Introduction

Recommended amount of physical activity (PA) can have a positive influence on risk factors associated with a wide range of non-communicable diseases, such as ischaemic heart disease (I. M. Lee et al., 2012), diabetes mellitus (Chimen et al., 2007), hypertension (Saftlas, Logsden-Sackett, Wang, Woolson, & Bracken, 2004; Warburton, Charlesworth, Ivey, Nettlefold, & Bredin, 2010), overweight and obesity (Swinburn et al., 2011), and depression (Goodwin, 2003). In the adult population, the recommended amount of PA represents at least 150 minutes of moderate to vigorous intensity PA a week (World Health Organization, 2010). Moreover, there is large evidence that accumulating 10,000 steps has positive health benefits (Pelclová, Gába, Tlučáková, & Pośpiech, 2012; Tudor-Locke et al., 2011). However, a large part of the population worldwide fails to meet the PA guidelines (Hallal et al., 2012). According to the data based on an IPAQ questionnaire survey, adults in the Czech Republic were classified as highly active (Mitáš, Ding, Frömel, & Kerr, 2014), which is also confirmed by accelerometer-based data in a study conducted by Cerin et al. (2016). There are some limits associated with both subjective (Cerin et al., 2016; Troiano et al., 2008) and also objective (Hills, Mokhtar, & Byrne, 2014) measures to estimate PA and steps. Given the limitation to assessment of PA and steps, it is necessary to continuously validate estimates of objective monitors.

Accelerometers are the most commonly used instruments in objectively measured PA intensity in epidemiologic studies (Crouter, Dellavalle, Haas, Frongillo, & Bassett, 2012). However, accelerometer-derived
A. Šimůnek et al. (e.g. ActiGraph) measures require special software and skills to understand and evaluate such data. The accelerometer-derived time spent in PA of various intensity levels is influenced by the selected epoch length, cut-point threshold and bout duration (Orme et al., 2014). On the other hand, the number of steps is a unique value and easy for interpretation. Compliance to hip-worn accelerometer or pedometer can be affected by (1) the respondent’s willingness to wear a pedometer or accelerometer for a specified period (Frömel, Svozil, Chmelík, Jakubec, & Groffik, 2016); (2) willingness to put the device back on after changing clothes and to remove the device for the period of sleeping, personal hygiene and for water sport (Fairclough et al., 2015). This can be resolved by using wrist activity trackers (WATs); these are accelerometer devices which attempt to provide users with an easy way to objectively monitor their PA (J. M. Lee, Kim, & Welk, 2014). WATs have become a widespread tool for the assessment of achieved PA; therefore, these devices must be subjected to studies to test their validity for PA assessment in free-living conditions. The Garmin Vivofit (Vivofit) and Polar Loop (Loop) are relatively new WATs that, to the best of our knowledge, have so far not been assessed for the validity of measuring daily step counts in free-living conditions. Thus, the objective of this study was to determine the validity of the Vivofit and Loop by comparing them to well-validated devices, i.e. ActiGraph GT3X+ accelerometer (ActiGraph) (Chen & Bassett, 2005) and Yamax Digiwalker SW 701 pedometer (Yamax) (Schneider, Crouter, Lukajic, & Bassett, 2003).

Methods

Participants

The study involved twenty adults (14 males, 6 females) aged 25–52 years. The participants were volunteer university staff with a predominantly sedentary occupation. The individuals were encouraged to respond via email to indicate their willingness to participate. The participants were excluded if they had experienced an injury or illness affecting their mobility. The study was approved by the institutional research Ethics Committee of the Faculty of Physical Culture, Palacký University Olomouc under reference number 3/2015. After an initial introduction of the purpose, content and objective of the study, all participants gave an informed consent to participate in the study.

Data collection

This study was a cross-sectional comparison of four objective step measurement methods. The participants were asked to wear the Vivofit and Loop on the wrist of their non-dominant hand, simultaneously with Yamax and ActiGraph on their right hip for seven consecutive days except during the period of sleeping and for water sports. The participants used a daily activity log to record time and step counts (recorded by Vivofit, Loop and Yamax) when the devices were put on and when they were taken off. Daily step counts measured by the ActiGraph for the period specified in the daily activity log were calculated using Meterplus software (Santech, San Diego, CA, USA). The participants had no access to any of the device software. Data collection was undertaken between December 2014 and February 2015.

