Comparing Monitoring Results of Two Motion Trackers for Geriatric Patients

Amelie ALTENBUCHNER[a,1], Vanessa MÜCKE[a] and Sonja HAUG[a]
[a] *Institute for Social Research and Technology Assessment (IST), Ostbayerische Technische Hochschule Regensburg (OTH), Germany*

Abstract. Background: Fitbit and Garmin motion tracker devices are highly used in research. The validity and reliability of these devices is proven for healthy adults between 18 and 64. Objectives: Comparing data output of two devices. Methods: Observational case study on a test track and in the domestic environment of a 80-year-old female multimorbide geriatric patient. Results: High significant correlation of the devices on the test track [r=.776, p≤.001, Bca-CI-95% (.618;.874), N=33], but significant different in the domestic environment over time (z=4.840, p≤.001). Conclusion: The dominant/non-dominant body side and further sources of error may play a role in monitoring steps with these devices.

Keywords. physical activity, steps, geriatric patients, health monitoring, motion tracker, wearables, parallel test reliability, concurrent validity

1. Introduction

When commercial motion trackers are used as research instruments in medical or health science related studies, devices of Garmin and Fitbit are on top of the list, as a systematic review showed in 2018 [1]. Sensor based monitoring using customized products with one point systems worn on the wrist [2] works well to display physical activity and the outcome *steps* showed to be more suitable than duration [3,4] or energy consumption [4]. In such study designs the observation of mobility should provide an objective and individualized data base [5]. This also makes it beneficial for geriatrics as *mobility* is part of basical geriatric assessments.Steps recording of Fitbit devices is highly valid as a systematic review [6] showed. A high correlation between Fitbit records and a research pedometer (*r*=.91) can be reported [4] and GARMIN devices showed similar correlations (*r*=.90) with a research pedometer through different conditions:on a test track, a treadmill and in 24 hours-home setting [7]. These results are valid for healthy subjects between 18 and 64 mainly under laboratory conditions. In recent years more and more observational studies based using data generated by commercial motion trackers are conducted on patients [8], older patients and in geriatric rehabilitation contexts [9–12]. In a systematic review [3] of seven studies, published between 2014 and 2018, with overall 290 subjects aged 70.2 (± 4.8) activity tracking was accurate, whereas steps showed more accuracy than duration. Fitbit devices showed a high intra-class-reliability (ICC .94) with a reference device [3]. Devices underestimated steps of the group of older adults. Wearing the sensor on the ankle improved observation and recording of steps [3]. Other

[1] Corresponding Author: Amelie Altenbuchner, Institut für Sozialforschung und Technikfolgenabschätzung (IST), Ostbayerische Technische Hochschule Regensburg (OTH), Seybothstraße 2, D-93053 Regensburg, E-Mail: Amelie.Altenbuchner@oth-regensburg.de

Researchers [13] evaluated monitoring of steps of 99 subjects (mean age: 78.9 ± 8.6 years) with and without walking impairment as well as under the condition walking, walking with a walking stick and with a walker. Under all these conditions average error was lower when subjects wore the sensor on the ankle and in all groups under all conditions the number of steps was underestimated.

Why is the number of steps regularly underestimated? The devices recognize motion via a three-axis accelerometer [14,15], an acceleration sensor or gyroscope belonging to the so-called Micro Electro Mechanical Systems (MEMS). In a nutshell, the sensor captures a signal and preprocesses certain characteristics of it to classify the type of motion e. g. *steps*. Characteristics could be time-intervals, frequencies, other statistical features and further components according to domain and target [2]. The manufacturers usually do not publish their procedures. The uncertainty is not located within the sensor but in the algorithm calculating the number of steps. It may be reasonably assumed that manufacturers do not target older and oldest adults and especially not geriatric patients as customer group. Hence, the algorithm estimating the number of steps is not designed for the (differing) walking behaviour of these customers.

In this paper we argue, that although the monitoring and recording of steps with commercial wrist-worn devices may not be accurate for older adults there are advantages of using it in field research projects: Missing values in mobility assessment can be reduced [9] and patients with cognitive impairment, who cannot follow instructions of geriatric mobility assessment can be included in studies [5]. In aim to reduce missing values and to include patients with cognitive impairment it is advisable to choose the wrist solution over the ankle, because patients wear it like a wristwatch [16] and do not forget to put a sensor on their stocking in a domestic environment scenario. Furthermore they eventually forget about their tracking device, as it is just like a wristwatch and behave more natural during time still under observational conditions. By using the outcome *steps* (per day) instead of a mobility assessment score or duration of activity or energy consumption, the display of mobility / activity should be easier to understand for patients, caregivers and nursing relatives [17]. It could be accompanied with a comprehensible guideline in the future. Most importantly mobility and the ability to walk is the linchpin in every geriatric treatment and the key to self-help, autonomy and quality of life for older and oldest adults, for whom multiple unexplored research areas exist that have to be filled with evidence.

