



Accuracy of commercially available heart rate monitors in athletes: a prospective study

Selena R. Pasadyn¹, Mohamad Soudan¹, Marc Gillinov², Penny Houghtaling³, Dermot Phelan¹, Nicole Gillinov¹, Barbara Bittel¹, Milind Y. Desai¹

¹Department of Cardiovascular Medicine, ²Department of Thoracic and Cardiovascular Surgery, ³Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA

Correspondence to: Milind Y. Desai, MD. Department of Cardiovascular Medicine, Cleveland Clinic, Desk J1-5, Cleveland, OH 44195, USA.
Email: Desaim2@ccf.org.

Abstract: Endurance athletes, particularly competitive runners, are using wrist worn devices with the heart rate (HR) feature to guide their training. However, few studies have assessed the effectiveness of these at high levels of exertion. The purpose of this study was to measure the accuracy of the HR monitor feature in four watches at six different treadmill speeds. This prospective study recruited 50 healthy, athletic adults (68% male, mean age of 29, and mean BMI of 23 kg/m²). All subjects wore a three lead ECG and Polar H7 chest strap monitor and two different randomly assigned wrist worn HR monitors. These included the Apple Watch III, Fitbit Ionic, Garmin Vivosmart HR, and Tom Tom Spark 3. Once all devices were on, they were asked to run at the following speeds on a treadmill (in mph): 4, 5, 6, 7, 8, and 9 for two min. HR was assessed on all devices and agreement among measurements determined with Lin's concordance correlation coefficient (CCC) (r_c). The Polar H7 chest strap had the greatest agreement with the ECG ($r_c=98$). This was followed by the Apple Watch III ($r_c=96$). The Fitbit Ionic, Garmin Vivosmart HR, and Tom Tom Spark 3 all had the same level of agreement ($r_c=89$). The Polar H7 chest strap was the most accurate, and the Apple Watch was superior among watches. For endurance athletes and their coaches, a chest strap device or Apple Watch may be the best choice for guiding workouts and performance.

Keywords: Athletes, heart rate (HR), wearable electronic devices

Submitted May 03, 2019. Accepted for publication Jun 06, 2019.

doi: 10.21037/cdt.2019.06.05

View this article at: <http://dx.doi.org/10.21037/cdt.2019.06.05>

Introduction

Wearable fitness tracking devices are a billion-dollar industry (1). Endurance athletes, particularly competitive runners, increasingly rely on wrist worn devices to guide their training (2,3). Distance, pace, performance zones and heart rate (HR) are just a few of the many aspects of physical activity that can be tracked through these optically based commercially available devices. In particular, the HR feature is employed by many distance runners as a marker of their fitness progress (4). Given that long distance runs, interval workouts, and even hill sprints are all part of training plans for competitive runners, the accuracy of HR monitoring becomes extremely important so that a balance is achieved across these different training loads.

Convenience and comfort of the wrist-based devices has enabled them to largely replace chest straps that employ electrodes that measure cardiac electrical activity.

Given that distance runners often have the physical and mental fortitude to push their bodies to extremes, the accuracy of these devices is important for idealized training and safety (5). Maximum HR levels prescribed by physicians or coaches are useful only in the setting of an accurate device for measuring them. There have been studies examining wrist worn HR monitor accuracy during aerobic activity, demonstrating their validity, but none at high levels of exertion (3,5-7). Also, no studies have examined the latest products on the market. Four updated wrist-worn devices commonly used by competitive distance runners include

Table 1 Baseline characteristics of participants (n=50)

Characteristic	N [%]	Mean \pm SD	15/50/85 percentile	Min/max
Age	–	29.5 \pm 9.32	20/27.5/39	18/56
Height (in)	–	68.6 \pm 3.54	65/69/72	62/75
Weight (lb)	–	153 \pm 23.5	128/152/176	107/220
BMI	–	22.8 \pm 2.4	20/23/25	18.5/28.3
Left wrist circum (cm)	–	16.2 \pm 1.12	15/16/17.5	14/18.5
Right wrist circum (cm)	–	16.3 \pm 1.06	15/16.5/17.5	14/18.5
Days per week subject performs aerobic exercise (including walking)	–	4.74 \pm 1.64	3/5/7	1/7
Female	16 [32]	–	–	–
Race				
Asian	7 [14]	–	–	–
Black	1 [2]	–	–	–
Hispanic	1 [2]	–	–	–
Other	3 [6]	–	–	–
White	38 [76]	–	–	–

the Apple Watch III, Fitbit Ionic, Garmin Vivosmart HR, and Tom Tom Spark 3. By understanding their accuracy as compared to a telemetry-based chest-strap monitor (Polar H7) and a three lead ECG, athletes can design specialized, effective, and safe training regimens.

