Journal of Sports Sciences

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: http://www.tandfonline.com/loi/rjsp20

Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities

Elizabeth A. Thomson, Kayla Nuss, Ashley Comstock, Steven Reinwald, Sophie Blake, Richard E. Pimentel, Brian L. Tracy & Kaigang Li

JOURNAL OF SPORTS SCIENCES
https://doi.org/10.1080/02640414.2018.1560644

Routledge
Taylor & Francis Group

Check for updates

PHYSICAL ACTIVITY, HEALTH AND EXERCISE

Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities

Elizabeth A. Thomson, Kayla Nuss, Ashley Comstock, Steven Reinwald, Sophie Blake, Richard E. Pimentel, Brian L. Tracy and Kaigang Li

Department of Health & Exercise Science, Colorado State University, Fort Collins, CO, USA

ABSTRACT

This study compared heart rate (HR) measurements for the Fitbit Charge HR 2 (Fitbit) and the Apple Watch devices with HR measurements for electrocardiogram (ECG). Thirty young adults (15/15 females/males, age 23.5 ± 3.0 years) completed the Bruce Protocol. HR measurements were recorded from the ECG and both devices every minute. Average HR for each participant was calculated for very light, light, moderate, vigorous and very vigorous intensities based on ECG-measured HR. A concordance correlation coefficient (CCC) was calculated to examine the strength of the relationship between ECG measured HR and HR measured by each device. Relative error rates (RER) were also calculated to indicate the difference between each device and ECG. An equivalence test was conducted to examine the equivalence of HRs measured by devices and ECG. The Apple Watch showed lower RER (2.4–5.1%) compared with the Fitbit (3.9–13.5%) for all exercise intensities. For both devices, the strongest relationship with ECG-measured HR was found for very light PA with very high CCC (>.90) and equivalence. The strength of the relationship declined as exercise intensity increased for both devices. These findings indicate that the accuracy of real-time HR monitoring by the Apple Watch and Fitbit Charge HR2 is reduced as exercise intensity increases.

ARTICLE HISTORY
Accepted 6 September 2018

KEYWORDS
Heart rate measurement; exercise intensity; wearable devices; Apple Watch; Fitbit Charge HR 2

Introduction

Exercise physiologists often prescribe exercise based on the FITT (Frequency, Intensity, Time and Type) principle (American College of Sports Medicine, 2013). Appropriate exercise intensity is a key component related to enhanced physical performance and favourable health outcomes such as longevity (Lee, Hsieh, & Paffenbarger, 1995) and cardiac health (Warburton, Nicol, & Bredin, 2006). Exercise intensity is an important variable in determining the health benefits of exercise and is a predictor of exercise-induced mortality (Warburton et al., 2006) and heart rate (HR) is a reasonable indicator of exercise intensity (Porcari, Bryant, & Comana, 2015) Therefore, accurate HR monitoring is beneficial to accurate exercise intensity monitoring. Incorrect HR readings may prompt a consumer to increase or decrease exercise intensity, which can have health ramifications and negatively affect exercise performance (Karvonen & Vuorimaa, 1988). An inaccurate HR reading may alter the perception of exercise intensity, and result in a consumer exercising outside of the actual targeted HR zone (American College of Sports Medicine, 2013; Karvonen & Vuorimaa, 1988). If HR is used to gauge exercise intensity and the consumer perceives that their exercising HR is lower than their actual HR while exercising at a vigorous intensity, the consumer may feel the need to increase their exercise intensity when it is not safe to do so.

The electrocardiogram (ECG) is the gold standard for HR monitoring; however, ECGs are limited to clinical and research settings and are not readily available to the consumer during exercise in real-world settings. In addition, other validated HR monitors, such as portable HR monitors and chest-strap HR monitors (Goodie, Larkin, & Schauss, 2000), are used to measure exercise intensity and are available to use in everyday settings. However, the inconvenience, the necessity of a receiver and the discomfort associated with these chest-strap devices are barriers for consumer use (Desilets & Mahar, 2016).

Wearable physical activity monitors (or fitness trackers) have improved greatly with the increased convenience and access of many features associated with physical activity (Lewis, Lyons, Jarvis, & Baillargeon, 2015). Numerous studies have examined many consumer based devices for validation of some variables (e.g., steps, distance, activity duration, and estimate energy expenditure) (Ferguson, Rowlands, Olds, & Maher, 2015; Gusmer, Bosch, Watkins, Ostrem, & Dengel, 2014; Imboden, Nelson, Kaminsky, & Montoye, 2017). However, HR assessment is a newer feature of these wearable physical activity monitors. The technology of photoplethysmography (PPG) measures HR using green and infra-red lights and photodetector(s) to detect changes in blood volume just below the skin surface (Kamal, Harness, Irving, & Mearns, 1989). This technology is used in both the Apple Watch (Apple Support, 2017) and the Fitbit Charge HR 2 (Fitbit, 2017) to measure HR.

Heart rate can be a difficult variable to measure in a wrist-worn device given the movement related variability during most activities. A number of studies that have examined the

CONTACT Kaigang Li kaigang.li@colostate.edu

validity of HR readings measured by several wearable devices have demonstrated variable results. Some studies concluded that overall HR accuracy is acceptable (detect accurate HR 86–87% of the time) (Parak & Korhonen, 2014), while other studies did not verify that the HR measured accuracy is acceptable. A previous study suggested that wearable devices, including the Fitbit Charge HR and the Apple Watch, underestimate HR measures (Kroll, Boyd, & Maslove, 2016; Wang et al., 2017). In addition, a few studies have determined that the PPG technology is not accurate for high intensity exercise (Cadmus-Bertram, Gangnon, Wirkus, Thraen-Borowski, & Gorzelitz-Liebhauser, 2017; Jo, Lewis, Directo, Kim, & Dolezal, 2016; Wang et al., 2017). Therefore, HR measurement in wearable devices is an area of research that warrants further investigation. Specifically, further investigation is necessary to determine the accuracy of HR measures detected by these devices under different exercise intensities.

