Are Wrist-based Heart Rate Monitors a Valid Tool for Fitness Professionals to Measure Training Intensity During Exercise Classes?

Korey Little[a], John C. Sieverdes[a], D. David Thomas[b,c], M. Blake Lineberger[a], Daniel B. Bornsteind, Marco Bergamine, & Wesley D. Dudgeon[a]*

[a]*Department of Health and Human Performance, College of Charleston, Charleston SC*
[b]*Informatics Programs, University of Illinois, Champaign, IL*
[c]*Stephens Family Clinical Research Institute, Carle Health, Urbana, IL*
[d]*DBornsteinSolutions, LLC, Norwich, VT*
[e]*Department of Medicine, University of Padova, Padova, Italy*

https://doi.org/10.33697/ajur.2023.088

Student email: littlekc@g.cofc.edu
Mentor email: sieverdesjc@cofc.edu, ddt3@illinois.edu, linebergerbb@cofc.edu, dan@dbornsteinsolutionsllc.com, marco.bergamin@unipd.it, dudgeonw@cofc.edu*

ABSTRACT
This article aims to inform personal trainers and group fitness coaches about the validity and utility of wrist-located heart rate (HR) monitors compared to chest-located HR monitors for training purposes. HR from four wrist-based optical sensor HR products (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, Mio Fuse) were compared against a Polar H7 chest strap & RS800cx receiver during nine activities. Two researchers visually observed HR during a protocol incorporating resting, standing, a grocery bag carry, and a 6-stage cycle ergometer protocol that reached maximal HR. Pearson's r and interclass correlations (ICC) in the sample (n=45, mean age=20.22 [SD 2.32]) resulted in the following: Mio Fuse r=.93, ICC=.97; Apple Watch 1 r=.91, ICC=.95; Fitbit Charge HR r=.83, ICC=.91; and Garmin Vivosmart HR r=.74, ICC=.85 (all p's <.001). Bland-Altman plots showed the lowest bias for the Mio (-3.30 bpm), followed by the Apple Watch (-2.82 (SD:14.6) bpm), Garmin (-2.99 (SD:23.9) bpm) with Fitbit having the highest bias (-8.13 (SD:20.6) bpm). No drift in bias was found for any device in successive HR categories (all *p's* >.09). Wrist-based HR monitors were deemed acceptable for fitness classes, though caution should be taken when interpreting any singular visually observed measurement point.

KEYWORDS
Smartwatch; Heart Rate Monitoring; Fitness; Fitness Watch; Validity; Exercise; Cycle Ergometer; Training; Intensity

INTRODUCTION

Over the past ten years participation in group fitness classes has been a method that individuals use to engage in regular physical activity in a social and non-competitive environment. Fitness classes may utilize a range of modalities (i.e., cycling, resistance, step, and Pilates), but cycling has become one of the most practiced activities in fitness centers for people regardless of their physical conditioning level.[1, 2] Participants may monitor their exercise intensity to ensure they are meeting workout goals, avoid over- or under-exertion, and to track changes in fitness over time. Various scales including variations of the Borg Rating of Perceived Exertion (BORG) and OMNI Picture System of Perceived Exertion (OMNI) are used to evaluate intensity levels based on rating physical effort,[3] however objectively assessing the quality of physical activity (PA) can be superior but also challenging. Wrist-based activity monitors that measure heart rate (HR) have been shown to be a practical tool to better understand the quality of PA in directed fitness classes.[4]

The utility of consumer activity trackers has gained mainstream popularity and been highly marketed in medical[5-8] and commercial settings.[3, 9, 10] The wearables market, consisting of smartwatches, chest straps, armbands, head-mounted displays, sensor-embedded clothing, and wrist-secured activity trackers contain various sensor technologies that monitor a variety of biofunctions (i.e., movement, sweat response, pulse oximetry, HR, etc.), but some of these wearables require a second device to see the biofeedback and can be uncomfortable to wear during a high-intensity exercise. In 2016, 102.4 million wearables were shipped with a year-over-year growth of 25.0%.[11] Projected growth of wearables, as reported by the International Data Corporation, was expected to reach 125.5 million shipments in 2017 and grow to 240.1 million by 2021.[12] These projections were an underestimation with double the shipments in 2021 at 533.6 million units.[9] Consumer demand and marketplace competition are driving the number of innovative features, but concerns over which and how biometrics are reported to consumers has prompted the scientific community to assess the validity of features and recommend how much confidence in the metrics should be afforded. The applicability of the reported data, often displayed on a screen, could vary over a variety of activities including light to vigorous activity and by populations ranging from children to adults. Typically, assessing the validity of either research-grade or commercial-grade physical activity monitors has focused on movement and estimating energy expenditure. Metrics range from accelerometer-derived estimates of light, moderate and vigorous physical activity minutes, steps, GPS tracking, to more recent algorithmic attempts at activity pattern recognition, which remain proprietary in most cases to the manufacturer.[13-15]

HR monitoring is a common feature in wearable devices. Part of the rationale for measuring HR is its strong correlation with oxygen consumption[16] which provides wearables manufacturers an estimation of PA intensity. Prior validation of HR wearable chest strap devices show a strong correlation (r = 0.99) when compared to electrocardiogram (ECG).[17, 18] Although well validated, chest strap devices introduce wearability issues such as comfort and are not easily integrated into daily exercise. Therefore, the ability to use a lower-burdening wrist-based HR monitor alternative would be advantageous to implement for guided group exercise classes.