The purpose of this study was to measure the accuracy of the HR monitor feature of four wrist-worn devices at six different treadmill speeds, including high exertion levels, to compare brand effectiveness and to determine the levels of physical exertion at which HR data is most accurate.

Methods

Participants

This prospective study recruited 50 healthy athletic adults 18 years or older from September 2018 through December 2018 (*Table 1*). Healthy athletic was defined as being able to run a mile in under 7 min and lacking any excluding health conditions as outlined below. Subjects were recruited using flyers placed around a hospital campus. The subjects were 68% male, had a mean age of 29, and mean BMI of 23 kg/m².

Subjects were assessed for their ability to perform at minimum a 12-min running protocol on a treadmill,

consisting of running at 4, 5, 6, 7, 8, and 9 mph at zero incline (8). Subjects were included if they could run a mile in 7 min or less. Subjects were excluded if they were <18 years old, had tattoos around their wrists or forearms, had a cardiac pacemaker, a heart rhythm disorder, known cardiovascular or lung disease, and/or were treated with beta-blockers or heart rhythm medications.

The Institutional Review Board approved the protocol, and all subjects provided written informed consent. The study was registered at clinicaltrials.gov (NCT03612063) before any trials were conducted.

HR monitors

In order to accurately assess HR in each subject to compare to the wrist worn monitors, a three lead ECG and Polar H7 chest strap monitor were used. The Mason-Likar electrode placement was used and allowed for the assessment of modified leads I, II, and III. The ECG was monitored on a Quinton Q-tel RMS telemetry system and ECG-based HR was determined by visual assessment by trained research personnel. Using a 3 lead ECG in this fashion is considered the gold standard for HR measurement (9). For the chest strap placed on the distal sternum, a cell phone transmitter

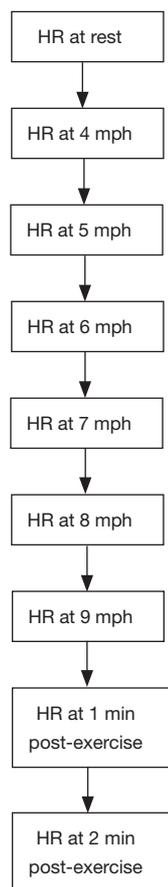


Figure 1 Demonstration of the running protocol that participants completed with time points where HR was recorded. HR, heart rate.

application (Polar Beat App) was used for the readings.

Participants were assigned at random using a computer program two different wrist-worn HR monitors, one for each wrist. Wearing a watch on the right or left wrist was also randomly determined. Previous studies demonstrate that in healthy individuals (i.e., no peripheral vascular disease) there is no difference in accuracy of the HR monitors based on which wrist they are worn on (3). Each of the 4 watches was assessed 25 times.

Each device measures HR via an optically obtained plethysmogram that is processed according to proprietary algorithms. In brief, this approach involves (I) shining a light on the skin; (II) assessing the light reflected back; (III) using the device's proprietary algorithm to determine changes in blood volume based upon reflected light; and (IV) calculating HR based upon oscillations in blood volume. The weights of the watches were as follows: 52.8 g (Apple Watch III), 50 g (FitBit Iconic), 31 g (Garmin Vivosmart

HR), and 76 g (Tom Tom Spark 3).

