PHYSICAL ACTIVITY PARTICIPATION AND OBESITY IN U.S. YOUTH: A MIMS EQUATING STUDY

BY

XIONG QIN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Kinesiology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Weimo Zhu, Chair
Professor Xiaotian Gao
Professor Ian Rice
Professor Synthia Sydnor

ABSTRACT

Physical activity (PA) is associated with many health benefits. To better understand PA participation and health-related fitness of U.S. children and youth, the National Health and Nutrition Examination Survey (NHANES) National Youth Fitness Survey (NNYFS) tracked free-living PA of children in U.S. using ActiGraph in 2012. NNYFS also conducted a set of fitness tests, including cardiorespiratory endurance, muscular strength and endurance, flexibility and body composition. However, PA data was not released using ActiGraph Count, or simply the Count, an established unit with intensity cutoff scores, but rather in a new unit called “Monitor-Independent Movement Summary (MIMS).” While MIMS provides versatility as it can be used with any brand of accelerometer, it has not been validated with criterion measures, nor the cutoff scores for PA intensities were set up.

This dissertations study consists of three studies and the purposes of these studies were to link MIMS with the ActiGraph Count, to validate the link and to determine the PA participation of U.S. children and youth by applying the link to the NNYFS data, respectively.

Specifically, Study 1 used test equating to derive a transformation link between MIMS and the Count by using a set of ActiGraph data, in which both MIMS and the Count could be calculated, and to create an estimate for the Count called “MIMS-EQ-Count”, where “EQ” represents “equated”. Study 2 used the indirect calorimeter data from the same data of Study 1 to determine the equivalence of MIMS-EQ-Count. Finally, Study 3 used the link developed in Study 1 and verified by Study 2 to transform the MIMS data in NNYFS to MIMS-EQ-

Count and to further evaluate the PA intensity, as well as to examine the relationship between body composition and PA participation.

As a result of this study, MIMS was successfully linked to the Count by using the equipercentile test-equating method and creating a table of transformation. Moreover, Study 2 proved that MIMS-EQ-Count and the Count have the similar relationship with VO_2, indicating that MIMS-EQ-Count could be a valid substitute for the Count. Finally, Study 3 found that the time in moderate-vigorous PA (MVPA) of U.S. children and youth decreased as age increased, but no clear differences were found between sex, as well as racial and BMI categories. This result contradicts the conclusions of previous studies that obese children and youth are less active. In addition, Study 3 found that U.S. children and youth were rather physically active and the rates to meet the PA recommendation, i.e., 60 minutes per day, were all close to or at 100%. This is clearly too high to be true and the supervising finding is likely caused by wearing accelerometers on wrist and the “sedentary active behavior”, such as when children move their wrists vigorously while playing video games, was mistakenly recognized as MVPA.

In conclusion, this dissertation study successfully found a way to transform MIMS into the Count. However, due to the test site complications and the MIMS algorithm, the NNYFS data is not applicable yet as future studies need to focus on the methodology for evaluation of data from wrist-worn accelerometers.

Keywords: Intensity, linking, validation, sedentary-active

ACKNOWLEDGEMENTS

I would like to thank the following individuals who made this dissertation possible:

To my academic advisor, Professor Weimo Zhu, who not only provided me with knowledge of the fields of Kinesiology and Kinesmetrics, but who also guided me on the path to becoming a stronger researcher.

To the professors in my committee: Professors Xiaotian Gao, Ian Rice, and Synthia Sydnor, who assisted me with plenty of valuable suggestions.

To Professor Lin Zhu at Guangzhou Sports University, China, who shared valuable data.

To my parents for their endless love and unconditional support in the pursuit of my dream.

To my friend, Siyuan Guo, who helped me with the mathematical obstacles.

TABLE OF CONTENTS

Chapter 1

Introduction

This chapter provides an introduction about the background of this study. First, the concept of physical activities (PA) will be introduced, including its definition, categorization and the dose components. Second, its relationship with health and obesity will be described, including the health benefits of PA, definition and measures of obesity, prevalence of obesity and effect of PA on obesity. Third, methods of measuring PA will be presented, and the evaluation through ActiGraph, a measurement device, will also be covered. This chapter will then discuss the relationship between PA and obesity, PA assessment and fitness assessment in children and youth. Further, this chapter will address the limitations of current national surveys in assessing PA and will propose the potential methods to solve the problems. Finally, the purpose, specific aims, corresponding hypotheses, and the plan of this study will be stated at the end of this chapter.

PA

PA is defined as any type of human body movement caused by the contraction of skeletal muscles resulting in energy expenditure (EE; The World Health Organization [WHO], 2020). All spontaneous activities in daily life could be considered as PA, e.g., occupational, sport participation, home and family care, transportation, leisure activities, and exercising etc. The most popular exercises are walking, running, cycling or conditional training (Ball et al., 2014).

Category of PA

PA could be categorized in multiple ways, usually by activity aim and fitness type:

By Activity Aim. PA includes household, occupational, conditional, sports or other (Caspersen et al., 1985).

Household and occupational PA are the activities aiming at accomplishing house works or works, such as gardening, cooking, transporting, occupational activities, instead of improving physical condition. Contracting skeletal muscles and consuming energy are only by-products. Today, the PA from work and housework becomes less due to the help of modern technology (Woessner et al., 2021), e.g., floor-sweeping robots replacing brooms, online news replacing newspaper, video games replacing real-life games.

Conditioning, known also as exercise, is "planned, structured, repetitive and purposive" PA designed for improving one or more aspects of physical fitness, including aerobic capacity, muscular strength and endurance, flexibility, etc. (Caspersen et al., 1985). Sport is a competitive PA. Some sports are also exercises when they are used for "planned, structured, repetitive and purposive" PA (Caspersen et al., 1985).

By Fitness Types. In exercises, people purposely employ specific activities to improve one or more fitness components. Fitness could be generally divided into two groups: Health-related and skill-related fitness. Since this study will focus on health, only health-related fitness will be described and used. Health-related fitness includes five components: Cardiorespiratory endurance, known also as aerobic capacity, muscular strength and endurance, body composition and flexibility (Caspersen et al., 1985). Exercise varies given different training goals. Exercises aiming to increase aerobic capacity are aerobic exercises, while strength exercises help with muscular strength and stretches improve the flexibility. Aerobic, strength and flexibility exercises are the major types of exercises practiced for health to increase different components of fitness. Some exercises focus on one aspect, e.g., running for aerobic, lifting for muscular strength, while other exercises cover multiple aspects, e.g., Taekwondo exercises could improve both muscular strength and flexibility.

Dose of PA

Due to lack of PA is one of the major health risk behaviors today, PA is becoming important and necessary. The dose of PA typically consisted of intensity, duration, frequency, and type. Intensity, especially for aerobic activities, is the rate at which energy is expended during PA participation, or in another words, how vigorously is the person performing PA. Intensity is often defined by volume of oxygen consumption and is often classified in the following categories: Sedentary, light, moderate, vigorous and very vigorous, with EE rate ranked from low to high, respectively. Duration refers to the period of PA. In some exercises, e.g., strength training, individuals use number of repetitions and sets, instead of time, to quantify the exercises, as it still establishes a period of activity. Frequency indicates how many times a day or week a person does exercise. Proper frequency gives needed resting time, preventing over-training or over-resting. Some of the commonly used frequencies are: Every day, every other day, or times a week. The type refers to the kind of exercise, activity or sport a person does. Common types for conditional training are: Aerobic, strength, flexibility, or weight-bearing exercises, etc. More detailed sub-types could be further applied, e.g., running, jogging, cycling, swimming, etc. in aerobic exercise.

PA and Health

Health is “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (WHO, 2006). PA has been confirmed to have substantial health benefits according to the 1996 Surgeon General’s Report on PA and Health (U.S. Department of Health and Human Services, 1996) including not only increased physical fitness and health (Blair et al., 1989; Janssen & LeBlanc, 2010), but also decreased mortality (Blair et al., 1995). Therefore, increasing PA would suggest a decreased risk for diseases and health conditions caused by physical inactivity (Pate et al., 1995). Because of these health benefits of PA, the U.S. Department of Health and Human Services published the first edition of *Physical Activity Guidelines for Americans* in 2008, and released the second edition in 2018, which provided specific dose recommendations of PA to every age group. The guideline played a key role in national PA policies (U.S. Department of Health and Human Services, 2018b).

Unfortunately, many people are still lacking PA, and sedentary behavior is one of the most popular unhealthy behaviors for the twenty-first century (Zhu & Owen, 2017). In all the states of U.S., for example, more than 15% of the people are physically inactive, with 7 states and 2 Territories have more than 30% people physically inactive (CDC, 2020a). As a result of lack of PA, many people suffered from both physical and mental health problems where obesity is one of the major ones.

Defining and Measuring Obesity

Obesity is a condition with excessive accumulation fat in the body. The commonly used methods for measuring obesity include dual X-ray absorption (DXA), bioimpedance, skinfold, and body mass index (BMI). DXA uses the absorption of energetic high-frequency electromagnetic radiation to measure body fat percentages, and it is considered as the criterion measurement of obesity since it has the highest reliability and validity. Bioimpedance method assesses body fat percentage by measuring the electrical current passing through the body. Bioimpedance could also provide the estimates of muscle mass, water percentage, and waist-to-hip fat ratio. The skinfold method uses the thickness of pinched skin as an index of body fat percentage. Thicker skinfold corresponds to higher body fat. Skinfold is normally measured at the triceps, calves, and abdominal or subscapular areas. The result is a variable of thickness that can be transformed into body fat percentage through prediction formulas (Slaughter et al., 1988). BMI is calculated by:

$$\mathrm{BMI} = \frac{Body\ weight\ in\ kg}{(Height\ in\ m)^2}, \qquad \text{(Formula 1.1)}$$

where “kg” represents kilogram and “m” represents meter. BMI is the most feasible means to estimate the extent of obesity.

Obesity Based on DXA Classification. Pasco et al. (2014) provided the cutoff scores of body fat percentage, measured by DXA to classify underweight, healthy weight, overweight and obese by age and sex among adults and old adults (see Table 1.1).

Table 1.1

Cutoff scores of body fat percentage (%BF) in adults and older adults

Age	Underweight		Healthy weight		Overweight		Obese	
	Men	Women	Men	Women	Men	Women	Men	Women
20–29	≤9.3	≤21.8	≤20.6	≤36.0	≤27.5	≤43.4	>27.5	>43.4
30–39	≤9.9	≤22.0	≤21.2	≤36.2	≤28.1	≤43.6	>28.1	>43.6
40–49	≤10.5	≤22.2	≤21.8	≤36.4	≤28.7	≤43.8	>28.7	>43.8
50–59	≤11.1	≤22.4	≤22.4	≤36.6	≤29.3	≤44.0	>29.3	>44.0
60–69	≤11.7	≤22.6	≤23.0	≤36.8	≤29.9	≤44.2	>29.9	>44.2
70–79	≤12.3	≤22.8	≤23.6	≤37.0	≤30.5	≤44.4	>30.5	>44.4
80–89	≤12.9	≤23.0	≤24.2	≤37.2	≤31.1	≤44.6	≥31.1	>44.6

Note: This table is from Pasco et al. (2014).

Cutoff scores of body fat percentage among children, teenagers and young adults from 5 to 19 years old have been provided by Laurson et al. (2011). This set of cutoff scores has been employed by FitnessGram®, a U.S. leading program of children and youth fitness assessment by the Cooper Institute.

Obesity Based on BMI Classification. In adults, the classification of obesity by BMI is: Underweight (BMI < 18.5), normal (BMI of 18.5 to < 25), overweight (BMI of 25 to < 30), and obesity (BMI >= 30) (CDC, 2021). Meanwhile, in children, the weight status is defined by age and sex specific BMI percentiles: Underweight (< 5th percentile), normal (5th to < 85th percentile), overweight (85th to < 95th percentile), and obese (> 95th percentile) (CDC, 2018).

Prevalence of Obesity

In 2017-18, based on the BMI data, the prevalence of U.S. adults' overweight and obesity rates were at 30.7% and 42.4%, respectively, which showed a significant increase from 1988-1994 when the corresponding rates were at 33.1% and 22.9%, respectively (Fryar et al., 2020a). Obesity is not only just an abnormal body composition, but also a trigger to many other health problems, including negative effects on individuals' physical and mental health (Cornette, 2011; Must et al., 1999).

PA and Obesity

As has been clearly stated by WHO (2020a), due to physical inactivity and an increased intake of energy-dense food, the essential causation of obesity is the unbalanced calorie intake and expenditure. More than half century ago, research already identified physically inactive is a risk factor of obesity (Chiriko & Stunkard, 1960). Dietz (2004) stated that physical inactivity is a robust predictor of obesity, as studies found that more TV time is correlated with higher obesity extent. Aerobic exercise with moderate to high intensity has effects on weight loss, and if exercise combines with diet restriction, the weight-loss effect will be better than diet alone (Chin et al., 2016). Due to the fact that excessive energy will lead to weight gain and more EE will reduce the weight gain, more PA always have effect on weight control.

Measuring PA

Knowing the level of PA participation helps build a dose-response relationship and enables intervention (Blair et al., 2004) and implement PA guidelines. To do so, all elements of PA dose, i.e., intensity, duration, frequency, and type, need to be measured and quantified. The type of PA and frequency could be observed or recalled, but the intensity and duration are more complicated and need some efforts to measure. Methods to measure PA could be classified as criterion and field measures, and the latter can be further classified as subjective and objective measures.

Criterion Measures

Criterion measures have high validity and reliability, but requiring more funding, time, and technological support. Two commonly used criterion measures are doubly labeled water (DLW) and indirect calorimeters.

The DLW method is based on the principle that after drinking a dose of DLW, $^{2}H_2\,^{18}O$, the two isotopes equilibrate with total body water (TBW) and then are eliminated differentially from the body. Deuterium (^{2}H) leaves the body as water, while ^{18}O leaves as water (H_2O) and carbon dioxide (CO_2). Therefore, CO_2 production can be calculated by subtracting ^{18}O elimination from ^{2}H elimination. The EE could be calculated based on the amount of CO_2 product.

DLW is typically used as a criterion measure of EE for PA in free-living conditions (Ekelund et al., 2001; Plasqui & Westerterp, 2007). During the assessment, participants take DLW that is labeled by isoforms of hydrogen and oxygen atoms in the water molecule. By analyzing the urine samples, the researchers are able to calculate respiratory CO_2, which could be further converted into estimated EE with standardized formula (Speakman et al., 2021). However, DLW method is limited due to its higher requirement of equipment, funding, time, and participant compliance. Furthermore, it only provides total EE and it is not readily available for most laboratories (Speakman, 1998).

Indirect calorimeters use a mask to measure gas metabolism during PA (Duffield et al., 2004) and is portable and can provide instant measurement without lengthy preparation. It measures the amount and component of expended air to further estimate consumption of oxygen and the intensity of exercise with a time stamp. The indirect calorimeter has been widely used in PA assessment studies, especially in the calibration and cutoff scores setup for objective measures, e.g., Santos-Lozano et al. (2013).

Field Measures

Field measures are the measurements that could be done in a non-laboratory setting. The methods of field measures could be based on recall, physiological response or movement of PA participations and are easy to set up and execute. The validity and reliability of the filed measures, however, varied from the method to method. In general, the field measures can be classified as subjective and objective measures:

Subjective Measures. Subjective measures quantify PA based mainly on recall or observation. The judgment could be made by participants or observers. Two commonly used formats of subjective measures are questionnaires and direct observations. Questionnaires ask participants to recall their daily PA, often including intensity, time, frequency and type. While questionnaires are straightforward and inexpensive, they are low in validity (Troiano et al., 2008) because people tend to report more than what they really did in PA. In comparison, direct observations use direct observations or videotapes to record children's behavior during activity and classify the intensity of each child's PA for every epoch (Pate et al., 2008; McKenzie, 2016). This method eliminates the subjectivity of participants, but introduces subjectivity and scoring burden of the observers.

Objective Measures. Objective measures quantify PA by physiological and physical variables. Their outcomes provide information on physiological or movement aspects. Commonly used objective measures include pedometers, heart rate monitors, accelerometers and multi-sensor devices.

The pedometer, which measures and records step counts, is an objective PA measure with a long history (Bassett et al., 2017; Tudor-Locke et al., 2005; Tudor-Locke et al., 2009). Some guidelines specify the number of steps required every day and the measurement of steps is critical (U.S. Department of Health and Human Services, 2018c). Total step count represents duration and step per minute represents intensity (Tudor-Locke et al., 2005). Most of current pedometers use accelerometers to grasp additional step related information, e.g., step intensity (Saint-Maurice et al., 2020).

Heart rate monitor is a type of wearable device that measures heart rate instantaneously. They are widely used as a means to determine PA intensity (Karvonen & Vuorimaa, 1988). Two types of heart rate monitors are widely adopted: The electrocardiogram (ECG) based and light based. The Polar chest strap, an ECG-based heart rate monitor, has similar accuracy to hospital devices (Kingsley et al., 2005). Meanwhile, light based monitors use the change of reflected light from the capillaries to infer heartbeats. Heart rate monitors are usually manufactured as a wristband or armband. However, its validity is not as high as ECG-based monitors (Cosoli et al., 2020).

Accelerometer measures the acceleration. Once is attached to human body, the accelerometer could accurately measure the acceleration of the attached body part. Often used body parts including wrist, forearm, upper-arm, waist, hip, and ankle, but the hip was the most frequent one in PA studies. Meanwhile, the wrist site has become more popular in commercial devices because of the convenience. Raw data captured by accelerometers includes acceleration time-series data, including 3 axes, and can be further used to estimate EE (Puyau et al., 2002) and exercise intensity (Freedson et al., 2005). Accelerometers were also used to recognize the type of PA through pattern recognition techniques (Maheedhar et al., 2016).

Finally, multi-sensor devices, often attached on armband or wristband, are a combination of two or more movement and physiological functions, which not only could include step count or heart rate functions, but also involve environmental sensors, e.g., the Global Positioning System (GPS), barometer, ultra violet index (UVI) sensors, or physiological sensors, e.g., skin temperature (Reinerman-Jones et al., 2017). Some PA trackers are embedded with multiple sensors and have acceptable validity in tracking daily PA (Evenson et al., 2015). In common available multi-sensor devices, accelerometer is widely used and embedded, playing an important role.

ActiGraph-related Studies

Although other accelerometers are available, ActiGraph, a device by ActiGraph LLC., is utilized most in the past research studies (McClain et al., 2010; Montoye et al., 2020). Two *Medicine and Science in Sports and Exercise* (MSSE) supplements (MSSE 2005 Nov; 37 (11 Supplement) MSSE & 2012 Jan; 37 (1 Supplement)) were, in fact, devoted specifically for ActiGraph. A search of “ActiGraph” in the PubMed database on August 28, 2021 showed more than 4,000 results (Web of Science and Scopus > 3500). As an outcome unit, the

ActiGraph Count, or simply “the Count”, is calculated from raw acceleration data captured by ActiGraph. The algorithm, computing raw accelerometer data into the Count, was developed by the ActiGraph Company to quantify the amount of PA a person completed in a certain epoch. However, the algorithm has not been released to the public. In general, higher Count values represent more PA, and vice versa (ActiGraph LLC., 2008). Furthermore, with in a fixed-length epoch, higher Count value represents higher intensity. Scores of the Counts per min (CPM) between different intensity categories, i.e., sedentary, light, moderate, vigorous and very vigorous, have been set up in many different age groups (Butte et al., 2014; Freedson et al., 1998; Freedson et al., 2005; Matthew, 2005; Montoye et al., 2020; Pate et al., 2006; Santos-Lozano et al., 2013; Welk, 2005). Starting in 2003, ActiGraph and its measurement unit “the Count” have been employed in U.S. national surveillance studies, e.g., National Health and Nutrition Examination Survey (NHANES). One drawback is the inconsistency in setting the intensity cutoff scores, making it difficult to compare the research findings across studies (Freedson et al., 2005).

At any rate, with the help of the Counts, ActiGraph has been employed into broad PA applications, including measuring PA intensities (Troiano et al. 2008), PA EE (Plasqui et al., 2013; Plasqui & Westerterp, 2007), sedentary time and moderate-vigorous PA (MVPA) time (Migueles et al., 2021), sleeping time (Monk et al., 1999) and to correlate PA measures with cardiovascular disease risks (Alhassan & Robinson, 2008).

ActiGraph Cutoff Score of Intensity

Intensity is the energy-expending rate. The more intense the PA is, the faster a person expends energy. Many variables, such as VO_2, MET, heart rate, could serve as the index of intensity independently, giving multiple means to measure intensity.

Intensity Indexes. VO_2 is the abbreviation of Volume of Oxygen (O_2), indicating the volume of oxygen consumed per time unit, e.g., per minute, or per time-weight unit, e.g., per minute per kilogram (kg). Higher VO_2 represents higher intensity.

