5-2017

Validity of Consumer-Based Physical Activity Monitors for Estimating Energy Expenditure in Youth

Andrew Blythe
University of Tennessee, Knoxville, ablythe@vols.utk.edu

Scott E. Crouter PhD
University of Tennessee, Knoxville, scrouter@utk.edu

Samuel R. LaMunion MS
University of Tennessee, Knoxville, slamunio@vols.utk.edu

Validity of Consumer-Based Physical Activity Monitors for Estimating Energy Expenditure in Youth

Chancellor's Honors Program Thesis

Andrew Blythe

Faculty Advisor: Dr. Scott E. Crouter

May 2017

Chapter 1: Introduction

Physical activity (PA) has been shown to reduce the risk of a variety of health issues in youth, such as, obesity (43), diabetes (38), and depression (22,38). The current PA recommendations for youth include obtaining a minimum of 60 minutes of moderate intensity PA per day (43). It is also recommended that muscle strengthening activities be included at least three days per week. Lastly, incorporating vigorous-intensity activities three days per week is advised (43).

As our understanding of the importance of PA in youth continues to grow, the ability to accurately track this behavior is a necessity; commonly expressed as an estimate of daily energy expenditure (EE) or time spent in sedentary behaviors, and light, moderate, and vigorous PA. Daily PA can also be examined using the following domains: occupational, domestic, transportation, and leisure. PA is defined as any use of skeletal muscle that results in increased EE above the resting value (5). The amount of energy expended to complete PA can be expressed as a metabolic equivalent of task, or MET. While an adult has an estimated resting metabolic rate (RMR) of 3.5 mL of oxygen per kg of body weight per minute, which equals 1 MET, the RMR of youth is higher than adults. Thus, a youth MET is defined as the oxygen cost of an activity divided by a measured or predicted RMR. Common prediction equations of RMR in youth include the Schofield and Harrell equations (34, 40). Factors influencing youth resting metabolic rate include growth, puberty and differences in body mass (19).

For measuring EE, doubly-labeled water, direct calorimetry, and indirect calorimetry are the gold standards (24,25,39). Doubly-labeled water is ideal for assessing total EE over an extended period of time. Direct and indirect calorimetry are ideal for shorter, more flexible periods of time. EE can also be estimated using a variety of objective measures. Objective measures include pedometers, heart rate monitors, and accelerometers. Of these objective measures, the most commonly used device is the accelerometer.

Accelerometers that are most often used in research include the ActiGraph (10,16,44), GENEActiv (14,31), activPAL (20,33), and Axivity (26). These accelerometers are specifically designed for research and clinical uses. Consumer-grade accelerometers are specifically marketed to the general population. Along with EE, these devices also provide information about steps taken, total distance traveled, and minutes of exercise. Importantly, the consumer models are often validated by a variety of methods including indirect calorimetry or by comparisons to research grade models (17). Some of these consumer PA monitors are the Apple Watch Series 2, the FitBit Charge 2, the Samsung Gear Fit 2, the Mymo Activity Tracker, and the Misfit Shine 2. Though consumer models have become increasingly popular, the research on these monitors focuses on adult populations (6,13,17,18,23). Because of this, validation studies need to be conducted to determine if these consumer models accurately estimate EE in youth.

Chapter 2: Literature Review

Measuring Energy Expenditure

Doubly-Labeled Water

Doubly-labeled water was developed by Lifson et, al. (39) in the 1940s for the use in animals. Applications for its use in human participants was slower to progress because of the initially high cost of the isotopes (39). That being said, the doubly-labeled water method is the "gold standard" for measuring EE in free-living participants. For this method, the participant is initially given $^{2}H^{18}O$ water. This water will then become isotopically equilibrated with the water throughout the body. When fluid is excreted from the body in urine, saliva, or sweat, the excretion rates of the ^{2}H and ^{18}O can be calculated. Because the ^{2}H isotope is only excreted in water while the ^{18}O isotope is excreted in both carbon dioxide and water, the difference in excretion rates yields total carbon dioxide produced. The rates of ^{2}H and ^{18}O disappearance can then be used to calculate the daily EE (39).

Doubly-labeled water has been shown to provide an accurate measure of EE, yielding a precision of 2-8% (39). It is also non-restrictive and can be used easily in a free-living environment. However, this method does require anywhere from one to three weeks to complete (39). Along with this, doubly labeled water does not provide information on intensity, duration, frequency, or type of activity. Because of these reasons, other methods are often used to measure EE.

Direct Calorimetry

Direct calorimetry is the process of measuring a person's metabolic rate by measuring their heat production. This method requires placing the participant in a sealed chamber that is insulated from the surrounding environment. By measuring the participant's total heat loss, the rate of EE can be estimated. This is because of the direct relationship between oxygen consumption and heat production (25). Currently, there are three common types of direct calorimetry units: isothermal, heat sink, and convection systems. While these systems are very accurate, they are expensive to build and operate (25). Furthermore, participant activity is limited to what is available inside the room.

Indirect Calorimetry

Indirect calorimetry involves measuring the gas exchange of a participant. By determining oxygen consumption and carbon dioxide production and knowing the oxidation reactions of carbohydrates and fats, EE can be calculated. The most common indirect calorimetry method is open-circuit spirometry. The original technique, and the technique used to validate newer technology, involved the participant wearing a mask that allowed air to be inspired, but any expired air would be funneled into a Douglas bag. This bag is then analyzed for percent oxygen and carbon dioxide (24). Currently, computer technology allows for breath-by-breath analysis of gas exchange as well as automatic calculations of oxygen consumption and carbon dioxide production. This set-up is referred to as a metabolic cart and is frequently used in both clinical and research settings. However, metabolic carts are restricted to a laboratory setting due to their lack of portability.

The Cosmed K4b^2, a portable metabolic system, is an indirect calorimeter designed to be used in a free-living environment. It is designed to be worn on the chest with a harness and has a battery pack mounted on the back. A face mask covers the mouth and nose of the user to seal against air leaks, and deliver expired air to the oxygen and carbon dioxide sensors through a sampling line. This device has been validated against the Douglas bag method for measuring oxygen consumption, carbon dioxide production, and ventilation, and had an error of <10% during stationary cycling (27).

Measurement of Physical Activity

Subjective Measures

Subjective measurement of PA involves the use of self-report surveys, diaries, and journal logs. Self-report surveys allow individuals to answer questions regarding their recent PA and often ask about different domains of PA; e.g., leisure time activity, or specific types of PA that participants do on a regular basis (e.g. walking or running) (41). Furthermore, self-report surveys are ideal for gathering information for population based studies, such as the National Health and Nutrition Examination Survey.

