ARTICLE

Inter-Instrument Reliability and Agreement of Fitbit Charge Measurements of Heart Rate and Activity at Rest, during the Modified Canadian Aerobic Fitness Test, and in Recovery

Goris Nazari, MSc, PhD (C), PT;[*] Joy C. MacDermid, PhD, PT;[†]*
Kathryn E. Sinden, PhD, R Kin;[‡] Julie Richardson, PhD, PT;[§] Ada Tang, PhD, PT[§]

ABSTRACT

Purpose: We determined the inter-instrument reliability and agreement parameters of the Fitbit Charge Heart Rate (Charge HR) device during three phases: rest, modified Canadian Aerobic Fitness Test (mCAFT), and recovery. **Method:** We recruited 60 participants for this cross-sectional measurement study using convenience and snowball sampling approaches. The performance of the Charge HR was assessed throughout the rest, mCAFT, and recovery phases. To establish inter-instrument reliability, the Charge HR variables – heart rate, steps taken, and energy expenditures – were compared with those for two other devices: the Zephyr BioHarness (ZB) for heart rate and the Fitbit One for steps taken and energy expenditure. Measurements were recorded every 30 seconds. **Results:** At rest, the inter-instrument intra-class correlation coefficient (ICC) (standard error of measurement [SEM]) for the Charge HR versus the ZB was $\geq$ 0.97 (range, min–max, 1.02–1.32). During the mCAFT and in recovery, the ICCs (SEMs) for the Charge HR and the ZB were $\geq$ 0.89 (range, min–max, 1.30–3.98) and $\geq$ 0.68 (range, min–max, 3.58–8.35), respectively. During the mCAFT only, the number of steps taken and the energy expenditure recorded by the Charge HR and the Fitbit One displayed ICCs (SEMs) of 0.97 (83.00) and 0.77 (14.70), respectively. The average agreement differences in heart rate in this pair-wise device comparison indicated mean differences of −0.20, 4.00, and 1.00 beats per minute at rest, during the mCAFT, and in recovery, respectively. **Conclusions:** The Charge HR heart rate variable demonstrated excellent inter-instrument reliability compared with the ZB and provided good levels of agreement. The steps taken and energy expenditure variables displayed excellent reliability measures between Charge HR and Fitbit One. Our findings may be used to capture field-based wireless measures of heart rate in various phases and provide information about possibly using the Charge HR and ZB devices interchangeably.

Key Words: Fitbit Charge; reproducibility of results; cardiorespiratory fitness; heart rate determination; fitness trackers.

RÉSUMÉ

Objectif : déterminer la corrélation interinstrument et les paramètres de concordance du bracelet Fitbit Charge pour la fréquence cardiaque (Charge FC) pendant trois phases : repos, physitest aérobie canadien modifié (PACm, un test de capacité aérobique) et récupération. **Méthodologie :** les chercheurs ont recruté 60 participants pour cette étude transversale de mesures faisant appel aux approches d'échantillonnage de commodité et en boule de neige. Ils ont évalué le rendement du bracelet Charge FC pendant les phases de repos, de PACm et de récupération. Pour établir la corrélation interinstrument, ils ont comparé les variables du bracelet Charge FC (fréquence cardiaque, pas effectués et dépense énergétique) à celles de deux autres dispositifs : le Zephyr BioHarness (ZB) pour la fréquence cardiaque et le Fitbit One pour les pas effectués et la dépense énergétique. Ils ont enregistré les mesures toutes les 30 secondes. **Résultats :** au repos, le coefficient de corrélation intraclasse (CCI) interinstrument (erreur-type de mesure [ETM]) du Charge FC par rapport au ZB était de $\geq$ 0,97 (plage minimum-maximum de 1,02 à 1,32). Pendant le PACm et la récupération, le CCI (ETM) du Charge FC par rapport au ZB était $\geq$ 0,89 (plage minimum-maximum de 1,30 à 3,98) et $\geq$ 0,68 (plage minimum-maximum de 3,58 à 8,35), respectivement. Pendant le PACm seulement, le nombre de pas effectués et les dépenses énergétiques enregistrés par le Charge FC et le Fitbit One indiquaient un CCI (ETM) de 0,97 (83,00) et 0,77 (14,70), respectivement. Selon les différences moyennes de concordance de la fréquence cardiaque de cette comparaison par paire, les différences moyennes étaient de −0,20, 4,00 et 1,00 battements à la minute (battements/min) au repos, pendant le PACm et la récupération, respectivement. **Conclusion :** la variable de la fréquence cardiaque du bracelet Charge FC présentait une excellente corrélation interinstrument par rapport au dispositif ZB et un bon niveau de concordance. Les variables des pas effectués et de la dépense énergétique présentaient d'excellentes mesures de concordance entre le bracelet Charge FC et le dispositif Fitbit One. Ces résultats peuvent servir à saisir les mesures sans fil de la fréquence cardiaque à diverses phases sur le terrain et fournir de l'information sur l'interchangeabilité possible du bracelet Charge FC et du dispositif ZB.

From the: *School of Physical Therapy, Western University; †Roth McFarlane Hand and Upper Limb Centre, St. Joseph's Hospital, London; ‡School of Kinesiology, Lakehead University, Thunder Bay; §School of Rehabilitation Sciences, McMaster University, Hamilton, Ont.

Correspondence to: Goris Nazari, School of Rehabilitation Sciences, McMaster University, IAHS 403, 1400 Main St. West, Hamilton, ON L8S 1C7; gnazari@uwo.ca.