The convergence of several technologies measuring environmental and biometric characteristics into wrist-worn devices has predominantly incorporated optical-based photoplethysmography (PPG) to measure HR by detecting blood volume changes through the skin's surface.[19] The utility of HR measures can be used to infer and quantify the extent of exertion during physical activity, allowing those leading and participating in group exercise to understand their level of effort. As with other biometric measures, HR measures should be independently rigorously evaluated prior to widespread use.[20]

Several consumer-based validity studies have published HR monitoring results.[10, 21-25] These studies have focused on resting[10, 21-24] and exercise modalities including treadmill,[10, 21-25] elliptical,[10] and cycle ergometry[10, 22] using one or multiple stages of increasing intensities. Overall, these studies demonstrate correlations (i.e., r ≥ 0.90) comparable against electrocardiogram and chest strap devices, but resulted in varying levels of residual error depending on the device or exercise selection ranging from 1.8% to 16.7%.[10, 22-25] Conclusions drawn from these studies reported wrist-based HR measures are more accurate during rest than for exercise[21, 23] and may demonstrate sufficient validity for casual, consumer-based use but the results have a limitation.[10, 21, 23, 25] Although various ranges of HR response have been studied using pre-defined absolute intensities during exercise protocols, upper limits of HR response have not been adequately sampled requiring additional research with younger groups who can attain near maximal HR values.[26]

The purpose of this study was to test the validity of four consumer-based wrist-based HR monitors against a chest-strap monitor using visual inspection across an expanded HR range compared to previous studies to examine validity and bias at higher HRs. Results from this study will inform users on the validity of wrist-based consumer-grade HR monitors for high-intensity exercise and help individuals compare which affordable wrist-based HR monitor is best for them.

MATERIALS AND PROCEDURES

Study Design

The overall study design was cross-sectional using a concurrent validity model that captured visually observed optical HRs from five devices during a nine-stage activity protocol. This included four consumer-available wrist-based optical HR devices with a Polar HR chest strap used as the criterion device. Pretesting was performed to ensure nominal values were obtained against the chest strap over a range of exercise intensities prior to the start of data collection. All procedures were approved by the University's ethical review board prior to completion of any research activities. Written informed consent was collected from each participant prior to the conduct of any research activities (IRB# IRB-2016-05-06-143848).

Participants

A convenience sample (n = 45) consisted of undergraduate university students recruited through flyers and classroom announcements. This sample was chosen due to the ability to motivate participants to elicit a near-maximal HR response. Inclusion criteria consisted of college students or university employees between 18 and 30 years of age, being male or female, and belonging to any race or ethnicity. Exclusion criteria included metabolic dysfunction (i.e., type I or II diabetes), implanted pacemaker, and cardiovascular or cerebrovascular conditions. Prior to signing the informed consent, the qualification to be included in the study was determined by passing the Physical Activity Readiness Questionnaire Plus (PAR-Q+) in addition to not taking medications that affect HR or blood pressure.[27] No participant experienced adverse events during the conduct of the study.

Devices

Selection of devices was limited to models available during the first quarter of 2016. The concurrent validity device was a Polar chest strap HR monitor (Model Polar H7 with polar RS800cx watch receiver; Polar Electro Inc., Kempele, Finland). HR measures from this model have been extensively used in other research studies.[17, 28] Studies indicate the Polar chest strap HR monitor exhibits high correlation and low error (r = 0.99, ICC > 0.999, error rate = 0.086%) against ECG measurements.[10, 21, 29] The four wrist-based HR devices included the Apple Watch (Apple) (1st generation, OS: 2.82 firmware: 57.11; Apple Inc., Cupertino, CA), Garmin Vivosmart HR (Garmin) (OS: 2.99, firmware: 93.85; Garmin Corp., Olathe, Kansas USA), Fitbit Charge HR (Fitbit) (OS: 80.77, firmware: 8.13, Fitbit; San Francisco, California USA) and Mio Fuse (Mio) (OS: 49.34, firmware: 1.18; Mio Global; Vancouver, British Columbia Canada). Each device used its own proprietary software. All were based on optical sensor technology to measure HR. For the duration of the study, auto-updates were disabled to maintain device firmware.

Development

The protocol was developed to include equal representation of HR measures along the total range of individuals' responses from resting to near maximum effort using maximum HR ranges (i.e., Astrand formula: max HR = 220 - age).[26] Both free-living stages (i.e., resting, standing, and a simple carry task) and structured exercises (i.e., using a cycle ergometer) were included in the study protocol. Cycle ergometer stages had increasing levels of intensity based upon individual HR ranges **(see Table 1)**. Five participants not included in the analysis were used to collect developmental HR responses to verify measurement procedures, time (i.e., 3 minutes) to establish a steady-state HR measure, device wear locations, and HR response ranges. In our preliminary testing, we found no effect of device location on HR when the manufacturers' instructions were followed. Therefore, device location and fitting were standardized to limit inadvertent pressing of the devices' buttons during the protocol. Since PPG measures HR at the capillary bed level, the authors surmised that counterbalancing devices with different placements was unnecessary compared to other wrist-based variables such as accelerometry, where measures are more susceptible to error due to biomechanical moment arms during movement. Measurement samples were visually taken during a short period (i.e., <10 seconds) with the participant's forearm motionless and parallel to the floor. Visual inspection was used in the methodology to capture HR data to mimic data collected in the field from students' perspectives. Participant burden was 1 hour over 1 visit and no incentive was provided.