Exercise protocol

In each subject, right and left wrist circumferences were measured using a tape measure to ensure that no wrist was too small for the watch strap. Wrist-worn monitors were affixed above the ulnar styloid. Once all devices were on, the resting HR was recorded for each device. Subjects were then asked to run at the following levels (in mph) on the treadmill: 4, 5, 6, 7, 8, and 9 at zero incline. This incline was chosen because it is most reflective of average elevation completed on a long endurance run. It also allowed us to assess the impact of speed on HR, while limiting other variables, as an increase in incline can raise HR. Subjects were asked to run for at least 2 min at each of these speeds. HR was assessed from the four devices at 2 min of activity at each level in order to ensure steady state HR had been achieved (3). Subjects were asked to hold the treadmill bars so watch readings could be documented over a period of approximately 5 s. Values were then entered into an IRB-approved database. Once HR was recorded, subjects were given the option to rest or move the treadmill speed to the next level. After completion of all six levels, HR was assessed post-exercise at one and two min. Preliminary studies were conducted on five subjects to ensure smooth function of the protocol (Figure 1).

The treadmill settings of 4, 5, 6, 7, 8, and 9 mph for an individual running for 2 min correspond to workloads of 7.5, 9.1, 10.7, 12.3, 13.9, and 15.5 metabolic equivalent of tasks (METs), respectively (10). Each subject exercised for at least 12 min total, with variable rest time.

Statistical methods

Sample size

Sample size was based on the use of Lin's concordance correlation coefficient (CCC) (r_c) to compare HR measurements with wearable, optically based HR monitors to those obtained with the ECG. Based on prior work, we deemed an $r_c > 0.8$ to represent acceptable accuracy in HR measurement (3). Generation of 25 pairs of data for each device was necessary to provide 90% power to determine a difference from r_c of 0.82 to r_c of 0.93.