Advances in technology have promoted the development of wearable physiological monitoring (WPM) devices, which are small, non-invasive, and easy to use. WPM devices can be used to monitor workouts, help set goals, and monitor individuals' progress toward those goals. Although professionals often use subjective assessments of exercise intensity, such as perception of effort and the talk test,[1,2] WPM devices make it easier to assess heart rate and verify exercise intensity because they are objective measures. In addition, for athletes, continuous heart rate monitoring provides valuable information because during the early stages of overtraining, the maximal and sub-maximal heart rates may be decreased, whereas resting and sleeping heart rates may be increased.[3] Moreover, WPM devices can be used to motivate people to increase their participation in physical activity.

These devices are mounted at the waist or worn on the wrist to provide continuous recording of the number of steps taken and total energy expenditure. The total number of steps taken in 1 day is then used to categorise individuals into physical activity levels: sedentary (<5,000 steps), low active (5,000–7,499 steps), somewhat active (7,500–9,999 steps), active (10,000–12,500 steps), and highly active (>12,500 steps).[4] Because WPM devices are becoming very popular, individuals need to know whether their monitor of choice is providing accurate information. Therefore, research into their measurement properties is warranted.

TYPES OF WEARABLE PHYSIOLOGICAL MONITORING DEVICES

Fitbit Charge Heart Rate

The Charge HR (approximate cost $160) is a wristband that provides a continuous automatic wrist-based heart rate, which is measured in beats per minute using an optical heart rate monitor. The device also provides the number of steps taken and energy expenditure (in calories) using a three-axis accelerometry sensor.[1] It synchronizes with personal computers; is powered by a lithium polymer battery, which lasts for 5 days on a 2-hour charge; and stores data for up to 30 days.[1]

Zephyr BioHarness

The ZB (approximate cost $650) consists of an adjustable chest belt featuring conductive fabric sensors and an electronic BioModule that snaps on to the belt. The device fits comfortably on the chest at the lower sternum for both men and women and weighs 85 grams.[5] It monitors and records physiological measures such as heart rate, respiratory rate, and estimated core temperature; posture; and activity level[5] by measuring the electrical cardiac impulses travelling through the sensors. The cardiac impulses are then sent to and processed by the BioModule as beats per minute.[5] On a 3-hour charge, the BioModule can record 26 hours of data and log up to 20 days' worth of data.

Fitbit One

The Fitbit One (approximate cost $100) is a small (48.0 × 19.3 × 9.6 mm), lightweight (8 g) advanced triaxial accelerometry-based device. It can be worn on the hip or in the front pocket of pants or shorts, and it tracks physical activity and measures sleep quality.[6] Its physical activity recording features include number of steps taken, energy expenditure (calories), number of floors climbed, and distance travelled (metres). It is powered by a lithium-ion polymer battery and stores data for up to 23 days; the captured data can be uploaded to a personal computer.[6]

Measurement Properties

Before using a WPM device, it is important to determine both its reliability and its validity.[7] *Reliability* pertains to a device's consistency in producing similar measurements, whereas *validity* refers to the degree to which a device measures what it is intended to measure.[8,9] However, the validity between two device measures does not constitute agreement.[10] Conceptual differences exist between validity and agreement parameters because a lack of agreement, in spite of high correlations, can indicate the presence of systemic error.[10] The measurement properties of two newly developed WPM devices, the ZB and the Fitbit One, tested against gold standard measures, have been established in the literature.[11–14]

Our previous systematic review of the measurement properties of the ZB yielded good- to excellent-quality evidence, suggesting that it could provide reliable and valid measurements of heart rate in multiple contexts. Moreover, it demonstrated good agreement with gold standard comparators (an electrocardiogram [ECG]), a finding that supports criterion validity.[11]

Table 1 compares the validity and agreement parameters of the Fitbit One in number of steps taken and energy expenditure with gold standard criterion measures: direct observational count and metabolic system–indirect calorimeter, respectively. Overall, the Fitbit One's steps taken and energy expenditure variables demonstrated very strong correlations and small mean differences compared with their gold standard measures.[12–14] Therefore, the fact that the ZB and Fitbit One compare favourably with the gold standard measures[11–14] may de-emphasize the need to establish the measurement properties of newer WPM devices against gold standard measures.

In our previous research, we reported the excellent intra-session and inter-session reliability measurements of the Charge HR heart rate and activity variables in three phases: rest, sub-maximal exercise, and recovery.[15] The current literature has a paucity of reports on the inter-instrument reliability and agreement properties of the Charge HR. Therefore, we aimed to establish the inter-instrument reliability and agreement parameters of the

Table 1 Summary of Validity and Agreement Properties of the Fitbit One

Study	Criterion measure	Testing protocol	Agreement bias	r/r_s	Sample
Lee et al.[13]	Portable metabolic system	Sedentary; walking, 4 km (2.50 miles) per hour; running, 8.8 km (5.50 miles) per hour; moderately vigorous activities	−26.00 calories	0.81	60 healthy participants: 30 men aged 28.6 (SD 6.40) y; 30 women aged 24.2 (SD 4.70) y
Takacs et al.[14]	Observer step count	Treadmill walking	−1.00 step	≥0.97	30 participants: 15 men, 15 women, aged 29.60 (SD 5.70) y
Diaz et al.[12]	Direct observation; indirect calorimeter	Walking or jogging 3.1– 8.4 km (1.90–5.20 miles) per hour	−3.10 to −0.30; −0.80–0.40	≥0.97/ ≥0.86	23 healthy adult participants: 10 men, 13 women, aged 20–54 y

r = Pearson's correlation coefficient; r_s = Spearman's rank correlation.

Charge HR compared with the ZB and Fitbit One. Specifically, we sought to determine (1) the inter-instrument reliability and agreement between the Charge HR and ZB heart rate variables at rest, during the modified Canadian Aerobic Fitness Test (mCAFT), and in the recovery period and (2) the inter-instrument reliability and agreement between the Charge HR and Fitbit One in number of steps taken and energy expenditure variables during the mCAFT.