To properly assess PA through self-report, surveys need to be validated for the population being studied. Surveys have been validated using doubly-labeled water and accelerometers as criterion measures (41). Surveys can be invalid because of the difficultly in accurately recalling information about past PA. Because of the difficulty of recalling information, other methods for determining participant participants' PA are often used or preferred (41).

Objective Measures

The most common objective measure used in PA research is an accelerometer. Accelerometers are small, easy to use devices that can be worn on multiple body locations (11). Because of this, they have become the standard method for predicting EE in a free-living environment (22). These devices record acceleration data based on movement and have memory capacity to store this data for extended periods of time. Researchers have developed a variety of prediction equations to convert the accelerometer data into estimates of EE (42). It is important to point out that there are a variety of accelerometer-based devices and multiple prediction equations for each device that have been validated (32,35,42).

Research-Grade Activity Monitors

ActiGraph GT9X Link*.* The ActiGraph GT9X Link is ActiGraph's newest PA monitor. Developed in 2014, the GT9X uses a triaxial accelerometer and is typically worn on the wrist or waist. It weighs 14 grams, has four gigabytes of memory, has a sample rate of 30-100 hertz, and is water resistant. This model also incorporates a gyroscope, magnetometer, and a second triaxial accelerometer, that record at 100 Hertz, to measure information about movement, rotation, and body position. Furthermore, to set up the device and download data after use, one must use the Actilife software (1).

ActiGraph's accelerometers are the most popular among the scientific community for studies measuring PA in adults or youth. This is because there are a variety of independently developed and validated ActiGraph-based EE prediction equations (9,10,16). Research on ActiGraph accelerometer outputs has resulted in prediction equations for sedentary time, intensity levels of PA and estimated EE (9,10,16). Historically, most of these prediction equations focused on ActiGraph's worn on the hip; however, in recent years there have been equations developed for ActiGraph's worn on the wrist (9,16). For the hip, one study found that three equations that were specifically designed for children and adolescents, the Puyau equation, the Trost equation, and the

Freedson child equation, did not accurately predict EE in youth (42). Another study on the hip found that the age-specific Freedson MET equation overestimated EE, ranging from +0.3 METS to +1.5 METs, compared to the Cosmed Quark b^2 during treadmill walking (44). As for the wrist, one study developed two separate equations using outputs from an ActiGraph GT3X, one equation used the vertical axis and the other equation used vector magnitude, that both predicted EE (9). Neither of these equations were significantly different from measured child-METs from the Cosmed $K4b^2$ with the vertical axis model having a -3.2% mean bias and the vector magnitude model having a -6.1% mean bias (9).

GENEActiv*.* The GENEActiv is a wrist-worn triaxial accelerometer that records raw data at a rate up to 100 Hertz, stores up to 0.5 gigabytes of raw data, is waterproof, and can continuously monitor activity for a month on one charge. This device provides a continuous output of unfiltered, raw data that can be analyzed and manipulated in their free, open source software (18). Furthermore, the GENEActiv has been validated for use in adults (14) and youth (31). For adults, one study found that the GENEActiv had excellent criterion validity using the Cosmed $K4b^2$ as the criterion. The GENEActiv had correlations of 0.86, 0.83, and 0.87 for the left wrist, right wrist, and hip, respectively (14). For youth, Phillips et al. (31) developed cut-points based on GENEActiv outputs that identified sedentary behaviors and moderate and vigorous intensity activities with >82% sensitivity and >83% specificity.

activPAL*.* The activPAL is a small, lightweight device designed to be worn on the anterior thigh, midway between the hip and knee. It is attached directly to the skin with small, adhesion patches and can be worn continuously up to seven days. This device uses a sampling frequency of 10 Hz, has 4 MB of memory, and uses a triaxial accelerometer to produce raw data. The activPAL classifies free-living activity into time spent sitting, standing, and walking and estimates of EE can be obtained (30). Furthermore, it is often used for medical research and has been validated against the ActiGraph for use in children (33). One study that involved youth participants found that the METs estimated from the activPAL (4 METs) were significantly lower than measured METs (4.5 METs) during walking activity ($p < 0.001$) (20).

Axivity*.* The newest model from Axivity is the AX3 triaxial logging accelerometer. It weighs 11g, has 512 Mb of flash memory, and has a sample rate of 12.5-3200 Hertz. The device can continuously monitor activity for up to 14 days at 100 Hertz and is water resistant up to 1.5 meters (2). The Axivity is marketed as being ideal for collecting longitudinal movement data and being used for activity recognition, motion measurement, and medical research. However, there is limited research on its validity for estimating EE in youth. That being said, one study in youth found that equations developed by Brandes et al. (4) and Hildebrand et al. (21), using the Axivity AX3 outputs, overestimated EE during walking compared to a Metamax II portable indirect calorimeter by an average of 4.5% and 12.0-14.1%, respectively (26).

Consumer-Grade Activity Monitors

Fitbit Charge 2*.* The Fitbit Charge 2 is Fitbit’s latest fitness tracking device. The Charge 2 has a built in triaxial accelerometer, heart rate tracker, and altimeter. It has a battery life of five days, stores detailed motion data for up to seven days, and will save daily totals for up to 30 days. The Charge 2 synchs to the user’s phone or computer via Bluetooth LE wireless technology to view detailed motion data. The Fitbit Charge 2 retails for $149.00 (15). The Fitbit Charge 2 provides the user with total EE in kcals, total steps taken for the day, distance traveled, minutes of exercise, and total flights of stairs ascended.

Fitbit devices are commonly used in research studies. Currently, the research examining the estimates of EE by the Fitbit devices is limited to adult populations. The Fitbit Zip, a hip-worn device, has been shown to be equivalent with the Oxycon Mobile, for estimating EE (23). However, the Fitbit Flex, a wrist-worn device, has been shown to not be equivalent with the Oxycon Mobile for estimating EE. The Fitbit Flex had an average estimated EE of 337.2 kcals compared to an average EE of 316.8 kcals from the Oxycon Mobile (3). Another study reported the Fitbit Charge HR had a mean absolute percent error (MAPE) of 36% for EE compared to measured EE by a Cosmed $K4b^2$ (6). For steps taken during treadmill walking and jogging, the Fitbit One (wrist-worn) and Fitbit Flex (hip-worn) had correlations with hand counted steps of 0.97-0.99 and 0.77-0.85, respectively (12).