Paired differences

Paired differences were calculated by subtracting the

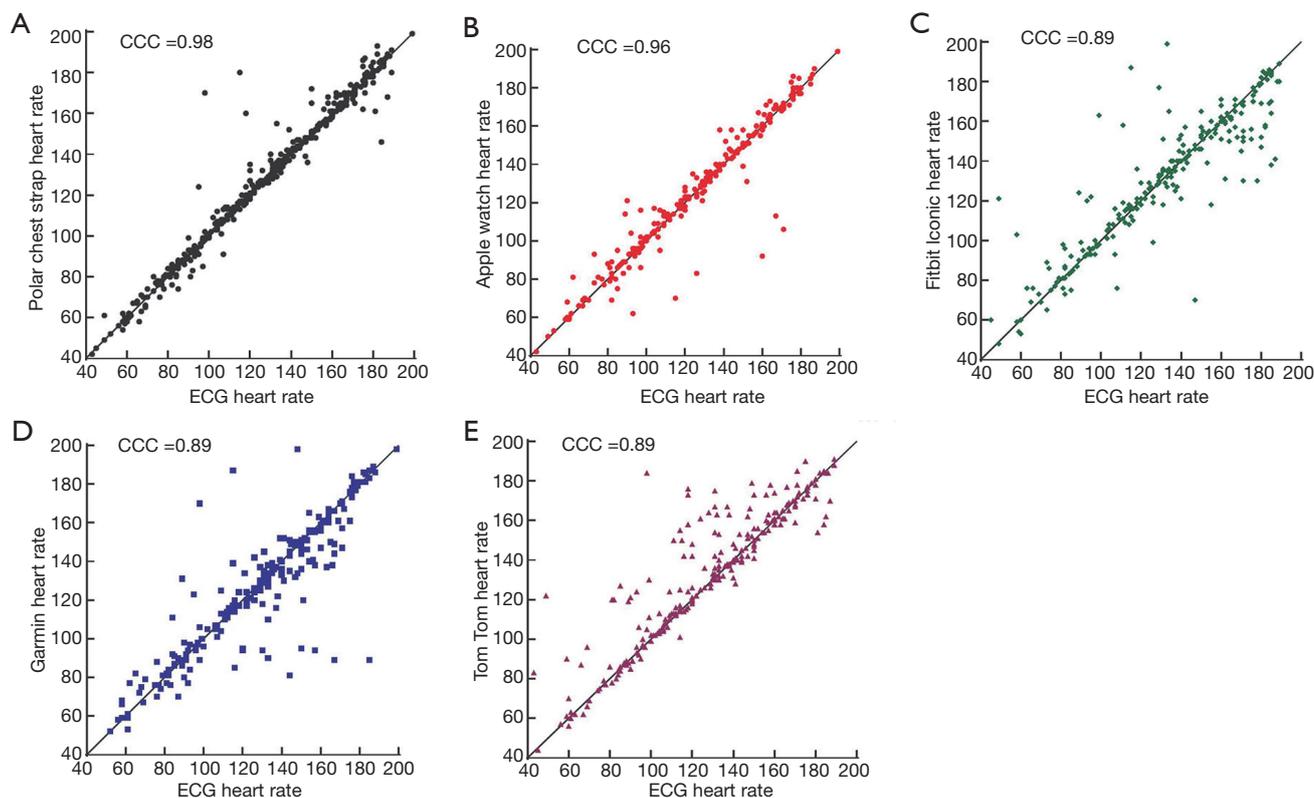


Figure 2 CCCs depicting agreement of device-measured HR with ECG. (A) Polar H7; (B) Apple Watch III; (C) Fitbit Iconic; (D) Garmin Vivosmart HR; (E) Tom Tom Spark 3. CCC, concordance correlation coefficient; HR, heart rate.

measured HR from the HR recorded on the ECG under each condition and at each time point.

Agreement

To measure agreement, Bland-Altman analysis was performed examining the differences against the means. This method uncovers any tendency for the variation to change with the magnitude of the measurement. Lin's concordance r_c were calculated to provide a measure of agreement for each device with the ECG (11).

Multivariable testing

Repeated measures mixed model analysis of variance was used to test the overall effect of the watches while adjusting for other covariates and taking into account the multiple measurements for each subject. Compound symmetry covariance structure was assumed. The first model was run with device only in the model. The second model included device and intensity of activity. The final

overall model included the additional collected covariables which included age, sex, race, wrist size, BMI, and height. Similar approaches were used to generate final models for determining factors related to (a) HR and (b) HR differences from ECG.

Data were analyzed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and R software version 3.2.3 (12).

Presentation

Continuous variables are reported as mean \pm SD, with median and percentile values. Categorical variables are reported as percent and frequency.

Results

HR data compared to ECG are shown in *Figures 2, 3*. Overall, the Polar H7 Chest Strap had the highest agreement with the ECG ($r_c=98$). This was followed by the Apple Watch III ($r_c=96$). The Fitbit Iconic, Garmin Vivosmart HR, and Tom Tom Spark 3 all had the same level

