5-2017

Step Counter Accuracy in Free-Living Environment Over a 24-Hour Period

Janelle Dedic
jdedic@vols.utk.edu

Step Counter Accuracy in Free-Living Environment Over a 24-Hour Period

Janelle Dedic

Chancellor's Honors Thesis

David R. Bassett, Jr., PhD.

May 8, 2017

Step Counter Accuracy in Free-Living Environment Over a 24-Hour Period

Janelle Dedic

Kinesiology Undergraduate Program, University of Tennessee, Knoxville

Introduction

As the demand for fitness trackers increases and the movement towards using exercise as a prescription in medicine gets closer to being put into action, the need for regulations on what it means to take a step is important, especially concerning step counters. Coupled with those regulations, the accuracy of step counters is of the utmost importance if the data they provide is going to be used as a guideline for health. In order to quantify daily physical activity, pedometers are used because they can be used across people of all ages, body types, and fitness levels. The accuracy of the amount of physical activity being exhibited based on each person's step count begins with a pedometer's reading.

A study done by Schneider et al. (2003) made the process of determining step-counter accuracy simple by having participants walk a set distance of 400m over ground while wearing different pedometers. This was easiest for monitoring distance, but also for testing accuracy because of the researchers present counting every step by hand. Overall this helped establish how to set up a baseline measurement for testing accuracy by using a set distance. This became more common as treadmills were used as baseline measurements in laboratory settings. For instance, Melanson et al. (2004) used a treadmill as a way to control speed while assessing step-counter accuracy and energy expenditure.

Foster et al. (2005) developed a study that broadened the scope of testing accuracy of pedometers. The StepWatch was assessed in both a treadmill setting and in a closed, set course in a hallway for a participant to travel. While the movements were not complex, adding the element of testing over ground walking was important for establishing reliability in a more lifelike environment.

The most important task when testing pedometers and fitness trackers is ensuring the reliability in a free-living, every day environment. Hickey et al. (2016) kept the treadmill as the comparison method, but added in simulated free-living activities such as cleaning a room, for example. These activities were more realistic for what an everyday person would perform on a daily basis. Common activities such as golfing, indoor housework, and outdoor yardwork were tested in a study done by Hendelman et al. (2000) as well, further broadening the areas of activity that need to be investigated.

While testing free-living environment step counts against treadmill step counts is practical in its use and easy to replicate, an issue arises when the step counts for the free-living environment are only based on the data from the step counter. Since the activities in a free-living environment are completely different from simply walking on the treadmill, it is difficult to equate the two step counts. With no way of reviewing the steps taken during the study, a level of relatability to real life is lost and there is a risk of losing the data recorded in a one-time trial, such as performing household chores. Testing children in a free-living environment after completing a treadmill protocol, Rosenkranz et al. (2011) implemented a video recording of the treadmill protocol in order to verify the steps taken and analyze the gait in terms of strides during an allotted amount of time. Even though the video was of the treadmill, a sense of reliability was rooted in the data obtained for the step counts, which were the basis of comparison.

The issue still persisting when measuring free-living environment activities is validity of steps counted in the environment. Using a researcher to count the steps by hand in person while watching the activity is reliable, but there is a higher reliability on the hand count of the researcher being correct considering the count is only done once. The best way to prevent the risk of this happening is recording the free-living environment activity on video so that the opportunity for multiple counts of steps by hand exists. Grant et al. (2008) and Dijkstra et al. (2008) included video recordings of free-living environment activities for older adults and patients with Parkinson's disease, respectively. The key parts about these studies were the presence of miscellaneous activities done in the free-living environment and the recording of those activities on video. The data obtained from the studies are some of the most accurate processes as far as obtaining accuracy from true free-living environment activities. Granted these studies still employed the use of a treadmill as a baseline measurement, the accuracy of multiple step counters could be tested in a more open environment. However, the lack of complete 'freedom' in the free-living environment was still missing to authenticate the step counts. The free-living activities performed were set up by the study to be done. A true test of free-living would include a broad range of activities being performed in a short amount of time outside of a laboratory setting.

