

RELIABILITY OF FITNESS TRACKERS AT DIFFERENT PRICES FOR MEASURING STEPS AND HEART RATE: A PILOT STUDY

Kayla Ricchio,^{1, A, B, D} Penny Lyter-Antonneau,^{1, A, D} José M. Palao^{1, A, B, C, D}

¹ Department of Health, Exercise Science and Sport Management. University of Wisconsin – Parkside, Kenosha, United States

^A Study Design; ^B Data Collection; ^C Statistical Analysis; ^D Manuscript Preparation; ^E Funds Collection

Address for correspondence:

José M. Palao

University of Wisconsin – Parkside. SAC D140A. Kenosha, WI 53,144, United States

E-mail: palaojm@gmail.com

Abstract The purpose of this pilot study was to assess the accuracy of steps and heart rate measurement of wrist fitness trackers at different prices. Four healthy college students voluntarily tested three wrist fitness bands and a sports watch (Xiaomi Mi Band, Fitbit Charge HRm, Fitbit Surge, and sports watch Polar M400). Subjects performed two sets of 10 series of 100 steps wearing the fitness trackers on an indoor track in two situations: walking and jogging. In the walking situation, the subjects wore a winter coat and gloves. The variables measured were the number of steps, the heart rate, and the level of error. The steps error percentage for all four devices was lower than 8%. The Fitbit Surge registered significantly more steps in the walking situation ($p < 0.001$). No significant differences were found in the steps measurements in the jogging situation ($p = 0.138$). In the jogging situation, significantly lower values in the heart rate measurements for the Xiaomi Mi Band, Fitbit Charger HR, and Fitbit Surge were found ($p < 0.001$). The results showed that the wearable fitness trackers were relatively accurate for tracking steps (on average, there was a level of error of 2–6%). The assessment of the steps was more accurate in the jogging situation (higher and faster arm swing) than in the walking situation, which involved wearing coats and gloves. The results showed that the wearable fitness trackers that were tested underestimate the heart rate with a level of error of approximately 6–11%. The step error was lower in the walking situation (less mobility of the devices). The price of the devices that were tested did not affect the accuracy of the steps and heart rate assessment. Further studies with a larger sample and more type of devices are needed to confirm these results.

Key words technology, assessment, physical activity, exercise

Introduction

Advances in technology have allowed the general population to easily utilize fitness trackers in their daily life. This type of instrument provides information about our physical activity and exercise. These data can help to establish a baseline of physical activity, monitor one's activities, and provide immediate feedback about it. In a review of the market options, it is possible to find a wide variety of fitness trackers with different characteristics and prices. Most of the fitness trackers allow for monitoring steps, displacement, and heart rate. Their prices range from less

than \$20 to more than \$300. These differences lead to questions about the reliability of the fitness trackers at the different price ranges that are available in the market during the different situations of our daily life. Fitness trackers measure the steps taken and the physical activity carried out using a 3D accelerometer, and they measure the heart rate using LED lights. It is important to keep in mind that these are not research or medical devices (Kroll, Boyd, Maslove, 2016). They are instruments developed and orientated for the general population. These devices are used by the general population to obtain feedback about their physical activity and exercise. However, these devices need to be accurate to properly guide one's workout and lifestyles. It is not clear how changes in our movement patterns affect the accuracy of our devices. For example, wrist fitness trackers indirectly assess the steps from the arm swing of the subject, which changes according to different walking and jogging movements.