MET refers to "metabolic equivalent" and one MET is the EE rate while a person is sitting quietly (Ainsworth et al., 2000). MET value is the ratio of instant EE to rest EE (Ainsworth et al., 2011). It is also used to originally define intensity: Sedentary = 1 to 1.5 METs, light = 1.6 to 2.9 METs, moderate = 3.0 to 6.0 METs, vigorous = 6.1 to 9.0 METs, and very vigorous > 9.0 METs (Ainsworth et al., 2011; Santos-Lozano et al., 2013). MET value is also used to describe average EE rate of a specific activity, e.g., brisk walking, swim for recreation, jogging etc. (Ainsworth et al., 2011).

Heart rate, usually in a percentage of maximum heart rate (HRmax) or heart rate reserved (HRR), is also used to define intensity. HRmax is the maximum heart rate a person can get when exercising extremely hard. It could be measured by specific tests or estimated by the formula:

$$HRmax = 220 - age, \qquad \text{(Formula 1.2)}$$

(Karvonen & Vuorimaa, 1988). HRR is the reserved part of heart rate. A target heart rate is normally higher than HRrest but lower than HRmax, which will only use a certain proportion, n%, of the HRR. HRmax and resting heart rate (HRrest) are needed to calculate target heart rate:

$$HRR = HRmax - HRrest, \qquad \text{(Formula 1.3)}$$

$$Target\ heart\ rate = HRR * n\% + HRrest, \qquad \text{(Formula 1.4)}$$

For example, if a person is 30 years old and HRrest is 65, then his/her target hear rate of lower bound of moderate intensity, i.e., 40% HRR, is calculate as follow:

$$HRmax = 220 - age = 220 - 30 = 190,$$

$$HRR = HRmax - HRrest = 190 - 65 = 125,$$

$$Target\ heart\ rate = HRR * n\% + HRrest = 125 * 40\% + 65 = 115.$$

Since a heart rate monitor is easy to wear and could provide heart rate values in real time, it is often used in daily life practice.

Rating of perceived exertion (RPE; scale from 6, very light, to 20, very hard) can also be used to estimate intensity based on a person's perceived intensity (Borg, 1982).

The criterion for different intensities by different variables is shown in Table 1.2.

Table 1.2

Criterion of intensities in different indexes

	MET	% VO_2max	% HRR	% HRmax	RPE
Sedentary	1.0–1.5	<37	<30	<57	<9
Light	1.5–2.9	37–45	30–39	57–63	9–11
Moderate	3.0–5.9	46–63	40–59	64–76	12–13
Vigorous	6.0–8.9	64–90	60–89	77–95	14–17
Very Vigorous	≥9.0	≥91	≥90	≥96	≥18

Note: Ainsworth et al. (2011); Santos-Lozano et al. (2013); American College of Sports Medicine (ACSM; 2018).

ActiGraph Cutoff Scores. Extensive efforts have been made to link the Counts with intensities. Plenty of studies adopted or reviewed the methods of setting cutoff score on the

Count, mostly CPM. A common approach is to use oxygen consumption as a criterion, either to develop a linear regression between oxygen consumption and the Count and thereafter to set up the cut-off scores (Puyau et al., 2002; Treuth et al., 2004) or to use a receiver operating characteristic (ROC) curve to determine the optimal cutoff scores directly (Evenson et al., 2008). Table 1.3 summarizes a few selected example studies of setting cutoff scores.

Table 1.3

Summary of selected calibration studies

Author (Year)	Age (yr.)	*N*/male	Method	CPM Range	Counts by Intensity Upper Bounds			
					S	L	M	V
Santos-Lozano et al. (2013)	12–16	31/19	ROC	NR	NR	2114	6548	11490
Freedson et al. (2005)	6–16	100/14	Reg	NR	NR	1400	5700	10000
Freedson et al. (1998)	23.9±4.0	50/25	Reg	1800-13800	NR	1952	5725	9499
Butte et al. (2014)	3–5	50/25	ROC	0-13077	820	3908	6112	NR
Puyau et al (2002)	6–16	26/14	Reg	NR	800	3200	8200	NR
Treuth et al (2004)	13–14	74/0	ROC	7856	100	2998	5198	NR
Evenson et al. (2008)	5–8	30	ROC	NR	100	2292	4008	NR

Note: CPM = count per minute; Reg = regression; NR = not reported; S = sedentary; L = light; M = moderate; V = vigorous.

PA and Childhood Health-related Fitness

PA has been proved to have positive impact on health-related fitness among children, i.e., aerobic capacity, muscular strength and endurance, flexibility and body composition.

Aerobic Capacity, Muscular Strength and Endurance, Flexibility

Aerobic capacity is the ability of using oxygen to provide energy for movement. Aerobic capacity could be indexed by maximal oxygen uptake (VO_2max), the maximum amount of oxygen that a person can utilize during intense or maximal exercise. Higher VO_2max represents higher aerobic capacity of a person and $\%VO_2max$ was used to define the intensity. In practice, fitness tests such as 1-mile run/walk, Progressive Aerobic Cardiovascular Endurance Run (PACER) test are used to estimate aerobic capacity. Aerobic capacity is observed significantly higher in children with higher habitual PA and less sitting and TV (Pate & Ross, 1987). Interventions of more PA also had significant effect on increasing aerobic capacity (Braun et al., 2017). However, only doing more walking was found not as effective (Morrow & Freedson, 1994), therefore, to improve aerobic capacity, PA with certain intensity is needed.

Muscular strength is the ability of muscle providing large force (DeSimone, 2016). Depending on the muscle groups involved, a specific test can be used to measure muscular strength, e.g., upper-body strength could be measured by bench press (Baker & Newton, 2004), push-ups (van den Tillaar & Ball, 2020), Young Men's Christian Association (YMCA) bench press (Ronai, 2020), and pull-ups tests (Baker & Newton, 2004). Lower-body strength could be measured by squat (Robbins et al., 2012) or leg extension (CDC, 2020b). Core strength could be estimated by curl-up test (Fry et al., 2015) and grip strength can be measured by a hand dynamometer. Strength training has a positive impact on body strength. Vigorous PA could also be helpful to develop children's lower body strength (Moliner-Urdiales et al., 2010).

Muscular endurance is the ability of muscle on reoccurring muscle contraction without rest and fatigue (DeSimone, 2016). In practice, muscular endurance is often measured together with muscular strength using the same tests like push-ups, modified pull-ups, curl-ups, plank tests, etc. Meeting the PA guidelines will give youth a higher possibility to have muscular endurance (Morrow et al., 2013).

Flexibility is the range of motion of different joints (Corbin & Noble, 1980). Sit-and-reach is one of the mostly used, e.g., in FitnessGram®, flexibility tests for back and hamstring flexibility (Baltaci et al., 2003; Wells & Dillon, 1952). Stretching exercises, combining different stretching moves, have a great impact on youth's flexibility (Kamandulis et al., 2013).

Body Composition

Body composition is the status of how much body fat, muscles and bone are in the body. Depending on the numbers of components employed, e.g., 2, 3 or 4 components, different models have been developed to estimate body composition. Usually, the body fat percentage is the key interest of body composition. Obesity is one of the unhealthy body-composition statuses among children. There is also prevalence of obesity in both children and adolescents in both U.S. and worldwide. For example, the obesity rates from 2 to 19 years old in U.S. increased from 5.2% in 1971–74 to 19.3% in 2017–18 and the overweight rate increased from 10.2% in 1971-74 to 16.1% in 2017-18 (Fryar et al., 2020b).

PA helps to fight against childhood obesity by reducing the sedentary time (Prentice-Dunn & Prentice-Dunn, 2012), lowering the risk of obesity (Hong et al., 2016), controlling weight (Wier et al., 2001), losing weight (Bulbul, 2020) and lowering the probability to remain overweight (Dhar & Robinson, 2016). So far, however, only a few large-scale studies (Moliner-Urdiales et al., 2009) examined the impact of PA on childhood obesity by using objective PA measures.

Other PA Benefits and PA Guidelines for Children

In addition to fight against obesity, regular PA offers many other benefits to children and youth, e.g., improve fitness (Janssen & LeBlanc, 2010) and cognitive function (Lubans et al., 2016), and reduce their risk of developing chronic diseases in late adulthood (Fernandes & Zanesco, 2010). Therefore, PA is critical to children and youth's health.

Because of PA's critical role to children and youth's health, governments and organizations have set guidelines to keep children and youth more active. For instance, in 2018, PA Guidelines for Americans suggested that a school-aged children or youth need 1-hour PA daily, which should include aerobic, muscle-strengthening and bone-strengthening exercises (U.S. Department of Health and Human Services, 2018a). WHO recommended that children and youth age 5 to 17 years old need to do, on average, 60 minutes PA a day at a moderate to high intensity level for at least three days a week. Sedentary behavior and screen time need also being limited (WHO, 2020b). Finally, ACSM also recommended 60 minutes of exercise per day and noted that at least three days a week needed to be at vigorous intensity level for the six to twelve years' age group (ACSM, 2019b) and for teenagers (ACSM, 2019a). Table 1.4 summarizes the guidelines.

Table 1.4

Dose recommendations for children and youth from 2018 PA Guideline, WHO and ACSM

	Intensity	Duration	Frequency	Type
PA Guideline 2018	NS	60 minutes	Everyday	Aerobic, muscle and bone-strengthening
WHO	MVPA	60 minutes	3 days a week	Aerobic, muscle and bone-strengthening
ACSM (teen)	NS	60 minutes	Everyday	NS
	Vigorous	60 minutes	3 days a week	
ACSM (Child)	NS	60 minutes	Everyday	NS
	Vigorous	60 minutes	3 days a week	

Note: NS = not specified.

Based on the Youth Risk Behavior Surveillance System (YRBSS) survey in 2019, however, less than half of U.S. high school youth met the guidelines of PA: Only 23.2% met the recommendation of 60 minutes daily PA, 49.5% did muscle strengthening exercises no less than 3 times a week and 16.5% met both aerobic and strength exercise recommendations (Merlo et al., 2020). Meanwhile, the statistics above have been questionable since the validity of the questionnaire employed by YRBSS was not high (Troiano et al., 2008). Objective measure of PA thus is necessary to determine the true PA participation of children and youth in U.S.

Past Efforts of Assessing U.S. PA and Fitness in Children and Youth

To better know PA of children and youth and build dose-response relationships, accurate assessment of free-living PA is very important. Measuring PA could not only give individuals feedback about their PA participation, but also could gradually build the dose-response relationship.

The PA methods used for adults have also been applied to assess the PA status in U.S.

children and youth. For example, the criterion measures of both DLW (Ekelund et al., 2001) and indirect calorimeters (Puyau et al., 2002; Treuth et al., 2004) had been used to calibrate accelerometers for children and youth's PA intensity cutoff scores. Subjective measures, i.e., questionnaires, had been used for many times in the surveys of U.S. children and youth's free-living PA participation. In fact, YRBSS was the only one that had been used for U.S. youth PA surveillance before 2012, although its validity was low (Clark et al., 2011; Troiano et al., 2008). Finally, efforts have been made to use objective measure on children and youth's PA participation, including Pedometers (Tudor-Locke et al., 2009, 2010), heart rate monitor (Epstein et al., 2001) and accelerometers (Trost et al., 2011; Wolff-Hughes et al., 2014).

Fitness of children and youth has been priority of any nation. Extensive efforts have been made in U.S. to develop youth fitness tests and corresponding standards since 1950s (Morrow et al., 2009). Over years, there has been shifting from skill-related fitness to health-related fitness and from norm-referenced standards to criterion-related standards. Well-established youth fitness tests or test batteries include the President's Challenge, FitnessGram®, and YMCA Youth Fitness Test. Corresponding to FitnessGram®, ActivityGram® was also developed by the Cooper Institute, but it was not widely used. To assess the fitness status in U.S. children and youth, National Children and Youth Fitness Study (NCYFS) in 1985 and NCYFS II in 1987 were conducted for 10–17 and 6–9 years old groups, respectively (Ross & Gilbert, 1985; Ross & Pate, 1987). The NCYFS studies also collected data of PA through questionnaire and the tested fitness components included aerobic capacity, muscular strength and endurance, flexibility and body composition. However, after NCYFS II in 1987, no national study has been conducted to study U.S. children and youth PA participation, their fitness status and the relationship between PA and fitness in this population for more than two decades.

A Welcomed New Effort

Finally, the NHANES National Youth Fitness Survey (NNYFS; CDC, 2020b) was conducted in which, along many fitness variables, the objective measure was used to measure the PA participation of U.S. children and youth. Specifically, (a) A large national sample (3 to 15 yr. old; measured 1640, male = 823; weighted 53669505, male = 51%), which represented the U.S. national children and youth population, was studied; (b) ActiGraph GT3X, a wrist-wearable accelerometer device, was used to measure the free-living PA participation of U.S. children and youth for one week; and (c) A set of fitness tests was also administered, including body composition, cardiovascular endurance, leg extension, modified pull-up, grip strength and plank. NNYFS was thus the first study that adopted both objective PA measures and fitness measures in a nationwide survey in U.S.

MIMS, a New, But Unusable Unit of Accelerometer Device

While NNYFS's PA data were collected using ActiGraph, the data, with a big surprise, were not released in the Count, but in a unit called "Monitor-Independent Movement Summary (MIMS)" (John et al., 2019). By definition, a MIMS is a unit defined by the accumulation of the acceleration over time (John et al., 2019). As mentioned earlier, ActiGraph and the Count were originally used in the PA related studied, including the national ones. Many other accelerometer devices were also later developed to assess PA, but except for a few, e.g., ActivPAL and ActiHeart, these devices had not been systematically validated and calibrated. As a result, the data from these different devices are not exchangeable since they were set onto different scales, which make comparing findings by different devices impossible.

To address this problem, MIMS was developed to provide a platform for different accelerometer devices (John et al., 2019). MIMS, similar idea to the Count, represents the accelerations caused by PA movement during a specific epoch, e.g., one minute, one hour or one day. MIMS per minute thus could be used to index the intensity of PA. The advantage of the MIMS algorithm, therefore, is compatible with the accelerometry data from all brands of devices, while the Count could be derived from the data only from ActiGraph accelerometers (John et al., 2019). The disadvantage is that intensity cutoff scores have been not established for MIMS, making it ineffective at assessing PA since some critical PA information, e.g., sedentary and MVPA time, will be missing. As a result, even though national youth PA data from NNYFS are available, no meaningful evaluations can be conducted at this time.

Equating: A Method May Help

To set PA intensity cutoff scores for MIMS, researchers either have to redo all the work that has been completed on the Count of ActiGraph, which might take another two decades or discover the relationship between MIMS and the Count and then transform MIMS into the Count to use all the previous Count-related-study outcomes, e.g., cutoff scores of PA intensity level. Fortunately, test equating may help for the latter choice.

Test equating is a statistical method to examine the relationship between two or more tests, or to transform test scores from different tests onto the same scale (Zhu, 1998). Test equating has been used in kinesiology to transform different fitness test scores onto the same scales (Zhu et al., 2010). It should also be used to discover the relationship between MIMS and the Count since for each 1-minute movement sequence, the Count and MIMS could be considered as two tests to measure the intensity. In this case, a participant in NNYFS is a student and MIMS and the Counts converted from MIMS are “scores” from two “tests”. If the conversion is successful, the data from any new device can first be calculated to MIMS, then link to MIMS-derived Counts so that Count-related information can be used e.g., EE (Puyau et al., 2002), risk of cardiovascular diseases (Boiarskaia, 2016), bone loading (Ren et al., 2021), etc., and study results could be crossly compared even if different devices were used.

Test equating uses different designs to collect data and the single-group and equivalent group designs are the most popular ones. In the single group design, one group of participants completed two tests. The advantage of this design is the group difference impact is removed. The disadvantage of this design is that participants’ performance might be impacted by fatigue or practice effect. Fatigue is negatively affecting performance and practice effect could positively affect performance. Equivalent-group design required two groups having similar abilities. While this design eliminates the fatigue and practice effect, it is often different to find equivalent groups in practice.

Two equating methods, i.e., linear and equipercentile, have been commonly used in testing equating practice. Linear equating uses linear function to transform between two tests. Linear equating could be used under both single-group design and equivalent-group design. The linear function is generated from the following rules between X and Y tests:

$$\frac{Y-mean(Y)}{SD(Y)} = \frac{X-mean(X)}{SD(X)}. \qquad \text{(Formula 1.5)}$$

Based on the condition, the transformation between X and Y scores could be yielded as:

$$Y = mean(Y) + \frac{(X-mean(X)) \times SD(Y)}{SD(X)}. \qquad \text{(Formula 1.6)}$$

Therefore, in a standard linear format, the transformation is:

$$Y = \frac{SD(Y)}{SD(X)} X + (mean(Y) - \frac{mean(X) \times SD(Y)}{SD(X)}). \qquad \text{(Formula 1.7)}$$

Advantage of linear equating is its computational convenience, i.e., the conversion could be easily accomplished with only means and SDs of both tests available. Disadvantage is that it is not able to accurately fit two scores with non-linear relationship. Figure 1.1 illustrates the foundation of linear equating.

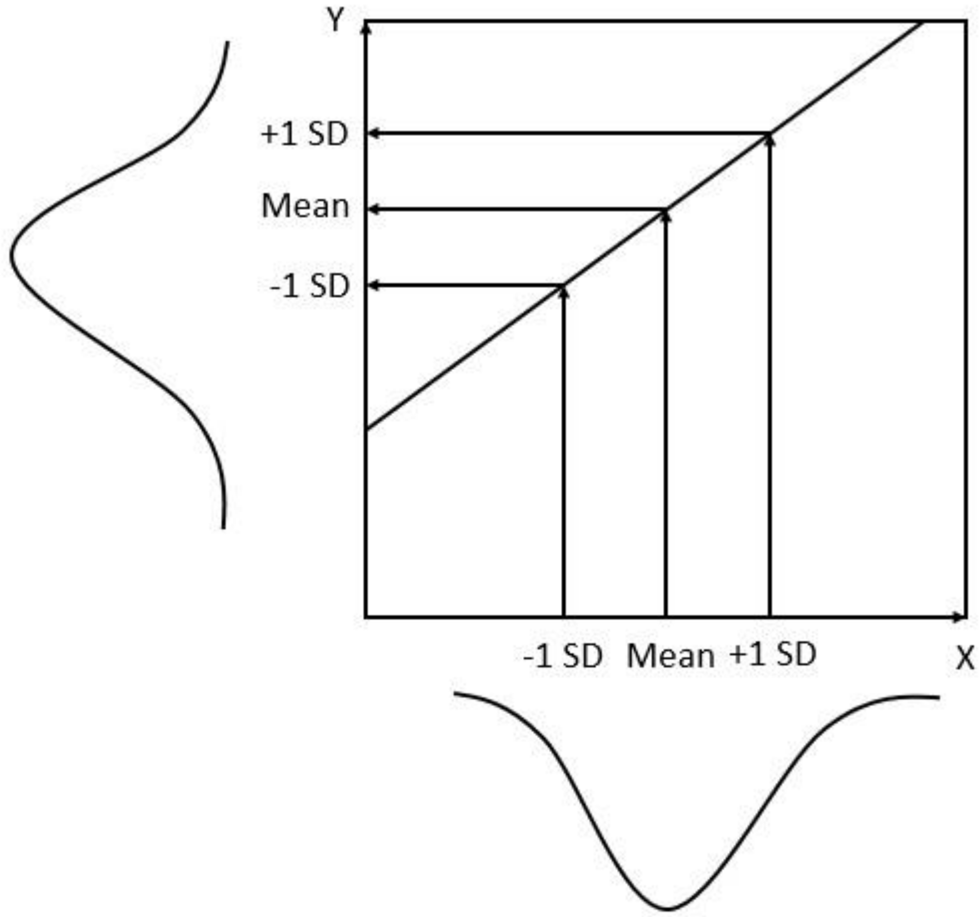

Figure 1.1 Illustration of linear equating

Equipercentile equating transforms test scores from one test to another based on percentiles. Equipercentile equating method could be used under both single- and equivalent-group design. Specifically, the score X in Test A will be equated to the Score Y in Test B whose percentile rank in Test B is the same as the X's percentile rank in Test A. The advantage of equipercentile is that the participants do not need to be same group, as long as two groups are equivalent, which, again, is sometime hard to find. The disadvantage of equipercentile equation is that its calculation requires more steps that the linear equating, but this can be easily addressed with modern computing power. Figure 1.2 illustrates the foundation of equipercentile equating.