The purpose of this study was to determine the validity of the HR measurement at different exercise intensities for two popular wearable devices with HR monitoring function, the Fitbit Charge HR 2 and the Apple Watch devices (Kahn, 2017). Measurements from each device were compared with the ECG, the criterion measure.

Methods

Subjects recruitment and health screening

All potential participants completed a health history questionnaire to minimize the risk of cardiovascular complications. The questionnaire was reviewed by research staff for exclusion criteria prior to scheduling a study test time. Exclusion criteria included a body mass index greater than 30, resting blood pressure greater than 140/90 mm/Hg (see the details in the Protocol below), a resting HR greater than 100 beats per minute (bpm) (to avoid tachycardia-induced cardiomyopathy, i.e., tachydysrhythmia, and see the details in the Protocol below), a major illness in the past, or an inability to exercise at a moderate and vigorous intensity (American College of Sports Medicine, 2013; Umana, Solares, & Alpert, 2003). A total of 30 eligible research participants (15/15 males/females) provided informed consent and participated in this single visit study. The procedures were reviewed and approved by the Colorado State University Institutional Review Board.

The data of all thirty recruited young adults (age: 23.50 ± 2.96 years, BMI: 22.84 ± 2.23 kg/m^2) were used for the analyses.

Protocol

On the test day, the participant completed the informed consent form and underwent assessment for height, weight and resting blood pressure. Blood pressure was measured while the participant rested in the supine position using an inflatable cuff, a sphygmomanometer and a stethoscope. Participants were then prepped for a 12-lead ECG (Q-Stress ECG, Mortara, Milwaukee, WI, USA). Research staff assisted participants with device placement; for consistency among participants, the Fitbit Charge HR 2 and Apple Watch were placed on the left and right wrists respectively, according to the product instructions. The device displays were oriented away from the participant so researchers could easily read the HR. Resting HR was recorded from the ECG as the participant was resting in the seated position for 3 minutes.

HR readings were taken from each device and from the ECG each minute for the entire duration of the exercise protocol. One HR reading was measured from the device and the ECG each minute to reflect how consumers use these devices during real time exercise. The exercise protocol began with three minutes of upright sitting in a chair, followed by two minutes of standing on the treadmill. After this five minute period, the Bruce protocol (Bruce, Kusumi, & Hosmer, 1973) was performed (three minute stages with speed and incline increasing every three minutes until volitional fatigue). Participants were not allowed to grip the handrails of the treadmill for the duration of the protocol. However, participants briefly rested their hands on the handrail for researchers to take a HR reading from each device every minute. After the HR reading, participant did not touch the handrails. When participants indicated they reached maximum exercise intensity, the maximum HR was obtained from the ECG as the highest HR measured by the ECG and researchers slowed the treadmill down to 2.5 mph and 0% incline. Participants walked for two minutes, then the speed was decreased to 1.7 mph for one minute, and then participants returned to sitting and remained stationary for two minutes. The Bruce protocol was chosen to produce consistent increases (every 3 minutes) in exercise intensity to test the devices response across a large range of HR. Maximal HR was not assessed prior to the Bruce protocol and was acquired after the Bruce protocol was completed. Participants indicated that they had previously run on a treadmill; however, experience level of treadmill running was not collected.

Measurements were recorded from the ECG and each device every minute during the entire protocol (warm-up, Bruce protocol and cool-down). From 50–55 seconds of each minute, the researcher on the left took a reading from the Fitbit Charge HR 2 and from the ECG. From the 55–60 seconds of each minute, the researcher on the right took a reading from the Apple Watch and from the ECG.

Heart rate measures

For all 30 participants, an average HR at each exercise intensity was calculated from the ECG-measured HR. Intensity categories were very light (<20% HRR), light (20–40% HRR), moderate (40–60% HRR), vigorous (60–85% HRR), and very vigorous (>85% HRR) (American College of Sports Medicine, 2013). The HR reserve [HRR] method, maximum HR minus resting HR, was used to calculate a HR range for each exercise intensity. For example, the light intensity range, the lower cut point of the range was calculated as the (max HR–min HR)*0.20 + min HR and the upper cut point of the range was calculated as the (max HR–min HR)*0.40 + min HR. The maximum HR and the minimum HR used for this calculation were from the ECG. The ECG displays the HR on the screen every 20 seconds based on the moment-to-moment R-R intervals. We assume that the ECG selects a stable HR in the last 10 seconds of each minute because the HR usually reached a stable value at a certain exercise intensity. Then, the ECG measured HRs that aligned with each exercise intensity were averaged. These same HR data points were then used to calculate an average for each exercise intensity for both wearable devices. HR readings that were identical to the cut points were included in the lower exercise intensity category. For example, if the cut point between moderate and vigorous intensity was 141 and a participant had a HR reading of 141 that HR reading was included in the calculation for the moderate intensity (lower exercise intensity) category.

Additionally, relative error rates (RERs) were also calculated for each device for each intensity range. The RER = (ECG HR measurement–Device HR measurement) * 100/ECG HR measurement was used to calculate a relative error rate for each intensity range (Shcherbina et al., 2017).

The authors of this paper are in favor of the artificially determined error rate of 5% as defining acceptable limits as suggested by Shcherbina and colleagues (Shcherbina et al., 2017).