Stage	Activity description	Estimated time (min)	Cumulative Time (min)
1	At rest (seated)	1	1
2	Standing desk task reading material	3	4
3	Grocery bag carry with 22 kg in each bag, one in each hand – 10 meter distance, placing bag down at each end.	3	7
4	Cycle (resting) (take measurement when HR<100) no pedaling	3	10
5	Cycle stage 1: target HR (90-120)	3	13
6	Cycle stage 2: target HR (121-135)	3	16
7	Cycle stage 3: target HR (136-150)	3	19
8	Cycle stage 4: target HR (151-175)	3	22
9	Cycle stage 5: target HR (176+) max effort (max 2 min	2	24

Table 1. Activity protocol.
Estimated time includes time to reach steady state, measurement and transfer to next stage. Simultaneous heart rate measures taken during last 10 seconds of each stage.

Protocol

After written informed consent, each participant completed a demographic questionnaire consisting of self-reported age, race, sex and ethnicity with laboratory measures for height (Seca Portable Stadiometer, model 213, Seca Gmbh & co., Hamburg, Germany), weight (weight scale; Seca Gmbh & co., Hamburg, Germany Model: 869 1321004) and wrist circumference (spring-loaded measurement tape; Gulick II, Country Technology, Inc. Gays Mills, WI). Prior to the protocol stages, each participant was fitted with devices in the following order **(See Figure 1)**. On the right arm, the Fitbit was located closer to the hand with the Apple Watch closer to the elbow. On the left arm, the Mio was closer to the hand with the Garmin closer to the elbow. The Polar chest strap was placed around each person's chest and secured per the manufacturer's instructions with the receiver held by a research assistant. Participants remained seated while devices were adjusted until all resting HRs were consistently displayed. During each measurement stage, two research assistants, positioned on either side of the participant, took HRs from all devices simultaneously within ten seconds of ceasing the free-living activity stages (stages 1-3) or while continuously pedaling during the cycle ergometer stages (stages 5-9). One research assistant would take the measurement from one hand, and the other would take the Polar HR monitor and the other two with the Polar HR taken first. If a device's HR measure was missing and or delayed for greater than ten seconds, then the data point was recorded as missing for that device.

Figure 1. Device Placement. Right arm (Fitbit closer to hand, Apple Watch). Left arm (Mio closer to hand, Garmin).

Data Analysis

Demographics were analyzed using descriptive statistics including means, standard deviations, and percentages. Normality was tested using Shapiro-Wilk testing. Missing data was not imputed in the analyses. Device HR output metrics were analyzed using Pearson's correlations with a significance level of 0.05. An a-priori Pearson's correlation value of $r \geq 0.80$ compared to the Polar chest strap was chosen as the acceptability cut-point as used by Gillinov and colleagues.[10] For Interclass-correlations (ICC), a value of above 0.70 was used as a cut-point, and for mean absolute percent error (MAPE), a value lower than 5% was used as reported in other validity studies.[4, 30] Other validity measures included mean absolute error (MAE), calculation of the bias and 95% limits-of-agreement using Bland-Altman plots for comparison to other studies. Secondary analysis of MAPE was stratified by the 9 activities sorted by increasing average HR to assess trends. Change in MAPE across intensity stages were analyzed using multi-variate analysis controlled for sex and race. Data was analyzed using IBM SPSS Statistics, Version 27.0 (Armonk, NY: IBM Corp.).

RESULTS

Forty-five participants were recruited and completed the protocol. Descriptive statistics are provided in **Table 2.** Participants were college-aged (mean: 20.22, Range: 18-28) and were predominantly White (93.0%) and female (82.2%). No participants had tattoos present at the wear location.

Variable	Measure
Age, years, mean (SD)	20.22 (2.32)
Sex	
Female, %	82.2%
Male, %	17.8%
Race	
White n, %	42, 93.3%
Black n, %	2, 4.4%
Asian n, %	1, 2.2%
Wrist size (cm), mean (SD)	15.59 (1.08)
BMI (kg/m^2), mean (SD)	26.07 (18.31)

Table 2. Descriptive Characteristics.
SD: standard deviation; n = 45

Table 3 reports the validity metrics from the study. The mean and standard deviation for HR were similar for all devices in the study. Pearson's correlation coefficients for the devices compared to the Polar HR monitor from strongest to weakest were ordered Mio (r = 0.93), Apple Watch (r = 0.91), Fitbit (r = 0.83), and Garmin (r = 0.74), respectively. All devices except the Garmin were higher than the predetermined cut-point of r = 0.80. Additional Pearson's correlations between devices showed the strongest association for the Apple and Mio devices (r = 0.90), followed closely by the Fitbit and Apple (r = 0.86) and Fitbit and Mio (r = 0.81), with moderate correlations between the Apple and Garmin (r = 0.74), Garmin and Mio (r = 0.71) and Fitbit and Garmin (r = 0.63). All associations were statistically significant at $p < 0.05$. Interclass correlations for each device were rated higher than the 0.70 cut-point with the same rankings as the Pearson's correlations (all ICC's > 0.85, $p<.001$). MAPE was relatively high for each device with the Apple Watch and Mio showing mean values at 5% or below. The Garmin showed the highest MAPE (12.07%).