The aim of this paper is to compare monitoring results of two different motion trackers. It is an explorative approach since the research field holds many uncertainties for the target group. Key parameters of the observation are *steps* and *distance* both recorded by the motion tracker and observed by the researchers, in aim to research their consistency. The sub study is part of an ongoing long-term study ‘Motion monitoring of geriatric trauma patients – Explorative Study on the rehabilitation process after hip fracture using sensor-based data’ [16,17] and was positively evaluated by an ethics committee of the University of Regensburg. This study examined motion tracker devices of two providers: Garmin vívosmart 4 and Fitbit Alta HR (Tab. 1).

2. Methods

2.1. Subject

Subject of the study is an 80-year-old resident of a nursing home, who uses a walker inside and in the surrounding area. Her mobility is limited by multi morbidity. She has a medical history of cardiac arrhythmia, breast cancer and knee surgery. The test person is 160 cm tall and weighs 80 kilograms.

2.2. Instruments

General operating principles of devices (Tab.1) are explained in the background section.

Table 1. List of instruments

Provider	Type	Model
Fitbit	Motion tracker	Alta HR
Garmin Ltd.	Motion tracker	Vivosmart 4

2.3. Experimental Setup

The test procedure can be classified in two sections. The first part contains the test under defined conditions (test track), the second part the test in the domestic environment.

On a pre-defined test track of 50 steps, which consisted of six rounds (Tab. 2), two tracker-based observation records and one researcher based observation record were created. Furthermore the aim of the first test was to determine the personal average stride length, as it is not possible to measure it regulary due to immobility conditions of older and oldest adults. The second part of the step monitoring took place in the domestic environment. The aim was to find out whether the motion trackers produced similar results under everyday conditions. Prior to recording, the subject was informed what aspects to consider during wearing the devices. One of the main aspects was that the device always had to be worn at the same time in order to make a direct and clear comparison. If possible, the subject had to record any deviations in her daily routine and report them in a final interview. If the subject did not want to wear the devices for a certain period of time, she should discard both in order to avoid large deviations in step and distance estimation. If large differences occurred, these were investigated using the daily statistics, so it was possible to detect possible one-sided wear and remove it from the total quantity. In addition, the daily statistics could be used to check whether the different motion trackers recorded at the same time interval, had the same sensitivity, and provided similar results in motion detection. In addition, the respondent should always wear the wristbands on the same side as specified in the settings in order to consider the arm dominant / non-dominant option. Swapping the sides may result in over- or under-counting of the steps. At the end of the procedure, a final interview was held in order to determine possible mistakes or abnormalities.

2.4. Data collection

The subject wore two different devices on the test track and over a time period of 33 days. On the dominant arm side (right) she wore a Fitbit device, on the non-dominant arm side (left) she wore a Garmin device. Data was anonymized and stored in the

corresponding application. The test person received an anonymised abbreviation, which could not be traced back to her person. The storage and use of the recorded data required the consent of the test person and was recorded and signed in a declaration.
The data collection process consisted of four steps: At the beginning personal settings need to be calibrated. Body size is an important parameter, since the devices estimate the personal stride length by it. Equally important is adjusting the dominant / non-dominant arm setting, as this influences the sensitivity of the activity tracker. The next step was to define a test track in the home of the subject that had to meet certain prepared requirements: Straightness, no clear gradient and solid ground. The test track test was carried out at 50-step intervals. After each run, the different values were reported and displayed in a table (Tab. 2).

2.5. Analysis

Analysis was carried out using Microsoft Excel 16 and IBM SPSS 24. A paired sample Wilcoxon sign rank test was performed to estimate a difference in the distributions of the two devices and in case of the test rack of the counted steps. Pearson's correlation coefficient was estimated to compare data of the devices.