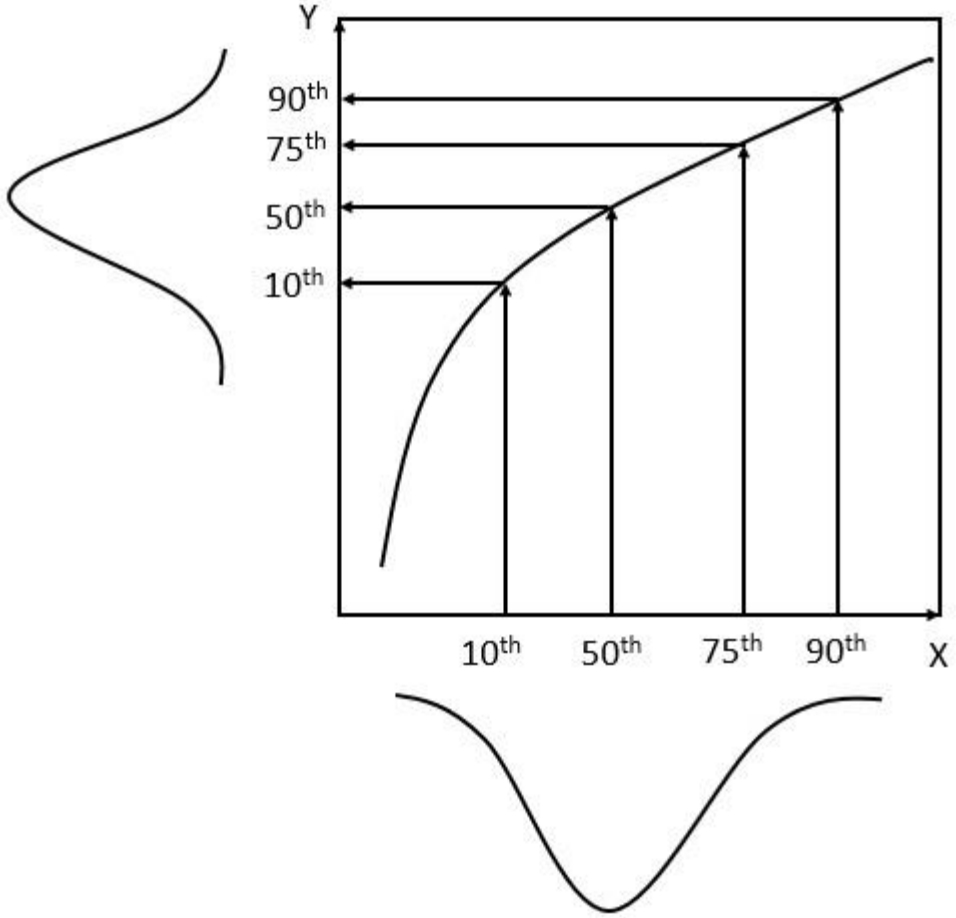

Figure 1.2 Illustration of equipercentile equating

Test equating has also the advantage of providing symmetrical model between two test scores over the commonly used statistical regression. Take linear regression as an example, using X as predictor and Y as outcome, the regression function is:

$$Y = f(X) = r\frac{SD(Y)}{SD(X)}X + intercept(Y), \qquad \text{(Formula 1.8)}$$

while using Y as predictor and X as outcome:

$$X = h(Y) = r\frac{SD(X)}{SD(Y)}Y + intercept(X). \qquad \text{(Formula 1.9)}$$

In both formula, *r* represents correlation between X and Y. It is obvious that the slopes in Formula 8 and 9 are not reverse to each other, but both with a term "*r*" to flatten the slope, making the chain gradually regress to the means. If X is predicted by the predicted Y predicted by original X, unless the correlation is strictly 1, the predicted X will not equal to the original X:

$$Predicted\ X = h(predicted\ Y) = h\big(f(X)\big) = r^2X + a\ intercept \neq X. \quad \text{(Formula 1.10)}$$

The difference in symmetricallity between regression and equating is illustrated in Figure 1.3.

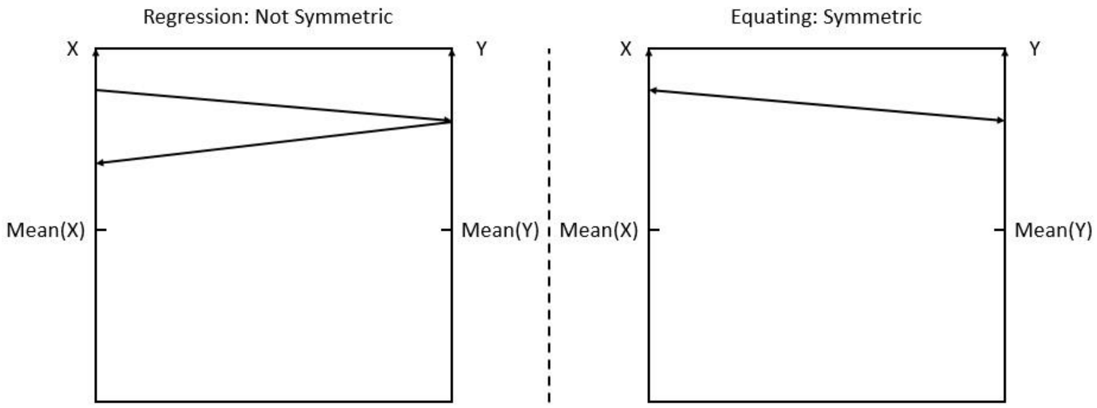

Figure 1.3 Difference between regression and equating

Needs of This Study

After a long delay, NNYFS is the first study to assess PA using objective measure and health-related fitness of U.S. children and youth. However, some critical information, e.g., PA intensity cutoff scores, have not been established for newly released MIMS units of the PA measure, which made PA data in NNYFS less meaningful. Linking MIMS to the Counts using test equating may solve this problem. To do so, three sub-studies are planned:

1. Study 1: Using the ActiGraph data set from a youth PA study (Liu et al., 2021), in which youth's PA data were collected by both ActiGraph and indirect calorimeters, to develop and validate the conversion relationship between MIMS and the Counts using test equating procedures.

2. Study 2: Using the EE data by indirect calorimeters from the study of Liu et al. (2021), to cross-validate the conversion relationship developed using the EE data collected by indirect calorimeter.
3. Study 3: Based on the MIMS-equated Count (MIMS-EQ-Count) developed and validated in Studies 1 and 2, to assess U.S. children and youth PA participation and examine its relationship with obesity using the NNYFS data.

Hypotheses

For this study, six hypotheses (1 in Study 1, 2 in Study 2, and 3 in Study 3) will be assumed and examined:

1. MIMS and the Count have a strong correlation and can be set onto the same scale using equating procedures. With the converted relationship, PA intensity can also be determined accurately based on the MIMS-EQ-Count.
2. EE represented by VO_2 will be similar by two counts within a specific PA intensity level. The "two counts", with lower-cased "c", represents two types of counts in this dissertation study, either the Count or MIMS-EQ-Count, while again "the Count" represents the ActiGraph Count.
3. The difference patterns of VO_2 across different PA intensity levels will be similar by two counts.
4. A high percentage of U.S. children and youth may not meet the daily 60-minutes-MVPA recommendation.
5. There is a difference among age, sex and racial/ethnic groups in PA participation in

U.S. children and youth.

6. For U.S. children and youth there is a negative relationship between their PA participation and obesity, and overweight and obesity groups likely spend less time on moderate, vigorous or very vigorous intensity PA.

Purpose/Specific Aims of This Study

The purpose of this study is, after connecting MIMS to the Count using test equating procedures and validating using the EE data, to assess free-living PA participation of U.S. children and youth and the relationship between PA participation and obesity using the 2012 NNYFS data.

To address hypotheses, this study has six (1 in Study 1, 2 in Study 2, and 3 in Study 3) specific aims:

1. Develop a formula linking MIMS and the Count using equating procedures and examine the conversion accuracy.
2. Examine the difference of VO_2 by two counts within a specific PA intensity level.
3. Examine the patterns of VO_2 across different PA intensity levels using the two counts.
4. Use the results from Study 1 and 2 to determine U.S. children and youth PA participation.
5. Examine the differences among age, sex and racial/ethnic groups in PA participation in U.S. children and youth.
6. Examine the relationship between PA participation and obesity of U.S. children and youth.

Significances of the Study

1. This will be the first study linking MIMS, a new unit, which could link all accelerometers on the same scale, to the Counts of ActiGraph, a PA unit whose validity and application have been well-studied and established.
2. This will also be the first study to evaluate the PA participation of U.S. children and youth using the latest PA objective data and examine its relationship with obesity.

References

ActiGraph LLC. (2008). *ActiGraph white paper: What is a count?* https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2017/11/26205758/ActiGraph-White-Paper_What-is-a-Count_.pdf

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: A second update of codes and MET values. *Medicine & Science in Sports & Exercise*, *43*(8), 1575–1581.

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W.L., Basset, D. R., Schmitz, K. H., Emplaincourt, P. O., Jacobs, D. R., & Leon, A. S. (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, *32*(Suppl. 1), 498–516.

Alhassan, S., & Robinson, T. N. (2008). Objectively measured physical activity and cardiovascular disease risk factors in African American girls. *Ethnicity & Disease*, *18*(4), 421–426.

American College of Sports Medicine (ACSM; 2019a). *Being active as a teen*. https://www.exerciseismedicine.org/assets/page_documents/EIM_Rx%20for%20Health_Teens.pdf

American College of Sports Medicine (ACSM; 2019b). *Being active with your young child*. https://www.exerciseismedicine.org/assets/page_documents/EIM_Rx%20for%20Health_Being%20Active%20with%20Your%20Young%20Child.pdf

American College of Sports Medicine (ACSM, 2018). Cardiorespiratory training programs. In R. Battista, K.L. Everett, T. Hargens, & M. Mayol, (Eds.), *ACSM's resources for the personal trainer* (5th ed., pp. 413–447). Wolters Kluwer.

Baker, D. G., & Newton, R. U. (2004). An analysis of the ratio and relationship between upper body pressing and pulling strength. *The Journal of Strength & Conditioning Research*, *18*(3), 594–598.

Ball, J. W., Bice, M. R., & Parry, T. (2014). Adults' motivation for physical activity: Differentiating motives for exercise, sport, and recreation. *Recreational Sports Journal*, *38*(2), 130–142.

Baltaci, G., Un, N., Tunay, V., Besler, A., & Gerçeker, S. (2003). Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. *British Journal of Sports Medicine*, *37*(1), 59–61.

Bassett, D. R., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step counting: A review of measurement considerations and health-related applications. *Sports Medicine*, *47*(7), 1303–1315.

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Journal of the American Medical Association*, *273*(14), 1093–1098.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W.

(1989). Physical fitness and all-cause mortality. *The Journal of the American Medical Association*, *262*(17), 2395–2401.

Blair, S. N., LaMonte, M. J., & Nichaman, M. Z. (2004). The evolution of physical activity recommendations: How much is enough? *The American Journal of Clinical Nutrition*, *79*(5), 913S–920S.

Boiarskaia, E. (2016). Recognizing cardiovascular disease patterns with machine learning using NHANES accelerometer determined physical activity data (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise, 14*(5), 377–381.

Braun, H. A., Kay, C. M., Cheung, P., Weiss, P. S., & Gazmararian, J. A. (2017). Impact of an elementary school-based intervention on physical activity time and aerobic capacity, Georgia, 2013-2014. *Public Health Reports*, *132*(Suppl. 2), 24S-32S.

Bülbül, S. (2020). Exercise in the treatment of childhood obesity. *Turkish Archives of Pediatrics*, *55*(1), 2–10.

Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of energy expenditure and physical activity in preschoolers. *Medicine & Science in Sports & Exercise*, *46*(6), 1216–1226.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126–131.

Centers for Disease Control and Prevention (CDC; 2021). *Defining adult overweight & obesity*. https://www.cdc.gov/obesity/adult/defining.html

Centers for Disease Control and Prevention (CDC; 2018). *Defining childhood obesity*. https://www.cdc.gov/obesity/childhood/defining.html

Center for Disease Control and Prevention (CDC; 2020a). *Adult Physical Inactivity Prevalence Maps by Race/Ethnicity*. https://www.cdc.gov/physicalactivity/data/inactivity-prevalence-maps/index.html

Center for Disease Control and Prevention (CDC; 2020b). *The NHANES National Youth Fitness Survey*. https://www.cdc.gov/nchs/nnyfs/index.htm

Chin, S. H., Kahathuduwa, C. N., & Binks, M. (2016). Physical activity and obesity: What we know and what we need to know. *Obesity Reviews*, *17*(12), 1226–1244.

Chirico, A. M., & Stunkard, A. J. (1960). Physical activity and human obesity. *New England Journal of Medicine*, *263*(19), 935–940.

Clark, B. K., Healy, G. N., Winkler, E. A., Gardiner, P. A., Sugiyama, T., Dunstan, D. W., Matthews, C. E., & Owen, N. (2011). Relationship of television time with accelerometer-derived sedentary time: NHANES. *Medicine & Science in Sports & Exercise*, *43*(5), 822–828.

Corbin, C. B., & Noble, L. (1980). Flexibility: A major component of physical fitness. *Journal of Physical Education & Recreation*, *51*(6), 23–60.

Cornette, R. E. (2011). The emotional impact of obesity on children. *Global Perspectives on Childhood Obesity*, 257–264.

Cosoli, G., Spinsante, S., & Scalise, L. (2020). Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. *Measurement*, *159*, 107789.

Crouter, S. E., Flynn, J. I., & Bassett Jr, D. R. (2015). Estimating physical activity in youth using a wrist accelerometer. *Medicine & Science in Sports & Exercise, 47*(5), 944–951.

DeSimone, G. T., Muscular Strength Versus Endurance Versus Power—What Is the Difference? *ACSM's Health & Fitness Journal*, *20*(5), 3–4.

Dhar, P., & Robinson, C. (2016). Physical activity and childhood obesity. *Applied Economics Letters*, *23*(8), 584–587.

Dietz, W. H. (2004). The effects of physical activity on obesity. *Quest*, *56*(1), 1–11.

Duffield, R., Dawson, B., Pinnington, H. C., & Wong, P. (2004). Accuracy and reliability of a Cosmed K4b2 portable gas analysis system. *Journal of Science and Medicine in Sport*, *7*(1), 11-22.

Ekelund, U. L. F., Sjöström, M., Yngve, A., Poortvliet, E., Nilsson, A., Froberg, K. Wedderkopp, N., & Westerterp, K. (2001). Physical activity assessed by activity monitor and doubly labeled water in children. *Medicine & Science in Sports & Exercise*, *33*(2), 275–281.

Epstein, L. H., Paluch, R. A., Kalakanis, L. E., Goldfield, G. S., Cerny, F. J., & Roemmich, J. N. (2001). How much activity do youth get? A quantitative review of heart-rate measured activity. *Pediatrics*, *108*(3), e44–e44.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences*, *26*(14), 1557–1565.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition & Physical Activity*, *12*(1), 1–22.

Fernandes, R. A., & Zanesco, A. (2010). Early physical activity promotes lower prevalence of chronic diseases in adulthood. *Hypertension Research*, *33*(9), 926–931.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, *30*(5), 777–781.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise*, *37*(11), S523–S530.

Fry, D. K., Huang, M., & Rodda, B. J. (2015). Core muscle strength and endurance measures in ambulatory persons with multiple sclerosis: validity and reliability. *International Journal of Rehabilitation Research*, *38*(3), 206–212.

Fryar, C. D., Carroll, M. D., Afful J. (2020a). Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960–1962 through 2017–2018. *NCHS Health E-Stats*.

Fryar, C. D., Carroll, M. D., Afful, J., & Division of Health and Examination Surveys (2020b). *Prevalence of overweight, obesity, and severe obesity among children and adolescents aged 2–19 years: United States, 1963–1965 through 2017–2018*.

https://www.cdc.gov/nchs/data/hestat/obesity-child-17-18/obesity-child.htm

Hong, I., Coker-Bolt, P., Anderson, K. R., Lee, D., & Velozo, C. A. (2016). Relationship between physical activity and overweight and obesity in children: Findings from the 2012 National Health and Nutrition Examination Survey National Youth Fitness Survey. *American Journal of Occupational Therapy*, *70*(5), 7005180060p1–7005180060p8.

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity*, *7*(1), 1–16.

John, D., Tang, Q., Albinali, F., & Intille, S. (2019). An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing. *Journal for the Measurement of Physical Behaviour*, *2*(4), 268–281.

Kamandulis, S., Emeljanovas, A., & Skurvydas, A. (2013). Stretching exercise volume for flexibility enhancement in secondary school children. *The Journal of Sports Medicine & Physical Fitness*, *53*(6), 687–692.

Karvonen, J., & Vuorimaa, T. (1988). Heart rate and exercise intensity during sports activities. *Sports Medicine*, *5*(5), 303–311.

Kim, Y., Beets, M. W., & Welk, G. J. (2012). Everything you wanted to know about selecting the “right” Actigraph accelerometer cut-points for youth, but…: A systematic review. *Journal of Science & Medicine in Sport*, *15*(4), 311–321.

Kingsley, M., Lewis, M., & Marson, R. (2005). Comparison of Polar 810s and an ambulatory ECG system. *International Journal of Sports Medicine*, *26*(1), 39–43.

Laurson, K. R., Eisenmann, J. C., & Welk, G. J. (2011). Body fat percentile curves for US children and adolescents. *American Journal of Preventive Medicine*, *41*(4), S87–S92.

Liu, J., Zhu, L., Xu, J., Su, Y., Sun, Y., Liao, J., Hu, S., & Liu, X. (2021). Diagnostic criteria and value of multiple-index joint diagnosis of exercise intensity in obese children and adolescents. Manuscript submitted for publication.

Lubans, D., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., Kelly, P., Smith, J., Raine, L., & Biddle, S. (2016). Physical activity for cognitive and mental health in youth: A systematic review of mechanisms. *Pediatrics*, *138*(3), e20161642.

Maheedhar, M., Gaurav, A., Jilla, V., Tiwari, V. N., & Narayanan, R. (2016). StayFit: A wearable application for Gym based power training. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6290–6293.

Matthew, C. E. (2005). Calibration of accelerometer output for adults. *Medicine & Science in Sports & Exercise*, *37*(Suppl. 11), S512–522.

McClain, J. J., Hart, T. L., Getz, R. S., & Tudor-Locke, C. (2010). Convergent validity of 3 low cost motion sensors with the ActiGraph accelerometer. *Journal of Physical Activity & Health*, *7*(5), 662–670.

McKenzie, T. L. (2016). Context matters: Systematic observation of place-based physical activity. *Research Quarterly for Exercise & Sport*, *87*(4), 334–341.

Merlo, C. L., Jones, S. E., Michael, S. L., Chen, T. J., Sliwa, S. A., Lee, S. H., Brener, N. D., Lee, S. M., & Park, S. (2020). Dietary and physical activity behaviors among high school students—Youth Risk Behavior Survey, United States, 2019. *Morbidity and Mortality Weekly Report—Supplements*, *69*(1), 64–76.

Migueles, J. H., Cadenas-Sanchez, C., Alcantara, J., Leal-Martín, J., Mañas, A., Ara, I., Glin, N. W., & Shiroma, E. J. (2021). Calibration and cross-validation of accelerometer cut-points to classify sedentary time and physical activity from hip and non-dominant and dominant wrists in older adults. *Sensors*, *21*(10), 3326.

Moliner-Urdiales, D., Ortega, F. B., Vicente-Rodriguez, G., Rey-Lopez, J. P., Gracia-Marco, L., Widhalm, K., Sjöström, M., Moreno, L. A., Castillo, M. J., & Ruiz, J. R. (2010). Association of physical activity with muscular strength and fat-free mass in adolescents: The HELENA study. *European Journal of Applied Physiology*, *109*(6), 1119–1127.

Moliner-Urdiales, D., Ruiz, J. R., Ortega, F. B., Rey-Lopez, J. P., Vicente-Rodriguez, G., Espana-Romero, V., Munguía-Izquierdo, D., Castillo, M. J., Sjöström, M., & Moreno, L. A. (2009). Association of objectively assessed physical activity with total and central body fat in Spanish adolescents; the HELENA Study. *International Journal of Obesity*, 33(10), 1126–1135.

Montoye, A. H., Clevenger, K. A., Pfeiffer, K. A., Nelson, M. B., Bock, J. M., Imboden, M. T., & Kaminsky, L. A. (2020). Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. *Journal of Sports Sciences*, *38*(22), 2569–2578.

Morrow, J. R., & Freedson, P. S. (1994). Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science*, *6*(4), 315–329.

Morrow, J. R., Tucker, J. S., Jackson, A. W., Martin, S. B., Greenleaf, C. A., & Petrie, T. A. (2013). Meeting physical activity guidelines and health-related fitness in youth. *American Journal of Preventive Medicine, 44*(5), 439–444.

Morrow, J. R., Zhu, W., Franks, D. B., Meredith, M. D., & Spain, C. (2009). 1958–2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise & Sport*, *80*(1), 1–11.

Must, A., Spadano, J., Coakley, E. H., Field, A. E., Colditz, G., & Dietz, W. H. (1999). The disease burden associated with overweight and obesity. *The Journal of the American Medical Association*, *282*(16), 1523–1529.