	Criterion: Polar HR	Fitbit	Apple	Garmin	Mio
HR, m(SD)	121.00(33.8)	115.1(36.5)	119.8(34.1)	120.1(30.0)	121(34.2)
Pearson's r	-	.834 (p<.001)	.913 (p<.001)	.739 (p<.001)	.934 (p<.001)
ICC	-	.909 (p<.001)	.954 (p<.001)	.845 (p<.001)	.966 (p<.001)
MAE	-	10.419	5.286	13.627	4.484
MAPE, %	-	8.563	4.386	12.066	4.088
Activity 1	84.2 (14.14)	3.570	3.477	9.396	4.390
Activity 2	92.8 (16.98)	7.738	4.227	12.326	3.272
Activity 3	99.0 (19.51)	8.883	4.415	28.706	6.187
Activity 4	94.3 (15.27)	9.820	4.037	12.144	8.628
Activity 5	112.8 (9.97)	10.185	6.206	10.679	2.165
Activity 6	128.8 (5.26)	10.632	1.456	4.471	2.448
Activity 7	144.9 (6.31)	6.009	4.083	7.060	1.819
Activity 8	162.7 (6.26)	1.040	.569	8.615	.554
Activity 9	178.3 (4.84)	4.076	.632	9.172	.479
p-trend	-	p=.40	p=.09	p=.42	p=.036
Bias, m(SD)	-	-8.134(20.60)	-2.816(14.57)	-2.987(23.94)	-3.303(25.31)
LOA 95% CI	-	(32.25, -48.52)	(25.74, -31.37)	(43.94, -49.91)	(24.12, -25.22)

Table 3. Validity and agreement measures between devices.
m(SD) = mean (standard deviation), ICC = interclass-correlation, MAE = mean absolute error
MAPE = mean absolute percent error. LOA = limits of agreement.

Examining MAPE by activity is visually depicted in **Figure 2**. In general, we see lower MAPE at rest (Activity 1) and at higher activity levels (Activity 8 and 9). Higher MAPE was seen in the Garmin, especially during Activity 4, followed by the Fitbit, which increased in moderate to higher HR activities then reduced at higher intensities. Trends for the Fitbit, Apple, and Garmin were flat as intensity increased but the Mio had a trend toward a smaller MAPE as intensity increased (p = .036). To provide a real-world metric to describe the error, time-aligned HRs from each device were compared to the Polar criterion to establish a percentage where readings were within 5 bpm of each other. Percentages were highest (i.e., higher is better) for the Mio at 85.2% and Apple Watch at 82.2%, followed by the Fibit at 66.6% and the lowest for the Garmin at 57.4%. As an example, one would interpret the data as: 85.2% of the time, the values from the Mio device were within 5 bpm of the Polar device.

Figure 3 reports Bland-Altman plots with 95% limits of agreement (LOA) for each device compared to the Polar chest monitor. Plots illustrate the distribution of HR residuals across the observed HR range for all 45 participants with HR estimates (y-axis: wrist-based device subtracted from the Polar (bpm)) relative to the mean of the two comparison devices (x-axis: mean of Polar and respective devices' (bpm)). Results show the LOA have the tightest clustering for the Mio (LOA: 24.12, -25.22), followed by the Apple Watch (LOA: 25.74, -31.37), Fitbit (LOA: 32.25, -48.52), and Garmin (LOA: 43.94, -49.91). Using all data, each device was negativly biased, underestimating HR compared to the Polar, with the lowest bias for the Mio (-.55), followed by the Apple Watch (-2.82), Garmin (-2.99) with the Fitbit having the highest bias (-8.13). Exploratory analysis was performed using a subjective author-determined value of +/- 20 bpm difference to mimic ignoring a very low or high HR value reported to the user. This resulted in the lowest adjusted bias values for the Mio (-.16), followed by the Garmin (-.28), Apple Watch (-.43) with the Fitbit having the highest bias (-1.71). Each device had the majority of their residuals around 0, though error values were more noticable for the Fitbit and the Garmin devices. In general, the shape of the residuals for all devices generally showed more error near the center of the HR distribution compared to the tails, with the majority of error approximately in the 90-130 bpm range. Multivariate regression showed all devices had slopes that were not different than 0 (all p's > 0.09), meaning no systematic bias was found as HR response increased. Further modeling controlling for race and/or sex showed no systematic bias (all p's > .10).

