

Research article

Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking

Edward Jo ¹✉, Kiana Lewis ¹, Dean Directo ¹, Michael J. Kim ¹ and Brett A. Dolezal ^{1,2}

¹ Human Performance Research Laboratory, Department of Kinesiology and Health Promotion, California State Polytechnic University Pomona, Pomona, CA, USA; ² Exercise Physiology Research Laboratory, David Geffen School of Medicine and Department of Physiology, University of California, Los Angeles, Los Angeles, CA, USA

Abstract

The purpose of this study was to examine the validity of HR measurements by two commercial-use activity trackers in comparison to ECG. Twenty-four healthy participants underwent the same 77-minute protocol during a single visit. Each participant completed an initial rest period of 15 minutes followed by 5 minute periods of each of the following activities: 60W and 120W cycling, walking, jogging, running, resisted arm raises, resisted lunges, and isometric plank. In between each exercise task was a 5-minute rest period. Each subject wore a Basis Peak (BPK) on one wrist and a Fitbit Charge HR (FB) on the opposite wrist. Criterion measurement of HR was administered by 12-lead ECG. Time synced data from each device and ECG were concurrently and electronically acquired throughout the entire 77-minute protocol. When examining data in aggregate, there was a strong correlation between BPK and ECG for HR ($r = 0.92$, $p < 0.001$) with a mean bias of -2.5 bpm (95% LoA 19.3, -24.4). The FB demonstrated a moderately strong correlation with ECG for HR ($r = 0.83$, $p < 0.001$) with an average mean bias of -8.8 bpm (95% LoA 24.2, -41.8). During physical efforts eliciting ECG HR > 116 bpm, the BPK demonstrated an $r = 0.77$ and mean bias = -4.9 bpm (95% LoA 21.3, -31.0) while the FB demonstrated an $r = 0.58$ and mean bias = -12.7 bpm (95% LoA 28.6, -54.0). The BPK satisfied validity criteria for HR monitors, however showed a marginal decline in accuracy with increasing physical effort (ECG HR > 116 bpm). The FB failed to satisfy validity criteria and demonstrated a substantial decrease in accuracy during higher exercise intensities.

Key words: Photoplethysmography, biosensor, biotechnology, fitness, cardiovascular .

Introduction

Digital health, specifically wearable biosensor technology, has recently undergone tremendous advancements in the sport, fitness, and health industries. Initially developed to augment personal fitness and performance with basic quantitative biofeedback, the newest generation of devices, i.e. “activity trackers”, provide real-world, immediate feedback on multiple biometrics related to the consumers’ physical activity, health, and exercise quality (Lyons et al., 2014). With continuous technological advancements in wearable biofeedback devices, the potential applications have also expanded to include medical surveillance, non-invasive medical care, and mobile health-wellness monitoring.

The pursuit for practical and accurate methods to assess personal health and physical activity continues to

lay emphasis on wearable and sophisticated biosensor technologies. It has been previously suggested that integrative biometric processing through multiple sensors (e.g. body temperature, skin galvanic response, heart rate, etc.) may support activity-specific prediction algorithms to accurately compute real-life energy expenditure and overall activity level (Chen and Bassett, 2005; Doherty and Oh, 2012; Gao et al., 2014). This has stimulated the adoption of multi-sensor technology in nascent activity trackers which has shown to outperform more rudimentary devices that utilize basic accelerometer data that record steps to infer movement, energy expenditure, and subsequently overall activity level. Moreover, with the inclusion of sophisticated reflective photoplethysmography (PPG) technology, contemporary activity trackers such as the Basis Peak™ (BPK) and Fitbit Charge HR™ (FB) have the capacity to use heart rate-derived algorithms to support energy expenditure estimations (Keytel et al., 2005; Luke et al., 1997; Wallen et al., 2016). Besides predicting energy expenditure, PPG-based heart rate detection allows for unobtrusive objective monitoring of physical exertion as well as systematic exercise prescriptions. The use of PPG technology for heart rate monitoring has shown acceptable validity, however there is inherent variability, indicating that the accuracy of these trackers is dependent on the specific device used and the type and intensity of activity. Therefore, PPG-based activity trackers remain under much scrutiny with respect to accuracy especially considering the number of extrinsic factors that may interrupt proper heart rate detection (i.e. ambient light, sweat, anatomical placement, movement, skin contact force) (Allen, 2007).

With the recent development of new types of activity trackers incorporating PPG-based heart rate sensors and multi-sensor technology, there needs to be a steady focus on improving overall device performance. With that said, validation of heart rate measurement accuracy for such multi-sensor activity trackers becomes a critical step in improving the consumer experience and at times, their safety. Notwithstanding, there is a paucity of rigorous, scientifically-based validation studies on the accuracy of multi-sensor trackers incorporating reflective PPG in reference to gold-standard methods for heart rate measurements, i.e. ECG. Therefore, the objective of this study was to determine the capability of two current generation activity trackers, the BPK and FB, to accurately measure heart rate during a variety of laboratory-based exercise tasks in reference to the ECG criterion measure. Moreo-

ver, this study aimed to assess the accuracy of heart rate measurements for these activity trackers specifically during differing levels of physical efforts (e.g. high vs. low ECG heart rate range) and exercise tasks (e.g. running, cycling, resistance exercise, isometric exercise, etc.).