METHODS

Sample size calculation

We based the sample size calculation on our previous ZB and Fitbit Charge reliability study, with a null hypothesis test–retest reliability value for the ICC of 0.80 and the expectation of obtaining a test–retest reliability ICC of 0.90.[15] On the basis of the calculation, we needed a sample size of 60 participants (see the Appendix).

Hypothesis

For the Charge HR heart rate measure, we formulated two hypothesis tests:

1. *Null hypothesis:* the inter-instrument reliability of the Charge HR versus the ZB heart rate measurements in all three phases will be less than or equal to 0.60 (R 0.60).
2. *Alternative hypothesis:* the inter-instrument reliability of the Charge HR versus the ZB heart rate measurements in all three phases will be greater than 0.80 ($R > 0.80$).

For the Charge HR steps taken and energy expenditure measurements, we formulated two other hypothesis tests:

1. *Null hypothesis:* the inter-instrument reliability of the Charge HR versus the Fitbit One steps taken and energy expenditure measurements will be less than or equal to 0.80 (R 0.80).
2. *Alternative hypothesis:* the inter-instrument reliability of the Charge HR versus the Fitbit One steps taken

and energy expenditure measurements will be greater than 0.80 ($R > 0.80$).

Sample

We used stratified convenience and snowball sampling approaches to recruit 60 healthy volunteers from among university students and staff (30 men, 30 women). We received ethics approval from the Hamilton Integrated Research Ethics Board (No. 0825), and all participants provided written, signed consent.

Inclusion criteria

We administered a Physical Activity Readiness Questionnaire (PAR–Q) to all potential participants aged 20–69 years, and those who answered "no" to all seven questions were eligible to take part in the study.[16] The ability to read, write, and communicate in English was also a requirement.

Exclusion criteria

Participants were excluded if they were unable to complete the PAR–Q because of a lack of English proficiency.[16]

Procedures

Standardized procedures were followed, whereby the participants were emailed the mCAFT preliminary participant instructions 48 hours before their visit. At their first visit, the principal investigator explained the study to the participants, administered the PAR–Q and obtained written, signed consent. Next, the participants' resting heart rate, resting blood pressure, height (in metres), and body weight (in kilograms) were recorded.[16] Then, all three devices – ZB, Charge HR, and Fitbit One – were fitted. Participants were required to remain seated for a 10-minute rest period. After this baseline rest period, the mCAFT activity phase began. This phase was followed by a 10-minute recovery period, during which the participants were again required to remain seated.

To determine the inter-instrument reliability of the Charge HR and establish its level of agreement with the ZB and the Fitbit One, participants' heart rate (in beats per minute) was recorded using the Charge HR and ZB

during the three phases. The heart rate measurements were recorded every 30 seconds during the rest, mCAFT, and recovery phases. This procedure generated 20 values during rest and recovery (10 minutes each) and 24 values during the mCAFT (12 minutes each). The number of steps taken and energy expenditure (calories) variables were recorded only during the mCAFT phase, not during the rest or recovery phases. The data sources for the heart rate measures for the Charge HR and ZB were synchronized and examined during the same 30-second time periods in all phases.

Primary outcome measure: heart rate

The mCAFT estimates the level of maximal oxygen uptake (VO2 max) in individuals completing a functional stair-climbing task.[17] It is a multistage sub-maximal step test consisting of eight stages,[16] each of which lasts for 3 minutes. The participants' initial stepping stage was determined on the basis of their age and gender. In addition, their 85% maximum heart rate was calculated on the basis of an age-predicted equation (220 – age).[16] The 85% maximum heart rate value is referred to as the *ceiling post-exercise heart rate*.[16] During a stepping session, the participants completed 3 minutes of stepping on a double 20.3-centimetre step stool at a predetermined cadence (foot-plants per minute) corresponding to the assigned stepping stage. The predetermined cadence ranged from a minimum of 66 to a maximum of 144 steps per minute. The average change in cadence ranged from 12 to 18 steps per minute.[16]

At the end of each 3-minute stepping session, the participants' heart rate was measured and compared with the predetermined post-exercise heart rate ceiling. If the measured heart rate did not equal or exceed the predetermined heart rate value, the participants proceeded to the next 3-minute stepping session.[16] The participants performed these progressively demanding 3-minute stepping sessions until they had achieved a heart rate that equalled or exceeded the ceiling post-exercise heart rate.[14] Once they had achieved the predetermined heart rate, the test was complete.

The mCAFT has been validated in comparisons with maximal treadmill testing (the gold standard measure) to predict VO2 max, with agreement (*r*) of 0.87 and 0.88 in samples of 129 and 154 participants aged 15–69 years, respectively.[17] Moreover, the mCAFT sub-maximal test has been identified as a valid, standardized, and feasible method for determining aerobic fitness.[17]

Inter-instrument reliability of the Charge HR

To estimate its inter-instrument reliability, we concurrently compared the Charge HR variables – heart rate, steps taken, and energy expenditure – with those from two other devices: the ZB for the heart rate variable and the Fitbit One for the steps taken and energy expenditure variables.

Statistical analysis

The demographic characteristics of the sample were stratified by gender, and then age, height, weight, BMI, and age-related heart rate maximum were described using means, standard deviations, and minimum and maximum scores. The ZB data file was exported into Microsoft Excel 2016 (Microsoft Corporation, Redmond, WA) and stratified into 30-second intervals. We collected Charge HR heart rate data through the Fitbit dashboard using the Fitbit app (Fitbit, Inc., San Francisco, CA) at 30-second intervals. Charge HR steps taken and energy expenditure data were extracted using the software suggested by Fitbit, Inc.