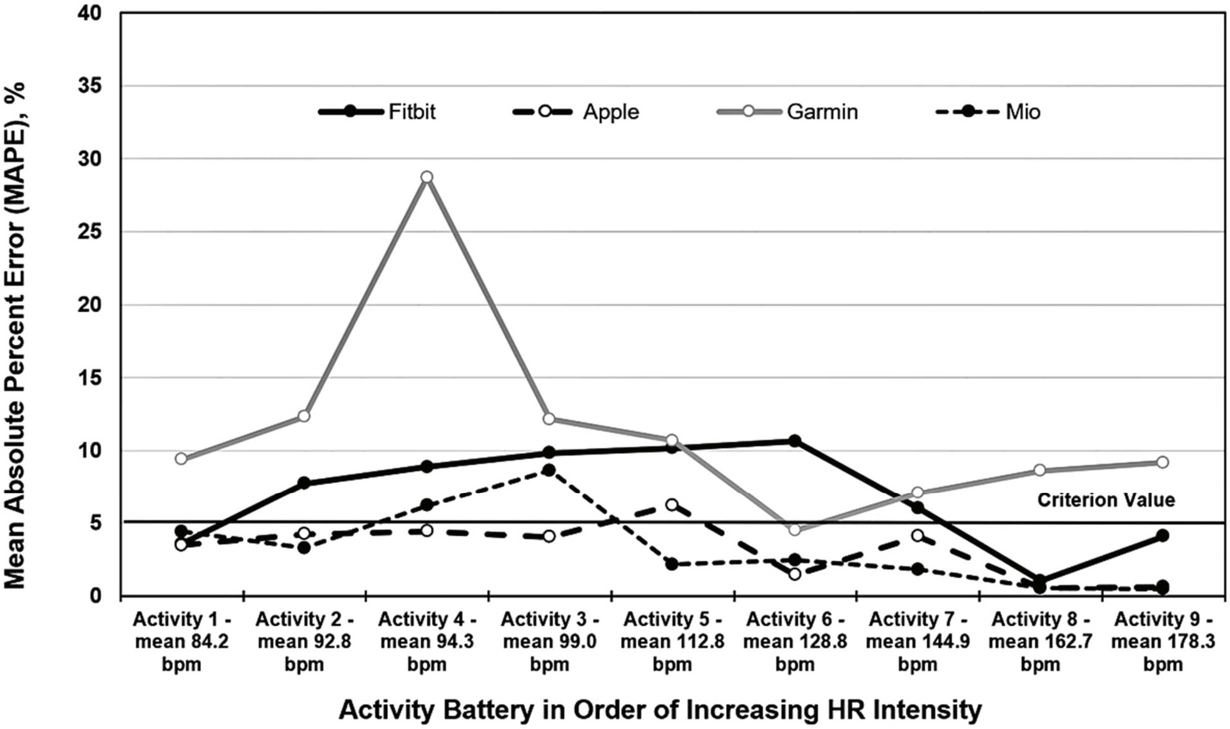

Figure 2. Mean absolute percent error for heart rate stratified by activity stage and device (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, and Mio Fuse compared to Polar H7 chest strap and RS800 cx receiver criterion device. Cycle values represent targeted HR ranges during cycle ergometry stages.

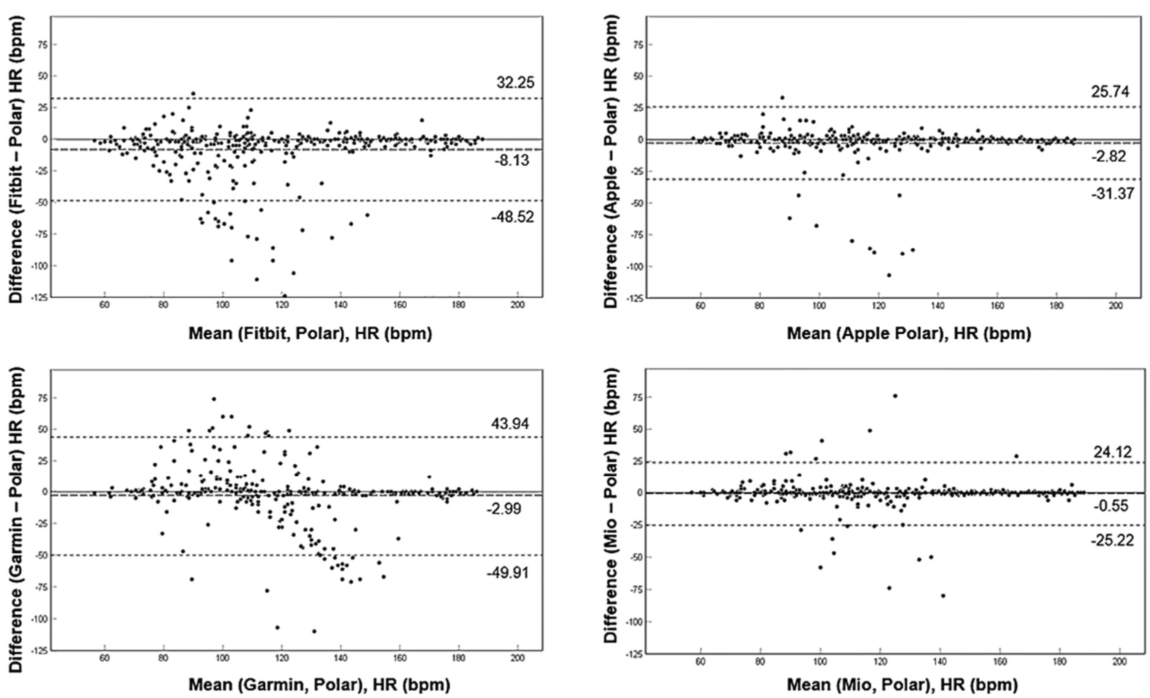

Figure 3. Bland Altman plots of each consumer wrist-based heart rate monitors (Fitbit Charge HR, Garmin Vivosmart HR, Apple Watch series 1, Mio Fuse) compared to Polar H7 chest strap and RS800 cx receiver showing overall bias (solid line) and 95% limits of agreement.