What testing the reliability of pedometers and step counters depends on is the criterion of comparison. In order to provide the best results, the count that is being used as a baseline, accurate measure needs to be accurate in itself. Previous studies that used a treadmill as a method of comparison cannot be extremely valid because the actions performed in each setting (i.e. the treadmill versus free-living) are different. Treadmills involve simple, straight movements while free-living environments most often include complex movements not done in the same plane of motion. A different approach to counting was done by Feito et al. (2012), Kooiman et al. (2015), and Tully et al. (2014), where the basis of comparison in each of the studies was a specific step counter. Each step counter used as comparison was researched and authenticated on reliability before determining if it would be used. However, the accuracy of the criterion step counters remains unknown. Comparisons to criterion devices describe differences between devices but do not directly assess the true step counts for a 24-hr period.

To maximize the opportunity for accurate step counts while also reducing the risks of inaccurate or unknown total daily step count, this study included video monitoring participants across all waking hours of one day while they also wore several step counting devices. Researchers then counted all steps captured in the videos and compared the total step counts to the step count estimates from each device. By incorporating video recordings of activities performed in the free-living environment, steps can be verified multiple times by researchers performing hand counts. Using these hand counts as the basis for comparison for multiple step counting devices, the risk of comparing step counting data to a criterion that is not accurate within itself is also removed.

Methods

Activity Monitors. Each participant wore a total of 14 step counters for the duration of the study. The locations of these devices included the wrists, hips, thighs, and ankles. The step-counters used were commercially available pedometers and accelerometers. These included the

ActiGraph GT9X, ActivPAL, Yamax Digi-Walker SW-200, Fitbit Zip, Fitbit Charge, New Lifestyles NL-2000, and the Step Watch.

The devices were placed on each participant according to the recommendations given by the manufacturers. The ActiGraph GT9X and Fitbit Charge were worn on both the dominant and non-dominant wrist to account for any differences that may appear due to increased movement of one wrist versus the other. The Fitbit Zip, DigiWalker, New Lifestyles, and ActiGraph GT9X were all worn on the hip in randomized order in four locations. The left and right lateral devices were worn in line with the respective anterior axillary fold and the medial devices were just medial to the lateral devices. The ActivPAL was attached to the midline of each thigh and was secured using a Tegaderm adhesive patch. Lastly, the Step Watch was attached to the ankle using a Velcro strap.

Protocol.

Each participant wore the 14 step-counters and a GoPro camera for all waking hours of one day. Initial recordings of steps on each device that displayed a step count was recorded by the participant immediately after placing them on the body. The participant also recorded step counts directly before taking off the devices before going to bed. In order to verify the ending step count total, a GoPro camera was worn by the participant for the day of the study. The camera was worn on the chest and angled towards the feet so that the video included the participant’s feet and other close surroundings. Video recording was taken at all waking hours of the day with the exception of the participant changing clothes or using the restroom. In these cases, a covering was placed over the camera to block the screen. Sound was not recorded with the videos to preserve the privacy of each participant. At the end of the day of the study, the total step counts were recorded for each step counter.

The videos from each participant were extracted from the GoPro and divided into 10-minute segments. This was done to help obtain a more accurate step-count by hand and to locate errors in the step-counters themselves after evaluating the end step-counts. For each video, two different research assistants watched and hand counted steps. In order to maximize accuracy, if the data reported by the assistants differed by more than 5% or by more than 6 steps, a third researcher watched the video segment and reported their counts. When two separate counts fell within the ranges listed above, the counts were averaged and recorded as the total step count for that participant.

Statistical Analysis.

One-sample t-tests with Bonferroni adjustment were used to compare each device's count with the hand counts. Significance was set at 0.05 and data was analyzed using SPSS Version 24 for Windows (SPSS Inc., Chicago, IL).

Results

The data from four participants in this study was examined after each had worn 20 step counting devices of varying location for the waking hours of a 24 hour period. Step totals from each device were reported and compared to a hand count taken by researchers to determine accuracy. Table 2 presents the average differences from the official hand count, a 95% confidence interval of the difference found, standard deviation, standard error from the mean, and each device's significance after performing one-sample t-tests.

While the ActiGraph GT9X LFE on the hip did not have the lowest difference from the mean, it did have the lowest standard error mean (1.1) and one of the lowest P values (0.001).

