EXAMINING THE RELABILITY AND VALIDITY OF THE FITBIT CHARGE 2™ TECHNOLOGY FOR HEART RATE DURING TREADMILL EXERCISE

Mahalie Hargis[1], Robert O. Powell[1], and Suzanne M. Konz[1]

Marshall University, Huntington, WV, United States[1]

The purpose of this study was to determine the reliability and validity of the Fitbit® Charge 2™ compared to a 4-lead ECG to monitor heart rate during exercise. All participants completed a VO_2Max test to determine the participants' cardiorespiratory fitness and appropriate work loads for the intensity protocol. Participants wore a Fitbit® Charge 2™ on the wrist and a 4-lead ECG. 16 participants were assigned to testing conditions: GPS stride length/manual stride length and hold/no-hold on the handrails of the treadmill. The participants completed a walking protocol including light (<40% of HRR), moderate (40-60% of HRR), and vigorous (>60% of HRR) intensities while wearing both devices. Intraclass Correlation Coefficient (ICC) was used to analyze the results. The ECG is a reliable and valid means to monitor heart rate. The Fitbit® Charge 2™ demonstrated poor reliability and validity to monitor heart rate.

KEY WORDS: Fitbit™, Fitbit® Charge 2™, ECG, heart rate.

INTRODUCTION: Wrist-worn monitors are predicted to account for 87% of wearable devices in 2018 (Smart, 2014). Among many capabilities, these devices record steps, calculate energy expenditure, and measure heart rate. One of the most popular brands for wearable fitness trackers is the Fitbit™. In 2016, Fitbit™ reported revenue of $2.17 billion; selling a total of 22.3 million connected health and fitness devices (Business Wire, 2017). The Fitbit® Charge 2™ lead the industry in health and fitness device sales from September through December of 2016 (Business Wire, 2017). Accuracy of tracking devices are affected by personal physiology, the location of wear, and type of movement (Lewis, Directo, Kim, & Dolezal, 2016). Customers using these devices to track health and fitness expect consistency and accuracy and unaware of potential inaccuracies which exposes them to health risks by the underestimation or overestimation of data (e.g. patients using this device to monitor heart rate for a prescribed HR range during exercise). Clinical setting are increasing usage to monitor patient activity levels as well as to assist patients in the management of their medical conditions (Walker, Hickey, & Freedson, 2016).

Devices like the Fitbit® Charge 2™ use a technology to detect heart rate called photoplethysmography (PPG) (Lai and Kim, 2015). PPG is a simple, accurate, low-cost technique using low-intensity light on the surface of the skin to non-invasively measure volumetric changes in the blood in the peripheral circulation (Lai and Kim, 2015; Li & Kim, 2017; Pietilä et al., 2017). PPG devices use LED light to penetrate the body. The PPG technology that Fitbit™ uses is PurePulse™ and uses green light (Smart, 2014). Fitbit® products are wrist-worn devices and are prone to motion artifacts due to movement and vibration when the electrode-skin interface is disrupted (Li & Kim, 2017).

While wrist-worn devices generally have a good reliability in monitoring step count, they traditional do not have high correlation for heart rate (Li & Kim, 2017). Studies have been performed testing the accuracy of Fitbit™ heart rate monitoring devices, but are limited. When compared to a Polar® RS400 HR chest strap, six activity trackers including the Fitbit® Charge HR™ provided an accurate measurement of HR during walking and running activities (Stahl, An, Dinkel, Noble, & Lee, 2016). In a study by Cadmus-Bertram, et al (2017), the Fitbit® Charge™ performed best at rest (within 5 beats) compared to an ECG. Accuracy decreased for the device with a variance of 20 to 40 beats per minute as activity intensity increased. However, Fitbit® PurePulse™ Trackers did not provide a valid measure of the users' heart rate during

moderate to high intensity exercise (Jo & Dolezal, 2016). The Fitbit® Charge HR™ had a mean absolute percentage error value of 6.2% and a Pearson product-moment correlation coefficient (r) of r=0.933 when compared to five other HR monitors (Stahl, et al., 2016). The purpose of this study was to determine the reliability and validity of the Fitbit® Charge 2™ compared to a 4-lead ECG to monitor heart rate during an exercise protocol of light, moderate, and vigorous intensities (Cadmus-Bertram, et al., 2017).

