Observing Step Accuracy with the Fitbit and iPhone Health Application

WearableTech
Rochester Institute of Technology
Rochester, NY
hci@rit.edu

ABSTRACT

Fitness and health tracking wearable devices are becoming more popular and innovative. Many Smartphones are beginning to have these capabilities as well. In most devices, there is a step count function, which measures the amount of steps a user takes. In this paper, the accuracy of steps was evaluated with the Fitbit wearable device and iPhone 6's built-in "Health" application. To answer the hypotheses, *There is no difference between Actual Steps and Fitbit Recorded Steps* and *There is no difference between Actual Steps and iPhone Recorded Steps,* an experiment was conducted with 10 participants testing both devices. Based on the results, it was observed that the iPhone was more accurate, however, the Fitbit was more consistent and precise.

1. INTRODUCTION

Health and fitness tracking devices allow users' to easily track their vitals and movements. Most wearable fitness tracking devices are worn on the users' wrists, while smartphone devices are commonly placed in users' pockets or in their bags. Both the Fitbit and the iPhone 6 and 6S have a variety of ways to track different health and fitness information. According to previous research, 70% of surveyed participants believed the Fitbit wearable device reports data accurately [1]. However, Ledger and McCaffrey [3] found that about one-third of all Americans who purchased the wearable devices for have stopped using it after just 6 months of use. This leads to the conclusions that either: the wearables have slowly failed to meet the expectations of their users (by reporting inaccurate data) or that users have replaced wearables with Smartphones (which provide similar fitness and health tracking functionality) [3]. This study aims to answer the two hypotheses: *There is no difference between actual steps and Fitbit recorded steps* and *There is no difference between actual steps and iPhone recorded steps*. Given the short amount of time, an experiment was conducted testing the two devices (the Fitbit One and iPhone 6) using ten participants.

2. METHODS

The "within-subject design method" was used for the experiment, which allowed testing of participants while wearing both devices (Fitbit on the wrist and iPhone in the pocket). This way each participant would test both devices simultaneously, and thus reduced the number of random biases.

This experiment was conducted in order to prove or disprove the hypotheses: *There is no difference between Actual Steps and Fitbit Recorded Steps* and *There is no difference between Actual Steps and iPhone Recorded Steps*. The independent variable was the actual number of steps the participant took and had two levels. The dependent variables were the Fitbit and iPhone recorded steps.

Prior to conducting the experiment, a path of 200 steps was mapped for the participants to walk. The path did not have any stairs or elevation increases. We used 200 steps to ensure step count readings would be apparent. All 10 participants were voluntarily chosen from the *HCIN 600: Human-Computer Interaction Research Methods* course at RIT. Each participant was tested individually. To keep things consistent, the Fitbit was placed on the right hand of each participant and the iPhone in the right back pocket. The number of steps on the Fitbit wearable device and the iPhone Health application was unable to be reset, so the initial step reading before the participant began walking was recorded and then subtracted by the number that the device ended on. Due to the sensitivity of both devices, the participant was asked not to move when the initial and end step count was recorded. To avoid participant count error, the same researcher walked alongside the participant and counted the participant's steps. One set of data was unusable because the Fitbit wearable device fell off the user's wrist and might have skewed the step reading.

3. RESULTS

Participants were 7 men and 3 women aged 23 to 56 years (men: M = 25.9, SD = 2.12; women: M = 34.3, SD = 18.77). Two 2-tailed paired t-tests were conducted from the data in order to measure the difference between Actual Steps vs. Fitbit Recorded Steps and Actual Steps vs. iPhone Recorded Steps. A 0.05 significance level was used. Figure 1 is a line graph with plots of our recorded data.

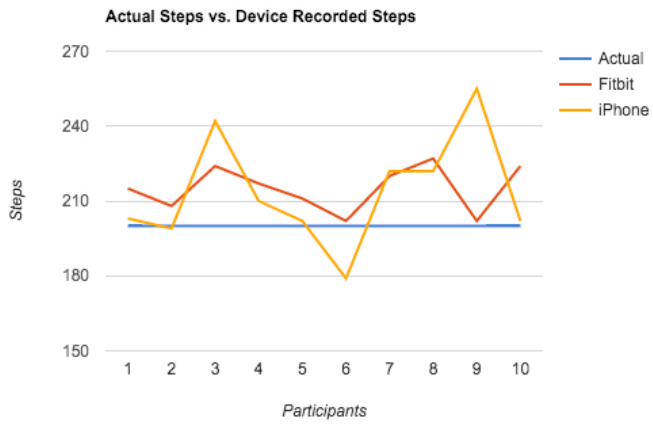

Figure 1. Line Graph of Actual Steps, Fitbit Recorded Steps, and iPhone Health Application Recorded Steps