Out of the 7 devices with the lowest standard mean error, 4 of them were located on the ankle (StepWatch Quick Stepping, Both Extremes, Default, & QS and Dynamic/Fidgety) and the other 3 were located on the hip (ActiGraph GT9X LFE, MAVM, & FitBit Zip). The devices worn on the wrists, dominant and non-dominant, tended to have higher standard error means. Figure 1 illustrates the variance in mean differences while showing the increasing standard error of the mean. Devices such as the StepWatch Quick Stepping and StepWatch Default had very low mean difference values, but did not show significant P values after a one-sample t-test.

In terms of P values, there was more variation in the location of devices that had the lowest values. There were 10 devices that had significant values ($P < 0.05$), presented in Table 1. Of those 10 devices, 6 of them were ActiGraphs with either a LFE or No-LFE setting on the hip and both dominant and non-dominant wrists.

Placement	Device	Mean	Std. Dev	Mean Difference	Significance (2-tailed)
Ankle	StepWatch QS & Dynamic/Fidgety	13181.5	4.9	-4.5	0.161
	StepWatch Both Extremes	13541.0	3.8	-2.5	0.288
	StepWatch Quick Stepping	13802.5	2.3	-1.2	0.373
	StepWatch Default	13972.5	4.4	0.5	0.85
Hip	*ActiGraph GT9X LFE Hip	16507.5	2.3	17.6	0.001
	*ActiGraph GT9X MAVM Hip	10733.8	3.0	-23.6	0.001
	*FitBit Zip	11300.3	4.5	-18.7	0.004
	*ActiGraph GT9X No LFE Hip	10495.3	7.4	-24.2	0.007
	New Lifestyles	11755.7	43.7	-34.5	0.213
	Yamax DigiWalker	10549.8	29.8	-22.9	0.222
Thigh	*ActivPAL Left	11118.5	5.7	-18.7	0.007
	*ActivPAL Right	10943.5	6.5	-19.5	0.009
Wrist	*ActiGraph GT9X LFE non-dominant wrist	21202.3	21.8	58.0	0.013
	*ActiGraph GT9X LFE Dominant Wrist	23741.0	37.7	83.0	0.022
	*Fitibt Charge non-dominant wrist	10902.3	11.6	-20.2	0.040
	*ActiGraph GT9X No LFE non-dominant wrist	12083.3	7.5	-12.1	0.048
	ActiGraph GT9X MAVM non-dominant wrist	11082.3	14.1	-21.7	0.054
	FitBit Charge dominant wrist	11825.8	11.6	-14.3	0.091
	ActiGraph GT9X MAVM dominant wrist	12250.3	17.9	-15.4	0.185
	ActiGraph GT9X No LFE dominant wrist	13552.3	14.4	-0.5	0.951

Table 1: Data for each device is presented according to placement on the body. Significant values are identified as those with a P < 0.05 and are denoted with (*).