Pasco, J. A., Holloway, K. L., Dobbins, A. G., Kotowicz, M. A., Williams, L. J., & Brennan, S. L. (2014). Body mass index and measures of body fat for defining obesity and underweight: A cross-sectional, population-based study. *BMC Obesity*, *1*(1), 1–7.

Pate, R. R., McIver, K., Dowda, M., Brown, W. H., & Addy, C. (2008). Directly observed physical activity levels in preschool children. *Journal of School Health*, *78*(8), 438–444.

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., King, A. C., Kriska, A., Leon, A. S., Marcus, B. H., Morris, J., Paffenbarger, R. S., Jr., Patrick, K., Pollock, M. L., Rippe, J. M., Sallis, J., &

Wilmore, J. H. (1995). Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *The Journal of the American Medical Association, 273*(5), 402–407.

Pate, R. R., & Ross, J. G. (1987). Factors associated with health-related fitness. *Journal of Physical Education, Recreation & Dance, 58*(9), 93–97.

Plasqui, G., Bonomi, A. G., & Westerterp, K. R. (2013). Daily physical activity assessment with accelerometers: New insights and validation studies. *Obesity Reviews*, *14*(6), 451–462.

Plasqui, G., & Westerterp, K. R. (2007). Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity, 15*(10), 2371–2379.

Prentice-Dunn, H., & Prentice-Dunn, S. (2012). Physical activity, sedentary behavior, and childhood obesity: A review of cross-sectional studies. *Psychology, Health & Medicine*, *17*(3), 255–273.

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, *10*(3), 150–157.

Monk, T. H., Buysse, D. J., & Rose, L. R. (1999). Wrist actigraphic measures of sleep in space. *Sleep*, *22*(7), 948–954.

Reinerman-Jones, L., Harris, J., & Watson, A. (2017). Considerations for using fitness trackers in psychophysiology research. In *International Conference on Human Interface & the Management of Information*, 598–606.

Ren, S., Chen, Z., Qin, X., Zhao, X., Yang, T. D., & Zhu, W. (2021). Measurement and evaluation of bone loading in physical activity: A systematic review. *Measurement in Physical Education & Exercise Science*, *25*(2), 149–162.

Robbins, D. W., Marshall, P. W., & McEwen, M. (2012). The effect of training volume on lower-body strength. *The Journal of Strength & Conditioning Research*, *26*(1), 34–39.

Ronai, P. (2020). The YMCA Bench Press Test. *ACSM's Health & Fitness Journal*, *24*(6), 33–36.

Ross, J., & Gilbert, G. (1985). The national children and youth study: A summary of findings. *Journal of Physical Education, Recreation & Dance*, *56*(1), 45–50.

Ross, J. G., & Pate, R. R. (1987). A summary of findings. *Journal of Physical Education, Recreation & Dance*, *58*(9), 51–56.

Saint-Maurice, P. F., Troiano, R. P., Bassett, D. R., Graubard, B. I., Carlson, S. A., Shiroma, E. J., Fulton, J. E., & Matthews, C. E. (2020). Association of daily step count and step intensity with mortality among US adults. *The Journal of the American Medical Association*, *323*(12), 1151–1160.

Santos-Lozano, A., Santin-Medeiros, F., Cardon, G., Torres-Luque, G., Bailon, R., Bergmeir, C., Ruiz, J, R., Lucia, A., & Garatachea, N. (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine, 34*(11), 975–982.

Slaughter, M. H., Lohman, T. G., Boileau, R., Horswill, C. A., Stillman, R. J., Van Loan, M. D., & Bemben, D. A. (1988). Skinfold equations for estimation of body fatness in children and youth. *Human Biology*, *60*(5), 709–723.

Speakman, J. R. (1998). The history and theory of the doubly labeled water technique. *The American Journal of Clinical Nutrition*, *68*(4), 932S–938S.

Speakman, J. R., Yamada, Y., Sagayama, H., Berman, E. S., Ainslie, P. N., Andersen, L. F., Anderson, L. J., Arab, L., Baddou, I., Bedu-Addo, K., Blaak, E. E., Blanc, S., Bonomi, A. G., Bouten, C. V. C., Bovet, P., Buchowski, M. S., Butte, N. F., Camps, S. G. J. A., Close, G. L., ... & Wong, W. W. (2021). A standard calculation methodology for human doubly labeled water studies. *Cell Reports Medicine*, *2*(2), 100203.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J., Going, S., Norman, J. E., & Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine & Science in Sports & Exercise*, *36*(7), 1259–1266.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, *40*(1), 181–188.

Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine & Science in Sports & Exercise*, *43*(7), 1360–1368.

Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer-determined steps per day in US children and youth. *Medicine & Science in Sports & Exercise*, *42*(12), 2244–2250.

Tudor-Locke, C., McClain, J. J., Hart, T. L., Sisson, S. B., & Washington, T. L. (2009). Pedometry methods for assessing free-living youth. *Research Quarterly for Exercise & Sport*, *80*(2), 175–184.

Tudor-Locke, C., Sisson, S. B., Collova, T., Lee, S. M., & Swan, P. D. (2005). Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Canadian Journal of Applied Physiology*, *30*(6), 666–676.

U.S. Department of Health and Human Services (2018a). Active children and adolescents. In *Physical activity guidelines for Americans, 2nd edition* (pp.46–54). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

U.S. Department of Health and Human Services (2018b). *Physical activity guidelines for Americans, 2nd edition* (pp.55–65). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

U.S. Department of Health and Human Services (2018c). *Physical activity guidelines for Americans, 2nd edition* (pp.111). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic

Disease Prevention and Health Promotion. https://www.cdc.gov/nccdphp/sgr/pdf/sgrfull.pdf

Van den Tillaar, R., & Ball, N. (2020). Push-ups are able to predict the bench press 1-RM and constitute an alternative for measuring maximum upper body strength based on load-velocity relationships. *Journal of Human Kinetics*, *73*, 7–18.

Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine & Science in Sports & Exercise*, *37*(Suppl. 11), S501–511.

Wells, K. F., & Dillon, E. K. (1952). The sit and reach—a test of back and leg flexibility. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *23*(1), 115–118.

Wier, L. T., Ayers, G. W., Jackson, A. S., Rossum, A. C., Poston, W. C., & Foreyt, J. P. (2001). Determining the amount of physical activity needed for long-term weight control. *International Journal of Obesity*, *25*(5), 613–621.

Woessner, M. N., Tacey, A., Levinger-Limor, A., Parker, A. G., Levinger, P., & Levinger, I. (2021). The evolution of technology and physical inactivity: The good, the bad, and the way forward. *Frontiers in Public Health*, 9, 672.

Wolff-Hughes, D. L., Bassett, D. R., & Fitzhugh, E. C. (2014). Population-referenced percentiles for waist-worn accelerometer-derived total activity counts in US youth: 2003–2006 NHANES. *PloS One*, *9*(12), e115915.

World Health Organization. (2006). *Constitution of the World Health Organization– Basic Documents*, Forty-fifth edition, Supplement, October 2006.

World Health Organization (WHO; 2020a). *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

World Health Organization (WHO; 2020b). *Physical activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise & Sport*, *69*(1), 11–23.

Zhu, W., & Owen, N. (2017). *Sedentary behavior and health: Concepts, assessments, and interventions*. Human Kinetics. Champaign, IL.

Zhu, W., Plowman, S. A., & Park, Y. (2010). A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise & Sport*, *81*(4), 400–409.

Chapter 2

Linking MIMS with ActiGraph Counts: An Equating Study

Introduction

Being able to accurately measure free-living physical activity (PA) is essential to public health and the prevention and treatment of chronic diseases. Accelerometer-based monitors are now the most commonly used free-living PA measures in both research and practice. Among them, ActiGraph, a device by ActiGraph LLC., has been used most often in the past research studies. In fact, two *Medicine & Science in Sports & Exercise* supplements (2005 Nov; 37 (11 Supplement) & 2012 Jan; 37 (1 Supplement)) have been devoted to measurement and statistical data analysis related issues of ActiGraph. A keyword search for "ActiGraph" in *PubMed* databases on August 28, 2021 has resulted in about 4000 research articles, and corresponded to more than 3500 in *Web of Science* and *Scopus*.

Derived from the raw acceleration data captured, ActiGraph uses a unit called "Count" to quantify the amount of PA a person completed in a specific epoch, with larger counts representing more PA, and vice versa (ActiGraph LLC., 2008). To help determine the intensity of PA, a time-based unit called "Count per minute (CPM)" has been used and, based on CPM, extensive efforts have been made to set PA intensity categories, including sedentary, light, moderate, vigorous and very vigorous and across different age groups (Butte et al., 2014; Freedson et al., 1998; Freedson et al., 2005; Montoye et al., 2020; Pate et al., 2006; Santos-Lozano et al., 2013). Starting from 2004, both ActiGraph and the Count have been employed in national surveillance studies, e.g., National Health and Nutrition

Examination Survey (NHANES), and are used further to predict chronic health risks (Boiarskaia, 2016).

Many other accelerometry devices have also been developed to assess PA, but most of them have not been systematically studied and set PA intensity cutoff scores. As a result, outputs from different devices are not comparable and study findings by different devices cannot be directly compared. To address this problem, a new unit called "Monitor-Independent Movement Summary (MIMS)" has been developed (John et al., 2019), which could convert the acceleration data from any accelerometry devices onto the MIMS. In fact, although the PA data were collected using ActiGraph in the 2012 NHANES National Youth Fitness Survey (NNYFS; CDC, 2020), the released data, with a surprise, were in the MIMS unit.

While MIMS provides a platform to help the data from different accelerometry devices on the same scale, MIMS unit itself is not ready to be used for PA research and practice. This is due to the fact that although a higher MIMS value could represent higher PA, no specific cutoff points for PA intensities have been defined for MIMS, as a result, critical PA information, such as sedentary, moderate and vigorous PA time, will be missing, which makes MIMS ineffective at assessing PA. To set PA intensity cutoff scores for MIMS, researchers have two options. The first is that they need to redo all the works that have been completed for the Count, which might take another two decades. The second option would be to develop the relationship between MIMS and the Count and then transform MIMS into the

Count to use all the previous Count-related-study outcomes, e.g., cutoff scores of different PA intensity levels. Fortunately, test equating may help for the latter choice.

Test equating is a statistical method to examine the relationship between two or more tests, or to transform test scores from different tests onto the same scale (Zhu, 1998). Test equating has been used in exercise science to transform different fitness test scores onto the same scales (Zhu et al., 2010). It should also be used to discover the relationship between MIMS and the Count since for each 1-minute movement sequence, the Count and MIMS could be considered as two tests to measure the intensity. In this case, an observation is not a participant in NNYFS anymore, but instead, a 1-minute-movement with MIMS and the Counts serving as two “scores” from two “tests”. Furthermore, if the conversion is successful, i.e., the Count could be derived from MIMS, plus the data from any new device can first be calculated to MIMS, the Count could be calculated from any devices, not limited to ActiGraph. Therefore, Count-related information can be used, e.g., energy expenditure (Puyau et al., 2002), risk of cardiovascular diseases (Boiarskaia, 2016), bone loading (Ren et al., 2021), etc., and study results could be crossly compared even if different devices were used.

By using a data set collected using ActiGraph for a group of youth with a similar age range of NNYFS, the purpose of this study was to link MIMS to the scale of the Count. Specifically, this study aimed at determining which test equating method was the more efficient in converting MIMS to the Count by examine accuracy of the conversion using both calibration and cross-validation samples.

Methods

Participants and Data Collection

The employed data were adopted from a PA study by Liu et al. (2021; in press), in which a total of 81 obese youth, aged from 10 to 17 years old, from Dalian, Liaoning Province, China, was recruited based on screening for overweight and obesity among school-age children and adolescents (ICAHPU et al., 2018). Height and weight of the participants were measured to validate their obesity status. The participants were asked to perform various types of PA: Squat, jump-rope, aerobic dance, walking and running. Before each participant started PA, an ActiGraph GT3X initialized as 30Hz was worn on waist and a portable calorimeter (Cortex Meta Max 3B) was wear on face to measure the oxygen assumption.

Participants started with squat, jump rope or aerobic dance for 1 minute, and thereafter, after 1-minute rest, they performed walking/running at 3 km/h, 4 km/h, 5 km/h, 6 km/h and 7 km/h, respectively, without rest between different speeds, in a 10-by-10-meter square space, marked by cone Eight cones were placed on the corners and in the middle of the edges of squares and set to a cadence of sound at the of speeds of 3, 4, 5, 6 and 7 km/h to help the participants maintain target speeds. Participants needed to arrive to every cone, covering five meters' distance, at the time a beep occurred. The participants and their parents were fully informed of the procedures and risks of the study and gave their consent ahead of data collection. The study was also approved by the local ethics committee.

Data Analysis

Compute the Count. The raw ActiGraph data were downloaded to computer through ActiLife, the analytical software by ActiGraph, to compute the Count, with epoch length 1 minute, and saved as a matrix in another file. Every row of the matrix represents 1 minute in natural time, e.g., 9:38:00 am CST to 9:38:59 am CST where the left boundary of each epoch is the moment of an integer minute. Every column represents a variable, including, timestamp, count on x-axis, count on y-axis, count on z-axis and vector magnitude (VM), i.e., the magnitude of the sum of the vectors of the Count on the three axes. Each element in the Count matrix represents the Count value on the corresponding axis and minute.

Compute MIMS. The raw ActiGraph data were also processed in R statistical software using the function "mims_unit" in an R package, "MIMSunit" (John et al., 2019), and, again, the epoch was set as 1 minute, to generate MIMS matrix. In the MIMS matrix, each row represents one minute, each column represents a variable, including timestamp, MIMS on x-axis, MIMS on y-axis, MIMS on z-axis and 3-axis MIMS, i.e., the sum of MIMS on each axis. Each element in the MIMS matrix represents the MIMS value on the corresponding axis and minute. Thereafter, the Count matrix and MIMS matrix were merged by timestamp.

Statistical Analysis. Since NNYFS released 3-axis-MIMS data, and the 3-axis could be used without any further direction specification of the device, the 3-axis MIMS and VM of Count (3-axis count) are the two variables used in this equating study. A total of 16 participants (20%) were randomly selected to form the testing data for the purpose of cross-validation while the rest 80% participants (n = 65) as the training data. The test data were then analyzed using both linear equating and equipercentile equating methods, and the derived relationship between Counts and MIMS was further examined using the testing data.

Smoothing. Postsmoothing was used on the raw equipercentile equating transformation. Cubic spline was used in the postsmoothing procedure following the suggestion from Kolen and Brennan (2004). First, the cubic spline was used with MIMS as independent variable (IV) and Count as dependent variable (DV):

$$Count = f(MIMS). \qquad \text{(Formula 2.1)}$$

Second, another cubic spline was fitted with Count as IV and MIMS as DV:

$$MIMS = g(Count). \qquad \text{(Formula 2.2)}$$

Third, the average of these two cubic splines was used as the final smoothed relationship, with MIMS as IV and Count as DV:

$$Count = \frac{f(MIMS)+g^{-1}(MIMS)}{2}. \qquad \text{(Formula 2.3)}$$

To determine if the smoothing has an impact, the rooted mean square error (RMSE) of the postsmoothed equipercentile equating, as well as the correlation and difference between raw and postsmoothed equipercentile equating, were computed. Statistical significance was set a priori at $p < 0.05$.

Validation. To determine the accuracy of test equating, i.e., if MIMS can be converted to the Count scale accurately, the mean and standard deviation (M$\pm$SD), as well as correlation, *t*-statistics and effect size of the original ActiGraphy and MIMS-equated Count (MIMS-EQ-Count), the Count predicted from MIMS, were computed for both training and testing datasets.

In addition, cutoff scores from previous studies were applied to validate the classifications by MIMS-EQ-Count. Two sets of cutoff scores of school aged children (Freedson et al., 2005; Santos-Lozano et al., 2013), one for adults (Freedson et al., 1998) and two for pre-school children (Butte et al., 2014; Pate et al., 2006) were adopted and compared. The details of these cutoff scores are summarized in Table 2.1. The agreement and kappa between classification using the Count and MIMS-EQ-Count were computed.

Table 2.1

Cutoff scores from previous studies

Cutoff	Author (Year; Population)	Participants (N/male)	Counts Range	Counts by Intensity Upper Bounds			
				Sedentary	Light	Moderate	Vigorous
1	Santos-Lozano et al. (2013; Children)	31/19	NA	NA	2114	6548	11490
2	Freedson et al. (2005; Children)	100/14	1-7867	NA	1400	5700	10000
3	Freedson et al. (1998; Adult)	50/25	1800-13800	NA	1952	5725	9499
4	Butte et al. (2014; Preschool)	50/25	0-13077	820	3908	6112	NA
5	Pate et al. (2006; Preschool)	29/13	0-6280	NA	1680	3368	NA

Note: NA = Not available.

Results

Demographics

The demographics of the participants, as well as the training and testing subgroups, are summarized in Table 2.2.

Table 2.2

Demographic Information of participations by group

	N	Age (yr.)	Height	Weight	BMI
Train					
Male	35	13.4±2.3	169.3±10.4	85.7±20	29.5±4.3
Female	30	14.1±2.2	161.7±7.6	75.1±9.6	28.6±2.1
Subtotal	65	13.7±2.2	165.8±9.9	80.8±16.8	29.1±3.5
Test					
Male	10	14.2±2	170.4±10.7	86.8±19.4	29.5±3.5
Female	6	11.5±1	160.3±7.2	70.2±16.1	27±4.3
Subtotal	16	13.2±2.1	166.6±10.5	80.6±19.5	28.6±3.8
Total					
Male	45	13.6±2.2	169.5±10.3	85.9±19.7	29.5±4.1
Female	36	13.7±2.3	161.5±7.4	74.3±10.8	28.4±2.6
Subtotal	81	13.6±2.2	165.9±10	80.8±17.3	29±3.6

Equating Statistical Summary

Table 2.3 summarized descriptive statistics, as well the correlation, *t*-test and effect size (Cohen's D) of the Count and MIMS-EQ-Count. The equipercentile equating provided much smaller standard errors while there is little difference between two methods in correlations (both high with $r > .80$), *t*-test (both had small t-values and $p > .05$), and effect size (both showed no effect size). Since all observations were distributed in the clear trend of positive correlation between MIMS and the Count, i.e., $r = 0.85$, and no outlier was identified; therefore, all the observations were included in the study.

Table 2.3

Descriptive, correlation, t-test and classification results between predicted count and true count

	M±SD	SEE (M±SD)	*r*	t	*p*	Cohen's D
Train						
True	3351±5424	NA	NA	NA	NA	NA
Linear	3351±5424	36851±84176	0.85	0	1	0
Equipercentile	3350±5426	18.5 ±18.7	0.84	-0.02	0.98	0.0004
PSE	3360±5410	18.5 ±18.7	0.84	0.21	0.84	0.003
Test						
True	3831±5737	NA	NA	NA	NA	NA
Linear	3975±5797	43786±93975	0.86	1.53	0.13	0.05
Equipercentile	3911±5785	19.1±18.2	0.86	0.83	0.41	0.03
PSE	3924±5769	19.1±18.2	0.86	0.98	0.33	0.03

Note: PSE = Postsmoothed equipercentile.

The RMSE of the postsmoothed equipercentile equating was 409.66 counts per minute, which was similar to the one of the raw equipercentile equating. The correlation between raw and smooth equating was .9995 (see also Figure 2.1) and there is no statistical difference between two equating (paired *t*-statistic = -0.25, effect size = 0.02, $p = .80 > .05$). Since there was no significant difference between raw and postsmoothed equipercentile equating, the final decision will be: Still use the raw equipercentile equating.

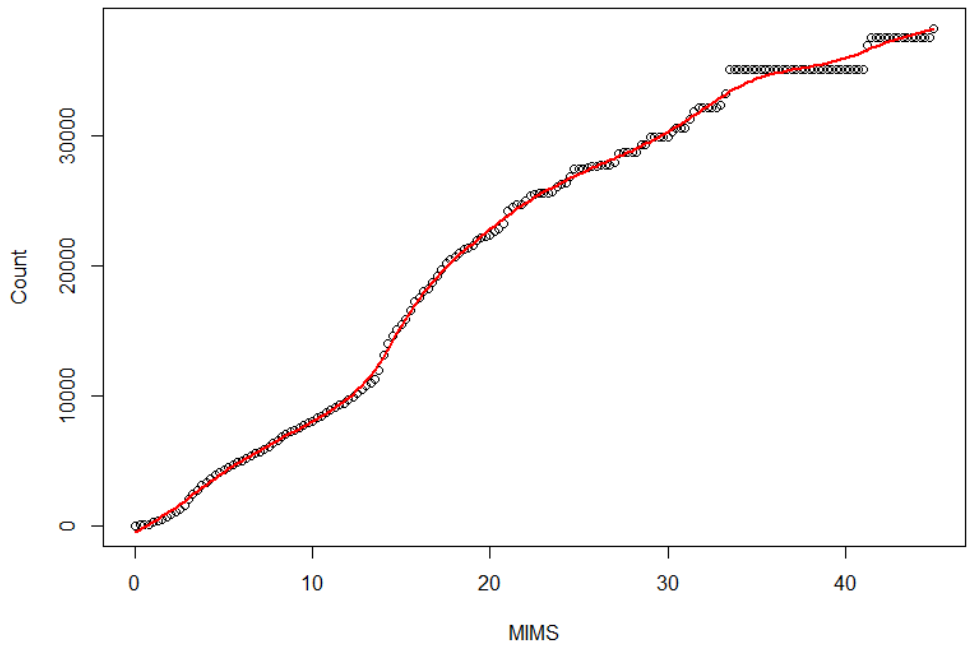

Figure 2.1 Equipercentile transformation: Raw (circled points) and postsmoothed (red line)

Classification and Cross-validation

Table 2.4 summarizes the classification agreement between the Count and MIMS-EQ-Count by each set cutoff scores and it was found that, although the agreement and Kappa coefficients of the two counts varied among the cutoff scores, the performances between linear and equipercentile methods were similar.