Methods

Experimental design

Twenty-four college-aged subjects (12 males and 12 females) were recruited to participate in this randomized, controlled, crossover design study. Through random sampling methods subjects were recruited via posted flyers as well as mass email and social media solicitations. Each subject underwent an identical 77-minute protocol during a single visit to the Human Performance Research Laboratory at California State Polytechnic University, Pomona (CPP). Initially, participants completed and signed an informed consent form and underwent assessment for basic anthropometric measures (i.e. height, bodyweight). Subsequently, subjects underwent preparatory steps for 12-lead ECG testing procedures for criterion measurements for heart rate (HR) (described below). Next, researchers placed the Basis Peak (BPK) and Fitbit Charge HR (FB) on opposite wrists. Wrist assignment (i.e. dominant vs. non-dominant) was alternated between subjects for counterbalance and to control for any confounding effects associated with wrist assignment. Care was taken to follow proper-use guidelines as suggested by the manufacturer for each device. Time synced heart rate data from each device (test devices and ECG) were concurrently and continuously acquired second-by-second throughout the entire 77-minute protocol for each participant. During data acquisition, participants completed an initial rest period (supine) of 15 minutes (data excluded) followed by 5 minute periods of each of the following activities: low cycling (60W resistance), intense cycling (120W), walk (3.0-3.5mph speed), jog (4.0-5.0mph), run (5.5-7.0mph), arm raises with self-selected added resistance using dumbbells (12 reps), lunges with self-selected added resistance using dumbbells (12 reps), and isometric plank (60 second hold). In between each exercise task, a 5-minute sitting rest period was implemented during which time HR data was continuously acquired and included in the analysis. This study was approved by the CPP Institutional Review Board (ethics board).

Participants

Twenty-four healthy participants (12 males and 12 females) volunteered in this study (age = 24.8 ± 2.1 years, weight = 71.3 ± 9.9 kg, height = 1.66 ± 0.07 m). Written informed consent was obtained after discussing the study procedures in detail, including the voluntary nature of participation and notification that the subject can withdraw at any time. Upon the subject's agreement to participate, a signed copy of the informed consent form was given to the researcher. Participants were included in the study if they were physically able to perform the exercise protocol with minimal risk for injury or health complications. Participants were excluded from the study if they reported or exhibited any significant medical conditions,

including cardiovascular or pulmonary disease that may limit ability to exercise or increase the cardiovascular risk of exercising.

Experimental procedures

Criterion measure electrocardiograph: A standard 12-lead electrocardiograph system (Cosmed C12x; Concord, CA, USA) integrated to a laboratory-grade metabolic testing system (ParvoMedics TrueOne[®] 2400; Salt Lake City, Utah, USA) was used to provide criterion measures of heart rate and energy expenditure (energy expenditure not included in current analysis). For ECG testing, the electrode placement sites were prepared by standardized procedures of cleaning, shaving, and abrading the skin to improve signal acquisition and to minimize noise artifact. Ten silver/silver-chloride self-adhesive electrodes (RA, LA, RL, LL, V1, V2, V3, V4, V5, and V6) were placed on the upper torso according to the Mason-Likar lead placement configuration for the 12-lead ECG (Mason and Likar, 1966). The ECG wires were secured to the body using elastic foam wrap and a standard ECG stress test belt to minimize electrode movement artifact. The ECG and indirect calorimetry system was integrated and thus metabolic and heart rate data acquisition was simultaneous and time-synced. Although, metabolic data were excluded in the current analysis, the integration of testing equipment was necessary to obtain exportable second-by-second HR records. Heart rate data per second was converted to beats per minute (BPM) automatically by the data acquisition software program prior to analysis.

Activity tracker procedures: The BPK and FB were attached to opposing wrists on the subject according to manufacturer instructions. Half of the subject pool wore the BPK on the dominant wrist and the FB on the non-dominant wrist. The other half of the subject pool wore the BPK on the non-dominant wrist and the FB on the dominant wrist. We implemented this randomized counterbalancing strategy to control for any potential confounding issues associated with the wrist on which the devices are placed. Data acquisition from each device along with ECG was time-synced according to a single master clock. Heart rate data were electronically downloaded from the BPK using a data extraction software program customized specifically to sync with the BPK via Bluetooth transmission upon completion of the testing protocol. For FB data acquisition, the "track exercise" function on the mobile device application (i.e. iOS app) interfaced with the FB device was used. This function allowed for time-synced GPS and HR data acquisition. Upon completion of the testing protocol, the GPS (.tcx) file linked to the "tracked" exercise was downloaded from the Fitbit server and subsequently imported as a .csv spreadsheet file. The spreadsheet displayed time-synced, second-by-second GPS and HR data. The GPS data were discarded while HR data were subsequently used for analysis. The mobile application settings for the FB were adjusted appropriately for each subject. Each device was confirmed to have full battery charge prior to testing. Heart rate data per second was converted to beats per minute (BPM) automatically by each device's respective data acquisition program prior to analysis.