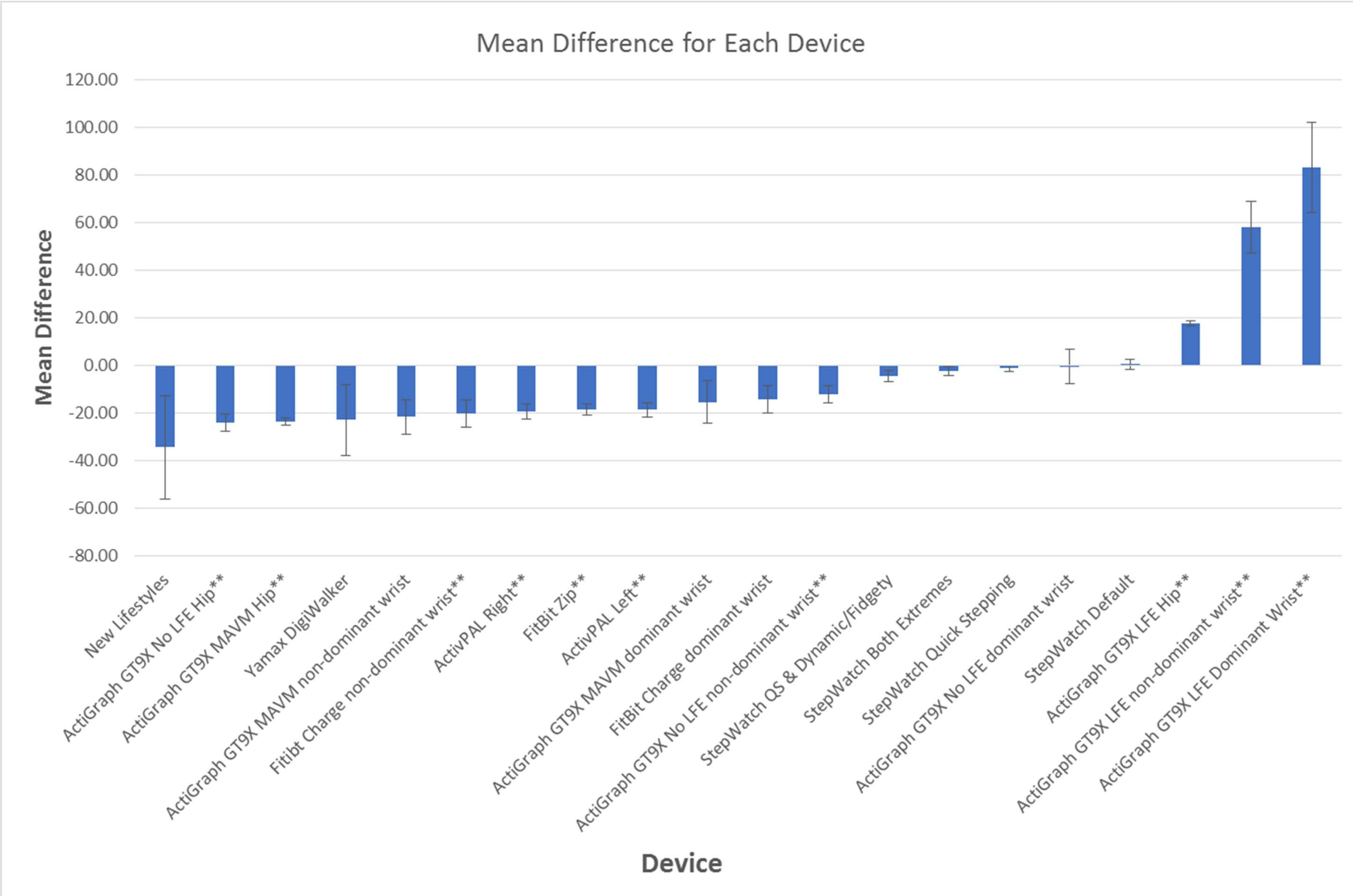

Figure 1: Mean step difference of each device is plotted in ascending order of the mean difference. Those devices with significant values (P < 0.05) are denoted with (**).

Discussion

Commercially sold step counting devices have become a popular and useful tool for measuring physical activity in today's increasingly active society. The accuracy of these step counting devices is an important part of quantifying the level of physical activity being performed for all different populations. As presented in this study, the placement of the device on the body has an influence on step count accuracy. In order to give the most accurate count of steps, these devices need to have high efficiency in counting on a specific place on the body.

After examining the results a pattern as to which part of the body had the most accurate placement for step counting became evident. Out of the seven devices with the lowest mean difference, none of them were on the wrists or upper extremities. A study done by Tudor-Locke et al. (2015) compared devices worn on the waist and wrists to determine which would be most accurate. Their results showed more accurate counts coming from those located on the waist rather than those on the wrists. However, because two different pedometers were being tested, it was difficult to apply these results to all situations for placement. The same situation applied to a study done by Simpson et al. (2015), except devices worn on the ankle and waist were being compared. In this instance, those worn on the ankle were more accurate than those on the waist, but only by a slight margin. Both locations still proved to have accurate readings for steps. Our data supports the findings by Simpson as well, showing that the devices worn on the ankle were some of the most accurate overall, including being more accurate than those worn on the hip.