Statistical analysis

Means and standard deviations of HR and relative error rate for each exercise intensity category were calculated for each device and the corresponding ECG. Paired-Samples Tests of Equivalence were performed to examine the equivalence or the lack of differences using two one-sided tests (TOST) approach between ECG-measured HR and Apple Watch- and Fitbit Charge HR 2-measured HR within each intensity category (Lakens, 2017). The statistical hypotheses of the equivalence test are: H_0: $\mu_W - \mu_E \le \theta_1$ or $\mu_W - \mu_E \ge \theta_2$ and H_1: $\theta_1 < \mu_W - \mu_E < \theta_2$, where μ_W represents the average HRs measured by test devices (Apple Watch or Fitbit), μ_E represents the average HRs measured by ECG (Schuirmann, 1987). The null hypothesis, H_0, states that μ_W and μ_E are *not* equivalent (presented with $p > .05$ and not equivalent) and the alternative hypothesis, H_1, states that they *are* equivalent (presented with $p \le .05$ and equivalent).

In the TOST procedure, an upper (θ_2) and lower (θ_1) equivalence bound are determined by medium Cohen's d (.5)×SD_{diff} (standard deviation of HR difference between wearable devices and ECG at each intensity). The medium Cohen's d was chosen considering the high correlations between HRs measured with wearable devices and ECG. Regression scatterplots and concordance correlation coefficient (CCC) were used to depict the strength of the relationship of ECG measured HR with Apple Watch and Fitbit Charge HR 2 measured HR for different intensities, overall, and split by sex. In addition, Bland-Altman plots were examined for the agreement between the Apple Watch/Fitbit and ECG. SAS (9.4, Cary, NC, USA) statistical software was used for all analyses.

Results

The overall, male, and female resting HRs are 68.33 ± 10.46, 70.13 ± 11.96, and 66.53 ± 8.77 BPM, respectively. And corresponding maximum HRs are 189.03 ± 6.64, 190.07 ± 8.36, and 188.00 ± 9.07 BPM.

For the very light intensity category (<20% HRR), average HR is 76.39 ± 11.74 BPM (ranged from 55.54–102.13 BPM); for the light category 20–40% HRR), average HR is 104.22 ± 8.64 BPM (ranged from 89.58–121.25 BPM); for the moderate category (40–60% HRR), average HR is 127.81 ± 9.74 BPM (ranged from 108.75–154.50 BPM; for the vigorous category (60–85% HRR), average HR is 157.14 ± 9.34 BPM (ranged from 133.34–76.50 BPM); and for the vigorous category (>85% HRR), average HR is 180.63 ± 9.31 BPM (ranged from 163.00–199.84 BPM).

Table 1 shows HR in BPM, the RERs for the Apple Watch and the ECG, and the results of the equivalence test. The results of equivalence test indicate that HR measurements between Apple Watch and ECG were equivalent only for the overall sample at very light intensity and for males at very vigorous intensity. None of other HR measurements between Apple Watch and ECG were equivalent for overall, males and females at any intensity. For the Apple Watch, most relative error rates were less than 5% compared to the ECG except moderate intensity for overall (5.13%), moderate (5.17%) and very vigorous (5.73%) intensities for males, and moderate (5.09%) and vigorous (5.13%) intensities for females.

Table 2 shows HR, the RERs, and the results of the equivalence test for the Fitbit Charge HR 2 and the ECG. The results of the equivalence test indicate that HR measurements between Fitbit Charge HR 2 and ECG were equivalent only for the overall sample

Table 1. HR means of Apple Watch device compared to ECG by exercise intensity.

		Apple Watch		ECG		RER	Equivalence test	
Sex	Intensity	Mean in BPM	SD	Mean in BPM	SD	%	Overall P value	Assessment result
Overall	Very Light	76.93	12.80	76.68	11.96	2.39	.02	Equivalent
	Light	102.62	9.92	104.20	8.36	3.70	.07	Not equivalent
	Moderate	121.78	13.41	127.83	9.46	5.13	.50	Not equivalent
	Vigorous	153.13	14.77	157.03	9.64	4.67	.15	Not equivalent
	Very Vigorous	177.00	12.05	179.90	11.15	4.74	.06	Not equivalent
Male	Very Light	79.55	14.63	78.59	13.55	2.39	.31	Not equivalent
	Light	102.52	9.79	103.98	8.41	4.05	.13	Not equivalent
	Moderate	123.99	15.46	130.38	10.66	5.17	.38	Not equivalent
	Vigorous	154.88	12.88	157.66	10.08	4.22	.16	Not equivalent
	Very Vigorous	179.54	14.49	179.68	13.20	5.73	.04	Equivalent
Female	Very Light	74.32	10.52	74.76	10.24	2.39	.10	Not equivalent
	Light	102.73	10.40	104.42	8.59	3.35	.16	Not equivalent
	Moderate	119.58	11.10	125.28	7.60	5.09	.77	Not equivalent
	Vigorous	151.37	16.71	156.40	9.49	5.13	.30	Not equivalent
	Very Vigorous	174.47	8.79	180.13	9.13	3.74	.71	Not equivalent

Beats per minute (BPM); Relative Error Rate = RER

Table 2. HR means of Fitbit charge HR 2 compared to ECG by exercise intensity.

		Fitbit		ECG		RER	Equivalence test	
Sex	Intensity	Mean in BPM	SD	Mean in BPM	SD	%	Overall P value	Assessment result
Overall	Very Light	75.42	10.11	76.10	11.58	4.91	.03	Equivalent
	Light	100.25	6.93	104.24	9.09	5.36	.64	Not equivalent
	Moderate	116.66	23.74	127.79	10.27	9.20	.53	Not equivalent
	Vigorous	153.13	14.77	157.25	9.80	11.01	.99	Not equivalent
	Very Vigorous	157.47	15.44	181.35	9.44	13.04	.99	Not equivalent
Male	Very Light	76.94	11.98	78.31	13.32	5.96	.16	Not equivalent
	Light	100.46	7.28	104.35	9.84	5.86	.51	Not equivalent
	Moderate	122.39	13.66	130.96	10.80	6.44	.80	Not equivalent
	Vigorous	143.00	13.61	159.39	9.58	10.09	.98	Not equivalent
	Very Vigorous	159.45	12.90	182.81	8.92	12.58	.99	Not equivalent
Female	Very Light	73.90	7.94	73.89	9.49	3.86	.04	Equivalent
	Light	100.05	6.82	104.12	8.62	4.86	.66	Not equivalent
	Moderate	110.52	30.57	124.41	8.81	12.16	.46	Not equivalent
	Vigorous	137.24	18.86	155.11	9.86	11.92	.88	Not equivalent
	Very Vigorous	155.49	17.87	179.89	10.02	13.50	.97	Not equivalent