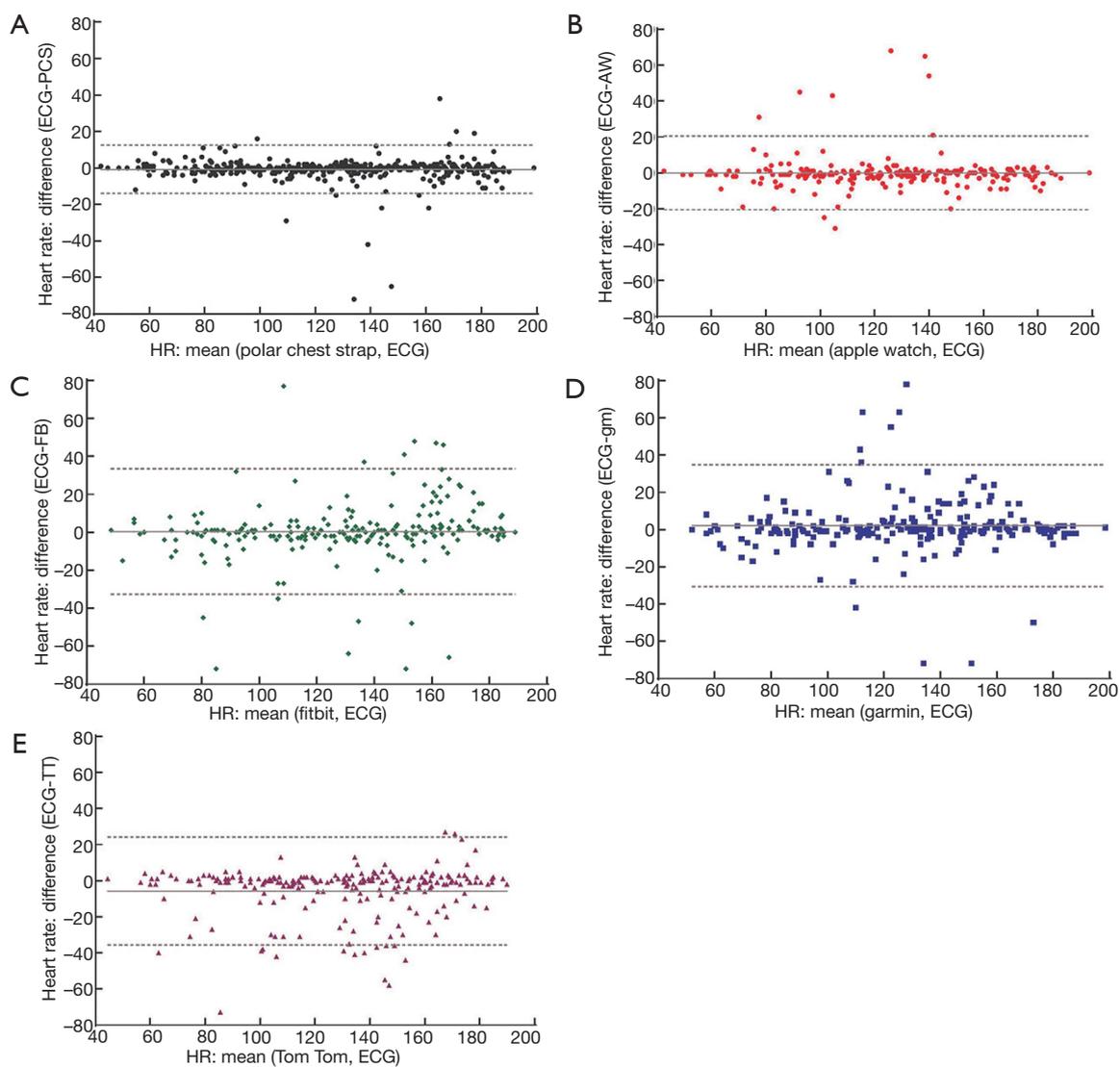


Figure 3 Bland-Altman plots and 95% limits of agreement with electrocardiographically measured HR. (A) Polar H7; (B) Apple Watch III; (C) Fitbit Ionic; (D) Garmin Vivosmart HR; (E) Tom Tom Spark 3. HR, heart rate.

of agreement with the ECG ($r_c=89$).

At rest, all devices measured accurately ($r_c \geq 85$). However, on the treadmill, accuracy of wrist-worn devices decreased as intensity increased. At 8 and 9 mph, none of the wrist-worn devices had $r_c \geq 70$. Apple Watch had the highest agreement under each condition.

Multivariable models

The final model confirmed that the Apple Watch III and Fitbit Ionic were the most accurate, with no statistical difference from ECG even after adjustment for other

factors. The P values refer to the difference from the ECG recording. The Garmin Vivosmart HR had a small underestimate of about 2 bpm ($P=0.07$). The Tom Tom Spark 3 overestimated HR on average by 6 bpm ($P<0.0001$).

BMI slightly altered accuracy ($P=0.01$) as did non-white race ($P=0.01$). HR accuracy was not influenced by sex ($P=0.9$) or age ($P=0.9$).

Discussion

Confirming prior research findings, we found that wrist worn devices are not as accurate as the Polar H7 Chest

Strap (3,13). However, for some individuals wearing a chest strap for long-distance endurance events is not comfortable or practical. This study demonstrates that in high intensity exercise the accuracy of all devices falls off, and the Apple Watch III comes the closest to the ECG standard.

In a previous study conducted by our group, agreement with the ECG was as follows: Polar H7 chest strap (0.99), Apple Watch (0.93), Fitbit Blaze (0.76), Garmin Forerunner 235 (0.92), and TomTom Spark Cardio (0.88) while on a treadmill walking and jogging at moderate speed (up to 6 mph). In that study, when biking, the Garmin and Apple Watch were acceptable ($r_c > 0.8$). On the elliptical trainer without arm levers, only the Apple Watch provided accurate readings ($r_c = 0.94$) (3). When comparing these findings to our current study, the superior accuracy of the Apple watch was replicated, and all other devices demonstrated improved accuracy on the treadmill compared to their prior versions. The Apple watch has also been shown to be superior to the Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2 in another study measuring accuracy across a variety of exercise intensities (14).

