Use of Consumer Monitors for Estimating Energy Expenditure in Youth

Samuel R. LaMunion[1], **Andrew L. Blythe**[1], **Paul R. Hibbing**[1], **Andrew S. Kaplan**[2], **Brandon J. Clendenin**[1], **Scott E. Crouter**[1]

[1]Department of Kinesiology, Recreation, and Sport Studies, The University of Tennessee, 1914 Andy Holt Ave, Knoxville, TN 37996

[2]Department of Kinesiology, University of Wisconsin-Milwaukee, 2400 E Hartford Ave, Milwaukee, WI 53211

Abstract

The purpose of this study was to compare EE estimates from five consumer physical activity monitors (PAMs) to indirect calorimetry in a sample of youth. Eighty-nine youth (mean(SD); age, 12.3(3.4) yrs; 50% female) performed 16 semi-structured activities. Activities were performed in duplicate across two visits. Participants wore a Cosmed K4b^2 (criterion for EE), an Apple Watch 2 (AW, left wrist), Mymo Tracker (MT, right hip), and Misfit Shine 2 devices (MSH, right hip; MSS, right shoe). Participants were randomized to wear a Samsung Gear Fit 2 (SG) or a Fitbit Charge 2 (FC) on the right wrist. Oxygen consumption was converted to EE by subtracting estimated basal EE (Schofield's equation) from the measured gross EE. EE from each visit was summed across the two visit days for comparison to the total EE recorded from the PAMs. All consumer PAMs estimated gross EE, except for AW (net Active EE). Paired t-tests were used to assess differences between estimated (PAM) and measured (K4b^2) EE. Mean absolute percent error (MAPE) was used to assess individual-level error. The MT was not significantly different from measured EE and was within 15.9 kcals of measured kcals (p = 0.764). Mean percent errors ranged from 3.5% (MT) to 48.2% (AW). MAPE ranged from 16.8% (MSH) to 49.9% (Mymo).

- Only the MT was not significantly different from measured EE but had the greatest individual error.

- The MSH had the lowest individual error.

- Caution is warranted when using consumer PAMs in youth for tracking EE.

Keywords

Physical Activity; Activity Tracker; Wearable Devices; Fitness Tracker; Fitbit; Apple Watch

---

Corresponding Author: Samuel R. LaMunion, Department of Kinesiology, Recreation, and Sport Studies, 1914 Andy Holt Ave, Rm 307, Knoxville, TN 37996, Telephone: 865-974-5651, Fax: 865-974-8981, slamunio@vols.utk.edu.

Introduction

It is currently recommended that children and adolescents obtain 60 minutes or more of moderate-to-vigorous physical activity (MVPA) daily (U.S. Department of Health and Human Services). However, it is estimated that only 21.6% of youth ages 6–19 years meet the physical activity (PA) guideline according to wearable device assessment of PA (National Physical Activity Plan Alliance 2018; Troiano et al. 2008). PA assessment, which often focuses on estimating energy expenditure (EE) or counting steps, can be useful for activity tracking and as well as for goal-setting to assist with behavior change. A frequently used tool to assess PA are wearable PA monitors (PAMs). These devices are sold commercially and use sensors (e.g. accelerometers, gyroscopes, and altimeters) to track motion and translate that motion into relevant estimates such as steps or EE. Consumer PAMs are widely popular, and there is corresponding interest in their validity, particularly for devices made by industry leaders such as Apple, Fitbit, and Samsung. However, there has been disproportional focus on PAM validity and use for adults (Bai et al. 2016; Chowdhury et al. 2017; Dominick et al. 2016; El-Amrawy and Nounou 2015; Evenson et al. 2015; Ferguson et al. 2015; Kaewkannate and Kim 2016; Kooiman et al. 2015; Lee et al. 2014; Nelson et al. 2016; Woodman et al. 2016) versus youth (Sirard et al. 2017), and it is important to assess the two groups separately (Ainsworth et al. 2018; Butte et al. 2018; McMurray et al. 2015; Pfeiffer et al. 2018).

As of October 2018, a search of ClinicalTrials.gov showed there were 15 youth-focused studies that used a Fitbit device (U.S. Department of Health and Human Services 2018). As consumer PAMs become more common in clinical trials (Wright et al. 2017) and interventions (Gaudet et al. 2017), it is important to understand how consumer PAMs perform in youth. Thus, the purpose of this study was to evaluate the validity of EE estimates from five consumer PAMs in youth during simulated free-living activities.

Methods

Participants

One hundred youth ages 6–18 years from the Knoxville, TN community and surrounding areas were recruited via flyers, e-mail, and word of mouth. Parents or guardians of participants simultaneously provided written informed consent and a completed health history questionnaire to determine participant eligibility to participate in the study. Eligible participants then provided written informed assent and were enrolled in the study. Participants were asked to abstain from eating and drinking (except water) for 3-hrs prior to each lab visit. All study procedures were reviewed and approved by The University of Tennessee, Knoxville Institutional Review Board.

Procedures

Participants were asked to come into the Applied Physiology Lab at The University of Tennessee, Knoxville on two separate days. During the first visit, participants had their height, seated height, weight, and body composition measured in light clothing and no socks or shoes. Height was measured using a wall-mounted stadiometer. Weight was measured

during a body composition assessment using a Tanita BIA BC-418 Bioelectrical Impedance Analyzer.