3.1 Fitbit Recorded Steps

The first test indicated a significant difference between Fitbit Recorded Steps **(M = 215, SD = 9.06)** (Table 1) and Actual Steps **(M = 200, SD = 0), t(10) = 5.24, p = .0005, d= 2.34** (Tables 2, 3)**.** The sample statistics can be found in Table 1. The sample correlations can be found in Table 2. The sample test can be found in Table 3.

Table 1. Paired Samples Statistics of the Fitbit Recorded Steps and Actual Steps

	Mean	N	S.D.	Std. Error Mean
Fitbit	215	10	9.06	2.86
Actual	200	10	0	0

Table 2. Paired Sample Correlations of the Fitbit Recorded Steps and Actual Steps

	N	Correlation	Sig
Fitbit & Actual	10	*0.76*	*0.0005*

Table 3. Paired Samples Test of the Fitbit Recorded Steps and Actual Steps

Mean	S.D.	Std. Error mean	Lower	Upper	t	df	Sig (2 tailed)
15	9.06	2.87	8.52	21.48	*5.24*	9	*0.0005*

3.2 iPhone Health Application Recorded Steps

The second test did not indicate a significant difference between iPhone Recorded Steps **(M = 213.6, SD = 22.28)** (Table 4) and Actual Steps **(M = 200, SD = 0), t(10) = 1.93, p = 0.086, d= 0.86** (Tables 5, 6)**.** The sample statistics can be found in Table 4. The sample correlations can be found in Table 5. The sample test can be found in Table 6.

Table 4. Paired Samples Statistics of the iPhone Recorded Steps and Actual Steps

	Mean	N	S.D.	Std. Error Mean
iPhone	213.6	10	22.28	7.05
Actual	200	10	0	0

Table 5. Paired Sample Correlations of the iPhone Recorded Steps and Actual Steps

	N	Correlation	Sig
Fitbit & Actual	10	*0.4*	*0.086*

Table 6. Paired Samples Test of the iPhone Recorded Steps and Actual Steps

Mean	S.D.	Std. error mean	Lower	Upper	t	df	Sig (2 tailed)
13.6	22.28	7.05	-2.34	29.54	*1.93*	9	*0.086*

4. DISCUSSION

This experiment was conducted in order to answer the two hypotheses: *There is no difference between actual steps and Fitbit recorded steps* and *There is no difference between actual steps and iPhone recorded steps*. According to the data analysis, there is a significant difference between Fitbit Recorded Steps and Actual Steps, however, there is not a significant difference between iPhone Recorded Steps and Actual Steps.

According to the t-tests, the mean of the iPhone Recorded Steps was closer to the Actual Steps than the Fitbit Recorded Steps, which made the iPhone a more accurate device in this experiment. The iPhone step count, however, had a much higher standard deviation of 22.28 (Table 4) than the Fitbit's (which was 9.06) (Table 1). The Fitbit was the more precise device in this experiment because it had less variant data (which is evident in Figure 1).

There are several factors that could have contributed to this result. The Fitbit comes in several different models, which all have different capabilities. The Fitbit One was used in the experiment, which has a 3-axis accelerometer, altimeter, and vibration motor [2]. The accelerometer tracks movement in all directions and then is converted into data. "By analyzing acceleration data, our trackers provide detailed information about frequency, duration, intensity, and patterns of movement to determine your steps taken, distance traveled, calories burned, and sleep quality" [6]. The altimeter measures altitude, which is useful for measuring floors climbed. The iPhone 6, however, has a gyroscope and GPS in addition to the 3-axis accelerometer. The gyroscope measures orientation and rotation, the GPS measures distance [7]. This data is processed in the iPhone's m7 processing chip [8].