Although our data is in agreement with these findings, the wrist placement for devices can be examined on its own without being compared to other devices. Differences exist between wrist devices worn on the dominant and non-dominant wrists due to ambulatory activity. The amount of movement on each arm and wrist can have an influence on the step count even if steps are not being taken in the activity. Chen et al. (2016) studied this specific relationship with tests done on treadmills and various daily living activities that would present opportunities for a dominant hand to be used. The results showed that those devices worn on the non-dominant wrist were more accurate than those worn on the dominant one. Those results were conflicting with our data, which showed most of the non-dominant wrist devices having significantly different values compared to the mean than those on the dominant wrist. This information is important for the population wearing these devices considering wrist-worn devices are one of the most popular that are available commercially.

The ActiGraph GT9X devices worn on the body warranted more examination of data because of the settings that can be applied to the device that would influence the step counts. These devices have a “low frequency extension” LFE that increases the device’s sensitivity to smaller, low-intensity movements being performed. Feito et al. (2014) examined the effects of using this setting on the devices in a walking and free-living environment. Overall, the LFE caused the device to overestimate steps taken in the free-living environment compared to the criterion method. These findings were consistent with the data obtained in our study. The ActiGraph with LFE on the hip overestimated the step count, while the ActiGraph without LFE on the hip underestimated those steps. However, the device with the LFE had a lower, in fact the lowest overall, mean difference. There seemed to be a larger difference present when the setting was applied to the ActiGraph devices worn on the dominant and non-dominant wrists. Those with LFE on either of the wrists had significantly different step counts from the mean, demonstrating that using this setting with devices on the wrist does not produce accurate results in comparison with not using the LFE setting. This is most likely due to the increased activity of the wrists that is separate from taking steps, causing there to be more movement detected than is actually being performed. The LFE setting on the ActiGraph, although useful and practical in nature, must be used depending on where the device is worn to give the most accurate results.

Similar to the ActiGraph, the StepWatch device that was mounted on the ankles had various settings that could be applied to increase or decrease the sensitivity of the device depending on the type of activity. Most of these specific devices worn in this study presented low error and were reasonably close to the criterion step count. As seen in Table 1, each setting of the device was close to the mean after performing a significance test. Each participant was performing different exercises and motions while wearing these devices, so the settings had a different influence on each participant's overall accuracy. Had there been more generalizability of activities done, there could have been a better comparison done between the data. But, the data in this study shows that the device is accurate with different settings. This accuracy comes from modifying the settings according to different activities, settings, and intensity of exercises, as shown by Toth et al. (2017). These settings must be modified to the type of activity and fitness level of the individual wearing them in order to obtain the highest accuracy of steps.

Overall, this study had many strengths that were progressive in the field of testing the accuracy of step counting in free-living environments. The use of the GoPro camera mounted on the participants' chests established the validation of steps being taken in this environment that was missing in previous studies. The steps take in this environment were not controlled by a lab protocol, course, or other setting that would require the participant not to use their free will to perform actions. This variability of activity performed by each participant provided a wide range of actions tested for accuracy. Instead of limiting the study to testing only a small set of free-living activities, a vast amount can be tested in the same study. Testing a large amount of devices at once also helped to limit the need for testing the free-living conditions for multiple devices in different studies. Precautions to validate the step counting were also taken by providing multiple counts if the researchers' counts differed too much in one instance. This establishes the reliability of the criterion of comparison, helping to eliminate other outside factors for step counts that may not be accurate.

The limitations of this study included the lack of diversity of participants in special populations. Although there was variability in age, occupation, and fitness level for current participants in the study, there weren't any participants included that had present health conditions or diseases of any sort. These conditions would have an influence on the types of movements performed, especially the speed at which they were performed. Thorup et al. (2017) compared the accuracy of step counters on healthy adults and those with cardiac disease to illustrate this difference. The results showed further testing was required because of the lower step count and lower activity speed of those cardiac patients. The inclusion of older adults or those with chronic diseases that hinder movement would be useful considering the consistent low accuracy of step counters at low walking speeds and low amounts of steps being taken. Feng et al. (2017) tested multiple devices at lower speeds to determine a possible issue, but was unsuccessful and suggested more testing to be done with these lower speeds. Further testing in a study set up similarly to the one conducted with free-living environments and video evidence will need to be done in future research to investigate improving accuracy at these lower step counts and intensities.

