EXAMINING THE RELIABILITY AND VALIDITY OF THE FITBIT® CHARGE 2™ FOR STEP COUNT 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™ to count steps during a treadmill protocol. The participants wore a Fitbit® Charge 2™ on their wrist. 16 participants who met inclusion criteria were assigned to the following testing conditions; GPS stride length/manual stride length and hold/no-hold onto the handrails of the treadmill. The participants completed a walking protocol that included light (<40% of HRR), moderate (40-60% of HRR), and vigorous (>60% of HRR) intensities. The Fitbit® Charge 2™ step count and the manually counted steps were documented. Intraclass Correlation Coefficients (ICC) analysis found that the Fitbit® Charge 2™ was not reliable or valid to count steps.

KEY WORDS: Fitbit™, Fitbit® Charge 2™, steps, stride length, intensity.

INTRODUCTION: One of the most popular brands for wearable fitness trackers is Fitbit™. Wrist-worn monitors are predicted to account for 87% of wearable devices in 2018 (Smart, 2014). In 2016, Fitbit™ reported revenue of $2.17 billion selling 22.3 million connected health and fitness devices (Business Wire, 2017). The Fitbit® Charge 2™ was the top selling connected health and fitness device for the months of September through December of 2016 (Business Wire, 2017). These devices are capable of recording steps, calculating energy expenditure and measuring heart rate. Customers using these devices to track health and fitness expect consistency and accuracy and are unaware of potential inaccuracies which may expose them to health risks. Tracking device accuracy is affected by personal physiology, the location of the sensor and the type of movement (Lewis, Directo, Kim, & Dolezal, 2016).
The step tracking ability of Fitbit™ devices has been previously investigated, with inconsistent results. The disparity causes confusion for consumers who want to know the reliability and validity of a product. The Fitbit® Charge HR™ has been found to have an Intraclass Correlation Coefficient (ICC) of 0.52 and the Fitbit® Surge™ an ICC of 0.73 for step count during low speed treadmill walking (Tophøj, Petersen, Sæbye, Baad-Hansen, & Wagner, 2018). Farina and Lowry (2018) have demonstrated poor agreement of the Fitbit™ devices with the wrist-worn ActiGraph® GT3X+s™ and waist-worn device, Misfit® Shine™. The Fitbit® Flex™ significantly undercounted walking steps with more variance but showed greater validity during jogging and stair climbing when compared to an Actigraph® GT3X+™ (Sushames, Edwards, Thompson, McDermott, & Gebel, 2016). When devices including the Fitbit® One™, Fitbit® Flex™, Fitbit® Charge HR™, and Jawbone® UP24™ were compared to an Actigraph™, the Fitbit® One™ was the strongest performer across all speeds (Chow, Thom, Wewege, Ward, & Parmenter, 2017). Limited research has been conducted on the Fitbit Charge 2™. The purpose of this study was to determine the reliability and validity of the Fitbit® Charge 2™ to count steps during a treadmill protocol (accounting for variables such as holding on to the treadmill and stride length).

METHODS: Participants: There were 16 participants (3 males, 13 females), aged 18-35 years, with a BMI < 30 and a physical activity level that ranged from sedentary to lightly active and/or with a VO_2Max of <40 ml/kg/min. Participants completed an informed consent that was approved by the ethics review board as well as a Health History Questionnaire and a Physical Activity Readiness Questionnaire (PAR-Q).
Protocol and Data Collection: A cross-sectional repeated measures cross over design. Baseline testing: All Participants completed a modified Balke-Ware treadmill walking protocol to assess

their cardiorespiratory fitness (CRF) and the appropriate work loads for Experimental Protocol (Protocol 2). The CRF assessment test started out at 0% grade and a walking speed of 3.3 mph. Throughout the baseline test, the treadmill maintained a constant speed of 3.3 mph and increased elevation by 2.5% every minute. If maximal elevation of the treadmill was achieved, the speed was increased by 0.1 mph 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 were recorded. Additionally, the participants wore a Fitbit® Charge 2™ on their wrist and a 4-lead ECG.

Test protocol: Sixteen qualified participants meeting the inclusion criteria were randomly assigned to one of two groups by drawing a card to determine test condition order. Stride length calculation conditions were manual stride length or GPS stride length. Hold or no-hold of handrails were conditions for walking on the treadmill. Calculation of the manual stride length consisted of taking the number of steps each participant took to walk 9.14 meters and dividing that quantity by 30. This stride length measurement was then input into the Fitbit™ app on the researcher's iPad which reduced inter-phone variability. The GPS stride length was calculated by the automatic stride length settings within the Fitbit™ app. 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. Participants switched testing conditions for the second testing session.