What differentiates fitness trackers from one another are their algorithms. Since the Fitbit and iPhone are worn and used in different ways, their algorithms will account for this. The Fitbit uses a threshold to indicate if the user is walking. The following quote is from the Fitbit website describing in detail how the threshold functions:

> *"If a motion and its subsequent acceleration measurement data meet the threshold, the motion will be counted as a step. If that threshold is not met, the algorithm won't count the motion as a step. Other factors can create enough acceleration to meet our threshold and therefore cause some over counting of steps, such as riding on a bumpy road. Equally, it's possible for the algorithm to undercount (not meet the required acceleration threshold"[7].*

Furthermore, according to the previous Fitbit survey a participant responded, "Fitbit does not register all step(s). Example if you are pushing a grocery cart or a baby carriage it does not measure the steps accurately. Your arm needs to swing" *(Quote 1, P10)* [1]. This is an example how the Fitbit accuracy may be limited and may have effected the experiment.

The iPhone 6 is secretive about their step counting algorithm for a variety of reasons. The following quote was from an "Apple Certified Associate for Mac Integration"[8].

> *"When you walk, the sensor suite collects datapoints based on how the device is moving by its position in space, and the velocity it senses as your body moves. Apple built a walking data profile based on inputs from xNumber of people in the Health R&D stage. this profile is used to compare the current datapoints collected by the handset… If your current datapoint profile matches the walking profile, you are walking. Each step can be counted by isolating the shock and velocity that is gained and then lost at each step gives a +1 step to the counter"[8].*

From these quotes, it is evident that the algorithms for both the Fitbit and iPhone function very differently, which contributes to the discrepancy in accuracy.

Limitations

The experiment had several limitations and confounding factors. The most recent Fitbit model, the Fitbit Surge, has a GPS and is therefore more comparable to the iPhone 6. Since the Fitbit One does not have GPS, this might haven given the iPhone an advantage.

Due to the short amount of time and the limited access to the Fitbit wearable device, there might have been a sampling error, which occurs when there is a small amount of participants. When there is a larger sample, generally there is also less variability.

Furthermore, since both devices did not have a reset data function, this could have caused a random error. Both the devices update steps by groups and do not load individual steps. There might have been a loading problem and did not report the correct step count.

5. CONCLUSIONS AND FUTURE WORK

The data gathered from our experiment presents a mixed result. Statically, iPhone's step counter is more accurate than the Fitbit's, however, both devices have the tendency of counting more steps than people actually walk, and the tracked data varies a lot. The result is unexpected based on the former study, since Fitbit users expressed satisfaction about the step tracker [1]. Although our experiment is limited by factors, is does indicate that research can be done to understand the operating mechanism, behavior and limitation about nowadays activity monitoring smart devices. In the future, gathering a larger sample size can be helpful to improve the quality of the study. It would be interesting in the future to investigate the reasons behind the exaggerated step count, and whether the algorithms between different companies affect the user's experience of health tracking devices.

REFERENCES

[1] Keller, Zhao, Kannekanti. Observing Lifestyle Changes with Fitbit Users. Rochester Institute of Technology. October 2015.

[2] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.* 12, 2 (March 2011), 74-82. DOI=http://dx.doi.org/10.1145/1964897.1964918

[3] Ledger, D., and McCaffrey, D. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. Endeavour Partners, LLC 93, 1 (2014), 36–45.

[4] Rayoung Yang, Eunice Shin, Mark W. Newman, and Mark S. Ackerman. 2015. When fitness trackers don't 'fit': end-user difficulties in the assessment of personal tracking device accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15). ACM, New York, NY, USA, 623-634. DOI=http://dx.doi.org/10.1145/2750858.2804269

[5] Amit Pande, Yunze Zeng, Aveek K. Das, Prasant Mohapatra, Sheridan Miyamoto, Edmund Seto, Erik K. Henricson, and Jay J. Han. 2013. Energy expenditure estimation with smartphone body sensors. In *Proceedings of the 8th International Conference on Body Area Networks* (BodyNets '13). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, 8-14. DOI=http://dx.doi.org/10.4108/icst.bodynets.2013.253699

[6] How it works: We explain how your fitness tracker measures your daily steps: 2015. *http://www.wareable.com/fitness-trackers/how-your-fitness-tracker-works-1449*. Accessed: 2015- 11- 18.

[7] How does my tracker count steps?: 2015. *http://help.fitbit.com/articles/en_US/Help_article/How-does-my-tracker-count-steps*. Accessed: 2015- 11- 18.

[8] How does iPhone's step counter work? - Quora: 2015. *https://www.quora.com/How-does-iPhones-step-counter-work*. Accessed: 2015- 11- 18.