Apple Watch Series 2*.* The Apple Watch Series 2 is Apple's latest generation smart watch. It is marketed as a fitness device allowing the user a way to monitor many aspects of their physical health. The device has a built-in GPS, heart rate sensor, accelerometer, and gyroscope. From these sensors, the Apple Watch Series 2 estimates net EE, total daily steps taken, distance traveled by walking and running, and active minutes of exercise. It has a battery life of up to 18 hours per charge and is water resistant up to 50 meters. The Apple Watch Series 2 retails at $369.00 (7).

While no studies have been done using the Series 2, there have been several studies examining the Series 1 in adult populations. One study compared the Fitbit

Charge HR, Jawbone UP24, Microsoft Band, and Apple Watch Series 1. Of these devices, the Apple Watch Series 1 had the closest estimate of EE and lowest MAPE (27%) compared to measured EE (6). In a separate study, during a 1,000 step walking protocol, the Apple Watch Series 1 was 99.5% accurate for steps taken, compared to hand counted steps (13).

Samsung Gear Fit 2. The Samsung Gear Fit 2 is Samsung's most recent sports band. The Gear Fit 2 has a built-in GPS, heart rate monitor, accelerometer, gyro sensor, and barometer. These sensors provide data about total estimated EE, total daily steps taken, and flights of stairs ascended. It has a battery life of 3-4 days, memory storage of 4 GB, and is water resistant. It uses Bluetooth technology to connect to the user's phone or computer. The device can estimate steps taken, EE, and quality of sleep. The Gear Fit 2 retails at $149.00 (36). Research on Samsung devices has been limited to adult populations and on step counting. There is no research on the Samsung Gear Fit 2, however, the Samsung Gear 1 was 94% accurate for counting steps compared to hand counted steps (13).

Mymo Activity Tracker. The Mymo Activity Tracker is a small, device that clips onto the user's clothes. It estimates steps taken, distance traveled, EE, and active minutes. This activity tracker has a 4-6 month battery life, 35-day active memory, connects wirelessly to the user's smartphone or computer using Bluetooth, and automatically synchs when the device is shaken (29). Currently there is no research available examining the validity of the Mymo.

Misfit Shine 2. The Misfit Shine 2 is Misfit's newest PA tracker. It features a triaxial accelerometer, magnetometer, and a sleep tracker. It estimates total daily steps taken, total EE, distance traveled during any activity, and specific activity types such as walking, running, and swimming. This device is worn on the wrist, hip, or shoe, uses replaceable watch batteries that last up to 6 months, and is water resistant up to 50 meters. The Misfit Shine 2 retails at $99.99 (28).

Research on Misfit devices has been limited to adult populations. One study reported the Misfit Shine was 99.7% accurate during a 1,000 step walking protocol for step counting, compared to hand counted steps (13). Another study reported that, compared to an Oxycon Mobile measure EE (316.8 kcals), the Misfit Shine overestimated EE (395.5 kcals), during a protocol consisting of sedentary activities, aerobic exercise, and resistance exercise (3).

Future Research

Currently, the amount of research on consumer PA monitors is constantly growing. However, there is no research available on the newest versions of these devices. Furthermore, research on EE for the older generations of these devices have been limited to adult populations and have not been examined in youth. These consumer monitors claim to use updated PA tracking technology and provide the most accurate data available. Because of these claims, these models need to be validated for estimating EE. On top of this, these devices need to be specifically validated in youth populations.

Chapter 3: Manuscript

Introduction

Accelerometers are one of the most common methods for tracking daily physical activity (PA). These devices are small, easy to use, and can be worn on multiple body locations (11). Because of these characteristics, they have become a widely-used method for estimating energy expenditure (EE) in free-living environments (23). Recently, accelerometer-based devices have gained popularity among the general population. These consumer activity monitors are a convenient way to provide the user with information about their PA (e.g. steps taken and calories burned) (3). Popular manufacturers of commonly used consumer PA monitors include Fitbit, Apple, Misfit and Samsung.

Many currently available consumer devices have been studied within the scientific community, but to date, research examining the estimates of EE from these devices has been limited to adult populations. For example, one study reported that the Fitbit Charge HR and the Apple Watch Series 1, both wrist-worn devices, had MAPEs of 36% and 27%, respectively, for EE compared to measured EE by the Cosmed $K4b^2$ during an eight activity protocol (6). These activities were typing, loading and unloading a dishwasher, sweeping, walking stairs, walking on a treadmill, walking with shopping bags, stationary cycling, and jogging on a treadmill (6). Another study found that the Fitbit Flex and the Misfit Shine, both wrist-worn devices, were not equivalent to the Oxycon Mobile for estimating EE with mean absolute percent errors (MAPE) of 16.7% and 30.4%, respectively (3). Lastly, one study showed the Fitbit Zip, a hip-worn device, was equivalent with the Oxycon Mobile, for estimating EE with an error of 10.1% (22).

These studies support a common trend in the literature that devices worn on the hip more accurately estimate EE compared to devices worn on the wrist.

While these studies show the validity of some of the most popular consumer PA monitors, they are limited to adult populations and older generations of these devices. There is no research on the validity of these devices for estimating EE in youth or on the newest generations of these devices (e.g. Apple Watch Series 2 and Fitbit Charge 2). Thus, the purpose of this study was to validate five consumer PA monitors (Apple Watch Series 2, Fitbit Charge 2, Misfit Shine 2, Mymo Activity Tracker, and Samsung Gear Fit 2) in youth, for estimating EE during 16 semi-structured activities.

Methods

Fifty youth, 6-18 years old, were recruited from the Knoxville community. Participants were screened for exclusion criteria using a health history questionnaire. Exclusion criteria included currently pregnant, asthma, or musculoskeletal injuries that would limit activity. Prior to participation in the study, written parental informed consent was obtained followed by written informed assent from the participant. This study was approved by the University of Tennessee Knoxville Institutional Review Board.

Participants were tested on two separate occasions while wearing a combination of research-grade (ActiGraph, activPAL, Axivity, GENEActiv) and consumer-grade (Apple Watch Series 2, Fitbit Charge 2, Samsung Gear Fit 2, Mymo Activity Tracker, and Misfit Shine 2) PA monitors on various body locations. For the purpose of this manuscript only the consumer PA monitors will be reported on. Participants were asked to abstain from vigorous exercise for 24-h before data collection, and abstain from eating and drinking (except water) for 4-h prior.