Following the anthropometric measurements, resting metabolic rate (RMR) was assessed during 30 minutes of supine rest using a Cosmed K4b^2. The RMR was done in a non-overnight fasted state. Afterwards, eight semi-structured activities were performed from a list of 16 activities including: supine rest, sitting in a reclined position, using the internet (e.g. checking e-mail, watching videos, reviewing homework assignments, etc.), reading a book, playing computer games (e.g. Tetris, Pac-Man, Slither.io), over-ground self-paced slow and brisk walking, sweeping (e.g. sweeping paper shreds off a hallway floor into a dustpan), dusting (e.g. wiping down a table/counter top with a rag and spray cleaner), over-ground self-paced running, playing catch with a football, self-paced stair climbing – ascending and descending, playing soccer (e.g. simulated soccer including one-on-one play with shooting, dribbling, defending, and passing), playing basketball (e.g. simulated basketball including one-on-one play with shooting, dribbling, defending, and passing; also included free-play shooting), stationary cycling at a moderate intensity, self-paced jumping jacks). The remaining eight activities were performed during the second lab visit. During each lab visit, the respective eight activities were performed twice, once for 60–90 seconds and once for 4–5 minutes. Participants selected the order in which they would perform the activities. A member of the research staff kept track of time and gave a verbal prompt when participants could transition to the next activity. Participants were allowed to start an activity when they were ready and were instructed to perform the activities how they would under normal conditions based on their interpretation of the activity title.

Each visit, a member of the research team selected eight activities the participant would perform. Participants were asked to choose the order of the activities, such that there were two non-consecutive bouts of each activity for that day, with one bout lasting 60–90 seconds (short bout) and the other lasting 4–5 minutes (long bout). Participants determined when to terminate each activity, within the required duration (i.e., 60–90 s for the short bout and 4–5 min for the long bout). Upon reaching the minimum activity duration (60-s or 4-min) the participant was informed by a member of the research team that they could transition to the next activity whenever they were ready. If the participant did not transition to a new activity before the maximum activity duration (90-s or 5-min) they were asked to transition to the next activity at that time. Within activities, a random number generator was used to determine whether the long or short bout was performed first. In total, there were 16 activity bouts per day (eight activities times two bouts each) for a total of 32 activity bouts.

During each visit, all participants wore a Cosmed K4b^2 portable indirect calorimeter as the criterion measure of EE. They also wore an Apple Watch 2 approximately 4" proximal to the styloid process of the left wrist, a Mymo Activity Tracker approximately 1" lateral of the right iliac crest mounted on an elastic belt using the manufacturer supplied clip, and a Misfit Shine 2 approximately 2" lateral of the right iliac crest and on the dorsal surface of the right shoe using the manufacturer's clip attachment. Participants were assigned to wear either a Fitbit Charge 2 or a Samsung Gearfit2 on the right wrist, approximately 4" proximal to the styloid process. The same device combination was worn by the participant during both testing visits.

This study was part of a larger ongoing study focused on research grade PAMs (ActiGraph, activPAL, Axivity, GENEActiv), which were worn simultaneously with the consumer PAMs. Thus, for the purpose of this manuscript only the consumer PA monitors are reported on and it should be noted that the wrist-worn consumer PAMs were not worn in the manufacturer specified wear locations due to being worn in tandem with the research-grade monitors that occupied the primary wear location during each trial.

Equipment

The Cosmed K4b^2 (Cosmed S.r.l., Rome, Italy) is a portable metabolic unit that measures oxygen (O_2) consumption and carbon dioxide (CO_2) production using a breath-by-breath system. The analyzer unit is mounted on the chest with a battery unit mounted on the middle to upper back using a manufacturer specific harness. A sampling line connects to a small turbine fixed on a sealed face mask to transport gas samples to the unit for analysis. This device has previously been validated against the gold standard Douglas bag method for measuring O_2 consumption, CO_2 production, and ventilation (McLaughlin et al. 2001). Prior to each testing session, the Cosmed K4b^2 was calibrated, which included: 1) a room air calibration, 2) a reference gas calibration using a mixture of 15.93% O_2 and 4.92% CO_2, 3) a volume calibration with a 3-L syringe, and 4) a delay calibration. All calibration procedures were performed according to manufacturer's specifications (Cosmed).

Each consumer PAM has a smartphone application or website that can be used during the setup process (i.e., inputting demographic information used by the algorithms) and the data retrieval process (i.e., recording EE values at the start and end of the trial). Although the electronic interface for each PAM is unique, they typically request date of birth, gender, height, and weight of the user. All device profiles were updated before each trial using the participant's demographic information. All consumer monitors were Bluetooth paired to an Apple iPhone 5s, except the Samsung Gearfit2, which was paired to a Samsung Galaxy S6. Prior to the start of each testing session, after entering participant data, initial starting EE values were recorded. Immediately following the end of the trial devices were synced and final EE values were retrieved. Table 1 provides device specific information and technical specifications available from the manufacturers for each consumer monitor. Please note that the Apple Watch 2 reports active calories (only during movement) and total daily calorie values. Both EE estimates are available in the Apple Activity App on the iPhone but only active calories are available in the native Apple Health App and on the Apple Watch. For the purposes of this study, only active calorie estimates are evaluated for the Apple Watch Series 2. Due to no explicit definition from Apple, the active calorie EE estimate is treated as a net EE estimate from all active time (i.e. all activities except supine rest, reclining, computer gaming, internet use, and book reading).

Data Processing

After each visit, the Cosmed K4b^2 breath-by-breath (BxB) data were exported from the Cosmed software in BxB and 30-s epoch formats and used as the criterion measure for EE (kcals). The K4b^2 was timestamped using a Windows system clock to obtain the exact start time for each testing day and EE was analyzed for the entire testing day, including transitions between activities. To calculate criterion gross EE from the K4b^2, each 30-s

epoch of VO_2 data (ml·min^{-1}) was divided by 1000 to get VO_2 in L/min, then multiplied by 4.867 kcals/L to get kcals·min^{-1}, and then divided by 2 to get a representative kcal/30-s value for each epoch. The kcal/30-s values were summed to get gross EE of the testing day. For the consumer PAMs gross EE for each day was obtained by subtracting the starting EE from the ending EE.