Beats per minute (BPM); Relative Error Rate = RER

and for females at very light intensity. None of other HR measurements between Fitbit Charge HR 2 and ECG were equivalent for the overall sample, males, and females at any intensity. For the Fitbit, the RER was 5.36% for light, 9.20% for moderate, 11.01% for vigorous, and 13.04% for very vigorous intensities. Among males, the relative error rates were greater than 5% for all intensity categories (very light = 5.96%, light = 5.86%, moderate = 6.44%, vigorous = 10.09% and very vigorous = 12.58%) and among females, the RERs were greater than 5% for moderate (12.16%), vigorous (11.92%), and very vigorous (13.50%) exercise intensities.

Figures 1 and 2 show the strength of the relationship between ECG HR, Apple Watch HR, and Fitbit Charge HR 2 HR in the regression plots for five different intensities. Concordance correlation coefficients and 95% confidence intervals are illustrated in corresponding regression scatterplots. For both the Apple Watch and the Fitbit Charge HR 2, the strongest relationship between measured HRs and the ECG measured HRs were found for the very light PA with high concordance correlation coefficients (>.90). This demonstrates a pattern that individual data values are very close to the regression lines for overall, male, and female models. The next strongest relationship was found for light PA (>.70 for Apple Watch and >.60 for Fitbit for overall, male and female participants). However, as the intensity of PA increased, the relationship of HR between each device and ECG became weaker, with greater deviation of individual values from the regression line for moderate, vigorous, and very vigorous intensities.

Compared with Fitbit Charge HR 2, the Apple Watch matched ECG more accurately for all exercise intensities for overall and within each sex in terms of both the Concordance correlation coefficients and the deviation of individual values from the regression line. Neither the Apple Watch HR nor the Fitbit Charge HR 2 HR measurements matched the actual HR during the vigorous and very vigorous intensity with an exception of the Apple Watch HR for females (Figure 1(n,o)).

As indicated by the means (Tables 1 and 2) and Bland-Altman plots (Figure 3), the Apple Watch and Fitbit Charge HR 2 underestimated the HR compared with the ECG in all intensity categories except very light PA.

Discussion

This study assessed the ability of two wearable devices, the Apple Watch and the Fitbit Charge HR 2, to accurately measure HRs in healthy young people performing treadmill exercise at incrementally increasing intensities. The main findings are that the accuracy of HR measures from both the Apple Watch and the Fitbit Charge HR 2 decreased as exercise intensity increased. Additionally, the Apple Watch revealed lower error rates for all exercise intensities and revealed significantly lower error rates for very light, moderate, vigorous, and very vigorous intensities) compared to the Fitbit Charge HR 2. In addition to this study, some studies have examined the Fitbit devices (e.g., Fitbit Surge, Fitbit Charge HR) and the Apple Watch for multiple outcome variables (Dooley, Golaszewski, & Bartholomew, 2017; Farina & Lowry, 2017; Leininger, Cook, & Adams, 2016), but to our knowledge, this is the first study to examine the validity of HR measurements of the Fitbit Charge HR 2.

Although there were significantly lower mean HR readings from the Apple Watch compared with the ECG, the average difference did not exceed 6 bpm. Our findings of relative error rates ranging from 2.39% to 5.73% agree with previous research that found that the Apple Watch demonstrated an overall lower relative HR error rate (from 1.1% to 6.7% overall) for sitting, walking, running, and cycling compared with other devices (Dooley et al., 2017; Shcherbina et al., 2017).

The Fitbit Charge HR 2 had greater relative error rates compared to Apple Watch at each intensity for the overall sample and within the male and female groups. Based on the acceptable amount of error (5%) (Shcherbina et al., 2017), most of the relative error values for Fitbit Charge HR 2 were greater than 5% (ranged from 3.86% to 13.50%) except for females at very light and light intensities only. For all exercise intensities, the Fitbit device underestimated HR compared to the ECG HR and statistically significant differences were shown for light, moderate, vigorous, and very vigorous intensities. These findings reveal a discrepancy in the accuracy of this Fitbit device based on exercise intensity and these findings are practically significant given the real-world application of these devices. Our findings also agree with Wang and colleagues who demonstrated that the accuracy of HR readings from