To indicate inter-instrument reliability, we report the Shrout and Fleiss Type 2,1 ICC and the standard error of measurement (SEM), along with the one-sided lower 95% confidence limit for the ICC and the one-sided upper 95% confidence limit for the SEM for the rest, mCAFT, and recovery phases. We determined that ICCs < 0.40 indicated poor reliability, ICCs ≥ 0.40 to < 0.75 indicated fair to good reliability, and ICCs ≥ 0.75 indicated excellent reliability.[18] To determine the level of agreement between the ZB and the Charge HR and between the Charge HR and the Fitbit One, we used MedCalc statistical software, Version 16.2.1 (MedCalc Software bvba, Seoul, Republic of Korea) to calculate Bland–Altman plots of the individual differences against the mean of the two measures for the three phases.[19]

We then summarized the individual agreement between each two devices by the mean difference and the 95% limits of agreement (LoA; ±1.96 times the standard deviation). To examine the average agreement and differences between each two devices, we tested the mean differences using a one-sample *t*-test; we report the mean differences, standard error of differences, *p*-values, and 95% CIs.[19] We performed analyses using IBM SPSS Statistics, Version 22.0 (IBM Corporation, Armonk, NY), and we considered a significance level of p 0.05 as statistically significant.

RESULTS

Sample

Table 2 displays the demographic characteristics of the participants stratified by gender, age, height, weight, BMI, and age-related heart rate maximum. The number of participants who completed each mCAFT stage was as follows: Stage 1, *n* = 3; Stage 2, *n* = 5; Stage 3, *n* = 19; Stage 4, *n* = 14; Stage 5, *n* = 14; Stage 6, *n* = 2; and Stage 7, *n* = 3.

Inter-instrument reliability between the Charge HR and ZB heart rate variable

At rest, the Charge HR and the ZB inter-instrument ICC was ≥0.97 (SEM range, min–max, 1.02–1.32). During the mCAFT phase and throughout the recovery phase, the Charge HR and the ZB ICCs were ≥0.89 (SEM range, min–max, 1.30–3.98) and ≥0.68 (SEM range, min–max, 3.58–8.35), respectively (see Table 3).

Table 2   Demographic Characteristics of the Sample

Characteristic	Women (n = 30)		Men (n = 30)	
	Mean (SD)	Range, min–max	Mean (SD)	Range, min–max
Age, y	48 (15)	23–68	48 (15)	21–68
Height, m	1.70 (0.05)	1.60–1.80	1.78 (0.06)	1.68–1.88
Weight, kg	69 (11)	50–100	79 (8.5)	64–97
BMI, kg/m^2	24 (3.5)	18–34	25 (2.3)	20–31
Age-related 85% heart rate maximum (bpm)	147 (13)	129–167	146 (13)	129–169

bpm = beats per minute.

Table 3   ICCs and SEMs for the Charge HR and ZB Heart Rate Variables in the Rest, mCAFT, and Recovery Phases

Measure	Time, min:s	Rest			mCAFT			Recovery		
		ICC	SEM	95% CI	ICC	SEM	95% CI	ICC	SEM	95% CI
1st	0:30	0.98	1.08	0.96, 73.70	0.89	3.98	0.83, 103.30	0.71	8.35	0.62, 147.86
2nd	1:00	0.98	1.10	0.96, 73.97	0.95	3.20	0.92, 113.77	0.83	6.47	0.73, 126.04
3rd	1:30	0.98	1.02	0.96, 73.67	0.98	2.25	0.96, 120.91	0.84	5.48	0.73, 114.34
4th	2:00	0.97	1.32	0.96, 74.40	0.96	3.72	0.93, 126.99	0.82	5.39	0.71, 108.56
5th	2:30	0.97	1.22	0.95, 73.70	0.98	2.65	0.97, 127.20	0.77	5.64	0.65, 106.29
6th	3:00	0.98	1.05	0.97, 73.57	0.98	2.97	0.97, 128.82	0.75	5.75	0.61, 101.77
7th	3:30	0.98	1.06	0.96, 73.74	0.96	2.96	0.93, 117.80	0.80	4.74	0.71, 98.29
8th	4:00	0.98	1.07	0.96, 73.80	0.97	2.84	0.95, 126.57	0.76	5.14	0.62, 98.28
9th	4:30	0.98	1.05	0.97, 73.81	0.97	2.99	0.96, 131.85	0.80	4.79	0.68, 95.58
10th	5:00	0.97	1.27	0.95, 73.95	0.96	3.68	0.93, 135.71	0.81	4.79	0.70, 95.60
11th	5:30	0.97	1.27	0.96, 74.20	0.98	2.63	0.97, 136.66	0.75	5.30	0.62, 95.39
12th	6:00	0.98	1.01	0.96, 73.59	0.99	1.91	0.99, 136.74	0.74	5.61	0.57, 95.49
13th	6:30	0.98	1.04	0.97, 73.59	0.97	2.11	0.94, 126.14	0.89	3.58	0.69, 91.82
14th	7:00	0.98	1.10	0.97, 73.67	0.96	2.52	0.93, 133.94	0.73	5.46	0.60, 94.99
15th	7:30	0.97	1.24	0.95, 73.90	0.93	3.23	0.88, 142.33	0.76	5.39	0.62, 94.06
16th	8:00	0.97	1.19	0.96, 74.05	0.93	3.25	0.88, 145.88	0.75	5.20	0.60, 94.19
17th	8:30	0.99	0.74	0.97, 73.36	0.96	2.56	0.92, 147.02	0.72	5.53	0.60, 94.24
18th	9:00	0.98	1.07	0.97, 74.01	0.95	2.73	0.91, 150.35	0.68	5.66	0.58, 94.09
19th	9:30	0.98	1.05	0.96, 73.76	0.99	1.30	0.98, 128.55	0.69	5.40	0.53, 92.59
20th	10:00	0.97	1.24	0.96, 73.93	0.99	1.55	0.96, 140.04	0.77	4.51	0.64, 89.84
21st	10:30	—	—	—	0.99	1.62	0.98, 147.18	—	—	—
22nd	11:00	—	—	—	0.99	1.53	0.98, 150.50	—	—	—
23rd	11:30	—	—	—	0.99	1.53	0.98, 152.50	—	—	—
24th	12:00	—	—	—	0.99	1.62	0.98, 153.48	—	—	—

Note: Dashes indicate no value (testing phase ended).
ICC = intra-class correlation coefficient; SEM = standard error of measurement; Charge HR = Fitbit Charge Heart Rate; ZB = Zephyr BioHarness; mCAFT = modified Canadian Aerobic Fitness Test.