On the first day of testing, height, seated height, weight, and body composition were measured in light clothing and no shoes. Height measurements were made using a mounted stadiometer and weight and body composition measures were made using bioelectrical impedance (Tanita BIA BC-418). All participants were fitted with an Apple Watch Series 2 approximately 3" proximal to the styloid process of the left wrist, a Mymo Activity Tracker and Misfit Shine 2 approximately 1" and 2" lateral of the right iliac crest, respectively, and a Misfit Shine 2 on the dorsal surface of both left and right shoes. Additionally, a Fitbit Charge 2 or Samsung Gear Fit 2 were randomly assigned to each participant to be worn approximately 3" proximal to the styloid process of the right wrist. The same combination was worn by the participant during all testing. Lastly, for measured energy expenditure each participant was fitted with a Cosmed K4b^2 portable indirect calorimeter mounted in a chest harness. Participants then complete 30 minutes of supine rest, to determine resting metabolic rate (RMR), followed by completion of a semi-structured PA routine consisting of eight activities. The eight activities were randomly assigned from a list of 16 activities, with the remaining eight activities to be performed by the participant during the second visit. The order the activities were performed was selected by the participant prior to testing. Each activity was completed two non-consecutive times with one bout lasting 60-90 seconds and the other bout lasting 4-5 minutes. After 60 seconds or 4 minutes of an activity, the participant was informed that they could transition to the next activity whenever they were ready. If the participant did not transition to a new activity before either the 90 second or 5-minute mark, they would be asked to transition to the next activity. The 16 activities performed are as follows:

1. supine rest
2. sitting in a reclined position
3. using the internet
4. reading a book
5. playing computer games
6. over-ground self-paced slow walk
7. sweeping
8. dusting
9. over-ground self-paced brisk walk
10. playing catch with a football
11. self-paced stair climbing
12. playing soccer
13. playing basketball.
14. stationary cycling at a moderate work rate
15. over-ground self-paced running
16. performing jumping jacks

The Cosmed K4b^2 (Cosmed, Italy) is a portable metabolic system engineered to measure oxygen consumption and carbon dioxide production on a breath-by-breath basis. The device is designed to be worn on the chest with a harness and have a battery pack mounted on the back. It is connected via a sampling line to a small turbine fixed on a face mask. This device has been validated against the Douglas bag method for measuring oxygen consumption, carbon dioxide production, and ventilation, and had an error of <10% during stationary cycling (27). Prior to each test, the Cosmed K4b^2 was calibrated, which consisted of room air calibration, reference gas calibration using a mixture of 15.93% O_2 and 4.92% CO_2, volume calibration with a 3-L syringe, and a delay calibration.

Detailed Information (such as components and functions) on the consumer PA monitors can be found in Table 1. The consumer monitors allow for a user profile to be created via either a smartphone application or website. These profiles typically request date of birth, gender, height, and weight. All device profiles were set-up using an investigator’s smart phone or computer prior to testing each day.

The Apple Watch Series 2 (Apple Inc., Cupertino, CA) is a small, lightweight, wrist-worn PA monitor. It is water resistant up to 50 m and has a battery life of up to 18 hours. Sensors within the device include a triaxial accelerometer, built-in GPS, heart rate sensor, and gyroscope. Data from these sensors are used to estimate net EE, steps taken, and distance traveled. Data was obtained using the MyWatch Activity Application from Apple.

The Fitbit Charge 2 (Fitbit, Inc., San Francisco, CA) is a small, lightweight, wrist-worn PA monitor. It is splash proof and has a battery life of up to 5 days. Sensors within the device include a triaxial accelerometer, altimeter, heart rate monitor, and sleep tracker. Data from these sensors are used to estimate gross EE, steps take, distance traveled, and flights of stairs ascended. Data was obtained using the Fitbit application.

The Samsung Gear Fit 2 (Samsung Electronics Co. LTD, South Korea) is a small, lightweight, wrist-worn PA monitor. It is water resistant up to 1.5 m and has a battery life of 3-4 days. Sensors within the device include a triaxial accelerometer, gyroscope, heart rate monitor, built-in GPS, and barometer. Data from these sensors are used to estimate gross EE, steps taken, and flights of stairs ascended. Data was obtained from the Samsung Gear Fit application.

The Mymo Activity Tracker (TupeloLife, India) is a small, lightweight, hip-worn PA monitor. It has a battery life of 4-6 months and can store data for up to 35 days. Information on the sensors within the device is currently unavailable from the manufacturer. Data from these sensors are used to estimate gross EE, steps taken, and distance traveled during ambulation. Data was obtained using the TupeloLife’s online website (mymolife.com) and TupeloLife application.

The Misfit Shine 2 (Misfit, Burlingame, CA) is a small, lightweight, PA monitor that can be worn on the wrist, hip, or shoes. It is water resistant up to 50 m and has a battery life of 6 months. Sensors within the device include a triaxial accelerometer, sleep tracker, and magnetometer. Data from these sensors are used to estimate gross EE, steps taken, and distance traveled. Data was obtained using the Misfit application.

Breath-by-breath VO_2 and VCO_2 from the Cosmed $K4b^2$ were used to compute EE (kcal), which was averaged over a 30-s period, and used as the criterion measure for EE. For the consumer devices, EE values were recorded on a data sheet immediately before and after each test, and the differences between the start EE and end EE were recorded as either net (Apple Watch Series 2) or gross EE (all other monitors). To compare the net EE of the Apple Watch Series 2 to the measured EE from the Cosmed $K4b^2$, the Schofield equation (38) was used to get a predicted RMR. The predicted RMR was converted to a one minute kcal value that was then multiplied by the total minutes of the test and subtracted from total EE of the Cosmed $K4b^2$ to get net EE. EE data were analyzed for the entire PA routine, including transitions, on day one and day two for both the Cosmed $K4b^2$ and the consumer devices. EE data from day one and day two (if available) were added together to obtain a single EE value for each device the participant wore.

All analyses were conducted using IBM SPSS statistics software version 22 (IBM, Armonk, NY). Data are presented as mean ± SD, unless otherwise noted. Equivalence testing was conducted comparing estimated EE from the consumer PA monitors to the measured EE from the Cosmed $K4b^2$. The equivalence zone was defined as the 90% confidence interval (CI) of the Cosmed $K4b^2$ instead of ± 10% of the criterion mean that has been previously used (3,23). This setup was chosen because ± 10% of the criterion mean fell within the 90% CI. Both the 90% CI and ± 10% of the Cosmed K4b^2 are included on the plots for reference. A device was considered equivalent with the Cosmed K4b^2 if the 90% CI of a device falls within the 90% CI of the Cosmed K4b^2. MAPE, relative to the measured EE from Cosmed K4b^2, was also calculated as an additional indicator of measurement error. The equivalence tests and MAPE were split by ages 6-12 years and ages 13-18 years. The analysis was split by age because the Fitbit and Samsung randomization order resulted in most Fitbit cases being in ages 6-12 years and most Samsung cases being in ages 13-18 years.