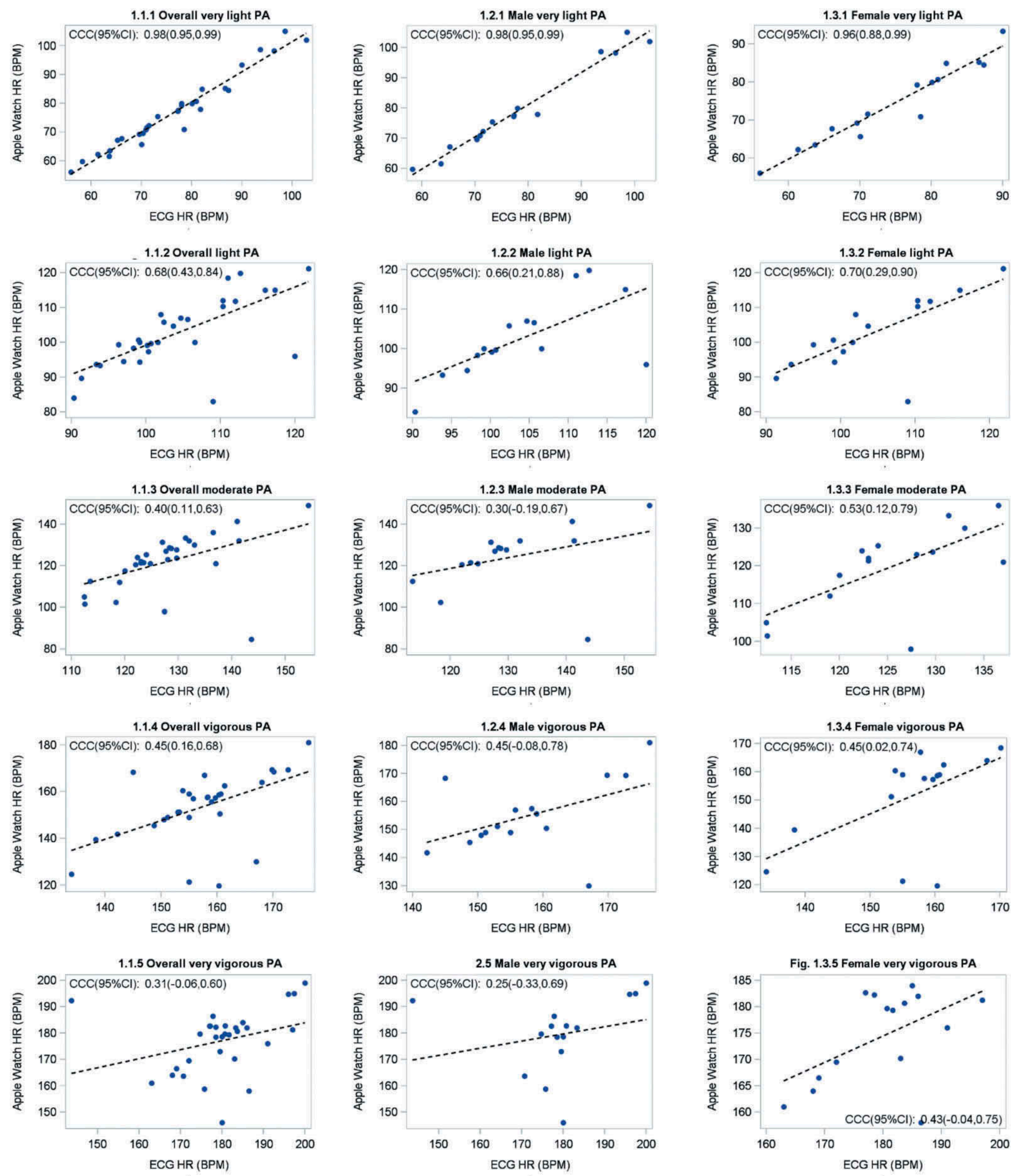

Figure 1. (a – o) Apple Watch vs. ECG heart rate regression plots.
CCC: Concordance Correlation Coefficient; CI: confidence interval; HR: heart rate; BPM: beats per minute

wrist-worn devices was best at rest and diminished with increasing exercise intensity in young adults exercising on a treadmill (Wang et al., 2017). The underestimated HR may be due to lagged outputs from the ECG especially when the HRs are expected to increase during an incremental load exercise.

Given that most previous studies (Cadmus-Bertram et al., 2017; Kroll et al., 2016; Wang et al., 2017) have found that many fitness trackers underestimate HR readings compared to ECG, our findings have added to the consensus that the Apple Watch and the Fitbit Charge HR 2 systematically underestimate HR readings during exercise performed by young healthy people.

As previous researchers has noted, detecting changes in HR at high intensities using PPG techniques can be challenging because during high intensity running, arm movement increases and previous studies have demonstrated that arm movement can interfere with PPG signalling (Zhang, Pi, & Liu, 2015; Zong & Jafari, 2015). Perhaps the Fitbit Charge HR 2 was more sensitive to artefact and to interference, which resulted in less accurate HR measurements obtained in this study. The greater sensitivity of the Fitbit Charge HR 2 compared to the Apple Watch may be due to the hardware differences, which always have an impact on the signal to noise ratio. The Apple Watch uses two green as well as infra-red light-emitting-diodes (LED) lights paired with light-sensitive photodiodes to detect the amount of blood flowing through the wrist (Apple Support, 2017; Kastrenakes, 2014). Whereas the Fitbit Charge HR 2 has two green LEDs but only one photodetector (Fitbit, 2017; Sarantos & Richards, 2016). The infra-red light and one more photodetector present on the Apple Watch could be used for filtering noise out, which may elevate the accuracy in measuring HR.

The novel aspect of this study was the analysis of HR measurement for exercise intensity tailored to each individual participant.