Table 2.4

Classification (light, moderate, vigorous and very vigorous) results between predicted count and true count

	Cutoff 1		Cutoff 2		Cutoff 3		Cutoff 4		Cutoff 5	
	p_0	Kappa	p_0	Kappa	p_0	Kappa	p_0	Kappa	p_0	Kappa
Train										
Linear	0.76	0.61	0.77	0.63	0.75	0.59	0.78	0.52	0.78	0.62
Equipercentile	0.73	0.53	0.77	0.62	0.73	0.55	0.77	0.48	0.78	0.62
PSE	0.74	0.56	0.76	0.62	0.74	0.56	0.77	0.49	0.78	0.63

Table 2.4 (cont.)

Test										
Linear	0.74	0.6	0.75	0.63	0.74	0.6	0.76	0.52	0.76	0.61
Equipercentile	0.71	0.53	0.74	0.61	0.71	0.55	0.76	0.51	0.77	0.61
PSE	0.73	0.57	0.75	0.63	0.72	0.56	0.76	0.51	0.77	0.61

Note: Details of Cutoff 1 – 5 can be found in Table 2.1; PSE = Postsmoothed equipercentile.

Table 2.5 summarizes the agreement and kappa between two counts to classify if an observation belongs to moderate and vigorous PA or not. It was found that the agreements between two counts under all cutoff scores were significantly improved. This is expected since only two PA intensity categories were classified. The performance between two equating methods was again very similar.

Table 2.5

Classification (MVPA or not) result between predicted count and true count

	Cutoff 1		Cutoff 2		Cutoff 3		Cutoff 4		Cutoff 5	
	p_0	Kappa	p_0	Kappa	p_0	Kappa	p_0	Kappa	p_0	Kappa
Train										
Linear	0.89	0.79	0.92	0.84	0.9	0.79	0.83	0.6	0.91	0.81
Equipercentile	0.87	0.73	0.92	0.85	0.89	0.77	0.84	0.6	0.9	0.79
PSE	0.88	0.76	0.92	0.84	0.89	0.77	0.84	0.6	0.91	0.82
Test										
Linear	0.89	0.77	0.92	0.85	0.9	0.79	0.82	0.6	0.91	0.82
Equipercentile	0.86	0.72	0.92	0.83	0.88	0.77	0.82	0.6	0.89	0.78
PSE	0.88	0.75	0.93	0.85	0.88	0.77	0.82	0.6	0.9	0.81

Note: MVPA = Moderate and vigorous physical activity; PSE = Postsmoothed equipercentile.

Based on the results of Tables 2.3–2.5, the conversion between the Count and MIMS based on the equipercentile equating method (Table 2.6) was used for the final conversion and the cutoff scores from Freedson et al. (2005), which provided the best agreement between two counts, were recommended for the analysis of the NNYFS PA data.

Table 2.6

Equipercentile equating based conversion table

MIMS	Count	SEE	MIMS	Count	SEE	MIMS	Count	SEE
0	0.05	0.02	15.25	15941.5	21.3	30.5	30629	3.5
0.25	71.5	11.8	15.5	16636	20.8	30.75	30629	3.5
0.5	79	47.2	15.75	17300	20.2	31	30672	3.3
0.75	128	47.4	16	17610	19.7	31.25	31304	3.2
1	299	47.5	16.25	18093	19.3	31.5	31936	3.0
1.25	415	47.5	16.5	18269.5	18.8	31.75	32154.5	2.8
1.5	523	23.8	16.75	18747	18.2	32	32154.5	2.8
1.75	669.75	23.8	17	19190	17.6	32.25	32154.5	2.8
2	866	47.6	17.25	19774	16.8	32.5	32154.5	2.8
2.25	1054	47.5	17.5	20251	16.1	32.75	32154.5	2.8
2.5	1283	47.4	17.75	20482.5	15.6	33	32373	2.6
2.75	1529.5	47.3	18	20741	15.2	33.25	33288	2.2
3	2017.75	23.5	18.25	20978	14.8	33.5	35133	2.0
3.25	2480.25	23.2	18.5	21311	14.5	33.75	35133	2.0
3.5	2794	23.0	18.75	21441.5	14.1	34	35133	2.0
3.75	3119	45.4	19	21628	13.7	34.25	35133	2.0
4	3360	44.7	19.25	21960	13.2	34.5	35133	2.0
4.25	3643	44.1	19.5	22224.5	12.8	34.75	35133	2.0
4.5	3890	43.4	19.75	22306.75	6.3	35	35133	2.0
4.75	4086	42.9	20	22358	12.4	35.25	35133	2.0
5	4282	42.4	20.25	22694.75	6.0	35.5	35133	2.0
5.25	4476	41.8	20.5	22900	11.6	35.75	35133	2.0
5.5	4672.75	20.6	20.75	23259.25	5.5	36	35133	2.0
5.75	4873	40.5	21	24216	10.4	36.25	35133	2.0
6	5036	39.8	21.25	24552	9.8	36.5	35133	2.0
6.25	5174	39.2	21.5	24701.5	9.3	36.75	35133	2.0
6.5	5368	38.6	21.75	24718.5	9.0	37	35133	2.0
6.75	5559	37.9	22	25015	8.7	37.25	35133	2.0
7	5736.5	37.3	22.25	25386	8.4	37.5	35133	2.0
7.25	5879.25	18.4	22.5	25490.5	8.3	37.75	35133	2.0
7.5	6066.5	36.1	22.75	25611.5	8.2	38	35133	2.0
7.75	6362	35.4	23	25635	8.0	38.25	35133	2.0
8	6607	34.7	23.25	25644	4.0	38.5	35133	2.0
8.25	6850	34.1	23.5	25705	7.6	38.75	35133	2.0
8.5	7027.75	16.8	23.75	26133.5	7.2	39	35133	2.0
8.75	7263.167	11.0	24	26280.25	3.4	39.25	35133	2.0
9	7384	32.7	24.25	26365	6.6	39.5	35133	2.0

Table 2.6 (cont.)

9.25	7565	32.3	24.5	26856	6.3	39.75	35133	2.0
9.5	7749	31.8	24.75	27445.5	6.0	40	35133	2.0
9.75	7933.25	15.7	25	27445.5	6.0	40.25	35133	2.0
10	8092.75	15.6	25.25	27445.5	6.0	40.5	35133	2.0
10.25	8322	30.8	25.5	27581	5.8	40.75	35133	2.0
10.5	8488.333	10.1	25.75	27676	5.6	41	35133	2.0
10.75	8698.5	30.0	26	27676	5.6	41.25	36978	1.7
11	8920.5	29.5	26.25	27741	5.5	41.5	37638	1.4
11.25	9138	29.0	26.5	27747.5	5.5	41.75	37638	1.4
11.5	9347	14.3	26.75	27754	5.4	42	37638	1.4
11.75	9430	28.4	27	27934	5.3	42.25	37638	1.4
12	9714	28.0	27.25	28626	5.0	42.5	37638	1.4
12.25	9947	27.6	27.5	28723.5	4.7	42.75	37638	1.4
12.5	10239	27.1	27.75	28723.5	4.7	43	37638	1.4
12.75	10494.75	13.3	28	28723.5	4.7	43.25	37638	1.4
13	10757.5	26.3	28.25	28730	4.5	43.5	37638	1.4
13.25	11008.25	12.9	28.5	29339.5	4.5	43.75	37638	1.4
13.5	11317.5	25.4	28.75	29339.5	4.5	44	37638	1.4
13.75	11926	24.7	29	29949	4.3	44.25	37638	1.4
14	13149	24.0	29.25	29958	4.2	44.5	37638	1.4
14.25	14052	23.4	29.5	29958	4.2	44.75	37638	1.4
14.5	14633	22.8	29.75	29958	4.2	45	38298	0.9
14.75	15074	22.3	30	29967	4.1			
15	15511	21.8	30.25	30312	3.7			

Discussion

To make any measure meaningful, outcomes must be set on a same scale so that corresponding evaluation standards, which could be norm-referenced, i.e., compare with a peer, or criterion referenced, i.e., compare with a criterion, could be used (Zhu, 2013). This is true for PA assessment practice. Using the accelerometer, one of the most popular objective PA measures, as an example, the raw data collected from the accelerometer must be set onto a scale whose relationship with PA intensity has been established. To establish this relationship, however, years’ efforts have to be made, which could be a challenge for a newly-developed accelerometer devices. In fact, except for ActiGraph, few devices, especially these commercial ones, has made efforts to develop and validate their PA intensity standards, which often made their reported values meaningless, at least questionable.

John et al. (2019) made a meaningful effort to try link all accelerometer devices on the same scale through the MIMS platform they created. As a result, the data collected in national studies by ActiGraph were released as MIMS. While this is a welcome attempt, a critical step is still missing, i.e., to set PA intensity standards for the MIMS scale. Without these standards, some critical information will be missed with analyzing the national data. As an example, a large MIMS value may represent a high intensity PA unit, but we don’t know if it represents a moderate, vigorous or very vigorous PA, which led to critical information related to PA such as duration of sedentary, light, moderate, vigorous and very vigorous PA cannot be calculated. While it could be done to set new PA intensity standards and link them with other health outcomes, it could take years’ effort and significant time and cost.

Fortunately, by applying test equating statistical methods, we are able to create a new measurement unit called “MIMS-equated Count” and link it with the Count of ActiGraph. In this way, all research findings based on the Count can be used by the MIMS platform. In fact, with a combination of MIMS, an open resource platform, and the results of this study, specifically the conversion table between MIMS and Count, a common problem in PA research, i.e., multiple devices with no validated PA intensity cutoff cores, was solved.

Postsmoothing helped the equating relationship becoming more continuous. In the present study, since the raw equipercentile equating is already very smooth and showing a clear trend, and postsmoothing did not give much difference, the raw equipercentile equating was still the final choice.The performance of postsmoothed relationship is also almost the same as the raw equipercentile. There are plenty of methods to smoothen the equating, including presmoothing and postsmoothing and more methods will be tried and examined in the future.

Figure 2.2 illustrates the advantages of MIMS and this study. In the past, whenever a new PA measure or device was created or developed, extensive works had to be made to link them to a criterion measure so that PA intensity cutoff scores can be created for the unit used by the new measure or device. With equating, the only thing needed is to administer the new measure or device and an established, valid and reliable measure with known PA intensity cutoff scores together and set the new measure and device onto the scale of the established one. Using MIMS in this study as an example, rather than link MIMS with a criterion measure and set PA intensity cutoff scores for it, which could take years' effort, we linked MIMS to the Count using test equating. Since MIMS is an open platform for other accelerometers, it provides some additional benefits. Say, a new accelerometer-based device was developed, the developers can simply transfer their data to MIMS (Step A in Figure 2.2), which can be easily further converted to MIMS-EQ-Count (Step B); thereafter, all accumulated research information for the Count, from PA intensity cutoff scores to the relationship between the Count and health outcomes, can be utilized by the new accelerometer (Step C).

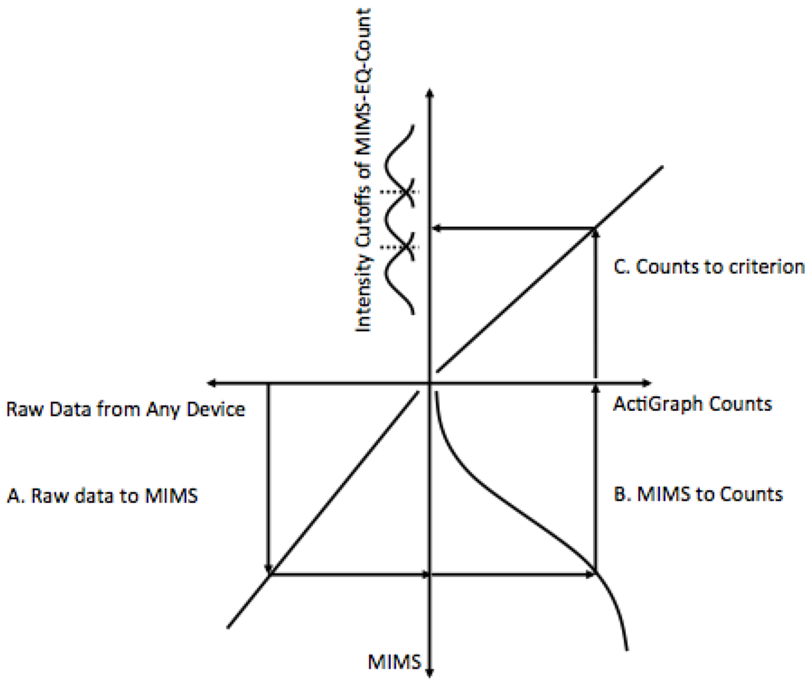

Figure 2.2 Illustration of the concept how MIMS and MIMS-EQ-Counts can solve the multiple devices and inconsistent comparison in PA research.

Limitations of this study should be noticed. First, the sample size was small, i.e., only about 80 participants were studied, which may not provide sufficient evidence for the conversion relationship developed when applying it to a larger group. Second, the participants were all young and overweight or obese, which may also threat the applicability of the conversion relationship, although the cutoff scores worked well. Last but not least, the number of PA tested was limited.

Conclusion

By applying the test-equating method, MIMS, which was used as the PA unit for national studies, could be linked onto the scale of the count. As a result, a new unit called

MIMS-EQ-Count was created and its compatibility with the count was verified. Together, any new accelerometry devices can transfer their data onto the MIMS platform, and from there to MIMS-EQ-Count, and finally to be able to use all useful research information developed for the count during the past several decades, thereby making new PA research findings trustworthy and exchangeable.

References

ActiGraph LLC. (2008). *ActiGraph white paper: What is a count?* https://s3.amazonaws.com/actigraphcorp.com/wp-content/uploads/2017/11/26205758/ActiGraph-White-Paper_What-is-a-Count_.pdf

Blair, S. N., Kohl, H. W., Barlow, C. E., Paffenbarger, R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Journal of the American Medical Association, 273*(14), 1093–1098.

Blair, S. N., Kohl, H. W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality. *The Journal of the American Medical Association, 262*(17), 2395–2401.

Blair, S. N., LaMonte, M. J., & Nichaman, M. Z. (2004). The evolution of physical activity recommendations: How much is enough? *The American Journal of Clinical Nutrition, 79*(5), 913S–920S.

Boiarskaia, E. (2016). Recognizing cardiovascular disease patterns with machine learning using NHANES accelerometer determined physical activity data (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of energy expenditure and physical activity in preschoolers. *Medicine & Science in Sports & Exercise, 46*(6), 1216–1226.

Center for Disease Control and Prevention. (CDC; 2020). The NHANES National Youth Fitness Survey. https://www.cdc.gov/nchs/nnyfs/index.htm

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise, 37*(11), S523–S530.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, *30*(5), 777–781.

Institute of Child and Adolescent Health of Peking University (ICAHPU), National Institute for Nutrition and Health (NINH) of Chinese Center for Disease Control and Prevention, & National Center for Women and Children's Health (NCWCH) of Chinese Center for Disease Control and Prevention. (2018). *Screening for overweight and obesity among school-age children and adolescents*. Beijing, China: National Health and Family Planning Commission of China (NHFPC).

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition & Physical Activity*, *7*(1), 1–16.

John, D., Tang, Q., Albinali, F., & Intille, S. (2019). An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing. *Journal for the Measurement of Physical Behaviour*, *2*(4), 268–281.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

Liu, J., Zhu, L., Xu, J., Su, Y., Sun, Y., Liao, J., Hu, S., & Liu, X. (2021). Diagnostic criteria and value of multiple-index joint diagnosis of exercise intensity in obese children and adolescents. Manuscript submitted for publication.

Montoye, A. H., Clevenger, K. A., Pfeiffer, K. A., Nelson, M. B., Bock, J. M., Imboden, M. T., & Kaminsky, L. A. (2020). Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. *Journal of Sports Sciences, 38*(22), 2569–2578.

Montoye HJ, Washburn R, Servais S, et al. (1983). Estimation of energy expenditure by a portable accelerometer. *Medicine and Science in Sports and Exercise, 15*, 403–7.

Pate, R. R., Almeida, M. J., McIver, K. L., Pfeiffer, K. A., & Dowda, M. (2006). Validation and calibration of an accelerometer in preschool children. *Obesity, 14*(11), 2000–2006.

Santos-Lozano, A., Santin-Medeiros, F., Cardon, G., Torres-Luque, G., Bailon, R., Bergmeir, C., Ruiz, J, R., Lucia, A., & Garatachea, N. (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine, 34*(11), 975–982.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, *40*(1), 181–188.

U.S. Department of Health and Human Services. (1996) *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. https://www.cdc.gov/nccdphp/sgr/pdf/sgrfull.pdf

Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine & Science in Sports & Exercise*, *37*(11), S501–S511.

World Health Organization (WHO; 2020). *Physical activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise & Sport*, *69*(1), 11–23.

Zhu, W., Plowman, S. A., & Park, Y. (2010). A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise & Sport*, *81*(4), 400–409.

Zhu, W. (2013). Science and art of setting performance standards and cutoff scores in kinesiology. *Research Quarterly for Exercise & Sport, 84*(4), 456–468.

Chapter 3

Cross-Validation of Cut-off Scores on the MIMS-EQ-Count

Introduction

Physical activity (PA) is any form of body movement caused by skeletal muscle contractions that leads to energy expenditure (WHO, 2020). Regular PA can provide substantial health benefits, which have been well-accumulated and confirmed since the 1996 Surgeon General's Report on PA and Health (U.S. Department of Health and Human Services, 1996). In children, PA can help to achieve lower bodyweight (Prentice-Dunn & Prentice-Dunn, 2012), better fitness (Janssen & LeBlanc, 2010), better cognitive function (Lubans et al., 2016), and lower the risk of developing chronic diseases in late adulthood (Fernandes & Zanesco, 2010). Because of the critical role PA performs for children's health, the 2018 PA Guidelines for Americans suggest that school-aged children and youth need one hour of PA daily, which should cover aerobic, muscle-strengthening, and bone-strengthening exercises (U.S. Department of Health and Human Services, 2018).

To better understand PA doses for children and youth, methods have been developed to measure PA and are divided into criterion and field measurements. Two commonly used criterion measurements are the doubly-labeled water (DLW) and the indirect calorimeter.

The indirect calorimeter estimates the rate of metabolism by measuring gas exchanges in humans or animals (Ferrannini, 1988). Specifically, the volumes of oxygen and carbon dioxide (VCO_2) are measured to calculate oxygen consumption (VO_2) and energy expenditure (EE). A consumption of one liter (L) of oxygen corresponds to 4.82 kilocalories (kcal). In practice though, to reduce the computing load, five kcal/L is often used (Adamakis et al., 2016). Previous research has often used COSMEMD K4b2, an indirect calorimeter with a mask and turbine, to measure the volume and composition of exhaled gas. These studies have utilized indirect calorimeter devices to measure children's basal EE (Wang et al., 2016), classroom EE and intensity (Honas et al., 2016), lifestyle EE (Aull et al., 2008), EE of common PA including rest (Harrell et al., 2005), and other types of EE. The validity (McLaughlin et al., 2001) and reliability (Darter et al., 2013) of the indirect calorimeter has been proven to be significant. Given their high validity and reliability, indirect calorimeters have also been used as a criteria measurement to calibrate the cut-off scores of accelerometers (Butte et al., 2014; Evenson et al., 2008; Freedson et al., 1998; Freedson et al., 2005; Puyau et al., 2002; Santos-Lozano et al., 2013; Treuth et al., 2004).