Statistical analysis

Three levels of statistical analysis were implemented to assess the level of accuracy of the consumer devices in reference to ECG: *i*) Pearson Product-Moment Correlation analysis was used to determine the strength of relationship between ECG and each of the consumer devices and whether the relationship was statistically significant. A significant correlation was determined if the *p*-value was less than 0.05 while the strength of correlation was determined by the correlation coefficient (*r*-value). The strength of correlation was determined as follows: 0.9-1.0 = strong, 0.8-0.89 = moderately strong, 0.7-0.79 = moderate, 0.6-0.69 = moderately weak, <0.59 = weak; *ii*) the Bland-Altman method was used to further assess the agreement between the consumer devices and ECG for heart rate measurements and whether the differences varied in a systematic or ambiguous way over the range of measurements (Bland and Altman, 1986). The mean bias between consumer heart rate device and ECG (=FB HR – ECG HR and =BPK HR – ECG HR) and the 95% limits of agreement (LoA; LoA = mean bias/difference \pm 1.96 standard deviation of the difference) was identified. Bland-Altman plots graphed the individual difference scores against the individual averages of the heart rate values obtained by the consumer device and ECG; and *iii*) Mean absolute differential between the consumer device and the ECG was computed. This represented the average difference score regardless of direction of the difference (i.e. regardless of under- or overestimation). All three levels of analysis were implemented on aggregate HR data, HR data above the mean ECG HR, HR data below the mean ECG HR, resting or recovery HR data, and task-specific HR data. Previous validation studies (Dolezal et al., 2014; Terbizan et al., 2002) have provided validity criteria for heart rate measurement as: 1) a correlation between ECG-derived heart rate and the heart rate measured by the test device of $r=0.90$ or greater, 2) a mean bias less than 3 beats/min, and 3) a standard error less than 5 bpm. Satisfaction of at least the first and second criteria together or all criteria was used to determine validation of the consumer devices for accurate HR measurement. However, results from all three aforementioned analyses were considered and discussed to provide a more inclusive examination on the accuracy of each device. Based on the assumption that the true correlation between ECG-derived HR and device-measured HR is associated with $\rho = 0.85$, a sample size of 24 subjects was determined (Wallen et al., 2016).

Results

Table 1. Results for basis peak.

	Aggregate Data (n = 87 340)	Data above ECG HR >116bpm (n = 41 315)	Data below ECG HR <117bpm (n = 46 025)
Basis Peak Mean HR, bpm (\pm SD)	115 (28)	137 (20)	94 (15)
ECG Mean HR, bpm (\pm SD)	117 (29)	142 (19)	94 (14)
Mean Absolute Difference, bpm (\pm SD)	6 (10)	7 (12)	5 (7)
Mean Absolute Difference, % (\pm SD)	5.3 (8.3)	5.3 (9.5)	5.3 (7.0)
Correlation (<i>r</i>)	.92*	.77 *	.84 *
Mean Bias, bpm (\pm SD)	-3 (11) [95% CI -3, -3]	-5 (13) [95% CI -5, -5]	-1 (8) [95% CI -1, -1]
95% Limits of Agreement (Upper, Lower)	19, -24	21, -31	15.6, -16.4
Standard Error of the Estimate (SEE)	11	12	8

SD= standard deviation, ECG= electrocardiograph, HR= heart rate, CI= confidence interval, n= per second ECG and Basis Peak data pairs. * $p < 0.0001$

Aggregate heart rate data

When examining time-synced ECG and activity tracker HR data in aggregate (inclusion of all resting and exercise conditions) (n = 87 340) (Figure 1), there was a strong positive correlation between ECG and BPK ($r = 0.92$, $p < 0.0001$) (Table 1) and a moderately strong positive correlation between ECG and FB ($r = 0.83$, $p < 0.0001$) (Table 2). In reference to ECG, BPK exhibited a mean bias of -3 ± 11 bpm (95% LoA 19, -24) while FB demonstrated a mean bias of -9 ± 17 bpm (95% LoA 24, -42).

HR data above mean ECG HR (>116 bpm)

Time-synced HR data pairs above the mean ECG HR (>116 bpm; n = 41 315 pairs) were analyzed (Figure 2). During conditions in which the ECG HR (true HR) exceeded 116 bpm, there was a moderate positive correlation between ECG and BPK ($r = 0.77$, $p < 0.05$) (Table 1) and a weak positive correlation between ECG and FB ($r = 0.58$, $p < 0.05$) (Table 2). In reference to ECG, BPK showed a mean bias of -5 ± 13 bpm (95% LoA 21, -31) and FB exhibited a mean bias of -13 ± 21 bpm (95% LoA 29, -54).

HR data below mean ECG HR (<117 bpm)

Time-synced HR data pairs below the mean ECG HR (<117 bpm; n = 46 025 pairs) were analyzed (Figure 3). During conditions in which the ECG HR (true HR) was below 117 bpm, there was a moderately strong positive correlation between ECG and BPK ($r = 0.84$, $p < 0.05$) (Table 1) and a moderate positive correlation between ECG and FB ($r = 0.73$, $p < 0.05$) (Table 2). Mean bias scores were -1 ± 8 bpm (95% LoA 16, -16) for BPK and -5 ± 11 bpm (95% LoA 16, -26) for FB.

