Accuracy and Reliability of the FitBit Charge™

Activity Tracker Among Older Adults

Michael Smith, M.S.
University of Central Oklahoma

Melissa Powers, Ph.D.
University of Central Oklahoma

Olivia Henderson, M.S.
University of Central Oklahoma

Theresa Bodman
University of Central Oklahoma

Abstract

Wearable technology, including activity trackers, remains a top fitness trend world-wide; however, the ability of individual activity trackers to accurately and consistently record physical activity remains unknown. The purpose of this study was to evaluate the accuracy and reliability of step counts from the Fitbit Charge™ among older adults. Thirteen participants with a mean age of 70.38 ±4.27 years volunteered to participate and completed all testing. Participants completed a 96 meter walk around a gymnasium while wearing the Fitbit Charge™ activity tracker. Step counts on the activity tracker were recorded before and after the walk to determine the step count. Each walk was also video recorded to determine the actual number of steps taken during the walk. Step counts from video analysis were confirmed by two researchers. The same procedures were repeated on a second, non-consecutive day of testing to determine the reliability of the activity tracker. Accuracy of the activity tracker was determined by comparing step counts reported by the Fitbit Charge™ to the observed step counts from each testing session. The reliability of the activity tracker was determined by correlation analysis and comparison of step counts from the first testing session to the second testing session. The activity tracker significantly underestimated observed steps at both testing sessions by 21.31 steps and 22.62 steps, respectively ($p < .05$). No difference in step count from the activity tracker was seen from session one to session two (151.85 steps vs. 152.54 steps, $p > .05$); however, the correlation between the two sessions was only moderate, $r = .55$, $p = .05$. Among older adults, the Fitbit Charge™ appears to underestimate steps taken even over a short distance. The reliability of the Fitbit Charge™ is questionable given only a moderate correlation between sessions. While preliminary, these results call into question the accuracy and reliability of daily step counts from the Fitbit Charge™. It is suggested that all new activity trackers to hit the market are given careful study to determine their ability to accurately measure daily activity.

Oklahoma AHPERD Vol 54, Issue 3 2017
Accuracy and Reliability of the FitBit Charge™

Introduction

The American College of Sports Medicine (ACSM) determined that wearable technology was the number one fitness trend for 2016 and 2017 (Thompson, 2015; Thompson, 2016). Tudor-Locke (2010) discussed that, along with public health recommendations for physical activity, an additional 3,000 to 4,000 steps per day may result in reduced body mass index and improved cardiovascular health. Activity trackers may help users with an increase in daily step counts (Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015). A wide variety of devices are available on today’s consumer marketplace; however, despite the popularity and prevalence of product development, little information is available on the accuracy or reliability of the devices produced. Lee and Finkelstein (2014) found that among six devices tested, only one of the device manufacturers (BodyMedia FIT) provided information on the accuracy of the metrics captured. BodyMedia has since been purchased by another company who has not provided information on the accuracy of its devices (Velazco, 2013). Given that devices are marketed to consumers with the promise of positive effects on health and health behavior (Lee & Finkelstein, 2014), more information on the validity and reliability of fitness tracking products should be readily available.

An increasing number of studies are being performed on consumer device accuracy. Stackpool, Porcari, Mikat, Gillette, and Foster (2014) tested five devices and found that no device tested differed by more than 10% for total steps during walking and running activities. An average of 4% underestimation was found among devices during running or walking which is under an acceptable 5% difference; however, under conditions where ambulation was abnormal, an average 18% difference was found (Stackpool et al., 2014). Among older adults, gait pattern differences and walking speed have been found to affect the accuracy of activity trackers (Korpan, Schafer, Wilson, & Webber (2015). Stackpool et al. (2014) found the Fitbit Ultra to be accurate during treadmill walking and elliptical exercise; however, there was a significant step count underestimation of 6% and 20% for running and agility-related testing.
respectively. Guo, Li, Kankanhalli, and Brown (2013) found the Fitbit Ultra One, Nike Fuelband, and iPhone Moves app to have 1.05% ± 2.26%, 7.79% ± 9.17%, and 27.28% ± 29.97% error respectively. The Fitbit Ultra One is a hip-worn device, the iPhone Moves app is an application run from a cellular phone, and the Nike Fuelband is a wrist-worn device. Variation in the error rates from hip-worn, wrist-worn, and cellular phone based activity trackers may indicate that device location is a factor in the accuracy of measurements. Most devices currently available to consumers are wrist worn. Chen, Kuo, Pellegrini, and Hsu (2016) found the Fitbit Flex, a wrist-worn device, to be more accurate for walking and running speeds ≥ 80 m•min⁻¹ and less accurate for walking speeds ≤ 54 m•min⁻¹ when compared to objectively measured step counts. The Fitbit Flex was found to underestimate steps by 6% at a treadmill walking speed of 54 m•min⁻¹ (Chen et al., 2016). Analysis revealed an underestimation in step counts for slower walking speeds, activities with restricted arm movements, and an overestimation in step counts for non-ambulatory activities such as folding laundry and playing a tablet game (Chen et al., 2016).

The aim of this study is to evaluate the accuracy and reliability of the Fitbit Charge wrist-worn activity tracker among older adults. There are few studies that evaluate Fitbit wrist worn devices for accuracy and reliability. Further, there are few studies that evaluate the Fitbit Charge among an older adult population specifically.

Methodology

Participants

Participants for this study were recruited from group fitness classes specifically designed for older adults. Eligibility criteria were that the participants were over 60 years of age and able to walk independently. Participants were not otherwise excluded from participation. Each participant was measured for height and weight, and asked their age.