3. Results

3.1. Test Track

Table 2. Monitored and counted steps on the test track

Test track	Steps (Garmin)	Steps (Fitbit)	Steps (Counted)
1	49	50	50
2	48	52	50
3	51	50	50
4	50	53	51
5	50	51	50
6	49	52	50

Table 3. Descriptive statistics of monitored and counted steps

	Steps (Garmin)	Steps (Fitbit)	Steps (Counted)
Mean	49,5	51,3	50,2
SD	1,05	1,21	0,41
Var	1,10	1,21	0,18
CI-95%	48,4< X <50,6	50,0< X <52,6	49,8< X <50,6

Table 2 shows the absolute counts of steps during six rounds on the test track. The biggest difference between the devices is four steps on test track no. 2. The biggest difference between steps counted and monitored by the researcher accounts for two steps. The Fitbit device appears to be more accurate as it is consistent with researchers monitoring in two rounds, whereas the Garmin device only reports the accurate number of steps in one round. Overall, the Garmin device had six deviations during the test track rounds and Fitbit's device had seven. According to the frequencies table, Table 3 shows the descriptive statistics of the six track rounds. The individual stride length of the test person cannot be calculated using the data obtained, as the distance covered is unknown and due to test and subject's condition it was impossible to measure it. The reasons are

explained in the chapter discussion. Therefore, the stride length of 66 cm calculated by the motion wristbands was chosen. There is no difference in the Median of the distribution of steps on the six test tracks (z=1.807, p=n.s.).

3.2. Domestic Environment

Figure 1 shows absolute step counts of both devices in a graph. It is noticeable that Garmin's motion tracker consistently monitored fewer steps than Fitbit's. The model of Fitbit was worn continuously on the dominant wrist. The distance curves (Fig. 2) of the different motion trackers show the same progressions as the step count. This suggests that the different trackers use similar step lengths for calculation. To justify this mathematically, the selected step length was assumed to be unknown. Thus, the average step length is 72 cm in the Garmin model and 69 cm in the Fitbit model. This corresponds to a difference of 3 cm. There is also a trend to reduced physical activity through time, which is represented in a significant moderate negative correlation of steps (in both devices) and time (Fitbit: r=-389, p=.025; Garmin: r=.-380, p=.029).
Figure 3 shows the correlation of the total number of steps. Pearson's coefficient [r=.776, p≤.001, *Bca-CI*-95% (.618;.874), N=33] can be interpreted as a strong positive correlation (Fig.3). Still a high significant difference in the Median of the distribution of steps (Fitbit=1534; Garmin=712) over time (33 days) (Fig.4) can be reported (z=4.840, p≤.001). Figure 4 also shows outliers and extreme values, which occured on days 1, 2, 6, 20, 26, 27 and 28. The correlation between the steps records of Garmin and Fitbit gets lower, if these days are left out in data cleansing: r=.631, p=.001, *Bca-CI*-95% (.331;.812), N=26.

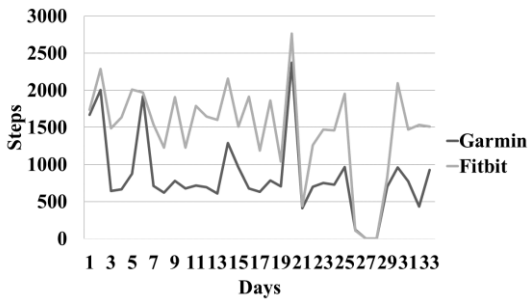

Figure 1. Steps monitored by Garmin and Fitbit devices

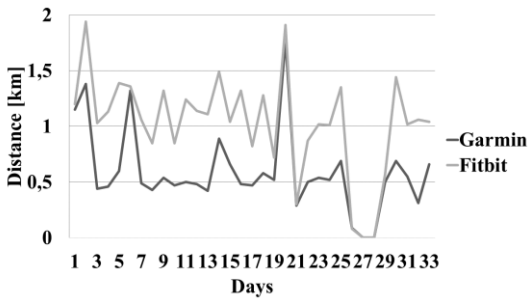

Figure 2. Distance monitored by Garmin and Fitbit devices

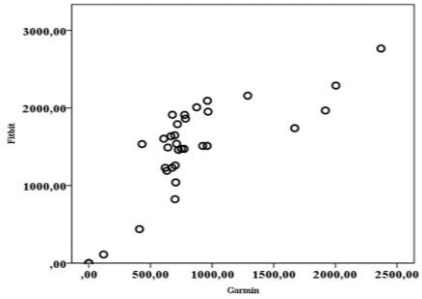

Figure 3. Scatterplot of correlation between steps monitored by Garmin and Fitbit devices