Task-specific Heart Rate comparison

Correlation coefficient, mean bias, and SEE for the BPK and FB for each exercise task and rest/recovery periods are indicated in Table 3. Across all exercise tasks and resting/recovery conditions, the BPK demonstrated significant ($p < 0.05$) moderate to strong correlations with ECG while the FB exhibited significant ($p < 0.05$) weak to moderately strong correlations. Both devices performed best during resting/recovery conditions as indicated by relatively strong correlations to ECG and small mean bias scores (BPK: $r = 0.96$, mean bias = -1 bpm; FB: $r = 0.83$, mean bias = -4 bpm). During exercise conditions, the BPK provided accurate HR readings during low intensity cycling, walking, jogging, and running. The accuracy for the BPK was slightly compromised during higher intensity

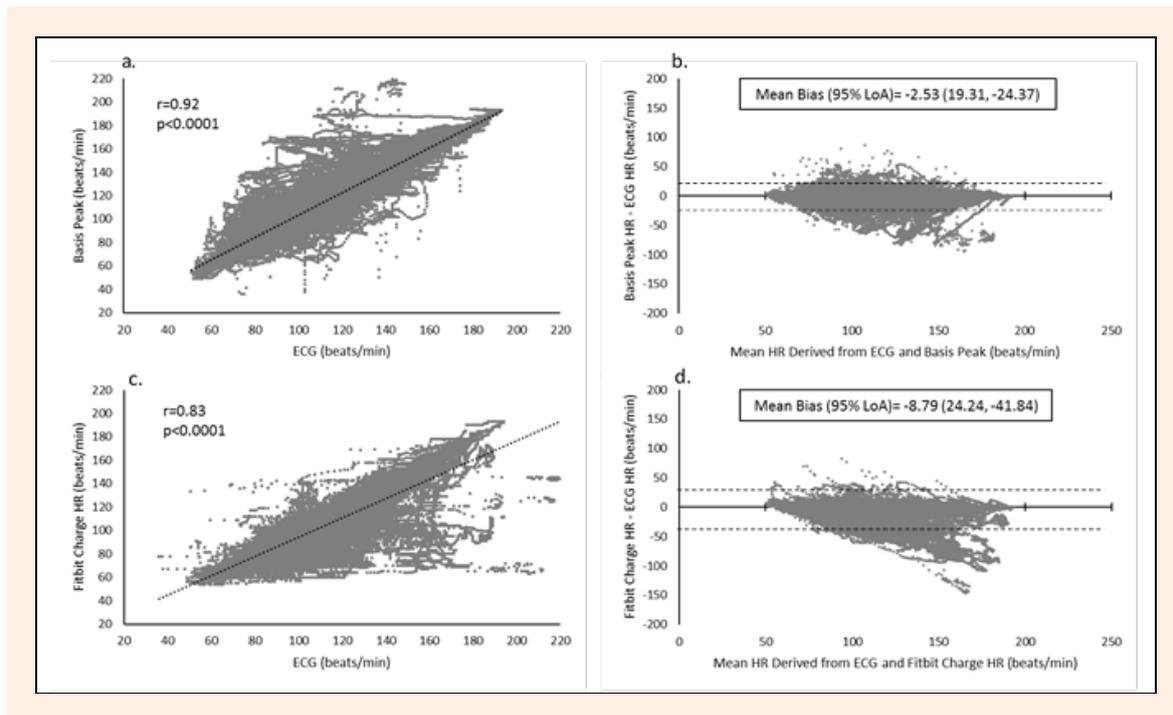


Figure 1. Results for aggregated data set for Basis Peak (1a and 1b) and Fitbit Charge HR (1c and 1d). Correlation between each test device and ECG and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement (LoA).

Table 2. Results for Fitbit charge HR.

	Aggregate Data (n = 87 340)	Data above ECG HR >116bpm (n = 41 315)	Data below ECG HR <117bpm (n = 46 025)
FB Mean HR, bpm (\pm SD)	108 (29)	130 (25)	89 (15)
ECG Mean HR, bpm (\pm SD)	117 (29)	142 (19)	94 (14)
Mean Absolute Difference, bpm (\pm SD)	11 (16)	14 (20)	8 (9)
Mean Absolute Difference, % (\pm SD)	9.8 (14.0)	11.3 (17.0)	8.5 (10.4)
Correlation (r)	.83*	.58 *	.73 *
Mean Bias, bpm (\pm SD) [95% CI]	-9 (17) [-9, -9]	-13 (21) [-13, -13]	-5.3 (11) [-5, -5]
95% Limits of Agreement (Upper, Lower)	24, -42	29, -54	16, -26
Standard Error of the Estimate (SEE)	16	20	10.0

FB = Fitbit Charge HR, SD= standard deviation, ECG= electrocardiograph, HR= heart rate, CI= confidence interval, n= per second ECG and FB data pairs. *p < 0.0001

cycling, resisted arm raises, and resisted lunges while the greatest detriment to performance occurred during the isometric plank ($r = 0.78$, mean bias = -5 bpm). During walking, jogging, and running tasks, the FB performed with strong correlation ($r = 0.94$ to 0.95) and excellent agreement to ECG (about -4 to -3 bpm bias). However, performance of the FB was exceptionally poor during cycling, resisted arm raises, lunges and isometric plank. In particular, the FB and ECG had the weakest correlation and agreement during the resisted lunges ($r = 0.28$, mean bias = -21 bpm) and isometric plank ($r = 0.26$, mean bias = -12 bpm).