Inter-instrument reliability between the Charge HR and the Fitbit One steps taken and energy expenditure variables

The number of steps taken and the energy expenditures recorded by the Charge HR and the Fitbit One, during the mCAFT phase only, demonstrated ICCs of 0.97 (SEM 83.00) and 0.77 (SEM 14.70), respectively (see Table 4).

Inter-instrument levels of agreement

The average agreement difference of heart rate in our pair-wise comparison of devices indicated small mean differences and narrow 95% CIs of –0.20 beats per minute

Table 4   ICCs and SEMs for the Charge HR and Fitbit One Steps Taken and Energy Expenditure Variables during the mCAFT Phase

Variable	Steps taken and energy expenditure		
	ICC	SEM	95% CI
Steps taken	0.97	83.00	0.96, 1,021.00
Energy expenditure	0.77	14.70	0.65, 140.00

ICC = intra-class correlation coefficient; SEM = standard error of measurement; Charge HR = Fitbit Charge Heart Rate; mCAFT = modified Canadian Aerobic Fitness Test.

(95% CI: –0.10, 0.30), 4.00 beats per minute (95% CI: 3.70, 4.30), and 1.00 beats per minute (95% CI: 0.55, 1.45) at rest, during the mCAFT, and in recovery, respectively (see Table 5). In addition, the steps taken and energy expenditure comparisons yielded mean differences of 79.40 steps (95% CI: 53.70, 105.10) and 39.20 calories (95% CI: 34.00, 44.40), respectively (see Table 5). However, when assessing individual levels of agreement in heart rate, the Bland–Altman plots displayed wider 95% LoA for the mCAFT (48.40, 40.30) and recovery phase s (39.80, 37.70) compared with at rest (5.20, 5.7; see Figure 1a–c).

Table 5   Sample *t*-Test of the Mean Difference between the Charge HR and Adopted Measures

Phase	*t*-test	Variable	Mean difference*	SE	95% CI
Rest	Charge HR vs. ZB	Heart rate (bpm)	–0.20	0.05	–0.10, 0.30
mCAFT	Charge HR vs. ZB	Heart rate (bpm)	4.00	0.16	3.70, 4.30
Recovery	Charge HR vs. ZB	Heart rate (bpm)	1.00	0.23	0.55, 1.45
mCAFT	Charge HR vs. Fitbit One	Steps	79.40	13.11	53.70, 105.10
mCAFT	Charge HR vs. Fitbit One	EE (calories)	39.20	2.66	34.00, 44.40

*p < 0.05.
Charge HR = Fitbit Charge Heart Rate; SE = standard error; ZB = Zephyr BioHarness; bpm = beats per minute; mCAFT = modified Canadian Aerobic Fitness Test; EE = energy expenditure.
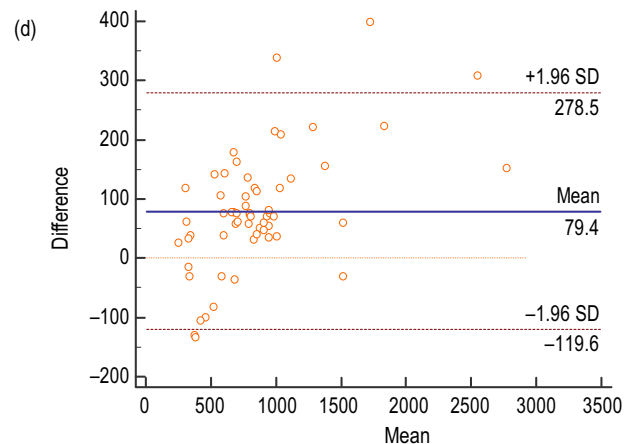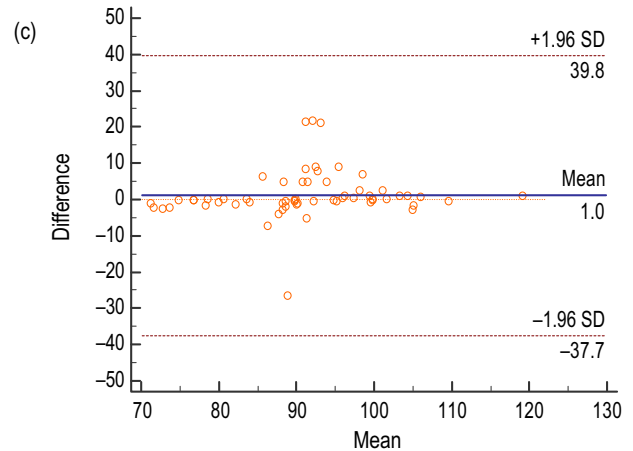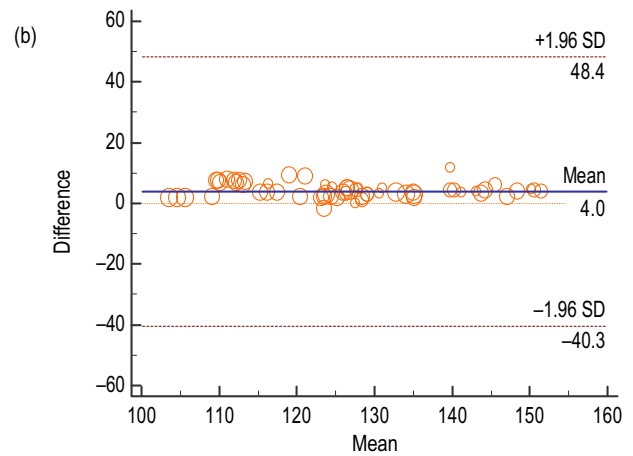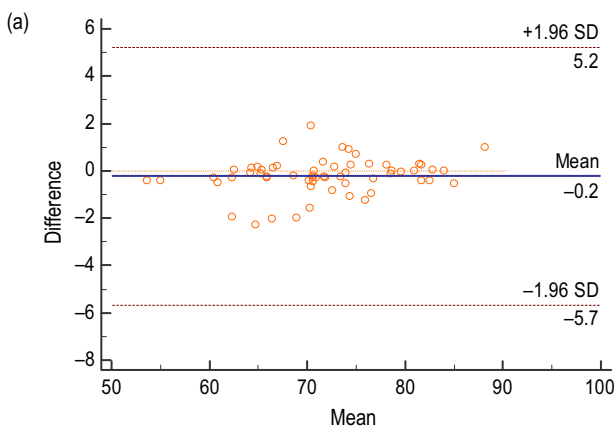