Wrist activity trackers

The Garmin Vivofit (Garmin, Schaffhausen, Switzerland) analyses data gathered by a 3-axis accelerometer by different algorithms using information such as height, weight, and age provided by the user (Mendoza, Han, Meyring-Wösten, Wilund, & Kotanko, 2015). The life of the CR1632 battery allows one year of monitoring of step counts, burned calories, PA time and quality of sleep. For heart rate monitoring the tracker must be paired with a chest strap (ANT+ compatible). The device is waterproof to a depth of 50 meters; therefore, activities in water can be carried out without removing it. The measured data are sent to a PC or a smart phone using Bluetooth smart and the Garmin Connect application (Maker, 2014). The price of the device is 99.99 USD.

Polar Loop (Polar Electro, Vantaa, Finland). The tracker uses an internal 3D accelerometer and wrist movement recorder to analyse the frequency and intensity of PA in accordance with the physical parameters of the user. The life of the chargeable lithium-polymer battery allows 5 to 7 days of monitoring of step counts, burned calories, level of PA, duration and quality of sleep. For heart rate monitoring the tracker must be paired with a chest strap (Bluetooth smart compatible). The device is waterproof to a depth of 20 meters; therefore, activities in water can be carried out without removing it. The measured data are sent to a PC using a charging USB cable or a smart phone using Bluetooth smart and the Polar FlowSync application (Maker, 2013). The price of the device is 59.95 USD.

Well-validated devices

Yamax Digiwalker SW-701 (Yamax, Kumamoto, Japan). The device is attached to the waist using a
clip. The Yamax pedometer is based on the principle of a horizontally positioned spring arm, which moves up and down during vertical movement of the waist. During each step, the movement of the arm opens and closes an electronic circuit, which represents a step count (Crouter, Schneider, Karabulut, & Bassett, 2003). Battery life allows 2 years of monitoring of step counts, distance covered during PA, amount of burned calories (Schneider et al., 2003). The device features high accuracy and reliability of step count measurement (Leicht, & Crowther, 2009; Schneider et al., 2003). The price of the device is 21.95 USD.

ActiGraph GT3X (ActiGraph, Pensacola, FL, USA). The device is attached to the waist using a clip. The ActiGraph accelerometer is a small electronic device that measures movement in three axes. The device monitors the intensity and duration of PA. Battery life allows PA monitoring for a period of 20 days. The GT3X accelerometer features high intra and inter tool reliability based on a vibration table test (Santos-Lozano et al., 2012). Due to its high validity it is the most frequently used instrument (of commonly manufactured accelerometers) for PA monitoring in children, adolescents, and adults (Chen & Bassett, 2005; De Vries, Baker, Hopman-Rock, Hirasing, & Van Mechelen, 2006). The price of the device is 250 USD.

Data analysis
The numbers of steps per day were averaged over valid days of recording. In accordance with a study by Van Dyck et al. (2015) a day was classified as valid if wear time of the ActiGraph exceeded 10 hours. The ActiGraph was set to collect raw acceleration signal at 30 Hz. These raw data were downloaded and converted with the Normal filter into 1-minute epochs using the ActiLife software 6.12 (ActiGraph, Pensacola, FL, USA). Descriptive statistics were conducted to summarize the demographic characteristics of the participants, including age, body mass index and daily step counts. The Shapiro-Wilk test was used to determine whether the data were normally distributed. A paired samples t-test was used to assess the differences in daily step counts recorded by the devices. Effect size was used to compare differences between the groups. Cohen (1988) classified effect sizes as small ($0.2 < d < 0.5$), medium ($0.5 \leq d < 0.8$) and large ($d \geq 0.8$). The mean absolute percentage errors (MAPE) for the number of steps for each WAT and well-validated device pair were calculated. MAPE were calculated as the mean of absolute differences between WATs and well-validated devices value divided by the well-validated devices value. The intraclass correlation coefficient (ICC) was used (absolute agreement, two-way random, single measure with 95% confidence agreement) to examine correlations between step counts derived from the Vivofit and Loop and from the Yamax and ActiGraph. The ICC used the following ratings – less than .20 is very weak, .20 to .39 is weak, .40 to .59 is moderate, .60 to .79 is strong and .80 or greater is a very strong correlation (Evans, 1996). Finally, Bland-Altman plots were created to evaluate the mean bias and the limits of agreement between the devices. All statistical analyses were performed using the IBM SPSS software (version 22.0; IBM, Armonk, NY, USA) with a significance level of alpha ≤ .05.