Discussion

The primary objective of this investigation was to assess the validity of heart rate measurements from two commercially available, wrist-worn activity trackers incorporating proprietary reflective photoplethysmographic biosensors, i.e. the Basis Peak™ and Fitbit Charge HR™. The accuracy of the Bpk and FB was evaluated in refer-

ence to the criterion measure ECG. When examining the data in aggregate ($n = 87,340$), the Bpk met the proposed validity criteria for heart rate detecting and monitoring devices ($r \geq 0.90$ and mean bias < 3 bpm) (Table 1, Figure 1). We observed a strong correlation ($r = 0.92$) between the Bpk and ECG with an acceptable mean bias of -3 bpm and an absolute differential from criterion measurements of approximately 5%. On the basis of the Bland-Altman analysis and the 95% limits of agreement (+19 to -24 bpm), the bias of the Bpk may be reasonably described as systematic (Figure 1) and therefore, may be used interchangeably with ECG for accurate HR measurements. In comparison, the Bpk demonstrated a similar level of performance to the Apple Watch which has been previously validated by data indicating a strong correlation ($r = 0.95$) with ECG and a small mean bias of -1 bpm (Wallen et al., 2016). Interestingly, the accuracy of the Bpk is slightly compromised with increasing physical effort (> 116 bpm per ECG). This was observed in the analysis of data pairs associated with an ECG HR above 116 bpm (mean ECG HR) ($n = 41 315$). Specifically, the correlation between

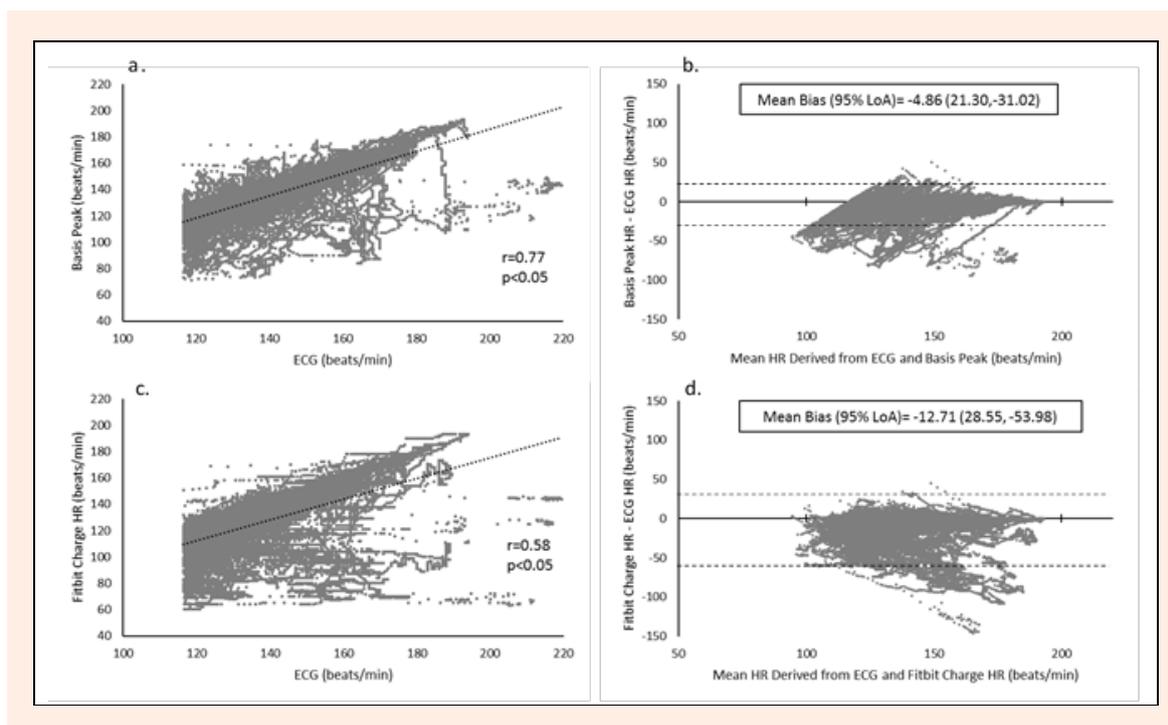


Figure 2. Results for data set above mean ECG heart rate (>116 bpm) for Basis Peak (2a and 2b) and Fitbit Charge HR (2c and 2d). Correlation between each test device and ECG and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement (LoA).