Figure 1   Bland–Altman plots displaying 95% LoA in pair-wise device comparisons: heart rate at rest, ZB and Charge HR (a); heart rate during mCAFT, ZB and Charge HR (b); heart rate during recovery, ZB and Charge HR (c); steps taken during mCAFT, Charge HR and Fitbit One (d); and energy expenditure during mCAFT, Charge HR and Fitbit One (e).

LoA = limits of agreement; ZB = Zephyr BioHarness; Charge HR = Fitbit Charge Heart Rate; mCAFT = modified Canadian Aerobic Fitness Test.
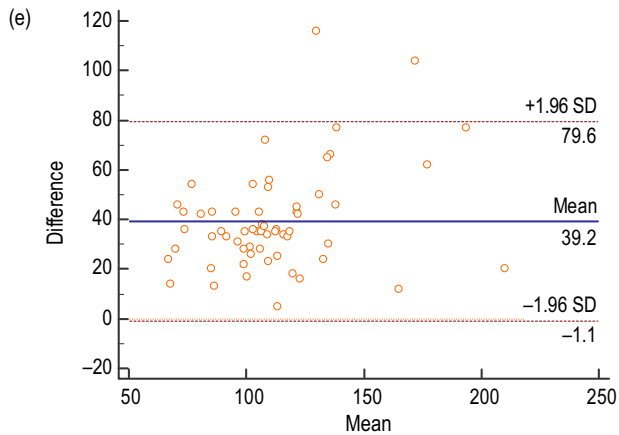
Figure 1   Continued

DISCUSSION

The purpose of this study was to determine the inter-instrument reliability and agreement parameters of the Charge HR device compared with the ZB and of the Charge HR compared with the Fitbit One at rest, during mCAFT, and throughout recovery. Our results supported, and we accepted, both of our a priori hypotheses of obtaining correlations of greater than 0.60 for the Charge HR compared with the ZB heart rate variable and correlations of greater than 0.80 for the Charge HR compared with the Fitbit One steps taken and energy expenditure variables.

Well-conducted systematic reviews aim to summarize and critically appraise a topic and provide the highest level of evidence from which inferences can be drawn. Our recent systematic review of the measurement properties of the ZB yielded good- to excellent-quality evidence that the ZB could provide reliable and valid measurements of heart rate.[11] Therefore, we compared the Charge HR measurements with those of the ZB to establish the Charge HR's inter-instrument reliability and agreement parameters in multiple phases.

Wang and colleagues assessed the validity of the Charge HR heart rate variable against an ECG under various physical exertions.[20] They reported overall correlation coefficients of 0.84 with the ECG during a treadmill testing protocol (resting and 3.2–9.6 km [2–6 miles] per hour) in 50 healthy participants (22 men, 28 women) with a mean age of 37 years. In addition, they found median differences of 7.2 and 6.4 beats per minute at 6.4 kilometres (4 miles) per hour and 9.6 kilometres (6 miles) per hour for men and women, respectively.[20] Similarly, Boudreaux and colleagues reported correlation coefficients ranging from 0.19 to 0.98 for the Charge HR assessed against an ECG during a treadmill graded exercise test in 30 college students (11 men, 19 women).[21]

Our results compared well with these findings. Moreover, our Charge HR and ZB heart rate correlation coeffi-

cients at rest ($r = 0.97-0.98$), during mCAFT ($r = 0.89-0.99$), and throughout recovery ($r = 0.68-0.89$) compared well with previously reported measures of the ZB compared with an ECG.[20–24]

We found agreement differences of 79.00 steps per every 3,500 steps (2.2% error) between the Charge HR and the Fitbit One. These differences were comparable with the findings in the study by Takacs and colleagues, who noted agreement differences of 5.00 steps per every 700 steps (0.70% error) between the Fitbit One and direct observation.[14] We noted agreement differences of 39.00 calories per every 250 calories (15.60% error) between the Charge HR and Fitbit One energy expenditure variable. This agreement difference was much larger than the finding by Lee and colleagues of 24.00 calories per every 700 calories (3.70% error) between the Fitbit One and the metabolic system during walking or running activities.[13] The larger percentage discrepancy in error of 15.60% in our study could be due to the fact that the Charge HR is worn on the wrist instead of at the waist. Because an accelerometer is worn on the wrist, excessive arm motions during an activity may contribute to larger accelerometry counts and, therefore, overestimation of energy expenditure.[25]

We used a sub-maximal fitness test as the stressor that would cause physiological changes after a period of rest. This activity phase was followed by recovery, which we considered to be a resumption of the rest phase. By including the three phases, and because of the way we took measurements, we expect that transitioning through the phases was considered unstable intervals. This would explain the high SEM value (3.98 beats per minute) at 0:30 seconds during the first measurement of sub-maximal testing (first 30 seconds) and the SEM of 8.35 beats per minute at 0:30 seconds during the first measurement (first 30 seconds) in the recovery phase. This result could also be because the Charge HR was unable to monitor this change accurately. However, smaller standard errors of measurement between the Charge HR and the ZB were observed as the testing progressed.