The Experimental Protocol (Protocol 2) was a walking protocol based on heart rate reserve (HRR) that included light (<40% of HRR), moderate (40-60% of HRR), and vigorous (>60% of HRR) intensities while wearing the Fitbit® Charge 2™ on the wrist (Lippincott, Williams, & Wilkins, 2014). This protocol had 5 stages with each stage lasting 3 minutes for a total of 15 minutes. Because walking speed is known to increase stride length and frequency, participants walked at 3 mph for stage 1 (light intensity); 3.5 mph for stage 2 (moderate intensity); 4 mph for stage 3 (vigorous intensity) then slowed back down to 3.5 mph (stage 4) and 3 mph (stage 5) to determine potential effects on validity. The grade of the treadmill remained at 0% for the entire protocol. Step count was recorded at the end of each protocol from the Fitbit® Charge 2™. All trials were filmed for later manual counting of steps.

Data Processing: The data were entered in a locked spreadsheet on Microsoft Excel. The data were transferred to SPSS version 22 for statistical analysis. Video footage was taken of each subject's testing protocol using a Sony video camera. Each video was transferred onto a Seagate Backup Plus (TB) and later viewed using RealPlayer. A GOGO Tally Counter was used to manually count steps.

Statistical Analyses: Intraclass Correlation Coefficient (ICC), type (1,k), was used to analyze the data. The level of significance used was a p-value of <0.05.

RESULTS: Participant's demographics are shown in Table 1. The means and standard deviations for the step count conditions are found in Table 2. Step count reliability and validity of the Fitbit Charge 2 indicated poor to good (See Table 2).

Table 1. Subject Demographics

	Sample (n=16)	Females (n=13)	Males (n=3)
Age	22.0 (1.9)	21.7 (1.1)	23.0 (4.3)
Weight (kg)	66.0 (11.1)	64.7 (11.6)	71.8 (8.2)
Height (cm)	169.4 (11.4)	166.8 (11.1)	180.3 (2.0)
BMI (kg/m^2)	22.9 (2.8)	23.1 (3.0)	21.9 (2.2)
VO_2Max (ml/kg/min)	34.2 (3.3)	33.1 (2.8)	38.6 (0.5)

Table 2. Step Count Means and Std. Deviations (N=16)

	No Hold (# steps)	Hold (# steps)
Fitbit Count	1766.4 ±149.4	1151.4 ± 557.5
Hand Step Count	1881.9 ±110.3	1896.0 ± 121.3
GPS stride setting	1491.4 ± 517.9	1888.8 ± 127.5
Manual Stride Setting	1426.4 ± 514.1	1889.1 ± 103.6

Table 3. Fitbit™ Step Count Reliability- ICC

Condition	*ICC Type*	*Single Measure of ICC*	F-test Significance
no hold, Fitbit vs video	1,2	0.721	0.001
hold, Fitbit vs video	1,2	0.274	0.144
GPS vs manual stride, Fitbit vs video	1,4	0.040	0.332
GPS vs manual stride, Fitbit vs video	1,2	0.328	0.099
manual stride vs manual stride, Fitbit vs video	1,2	0.041	0.437
GPS vs manual stride, Fitbit vs Fitbit	1,2	-0.301	0.880

DISCUSSION: The Fitbit® Charge 2™ was found to be poor in reliability and validity in tracking step count. The strength of reliability in all but one condition was poor. All ICC single measure values were .721 or lower. The strength of validity was poor for the Fitbit® Charge 2™ where both ICC single measure values were .247 or lower. The results of this study are in alignment with other research regarding Fitbit products. Physical activity trackers are most reliable when walking on a treadmill and at low-intensity (Alina, et al., 2017). The Fitbit® Zip™ produced an ICC of .92 while the Fitbit Flex demonstrated an ICC of .81 during a test-retest of activity trackers (Kooiman, Dontje, Sprenger, Krijnen, van der Schans, & de Groot, 2015). The Fitbit® Zip™ has been shown to be reliable and valid to record preschool children's step count in a childcare setting (Sharp, Mackintosh, Erjavec, Pascoe, & Horne, 2017) as well as for persons with stroke, however, slower walking speeds were associated with greater undercounting of steps (Schaffer, Holzapfel, Fulk, & Bosch, 2017).