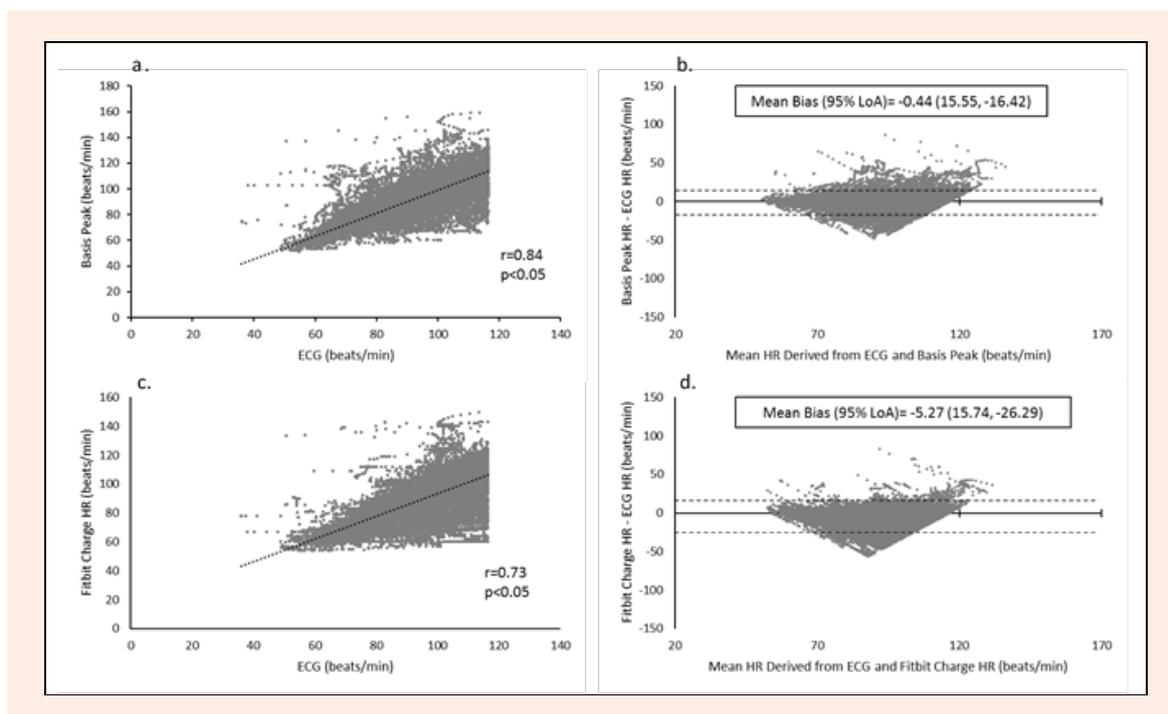


Figure 3. Results for data set below mean ECG heart rate (<117 bpm) for Basis Peak (3a and 3b) and Fitbit Charge HR (3c and 3d). Correlation between each test device and ECG and Bland-Altman Plots indicating mean bias scores and 95% limits of agreement (LoA).

BPK and ECG weakened per coefficient of 0.77, and the mean bias was slightly exacerbated to -5 bpm. Nonetheless, our data set above the ECG mean showed strong agreement between the heart rates derived from the BPK and the criterion measure of the ECG. When considering resting or recovery conditions and each individual mode of exercise, the BPK performed relatively accurately with

very strong agreement to ECG (especially during rest/recovery) (Table 3). Only during intense cycling and the isometric plank did the BPK demonstrate a minor decrease in performance. These outcomes are consistent with previous results for other PPG-based devices (Mio Alpha and Scosche myRhythm) that revealed a task-specific variation in device performance (Parak and

Table 3. Activity-specific results for Basis Peak and Fitbit Charge HR.

		Aggregate	Rest	Low Cycle	Intense Cycle	Walk	Jog	Run	Arm Raises	Lunges	Plank
Basis Peak	Correlation (r)	.92*	.96 *	.90*	.84*	.94*	.94*	.94*	.88*	.88*	.78*
	Mean Bias	-2.53	-20	-52	-7.42	-4.4	-1.79	-2.20	-4.95	-4.63	-4.69
	SE	11.07	6.04	8.79	16.25	8.16	9.63	10.38	10.13	9.40	14.46
Fitbit Charge HR	Correlation (r)	.83 *	.83*	.55*	.66*	.94*	.94*	.95*	.63*	.28*	.26*
	Mean Bias	-8.79	-3.73	-12.67	-14.20	-4.85	-3.36	-3.98	-20.55	-20.99	-12.10
	SE	15.79	10.66	14.90	23.40	8.59	7.71	10.33	12.65	20.11	15.48

Standard error (SE). Aggregate (n = 87 340); Rest (n = 6 776); Low Cycle (n = 14 155); Intense Cycle (n = 14 297); Walk (n = 16 039); Jog (n = 12 497); Run (n = 12 996); Arm Raises (n = 4 146); Lunges (n = 3 811); Plank (n = 1 623). * p < 0.05

Korhonen, 2014). For instance, the Mio Alpha showed the greatest error (-4.8% error) during cycling while the Scosche myRhythm was during the walking exercise (-3.1% error).

The FB presented with weaker correlation ($r = 0.83$) and less agreement (mean bias = -9 bpm, 95% LoA 24, -42) to ECG than the Bpk when examining the aggregated data set (n = 87 340). Despite a moderately strong correlation, the resulting r-value and mean bias score failed to meet the proposed validity criteria for accurate heart rate measurements. Furthermore, the Bland-Altman analysis for the aggregated data set reflects a large tendency for HR underestimation by the FB especially at higher ends of the mean HR spectrum. Thus, the FB may not be considered interchangeable with ECG for accurate measurement of HR. These findings corroborate previous FB results by Wallen et al. (2016) which demonstrated a -9 bpm bias, 95% LoA (7, -26 bpm), and a correlation coefficient of 0.81. Moreover, we observed severely diminished performance during physical activities eliciting higher ECG heart rates (i.e. >116 bpm). The very weak correlation ($r = 0.58$) together with the large mean bias (= -13 bpm) and high standard error (= 20.1) strongly suggest the FB to be an inaccurate means of monitoring HR with increasing physical exertion. The FB appeared to perform with improved accuracy during conditions corresponding to lower ECG heart rates based on a small average bias of -5.3 bpm. However, even during lighter physical exertion, per mean ECG HR, the substandard correlation strength ($r = 0.73$) may not suffice to substantiate accuracy. Specifically during resting or recovery situations, the FB demonstrated a moderately strong correlation with ECG ($r = 0.83$) and underestimated HR by only 3.7 bpm on average (Table 3). The FB performed with moderate accuracy and agreement to ECG during the walk, jog, and run as reflected by strong correlations and relatively low mean bias scores (~ -4 bpm). During the isometric plank, resisted lunges, and cycling, however, the FB demonstrated very weak to moderately weak correlations and large mean bias scores. Thus, in addition to activities of higher physical exertion, specific exercise tasks also appear to dramatically attenuate the performance of the FB for accurate heart rate tracking.