For comparison to the Apple Watch 2 active calories, net active kcals were calculated. In order to retain as much active EE data as possible, K4b^2 BxB data (VO_2 in ml/min) were used in place of 30-s epochs since activities did not start and stop on the minute or half minute. Thus, the breath-by-breath data were first timestamped with activity codes and then all data points with a sedentary activity label (i.e. supine rest, reclining, computer gaming, internet use, and book reading) were removed leaving only active minutes and transition periods. To calculate kcals per breath, K4b^2 BxB data were divided by 1000 to get VO_2 in L/min, multiplied by 4.867 kcals/L to get kcals·min^{-1}, and then multiplied by the length of that breath in seconds divided by 60 seconds. These kcals/breath values were summed to get gross EE for the entire active bout length. Basal metabolic rate (BMR; kcals/day) was predicted using the Schofield equations (Food and Agriculture Organization of the United Nations 2004; Schofield 1985). BMR in kcals/day was divided by 1440 minutes to get kcals·min^{-1}, then multiplied by the total active minutes. Total bout BMR (active time only) was then subtracted from total gross K4b^2 EE (active time only) to get net active EE for the testing day. For Apple Watch 2, net active EE for each day was obtained by subtracting the starting EE from the ending EE from the active time.

To summarize the EE data from each PAM, EE data for both the K4b^2 and each consumer monitor were summed across both visits to get a total EE value representative of both testing days. Since activities were not performed in the same order on the same day, only participants that completed both visits were included in the final sample to prevent unbalanced inclusion of some activities over others. The total EE across both testing days was used for the subsequent analysis. Net active EE was used for the Apple Watch 2, all other consumer PAMs used gross EE estimates.

Statistical Analysis

Analyses were conducted jointly using IBM SPSS statistical software version 24 (IBM Corporation, Armonk, NY) and R statistical software. Data are presented as mean (SD). Participants were excluded from final analyses for: 1) not completing 2 visits (n = 7) or 2) declining to wear the K4b^2 on one or both visits (n = 4). Of the remaining 89 participants, there were additional synchronization errors for the consumer PAMs resulting in exclusion of participants that wore the Apple Watch 2 (n = 23), Fitbit Charge 2 (n = 10), Samsung Gearfit2 (n = 17), Misfit Shine 2 – Hip (n = 68), Misfit Shine 2 – Shoe (n = 70), and Mymo (n = 21). Thus, the final analytic sample for each device was: Apple Watch 2 (n = 64), Fitbit Charge 2 (n = 38), Samsung Gearfit2 (n = 23), Misfit Shine 2 – Hip (n = 21), Misfit Shine 2 – Shoe (n = 19), and Mymo (n = 65). Due to the unequal sample sizes, each device was examined separately.

Three analyses were done for each consumer PAM: 1) paired samples t-test comparing the estimated EE of each consumer PAM and measured EE from the K4b^2, 2) a 2×2 ANOVA

(EE estimate $\times$ age group [dichotomized by 6–12 and 13–18 years old, and 3) a 2$\times$2 ANOVA (EE estimate $\times$ sex). Performance was further assessed using Bland-Altman plots and mean absolute percent error (MAPE). Significance was set at alpha = 0.05.

Results

Physical characteristics of participants are shown in Table 2. Across both days of measurement, participants wore PAMs for a mean of 142 minutes. The Mymo was the only consumer PAM to not be significantly different from the K4b^2 (F = 0.07; p = 0.764) (Table 3). For net EE, the Apple Watch 2 significantly underestimated measured net active EE by 45.6%. For gross EE, the Fitbit Charge 2 significantly overestimated measured gross EE by 32.2% while the Misfit Shine 2 – Hip, Misfit Shine 2 – Shoe, and Samsung Gearfit2 significantly underestimated measured gross EE by 12.3–29.8%. There were no statistically significant interactions for EE $\times$ sex (p > 0.05) and only the Mymo had a statistically significant interactions for EE $\times$ age (F = 19.22; p < 0.001). The Mymo overestimated gross EE by 119.9 kcals (31.4%) for the 6–12 year age group and underestimated gross EE by 32.7 kcals (7.2%) for the 13–18 year group. MAPEs ranged from 16.8% (Misfit Hip) to 49.9% (Mymo) (Figure 1).

The Bland-Altman plots (Figure 2) showed mean biases ranging from –137.7 kcals (Fitbit) to 143.2 kcals (Samsung) for total kcals. The Apple Watch which reports net active kcals had the highest mean bias 174.7 kcals of all consumer PAMs. Conversely, the Apple Watch 2 had the narrowest 95% limits of agreement (98.1 – 251.2 kcals) of all wrist-worn consumer PAMs examined. The Mymo had the lowest mean bias of –15.9 kcals but had the second widest 95% limits of agreement (–272.6, 240.7 kcals) behind only the Fitbit Charge 2 (–470.1, 194.8 kcals), indicating large individual error. Additionally, the wrist-worn consumer PAMs had considerably higher mean biases (all > 140 kcals) compared to the hip- and shoe-worn devices (all < 70 kcals). Both the hip- and shoe-worn Misfit Shine 2 devices had relatively low mean biases (55.9 and 63.5 kcals, respectively), as well as comparatively low limits of agreement, indicating comparatively low group- and individual-level error.