The 3D accelerometer of the fitness trackers measures the movement done by the person in the different planes of the space. A wrist fitness tracker measures the wrist movement when the subject moves. These data allow us to indirectly estimate the steps taken through the displacement of the wrist and the anthropometric characteristics of the person. Each company uses different algorithms to calculate the steps from the movement measured by the device. However, the calculation does not have the possibility to consider whether the movement is affected due to our clothing, because we are carrying something, etc. Previous studies have not shown differences in the accuracy of the step measurement regarding the prices of the fitness tracker (El-Amrawy, Nounou, 2015). These studies found an error of underestimation between 2% and 30% in the step measurement between different brands and models (El-Amrawy, Nounou, 2015; Montoye et al., 2016; Nelson et al., 2016). Using the indirect estimation of the steps, the time, the type of movement, and the information about the subject regarding weight and age, the devices estimate the caloric expenditure done by the subject. The current research available about this indirect estimation shows that fitness trackers are not reliable to estimate the energy expenditure at rest or in exercise situations (Brazeau et al., 2016; Evenson, Goto, Furberg, 2015; Montoye et al., 2016; Nelson et al., 2016; Sasaki et al., 2014).

To assess the heart rate, the fitness trackers measure the reflection of their LED lights on the wearer's skin to detect blood volume changes in the capillaries and, from that, estimate the heart rate. Not all LED lights' reflection has the same precision in the measurement. Green LED lights are less precise, are cheaper, and consume less battery than red LED lights (El-Amrawy, Nounou, 2015). Fitness trackers use heart rate, age, and gender to estimate the intensity of the activity the subject is doing. Several studies have shown that the estimation is not accurate for most of the devices and they have a 15–20% error (El-Amrawy, Nounou, 2015). The accuracy is greater when the fitness tracker combines the use of green and red LED lights and when a higher number of LED lights are used. The accuracy of the heart rate assessment improves when the measure is done in laboratory-based activities (Shcherbina et al, 2017). Fitness trackers are intended to provide information about physical activity and fitness to the general population. The users of these devices need to know how precise the information that fitness trackers provide is. This information, together with the price, is critical in deciding whether the different fitness trackers are worth buying (ratio of assessment accuracy to price). Previous research has shown that they are not accurate for monitoring patients in hospital settings as early warning systems (Kroll, Boyd, Maslove, 2016). The results of this study will provide information regarding the accuracy that devices at different price points provide and whether they can be used to guide exercise. The purpose of this pilot study was to assess how reliable the step and heart rate measurements of wrist fitness trackers at different price points are.

Material and methods

Four healthy and active college students voluntarily participated in this pilot study (mean age, 21.7 \pm 1.2 years old; height, 1.69 m \pm 0.06 m; weight, 70.5 \pm 6.3 kg). A snowball sampling technique was used to select the sample (Trochim, Donnelly, 2001). The participants completed a written informed consent, PARQ+ (Warburton, Jamnik, Bredin, Gledhill, 2011), and weekly levels of physical activity and exercise before beginning the study. No financial or academic incentive was given to the students for participating in the study. Three wrist fitness bands (Xiaomi Mi Band, Fitbit Charge HRm, Fitbit Surge) and one sports watch (Polar M400) were analyzed in this pilot study (Table 1). The criteria used to select the fitness trackers was the price. A device costing less than \$25 (Xiaomi Mi Band), a device around \$120–150 (Fitbit Charge HRm), and a device around \$200–250 (Fitness Surge) were selected for this study. The Polar M400 with a pulse sensor H7 was selected because it was a sports watch with step count and a chest band to monitor the heart rate (\$150–200). The four devices used a 3-axis accelerometer to assess the steps. The Xiaomi Mi Band, Fitbit Charge HRm, and Fitbit Surge used two LED green lights to assess the heart rate. The study was pre-approved by the Institutional Review Board of the principal researcher and followed ethical standards of the Helsinki Declaration.