Results

Physical characteristics of the participants are shown in Table 2. Of the 50 participants, 31 completed both days of testing and 19 completed only day one of testing. Device synchronization error during initialization or download resulted in missing data for: 8 Apple Watch Series 2, 7 Misfit Shine 2 hip, 6 Misfit Shine 2 left shoe, 6 Misfit Shine 2 right shoe, 5 Mymo Activity Tracker, 4 Fitbit Charge 2, and 7 Samsung Gear Fit 2.

The mean total calories ± SD from the Cosmed K4b^2 and the consumer activity monitors can be found in Table 3. For both age groups, none of the devices were equivalent to the Cosmed K4b^2 (figures 1-2). The MAPE for each device, split by age group, can be found in figure 3. The MAPE for ages 6-12 years had a range of 30.1% (Misfit Shine worn on the right shoe) to 59.3% (Fitbit Charge 2). The MAPE for ages 13-18 had a range of 37.4% (Fitbit Charge 2) to 61.8% (Apple Watch Series 2).

Discussion

The primary finding from the current study was that the estimated EE from the newest-generation, consumer PA monitors were not equivalent to the measured EE from the Cosmed $K4b^2$ in youth. The current study also found that, even though none of the monitors were equivalent, that monitors worn at the hip and shoe had less error than the monitors worn on the wrist. Furthermore, all the devices had MAPE of at least 30% regardless of wear location.

Due to the lack of studies examining estimates of EE from consumer devices in youth, direct comparisons of validity are limited to studies using adult populations. The Fitbit Charge HR, a predecessor to the Fitbit Charge 2, had a MAPE of 36% for EE compared to the criterion of measured EE by the Cosmed $K4b^2$ (6). This error is comparable to the MAPE of the Fitbit Charge 2 of 37.4% for ages 13-18 years. The Apple Watch Series 1, the predecessor to the Apple Watch Series 2, had a MAPE of 27% for EE compared to the criterion of measured EE by the Cosmed $K4b^2$ (6). This error is much lower than the MAPE of the Apple Watch Series 2 of 51.0% for ages 6-12 years and 61.8% for ages 13-18 years. The Misfit Shine, the predecessor to the Misfit Shine 2, had a MAPE of 30.4% for EE compared to the measured EE by the Oxycon Mobile (3). This error is comparable to the MAPE of the Misfit Shine 2 of 30.1-32.7% depending on device placement for ages 6-12 years. Though it is difficult to draw comparisons between studies due to differing participant populations and methods used, the Fitbit and Misfit devices both have consistently lower error than the other devices.

Though some consumer activity monitors have been validated in adults, the results of this study indicate these devices are not valid for youth. Despite their growing popularity among the general population, using consumer monitors to estimate EE in youth is not recommended. These devices under- and over-estimate EE and have large individual errors. Consumer's should be aware of these issues when using these devices to provide EE information about their daily PA. However, because these devices have a wide range of other functions (e.g. step counting and distance traveled), further research should be conducted to validate their other uses.

Strengths of this study include the criterion measurement of EE using the Cosmed $K4b^2$, having a large sample indicative of the local population, and having an even distribution of ages ranging from 6-18 years. One limitation of this study was the device placement site. The device placement of the wrist-worn monitors was approximately 2-3" above the location recommended by the manufacturer

The primary finding from this study was that the consumer PA monitors' estimates of EE were not equivalent to the Cosmed $K4b^2$ and had a MAPE ranging from 30-62%. Overall, the Misfit Shine 2, regardless of placement, had the lowest error for ages 6-12 years and the Fitbit Charge 2 had the lowest error for ages 13-18 years. However, because of the errors were still large, it is not recommended that any of these devices be used to estimate EE in youth.

Table 1: Information on consumer physical activity monitors.

	Fitbit Charge 2	Apple Watch Series 2	Samsung Gear Fit 2	Mymo Activity Tracker	Misfit Shine 2
Accelerometer	Yes	Yes	Yes	NA*	Yes
Gyroscope	No	Yes	Yes	NA*	No
Heart Rate Monitor	Yes	Yes	Yes	NA*	No
Built-in GPS	No	Yes	Yes	NA*	No
Sleep Tracker	Yes	No	No	NA*	Yes
Magnetometer	No	No	No	NA*	Yes
Altimeter	Yes	No	No	NA*	No
Barometer	No	No	Yes	NA*	No
Water Resistant	Splash Proof	50 meters	1.5 meters	NA*	50 meters
Battery Life	5 days	18 hours	3-4 days	4-6 months	6 months
Battery Type	rechargeable	rechargeable	200 mAh rechargeable	CR2032	CR2032
Sync Technology	Bluetooth 4.0	Bluetooth 4.0	Bluetooth 4.2	Bluetooth 4.0	Bluetooth 4.1
Compatibility	iOS, Android	iOS	Android, iOS	iOS, Android	iOS, Android
Screen	OLED	Touch OLED	Touch sAMOLED	No	No
Wear Location	Wrist	Wrist	Wrist	Hip	Wrist, Hip, Shoe
Price	$150 USD	$369+ USD	$180 USD	$125 USD	$100 USD
Energy Expenditure (kcals)	Gross	Net	Gross	Gross	Gross
Steps	Total	Total	Total	Total	Total
Distance (mi)	Total**	Walk/Run only	Not Provided	Total**	Total**
Minutes of Exercise	Active***	Active	Not Provided	Active	Not Provided
Flights of Stairs (ascended)	Yes	No	Yes	No	No

OLED, organic light-emitting diodes
sAMOLED, super active-matrix organic light-emitting diode
*Not available from manufacturer.
**Total distance covered during any activity.
***Active minutes only record for bouts longer than 10 minutes

Table 2. Physical characteristics of participants split by gender and age group.