Results
Twenty adult participants were included with mean age 34.0 ± 6.3 years and mean BMI 24.3 ± 4.0 kg/m². According to the BMI international classification of adults (World Health Organization, 2006), 3 participants were classified as underweight, 13 participants were within the normal range and 4 were overweight. All participants were able to complete the survey. The average number of monitored days was for Yamax 7 ± 0 days, Vivofit 6.85 ± 0.47, ActiGraph 6.65 ± 0.89, Loop 6.45 ± 1.25. The highest daily average number of steps was measured by the Loop (11,995 ± 3,603), followed by Yamax (10,989 ± 2,692), Vivofit (10,449 ± 2,515), and ActiGraph (9,346 ± 2,264). Table 1 presents differences in number of steps between devices using paired samples t-test. The Vivofit non-significantly underestimated step counts by 540 daily step counts ($p = .06; d = 0.2$) compared with Yamax, while significantly overestimated step counts by 1,102 daily step counts ($p < .01; d = 0.5$) compared with ActiGraph. The Loop significantly overestimated the daily step counts compared with both the Yamax (by 1,006 steps per day; $p = .04; d = 0.3$) and ActiGraph (by 2,649 steps per day; $p < .01; d = 0.9$). Table 1 shows MAPEs for the WATs. The smallest MAPE values were found when comparing Vivofit and Yamax (−4.0%) and highest for comparison of Loop and ActiGraph (28.0%).

According to the ICC, the average daily step counts measured by the Vivofit showed a very strong correlation with both Yamax (ICC = .94, 95% CI: 0.83–0.98) and ActiGraph (ICC = .89, 95% CI: 0.40–0.97). The average daily step counts measured by the Loop also showed a very strong correlation with Yamax (ICC = .89, 95% CI: 0.67–0.96) and a strong correlation with ActiGraph (ICC = .70, 95% CI: 0.17–0.91). The Bland-Altman plots and analysis of step counts recorded by the WATs and well-validated devices are displayed in Figure 1. A comparison of the Loop with both well-validated devices provided wider limits of agreements than the Vivofit. The broadest limits of
agreement were determined for the Loop compared to ActiGraph (8,401 steps) and the narrowest limits were for Garmin compared to ActiGraph (4,597). Additionally, we found that the differences between the Loop and both the Yamax ($r = .50; p = .02$) and ActiGraph ($r = .70; p < .01$) increased as the mean value of daily step counts (x-axis) increased.

**Table 1**

<table>
<thead>
<tr>
<th>Devices</th>
<th>Paired differences</th>
<th>$t$-value</th>
<th>Cohen's $d$</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vívofit vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamax</td>
<td>-540</td>
<td>-2.0</td>
<td>0.2</td>
<td>-4.0</td>
</tr>
<tr>
<td>ActiGraph</td>
<td>1,102</td>
<td>4.2</td>
<td>0.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Loop vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamax</td>
<td>1,106</td>
<td>2.4</td>
<td>0.3</td>
<td>8.9</td>
</tr>
<tr>
<td>ActiGraph</td>
<td>2,649</td>
<td>5.5</td>
<td>0.9</td>
<td>28.0</td>
</tr>
</tbody>
</table>

**Figure 1.** Bland-Altman plots with 95% limits of agreement comparing steps per day between wrist activity trackers and well-validated devices.

**Discussion**

This study investigated the validity of two WATs compared with well-validated devices during free-living conditions. In general, Vívofit was more accurate in measuring step counts, and may be a suitable alternative to well-validated devices.
The leader in the worldwide wearable device market (Q3 2015) is Fitbit, followed by Apple, Xiaomi, and Garmin (Ritcher, 2015). Studies reporting on step counting metric validity for Vivofit and Loop are lacking. Therefore, we discussed our results with findings reported in studies that examined accuracy between other types of activity trackers and the ActiGraph GT3X accelerometer or Yamax SW 701 pedometer.

Firstly, we assessed the differences in the measured daily step counts across the devices. There were no significant differences between step counts measured by the Vivofit compared with the Yamax, while both the Vivofit and the Loop recorded significantly higher daily step counts compared with the ActiGraph. This finding is in line with results of a study conducted by Tully, McBride, Heron, and Hunter (2014) and those for Loop in line with study by J. A. Lee, Williams, Brown, and Laurson (2015).

The second method to examine the validity was to assess systematic differences. The MAPE was smallest between Vivofit and Yamax and highest between Loop and ActiGraph. Based on a study conducted by Tudor-Locke et al. (2006) an acceptable mean deviation in free-living conditions is 10%. Compared with Yamax, both Vivofit (–4.0%) and Loop (+8.9%) achieved this criterion. Compared with ActiGraph, Vivofit showed a smaller mean deviation than Loop (12.5% vs. 28.0%). The result for Vivofit is in line with a study by Paul et al. (2015), where the percentage difference for Fitbit average daily step counts compared with ActiGraph was 13%. The findings for Loop are similar those published by J. A. Lee et al. (2015), where Polar Active Accelerometer overestimated steps each day by 40.8% to 50.0%. In a study by Kooiman et al. (2015), WATs Misfit Shine, Jawbone UP and Fitbit Flex demonstrated smaller MAPEs (from 1.1 to 7.9) compared with the present study. However, these lower MAPEs can be partly explained by a shorter period of measurement (1 day).