Procedure

A Fitbit Charge was configured for each individual by setting the sex, age, height, and weight into the application. The device was then synchronized and placed on the participant’s preferred wrist.
Accuracy and Reliability of the FitBit Charge™

before starting a 96-meter walk. No other devices were on the same wrist as the Fitbit Charge during the
testing procedure. Participants were instructed to follow the baseline of a National College Athletic
Associate (NCAA) regulation basketball court to the midcourt line, across to the opposite baseline, and
back to the starting point following the baseline. This half-court walk was completed twice before being
instructed to stop at the starting point for step data collection. The participants were videoed so that the
researchers could objectively count the number of steps taken during the walk. The same procedure was
completed during the second session a minimum of 24 hours later.

Data Analysis

Step data for both walking sessions were recorded for comparison to objective step counts. Three
researchers reviewed video footage of each participant to objectively count the number of steps taken.
Upon agreement on the number of steps for each video, the objective step counts at each session were
recorded for statistical analysis. Accuracy of the device was determined by comparing step counts from
the device to the objective step count at each session using paired samples $t$ tests. The reliability of the Fit
Bit Charge™ was determined by calculating the Pearson Product Moment Correlation Coefficient for
device steps counts between the two sessions. The researchers also compared device step counts collected
at the first session to those collected at the second testing session using a paired samples $t$ test.

Results

Thirteen participants with a mean age of 70.38 ± 4.27 years volunteered to participate and
completed all testing. Participants had a mean weight of 167.59 ± 37.35 pounds and mean height of 64.72
± 3.11 inches.

The results indicated that step counts recorded by the device were significantly lower than those
observed by video analysis at both sessions ($p < .05$; Table 1). At session 1, observed step counts were
higher by 14.0% or 21.30 steps ($p = .013$). A similar difference of 14.8% or 22.61 steps was seen at
session 2 ($p = .007$). Figure 1 displays the difference in device counted steps and observed steps from
session 1 and session 2.
Table 1

Device vs. Observed Step Counts from Session 1 and 2

<table>
<thead>
<tr>
<th></th>
<th>Device Step Count</th>
<th>Observed Step Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Session 1</td>
<td>151.85</td>
<td>37.18</td>
</tr>
<tr>
<td>Session 2</td>
<td>152.54</td>
<td>31.97</td>
</tr>
</tbody>
</table>

* p < .05

Note. Mean Diff = mean difference

![Bar Chart](image)

Figure 1. Mean number of steps counted by devices compared to observed step counts from sessions 1 and 2.

A significant, moderate correlation was observed between the device step count at session 1 and the device step count at session 2 ($r = .55$, $p = .05$). A mean difference in device step counts between session 1 and session 2 was not found ($p > .05$); however, individual differences between the testing sessions varied widely from 30 more steps recorded at session 1 to 99 more steps recorded at session 2 (range = 129). Variability in observed step counts was not as high ranging from 10 more steps observed at session 1 to 34 more steps observed at session 2 (range = 44).
Accuracy and Reliability of the FitBit Charge™

Discussion

The results of this investigation indicate that step counts recorded by the Fitbit Charge™ underestimate observed step counts by at least 14% among older adults. This is higher than the average of 4% underestimation of 5 devices reported by Stackpool et al. (2014) and 6% underestimation of the Fitbit Flex by Chen et al. (2016). In fact, Stackpool and colleagues found no device to underestimate steps by more than 10% except under abnormal conditions; however, Chen et al. (2016) found a 35% underestimation in the Fitbit Flex when participants performed walking activity with restricted arm movements, such as walking with a stroller. Guo et al. (2013) tested walking speeds on a flat 400 m outdoor running track. Conditions of the track and outdoor environment are unknown. Further, there were a small number of participants who completed multiple laps for step count comparisons rather than multiple participants over a fixed time and distance. In the current study, participants walked on a flat gymnasium floor with no obstacles or distractions (i.e. normal conditions). In all studies reviewed, participants were between the ages of 18 and 45 years, which misses a large segment of the total population. Participants in the current study were over age 60 years.

Step counts recorded by the device at two sessions were not different; however, wide variability was recorded in step counts from the device at each session. Steps counts at the first session correlated only moderately with step counts recorded at the second session. These low correlations are likely due to wide variability in steps recorded in session 1 and session 2 (as much as 99 steps different). Sushames, Edwards, Thompson, Mcdermott, and Gebel (2016) found that the Fitbit Flex worn on the wrist had weak to moderate test-retest reliability when comparing step counts during varied physical activities such as walking, walking on an incline, jogging, and free living activities. Analysis revealed similar results to previous studies in that slower walking speeds resulted in underestimated step counts (Sushames et al., 2016). Further, the results from the Fitbit Flex were inconsistent among testing periods and activities (Sushames et al., 2016).
Accuracy and Reliability of the FitBit Charge™

In conclusion, the Fitbit Charge™ appears to underestimate steps taken by older adults even over a short distance. The reliability of the Fitbit Charge™ is questionable given only a moderate correlation between sessions and wide variability from one session to another. While preliminary, these results call into question the accuracy and reliability of daily step counts from the Fitbit Charge™ particularly among older adults. Because new activity trackers hit the market each day and are often replaced before their accuracy and reliability can be carefully studied, it is recommended that users interpret step counts with caution. Daily monitoring of step counts may motivate individuals to engage in physical activity, but widely varying step counts could also foster frustration and disappointment.
Accuracy and Reliability of the FitBit Charge™

References


*Oklahoma AHPERD Vol 54, Issue 3 2017*
Accuracy and Reliability of the FitBit Charge™

good health and how confident are we in this number? Current Cardiovascular Risk Reports, 4, 271–276. doi:10.1007/s12170-010-0109-5.