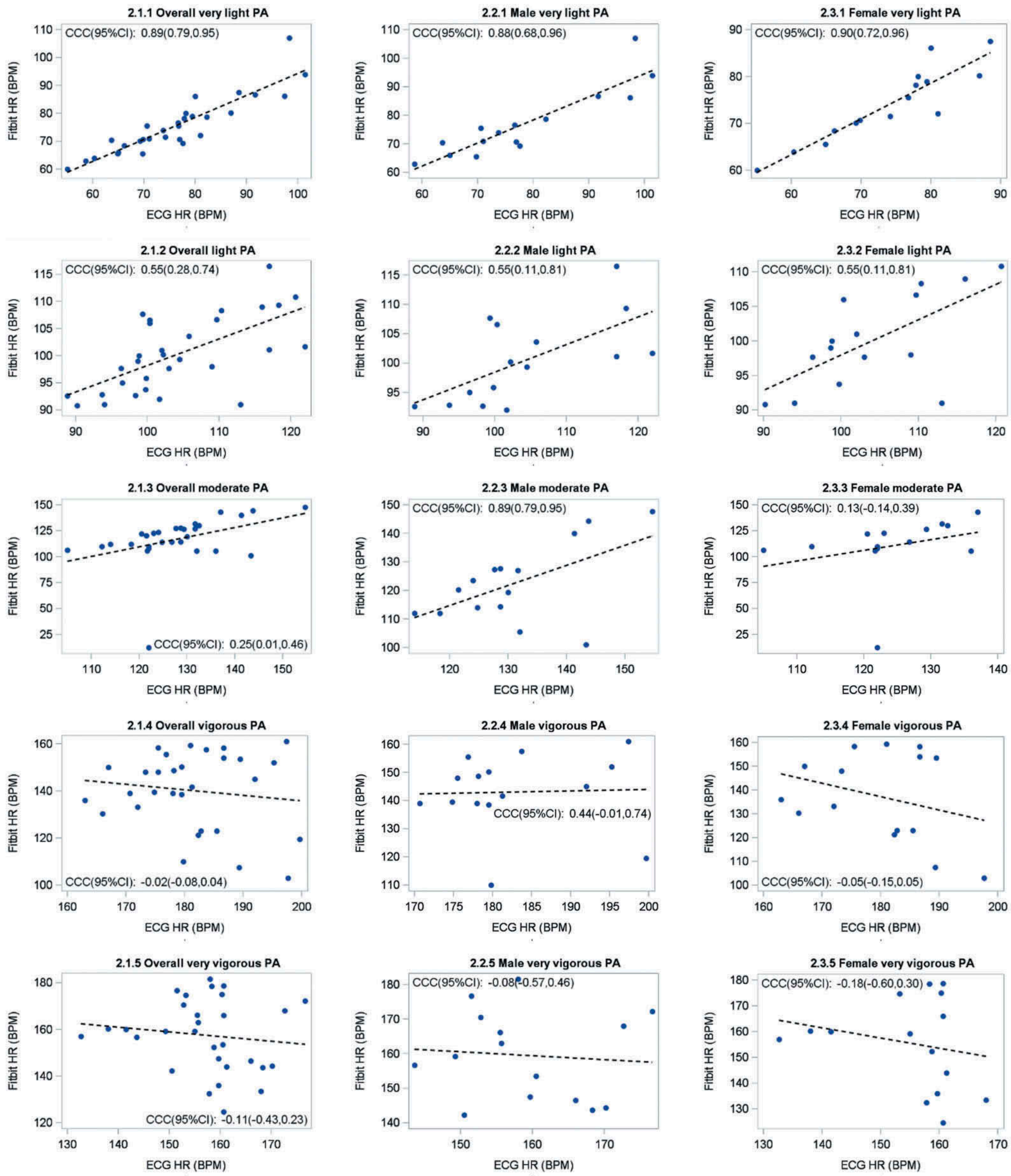

Figure 2. (a – o) Fitbit vs. ECG heart rate regression plots.
CCC: Concordance Correlation Coefficient; CI: confidence interval; HR: heart rate; BPM: beats per minute

Some previous studies (Shcherbina et al., 2017; Stackpool, Porcacri, Mikat, Gillette, & Foster, 2014) have tailored exercise intensity to participants but the measure was based on perception of intensity (e.g., Borg Rating of Perceived Exertion), which is not an objective measure. Because knowledge of exercise intensity is essential for assessing physical performance, the determination of distinct exercise intensities based on objective measures enhances the validity of this study. Our study utilized the HRR method and intensity ranges determined by ACSM to calculate individual intensity ranges for each participant. This approach determined exercise intensities based on resting and maximum HR readings, both of which were measured objectively by the ECG.

The current study also retrieved HR readings from each device and the ECG every minute (i.e., three readings for each stage) which differs from some previous studies in which HR readings were retrieved once per stage at 5 minute intervals (Shcherbina et al., 2017). Therefore, the current study addressed the need to compare these devices at fairly frequent intervals, in real-time, as a consumer would during exercise.

Some limitations exist for this study. First, only healthy subjects between the ages of 18–30 years old were recruited for participation; therefore, the findings of this study may not be generalized to other populations. Second, this study was conducted in a laboratory setting while running and therefore may not be generalizable to free-living activities. Future research should consider multiple activities to determine the accuracy of these measurements for a wide range of activities. Third, these devices were tested on each participant only one time. Each participant wore both devices simultaneously for about 30–60 minutes (for the duration of the protocol) for intentional exercise activities. This does not completely reflect the intended application of these devices; these devices are intended to be worn continuously and capture a wide range of physical activities. Therefore, future