Discussion

To our knowledge, this is the first study to examine estimates of EE from consumer PAMs in youth. The primary finding from this study is that estimated EE from all consumer PAMs were significantly different from measured EE in youth, except for the Mymo. Additionally, of interest, hip- and shoe-worn devices tended to have lower group-level error compared to wrist-worn devices as mean biases were consistently lower at non-wrist attachment sites. However, all the consumer PAMs had large individual error, with the highest MAPE coming from the hip worn Mymo.

Due to a lack of studies in youth, comparisons from the current study are limited to studies in adult populations. Several studies in adults have used similar semi-structured laboratory-based protocols to examine bout-level estimated EE from consumer PAMs. Of these studies, all reported lower MAPEs than what were found in the current study. It should also be noted that many of these studies used previous generations of the PAMs used in the current study,

however it is unclear if or when wearable manufacturers update EE prediction algorithms when new models are released. In general, Fitbit and Apple Watch devices had MAPEs ranging from 17%−36% in adults, (Bai et al. 2016; Chowdhury et al. 2017; Nelson et al. 2016) compared to MAPEs of >39% in the present study. Paired with the high mean biases and wide 95% limits of agreement it is clear that there is high group- and individual-level error for the Fitbit and Apple Watch models used in the present study. The Misfit Shine in adults has previously been shown to have a MAPE of 30%, (Chowdhury et al. 2017; Nelson et al. 2016), however in the current study MAPEs for the hip and shoe worn Misfit Shine 2 were <18%, the lowest of all consumer PAMs examined. Additionally, both the hip- and shoe-worn Misfit devices had low mean biases and comparatively narrow limits of agreement in the present study indicating overall these devices had low group- and individual-level error. It should be noted that this may be due to the relatively small sample size for these devices. There are no previous studies that have examined the Mymo or Samsung GearFit2 for estimating EE, however performance was similar to the other consumer PAMs examined in the current study.

One finding worth noting is that the present study reaffirms previous findings in adults that errors for estimating EE using consumer PAMs is generally greater when devices are worn on the wrist compared to the hip (Lee et al. 2014). The current study showed similar trends with greater mean biases for the wrist-worn devices compared to the hip-worn devices. However, the mean biases for the wrist-worn consumer PAMs in the current study were greater than mean biases reported using wrist-worn consumer PAMs in previous studies in adults (Bai et al. 2016; Chowdhury et al. 2017). This may be attributable to the more sporadic and intermittent movement patterns of youth compared to adults. Additional consideration should be given to algorithms used by each consumer PAM. It is unclear what data the algorithms use to make EE predictions and it is possible that a single attachment site-agnostic algorithm is used for all devices produced by a given manufacturer which could lead to high individual-level error at some attachment sites such as the wrist if the algorithm is not developed and optimized for the wrist. Conversely, it is also possible that a given manufacturer may have site-specific algorithms (e.g. hip and wrist) depending on the model and the specified attachment site which should theoretically improve the EE estimates and lower individual-level error within an attachment site. This evidence supports using caution when examining consumer PAM validation studies and to take device attachment site into account when interpreting validation studies in adults.

A recent meta-analysis by O'Driscoll et al. (O'Driscoll et al. 2018) examined the ability of consumer PAMs to estimate EE in adults. A total of 104 effect sizes were examined across 40 different devices. Several devices reported in the meta-analysis were the same or similar generation to monitors used in the present study. Specifically, The Apple Watch Series 2, Samsung Gear, and Misfit Shine were all found to underestimated EE in adults, which was consistent with performance of these consumer PAMs in the current youth study. In contrast, the Fitbit Charge 2 was found to underestimated EE in adults, while in the current youth study the Fitbit Charge 2 significantly overestimated EE. Taken together, the findings from previous consumer PAM studies in adults and the current youth study, show that in general the PAMs tend to over- or under-estimate EE in the same direction at a group level. However, the individual errors for consumer PAM estimates of EE tend to be lower in the

adult studies compared to the current youth study and supports the need for independent validations of consumer PAMs in youth.

While these devices may not be extensively used in the research community, there are several points to be made regarding the usability and interfacing of consumer PAMs for those that do use them for research purposes or are considering implementing consumer PAMs in their research. One challenge in consumer PAM validation studies is that the devices are designed to be worn by a single user. Error can occur from having to switch profile information between research trials if the devices are not carefully synced. For the Apple Watch Series 2 the smartphone functionality is divided between the "Watch" application, the native Apple "Health" application, and the "Activity" application. This distributed functionality can create confusion for the user and make validation research more challenging to carry out, hence the large number of synchronization errors with this device. However, the watch interface provides access to summary values in a single location, a user-friendly feature if only the watch is needed to record metrics of interest. An additional challenge with the Apple Watch is the lack of transparency about how "active" calories are being defined. This creates confusion about how to classify and compare the single summary EE value when conducting validation studies of the metrics consumers are using. Similarly, the Samsung GearFit2 has usability challenges of its own, where functionality is divided between the native Samsung "sHealth" and the "Gear" applications. Fortunately, for both devices, once user profile information is updated and devices are synced, data can be viewed on the smartwatch screen. However, it is not always clear whether the smartwatch screen and the application interface display the same values, nor is it clear which one should be recorded.

A limitation of most Fitbit models is that users cannot create an account if they are under the age of 13 years old. This can create an issue when editing the device profile for users under the age of 13. Fitbit recently released a youth-specific wrist-worn model called the Fitbit Ace. This device is designed for use in youth ages 8 years and older and features functionality and aesthetic similar to that of the Fitbit Alta but does not report EE; only steps, sleep, and active minutes (the accumulation of activity across the day to see if the user can reach the recommended 60 or more minutes of MVPA). Along with the release of the new device comes a "parent view" in the smartphone application. This allows the parent or guardian to setup and monitor the profile for youth users. Youth users are able to access their data through the "kid view" feature, a secure view that can only be edited by the primary account holder in parent view.