METHODS: Study Design: The study is a cross sectional repeated measure cross-over design.
Participants: A total of 16 participants who were males (n=3) and females (n=13) completed the study. Participants were aged 18-35-years-old, had a BMI < 30, and had a physical activity level that ranged from sedentary to lightly active and/or had a VO_2Max of <40 ml/kg/min. Participants completed an informed consent that was approved by the IRB. The participants also completed a Health History Questionnaire and a Physical Activity Readiness Questionnaire (PAR-Q).
Protocol and Data Collection: *Baseline testing:* All participants completed a VO_2Max test in order to determine the participants' cardiorespiratory fitness and the appropriate work loads for Protocol 2. The participants wore a Fitbit® Charge 2™ on their wrist and a 4-lead ECG. Participants completed a modified Bruce treadmill protocol by starting out at 0% grade and a walking speed of 1.5 m/s. Throughout the baseline test, the treadmill maintained a constant speed of 1.5 m/s and increased in incline by 2.5% each additional minute. Once the incline on the treadmill reached its maximum, the speed was increased by 0.004 m/s each minute. The testing continued until the participant chose to terminate the test due to fatigue, the participant was ready to stop, or abnormal heart rate responses occured with work rate. Heart rate was recorded at the end of each minute for each device.
Testing protocol: 16 qualified participants who met inclusion criteria were randomly assigned to one of two groups by drawing a card to determine testing condition order. Hold or no-hold of handrails were conditions for walking on the treadmill. Hold and no-hold conditions referred to the participants either holding or not holding onto the handrails of the treadmill. Participants switched testing conditions for the second testing session.
The testing protocol was a walking protocol that included light (<40% of HRR), moderate (40-60% of HRR), and vigorous (>60% of HRR) intensities while wearing both devices. (Lippincott, Williams, & Wilkins, 2014), Protocol 2 had 5 stages and each stage lasted 3 minutes lasting a total of 15 minutes. Participants walked at 1.3 m\s for stage 1 (light intensity); 1.6 m/s for stage 2 (moderate intensity); 1.8 m/s for stage 3 (vigorous intensity) then slowed back down to 1.6 m/s (stage 4) and 1.3 m\s (stage 5). The grade of the treadmill remained at 0% the entire protocol. Heart rate was recorded at the end of each minute for each device.
Data Processing: The heart rate data was entered in a locked spreadsheet on Microsoft Excel. The data was grouped by device, bar-hold condition, and protocol intensity. The data was transferred to SPSS version 22 for statistical analysis. The data was transferred to SPSS version 22 (IBM, Armonk, New York) for statistical analysis.
Statistical Analyses: Intraclass Correlation Coefficient (ICC), types (1,k) and (2,k), were used to analyze the data SPSS version 22. The dependent variable for the ICC analysis was heart rate. The level of significance was set at a p-value of <0.05.

RESULTS: ECG reliability indicated good to excellent results (See Table 1). The single measure ICC (1, k) for both hold-type conditions was 0.763, 95% CI [0.627, 0.887] ($F(15)= 58.815$, $p<0.000$). The Fitbit Charge2 demonstrated poor reliability with the study protocol (See Table 2).