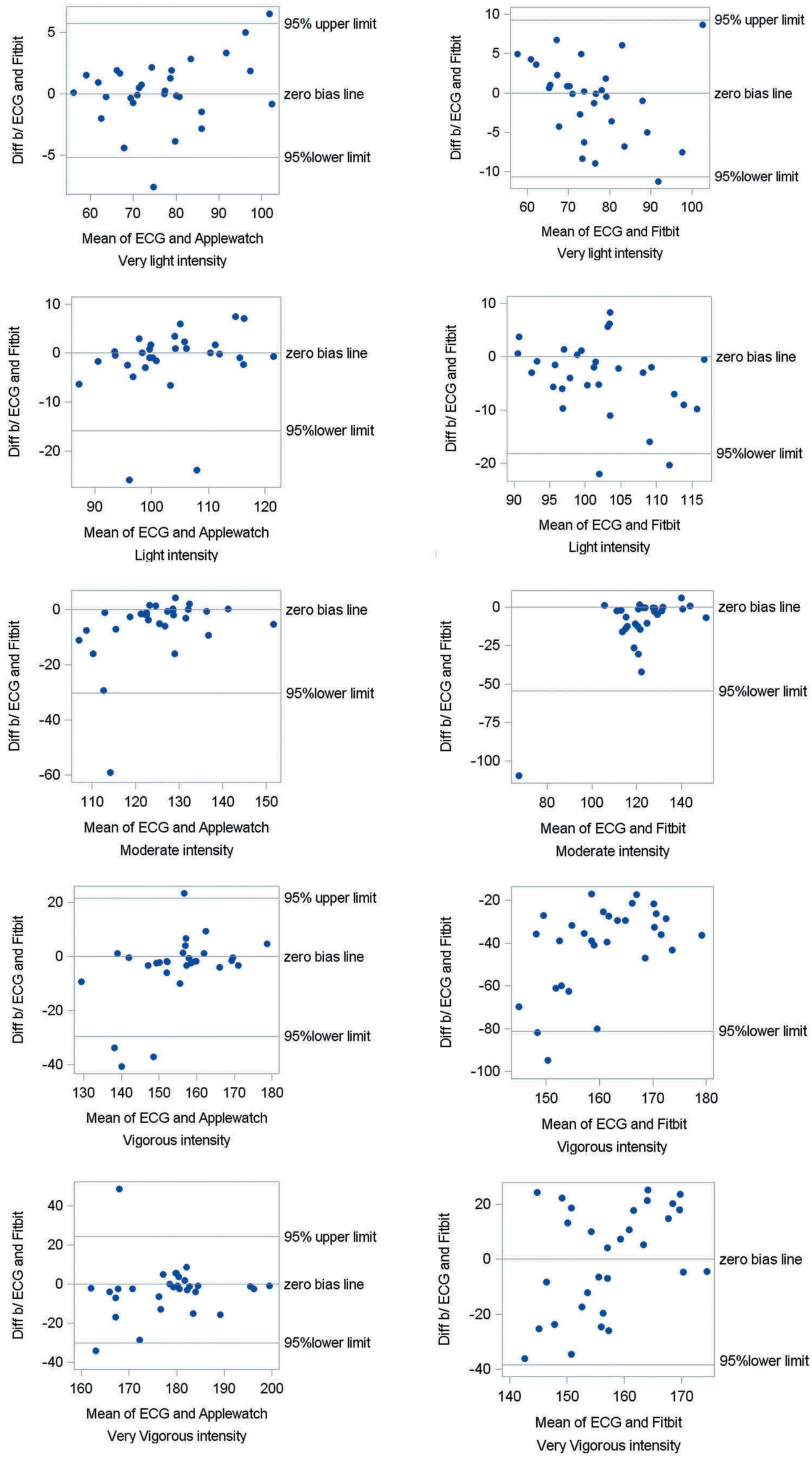

Figure 3. Bland Altman plots indicating accuracy between the Apple Watch/Fitbit and ECG.

studies should be designed to validate both devices for the measurements that are calculated and provided for the consumer in daily life. Fourth, the devices were orientated away from the exercising participant. This was done to allow the researchers to take HR readings from the devices without disrupting the exercise of the participant. Although the Apple Watch User Guide declares that the Apple Watch can be oriented backwards without influencing the readings (Apple Watch User Guide, n.d.), the results still need to be interpreted with caution as some studies report that HR measurements may be influenced by effective PPG sensor placement for reflected red and green light, and infrared wristband-type photoplethysmography (Lee, Shin, & Hahm, 2016). Fifth, we did not have access to the raw data of studied devices, which did not allow us to take into account the moving averaging rate and may have caused some errors in matching HR between studied devices and ECG due to the incapacity to determine the second-by-second time stamp. However, the research assistants were well trained and we are very confident that the human errors were minimal. Sixth, The ECG displays the HR on the screen every 20 seconds, therefore an output lag might exist when reading the HR from the ECG every minute and the output lag might have attenuated the quality check of HR output from the ECG as the reference. However, the errors due to the time lag may be counterbalanced by the steady state of workload at which HR reached a steady state. In our study, the Bruce protocol was used and the speed and incline were be kept at the same level for three minutes before increasing to another level. Because the HRs were taken at the end of each minute, we assumed that the HRs have been relatively stable (Fletcher et al., 2001); therefore, the time lag may not have impacted the actual average of HR too much during the 20 seconds or so.

Conclusions

These findings indicate that the accuracy of real-time HR monitoring by the Apple Watch and the Fitbit Charge HR2 is reduced as exercise intensity increases. Both Apple Watch and Fitbit Charge HR 2 provided valid HR measurement for very light exercise only based on the equivalence test results. However, Apple Watch performed better than Fitbit Charge HR 2 based on the error rates (Apple Watch: from 2% to 5.73% vs. Fitbit Charge HR 2: from 9% to 13%) suggesting that the former has more accurate HR assessment compared to the latter in practice. The Apple Watch and Fitbit Charge HR 2 may be more appropriate to be used to track fitness for recreational purpose if a strict HR monitoring is not required. However, researchers and practitioners should use these devices with caution for research purposes especially when dose-response relationship between heart rate and intensity is the research focus.

Acknowledgments

We appreciate Laurie Biela and Nathan Grimm in preparing the equipment and training team staff for this study. We also thank Kayla Foster, Kaitlyn Heitz, Kathryn Sowell, and Monika Teter in conducting partial pilot test for this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Kayla Nuss http://orcid.org/0000-0002-7914-5944
Richard E. Pimentel http://orcid.org/0000-0002-8734-1996

References

American College of Sports Medicine. (2013). *ACSM's guidelines for exercise testing and prescription*. Baltimore, MD: Lippincott Williams & Wilkins.

Apple Support. (2017). How Apple Watch measures your heart rate. Retrieved from https://support.apple.com/en-us/HT204666.

