs R – Rec	21	T2	Jog	0.96 (0.91 to 0.98)	0.44 (0.34 to 0.60)
L vs R – Rec	21	T2	Run	0.99 (0.98 to 1.00)	0.20 (0.16 to 0.26)

T1: Trial 1; T2: Trial 2; ICC: intraclass correlation; CI: confidence interval; STE: standardised typical error; Rec: recovery.

device), suggesting that HR is more reliable within a given exercise session than between sessions. Future studies could examine the independent and combined contribution of both blood flow and movement to the variation in HR measured by the Apple Watch.

In summary, the Apple Watch has very good validity during walking and good validity in recovery from walking. However, the validity of measuring HR decreases with increasing exercise intensity. Caution should be employed when interpreting HR data obtained with the Apple Watch during jogging and running. The proportion of HR values actually measured by the Apple Watch decreases with increasing exercise intensity and particularly during the first minute of measurement. The intra-device reliability is good during walking and in recovery from walking and improved with increasing exercise intensity. The inter-device reliability is very good.

Acknowledgements

We would like to thank Wagar Khalil for his help with data collection and all of the participants in our study for their time and commitment.

References

[1] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007; 28: R1–R39

[2] Anastasopoulou P, Tubic M, Schmidt S, Neumann R, Woll A, Härtel S. Validation and comparison of two methods to assess human energy expenditure during free-living activities. PLoS One 2014; 9: e90606

[3] Canalys. Apple Watch has its best quarter and takes nearly 80 % of total smartwatch revenue in Q4. Available from: https://www.canalys.com/newsroom/media-alert-apple-watch-has-its-best-quarter-and-takes-nearly-80-total-smartwatch-revenue-q

[4] Cole CR, Blackstone EH, Pashkow FJ, Snader CE, Lauer MS. Heart-rate recovery immediately after exercise as a predictor of mortality. N Engl J Med 1999; 341: 1351–1357

[5] Crisafulli A, Pittau G, Lorrai L, Carcassi AM, Cominu M, Tocco F, Melis F, Concu A. Poor reliability of heart rate monitoring to assess oxygen uptake during field training. Int J Sports Med 2006; 27: 55–59

[6] Harriss DJ, Atkinson G. Ethical standards in sport and exercise science research. Int J Sports Med 2009; 30: 701–702

[7] Hopkins WG. Spreadsheets for analysis of validity and reliability. Available from: sportsci.org/2015/ValidRely.htm

[8] Hopkins WG. Validity thresholds and error rates for test measures used to assess individuals. 21st Annual Congress of the European College of Sport Science; 2016; Vienna, Austria

[9] Johnson JM. Physical training and the control of skin blood flow. Med Sci Sports Exerc 1998; 30: 382–386

[10] Lee CM, Gorelick M. Validity of the smarthealth watch to measure heart rate during rest and exercise. Meas Phys Educ Exerc Sci 2011; 15: 18–25

[11] Priyadarsini P, Priyameenakshi K. The non-invasive PPG method of heart rate monitoring in smart phone device. Int J Innov Res Sci Eng Technol 2015; 4: 894–898

[12] Shetler K, Marcus R, Froelicher VF, Vora S, Kalisetti D, Prakash M, Do D, Myers J. Heart rate recovery: Validation and methodologic issues. J Am Coll Cardiol 2001; 38: 1980–1987

[13] Tocco F, Sanna I, Mulliri G, Magnani S, Todde F, Mura R, Ghiani G, Concu A, Melis F, Crisafulli A. Heart rate unreliability during interval training recovery in middle distance runners. J Sports Sci Med 2015; 14: 466–472

[14] Vanderlei LCM, Silva RA, Pastre CM, Azevedo FM, Godoy MF. Comparison of the Polar S810i monitor and the ECG for the analysis of heart rate variability in the time and frequency domains. Braz J Med Biol Res 2008; 41: 854–859

[15] Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016; 11: e0154420

[16] Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, Gillinov M. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2016, doi:10.1001/jamacardio.2016.3340

[17] Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L, Panel E. Assessment of physical activity – a review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. Eur J Cardiovasc Prev Rehabil 2010; 17: 127–139

[18] Weippert M, Kumar M, Kreuzfeld S, Arndt D, Rieger A, Stoll R. Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. Eur J Appl Physiol 2010; 109: 779–786