Lack of reliability and validity of the Fitbit® Charge 2™ could be the result of several factors. One possible factor is the placement of the device on the wrist, as this is not an optimal location to count steps compared to other placements (i.e., hip pedometers). Excellent reliability and validity were found for measurements taken from accelerometers mounted at the waist and shank during running and at the thigh and shank during bicycling (Gatti, Stratford, Brenneman, & Maly, 2016). Chow et al., (2017) determined waist-worn devices achieved better accuracy than those on the wrist across all gait speeds. Moreover, the device is likely not optimally situated to effectively pick up stride changes. Studies comparing wrist vs. hip worn step counters demonstrate a lack of accuracy with wrist worn trackers compared to either hip or ankle (Tudor-Locke, et al., 2015; Simpson, et al., 2015). Further, movement differences of the dominant vs. non-dominant hand may also influence counts on both treadmills and free-living environment (Chen, et al., 2016;

CONCLUSION: The reliability and validity of Fitbit™ devices to count steps is important for consumers who are investing in these products. The general population and individuals who are managing their weight may be using the Fitbit® Charge 2™ for tracking steps. Overestimation of steps by the device could result in early termination of exercise and loss in potential benefits. Future research should be done to improve Fitbit™ technology. Other wearable forms of devices could also be used as an alternative to count steps.

REFERENCES:

Alinia, P., Cain, C., Fallahzadeh, R., Shahrokni, A., Cook, D., & Ghasemzadeh, H. (2017). How Accurate Is Your Activity Tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities. *JMIR mHealth and uHealth*, *5*(8).

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*. 13(2).

Chow, J. J., Thom, J. M., Wewege, M. A., Ward, R. E., & Parmenter, B. J. (2017). Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. *Gait & Posture*, *57*, 199-203.

Farina, N., & Lowry, R. G. (2018). The validity of consumer-level activity monitors in healthy older adults in free-living conditions, *Journal of Aging and Physical Activity*, *26*(1), 128-135.

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

Gatti, A. A., Stratford, P. W., Brenneman, E. C., & Maly, M. R. (2016). GT3X+ accelerometer placement affects the reliability of step-counts measured during running and pedal-revolution counts measured during bicycling. *Journal of Sports Sciences*, *34*(12), 1168-1175.

Kooiman, T. J., Dontje, M. L., Sprenger, S. R., Krijnen, W. P., van der Schans, C. P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation. 7*(1), 24.

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).

Schaffer, S. D., Holzapfel, S. D., Fulk, G., & Bosch, P. R. (2017). Step count accuracy and reliability of two activity tracking devices in people after stroke. *Physiotherapy Theory and Practice*, 33(10), 788-796.

Sharp, C. A., Mackintosh, K. A., Erjavec, M., Pascoe, D. M., & Horne, P. J. (2017). Validity and reliability of the Fitbit Zip as a measure of preschool children's step count. *BMJ Open Sport & Exercise Medicine*, 3(1), e000272.

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

Smith, M., Powers, M., Henderson, O., & Bodman, T. (2017). Accuracy and Reliability of the FitBit ChargeTM Activity Tracker Among Older Adults. *Journal of the Oklahoma Association for Health, Physical Education, Recreation, and Dance*, 54(3), 29-37.

Sushames, A., Edwards, A., Thompson, F., McDermott, R., & Gebel, K. (2016). Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PloS one*, 11(9), e0161224.

Tophøj, K. H., Petersen, M. G., Sæbye, C., Baad-Hansen, T., & Wagner, S. (2018). Validity and Reliability Evaluation of Four Commercial Activity Trackers' Step Counting Performance. *Telemedicine and e-Health*.

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.

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.

Tudor-Locke, C., Barreira, T.V., & Schuna,J.M. (2015). Comparison of Step outputs for waist and wrist accelerometer attachment sites. Medicine and Science in Sports and Exercise, 47(7), 834-842.

Simpson, L.A., Eng, J.J., Klassen, T.D., Lim, S.B., Louie, D.R., Parapilly, B., Sakakibara, B.M, & Zbogar, D. (2015). Capturing Step Counts at Slow Walking Speeds in Older Adults: Comparison of Ankle and Waist Placement of Measuring Device. *Journal of Rehabilitation Medicine*, 47, 830-835.

Chen, M. Kuo, C. Pelligrini, C. & Hsu, M. (2016). Accuracy of Wristband Activity Monitors during Ambulation and Activities. Medicine and Science in Sports and Exercise, 48(10), 1942-1949.