Misfit allows users to purchase multiple devices and wear them at different attachment sites simultaneously. However, this creates a problem if all devices are synced to the same account. The estimates given by the devices will be linked to the last device synced and its associated attachment site, so even when wearing multiple devices at different attachment sites, all estimates come out to be approximately the same. This became evident in the current study when syncing devices from multiple attachment sites to the same participant account and obtaining the same EE estimates from the devices, regardless of attachment site. This can be overcome by creating separate user accounts for each device being worn and syncing separately. When connected to separate user accounts, different estimates for each

attachment site were seen, as was expected. This issue was not recognized until more than two-thirds through completion of the study resulting in fewer cases for the Misfit Shine 2 compared to other consumer PAMs. However, once this issue was resolved and the new system established there was minimal data loss due to synchronization errors.

The Mymo is a PAM advertised as being suitable for researchers. It is primarily used as part of a suite of medical and clinical tools and resources. Thus, a challenge with the Mymo is that it is difficult to purchase individual monitors, and the electronic interface is underdeveloped. The lack of a screen on the device makes regular syncing a requirement to track activity and retrieve metrics of interest. Additionally, user profile information must be edited on the device website and cannot be changed within the smartphone application limiting overall usability and functionality.

The main strengths of this study include the use of multiple widely-used devices, and the comparison to a criterion measure of EE (indirect calorimetry). Other strengths include a large sample size, broad age range, and even distribution of boys and girls. Limitations to this study include the RMR measurement and the wear location of the consumer PAMs. The RMR measurement used in this study was not a true resting measurement since it was done mid-day in a non-overnight fasted state. This difference from a standard overnight fasted RMR could result in higher than measured expected resting values. Standard practice is to wear devices according to manufacturer recommendations. The current study placed wrist-worn devices in different positions than recommended by the manufacturers, which could impact EE estimates. The magnitude to which EE estimates may be influenced by the change in placement is unknown but must be disclosed. As previously mentioned, due to the proprietary nature of these devices it is difficult to speculate how the estimates of EE are made and know where errors might occur. Intuitively, changing the placement of the device further up the limb as done in the present study, would change the acceleration experienced by the device and thus could make a difference in the magnitude of sensor signal (e.g. dampening the response) and the subsequent estimates made using that sensor data. Future research should focus on quantifying differences that may occur by placing monitors in attachment sites outside of what is recommended by manufacturers as the case for wearing multiple devices simultaneously at a single attachment site may occur again in research settings. An additional limitation is that the most current models of all the devices was not examined. This is problematic for ongoing research protocols as device manufacturers routinely release new models; however, most of the new models provide updates to functionality, hardware, software features, etc. It is not clear if manufacturers update EE prediction algorithms with each model release but based on some consistency with previous adult studies using older models then used in the current study it does not appear the EE predictions are routinely updated when new firmware is released. Future work should investigate multiple generations of the same device to examine if there are changes in the EE predictions.

Conclusions

In youth, consumer PAMs have large group- and individual-level errors for estimating EE, compared to measured EE, thus they should be used with caution. While similar trends in

youth-specific EE estimates were seen in this study, compared to previous studies in adults, both group level and individual error was greater in youth further indicating the need for independent validations in each population. Validation studies should be continuously conducted as the cycle of technological advancement results in the annual release of new generations of consumer PAMs.

Acknowledgements

References

Ainsworth BE, Watson KB, Ridley K, Pfeiffer KA, Herrmann SD, Crouter SE, et al. 2018 Utility of the Youth Compendium of Physical Activities. Res. Q. Exerc. Sport, 89(3): 273–281. doi:10.1080/02701367.2018.1487754. [PubMed: 30019994]

Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, et al. 2016 Comparison of Consumer and Research Monitors under Semistructured Settings. Med. Sci. Sports Exerc 48(1): 151–158. doi:10.1249/MSS.0000000000000727. [PubMed: 26154336]

Butte NF, Watson KB, Ridley K, Zakeri IF, McMurray RG, Pfeiffer KA, et al. 2018 A Youth Compendium of Physical Activities: Activity Codes and Metabolic Intensities. Med. Sci. Sports Exerc 50(2): 246–256. doi:10.1249/MSS.0000000000001430. [PubMed: 28938248]

Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, and Thompson D 2017 Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One, 12(2): e0171720. doi:10.1371/journal.pone.0171720. [PubMed: 28234979]

Cosmed. Cosmed K4b2 User Guide. Rome, Italy.

Dominick GM, Winfree KN, Pohlig RT, and Papas MA 2016 Physical Activity Assessment Between Consumer- and Research-Grade Accelerometers: A Comparative Study in Free-Living Conditions. JMIR Mhealth Uhealth, 4(3): e110. doi:10.2196/mhealth.6281. [PubMed: 27644334]

El-Amrawy F, and Nounou MI 2015 Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? Healthc. Inform. Res 21(4): 315–320. doi:10.4258/hir.2015.21.4.315. [PubMed: 26618039]

Evenson KR, Goto MM, and Furberg RD 2015 Systematic review of the validity and reliability of consumer-wearable activity trackers. Int. J. Behav. Nutr. Phys. Act 12: 159. doi:10.1186/s12966-015-0314-1. [PubMed: 26684758]

Ferguson T, Rowlands AV, Olds T, and Maher C 2015 The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int. J. Behav. Nutr. Phys. Act 12: 42. doi:10.1186/s12966-015-0201-9. [PubMed: 25890168]

Food and Agriculture Organization of the United Nations. 2004 Human Energy Requirements. Available from http://www.fao.org/docrep/007/y5686e/y5686e07.htm [accessed October 7 2018].