	Male (n=28)		Female (n=22)		Total (N=50)
	6-12 (n=13)	13-18 (n=15)	6-12 (n=7)	13-18 (n=15)	
Age (yrs)	10.5 ± 2.1	15.4 ± 1.2	10.1 ± 2.2	15.7 ± 1.0	13.5 ± 3.0
Height (cm)	142.0 ± 11.2	171.6 ± 9.9	136.9 ± 17.1	163.6 ± 5.9	156.7 ± 17.3
Weight (kg)	35.6 ± 8.0	64.7 ± 16.8	32.2 ± 12.0	62.0 ± 12.6	51.8 ± 19.1
BMI (kg/m^2)	17.4 ± 2.2	21.9 ± 5.1	16.8 ± 3.1	23.3 ± 5.3	20.4 ± 5.0
CDC BMI Percentile (%)	51.5 ± 27.6	52.8 ± 28.0	40.6 ± 33.2	63.8 ± 27.2	54.0 ± 28.5
Overweight (%)	15.4	0.0	14.3	20.0	12.2
Obese (%)	0.0	20.0	0.0	13.3	10.2

Values are mean ± SD.
BMI, body mass index.

Table 3. Mean ± SD energy expenditure (kcals) for ages 6-12 years and 13-18 years.

	6-12 year olds	13-18 year olds
Gross Kcals		
Cosmed	288.3 ± 110.5	452.3 ± 140.6
Fitbit	442.5 ± 150.7	454.9 ± 423.0
Misfit Hip	340.0 ± 158.0	469.3 ± 314.1
Misfit Left Shoe	327.4 ± 146.4	475.8 ± 322.7
Misfit Right Shoe	335.3 ± 144.8	498.0 ± 333.5
Mymo	343.9 ± 153.7	363.3 ± 189.2
Samsung	198.0 ± 77.8	286.8 ± 130.0
Net Kcals		
Cosmed	187.0 ± 82.8	308.1 ± 116.2
Apple	84.5 ± 39.2	140.2 ± 60.2

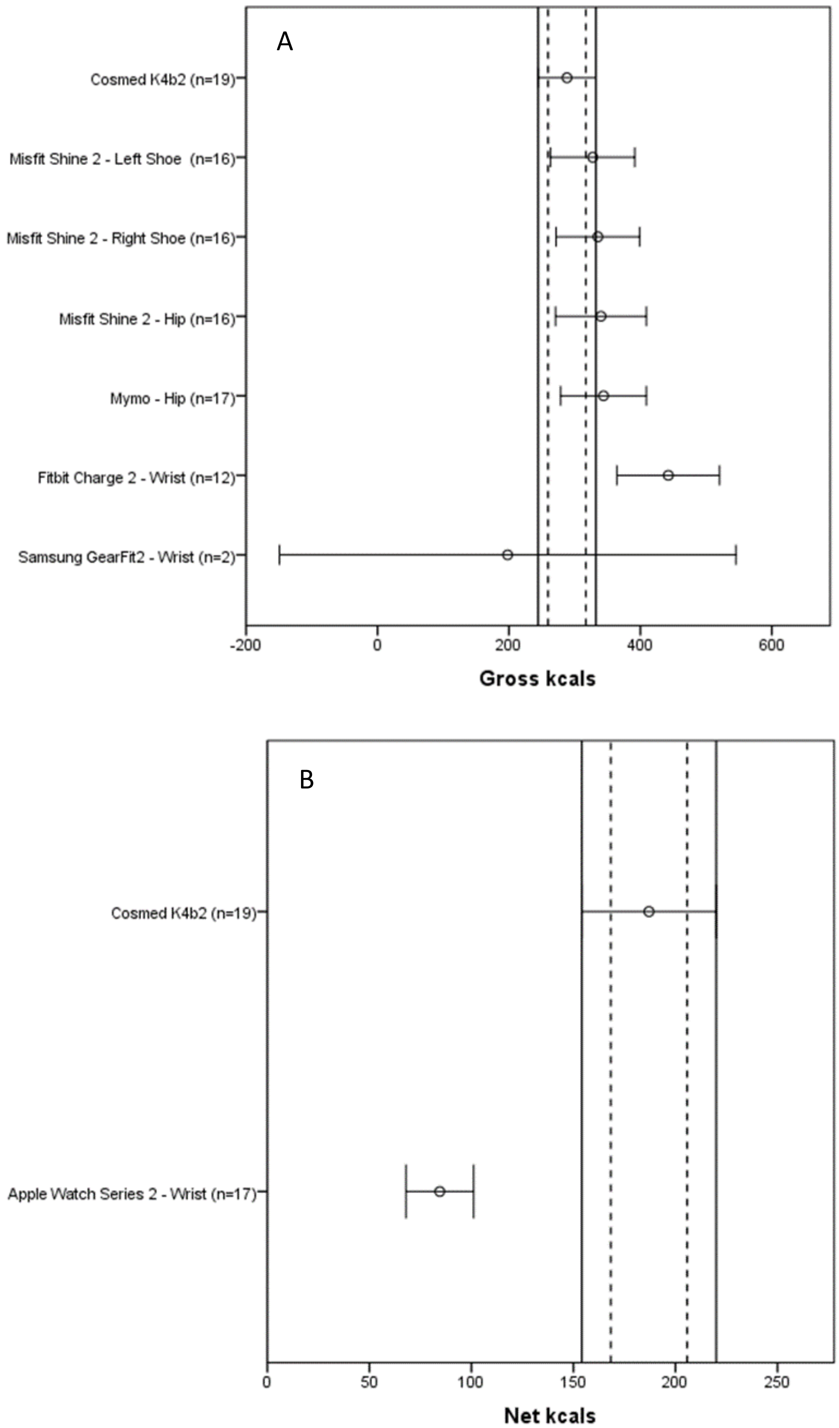

Figure 1. Agreement between the Cosmed $K4b^2$ and A) consumer monitors that provide estimates of gross EE and B) consumer monitors that provide estimates of net EE for participants aged 6-12 years. Dark vertical lines indicate the equivalence zone from the Cosmed $K4b^2$. Dotted vertical lines indicate ± 10% of the mean measured EE of the Cosmed $K4b^2$.