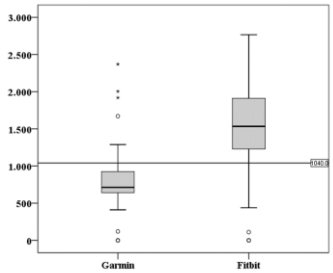

Figure 4. Box-Plot Distribution of steps monitored by Garmin and Fitbit devices

4. Discussion

In summary, the high correlation coefficient shows consistency in the data output of the devices. However, both appear to be inaccurate when we analyse the monitored results. The Garmin device, which was worn on the non-dominant wrist monitored less steps than the Fitbit device on the dominant arm side. This result only holds in domestic settings and not on the test tracks. On the test track step observation by the two devices was not significantly different from each other and from researcher's observation. The differences of monitoring on the dominant and non-dominant body side could be linked to movement in everyday conditions. During the time period of 33 days no data on the correct value of steps exists due to the research design. To assess the accuracy of the mean number of steps per day (Garmin:832.42, Fitbit: 1473.87) of the subject, only few publications can be taken into account. 49 Subjects (79.5 $\pm$ 7.8 years) showed a mean of 2467.7, CI-95% [617.1, 6820.4] during one week at roundabout 12 months after hip fracture [11]. The subject in this paper's study had no hip fracture, however shows multimorbidity and and is a geriatric patient, which comes along with a certain grade of immobility. There are limitations of the study which are discussed here. Firstly, a greater sample would have provided information more accurate. Secondly, it was impossible to measure the individual distances of each of the six test tracks and it was not manageable to switch arm sides of the devices, which may have had enabled to test effects of dominant and non-dominant side. These limitations are linked to the special challenges of the target group 'geriatric patients', which brings along the concept of an age-related vulnerability and frailty [18] and affects methodological and ethical considerations for

research [19]. Therefore, participating subjects are on a higher risk than healthy persons. At the same time, it is necessary to obtain more information on their health-related behaviour, especially in their domestic environment. As there is no best practice solution for the target group at this point, an adequate research design needs to provide flexible but well documented answers to the requirements of the patients [20]. For example participating in six test tracks of 50 steps showed out to be a physical and mental challenge for the subject. In order to save time and avoid both risk and drop out, it was not possible to measure the distance as planned. Additionally, the test track must have been performing in the domestic environment. Being a subject in a study also contains a certain kind of burden and for reasons of research ethics, this should not be pushed to the limits for a frail subject.

It is necessary to recognize sources of errors of the devices for future research, which can be of systematic and unsystematic type. Reasons for systematic errors in commercial motion trackers could be stride length and body height (p. e. incorrect setting), gait (e. g. clinical picture, aids) or installation (e. g. dominant vs. non-dominant arm). Reasons for unsystematic errors would be transmission errors (e. g. absentee synchronization), reading errors (e. g. time-delayed synchronization, wrong date) and variation in stride length (e. g. different speed, gradient, footwear). Results of Floegel et al. [13] show that use of walker results in a lower recording of number of steps. In this case it is also helpful to wear the motion tracker in a different place than the wrist during this time. However, if the study design includes also heart rate outcomes, the fact of underestimation can be taken into account when evaluating data or the daily target. The same holds for long-term studies where participants should never take their trackers of. In clinical surroundings it is often simply impossible to wear a tracker on the dominant body side due to pain or infusion for example. In aim to research geriatric patients and support their quality of life and mobility it is necessary to explore their mobility and their walking behaviour at all [17]. At the moment the use of commercial motion trackers seems to be an ‘as-good-as-possible-instrument’ for the research in domestic environments, if the following conditions are met: Exclude unsystematic errors, detect systematic errors and include this knowledge in consideration of study design, data management and analysis.

References

[1] A. Henriksen, M. Haugen Mikalsen, A. Z. Woldaregay, et al., Using Motion Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables, Journal of medical Internet research 20(e1109) (2018).

[2] G. Bieber, Methodik zur mobilen Erfassung körperlicher Aktivität mittels Beschleunigungssensoren. Dissertationsschrift Universität Rostock, Rostock, 2014.

[3] N. Straiton, M. Alharbi, A. Bauman, L. Neubeck, J. Gullick, R. Bhindi, R. Gallagher, The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review, Maturitas 112 (2018), 85–93.