Field measurements consist of subjective, e.g., questionnaire, and objective components, e.g., pedometer, heart rate monitor, accelerometer and multi-sensor devices. The ActiGraph, an accelerometer by ActiGraph LLC, is the most studied accelerometer device that two supplements of *Medicine and Science in Sports and Exercise* (MSSE) (MSSE 2005 Nov; 37 (11 Suppl) MSSE 2012 Jan; 37 (1Suppl)) were, in fact, devoted specifically to the discussion of ActiGraph. A database search of "ActiGraph" by August 28, 2021, found more than 4,000 results in *PubMed* and more than 3,500 in *Web of Science* and in *Scopus*. Specifically, the ActiGraph Count is a unit used by ActiGraph to estimate amounts of PA. Using the count per minute (CPM) to estimate intensity has been proven to have high validity (Plasqui & Westerterp, 2007).

A typical way to use CPM while evaluating PA is to calibrate CPM with indirect calorimeters and to set up the cut-off scores for different PA intensities. Based on the definition by Ainsworth et al. (2011) and Santos-Lozano et al. (2013), the levels of intensity are defined by metabolic equivalent (MET): sedentary = 1 to 1.5 METs, light = 1.6 to 2.9 METs, moderate = 3.0 to 6.0 METs, vigorous = 6.1 to 9.0 METs, and very vigorous > 9.0 METs, where MET represents the "metabolic equivalent" and one MET refers to the intensity of quiet sitting (Ainsworth et al., 2000). Since the indirect calorimeter, e.g., K4, has a direct measurement of VO_2, which could be further calculated into MET values, calibrating alongside an indirect calorimeter is the best choice to set up cut-off scores for different intensity levels. Two of these data analysis methods which are used to help set cut-off scores are regression and receiver operating characteristic (ROC) curves.

The regression method involves the fitting of a linear regression function with counts as predictors and MET values as outcomes. Then 3 METs, 6 METs and 9 METs are plugged into the regression function to find out the corresponding values of the count. Count values corresponding to 3, 6 and 9 METs become the cut-off scores among light, moderate, vigorous, and very vigorous intensity activity (Freedson et al., 1998; Freedson et al., 2005; Puyau et al., 2002). If the regression method uses VO_2 as the outcome, then VO_2 values corresponding to 3, 6 and 9 METs will be calculated and corresponding values of the count can be found.

The ROC method is used to classify an observation into dichotomous categories, i.e., either positive ("1") or negative ("0"). For example, in fitness testing, a participant needs to be evaluated as fit or unfit. This is through setting up a cut-off score of field measurements between two categories. If an observation is greater than the cut-off score, it will be classified in the positive class, and if lower, in the negative class. Some observations are classified correctly, i.e., true positives (TP) and true negatives (TN), but some can be misclassified, i.e., false positives (FP, the negative observation that has been classified as positive) and false negatives (FN). Standard measurements are always used to represent ground truth, while field measurements with cut-off scores are used to duplicate true classifications. A good classification needs to be as correct as possible and has: usually a high sensitivity (TP/(TP+FN)) and a high specificity (TN/TN+FP). However, there is a trade-off between sensitivity and specificity while setting cut-off scores. Lower cut-off scores lead to higher sensitivity, but lower specificity and vice versa.

The ROC curve illustrates the relationship between sensitivity and specificity when the cut-off score changes from the bottom to the top, by setting the x-axis as 1-specificity and y-axis as sensitivity. On the ROC curve, an optimal cut-off score could be located where both sensitivity and specificity are acceptably high. While classifying PA intensities, multiple cut-off scores need to be set and therefore, some additional work needs to be done. For a cut-off score between light and moderate intensities, light will be "0" and moderate, vigorous, and very vigorous will all be coded as "1", making the multiple ordinal classes dichotomous. Then one can find the cut-off score based on the ROC curve (Sirard et al., 2005). The same principle applies when finding the cut-off scores between moderate and vigorous intensities, as light and moderate are coded as "0" while vigorous and very vigorous are coded as "1".

Using either of the regression or ROC methods, extensive efforts have been made to set the PA intensity cut-off scores for ActiGraph, in which the count is used as a field measurement or predictor and the MET value from indirect calorimeter is used as the criterion (Butte et al., 2014; Evenson et al., 2008; Santos-Lozano et al., 2013; Treuth et al., 2004). Table 3.1 summarizes a few selected studies setting up cut-off scores for the count using either regression or ROC methods.

Table 3.1

Calibration studies of ActiGraph

Authors (Year)	Age (yr.)	*N*/male	Method	CPM Range	Counts by Intensity Upper Bounds			
					S	L	M	V
Santos-Lozano et al. (2013)	12–16	31/19	ROC	NR	NR	2114	6548	11490
Freedson et al. (2005)	6–16	100/14	Reg	NR	NR	1400	5700	10000
Freedson et al. (1998)	23.9±4.0	50/25	Reg	1800-13800	NR	1952	5725	9499
Butte et al. (2014)	3–5	50/25	ROC	0-13077	820	3908	6112	NR
Puyau et al (2002)	6–16	26/14	Reg	NR	800	3200	8200	NR
Treuth et al (2004)	13–14	74/0	ROC	7856	100	2998	5198	NR
Evenson et al. (2008)	5–8	30	ROC	NR	100	2292	4008	NR

Note: CPM = Count per minute; Reg = regression; NR = not reported; S = sedentary; L = light; M = moderate; V = vigorous.

While PA and physical fitness are critical to the health of children and youth, no national study for them has been conducted for more than two decades in U.S. since the *National Children and Youth Fitness Study* I and II in 1985 and 1987, respectively. Fortunately, the NHANES National Youth Fitness Survey (NNYFS; CDC, 2020) was conducted in 2012, in which the ActiGraph was used to measure, record, and track the PA of U.S. youth and a set of health-related fitness components, including body composition, were also measured. The data of NNYFS represents the U.S. national children and youth population (3 to 15 years old; Male = 823, 50%; after weighting N=53,669,505, Male = 51%) and their PA participations were measured by wearing ActiGraph devices on their wrists for one week. However, the data was not released in the count, but in a new unit called “Monitor-Independent Movement Summary, MIMS” (John et al., 2019). Since MIMS’ relationship with PA intensity has not been established, the released data could not be used to determine U.S. children and youth’s PA participation since some critical information, such as the moderate and vigorous PA time, could not be calculated.

While MIMS was developed to provide a mutual platform for all different accelerometer devices on the same scale so that the measurement results by these devices can be compared with each other, MIMS is not ready to use since it has not been linked with PA intensity cut-off scores. To solve this problem, Study 1 in this dissertation study has linked MIMS and the count together using test equating, which helped to create a count-equivalent unit called MIMS-equated Count (MIMS-EQ-Count). As a result, by converting MIMS to MIMS-EQ-Count, the existing PA intensity cut-off scores based on the count can be used and the data of NNYFS can be analyzed to determine PA participation of U.S. children and youth.

Figure 3.1 illustrates the advantages of the MIMS-EQ-Count for PA research using wearable devices. In the past, whenever a new PA measure or device was created or developed, extensive work had to be completed to link them to a criterion measurement so that PA intensity cut-off scores could be created for the unit to be used by the new device.

With equating these, the only aspect needed is the administration of the new device and an established, valid, and reliable standard with known PA intensity cut-off scores. These can be used together to set a new device onto the scale of the established one. Using MIMS in this study as an example, rather than linking MIMS with a criterion measurement and setting PA intensity cut-off scores for it (a significant investment of time), we linked MIMS to the count using test equating. Since MIMS is an open platform for other accelerometers, it provides some additional benefits. For instance, if a new accelerometer-based device was developed, developers could simply transfer their data to MIMS (Step A in Figure 3.1), which could be easily further converted to MIMS-EQ-Count (Step B); thereafter, all accumulated research information for the count, from PA intensity cut-off scores to the relationship between the count and health outcomes, could be utilized by the new accelerometer (Step C).

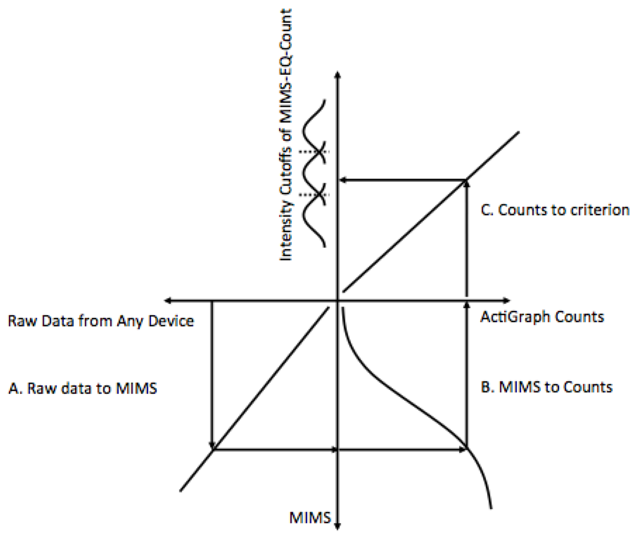

Figure 3.1 Illustration of the concept how MIMS and MIMS-EQ-Counts can solve the issue of multiple devices and inconsistent comparison in PA research.

The study by Liu et al. (2021, in press), whose data were used in Study 1 for the test equating, in fact, also collected VO_2, essentially energy expenditure (EE) or MET, data using the Meta Max 3B portable indirect calorimeter. This EE data provides another opportunity to further cross-validate the conversion relationship developed in Study 1.

Instead of using traditional regression or ROC methods to find PA intensity cut-off scores of MIMS directly from the EE data, this study focuses on proving the same role played by MIMS-EQ-Count as by the original count. Therefore, the purpose of this study is, using EE data, to examine whether or not MIMS-EQ-Count could have the same performance of the original count. Specifically, after classifying every observation, i.e., one-minute-movement, by both counts, the EE results under different PA intensity categories should be similar. We thus set our research hypotheses as follows:

1. EE represented by VO_2 will be similar by two counts within a specific PA intensity level.

2. The patterns of VO_2 across different PA intensity levels will be similar by two counts.

Methods

Data Used in This Study

The EE data collected in the Liu et al. (2021, in press) was used for this study. In that study, a total of 81 obese youth, aged from 10 to 17 years old, from Dalian, Liaoning Province, China, were recruited based on screening standards for obesity among school-age children and adolescents (ICAHPU et al., 2018). The heights and weights of participants were measured to verify their obesity status. The participants were asked to perform various types of PA: squats, jump rope, aerobic dance, walking, and running. Before each participant started PA, an ActiGraph GT3X initialized as 30 Hz was worn on the waist in addition to a portable calorimeter (Cortex Meta Max 3B) which was worn on the face to measure oxygen consumption.

Participants started with squats, jump rope or aerobic dance for one minute, and thereafter, after a one-minute rest, they performed walking/running at 3 km/h, 4 km/h, 5 km/h, 6 km/h and 7 km/h, respectively, without rest between different speeds, in a 10-by-10-meter square space, marked by cones. Eight cones were put on the corners and in the middle of the edges of squares. A musical cadence for speed of 3, 4, 5, 6 and 7 km/h was played to help the participants keeping target speeds: The participants needed to arrive to every cone, covering five meters' distance, at the time a beep occurred. The participants and their parents were fully informed of the procedures and risks of the study and gave their consents before data was collected. In addition, the study was approved by the local ethics committee. ActiLife, the corresponding software was used to initialize and download the data. The key variables from the EE data included VO_2 and VCO_2, but only VO_2 was used for this study.

Data Analysis

Data Cleaning. A timestamp of Cortex Meta Max 3B and a timestamp of ActiGraph were matched since Cortex Meta Max 3B uses relative time, i.e., starting from 00:00:00, and ActiGraph uses absolute time, i.e., the time of the corresponding time zone, in this case, China Standard Time. Since the Cortex Meta Max 3B provides the VO_2 value for every breath, the VO_2 of every minute need to be calculated. It was merged with MIMS, the count, and MIMS-EQ-Count by timestamp. The data matrix was finally constructed where each row represented an observation, i.e., the 1-minute movement, each column represented a variable, i.e., a feature of the observation, and included the MIMS value, the count value, the MIMS-EQ-Count value, and the timestamp.

Variable Calculation. Table 2.6 developed in Study 1, from which MIMS can be converted to MIMS-EQ-Count, was used to calculate MIMS-EQ-Count of each observation.

Cut-off Scores Employed. This study used 2 sets of cut-off scores for both counts to estimate intensity classes, from Freedson et al. (2005; Cutoff scores 1: 1400 and 5700 to distinguish Light, Moderate and Vigorous, respectively) and Santos-Lozano et al. (2013; Cutoff scores 2: 2114 and 6548). This study also used 3 METs and 6 METs, with 4.35±0.83 ml/min/kg (Liu et al., 2021) as 1 MET, i.e., 13.05 and 26.1 ml/min/kg, respectively, as cutoff score of VO_2. Classification results were an ordinal categorical variable, with 0 = Light, 1 = Moderate and 2 = Vigorous. The classes are shown in Table 3.2.

Table 3.2

Illustration of four new variables

		Light	Moderate	Vigorous
Cutoff 1	Count	0	1	2
	MIMS-EQ-Count	0	1	2
Cutoff 2	Count	0	1	2
	MIMS-EQ-Count	0	1	2
Cutoff VO_2	VO_2	0	1	2

Five new variables were calculated to represent the intensity categories:

1. ClassID1C: Use cutoff 1 to classify the Count.

2. ClassID1M: Use cutoff 1 to classify the MIMS-EQ-Count.

3. ClassID2C: Use cutoff 2 to classify the Count.

4. ClassID2M: Use cutoff 2 to classify the MIMS-EQ-Count.

5. ClassIDgas: Use cutoff score of VO_2 to classify VO_2.

Statistical Analysis. This study explored the relationship between VO_2 and different counts (the Count and MIMS-EQ-Count), in addition to different intensities (light, moderate, vigorous and very vigorous). The dependent variable was VO_2 and the independent variables were the counts and intensities.

To address the hypothesis, the following analysis was conducted under each set of cut-off scores with VO_2 as the dependent variable, and different counts and class as independent variables.

To address Hypothesis 1:

To examine the differences by two counts, t-tests was conducted. Statistical significance was set a priori at $p < 0.05$. The effect size was calculated by:

$$Effect\ size = Cohen's\ D = \frac{Mean\ difference}{SD_{pooled}}, \qquad \text{(Formula 3.1)}$$

where SD represents "standard deviation". The effect size was considered to be very small (0.01), small (0.2), medium (0.5), large (0.8), very large (1.2) and huge (2.0) given different values (Sawilowsky, 2009).

To address Hypothesis 2:

Contingency table, illustrated in Table 3.3, and related statistics were conducted between each count and VO_2.

Table 3.3

Illustration of the contingency table

		Count Intensity Classes		
		Light	Moderate	Vigorous
VO_2 Intensity Classes	Light	$Percentage_{00}$	$Percentage_{01}$	$Percentage_{02}$
	Moderate	$Percentage_{10}$	$Percentage_{11}$	$Percentage_{12}$
	Vigorous	$Percentage_{20}$	$Percentage_{21}$	$Percentage_{22}$

The agreement was calculated as:

$$P_0 = \text{Percentage}_{00} + \text{Percentage}_{11} + \text{Percentage}_{22}, \quad \text{(Formula 3.2)}$$

Kappa was also calculated as:

$$K = \frac{P_0 - P_e}{1 - P_e}, \quad \text{(Formula 3.3)}$$

where

$$P_e = \sum_{i=1}^{3} sum(column_i) \times sum(row_i). \quad \text{(Formula 3.4)}$$

The data analysis was completed using software R 3.6.2.

Results

The results are shown in Table 3.4 to 3.9 and Figure 3.2 and 3.3. Since the resting VO_2 of the participants was (M±SD) 4.35±0.83 ml/min/kg (Liu et al., 2021), the VO_2 cut-off scores were set at 13.05 and 26.1 corresponding to 3 METs and 6 METs (Ainsworth et al., 2011), respectively.

Table 3.4 and 3.5 show the results related to Hypothesis 1, i.e., EE represented by VO_2, were similar by two counts within a specific PA intensity level. Specifically, Table 3.4 shows the descriptive statistics of VO_2 of participants classified into each intensity level by the count and MIMS-EQ-Count, and the *t*-test results between two count groups under the same intensity level. Cut-off score 1, i.e., 1400 and 5700, was used to classify the count and

MIMS-EQ-Count. In each intensity level classified by two counts, the VO_2 values were similar, with less than 2 ml/min/kg difference. In effect, all sizes were smaller than 0.3, which corresponded to small (0.2) and was much less than the medium effect size (0.5).

Table 3.4

VO_2 in light, moderate and vigorous intensity levels classified by the count and MIMS-EQ-Count through cut-off score 1

VO_2 (ml/min/kg)		ActiGraph Count Based			MIMS-EQ-Count Based			*t*-test		
		N	VO_2	MET	*N*	VO_2	MET	*t*	*p*	ES
Count	Light	26	14.4±6.3	3.3±1.4	165	15.4±3.7	3.5±0.9	-0.82	0.42	0.25
Intensity	Moderate	660	16.4±6.8	3.8±1.6	755	18.3±7.7	4.2±1.8	-4.86	< 0.05	0.26
Classes	Vigorous	738	24.3±7.2	5.6±1.7	504	25.3±7.2	5.8±1.7	-2.55	< 0.05	0.15

Note: ES = effect size.

Table 3.5 summarizes the same analysis as in Table 3.4 but used cut-off score 2, i.e., 2114 and 6548. The effect sizes also showed less than medium effect (0.5) with the largest one 0.39 in moderate intensity, and a 0.27 of light intensity which was close to small (0.2), and a 0.09 which was close to very small (0.01; Sawilowsky, 2009).

Table 3.5

VO_2 in light, moderate and vigorous intensity levels classified by the Count and MIMS-EQ-Count through cut-off score 2

VO_2 (ml/min/kg) Under		ActiGraph Count-Based			MIMS-EQ-Count-Based			*t*-test		
Cut-off 2		*N*	VO_2	MET	*N*	VO_2	MET	*t*	*p*	ES
Count	Light	52	16.1±7.6	3.7±1.7	341	14.9±3.8	3.4±0.9	1.14	0.26	0.27
Intensity	Moderate	751	17.0±7.0	3.9±1.6	652	20.0±8.2	4.6±1.9	-7.1	< 0.05	0.39
Classes	Vigorous	621	24.9±7.0	5.7±1.6	431	25.6±7.0	5.9±1.6	-1.46	0.14	0.09

Note: ES = effect size.

Table 3.6 shows the descriptive statistics of the count and MIMS-EQ-Count, and the *t*-test results between the count and MIMS-EQ-Count by each intensity level classified by VO_2

values corresponding to 3 METs and 6 METs based on Ainsworth et al. (2011). The count group was higher than MIMS-EQ-Count group in each VO_2 intensity class, but effect sizes, represented by pooled Cohen's D, were less than medium, with one close to medium effect size and the other two close to small (0.2; Sawilowsky, 2009).

Table 3.6

The Count and MIMS-EQ-Count under each intensity level classified by VO_2

Count		ActiGraph Count			MIMS-EQ-Count			*t*-test		
		N	M	SD	N	M	SD	*t*	*p*	ES
VO_2	Light	300	4200	2680	300	3389	3150	3.4	< 0.05	0.28
Intensity	Moderate	705	7710	5478	705	5276	5320	8.5	< 0.05	0.45
Classes	Vigorous	419	12107	6975	419	10308	6775	3.8	< 0.05	0.26

Note: ES = effect size.

Table 3.7 summarizes further examination of the proportion of observations, originally from light, moderate and vigorous classes by the Count, being classified into light, moderate and vigorous classes by MIMS-EQ-Count. Under both cut-off scores, with in each intensity level classified by the count, most observations were classified into the same intensity level by MIMS-EQ-Count, with only a small proportion being classified differently.

Table 3.7

Proportion of each the Count-based intensity going into MIMS-EQ-Count based intensity

	ActiGraph Count Based	Frequency	MIMS-EQ-Count Based	Frequency	Proportion
Cut-off Score 1	Light	26	Light	15	58%
			Moderate	11	42%
			Vigorous	0	0%
	Moderate	660	Light	135	20%
			Moderate	497	75%
			Vigorous	28	4%
	Vigorous	738	Light	15	2%
			Moderate	247	33%

Table 3.7 (cont.)

			Vigorous	476	64%
Cut-off Score 2	Light	52	Light	33	63%
			Moderate	19	37%
			Vigorous	0	0%
	Moderate	751	Light	298	40%
			Moderate	422	56%
			Vigorous	31	4%
	Vigorous	621	Light	10	2%
			Moderate	211	34%
			Vigorous	400	64%

To address Hypothesis 2, i.e., the patterns of VO_2 across different PA intensity levels were similar by two counts, Table 3.8 and 3.9, and Figure 3.2 and 3.3 were generated. Table 3.8 shows the contingency tables between VO_2 and MIMS-EQ-Count, and between VO_2 and the count. Cut-off score 1 was used to classify the two counts. The agreement of the count and VO_2 was 0.48 and kappa was 0.15, while the agreement of MIMS-EQ-Count and VO_2 was 0.49 and 0.17, correspondingly.