Gaudet J, Gallant F, and Belanger M 2017 A Bit of Fit: Minimalist Intervention in Adolescents Based on a Physical Activity Tracker. JMIR Mhealth Uhealth, 5(7): e92. doi:10.2196/mhealth.7647. [PubMed: 28684384]

Kaewkannate K, and Kim S 2016 A comparison of wearable fitness devices. BMC Public Health, 16: 433. doi:10.1186/s12889-016-3059-0. [PubMed: 27220855]

Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, and de Groot M 2015 Reliability and validity of ten consumer activity trackers. BMC Sports Sci. Med. Rehabil 7: 24. doi:10.1186/s13102-015-0018-5. [PubMed: 26464801]

Lee JM, Kim Y, and Welk GJ 2014 Validity of consumer-based physical activity monitors. Med. Sci. Sports Exerc 46(9): 1840–1848. doi:10.1249/MSS.0000000000000287. [PubMed: 24777201]

McLaughlin JE, King GA, Howley ET, Bassett DR Jr., and Ainsworth BE 2001 Validation of the COSMED K4 b2 portable metabolic system. Int. J. Sports Med 22(4): 280–284. doi:10.1055/s-2001-13816. [PubMed: 11414671]

McMurray RG, Butte NF, Crouter SE, Trost SG, Pfeiffer KA, Bassett DR, et al. 2015 Exploring Metrics to Express Energy Expenditure of Physical Activity in Youth. PLoS One, 10(6): e0130869. doi:10.1371/journal.pone.0130869. [PubMed: 26102204]

National Physical Activity Plan Alliance. 2018 The United States Report Card on Physical Activity for Children and Youth.

Nelson MB, Kaminsky LA, Dickin DC, and Montoye AH 2016 Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types. Med Sci Sports Exerc. 48(8): 1619–1628. doi:10.1249/MSS.0000000000000933. [PubMed: 27015387]

O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. 2018 How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br. J. Sports Med doi:10.1136/bjsports-2018-099643.

Pfeiffer KA, Watson KB, McMurray RG, Bassett DR, Butte NF, Crouter SE, et al. 2018 Energy Cost Expression for a Youth Compendium of Physical Activities: Rationale for Using Age Groups. Pediatr. Exerc. Sci 30(1): 142–149. doi:10.1123/pes.2016-0249. [PubMed: 28787244]

Schofield WN 1985 Predicting basal metabolic rate, new standards and review of previous work. Human nutrition. Clin. Nutr 39 Suppl 1: 5–41.

Sirard JR, Masteller B, Freedson PS, Mendoza A, and Hickey A 2017 Youth Oriented Activity Trackers: Comprehensive Laboratory- and Field-Based Validation. J. Med. Internet Res 19(7): e250. doi:10.2196/jmir.6360. [PubMed: 28724509]

Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, and McDowell M 2008 Physical activity in the United States measured by accelerometer. Med. Sci. Sports Exerc 40(1): 181–188. doi:10.1249/mss.0b013e31815a51b3. [PubMed: 18091006]

U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans 2nd Edition Washington, D.C.

U.S. Department of Health and Human Services. 2018 ClinicalTrials.gov - Fitbit, Physical Activity, United States Available from https://clinicaltrials.gov/ct2/results?cond=Physical +Activity&term=Fitbit&type=&rslt=&age_v=&age=0&gndr=&intr=&titles=&outc=&spons=&lea d=&id=&cntry=US&state=&city=&dist=&locn=&strd_s=&strd_e=&prcd_s=&prcd_e=&sfpd_s= &sfpd_e=&lupd_s=&lupd_e=&sort= [accessed September 25 2018].

Woodman JA, Crouter SE, Bassett DR, Fitzhugh EC, and Boyer WR 2016 Accuracy of Consumer Monitors for Estimating Energy Expenditure and Activity Type. Med. Sci. Sports Exerc doi:10.1249/mss.0000000000001090.