Table 1. ECG Heart Rate Reliability (1,k)

Condition	ICC [95% C]	F-test Significance
Both hold conditions	0.763 [0.627, 0.887]	0.000
Holding bar	0.782 [0.642, 0.900]	0.000
No hold of bar	0.899 [0.819, 0.957]	0.000

Table 2. Fitbit™ Heart Rate Reliability (1,k)

Condition	ICC [95% C]	F-test Significance
Both hold conditions	0.213 [0.105, 0.423]	0.000
Holding bar	0.487 [0.303, 0.714]	0.000
No hold of bar	0.370 [0.198, 0.616]	0.000

Table 3. Fitbit™ vs. ECG Heart Rate Reliability (2,k)

Condition	ICC [95% C]	F-test Significance
ECG vs Fitbit™ Holding bar	*0.312* [0.188, 0.531]	0.000
ECG vs Fitbit™ No hold of bar	*0.591* [0.387, 0.794]	0.000

The single measure ICC (1, k) both hold-type conditions was .213, 95% CI [0.105, 0.423] ($F(15)= 5.859$, $p<0.000$). A poor degree of reliability was found between the comparison of Fitbit and ECG measurements (See Table 3). The single measure ICC was 0.312, 95% CI [0.105, 0.423] ($F(15)=15.705$, $p<.000$). The ICC was 0.591, 95% CI [0.387, 0.794] ($F(15)= 60.423$, $p<0.000$).

DISCUSSION: The gold standard ECG testing conditions produced ICC's of 0.899 when not holding the handrails and 0.782 when holding the handrails. Fitbit® Charge 2™ was determined to be of poor reliability and validity to detect heart rate. Holding the handrails (0.487) produced higher reliability than not (0.307). The lower reliability for not holding the handrails was a surprise as this condition allowed for a more natural swing of the arms to occur during the exercise protocol. Most individual do not hold on to treadmill when using it. The bar-hold condition does, however, address a small population that does hold on for balance or stability during treadmill use. The strength of validity was poor to fair for the Fitbit® Charge 2™ when compared to the ECG. The ICC values were of poor agreement between the two devices indicating that the Fitbit Charge 2™ provides poor results for both bar-hold conditions.

The results of this study are in agreement with other study results. Benedetto et al. (2018), Wang et al., (2017), and Wallen et al (2016) found that heart rate measurements taken by Fitbit™ products were routinely different than ECG. The average heart rate recorded by the Fitbit™ was approximately 6 beats per minutes (BPM) slower than the average with variance causing a possible underestimation of 30 bpm when compared to ECG (Benedetto, 2018). Heart rate values were underestimated by 34 bpm or overestimated by 39 bpm (wang et al., 2017). Wallen et al (2016) determined that heart rate underestimation occurred at a rate of 1-9%. However, Stahl, et al. (2016) determined that wrist-based HR monitors provide accurate measurement of HR during walking and running activities. Wrist-worn fitness trackers measure heart best when users are at rest and accuracy varied with activity. Chest strap-monitors should be used when accuracy of heart monitoring is needed (Wang, et al., 2017).

Lack of reliability and validity of the Fitbit® Charge 2™ is due to a variety of reasons. Changes in skin color, tattoos, temperature, perspiration, erratic movement, skin pressure variations, and excess hair contribute to the error in PPG technology (Scully, 2011). The optimal LED color choice based on ethnicity is a yellow LED signal with a 3.29% error for Caucasians and Asians and a green LED with a 2.98 % error for Middle Eastern and Indian subjects (Lai and Kim, 2015). Fitbit® Charge 2™ uses a green LED signal. The subjects in this study were all Caucasian. Profuse sweating impacts the ability of wrist-worn devices to properly couple with shin leading to decreased reliability (Li & Kim, 2017).