Table 1. Type, price, and characteristics of the wrist fitness trackers and the sports watch tested (Xiaomi Mi Band, Fitbit Charge HR, Fitbit Surge, and Polar M400)

	Xiaomi Mi Band S1	Fitbit Charger HR	Fitbit Surge	Polar M400
Type	Wrist fitness tracker	Wrist fitness tracker	Wrist fitness tracker	Sports watch
Price	\$15–20	\$120–150	\$200–250	\$180–200
Battery life	20–30 days	3–5 days	3–5 days	20–30 days
Display	No	Yes	Yes	Yes
3-D accelerometer	Yes	Yes	Yes	Yes
LED Lights	2 LED green lights	2 LED green lights	2 LED green lights	No, chest band

Subjects performed two sets of 10 sets of 100 steps wearing the fitness trackers on an indoor track in two situations: walking and jogging. Participants walked and jogged at a normal self-selected pace. In the walking situation, the subjects wore winter coats and gloves. The Xiaomi Mi band and Fitbit Surge were worn on the dominant wrist and the Fitbit Charge HR and Polar M400 were worn on the non-dominant wrist. Additionally, the subjects wore a chest band to monitor the heart rate with the Polar M400. The variables measured were the total steps and the heart rate after every 100 steps. Before starting the data collection, the subjects' information related to gender, height, weight, age, and arm used to wear the devices were introduced in the app or in the device. A test lap on the indoor track was carried out before the data collection to ensure that step counting and heart rate measurements were working in the devices. Use recommendations, updated app and updated firmware (Spring, 2016) for each device were used. At every 10-step interval, two researchers counted aloud the number of steps. If the researchers' counts did not match, the trial was started again. After each set, one researcher recorded the steps and heart rate from each device. The reliability of this method was tested by comparing a video recording for two walking and running video sequences of one of the subjects. An intra-class correlation coefficient value of 0.99 was obtained through the analyses of the video sequences and the steps recorded by the researchers. For the Xioami Mi Band, the data was obtained for its app (Android Mi Fit, version 2.2.3). After the data collection, the data were transferred to a spreadsheet.

The error for the steps and for the heart rate for each device were calculated. For the steps, the researchers' counting was used as the reference to evaluate the step measurement by the devices. For the heart rate, the Polar M400 with the chest band was used to evaluate the heart rate measurement of the wrist fitness trackers (Nunan, et al. 2009; Terbizan, Dolezal, Albano, 2002). A descriptive analysis (mean, standard deviation, and percentages) and an inferential analysis were performed using SPSS statistical analysis software (SPSS version 23.0, Chicago, USA). The Kolmogorov-Smirnov test was used to analyse the normality of the sample. Since the sample was non-parametric, a Wilcoxon Test was used to compared the differences between the step and heart rate assessments of the different devices. The level of statistical significance was set at $p \leq 0.05$.

Results

Table 2 and Figure 1 show the results of the step count trials for the wrist fitness trackers and the sports watch. The Polar M400 sports watch counted significantly fewer steps in the walking situation ($p < 0.001$). The Fitbit Surge counted significantly more steps in the walking situation ($p < 0.001$). No significant differences were found in the steps measurements in the jogging situation ($p < 0.138$). The Xiaomi Mi Band and Fitbit Charge HR presented the smallest levels of error in the step count. Both counted more steps for the walking situation and fewer steps for the jogging situation. Therefore, the total step count error was compensated, and the total error was less than 2%. The error percentage for each of the devices studied was lower than 8%. The sports watch presented greater variability than the fitness trackers in all the situations (average of 28.6 steps).

Table 3 and Figure 2 show the degree of error in the heart rate measurements of the fitness trackers when compared to the Polar M400 sports watch, which was used as a reference. In the walking situation, significantly lower values in the heart rate measurements for the Fitbit Charge HR were found ($p < 0.001$). In the jogging situation, significantly lower values in the heart rate measurements for the Xiaomi Mi Band, Fitbit Charger HR, and Fitbit Surge were found ($p < 0.001$). All the wrist fitness trackers registered fewer beats per minute than the Polar M400. The percentage of error was between 5 and 11% for each of the devices studied.