Wright SP, Hall Brown TS, Collier SR, and Sandberg K 2017 How consumer physical activity monitors could transform human physiology research. Am. J. Physiol. Regul. Integr. Comp. Physiol 312(3): R358–R367. doi:10.1152/ajpregu.00349.2016. [PubMed: 28052867]
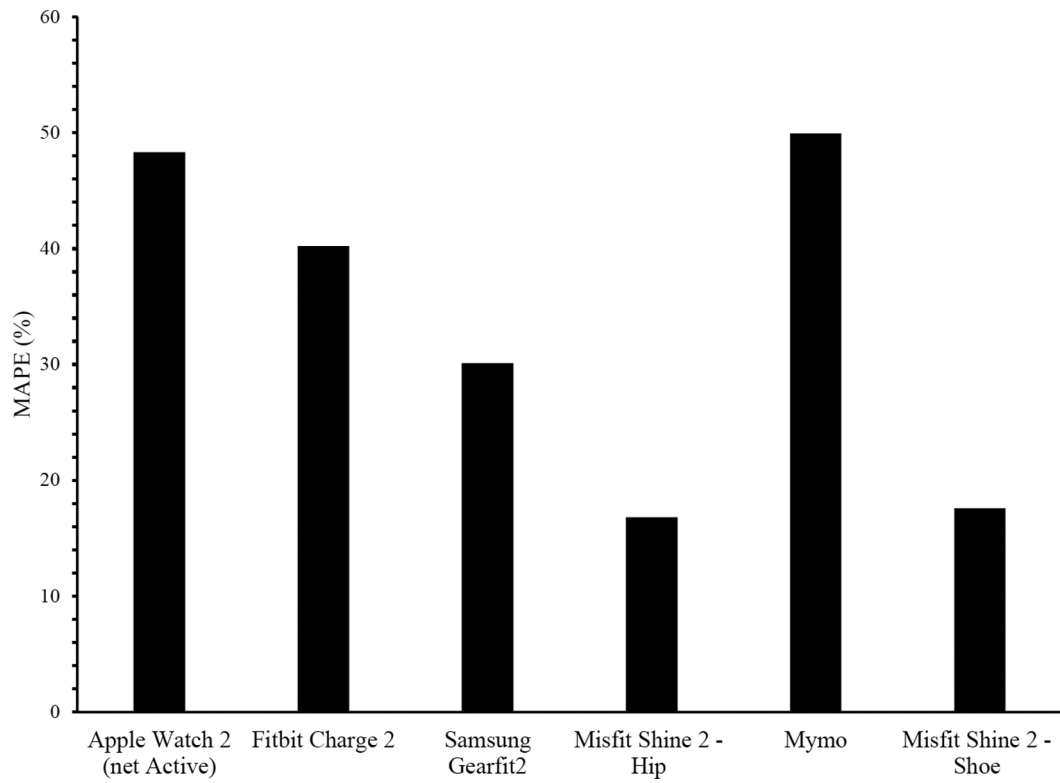
Figure 1.
Mean absolute percent error (MAPE) for each consumer physical activity monitor during an average of 148 minutes of assessment. (Apple Watch 2 is net EE; all other monitors are gross EE).
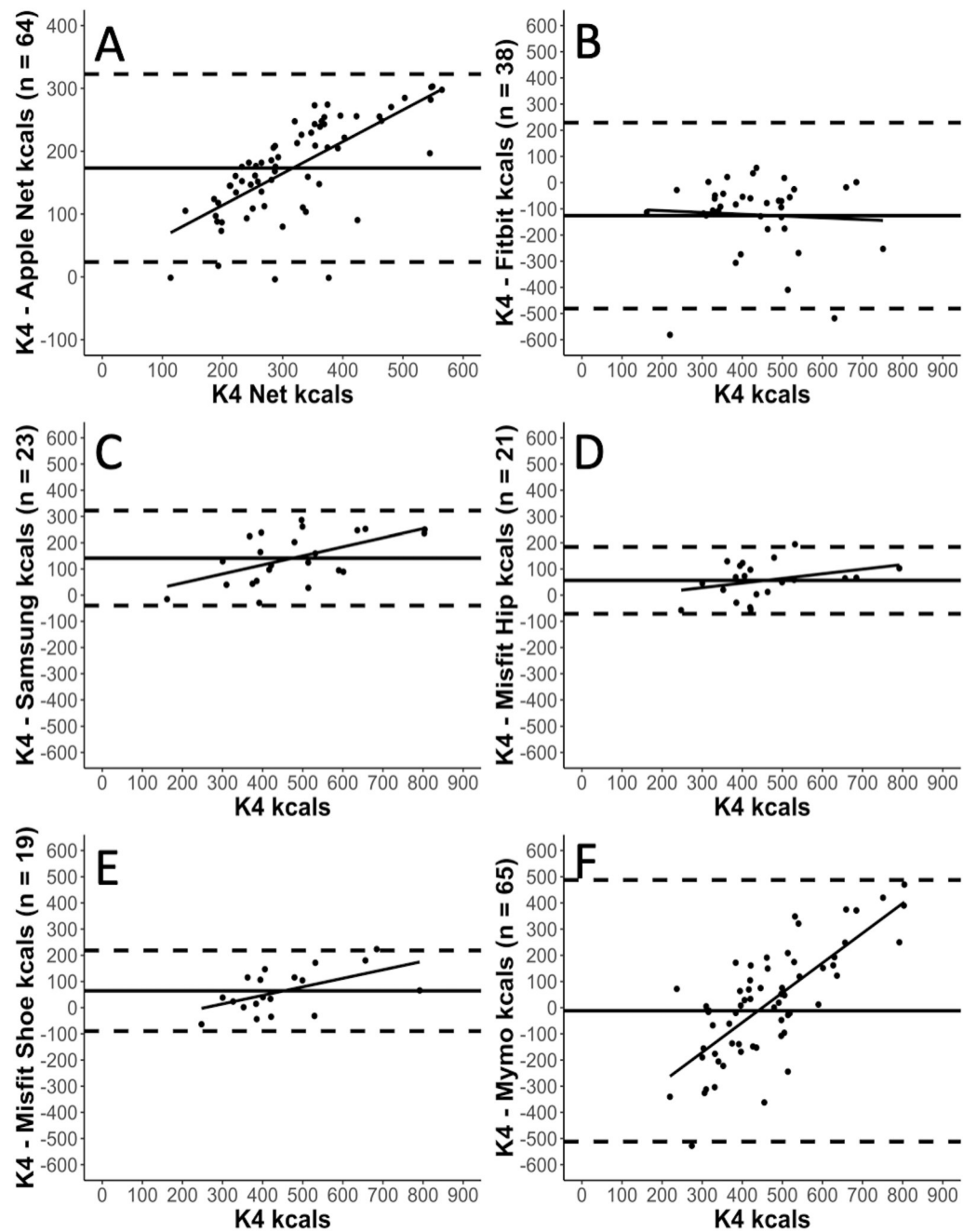
Figure 2.
Bland-Altman plots depicting error scores (measured minus estimation) for energy expenditure (EE) estimates made using consumer physical activity monitors compared to criterion measured Cosmed K4b^2 EE: A) Apple Watch 2 (net active EE), B) Fitbit Charge 2, C) Samsung GearFit2, D) Misfit Shine 2 – Hip, E) Misfit Shine 2 – Shoe, F) Mymo Activity Tracker.