The ICC with measures of consistency were used to determine the strength and consistency of the associations between daily step counts recorded by Vivofit, Loop Yamax and ActiGraph. The results demonstrated that the Vivofit exhibits a higher ICC compared with both the Yamax and ActiGraph than the Loop compared with both the Yamax and ActiGraph. The high value of ICCs between activity trackers and well-validated devices was consistent with a study by Paul et al. (2015), which concluded that Fitbit activity trackers showed excellent agreement with the ActiGraph in average steps per day over 7 days. Similarly, in a study conducted by Kooiman et al. (2015), WAT Fitbit Flex showed very strong (ICC$s = .96$) correlations with ActivPAL in free-living conditions. Evenson, Goto, and Furberg (2015) agree that ICCs of consumer-wearable activity trackers with accelerometers are generally high (ICC ≥ .80). Takacs et al. (2013) tested the Fitbit fitness bracelet on 30 participants performing PA on a treadmill and found a perfect agreement and inter-device reliability regardless of walking speed or device location. Ferguson, Rowlands, Olds, and Maher (2015) found a perfect step agreement between Withings Pulse and the ActiGraph GT3X+ accelerometer.

Finally, the level of agreement of Vivofit and Loop with Yamax and ActiGraph was assessed by Bland-Altman plots. The differences between upper and lower limits of agreement were smallest between the Vivofit and the ActiGraph and highest between the Loop and the ActiGraph (Figure 1). Our results, particularly for the Loop, are not in accordance with results by Ferguson et al. (2015), who found smaller differences between the upper and lower limits of agreement between the Withings Pulse and the ActiGraph.

Consistently with a study by Kooiman et al. (2015), despite the low MAPEs, larger individual differences (particularly for Loop) are obvious, which decreases the validity. However, the wear time in a study conducted by Ferguson et al. (2015) and by Kooiman et al. (2015) was 48 hours and 7.5 hours, respectively, which may explain the smaller differences in outputs from Bland-Altman plots. The important finding was that the Loop appeared to overestimate daily step counts in individuals who took more steps during a day. The findings indicate that, independently on PA level, the Vivofit can measure step counts in free-living conditions more accurately than the Loop.

The present study has some limitations. Firstly, a challenge concerning the interpretation of the results of the study included some discrepant findings. Vivofit non-significantly underestimated step counts compared with Yamax, but significantly overestimated step counts compared with ActiGraph. This disparity in our findings is caused by significant differences in average daily number of steps between well-validated Yamax and ActiGraph. This variance in reporting number of steps may be a cause of differences in the mechanism inside pedometers (pendulum) and accelerometers (micro electro-mechanical systems) (J. A. Lee et al., 2015). Secondly, we used a small sample, which can limit the generalizability of the results. The third limitation was that our study did not examine the validity of the devices in the assessment of total energy expenditure, duration of sleep, and time spent in MVPA. These functions have previously been, in a different type of trackers, subject to validation studies (Bai et al., 2016; Ferguson et al., 2015; Kooiman et al., 2015; J. M. Lee et al., 2014). These functions, especially total energy expenditure and time spent in MVPA, might be very
significant in various intervention studies. For providing complete characteristics of the devices, it is necessary to validate these variables in future studies. The fourth limitation is that the validity of activity trackers can be affected by various positions of the devices on the body, especially in free-living conditions (Atallah, Lo, King, & Yang, 2011; Kooiman et al., 2015; J. A. Lee et al., 2015).

Finally, the study participants were university staff with a mostly sedentary occupation, therefore validation in other population groups is required.

Conclusions

The Vivofit showed higher validity than the Loop in measuring daily step counts in free-living conditions. The findings of this study might assist individuals and researchers to make evidence based decisions about using wearable activity trackers for measuring daily numbers of steps.

Acknowledgment

This paper was supported by the research grant of Palacký University Olomouc “New technologies and approaches to physical activity monitoring: Utilization in kinanthropology research” (No. IGA_FTK_2015_003) and by the research grant of the Czech Science Foundation “The objectification of comprehensive monitoring of school mental and physical strain in adolescents in the context of physical and mental condition” (No. 13-32935S).

Conflict of interest

There were no conflicts of interest.

References


Validity of Garmin Vivofit and Polar Loop