[4] G. M. Dominick, K. N. Winfree, R. T. Pohlig, M. A. Papas, Physical Activity Assessment Between Consumer- and Research-Grade Accelerometers: A Comparative Study in Free-Living Conditions, JMIR mHealth and uHealth 4(e110) (2016); doi: 10.2196/mhealth.6281.

[5] T. Fleiner, P. Haussermann, S. Mellone, W. Zijlstra, Sensor-based assessment of mobility-related behavior in dementia: feasibility and relevance in a hospital context, International psychogeriatrics 28(10) (2016), 1687–1694.

[6] K. R. Evenson, M. M. Goto, R. D. Furberg, Systematic review of the validity and reliability of consumer-wearable activity trackers, The international journal of behavioral nutrition and physical activity 12 (2015), 1–22.

[7] H.-S. An, G. C. Jones, S.-K. Kang, et al., How valid are wearable physical activity trackers for measuring steps?, European journal of sport science 17(3) (2017), 360–368.
[8] R. K. Walker, A. M. Hickey, P. S. Freedson, Advantages and Limitations of Wearable Activity Trackers: Considerations for Patients and Clinicians, Clinical journal of oncology nursing 20(6) (2016), 606–610.
[9] P. Benzinger, U. Lindemann, C. Becker, et al., Geriatric rehabilitation after hip fracture. Role of body-fixed sensor measurements of physical activity, Zeitschrift fur Gerontologie und Geriatrie 47(3) (2014), 236–242.
[10] S. J. Davenport, M. Arnold, C. Hua, et al., Physical Activity Levels During Acute Inpatient Admission After Hip Fracture are Very Low, Physiotherapy research international 20(3) (2015), 174–181.
[11] L. Fleig, M. M. McAllister, P. Brasher, W. L. Cook, et al., Sedentary Behavior and Physical Activity Patterns in Older Adults After Hip Fracture: A Call to Action, Journal of aging and physical activity 24(1) (2016), 79–84.
[12] P. D. O'Halloran, N. Shields, F. Blackstock, et al., Motivational interviewing increases physical activity and self-efficacy in people living in the community after hip fracture: a randomized controlled trial, Clinical rehabilitation 30(11), (2016), 1108–1119.
[13] T. A. Floegel, A. Florez-Pregonero, E. B. Hekler, et al.,Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities, The journals of gerontology. Series A, Biological sciences and medical sciences 72(3) (2017), 229–236.
[14] Fitbit Inc., Wie zählt mein Fitbit-Gerät Schritte?, https://help.fitbit.com/articles/de/Help_article/1141#steps. Last access:1.11.2019.
[15] GARMIN, Technische Daten des vívofit 3, https://www8.garmin.com/manuals/webhelp/vivofit3/DE-DE/GUID-FA3C58BF-1124-4DE5-82CB-0DD6F461EF2E.html. Last access:1.11.2019.
[16] A. Altenbuchner, S. Haug, R. Kretschmer, K. Weber, How to Measure Physical Motion and the Impact of Individualized Feedback in the Field of Rehabilitation of Geriatric Trauma Patients. in: Health Informatics Meets eHealth: Biomedical Meets eHealth - From Sensors to Decisions. Proceedings of the 12th eHealth Conference, G. Schreier, D. Hayn (eds.), Amsterdam, 2018. pp. 226–232.
[17] A. Altenbuchner, S. Haug, K. Weber, Exploratory Analysis of Motion Tracking Data in the Rehabilitation Process of Geriatric Trauma Patients, in: dHealth 2019 - from eHealth to dHealth: Proceedings of the 13th Health Informatics Meets Digital Health Conference, D. Hayn, A. Eggerth, G. Schreier (eds.), Amsterdam, 2019. pp. 138–145.
[18] C. C. Sieber, Der ältere Patient--wer ist das?, Der Internist 48(11) (2007), 1190, 1192-4.
[19] Deutsche Akademie der Wissenschaften, Medizinische Versorgung im Alter: Welche Evidenz brauchen wir? Stellungnahme, Köthen, 2015.
[20] A. Altenbuchner, K. Weber, Geriatric trauma patients as research subjects in a technology-driven research project, in: Aging between Aging between Participation and Simulation: Ethical Dimensions of Social Assistive Technologies, J. Haltaufderheide, J. Hovemann, J. Vollmann (eds.), Berlin, 2020, pp. 87–103.