DISCUSSION

This study examined the concurrent validity of four commercially available, consumer-based, wrist-worn, HR activity monitors to inform group exercise leaders, coaches, personal trainers, and participants on the validity of HR measures across a wide range of HRs. Overall, our findings suggest moderately acceptable validity for the wrist-based monitors compared to the chest-based criterion with certain models exhibiting higher error than others. The four monitors showed moderate to high correlations and ICC's but MAPE varied with HR intensity. The device with the lowest overall error, bias, and correlations was the Mio Fuse closely followed by the Apple Watch with higher error for the Fitbit and Garmin devices. Overall, the Mio and Apple Watch had comparable and lowest bias, highest correlation, and narrowest limits of agreement.

This article specifically targets the validity of wrist-based devices at near maximum HR. All devices were found to have no slope regarding bias as HR intensity increased instilling confidence that higher intensities will not be over or underestimated more than lower HR ranges. However, random variability of HR observations outside of 5 bpm was endemic for all devices. This was most prevalent in the Garmin device but relevant for all devices, especially during moderate-intensity exercise. We hypothesize several sources of error. One is the onset of sweat or change in perfusion of tissues may introduce artifacts through altering the optical sensor reflection or changes in skin permeability.[31] Error may also be influenced by nitric oxide release, causing vasodilation that alter skin thickness.[31]

Our results corroborated other authors' findings with similar device iterations of their time. In Wang et al., 2017, several wrist-worn devices were compared to ECG during a treadmill protocol. They found the Apple Watch series 1 correlation to be $r = 0.91$, which was identical to our findings ($r = 0.91$) with comparable results for the Mio Fuse ($r = 0.91$ compared to the present study findings: $r = 0.93$). They also reported a wide 95% LOA for all measured devices including the Fitbit Charge HR (range = 73.0), while we found a higher range (range = 93.9).[21] An article by Cadmus-Bertram et al., 2017 reported a similar unacceptable 95% LOA for the Fitbit HR (bias: -2.5, range: 77.0) and Mio Fuse (bias: 1.8, range: 48.5) during exercise, though improved at rest.[23] Similar to Cadmus-Bertram et al., our findings also showed a large negative bias in the Fitbit device (bias = -8.13 (SD 20.6)) with a wide 95% LOA. Stahl et al., 2016 reported acceptable findings during walking and running treadmill activities with all correlations > 0.92.[24] These values were higher compared to the other studies' correlations, including ours.[10, 21] In Stahl et al., the authors analyzed several devices including the Mio Alpha (MAPE = 4.6%, $r = 0.93$) and the Fitbit Charge HR against a Polar RS400 chest strap (MAPE = 6.2%, $r = 0.93$). Their study found lower error rates compared to the current study which may be attributed to activity selection and intensities observed.[24] Gillinov et al., 2017 used a criterion of $r > 0.80$ as a measure of acceptability using running, biking, and elliptical stages that were continuously measured and downloaded for comparison.[10] Across activities, only the Apple Watch, TomTom Spark and Garmin Forerunner were acceptable where the Fitbit Blaze and Scosche Rhythm+ were not.[10] Using their criteria, only the Garmin ($r = 0.74$) would not reach acceptability in our findings. Their results showed the Apple Watch had the lowest MAPE (4.1%) which was similar to our findings (MAPE = 4.39%). Shcherbina et al., 2017, also reported the lowest amount of error for the Apple Watch compared to ECG when comparing six consumer devices including a Fitbit Surge and Mio Alpha.[22] Overall, other studies reported that HR measured on consumer devices were more accurate at rest, had reasonable error ratings, and generally wide 95% LOA when using them for exercise.[10, 21-24] Potential differences in studies using identical devices may be attributed to the type of measurement method used (i.e., downloaded vs. visual observed), inter-device reliability, differences in activity protocol, or different versions of firmware/software that was used. Studies should report the firmware and operating system version of devices so comparisons can be made in future studies.

Overall, the results of this study are consistent with other studies indicating that wrist-worn consumer-level optical HR monitors may provide an adequate measure of HR for rest and exercise during recreational uses but warrant caution in clinical applications for those with cardiac conditions.[10, 21, 22, 32, 33] In comparison to more recent studies,[25, 32, 33] the validity has not improved over older models; thus variability may be more inherent in the measurement technology or wear and that major improvements may not be realized until newer technologies are developed. This implies that newer, more expensive devices are not superior to older iterations of the same product. They offer the same level of HR monitoring accuracy as the newest devices but at a more affordable price. At the time of this study, there was little difference between the prices of the devices used making each of them a good choice for affordable HR monitoring during exercise, or at rest.