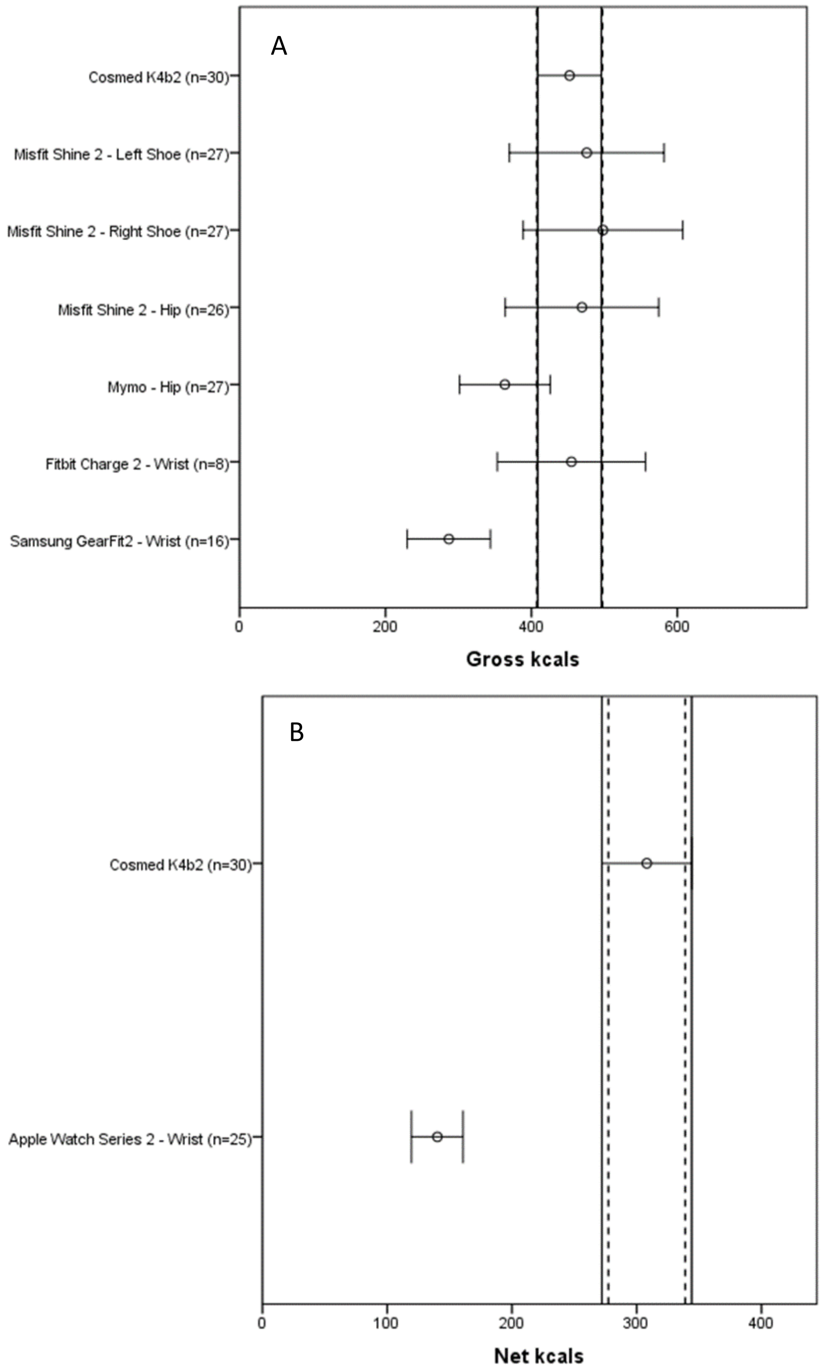

Figure 2. Agreement between the Cosmed K4b^2 and A) consumer monitors that provide estimates of gross EE and B) consumer monitors that provide estimates of net EE for participants aged 13-18 years. Dark vertical lines indicate the equivalence zone from the Cosmed K4b^2. Dotted vertical lines indicate ± 10% of the mean measured EE of the Cosmed K4b^2.

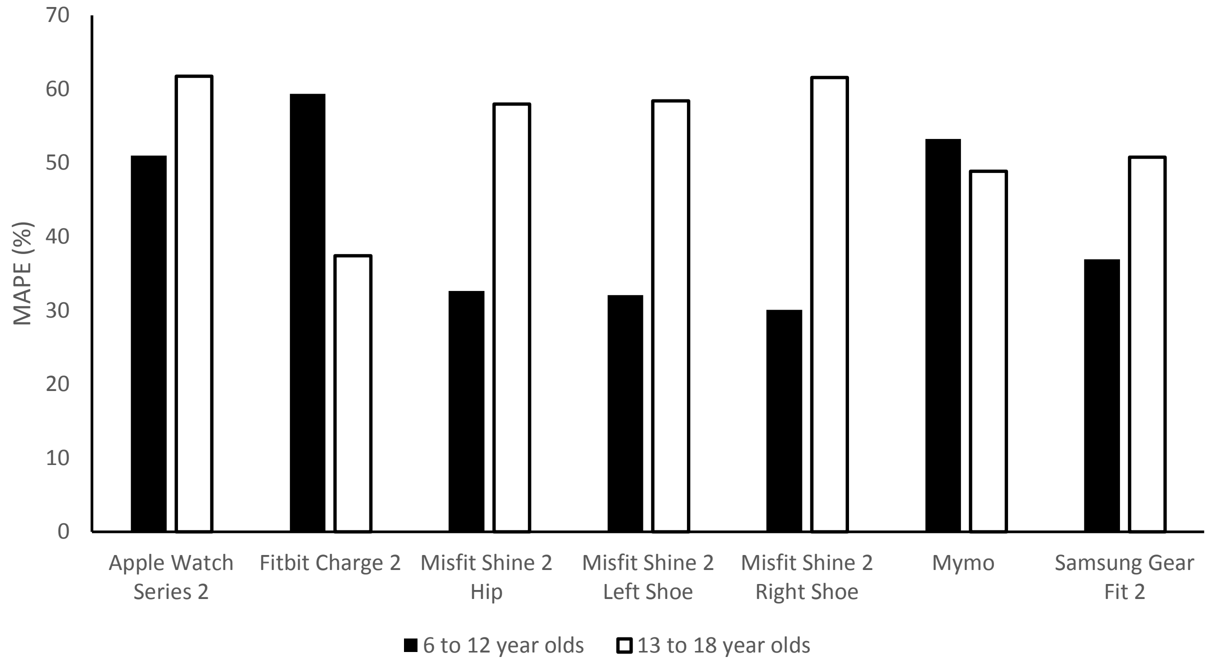

Figure 3. Mean absolute percent error (MAPE) of the consumer monitors compared to the Cosmed K4b^2 for ages 6-12 years and 13-18 years.

References

1. ActiGraph Website [Internet]. Pensacola (FL): ActiGraph GT9X Link; [cited 2017 May 1]. Available from: http://actigraphcorp.com/products-showcase /activity-monitors/actigraph-link/.