Table 2. Errors in number of steps measured by the different devices tested (Xiaomi Mi Band, Fitbit Charge HR, Fitbit Surge, and Polar M400 sports watch) in the situations of walking and jogging (values expressed in steps)

Situation	Xiaomi Mi Band S1		Fitbit Charger HR		Fitbit Surge		Polar M400	
	average	SD	average	SD	average	SD	average	SD
Walking	4.8	4.5	7.7*	5.5	7.6*	6.3	-8.7*	27.9
Jogging	-3.9	4.0	5.3	5.9	3.3	6.2	-2.8	29.3
Total	0.9	4.2	1.7	4.2	5.4	6.2	-5.8	28.6

* Significantly different from steps reference assessment ($p < 0.05$).

Table 3. Average heart rate measurement error for the different wrist fitness trackers (Xiaomi Mi Band, Fitbit Charge HR, and Fitbit Surge) in the situations of walking and jogging (values expressed in heartbeats per minute)

Situation	Xiaomi Mi Band		Fitbit Charger HR		Fitbit Surge	
	error	SD	error	SD	error	SD
Walking	-2.7	13.5	-9.5*	10.7	-0.5	8.5
Jogging	-11.7*	12.3	-17.9*	13.43	-14.6*	8.2
Total	-7.2*	13.0	-13.7*	12.0	-7.6*	8.4

* Significant differences found with heart rate reference assessment ($p < 0.05$).

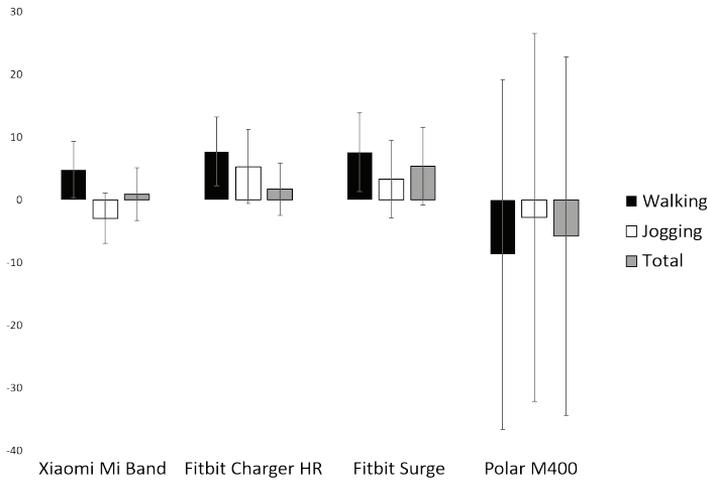


Figure 1. Average steps error for the different wrist fitness trackers and the sports watch (Xiaomi Mi Band, Fitbit Charge HR, Fitbit Surge, and Sports watch Polar M400) in the situations of walking and jogging (values expressed in steps)

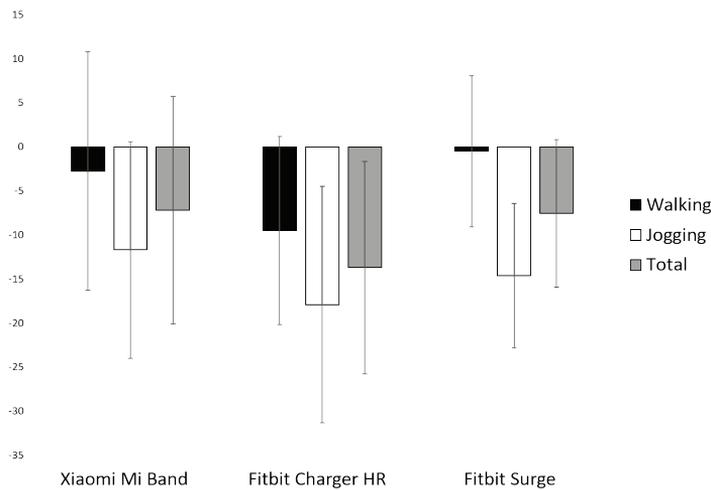


Figure 2. Average heart rate measurement error for the different wrist fitness trackers (Xiaomi Mi Band, Fitbit Charge HR, and Fitbit Surge) in the situations of walking and jogging (values expressed in heartbeats per minute)