Table 3.8

Contingency table between each count and VO_2 under cut-off score 1

N, Under Cut-off 1		ActiGraph Count Classes			MIMS-EQ-Count Classes		
		Light	Moderate	Vigorous	Light	Moderate	Vigorous
VO_2	Light	11	240	49	41	224	35
Intensity	Moderate	13	347	345	121	379	205
Classes	Vigorous	2	73	344	3	152	264
		$P_0 = 0.49$		K = 0.17	$P_0 = 0.48$		K = 0.15

Note: K = Kappa.

Table 3.9 summarizes the same analyzes as in Table 3.8 but used cut-off score 2 instead, with the agreement = 0.46 and kappa = 0.15 of MIMS-EQ-Count comparing to 0.52 and 0.20 of the Count, respectively, still showing the similar relationship. Although under a different cut-off score, the patterns were illustrated similar by the Count and MIMS-EQ-Count, as was described above.

Table 3.9

Contingency table between each count and VO_2 under cut-off score 2

N, Under Cutoff 2		ActiGraph Count Classes			MIMS-EQ-Count Classes		
		Light	Moderate	Vigorous	Light	Moderate	Vigorous
VO_2	Light	22	246	32	118	156	26
Intensity	Moderate	24	402	279	218	310	177
Classes	Vigorous	6	103	310	5	186	228
		$P_0 = 0.52$		K = 0.20	$P_0 = 0.46$		K = 0.15

Note: K = Kappa.

The results of Tables 3.8 and 3.9 indicated that MIMS-EQ-Count interacts with the VO_2 in three similar patterns that the count interacts with VO_2. One pattern noticed that there were very little misclassifications between light and vigorous categories. Another pattern that was observed was that most VO_2-classified light PA minutes were put into moderate levels by both counts, and some VO_2-classified moderate PA minutes were put into vigorous levels by both counts. This could be caused by the beginning of each higher speed of walking/running, when the energy expenditure system had not been stabilized at a higher level of intensity, but the kinematic movement was already intense. Finally, the third pattern was a small proportion of minutes classified as vigorous by VO_2 which had been put into moderate region. A possible explanation for these occurrences were: these minutes happened when a participant just finished their last item, 7 km/h running, and moved into the cool down phase when they were still in very intense breathing. Although both agreements are not substantially high, the relationships to VO_2 were almost the same between the count and MIMS-EQ-Count.

Figure 3.2 illustrates the scatter plot of the count against VO_2, while Figure 3.3 illustrates MIMS-EQ-Count against VO_2. The two plots show that the correlations (r = .5

and .49) and shape of the distribution between VO_2 and either count were similar, indicating again that the MIMS-EQ-Count and the count had similar interactions with given VO_2 values.

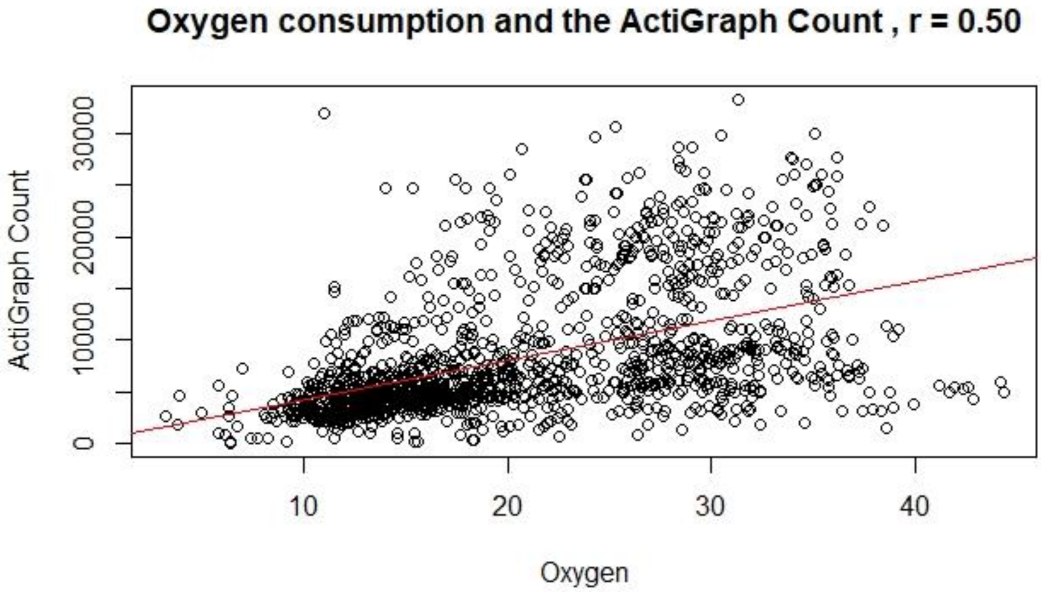

Figure 3.2 Relationship between the Count and VO_2

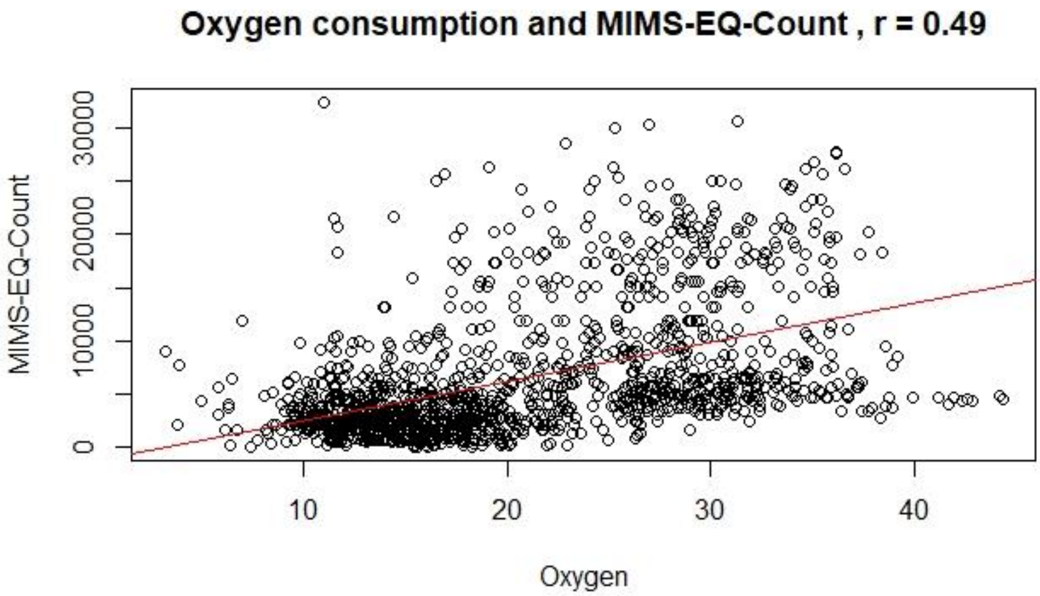

Figure 3.3 Relationship between MIMS-EQ-Count and VO_2

Together, these results supported a finding for Study 1 that MIMS-EQ-Count could be a valid substitution of the Count.

Discussion

The major finding of this cross-validation study was that the MIMS-EQ-Count has the similar performance with that of the count. The performance was examined by VO_2, a measurement representing PA energy expenditure or PA intensity. Specifically, Hypothesis 1, i.e., EE represented by VO_2 was similar by two counts within a specific PA intensity level, and Hypothesis 2, i.e., the patterns of VO_2 across different PA intensity levels was similar by two counts, were proven by the results.

The current result showed MIMS could be used to calculate. According to John et al. (2019), MIMS was invented to be an open-source algorithm to summarize accelerometer time-signals, which takes into consideration the different sampling rates of accelerometer devices and different range of acceleration, i.e., from ± 2g to ±8g. It could be used to measure walking or running in different speeds (John et al., 2019). However, without developing its own or linking to existing PA intensity, MIMS's application value is therefore limited.

Fortunately, this study further confirmed the applicability of the test-equating method to be used in kinesiology research, especially when changing the subjects from human participants into realistic movements for one minute. The test-equating method was originally designed for making scores in different forms of tests interchangeable (Kolen & Brennan, 2004). Later, it was adopted into kinesiology research, specifically, the fitness test (Zhu, 1998; Zhu et al., 2008), which is different from paper-pencil test. In fitness test, even the participants know the test content, like one-mile run/walk or push-up test, they might still not have a high score due to the limitation of fitness. This study, together with Study 1, further extended the usage of test-equating methods to PA measurement, where the target variable is not the ability of students, but the summary of some physical measures, e.g., acceleration. The successful application of this study indicates the great potential of applying test-equating methods into wider ranges of areas in exercise science.

This study was different from the traditional research due to finding cut-off scores by building the relationship between MIMS, or count, and VO_2 using ROC method (Butte et al., 2014; Evenson et al., 2008; Santos-Lozano et al., 2013; Treuth et al., 2004) or regression (Freedson et al., 1998; Freedson et al., 2005; Puyau et al., 2002). In lieu of initiating a new set of cut-off scores, which might disagree with other existing sets, this study took advantage of previously available cut-off scores to validate the MIMS-Count transformation. Meanwhile, several limitations of this study should be acknowledged:

1. The types of PA were limited. Only walking and running were included in Liu's study (in press), which might limit the acceleration and VO_2 relationship. If more activity type could be engaged, the algorithm should be better duplicated in a free-living situation.
2. The participants were only obese children and youth, which could also affect the mode of acceleration-VO_2 relationship. Studies showed that metabolic rates are different between obese and non-obese children in terms of both resting EE (Rodriguez et al., 2002) and total EE of free-living (Maffeis et al., 1995). As a result, the variety of acceleration-VO_2 relationship has not covered all the possibilities.

3. This study used the hip as a body part for attaching the sensor, which might also affect the rates in comparison to accelerometers attached to other sites, e.g., wrist, upper arm, ankle, etc., since, for instance, hip and wrist values of the count were shown to be different, and, furthermore, the differences were also distinct given the type of activities (John et al., 2019; Loprinzi & Smith, 2017).
4. Rather than using each participant's resting VO_2 values to compute PA MET values, a uniformed VO_2 value, i.e., 4.35 ml/min/kg, was used. As a result, the agreement between the accelerometer-based classification and VO_2-MET based classification may be negatively impacted.

In the future, more diversity needs to be introduced into the database, to include more types of activities, e.g., basketball, martial arts, leisure sports and household/schoolwork activities. In addition, a wider range of BMI of participants could be utilized as well as testing of accelerometers attached to different body parts. Finally, more age groups could be tested, including preschool children, young and older adults.

Conclusion

By using VO_2 to cross-validate the relationship between the count and MIMS-EQ-Count, this study proved that the MIMS-EQ-Count performed similar to the count in PA assessment. Specifically, EE represented by VO_2 was similar by two counts within a specific PA intensity level and the patterns of VO_2 across different PA intensity levels were also similar by two counts. We can therefore conclude based on the derivation of test-equating analysis that the MIMS-EQ-Count could be a valid substitute of the count and the cut-off scores developed for the count could be also used for the MIMS-EQ-Count.

References

Adamakis, M., Zounhia, K., Karteroliotis, K., & Koskolou, M. (2016). Objective measures for the assessment of energy expenditure: A historical overview. *Journal of Scientific Research & Reports*, *12*(3), 1–9.

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: A second update of codes and MET values. *Medicine & Science in Sports & Exercise*, *43*(8), 1575–1581.

Aull, J. L., Rowe, D. A., Hickner, R. C., Malinauskas, B. M., & Mahar, M. T. (2008). Energy expenditure of obese, overweight, and normal weight females during lifestyle physical activities. *International Journal of Pediatric Obesity*, *3*(3), 177–185.

Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of energy expenditure and physical activity in preschoolers. *Medicine & Science in Sports & Exercise*, *46*(6), 1216–1226.

Darter, B. J., Rodriguez, K. M., & Wilken, J. M. (2013). Test-retest reliability and minimum detectable change using the k4b2: Oxygen consumption, gait efficiency, and heart rate for healthy adults during submaximal walking. *Research Quarterly for Exercise & Sport*, *84*(2), 223–231.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences*, *26*(14), 1557–1565.

Fernandes, R. A., & Zanesco, A. (2010). Early physical activity promotes lower prevalence of chronic diseases in adulthood. *Hypertension Research*, *33*(9), 926–931.

Ferrannini, E. (1988). The theoretical bases of indirect calorimetry: A review. *Metabolism*, *37*(3), 287–301.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, *30*(5), 777–781.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise*, *37*(11), S523–S530.

Honas, J. J., Willis, E. A., Herrmann, S. D., Greene, J. L., Washburn, R. A., & Donnelly, J. E. (2016). Energy expenditure and intensity of classroom physical activity in elementary school children. *Journal of Physical Activity & Health*, *13*(Suppl. 1), S53–S56.

Harrell, J. S., McMurray, R. G., Baggett, C. D., Pennell, M. L., Pearce, P. F., & Bangdiwala, S. I. (2005). Energy costs of physical activities in children and adolescents. *Medicine & Science in Sports & Exercise*, *37*(2), 329–336.

Institute of Child and Adolescent Health of Peking University (ICAHPU), National Institute for Nutrition and Health (NINH) of Chinese Center for Disease Control and Prevention, & National Center for Women and Children's Health (NCWCH) of Chinese Center for Disease Control and Prevention. (2018). *Screening for overweight and obesity among school-age children and adolescents*. Beijing, China: National Health and Family

Planning Commission of China (NHFPC).

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition & Physical Activity*, *7*(1), 1–16.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

Liu, J., Zhu, L., Xu, J., Su, Y., Sun, Y., Liao, J., Hu, S., & Liu, X. (2021). Diagnostic criteria and value of multiple-index joint diagnosis of exercise intensity in obese children and adolescents. Manuscript submitted for publication.

Loprinzi, P. D., & Smith, B. (2017). Comparison between wrist-worn and waist-worn accelerometry. *Journal of Physical Activity and Health*, *14*(7), 539-545.

Lubans, D., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., Kelly, P., Smith, J., Raine, L., & Biddle, S. (2016). Physical activity for cognitive and mental health in youth: A systematic review of mechanisms. *Pediatrics*, *138*(3), e20161642.

Maffeis, C., Pinelli, L., Zaffanello, M., Schena, F., Iacumin, P., & Schutz, Y. (1995). Daily energy expenditure in free-living conditions in obese and non-obese children: comparison of doubly labelled water (2H2 (18) O) method and heart-rate monitoring. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity*, *19*(9), 671-677.

McLaughlin, J. E., King, G. A., Howley, E. T., Bassett Jr, D. R., & Ainsworth, B. E. (2001). Validation of the COSMED K4b2 portable metabolic system. *International Journal of Sports Medicine*, *22*(4), 280–284.

Prentice-Dunn, H., & Prentice-Dunn, S. (2012). Physical activity, sedentary behavior, and childhood obesity: A review of cross-sectional studies. *Psychology, Health & Medicine*, *17*(3), 255–273.

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, *10*(3), 150–157.

Rodriguez, G., Moreno, L. A., Sarria, A., Pineda, I., Fleta, J., Pérez-González, J. M., & Bueno, M. (2002). Determinants of resting energy expenditure in obese and non-obese children and adolescents. *Journal of Physiology and Biochemistry*, *58*(1), 9-15.

Santos-Lozano, A., Santin-Medeiros, F., Cardon, G., Torres-Luque, G., Bailon, R., Bergmeir, C., Ruiz, J, R., Lucia, A., & Garatachea, N. (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine, 34*(11), 975–982.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical methods*, *8*(2), 597–599.

Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and evaluation of an objective measure of physical activity in preschool children. *Journal of Physical Activity and Health, 2*(3), 345–357.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J., Going, S., Norman, J. E., & Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine & Science in Sports & Exercise*, *36*(7), 1259–1266.

U.S. Department of Health and Human Services (2018). Active children and adolescents. In *Physical activity guidelines for Americans, 2nd edition* (pp.46–54). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. https://www.cdc.gov/nccdphp/sgr/pdf/sgrfull.pdf

Wang, Z., Wu, H., Stone, W. S., Zhuang, J., Qiu, L., Xu, X., Wang, Y., Zhao, Z., Han, F., & Zhao, Z. (2016). Body weight and basal metabolic rate in childhood narcolepsy: A longitudinal study. *Sleep Medicine*, *25*, 139–144.

World Health Organization (WHO; 2020b). *Physical activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise & Sport*, *69*(1), 11–23.

Zhu, W., Plowman, S. A., & Park, Y. (2010). A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise & Sport*, *81*(4), 400–409.

Chapter 4

Are American Children and Youth Really Physically Active?

Introduction

Physical activity (PA) is very important to children and youth since regular PA offers many benefits to them, ranging from the improvement of fitness (Janssen & LeBlanc, 2010), cognitive function (Lubans et al., 2016), and the reduction of risk for developing chronic diseases in late adulthood (Fernandes & Zanesco, 2010). PA could also be helpful to fight against obesity, a major health issue among children and youth, by reducing sedentary time (Prentice-Dunn & Prentice-Dunn, 2012), lowering the risk of obesity (Hong et al., 2016), controlling weight (Wier et al., 2001), losing weight (Bulbul, 2020) and lowering the probability of remaining overweight (Dhar & Robinson, 2016). Therefore, having a certain amount PA in children's daily life is critical to keep healthy. The PA Guidelines for Americans suggests that children and youth need to have at least 60-minutes of moderate to vigorous PA (MVPA) everyday (U.S. Department of Health and Human Services, 2018).

However, recent trends suggest that children and youth lack the necessary PA required for their age group. Too many today are attracted by screens, which leads to a huge loss of time for PA. Led by the World Health Organization (WHO), a recent global study, which assessed the PA of 1.6 million children and youth across 146 countries, found that only a very small proportion of participants engaged in enough PA. Only 19% of the children and youth aged 11–17 years old were sufficiently physically active (Guthold et al., 2020). Moreover, an unbalanced participation between boys and girls was also observed. In general, girls (15.3%)

were less physically active than boys (22.4%). Furthermore, girls' proportion of sufficient PA status stayed almost the same (14.9% to 15.3%) between 2001 and 2016, whereas boys showed some improvement (19.9% to 22.4%) (Guthold et al., 2020). Should this trend continue, it will be extremely difficult to achieve the Global Action Plan's goal for PA 2018–2030, i.e., by 2030, to reduce 15% of the PA's insufficient rates from a high in 2018, e.g., as high as 81% of youth (aged 11–17 yr.) did not meet the PA recommendations by WHO (2018). As a result, researchers have suggested that not only should youth literacy regarding the importance of PA be encouraged, in addition to participation, but also families, schools, and policy makers need to work to create a better environment that inspires youth to have more PA (Guthold et al., 2020).

The U.S. follows this similar trend of lacking PA. According to the 2018 U.S. Report Card on PA for children and youth, an overall low PA participation has been observed (Katzmarzyk et al., 2018). Notably:

1. Only 24% of children and youth from six to seventeen years old met the PA guideline of 60 min/day and only 33% of the two to nineteen years old category met the sedentary guideline of less than two hours' screen time.
2. School time dedicated to physical education (PE) also did not provide enough of an impact for students to be active. Half (52%) of high school students in schools attended at least one PE class every week, but only 30% attended five days a week.
3. Similarly low PA statuses also existed in active transportation, i.e., only 23% of the students aged 12–19 walk or bike to school five to seven days a week.

PA participation can also differ dependent upon subpopulations, e.g., by age, sex, or race/ethnicity (Saffer et al., 2013; Bowser et al., 2016). Indeed, the low disparity of PA participation among different subpopulations was confirmed by the 2018 U.S. Report Card on PA where:

1. The proportions of meeting the PA recommendation were different between boys and girls, e.g., 45% of boys and 32% of girls reported active transportation (Katzmarzyk et al., 2018).
2. Age disparities exist: Among six to eleven years old children, 43% met the 60-minute guideline, but less than 10% in the 12–15 and 16–19 years old category of children and youth met the guideline. Sedentary behavior varied significantly among the different age groups. For the guideline of less than two hours of screen time, 35% of six to eleven, and 31% of 12–19-year-olds met the guideline, showing a correlation between sedentary time and increased age (Katzmarzyk et al., 2018).
3. There is also a racial and ethnic disparity in PA participation among children and youth. The proportions who met the screen time guidelines among White, Hispanic/Mexican American, Asian and African American were significantly different: 35%, 32%, 30% and 25%, respectively (Katzmarzyk et al., 2018).

As a result, the overall U.S. PA participation prevalence was graded as “D-” by the committee in 2016 (Katzmarzyk et al., 2016), and remained at the same score in 2018 (Katzmarzyk et al., 2018).