Table 1.

Consumer monitor device information and technical specifications.

	Apple Watch Series 2	Fitbit Charge 2	Misfit Shine 2	Mymo	Samsung GearFit2
Website	https://www.apple.com/watch/	https://www.fitbit.com/charge2	https://misfit.com/misfit-shine-2	https://tupelolife.com	https://www.samsung.com/global/galaxy/gear-fit2/
Cost (USD)	$369+	$150	$100	$125	$180
Smartphone/Web Application	Apple Health (native)	Fitbit	Misfit	Tupelolife	SHealth (native)
Sensors					
Accelerometer	✓	✓	✓	✓	✓
Gyroscope	✓				✓
Heart Rate Monitor	✓	✓			✓
Built-in GPS	✓				✓
Magnetometer			✓		
Altimeter		✓			
Barometer					✓
Sleep Tracking		✓	✓		
Device Specs from Manufacturer					
Water Resistance	50m	Splash proof	50m		1.5m
Battery Life	~18 hours	~5 days	up to 6 months	up to 6 months	3–4 days
Battery Type	rechargeable	rechargeable	coin cell -CR2032	coin cell-CR2032	rechargeable
Sync Technology	Bluetooth	Bluetooth	Bluetooth	Bluetooth	Bluetooth
Compatibility	iOS	iOS, Android	iOS, Android	iOS, Android	Android
Display	✓	✓			✓
Attachment Sites	Wrist	Wrist	Hip, Wrist, Shoe, Pocket, Necklace	Hip	Wrist
Device Metrics					
Energy Expenditure (kcals)	Net Active Calories Gross Total Calories	Gross	Gross	Gross	Gross
Steps	✓	✓	✓	✓	✓
Distance	✓	✓	✓	✓	✓
Minutes of Physical Activity	✓	✓	✓	✓	
Flights of Stairs		✓			
Proprietary Metric			Points		✓

Table 2.

Physical characteristics of participants

	6 – 12 years old (n = 50)	13 – 18 years old (n = 39)	All Participants (N = 89)
Age (years; mean (SD))	9.7 (2.1)	15.6 (1.3)	12.3 (3.4)
Female (n (%))	25 (50.0%)	20 (51.3%)	45 (50.6%)
Height (cm; mean (SD))	139.4 (14.7)	167.0 (9.3)	151.5 (18.7)
Weight (kg; mean (SD))	33.6 (9.8)	61.0 (14.9)	45.6 (18.3)
BMI classification (n (%))			
< 5th percentile	3 (6.0%)	1 (2.6%)	4 (4.5%)
5th - < 85th percentile	41 (82.0%)	31 (79.5%)	72 (80.9%)
85th - < 95th percentile	6 (12.0%)	2 (5.1%)	8 (9.0%)
95th percentile	0 (0.0%)	5 (12.8%)	5 (5.6%)

Table 3.

Summary statistics and statistical comparisons of energy expenditure (EE) for each consumer monitor and the Cosmed K4b^2 across approximately 148 minutes of measurement.

Monitor	N	K4b^2 Measured EE (kcals)	Predicted EE (kcals)	Mean Difference (measured-predicted)	Lower 95% Limit of Agreement	Upper 95% Limit of Agreement
Apple Watch Series 2 (net active EE)						
*All Participants**	64	269.9 (82.8)	146.8 (76.3)	121.8	69.9	331.3
Age Group						
6 – 12 (y)	33	220.5 (49.8)	94.9 (26.3)			
13 – 18 (y)	31	322.5 (78.8)	201.9 (73.3)			
Fitbit Charge 2 (gross EE)						
*All Participants**	38	426.8 (129.5)	564.5 (205.3)	−137.7	−470.1	194.8
Age Group						
6 – 12 (y)	26	374.9 (96.9)	500.4 (178.1)			
13 – 18 (y)	12	539.5 (121.9)	703.6 (197.3)			
Samsung GearFit2 (gross EE)						
*All Participants**	23	480.2 (154.1)	337.0 (128.9)	143.2	−44.9	331.3
Age Group						
6 – 12 (y)	13	382.6 (90.7)	252.1 (82.5)			
13 – 18 (y)	10	607.1 (124.1)	447.5 (86.7)			
Misfit Shine 2 – Hip (gross EE)						
*All Participants**	21	455.4 (128.4)	399.4 (123.8)	56.0	−77.9	189.8
Age Group						
6 – 12 (y)	12	390.2 (70.0)	331.1 (69.2)			
13 – 18 (y)	9	542.3 (140.0)	490.6 (124.0)			
Misfit Shine 2 – Shoe (gross EE)						
*All Participants**	19	451.1 (138.3)	387.6 (116.2)	63.5	−98.6	225.5
Age Group						
6 – 12 (y)	11	377.6 (73.0)	324.6 (59.7)			
13 – 18 (y)	8	552.2 (146.3)	474.4 (121.5)			
Mymo (gross EE)						
All Participants	65	455.9 (142.7)	471.9 (178.5)	−15.9	−519.1	487.2

Monitor	N	K4b[2] Measured EE (kcals)	Predicted EE (kcals)	Mean Difference (measured-predicted)	Lower 95% Limit of Agreement	Upper 95% Limit of Agreement
Age Group[#]						
6 – 12 (y)	40	382.4 (92.3)	502.3 (185.0)			
13 – 18 (y)	25	455.9 (130.6)	423.2 (178.5)			