Discussion

The purpose of this pilot study was to assess how accurate the steps and heart rate measurements of common wrist fitness trackers at different price ranges are. The results showed that the fitness trackers were relatively accurate for tracking steps with an average level of error of 2–6%. The sports watch studied presented

high variability in step counting (almost ± 30 of 100 steps). The devices tracked the steps in the jogging situation better than in the walking situation. These differences could be related to the shorter and slower arms swing when the subjects walked with winter coats and gloves. Fitness trackers indirectly measure the steps through the interpretation of the data collected by a 3-axis accelerometer of the movement done by the subjects' wrist. Each company has an algorithm that takes into consideration the wrist acceleration to establish whether a step was taken or not. The type of arm swing could affect the wrist acceleration generated and whether the threshold of the fitness tracker is reached. Regarding the prices, no significant differences were found between devices of different price ranges. The cheaper devices presented similar or lower levels of error than the more expensive devices studied. Fewer errors were found when compared to previous studies about the accuracy of the assessment of the fitness tracker (El-Amrawy, Nounou, 2015; Evenson, Goto, Furberg, 2015; Kooiman et al., 2015; Montoye et al., 2016). From a practical perspective, these results show the use of inexpensive devices is adequate for the general population to monitor physical activity due to their acceptable level of assessment and affordable price.

Regarding the heart rate assessment, the results showed that the wrist fitness trackers underestimate the heart rate when compared to the sports watch with a chest band. The level of error of the wrist fitness trackers was around 6–11% on average. The chest band was used as a reference since it provides valid measurements to monitor heart rate variability for walking and jogging (Nunan et al., 2009; Terbizan, Dolezal, Albano, 2002). The fitness trackers analyzed used two LED green lights and analyzed the reflection of the skin to detect changes in expansion and contraction of the capillaries when the heart beats. Algorithms interpret the data from the LED green lights' reflection in order to calculate the heart rate. Previous studies have found similar levels of error when LED green lights are used (El-Amrawy, Nounou, 2015). Another possible problem of the assessment was that the device moved along the wrist during the assessment, although criteria proposed by each company were followed. Most of the devices presented better values in the walking situation, in which the subjects used a winter coat and gloves. This type of clothing could help to maintain the same position of the fitness tracker during the assessment. In the jogging situation, when the arm swing is higher and faster, the devices had a higher amount of error in the assessment. The results from previous studies have found a similar problem with the chest band. At high speeds, the devices move, and they do not properly assess the heart rate (Terbizan, Dolezal, Albano, 2002). The same problem was found with fitness trackers, but it occurs also at lower speeds, because the device sits more loosely on the body. Regarding the prices, no differences were found in the accuracy of the assessment. The cheaper and the more expensive devices presented similar levels of error. From a practical perspective, these results showed how the use of the fitness tracker should not be used as a guide for an aerobic workout. The level of error does not allow one to accurately use it as a method to establish the intensity of the exercise.

In summary, the fitness trackers analyzed better assessed the step count than the heart rate. The devices studied underestimate the heart rate of the subject by approximately 10–15 beats per minute. It must be considered that this paper is an exploratory study and it has some limitations. Possible limitations of this study are: small number of participants, sample characteristics (young active adults), small number of devices, and wearing multiple fitness bands at the same time, which might affect the assessment. These limitations do not allow us to generalize the results. However, the paper presents some insights related to the accuracy of steps and heart rate, specifically related to the price of these fitness trackers. The findings of the study show how the different activities, clothing, etc. can influence their assessment. The current rapid evolution of our technology makes it necessary to constantly

evaluate the accuracy of the available fitness trackers. Future studies need to evaluate the accuracy of the new models of the fitness trackers (hardware and software) and the criteria general population use to select them.