There were several limitations in the current study. This study was designed as a concurrent validity study and did not use the gold standard criterion, ECG, to assess HR measures. However, Polar devices have been shown to be highly correlated with ECG measures ($r = 0.99$).[28] We did not test for HR max, instead used the participants' calculated HR max. We felt this was acceptable with the younger population, however, if an older population or special population were to be tested, calculated HR max may not be warranted. Another limitation was that only one sample from each device model was used in this study. Therefore, we could not perform reliability analyses. The model year of our devices was from 2016 and do not reflect the potential improvements of current software or hardware changes that are more prevalent at the time of this publication. PPG sensor technology has not drastically changed and we surmise that our findings should be relevant for current devices using this technology. Visual inspection was the primary measurement method used in this study to mimic end-user experience. Although this method introduces error, it has been used in several other HR validity studies.[10, 21, 23, 24] It is important to note that the results from this study can only be generalized to HR measures taken with the wrist in a steady position such as when one visually inspects their device and not with the wrist moving which likely introduces other HR artifacts.[34] Despite these limitations, the strengths of this study include a sample size that was comparable to other widely cited validity studies and exercise protocols with a sample that allowed for assessment of a wider range of HR responses.

CONCLUSION

This study replicates but also extends findings from other authors on the use of wrist-based HR monitors to near-maximum HR values for PA. With newer models, we cannot expect a large improvement in measurement accuracy, but a shift may be realized as sensor technologies advance. Fitness professionals can utilize wrist-located devices and still maintain a level of accuracy appropriate for recreational use. A trade-off in accuracy and error compared to ease of use may be relevant for personal trainers and fitness coaches to incorporate an efficient measure of intensity during exercise classes.

ACKNOWLEDGEMENTS

We thank the following student research assistants for data collection and development activities: Natalie Sealover, Allie M. Hoover, Heather M. Puleo, Carson McGinn, Quinten Miller, Katherine Kirven, Anna Helm, Cameron Helmer, Katherine White, Tyler Bailey, and Samuel Greene.

REFERENCES

1. Wickham JB, Mullen NJ, Whyte DG, Cannon J. Comparison of energy expenditure and heart rate responses between three commercial group fitness classes. *Journal of Science and Medicine in Sport*. Jul 2017;20(7):667-671. *doi:10.1016/j.jsams.2016.11.012*
2. Chavarrias M, Carlos-Vivas J, Collado-Mateo D, Perez-Gomez J. Health Benefits of Indoor Cycling: A Systematic Review. *Medicina (Kaunas)*. Aug 8 2019;55(8) *doi:10.3390/medicina55080452*
3. Canario-Lemos R, Vilaca-Alves J, Moreira T, et al. Are Heart Rate and Rating of Perceived Exertion Effective to Control Indoor Cycling Intensity? *International Journal of Environmental Research and Public Health*. Jul 4 2020;17(13)*doi:10.3390/ijerph17134824*
4. Brazendale K, Decker L, Hunt ET, et al. Validity and Wearability of Consumer-based Fitness Trackers in Free-living Children. *International Journal of Exercise Science*. 2019;12(5):471-482.
5. Mertz L. Convergence revolution comes to wearables: Multiple advances are taking biosensor networks to the next level in health care. *IEEE Pulse*. Jan-Feb 2016;7(1):13-7. *doi:10.1109/MPUL.2015.2498475*
6. Mombers C, Legako K, Gilchrist A. Identifying medical wearables and sensor technologies that deliver data on clinical endpoints. *British Journal of Clinical Pharmacology*. Feb 2016;81(2):196-8. *doi:10.1111/bcp.12818*
7. Munos B, Baker PC, Bot BM, et al. Mobile health: The power of wearables, sensors, and apps to transform clinical trials. *Annals of the New York Academy of Sciences*. Jul 2016;1375(1):3-18. *doi:10.1111/nyas.13117*
8. Pellanda EC, Pellanda LC. Primordial prevention and wearable health devices: The wearables in cardiology. *Arquivos Brasileiros de Cardiologia*. Jun 2016;106(6):455-6. *doi:10.5935/abc.20160094*
9. Llamas R, Shirer M, Ubrani J. Wearables Deliver Double-Digit Growth for Both Q4 and the Full Year 2021, According to IDC. International Dara Corporation (IDC). June, 29th, 2022, Updated March 9th, 2022. *https://www.idc.com/getdoc.jsp?containerId=prUS48935722*
10. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Medicine and Science in Sports and Exercise*. Aug 2017;49(8):1697-1703. *doi:10.1249/MSS.0000000000001284*
11. IDC. *Wearables aren't dead, they're just shifting focus as the market grows 16.9% in the fourth quarter, according to IDC*. 2017. *IDC Worldwide Quarterly Wearable Device Tracker*. March 2nd, 2017. Accessed Accessed: September 4th, 2017. *https://www.idc.com/getdoc.jsp?containerId=prUS42342317*
12. Llamas R, Ubrani J, Shirer M. *Worldwide wearables market to nearly double by 2021, according to IDC*. 2017. *Worldwide Quarterly Wearable Device Tracker*. June 21, 2017. Accessed Accessed: September 28th, 2017. *https://www.idc.com/getdoc.jsp?containerId=prUS42818517*
13. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health*. 2006;97 (Suppl. 1):S10-S16. NOT IN FILE.
14. Welk GJ, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI Actigraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity (Silver Spring)*. Apr 2007;15(4):918-28. *doi:10.1038/oby.2007.624*
15. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of Medical Engineering and Technology*. Apr 2017;41(3):208-215. *doi:10.1080/03091902.2016.1253795*