Bland and Altman put forward two aspects of agreement.[19] In average agreement estimation, the mean difference ± 1.96 × SE of differences is reported. In individual agreement, the 95% LoA as mean difference ± 1.96 × SD of differences, along with Bland–Altman plots, is computed.[10] In our study, small mean differences and narrow 95% CIs indicated the presence of non-significant systematic differences between two devices when recording heart rate measurements. The wider LoA during the mCAFT and throughout recovery could be due to (1) a large variability in our study sample and (2) the nature of mCAFT testing. Because mCAFT is a sub-maximal test, it requires participants to achieve 85% of their age-related maximum heart rate. Moreover, we included participants aged 21–68 years and calculated 85% age-related

maximum heart rates of 129.00–169.00 beats per minute for men and 129.00–167.00 beats per minute for women. We believe that these two factors could have contributed to the wider LoA. In addition, during a period of proper acclimatization (rest phase of 10 minutes), in which participants did not have to achieve a specific age-related maximum heart rate, much narrower LoAs were reported. Therefore, when interpreting heart rate LoA during the mCAFT and recovery phases, both the nature of mCAFT testing and the large variability in our sample must be considered.

Our study findings have implications for the field of continuous heart rate monitoring. By establishing the inter-instrument reliability of the Charge HR against the ZB device, more accurate real-time field-based wireless measurements of heart rate using small, portable, and less costly devices, in both stable and sub-maximal conditions, can be obtained. Moreover, assessing the agreement parameters between the Charge HR and ZB heart rate measurements provided information about their possible interchangeable use.

The strengths of this study include the support generated for the inter-instrument reliability and agreement parameters of the Charge HR device during both stable conditions (rest and recovery) and a standardized sub-maximal fitness test. In addition, we sampled a large number of participants of both genders and across a wide age range. Nevertheless, limitations to our approach should be considered when interpreting our findings. First, we did not evaluate the performance of the Charge HR against a gold standard (calibration). However, the properties of the criterion measurements used in our study have been reported in the literature and have been deemed valid. Second, some might consider it a limitation that we used a sub-maximal fitness test for the activity phase instead of a maximal test. We did this because the aim of our study was not to determine aerobic capacity levels by directly measuring VO2 max levels (maximal testing) but rather to use a sub-maximal, standardized, feasible, and functional test or activity that could be administered without specialized equipment. Finally, although we studied the reliability and agreement parameters of the Charge HR during human performance, we did not assess its ability to accurately detect change when it occurred.

CONCLUSION

The Charge HR device heart rate variable demonstrated excellent inter-instrument reliability when compared with the ZB in our healthy cohort at rest, while performing a sub-maximal aerobic fitness test, and during recovery. Similarly, the steps taken and energy expenditure variables displayed excellent reliability between the Charge HR and Fitbit One. In addition, comparing the heart rate measurements of the Charge HR and ZB

throughout these phases provided valuable information regarding their possible interchangeable use.

KEY MESSAGES

What is already known on this topic

Both the absolute and the relative reliability parameters of the Fitbit Charge Heart Rate (Charge HR) wristband device have been established among healthy participants.

What this study adds

This study aimed to establish the inter-instrument reliability and agreement parameters of the Charge HR against Zephyr Bioharness and Fitbit One devices among healthy participants at rest, during a sub-maximal test, and throughout recovery at 30-second intervals. We found that the Charge HR device heart rate variable displayed excellent inter-instrument reliability when compared with the Zephyr Bioharness in healthy participants throughout the three phases. Similar results were found between the Charge HR and Fitbit One in terms of steps taken and energy expenditure variables.

REFERENCES

1. Fitbit [homepage on the Internet]. San Francisco: Fitbit; 2015 [cited 2016 Apr]. Fitbit Charge HR Wireless Heart Rate + Activity Wrist Band. Fitbit Charge HR TM Product Manual Version 1.0. Available from: https://staticcs.fitbit.com/content/assets/help/manuals/manual_charge_hr_en_US.pdf.
2. American College of Sports Medicine. Health-related physical fitness assessment manual. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013.
3. Jeukendrup A, Van Dieme A. Heart rate monitoring during training and competition in cyclist. J Sports Sci. 1998;16(Supplement 1):S91–S99. https://doi.org/10.1080/026404198366722. Medline:22587722
4. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr. 2014;1(5). https://doi.org/10.3389/fnut.2014.00005. Medline:25988109
5. Zephyr Technology. BioHarness 3.0 user manual [Internet]. Annapolis (MD): Zephyr Technology; 2012 [cited 2016 Apr]. Available from: https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf.
6. Fitbit. Fitbit one wireless activity + sleep tracker: user manual [Internet]. San Francisco: Fitbit; 2012. Available from: https://staticcs.fitbit.com/content/assets/help/manuals/manual_one_en_US.pdf.
7. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. New York: Oxford University Press; 1995.
8. Brazier J, Deverill M. A checklist for judging preference-based measures of health-related quality of life: learning from psychometrics. Health Econ. 1999;8(1):41–51. Medline:10082142
9. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Norwalk (CT): Appleton & Lange; 1993.
10. Bunce C. Correlation, agreement, and Bland-Altman analysis: statistical analysis of method comparison studies. Am J Ophthalmol. 2009;148(1):4–6. https://doi.org/10.1016/j.ajo.2008.09.032. Medline:19540984
11. Nazari G, Bobos P, MacDermid JC, et al. Psychometric properties of the Zephyr Bioharness device: a systematic review. BMC Sports Sci Med Rehabil. 2018;10:6. https://doi.org/10.1186/s13102-018-0094-4. Medline:29484191