Another factor leading to lack of reliability is that the placement of the device was not in the most optimal area to effectively read changes on the skin. Instructions are to place the device three finger widths above the wrist (Fitbit, n.d.). Salazar, Lucio, & Funk (2017) found a significant difference in heart rate results when tracking device is not placed in manufacturer recommended position. An unstable position or movement of the wrist causes inconsistency in the interaction of the PPG sensors and the skin (Benedetto, 2017). Watch-type devices are desirable for many but are susceptible and associated with strong and frequent motion artifacts (Lai and Kim, 2015). A need exists to decrease the impact of external factors (e.g. tattoos, device placement, temperature, perspiration, erratic movement, skin pressure variations, and excess hair) that impact the reliability and validity of fitness tracking devices (Jo, et al., 2016).

CONCLUSION: The reliability and validity of Fitbit™ devices to monitor heart rate is important for consumers who are investing in these products. Those who use Fitbit Charge 2™ devices in a cardiac rehabilitation setting could be putting their safety at risk. Invalid readings of a patient's heart rate who has known cardiovascular disease could be dangerous. Those who use Fitbit Charge 2™ devices in a sport performance setting could also be affected. Invalid readings of a player's heart rate could lead them to training within incorrect training zones. For instance, training at a lower intensity could limit the amount of benefits an athlete should gain from their exercise program. Future research should aim to improve Fitbit™ technology. Other wearable forms of devices could also be used as an alternative to monitor heart rate.

REFERENCES:

Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D. C., Pensabene, V., & Actis, P. (2018). Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS ONE*. Vol. 13 Issue 2.

Cadmus-Bertram, L., Gangnon, R., Wirkus, E. J., Thraen-Borowski, K. M., & Gorzelitz-Liebhauser, J. (2017). The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. *Annals of Internal Medicine*, 166(8), 610-612.

Fitbit. (n.d.) Fitbit Charge HR Product Manual. San Francisco.

Fitbit Reports $574M Q416 and $2.17B FY16 Revenue, Sells 6.5M devices in Q416 and 22.3M devices in FY16 (2017, February). *Business Wire.* Retrieved from https://www.businesswire.com/news/home/20170222006497/en/Fitbit-Reports-574M-Q416-2.17B-FY16-Revenue

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

Jo, E., & Dolezal, B. A. (2016). Validation of the Fitbit® Surge™ and Charge HR™ fitness trackers.

Lai, P. H., & Kim, I. (2015). Lightweight wrist photoplethysmography for heavy exercise: motion robust heart rate monitoring algorithm. *Healthcare Technology Letters*, *2*(1), 6-11.

Li, M., & Kim, Y. T. (2017). Design of a Wireless Sensor System with the Algorithms of Heart Rate and Agility Index for Athlete Evaluation. *Sensors*, 17(10), 2373.

Pietilä, J., Mehrang, S., Tolonen, J., Helander, E., Jimison, H., Pavel, M., & Korhonen, I. (2017). Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities. In *EMBEC & NBC 2017* (pp. 145-148). Springer, Singapore.

Salazar, V. E., Lucio, N. D., & Funk, M. D. (2017). Accuracy of Fitbit Charge 2 Worn at Different Wrist Locations During Exercise. In *International Journal of Exercise Science: Conference Proceedings* (Vol. 2, No. 9, p. 41).

Scully, C. G., Lee, J., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2012). Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Transactions on Biomedical Engineering*, 59(2), 303-306.

Smart watches and Smart Bands Dominate Fast-Growing Wearables Market. (2014). *CCS Insight.* Retrieved from http://www.ccsinsight.com/press/company-news/1944-smartwatchesand-smart-bands-dominate-fast-growing-wearables-market.

Stahl, S. E., An, H. S., Dinkel, D. M., Noble, J. M., & Lee, J. M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & Exercise Medicine*, *2*(1), e000106.

Walker, R. K., Hickey, A. M., & Freedson, P. S. (2016). Advantages and limitations of wearable activity trackers: Considerations for patients and clinicians. *Clinical Journal of Oncology Nursing*, 20, 606-610.

Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisløff, U., & Coombes, J. S. (2016). Accuracy of heart rate watches: implications for weight management. *PLoS ONE*, 11(5), e0154420.

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