Conclusions

The results show that the price is not a factor to consider in the selection of the fitness trackers related to counting the steps. All wrist fitness trackers studied provided acceptable measurements of the steps. The devices studied did not provide reliable information related to heart rate. These devices can provide complementary information about the physical activity done (steps), but they should not be used as a way to establish the aerobic workout intensity alone (heart rate). The results of this study cannot be generalized due to the small sample used (subjects and devices). However, the present pilot study shows differences in the assessment of the devices studied in two common situations of life: walking while wearing a coat and gloves and jogging. The current advances in technology and fitness trackers require the constant evaluation of the accuracy of the new devices that provide us feedback regarding our physical activity and exercise.

Acknowledgements

The authors have no affiliations with, or financial interest with the companies of the fitness trackers tested in this study.

References

- Brazeau, A.S., Beaudoin, N., Bélisle, V., Messier, V., Karelis, A.D., Rabasa-Lhoret, R. (2016). Validation and reliability of two activity monitors for energy expenditure assessment. *Journal of Science and Medicine in Sport*, 19 (1), 46–50.
- El-Amrawy, F., Nounou, M.I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research*, 21 (4), 315–320.
- Evenson, K.R., Goto, M.M., Furberg, R.D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12 (1), 159.
- 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.
- Kroll, R.R., Boyd, J.G., Maslove, D.M. (2016). Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. *Journal of Medical Internet Research*, 18 (9).
- Montoye, A.H., Tuttle, M.S., Nelson, M.B., Steeves, J.A., Kaminsky, L.A. (2016). Abstract P121: Consumer-based physical activity monitors and research-grade accelerometers underestimate physical activity in a semi-structured setting. *Circulation*, 133 (Suppl 1), AP121-AP121.
- Murakami, H., Kawakami, R., Nakae, S., Nakata, Y., Ishikawa-Takata, K., Tanaka, S., Miyachi, M. (2016). Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. *JAMA Internal Medicine*, 176 (5), 702–703.
- Nelson, M.B., Kaminsky, L.A., Dickin, D.C., Montoye, A.H. (2016). Validity of consumer-based physical activity monitors for specific activity types. *Medicine & Science in Sports & Exercise*, 48 (8), 1619–1628.
- Nunan, D., Donovan, G., Jakovljevic, D.G., Hodges, L.D., Sandercock, G.R., Brodie, D.A. (2009). Validity and reliability of short-term heart-rate variability from the Polar S810. *Medicine & Science in Sports & Exercise*, 41 (1), 243–250.
- Sasaki, J.E., Hickey, A., Mavilia, M., Tedesco, J., John, D., Kozey Keadle, S., Freedson, P.S. (2014). Validation of the Fitbit Wireless Activity Tracker® for Prediction of Energy Expenditure. *Journal of Physical Activity and Health*, 12, 149–154.
- Shcherbina, A., Mattsson, C.M., Waggott, D., Salisbury, H., Christle, J.W., Hastie, T., Wheeler, M.T., Ashley, E.A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, 7 (2), 3.

- Terbizan, D.J., Dolezal, B.A., Albano, C. (2002). Validity of seven commercially available heart rate monitors. *Measurement in Physical Education and Exercise Science*, 6 (4), 243–247.
- Trochim, W.M., Donnelly, J.P. (2001). *Research methods knowledge base* (2nd edition). Retrieved from: <http://www.socialresearchmethods.net/kb/index.php>.
- Warburton, D.E.R., Jamnik, V.K., Bredin, S.S.D., Gledhill, N. (2011). The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). *Health & Fitness Journal of Canada*, 4 (2), 3–23.

Cite this article as: Ricchio, K., Lyter-Antonneau, P., Palao, J.M. (2018). Reliability of fitness trackers at different prices for measuring steps and heart rate: a pilot study. *Central European Journal of Sport Sciences and Medicine*, 4 (24), 57–64. DOI: 10.18276/cej.2018.4-06.