2. Axivity Website [Internet]. (United Kingdom): Axivity AX3; [cited 2017 May 1]. Available from: http://axivity.com/product/ax3.

3. Bai Y, Welk GJ, Nam YH, et al. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc*. 2016;*48*(1), 151-8.

4. Brandes M, Van Hees VT, Hannover V, Brage S. Estimating energy expenditure from raw accelerometry in three types of locomotion. *Med Sci Sports Exerc.* 2012;44(11), 2235-42

5. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* 1985;*100*(2), 126.

6. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. *PloS One*. 2017;*12*(2), e0171720.

7. Apple Website [Internet]. Cupertino (CA): Compare Apple Watch Models; [cited 2017 May 1]. Available from: http://www.apple.com/watch/compare/.

8. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. *J Appl Physiol. 2008;105*(3), 977-987.

9. Crouter SE, Flynn JI, Bassett DR. Estimating physical activity in youth using a wrist accelerometer. *Med Sci Sports Exerc*. 2015;*47*(5), 944.

10. Crouter SE, Horton M, Bassett DR. Validity of ActiGraph child-specific equations during various physical activities. *Med Sci Sports Exerc*. 2013;45(7):1403-1409

11. Dannecker KL, Sazonova NA, Melanson EL, Sazonoz ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*. 2013;*45*(11), 2105.

12. Diaz KM, Krupka DJ, Chang MJ, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *Int J Cardiol.* 2015;*185*, 138-140.

13. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthc Inform Res.* 2015;*21*(4), 315-320.

14. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG.Validation of the GENEA Accelerometer. *Med Sci Sports Exerc.* 2011;43(6), 1085-1093.

15. Fitbit Website [Internet]. San Francisco (CA): Fitbit Charge 2; [cited 2017 May 1]. Available from: https://www.fitbit.com/shop/charge2?activeFeature=specs.

16. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. *Med Sci Sports Exerc*. 2005;*37*(11), S523.

17. Ferguson T, Rowlands AV, Olds T, Maher C. 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.* 2015;*12*(1):42.

18. GENEActiv Website [Internet]. Huntingdon (UK): GENEActiv; [cited 2017 May 1]. Available from: http://www.geneactiv.org/using-geneactiv/why-geneactiv/.

19. Harrell JS, McMurray RG, Baggett CD, Pennell ML, Pearce PF, Bangdiwala SI. Energy costs of physical activities in children and adolescents. *Med Sci Sports Exerc*. 2005;*37*(2), 329-36.

20. Harrington DM, Welk GJ, Donnelly AE. Validation of MET estimates and step measurement using the activPAL physical activity logger. *J Sports Sci.* 2011;*29*(6), 627-633.

21. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Med Sci Sports Exerc*, 2014;*46*(9), 1816-24.

22. Janssen I, LeBlanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int J Behav Nutr Phys Act*. 2010;*7*(1), 40.

23. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc*. 2014;*46*(9), 1840-8.

24. Leininger LJ, Cook BJ, Adams KJ. Validation and Accuracy of FITBIT Charge: A Pilot Study in a University Worksite Walking Program. *J Fit Res.* 2016;*5*.

25. Levine JA. Measurement of energy expenditure. *Public Health Nutr*. 2005;*8*(7a), 1123-1132.

26. Maalen-Johansen IK. Investigation of energy expenditure during walking in children with cerebral palsy and typically developing children using raw acceleration data (Master's thesis, NTNU). *NTNU Open.* 2016.

27. McLaughlin JE, King GA, Howley ET, Bassett DR, Ainsworth BE. Validation of the COSMED K4 b2 portable metabolic system. *Int J Sports Med.* 2001;*22*(04), 280-284.

28. Misfit Websit [Internet]. Burlingame (CA): Misfit Shine 2; [cited 2017 May 1]. Available from: https://misfit.com/products/misfit-shine-2.

29. TupeloLife Website [Internet]. (India): Mymo Activity Tracker; [cited 2017 May1]. Available from: http://www.tupelolife.com/mymo/.

30. PALtechnologies Website [Internet]. (Scotland): activPAL; [cited 2017 May 1]. Available from: http://www.paltech.plus.com/ products.htm.

31. Phillips LR, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *J Sci Med Med Sport.* 2013;*16*(2), 124-128.

32. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. *Med Sci Sports Exerc*. 2004;*36*(9), 1625-1631.

33. Ridgers ND, Salmon J, Ridley K, O'Connell E, Arundell L, Timperio A. Agreement between activPAL and ActiGraph for assessing children's sedentary time. *Int J Behav Nurt Phys Act.* 2012;*9*(1), 15.

34. Ridley K, Ainsworth BE, Olds TS. Development of a compendium of energy expenditures for youth. *Int J Behav Nurt Phys Act.* 2008;*5*(1), 45.

35. Rowlands AV, Rennie K, Kozarski R, Stanley RM, Eston RG, Parfitt GC, Old TS. Children's physical activity assessed with wrist-and hip-worn accelerometers.*Med Sci Sports Exerc*. 2014;46(12):2308-16

36. Samsung Website [Internet]. (South Korea): Samsung Gear Fit 2; [cited 2017 May 1]. Available from: http://www.samsung.com/us/mobile/ wearables/smart-fitness-bands /samsung-gear-fit2-small-black smr3600danxar /#benefits.

37. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. ActiGraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med.* 2013;*34*(11), 975-982.

38. Sardinha LB, Andersen LB, Anderssen, SA, Quiterio AL, Ornelas R, Froberg K, Riddoch CJ, Ekelund U. Objectively measured time spent sedentary is associated with insulin resistance independent of overall and central body fat in 9-to 10-year-old Portuguese children. *Diabetes Care*. 2008;*31*(3), 569-575.

39. Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *J Nutr.* 1988;1*18*(11), 1278-1289.

40. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr.* 1984;*39*, 5-41.

41. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med.* 2003;*37*(3), 197-206.

42. Trost SG, Way R, Okely AD. Predictive validity of three ActiGraph energy expenditure equations for children. *Med Sci Sports Exerc*. 2006;*38*(2), 380.

43. U.S. Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans*. Rockville, MD: U.S. Department of Health and Human Services, Office of the Surgeon General; 2008. Available from: U.S. GPO, Washington.

44. Wickel EE, Eisenmann JC, Welk GJ. Predictive validity of an age-specific MET equation among youth of varying body size. *Eur J Appl Phys*. 2007;101(5), 555.