It is evident based on our data as well as others' (Parak and Korhonen, 2014) that PPG-based HR monitors experience reduced accuracy during elevated physical exertion and specific exercise tasks, two plausibly interrelated factors. Moreover, the degree of detriment and the exercise type related to poor performance varies among

the different PPG devices. To speculate on the potential factors impeding accurate measurements during exercise, especially for the FB, we examine the commonalities of the exercise tasks eliciting the worst device performance, i.e. cycling, resisted arm raises, resisted lunges, and isometric plank. Each of these exercises involve sustained or repetitive contractions of forearm skeletal muscles which may influence the efficacy by which the optical sensors acquire sufficient photoplethymographic signals for accurate HR computation. Previous evidence suggests that the contact or compression force between the sensor and the measurement site (i.e. skin surface of the wrist) significantly affects the waveform and thereby the quality of the photoplethymographic signal (Allen, 2007; Rafolt and Gallasch, 2004; Teng and Zhang, 2004). Moreover, increased compression of the wrist against the PPG sensor may exacerbate contact-related noise artifact, ultimately disrupting signal quality. This in turn would impede heart rate detection and preclude an accurate measurement (Teng and Zhang, 2004). Accordingly, manufacturer instructions for wrist-worn PPG devices often recommend the user to refrain from overtightening the strap as to avoid large sensor to skin contact force or compression. During specific exercises involving sustained or repeated forearm muscle contractions as noted above, the contact force between the device and skin is likely increased. Elevated contact force between the FB and the skin may be a plausible explanation for the lack of performance evident during these specific exercises. However, this contention is merely speculative as the effects of contact force were not the scope of the present investigation. Nonetheless, each device was worn by the subjects of the present investigation according to manufacturer recommendations. Thus, the activity trackers under investigation, particularly the FB, may require further scrutiny as it relates to 1) whether or not contact force between the skin and sensor varies across different exercises, and 2) if so, whether or not changing contact force with varying exercises alters the PPG signal quality and device accuracy.

Moreover, skin color has also been previously suggested as a factor affecting characteristics of PPG signals and thus algorithm performance (Allen, 2007; Butler et al., 2016). Although, it remains uncertain as to the level of technical control these devices incorporate to address skin color-related artifact, prior evidence suggests that PPG-based devices may detect pulsation across all skin types and that a greater signal resolution is obtained using a green light wavelength at rest and during exercise (Fallow et al., 2013). Regardless, the current data may be

limited as skin color was not accounted for within our methodology. This, however, may be balanced by the ample size of our data pool resulting from comparatively high frequency sampling rate. Future validation work should incorporate separate analyses specific to subject skin color as determined systematically by, for example, the Fitzpatrick Scale (Fitzpatrick, 1988).

Overall, heart rate monitoring on the dorsal wrist using reflective PPG sensors has apparent and evident limitations (Rafolt and Gallasch, 2004; Teng and Zhang, 2004). Traditional methods of personal heart rate tracking via electrocardiac detection using chest strap sensors with external monitors (e.g. Polar technology) have consistently demonstrated accuracy and high agreeability to ECG (Terbizan et al., 2002). Nonetheless, industry continues to lay focus on streamlining personal biometric and activity monitoring into a single wrist-worn device to enhance practicality and versatility. Thus, efforts to improve PPG-based HR tracking in multi-sensor, multi-function activity trackers require continual focus on improving the overall control of extrinsic factors that have shown to interrupt PPG-signals and accurate HR computation.

Conclusion

In summary, Basis Peak™ and Fitbit Charge HR™ perform with better accuracy during rest and lighter physical exertion versus moderate to higher intensities of activity. Also, each device exhibits varying performance under specific exercise tasks. Based on the proposed validity criteria, the current findings support validation of the BPK for accurate heart rate tracking, although further technological modifications may be necessary to resolve the slight decrease in performance with elevated physical effort. As for the FB, the data fail to support validation for accurate HR measurements. The FB may be permissible as an accurate heart rate monitor only when used during rest, recovery, or low physical exertion. Contrastingly, the FB failed to provide accurate HR readings with increasing physical effort and during specific exercise tasks that involve sustained or repeated forearm contractions.

Acknowledgements

The present study was conducted with technical and equipment support from Basis, an Intel Company. EJ was the principal investigator, KL, DD, and BD were co-investigators, and MJK assisted with data set organization. The investigators would like to acknowledge all participants for their time and efforts and Francis Figuerres for his assistance in the study. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation. All investigators of the present study are independent university researchers and conducted the current protocol without any bias towards the presented results.