Apple Watch User Guide. (n.d.). Change language and orientation. Retrieved 2017, from http://help.apple.com/watch/#/apd0bf18f46b.

Bruce, R. A., Kusumi, F., & Hosmer, D. (1973). Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal*, *85*(4), 546–562.

Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The accuracy of heart rate monitoring by some wrist-worn activity trackers. *Annals of Internal Medicine*, *166*(8), 610–612.

Desilets, P.-A., & Mahar, M. T. (2016). Evaluation of the basis band fitness tracker. *International Journal of Exercise Science*, *9*(3), 2.

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating accuracy at exercise intensities: A comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR mHealth uHealth*, *5*(3). doi:10.2196/mhealth.7043

Farina, N., & Lowry, R. G. (2017). The validity of consumer-level activity monitors in healthy older adults in free-living conditions. *Journal of Aging and Physical Activity*, 1–23. doi:10.1123/japa.2016-0344

Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *The International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 42.

Fitbit. (2017). Know your heart with purePulse. Retrieved from https://www.fitbit.com/technology#.

Fletcher, G. F., Balady, G. J., Amsterdam, E. A., Chaitman, B., Eckel, R., Fleg, J., … Rodney, R. (2001). Exercise standards for testing and training: A statement for healthcare professionals from the American heart association. *Circulation*, *104*(14), 1694–1740.

Goodie, J. L., Larkin, K. T., & Schauss, S. (2000). Validation of the polar heart rate monitor for assessing heart rate during physical and mental stress. *Journal of Psychophysiology*, *14*(3), 159.

Gusmer, R., Bosch, T., Watkins, A., Ostrem, J., & Dengel, D. (2014). Comparison of FitBit® ultra to ActiGraph™ GT1M for assessment of physical activity in young adults during treadmill walking. *The Open Sports Medicine Journal*, *8*(1). doi:10.2174/1874387001408010011

Imboden, M. T., Nelson, M. B., Kaminsky, L. A., & Montoye, A. H. (2017). Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. *British Journal of Sports Medicine*. doi:10.1136/bjsports-2016-096990

Jo, E., Lewis, K., Directo, D., Kim, M. J., & Dolezal, B. A. (2016). Validation of biofeedback wearables for photoplethysmographic heart rate tracking. *Journal of Sports Science & Medicine*, *15*(3), 540.

Kahn, J. (2017). IDC: Apple Watch's record Q4 makes it #3 in wearables market as leader Fitbit declines, #2 Xiaomi widens gap. Retrieved from https://9to5mac.com/2017/03/02/idc-apple-watch-wearables-fitbit-xiaomi-market-share/.

Kamal, A., Harness, J., Irving, G., & Mearns, A. (1989). Skin photoplethysmography—A review. *Computer Methods and Programs in Biomedicine*, *28*(4), 257–269.

Karvonen, J., & Vuorimaa, T. (1988). Heart rate and exercise intensity during sports activities. Practical application. *Sports Medicine*, *5*(5), 303–311.

Kastrenakes, J. (2014). Apple Watch uses four sensors to detect your pulse. Retrieved from https://www.theverge.com/2014/9/9/6126991/apple-watch-four-back-sensors-detect-activity.

Kroll, R. R., Boyd, J. G., & Maslove, D. M. (2016). Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: A prospective observational study. *Journal of Medical Internet Research*, *18*(9), e253.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Lee, I.-M., Hsieh, -C.-C., & Paffenbarger, R. S. (1995). Exercise intensity and longevity in men: The Harvard alumni health study. *Journal of the American Medical Association*, *273*(15), 1179–1184.

Lee, S., Shin, H., & Hahm, C. (2016). *Effective PPG sensor placement for reflected red and green light, and infrared wristband-type photoplethysmography*. Advanced Communication Technology (ICACT), 2016 18th International Conference on Advanced Communication Technology (ICACT), in Pyeongchang, South Korea.

Leininger, L. J., Cook, B. J., & Adams, K. J. (2016). Validation and accuracy of FITBIT charge: A pilot study in a University worksite walking program. *Journal of Fitness Research*, *5*(2), 3–9.

Lewis, Z. H., Lyons, E. J., Jarvis, J. M., & Baillargeon, J. (2015). Using an electronic activity monitor system as an intervention modality: A systematic review. *BMC Public Health*, *15*(1), 585.

Parak, J., & Korhonen, I. (2014). *Evaluation of wearable consumer heart rate monitors based on photopletysmography*. Engineering in Medicine and Biology Society (EMBC), 2014 Chicago, IL: 36th Annual International Conference of the IEEE.

Porcari, J., Bryant, C., & Comana, F. (2015). *Exercise physiology*. Philadelphia, PA: FA Davis.

Sarantos, C. H., & Richards, P. W. (2016). *Heart rate sensor with high-aspect-ratio photodetector element*. Google Patents.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680.

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., … Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, *7*(2), 3.

Stackpool, C. M., Porcacri, J. P., Mikat, R. P., Gillette, C., & Foster, C. (2014). The accuracy of various activity trackers in estimating steps taken and energy expenditure. *Journal of Fitness Research*, *3*(3), 1-12.

Umana, E., Solares, C. A., & Alpert, M. A. (2003). Tachycardia-induced cardiomyopathy. *The American Journal of Medicine*, *114*(1), 51–55.

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *The Journal of the American Medical Association Cardiology*, *2*(1), 104–106.

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: The evidence. *Canadian Medical Association Journal*, *174*(6), 801–809.

Zhang, Z., Pi, Z., & Liu, B. (2015). TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Bio-Medical Engineering*, *62*(2), 522–531.

Zong, C., & Jafari, R. (2015). *Robust heart rate estimation using wrist-based PPG signals in the presence of intense physical activities*. Engineering in Medicine and Biology Society (EMBC), 2015 Milan, Italy: 37th Annual International Conference of the IEEE.