16. Bernard T, Gavarry O, Bermon S, Giacomoni M, Marconnet P, Falgairette G. Relationships between oxygen consumption and heart rate in transitory and steady states of exercise and during recovery: Influence of type of exercise. *European Journal of Applied Physiology and Occupational Physiology*. 1997;75(2):170-6.
17. Laukkanen RM, Virtanen PK. Heart rate monitors: state of the art. *Journal of Sports Sciences*. Jan 1998;(Suppl. 16):S3-7. *doi:10.1080/026404198366920*
18. Gregoski MJ, Mueller M, Vertegel A, et al. Development and validation of a smartphone heart rate acquisition application for health promotion and wellness telehealth applications. *International Journal of Telemedicine and Applications*. 2012;2012:696324. *doi:10.1155/2012/696324*
19. Silva I, Lee J, Mark R. Photoplethysmograph quality estimation through multichannel filtering. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*. 2011;2011:4361-4. *doi:10.1109/IEMBS.2011.6091082*
20. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral and Physical Activity*. Dec 18 2015;12:159. *doi:10.1186/s12966-015-0314-1*
21. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *Journal of the American Medical Association Cardiology*. Jan 01 2017;2(1):104-106. *doi:10.1001/jamacardio.2016.3340*
22. 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. *Journal of Personalized Medicine*. May 24 2017;7(2):E3. *doi:10.3390/jpm7020003*
23. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. *Annuals of Internal Medicine*. Apr 18 2017;166(8):610-612. *doi:10.7326/L16-0353*
24. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *British Medical Journal Open Sport and Exercise Medicine*. 2016;2(1):e000106. *doi:10.1136/bmjsem-2015-000106*
25. Jagim AR, Koch-Gallup N, Camic CL, et al. The accuracy of fitness watches for the measurement of heart rate and energy expenditure during moderate intensity exercise. *The Journal of Sports Medicine and Physical Fitness*. Feb 2021;61(2):205-211. *doi:10.23736/S0022-4707.20.11151-4*
26. Oja P, Partanen T, Teraslinna P. The validity of three indirect methods of measuring oxygen uptake and physical fitness. *The Journal of Sports Medicine and Physical Fitness*. Jun 1970;10(2):67-71.
27. Warburton DER, Jamnik VK, Bredin SS, Gledhill N. The Physical Activity Readiness Questionnaire for everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). *Health and Fitness Journal of Canada*. 2011;4(2):3-23.
28. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. *European Journal of Applied Physiology*. Mar 2016;116(3):563-71. *doi:10.1007/s00421-015-3303-9*
29. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart rate variability recording with smart phone photoplethysmographic, Polar H7 chest strap and electrocardiogram methods. *International Journal of Sports Physiology and Performance*. Mar 14 2017:1-17. *doi:10.1123/ijspp.2016-0668*
30. Navalta JW, Ramirez GG, Maxwell C, Radzak KN, McGinnis GR. Validity and Reliability of Three Commercially Available Smart Sports Bras during Treadmill Walking and Running. *Scientific Reports - Nature*. Apr 30 2020;10(1):7397. *doi:10.1038/s41598-020-64185-z*
31. Jeong IC, Yoon H, Kang H, Yeom H. Effects of skin surface temperature on photoplethysmograph. *Journal of Healthcare Engineering*. 2014;5(4):429-38. *doi:10.1260/2040-2295.5.4.429*
32. Navalta JW, Montes J, Bodell NG, Salatto RW, Manning JW, DeBeliso M. Concurrent heart rate validity of wearable technology devices during trail running. *PLoS One*. 2020;15(8):e0238569. *doi:10.1371/journal.pone.0238569*
33. Fuller D, Colwell E, Low J, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR Mhealth Uhealth*. Sep 8 2020;8(9):e18694. *doi:10.2196/18694*
34. Zhu S, Tan K, Zhang X, Liu Z, Liu B. MICROST: A mixed approach for heart rate monitoring during intensive physical exercise using wrist-type PPG Signals. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*. 2015;2015:2347-50. *doi:10.1109/EMBC.2015.7318864*

ABOUT STUDENT AUTHORS

Korey Little is an Exercise Science Major in the Department of Health and Human Performance at the College of Charleston and is expected to graduate with his B.S. in Exercise Science in May 2023. He plans to work in medical research prior to obtaining a DPT.

PRESS SUMMARY

Heart rate is one of the most important things to monitor during exercise as it provides a better understanding of what is happening inside of the body and how hard a person is working. In group exercise classes, the way coaches and trainers control the intensity is usually a combination of using a preplanned exercise program and music. Adding the ability to measure heart rate would greatly increase the quality of exercise for everyone involved. Typical chest-located heart monitors can be intrusive and difficult to fit comfortably. Having the ability to wear a heart rate monitor on the wrist affords reasonable accuracy and provides more access to group exercise classes or personal training sessions. We tested four different wrist-based heart rate monitors (Mio fuse, Apple watch 1, Fitbit Charge, Garmin Vivosmart) on 45 college-aged participants and found that the Mio Fuse was the most accurate and closest to a medical-grade heart rate monitor followed closely by the Apple watch 1. Most importantly, there was no bias in heart rate as intensity increased to maximum values in all devices. All the watches were accurate enough to be used, but we found that the initial reading may be incorrect and a second or third look may be necessary to increase confidence in observed heart rates.