References

- Allen, J. (2007) Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement* **28**, R1-39.
- Bland, J.M. and Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307-310.
- Butler, M.J., Crowe, J.A., Hayes-Gill, B.R. and Rodmell, P.I. (2016) Motion limitations of non-contact photoplethysmography due to the optical and topological properties of skin. *Physiological Measurement* **37**, N27-37.
- Chen, K.Y. and Bassett, D.R. (2005) The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise* **37**, S490-500.
- Doherty, S.T. and Oh, P. (2012) A multi-sensor monitoring system of human physiology and daily activities. *Telemedicine Journal and E Health* **18**, 185-192.
- Dolezal, B.A., Boland, D.M., Carney, J., Abrazado, M., Smith, D.L. and Cooper, C.B. (2014) Validation of heart rate derived from a physiological status monitor-embedded compression shirt against criterion ECG. *Journal of Occupational and Environmental Hygiene* **11**, 833-839.
- Fallow, B.A., Tarumi, T. and Tanaka, H. (2013) Influence of skin type and wavelength on light wave reflectance. *Journal of Clinical Monitoring and Computing* **27**, 313-317.
- Fitzpatrick, T.B. (1988) The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatological Research* **124**, 869-871.
- Gao, L., Bourke, A.K. and Nelson, J. (2014) Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering and Physics* **36**, 779-785, Jun.
- Keytel, L.R., Goedecke, J.H., Noakes, T.D., Hiiloskorpi, H., Laukkanen, R., van der Merwe, L. and Lambert, E.V. (2005) Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *Journal of Sports Science* **23**, 289-297.
- Luke, A., Maki, K.C., Barkey, N., Cooper, R. and McGee, D. (1997) Simultaneous monitoring of heart rate and motion to assess energy expenditure. *Medicine and Science in Sports and Exercise* **29**, 144-148.
- Lyons, E.J., Lewis, Z.H., Mayrsohn, B.G. and Rowland, J.L. (2014) Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *Journal of Medical Internet Research* **16**, e192.
- Mason, R.E. and Likar, I. (1966) A new system of multiple-lead exercise electrocardiography. *American Heart Journal* **71**, 196-205.
- Parak, J. and Korhonen, I. (2014) Evaluation of wearable consumer heart rate monitors based on photoplethysmography. *Conf Proc IEEE Engineering in Medicine and Biology Society* **2014**, 3670-3673.
- Rafolt, D. and Gallasch, E. (2004) Influence of contact forces on wrist photoplethysmography--prestudy for a wearable patient monitor. *Biomedical Technology (Berl)* **49**, 22-26.
- Teng, X.F. and Zhang, Y.T. (2004) The effect of contacting force on photoplethysmographic signals. *Physiological Measurement* **25**, 1323-1335.
- Terbizan, D., Dolezal, B. and Albano, C. (2002) Validity of Seven Commercially Available Heart Rate Monitors. *Measurement in Physical Education and Exercise Science* **6**, 243-247.
- Wallen, M.P., Gomersall, S.R., Keating, S.E., Wisloff, U. and Coombes, J.S. (2016) Accuracy of Heart Rate Watches: Implications for Weight Management. *PLoS One* **11**, e0154420.

Key points

- Modern day wearable multi-sensor activity trackers incorporate reflective photoplethymography (PPG) for heart rate detection and monitoring at the dorsal wrist.
- This study examined the validity of two PPG-based activity trackers, the Basis Peak and Fitbit Charge HR.
- The Basis Peak performed with accuracy compared with ECG and results substantiate validation of heart rate measurements. There was a slight decrease in performance during higher levels of physical exertion.
- The Fitbit Charge HR performed with poor accuracy compared with ECG especially during higher physical exertion and specific exercise tasks. The Fitbit Charge HR was not validated for heart rate monitoring, although better accuracy was observed during resting or recovery conditions.

AUTHOR BIOGRAPHY

Edward JO

Employment

Assistant Professor at California State Polytechnic University, Pomona and the Director of the CPP Human Performance Research Laboratory.

Degree

PhD

Research interests

Applied physiology, nutrition, and metabolism in human health and performance.

E-mail: ejo@cpp.edu

Kiana LEWIS

Employment

Human Performance Research Laboratory, Department of Kinesiology and Health Promotion, California State Polytechnic University Pomona, Pomona, CA, USA

Degree

MSc

Research interests

Applied physiology in human health and performance.

E-mail: klewis425@hotmail.com

Dean DIRECTO

Employment

Human Performance Research Laboratory, Department of Kinesiology and Health Promotion, California State Polytechnic University Pomona, Pomona, CA, USA

Degree

MSc

Research interests

Physiological adaptations to exercise training and nutritional interventions.

E-mail: dean.directo@gmail.com

Michael June Young KIM

Employment

Human Performance Research Laboratory, Department of Kinesiology and Health Promotion, California State Polytechnic University Pomona, Pomona, CA, USA

Degree

BSc

Research interests

Data collection, storage, and analysis and dissemination of data.

E-mail: michaeljunekim@gmail.com

Brett A. DOLEZAL

Employment

Associate Director of the Exercise Physiology Research Laboratory at University of California, Los Angeles in the David Geffen School of Medicine.

Degree

PhD

Research interests

Exercise physiology

E-mail: drbducla@gmail.com

✉ Edward Jo

Human Performance Research Laboratory, Department of Kinesiology and Health Promotion, California State Polytechnic University Pomona, Pomona, CA, USA