Two other studies demonstrated lower HR monitor accuracy during more vigorous exercise, specifically with the Fit Bit Charge HR, which mirrored our findings (6,15). Overall, these data provide evidence-based support for athletes concerns regarding the accuracy of currently available devices at a variety of training intensity.

Limitations

Indoor treadmill running may produce different results when compared to running outdoors or on different terrain. In addition, the measurements used were recorded during athlete grip on the treadmill handrail, which may not reflect realistic training conditions with free arm motion. Also, there may have been small errors in capturing the data because visualization of the value was used instead of an electronic, time stamped approach, which is not yet available for all of these devices. We recognize that visualization is not exact, which is why we used two trained research personnel to observe and record HR during every trial. Finally, we did not directly compare the devices between each other for statistical significance, although the individual comparisons to the ECG enable inferences concerning relative accuracy of the devices.

Conclusions

Individuals competing in extremely vigorous activity need to be able to track exertion levels and design training plans that are appropriate, in order to avoid sequelae such as overtraining syndrome, burn out, and injury. This study demonstrates a moderate to high level of accuracy of four watches for monitoring HR across many treadmill speeds. If accuracy is imperative, a chest strap or the Apple Watch III may be the best choice.

Acknowledgments

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The Institutional Review Board approved the protocol, and all subjects provided written informed consent. The study was registered at clinicaltrials.gov (NCT03612063) before any trials were conducted.

References

1. El-Amrawy F, Nounou MI. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? *Healthc Inform Res* 2015;21:315-20.
2. Diaz KM, Krupka DJ, Chang MJ, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol* 2015;185:138-40.
3. Gillinov S, Etiwy M, Wang R, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Med Sci Sports Exerc* 2017;49:1697-703.
4. Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. *Sports Med* 2003;33:517-38.
5. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA* 2015;313:625-6.
6. Wang R, Blackburn G, Desai M, et al. Accuracy of Wrist-

- Worn Heart Rate Monitors. *JAMA Cardiol* 2017;2:104-6.
7. Crouter SE, Albright C, Bassett DR Jr. Accuracy of polar S410 heart rate monitor to estimate energy cost of exercise. *Med Sci Sports Exerc* 2004;36:1433-9 .
 8. National Academy of Sports Medicine Data Collection Sheet [cited 2016 April 5]. Available online: [https://www.nasm.org/docs/default-source/PDF/nasm_par-q-\(pdf-21k\).pdf](https://www.nasm.org/docs/default-source/PDF/nasm_par-q-(pdf-21k).pdf)
 9. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-68.
 10. Murakami H, Kawakami R, Nakae S, et al. Accuracy of Wearable Devices for Estimating Total Energy Expenditure: Comparison With Metabolic Chamber and Doubly Labeled Water Method. *JAMA Intern Med* 2016;176:702-3.
 11. Bland JM, Altman DG. Agreed statistics: measurement method comparison. *Anesthesiology* 2012;116:182-5.
 12. epiR: Tools for the Analysis of Epidemiological Data [cited 2016 April 4]. Available online: <http://cran.r-project.org/web/packages/epiR>
 13. Hough P, Glaister M, Pledger A. The accuracy of wrist-worn heart rate monitors across a range of exercise intensities. *J Phys Act Res* 2017;2:112-6.
 14. Jo E, Lewis K, Directo D, Kim MJ, et al. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. *J Sports Sci Med* 2016;15:540-7.
 15. Shcherbina A, Mattsson CM, Waggott D, et al. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *J Pers Med* 2017;7:3.

Cite this article as: Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, Bittel B, Desai MY. Accuracy of commercially available heart rate monitors in athletes: a prospective study. *Cardiovasc Diagn Ther* 2019;9(4):379-385. doi: 10.21037/cdt.2019.06.05