Overall, this study improved the efficiency of testing the accuracy of step counting devices. The data presented will be useful in further determining which step counting devices will provide the most accurate results depending on the type of activity being performed by the individual. This will improve the recommendations and general fitness goals for individuals that use step counting devices to measure their physical activity level. The growing popularity of fitness tracking devices collaborates with the increasingly active society we live in today. Establishing reliable and accurate fitness tracking devices will provide other benefits for individuals wearing them such as showing progress, setting goals, and increasing their overall fitness level.

References

1. Schneider, P., Crouter, S.E., Lukajic, O., & Bassett, D.R. (2003). Accuracy and Reliability of 10 Pedometers for Measuring Steps over a 400-m Walk. *Medicine in Science in Sports & Exercise, 35*(10), 1779-1784.

2. Melanson, E.L., Knoll, J.R., Bell, M.L., Donahoo, W.T., Hill. J.O., Nysse, L.J., Lanningham-Foster, L., Peters, J.C., & Levin, J.A. (2004). Commercially available pedometers: considerations for accurate step counting. *Preventive Medicine, 39*, 361-368.

3. Foster, R.C., Lanningham-Foster, L.M., Manohar, C., McCrady, S.K., Nysse, L.J., Kaufman, K.R., Padgett, D.J., & Levine, J.A. (2005). Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Preventive Medicine, 41*(3-4), 778-783.

4. Hickey, A., John, D., Sasaki, J., Mavilia, M., & Freedson, P. (2016). Validity of Activity Monitor Step Detection Is Related to Movement Patterns. *Journal of Physical Activity and Health, 13*, 145-153.

5. Rosenkranz, R.R., Rosenkranz, S.K., & Weber, C. (2011). Validity of the Actical Accelerometer Step-Count Function in Children. *Pediatric Exercise Science, 23*, 355-365.

6. Grant, P.M., Dall, P.M., Mitchell, S.L., & Granat, M.H. (2008). Activity-Monitor Accuracy in Measuring Step Number and Cadence in Community-Dwelling Older Adults. *Journal of Aging and Physical Activity, 16*, 201-214.

7. Dijkstra, B., Zilstra, W., Scherder, E., & Kamsma, Y. (2008). Detection of walking periods and number of steps in older adults and patients with Parkinson's disease: accuracy of a pedometer and an accelerometry-based method. *Age and Ageing, 37*, 436-441.

8. Feito, Y., Bassett, D.R., & Thompson, D.L. (2012). Evaluation of Activity Monitors in Controlled and Free-Living Environments. *Medicine in Science in Sports & Exercise, 44*(4), 733-741.

9. Kooiman, T., 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*(24).

10. Tully, M.A., McBride, C., Heron, L., & Hunter, R.F. (2014). The validation of Fitbit Zip physical activity monitor as a measure of free-living physical activity. *BMC Research Notes, 7*(952).

11. Tudor-Locke, C., Barreira, T.V., & Schuna, J.M. (2015). Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine in Science in Sports Exercise, 47*(7), 839-842.

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

13. Simpson, L.A., Eng, J.J., Klassen, T.D., Lim, S.B., Louie, D.R., Parappilly, 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.

14. Feito, Y., Garner, H.R., & Bassett, D.R. (2014). Evaluation of ActiGraph's Low-Frequency Filter in Laboratory and Free-Living Environments. *Medicine in Science in Sports Exercise, 47*(1), 211-217.

15. Toth, L.P., Bassett, D.R., Crouter, S.E., Overstreet, B.S., LaMunion, S.R., Park, S., Notta, S.N., & Springer, C.M. (2017). StepWatch accuracy during walking, running, and intermittent activities. *Gait & Posture, 52*, 165-170.

16. Feng, Y., Wong, C.K., Janeja, V., Kuber, R., & Mentis, H.M. (2017). Comparison of tri-axial accelerometers step-count accuracy in slow walking conditions. *Gait & Posture, 53*, 11-16.

17. Thorup, C.B., Andreasen, J.J., Sorensen, E.E., Gronkjaer, M., Dinesen, B.I., & Hansen, J. (2017). Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. *BMJ Open, 7*.