12. 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. https://doi.org/10.1016/j.ijcard.2015.03.038. Medline:25795203

13. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc. 2014;46(9):1840–8.

14. Takacs J, Pollock CL, Guenther JR, et al. Validation of the Fitbit One activity monitor device during treadmill walking. J Sci Med Sport. 2014;17(5):496–500. https://doi.org/10.1016/j.jsams.2013.10.241. Medline:24268570

15. Nazari G, MacDermid JC, Kathryn SE, et al. Reliability of Zephyr Bioharness and Fitbit Charge measures of heart rate and activity at rest, during the modified Canadian Aerobic Fitness Test and recovery. J Strength Cond Res. 2017; 33(2):559–571. https://doi.org/10.1519/JSC.0000000000001842.

16. Canadian Society for Exercise Physiology. The Canadian physical activity, fitness and lifestyle appraisal. 3rd ed. Ottawa: Canadian Society for Exercise Physiology; 2010.

17. Weller IM, Thomas SG, Gledhill N, et al. A study to validate the modified Canadian Aerobic Fitness Test. Can J Appl Physiol. 1994;20 (2):211–21. Medline:7640647.

18. Rosner, B. Fundamentals of biostatistics. 6th ed. Boston: Duxbury Press; 2005.

19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327 (8476):307–10. https://doi.org/10.1016/S0140-6736(86)90837-8. Medline:2868172

20. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol. 2017;2(1):104–6. https://doi.org/10.1001/jamacardio.2016.3340. Medline:27732703

21. Boudreaux BD, Cormier CL, Williams BM, et al. Accuracy of wearable devices for determining physiological measures during different physical activities. Med Sci Sports Exerc. 2017;49(5S):762. https://doi.org/10.1249/01.mss.0000519027.48408.29.

22. Dolezal BA, Boland DM, Carney J, et al. Validation of heart rate derived from a physiological status monitor-embedded compression shirt against criterion ECG. J Occup Environ Hyg. 2014;11(12):833–9. https://doi.org/10.1080/15459624.2014.925114. Medline:24896644

23. Flanagan SD, Comstock B, Dupont WH, et al. Concurrent validity of the Armour39 heart rate monitor strap. J Strength Cond Res. 2014;28 (3):870–3. https://doi.org/10.1519/JSC.0b013e3182a16d38. Medline:23860286

24. Kim JH, Roberge R, Powell JB, et al. Measurement accuracy of heart rate and respiratory rate during graded exercise and sustained exercise in the heat using the Zephyr Bioharness TM. Int J Sports Med. 2013;34(6):497–501. https://doi.org/10.1055/s-0032-1327661. Medline:23175181

25. Bouten CV, Westerterp KR, Verduin M, et al. Assessment of energy expenditure for physical activity using a triaxial accelerometer. Med Sci Sports Exerc. 1994;26(12):1516–23. Medline:7869887

APPENDIX: SAMPLE SIZE CALCULATION FOR RELATIVE RELIABILITY HYPOTHESIS TESTING

The estimated required sample size for reliability hypothesis testing is calculated as follows:

$$n = 0.5k\frac{(Z\propto + Z\beta)^2}{\delta^2\ (k-1)} + 2$$

where n = number of occasions, k = 2; $Z = \propto$ is the tabled Z-value associated with the α value of interest (a Z-value [one-tailed] of 0.05 = 1.645); is the Z-β value associated with a Type II error (the Z-value [one-tailed] for a β of 0.20 = 0.842); and δ is the difference between the null hypothesis Z-transformed R value and the expected Z-transformed R value.

$$\delta = ZR\,expected - ZR\,null,$$

where *ZR expected* is the Z-value associated with the reliability one hopes to obtain in one's study,

$$ZR\,expected = 0.5\,natural\log\frac{1 + (k-1)\,R\,expected}{1 - R\,expected},$$

where *ZR lower limit* is the lower confidence limit for the desired CI width.

$$ZR\,lower\,limit = 0.5\,natural\log\frac{1 + (k-1)\,R\,lower\,limit}{1 - R\,lower\,limit}$$

CALCULATION

1. $ZR\,expected = 0.5\,natural\log\frac{1 + (2-1)\,0.90}{1 - 0.90}$

$ZR\,expected = 1.47$

2. $R\,lower\,limit = 0.9 - 0.10$

$R\,lower\,limit = 0.80$

3. $ZR\,null = 0.5\,natural\log\frac{1 + (2-1)\,0.80}{1 - 0.80}$

$ZR\,null = 1.09$

4. $\delta = 1.47 - 1.09 = 0.38$

$\delta^2 = 0.38^2 = 0.14$

5. $n = 0.5\,k\frac{(Z\propto + Z\beta)^2}{\delta^2\,(k-1)} + 2$

$n = 0.5\,(2)\frac{(1.645 + 0.842)^2}{0.14\,(2-1)} + 2$

$n = 46\,patients$

6. Estimating an expected dropout of 20%:

$n = \frac{46}{0.80} = 57.5$

Therefore, a sample size of 60 participants will be required.