The data for the Report Card was synthesized by a Report Card Research Advisory

Committee organized by the National Physical Activity Plan Alliance (NPAPA; www.physicalactivityplan.org) using a set of national datasets, including the 2017 Youth Risk behavior Surveillance System (YRBSS), the 2016 National Survey of Children's Health (NSCH), 2005–06, 2011–12, the 2015–16 National Health and Nutrition Examination Survey (NHANES), the 2012 NHANES National Youth Fitness Survey (NNYFS), the State of Play 2017 Report, the 2016 School Health Policies and Practices Study (SHPPS), the 2016 School Health Profiles, and the 2017 United States Report Card on Walking and Walkable Communities (Katzmarzyk et al., 2018). Except for the NHANES where PA data was objectively measured, others collected data through questionnaires. As found by Troiano et al. (2008), PA recorded through the use of questionnaires may not be as accurate since students tend to over-report their PA participation. There is an urgency to verify the PA participation of U.S. children and youth using more valid and reliable measures, for instance such as objective measures.

Fortunately, a welcome effort was made by the Center for Disease Control and Prevention in 2012. NNYFS in that year used objective measures to track the PA of U.S. youth, along with many fitness variables. Specifically, the PA of a representative sample of three to fifteen-year-olds measured 1,640 children and youth (Male = 823, 50%). With weighting, that sample represented a total of 53,669,505 (Male = 51%) children and youth in U.S. ActiGraph GT3X, a wrist-wearable accelerometer device, was used to record the one-week-free-living-PA data of the participants. In addition, NNYFS also conducted fitness tests to assess a set of fitness components, including body composition, cardiovascular endurance, and muscular strength and endurance (specifically leg extension, modified pull-up, grip strength and plank). NNYFS was, in fact, the first study that adopted both an objective PA measure, i.e., ActiGraph, and a set of fitness measures in a nationwide survey. Although the study was conducted in 2012, the PA datasets of ActiGraph were not released until November 2020. Unfortunately, and surprisingly, the data was not released in the count of ActiGraph, but in a unit called “Monitor-Independent Movement Summary, MIMS” (John et al., 2019), whose relationship with PA intensity had not yet been established.

Fortunately, with the efforts made in Study 1 and Study 2 of my dissertation study, MIMS could be connected to the scale of the count, thus creating a new PA unit called MIMS-EQ-Count. With the MIMS-EQ-Count, we were able to determine the PA participation of U.S. children and youth.

Hypotheses

Three hypotheses were assumed and examined in this study:

1. A high percentage of U.S. children and youth do not meet the daily 60-minutes-MVPA recommendations.
2. There is a difference among age, sex, and racial/ethnic groups in PA participation in U.S. children and youth.
3. For U.S. children and youth there is a negative relationship between their PA participation and obesity, with overweight and obese groups likely spending less time on moderate, vigorous, or very vigorous intensity PA.

Purpose

The purpose of this study was to determine the PA participation of U.S. children and youth by using a new set of U.S. national data, in which PA was measured by ActiGraph, an objective PA measurement.

Specific Aims

To address the hypotheses of this study, three specific aims were set:

1. To determine the PA participation of U.S. children and youth using the MIMS-EQ-Count developed and validated from Study 1 and 2.
2. To determine the difference in PA participation between age, sex, and racial groups.
3. To examine the relationship between PA participation and obesity in U.S children and youth.

Methods

Datasets

Datasets of the NNYFS were used for this study, including demographic, body measure, and PA in MIMS by minutes. Only participants aged six to 15 years old were included in the study (N = 1149, 50% boys; weighted N = 41,159,799, 51% boys) and their demographics and characteristics are summarized in Tables 4.1 and 4.2 below:

Table 4.1

Summary of participants in Study 3

Age	Boys					Girls				
	N	*N* (Weighted)	BMI	OW (%)	OB (%)	*N*	*N* (Weighted)	BMI	OW (%)	OB (%)
6	65	2549252	16.6±2.2	7.1	15.7	59	1993737	16.6±3.1	15.4	11.2

Table 4.1 (cont.)

7	64	2329867	17.5±3.0	11.5	17.3	61	2112528	17.3±3.9	9.5	18.0
8	54	1799780	18.4±3.5	13.9	24.0	62	1968597	18.0±3.4	22.9	10.1
9	48	2030594	20.7±5.9	16.9	31.7	50	1783539	19.0±4.0	18.9	17.8
10	59	2191258	20.7±5.7	18.1	24.9	65	2115912	19.4±4.2	21.9	15.3
11	51	1674996	20.9±5.1	27.6	17.5	60	2064863	20.2±4.4	20.2	13.0
12	70	2571404	21.6±4.4	33.2	22.2	60	2119446	22.1±5.4	25.9	16.3
13	52	1929297	23.2±5.9	19.4	27.2	59	2113023	23.3±6.6	29.3	14.1
14	57	1968170	22.2±5.1	0.8	26.7	63	2409363	22.7±4.6	12.1	13.2
15	51	2018301	23.5±4.7	12.7	20.9	39	1415872	23.9±5.8	29.7	16.6

Table 4.2

Proportion of each race/ethnicity

Race	MA	OH	NHW	NHB	Other
Boys					
6	13%	7%	56%	11%	12%
7	14%	9%	51%	11%	15%
8	17%	12%	48%	18%	4%
9	6%	7%	65%	16%	7%
10	19%	11%	49%	10%	10%
11	16%	9%	52%	15%	8%
12	10%	12%	58%	14%	7%
13	11%	8%	52%	18%	10%
14	11%	13%	53%	17%	7%
15	15%	7%	63%	11%	4%
Girls					
6	15%	10%	54%	14%	6%
7	13%	10%	48%	15%	15%
8	16%	13%	48%	16%	7%
9	12%	3%	48%	15%	22%
10	12%	13%	54%	12%	9%
11	10%	13%	47%	11%	18%
12	14%	9%	58%	15%	5%
13	6%	13%	54%	18%	9%
14	10%	12%	61%	9%	8%
15	20%	7%	48%	20%	6%

Note: MA = Mexican American; OH = Other Hispanic; NHW = Non-Hispanic White; NHB = Non-Hispanic Black.

Measures

Multiple subsets of data were used and extracted variables of “age” and “gender” from the dataset “Demographic.” Height, weight, BMI, BMI category were also extracted from the “body measure” dataset. PA data in the MIMS unit with one-minute-epochs were also extracted.

Data Processing and Preparation

The following steps were taken to process the data and get them ready for the data analyses:

1. MIMS of each minute was transformed into the MIMS-EQ-Count as outlined in Study 1.
2. The MIMS-EQ-Count of each minute was classified into light, moderate, and vigorous intensity, and coded as ordinal data, “0”, “1”, “2” using the cut-off scores from Chandler et al. (2015), i.e., ≤ 9815, 9816 to 23627, and ≥ 23628, respectively.
3. For each day, each participant counted the number of minutes of light, moderate, and vigorous intensity PA.
4. The daily average time in light, moderate and vigorous intensities were subsequently calculated.
5. Categories of age, gender, height, weight, BMI, BMI category, daily light, moderate, and vigorous PA time based on participants’ identification number (variable name: SEQN, respondent sequence number) were merged. Each participant served as one observation.

Data Analysis

The following statistical analyses steps were followed to address Specific Aims 1, 2, and

3, respectively:

To address Specific Aim 1:

1. Sorting of the observations by age and sex.

2. Computation of mean (M), standard deviation (SD) of moderate, vigorous and very vigorous PA by age and sex.

3. Computation of mean (M), standard deviation (SD) of MVPA time of every observation.

4. Counting of the number of days of each observation having MVPA, i.e., moderate, vigorous or very vigorous PA, no less than 60-minutes.

5. Counting of the number of observations having all seven days more than 60-minutes MVPA in each sort by age and sex.

6. Computation of the proportion of participants satisfying 60-minute-MVPA by age and sex group.

To address Specific Aim 2:

7. Comparison of age difference by sex, sex difference by age, race difference by age and sex were examined by analysis of variance (ANOVA) with the average-daily MVPA used as a dependent variable. The statistical significance was set a priori at $p < 0.05$. The effect size was calculated by:

$$Effect\ Size = \eta^2 = \frac{SS_{between\ groups}}{SS_{total}}, \qquad \text{(Formula 4.1)}$$

where SS represents "sum of square". An effect size was considered as small effect (≥ 0.01), medium effect (≥ 0.06) and large effect (≥ 0.14) based on different values (Richardson, 2011).

To address Specific Aim 3:

8. Under each age-sex-racial/ethnic group, correlations between MVPA time and BMI were examined. Correlation was evaluated based on the criterion given by Safrit and Wood (1995): No relationship (±0 to 0.19), low (±0.2 to 0.39), moderate (±0.4 to 0.59), moderately high (±0.6 to 0.79) and high (±0.8 to 1.0)

9. Analysis of variance (ANOVA) in each sort was studied, using BMI categories (normal, overweight and obese) as independent variables and a daily average MVPA time as dependent variable. The statistical significance was set a priori at $p < 0.05$. The effect size was calculated by *Formula 1*. An effect size was considered as small effect (≥ 0.01), medium effect (≥ 0.06) and large effect (≥ 0.14) based on different values (Richardson, 2011). This study used the R version 3.6.2 to complete data analysis.

Results

To address Specific Aim 1, i.e., to determine the PA participation of U.S. children and youth using the MIMS-EQ-Count, developed and validated from Study 1 and 2, Tables 4.3, 4.4 and 4.5 were generated. Table 4.3 summarizes the daily MVPA time in minute by age, sex and BMI category.

Table 4.3

Descriptive statistics of MVPA time by age, sex and BMI category

	Male				Female			
Age	Normal	OW	Obese	Total	Normal	OW	Obese	Total
6	623±80	623±118	616±37	622±77	620±67	634±84	647±56	626±67
7	602±76	620±66	561±81	597±77	619±72	621±93	622±45	617±72
8	593±62	587±86	555±70	583±70	621±62	612±100	650±63	619±74
9	574±63	503±102	612±74	571±84	582±71	599±81	526±145	574±95

Table 4.3 (cont.)

10	561±78	518±69	520±56	541±75	574±74	564±93	569±78	571±79
11	497±98	550±89	538±94	518±98	526±93	547±68	562±101	535±91
12	505±131	488±110	454±93	488±112	504±80	517±88	516±80	501±99
13	479±121	479±43	451±98	466±106	477±100	460±85	483±86	470±95
14	406±109	417±111	425±104	412±108	444±112	437±71	473±105	447±106
15	411±93	388±42	407±81	410±85	438±105	428±78	446±77	425±99

Note: OW = overweight.

Table 4.4 summarizes the descriptive statistics of MVPA time. Some categories do not have sample and have been marked as not applicable (NA). Some other categories had only 1 sample before weighted so the standard deviation showed “0”.

Table 4.4

Descriptive statistics of MVPA time

		Male				Female			
Age	Race	Normal	Over	Obese	Total	Normal	Over	Obese	Total
6	1	618±65	634±0	540±0	607±61	575±81	690±72	634±0	613±90
	2	575±72	NA	644±39	609±68	639±45	639±73	NA	639±54
	3	634±84	668±29	612±12	634±78	634±52	597±54	668±49	634±53
	4	587±62	286±0	643±21	583±95	580±83	599±98	NA	586±88
	5	618±72	NA	621±0	618±64	658±46	NA	564±0	633±57
	Total	623±80	623±118	616±37	622±77	620±67	634±84	647±56	626±67
7	1	630±57	NA	602±61	623±59	645±70	630±14	643±0	641±58
	2	662±29	589±107	670±0	647±69	602±57	706±63	669±0	629±70
	3	601±57	686±0	533±75	589±70	634±63	614±0	623±53	627±63
	4	636±68	595±5	614±0	625±60	618±29	541±150	633±3	600±73
	5	497±99	619±46	NA	544±99	551±91	NA	575±0	555±84
	Total	602±76	620±66	561±81	597±77	619±72	621±93	622±45	617±72
8	1	595±0	547±49	563±60	563±52	587±59	557±79	NA	568±70
	2	607±108	539±12	NA	580±91	611±41	636±132	582±2	605±89
	3	590±31	705±65	559±73	589±62	620±62	620±95	695±0	624±69
	4	595±86	NA	567±0	593±84	642±72	659±12	667±66	654±61
	5	566±0	NA	471±0	516±47	650±36	NA	NA	650±36
	Total	593±62	587±86	555±70	583±70	621±62	612±100	650±63	619±74
9	1	516±0	NA	678±0	591±81	580±63	463±0	515±97	547±82
	2	613±34	508±43	659±0	574±71	611±53	NA	644±0	619±49
	3	562±65	500±127	607±78	567±89	564±71	605±71	635±6	584±73

Table 4.4 (cont.)

	4	598±61	592±8	594±41	596±53	627±49	NA	449±190	539±164
	5	615±0	449±49	NA	533±90	591±74	712±0	528±0	580±80
	Total	574±63	503±102	612±74	571±84	582±71	599±81	526±145	574±95
10	1	596±88	465±0	528±67	557±87	606±99	469±0	679±0	591±103
	2	506±93	579±0	512±11	512±78	577±80	689±0	670±56	596±84
	3	555±58	521±71	483±25	530±65	568±69	571±99	583±25	570±75
	4	612±112	NA	561±62	590±97	567±59	617±11	565±64	573±59
	5	553±4	NA	559±0	555±4	596±0	508±7	485±61	517±57
	Total	561±78	518±69	520±56	541±75	574±74	564±93	569±78	571±79
11	1	593±92	460±42	462±93	505±105	470±60	497±36	359±0	463±65
	2	585±97	575±68	NA	581±88	522±171	480±9	534±118	515±141
	3	492±59	554±100	599±40	527±81	526±80	575±66	603±66	549±80
	4	509±113	576±43	510±0	537±91	530±77	586±0	525±36	533±69
	5	385±75	NaN±0	NA	385±75	541±83	499±0	640±0	546±82
	Total	497±98	550±89	538±94	518±98	526±93	547±68	562±101	535±91
12	1	505±59	676±0	334±0	502±90	519±80	472±0	448±62	488±76
	2	389±267	364±122	498±85	442±178	495±41	678±0	622±28	535±79
	3	495±103	496±108	430±91	485±103	488±81	489±71	544±63	483±104
	4	552±66	537±0	479±72	518±69	561±87	622±55	516±80	561±88
	5	645±10	445±59	NA	511±106	566±0	NA	459±0	541±45
	Total	505±131	488±110	454±93	488±112	504±80	517±88	516±80	501±99
13	1	524±0	460±29	464±0	473±34	512±84	403±101	NA	463±107
	2	471±52	451±12	614±0	494±72	460±91	353±10	468±0	437±87
	3	431±93	533±11	410±102	432±101	470±97	469±56	432±28	466±79
	4	539±147	NA	497±34	520±131	555±60	506±0	529±103	544±78
	5	499±125	438±0	478±0	477±99	386±91	540±133	NA	424±126
	Total	479±121	479±43	451±98	466±106	477±100	460±85	483±86	470±95
14	1	421±74	NA	485±108	446±94	489±62	426±50	555±0	472±68
	2	451±77	NA	301±88	393±109	466±183	520±0	524±0	481±158
	3	405±127	NA	429±66	409±119	437±110	427±76	444±109	437±106
	4	384±73	594±0	454±84	428±92	383±32	NA	577±0	425±85
	5	378±70	348±0	NA	370±62	472±49	NA	NA	472±49
	Total	406±109	417±111	425±104	412±108	444±112	437±71	473±105	447±106
15	1	371±53	408±43	493±0	393±59	466±58	425±6	565±0	438±89
	2	511±35	379±0	441±0	468±60	550±0	NA	440±23	482±56
	3	400±93	373±40	367±59	395±82	382±98	414±87	370±0	391±87
	4	473±108	NA	511±64	485±97	477±113	NA	536±22	459±127
	5	378±63	NA	NA	378±63	NA	501±35	486±0	498±32
	Total	411±93	388±42	407±81	410±85	438±105	428±78	446±77	425±99

Note: NA = not applicable; In race: 1 = Mexican American; 2 = Other Hispanic; 3 = Non-

Hispanic White; 4 = Non-Hispanic Black; 5 = Other Race-including Multi-Racial.

Table 4.5 summarizes the proportion of satisfying the recommendation of 60 minutes of MVPA every day as has been suggested by WHO and 2018 PA Guidelines for America (WHO, 2020; U.S. Department of Health and Human Services, 2018). It is clear that the passing rates were much higher than it was expected.

Table 4.5

Passing rate of each group

	Male	Female	Total
6	1	1	1
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	1
12	0.988	1	0.994
13	1	1	1
14	1	1	1
15	1	1	1

To address Specific Aim 2, i.e., to determine the PA participation difference between age, sex, and racial/ethnic groups, comparisons were operated in multiple ways, and are displayed in Figure 4.1 and Tables 4.6 and 4.7. Figure 4.1 illustrates the trend of PA participation through age by each sex. A clear decreasing trend was observed.

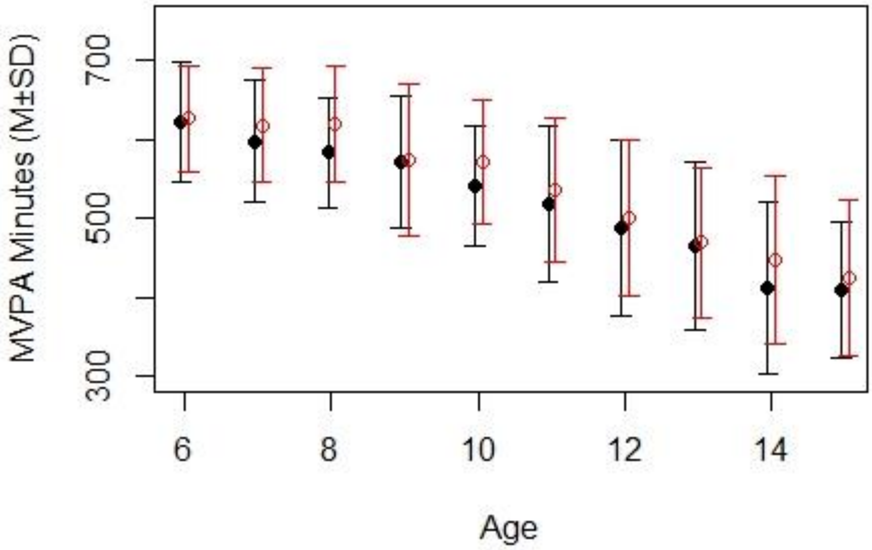

Figure 4.1 Trend of MVPA time through age by sex (Solid points: Male; Circled points: Female)

Comparisons among different population groups were also examined. The age has a large effect on the MVPA time in both males and females, with an effect size of 0.375 and 0.326 respectively. Given a specific age, however, the different between males and females were relatively small (see Table 4.6). A two-way ANOVA was completed using the age and sex as independent variable and MVPA time as dependent variable. The effect size (η^2) was 0.1185. Since the comparisons, i.e., t-test and ANOVA, were based on the weighted samples and the weights were large, representing the children and youth across the whole country, the *t*- and *F*-statistics were bound to be extremely large and therefore the *p*-values were extremely close to zero. As a result, *t, F* and *p* are not specifically reported each time because they could always show a significant difference.

Table 4.6

Effect size indicating difference between sex by age

Age	6	7	8	9	10	11	12	13	14	15
Cohen's D	0.113	0.208	0.477	0.005	0.345	0.086	0.134	0.011	0.286	0.156
η^2	0.001	0.017	0.057	0	0.037	0.008	0.004	0	0.027	0.006

Racial differences by age and sex were compared, and the results are shown in Table 4.7. There was no MVPA time difference between different races by age and sex.

Table 4.7

Effect size of race by age and sex

η^2	Male	Female
6	0	0
7	0.064	0.091
8	0.005	0.114
9	0.001	0.001
10	0.002	0.049
11	0.047	0.032
12	0.007	0.014
13	0.007	0.007
14	0.009	0.004
15	0.006	0.001

To address Specific Aim 3, i.e., to examine the relationship between PA participation and obesity in U.S children and youth, Tables 4.8, 4.9 and 4.10 were generated. Table 4.8 summarizes the effect size, i.e., η^2, of BMI category on MVPA time, as well as the correlation between BMI and MVPA time, by age and gender. No differences or only small difference of MVPA time were observed between different BMI categories.

Table 4.8

Effect size and correlation by age and sex

	Male		Female	
	η^2	*r*	η^2	*r*
6	0.001	0.087	0.015	0.203
7	0.032	-0.098	0.007	0.105
8	0.047	-0.176	0.011	0.189
9	0.048	0.211	0.022	-0.263
10	0.031	-0.126	0.002	0.102
11	0.04	0.095	0.021	0.062
12	0.026	-0.119	0.046	0.265
13	0	-0.184	0.001	0.019
14	0.006	0.03	0.002	0.064
15	0.009	-0.12	0.026	0.237

Table 4.9 summarizes η^2 of BMI category on MVPA time by age, gender and race. There was no difference between different BMI categories by age, sex and race.

Table 4.9

Effect size by age, sex and race

	Male					Female				
Race	1	2	3	4	5	1	2	3	4	5
6	0.098	0.26	0.002	0.023	0	0.191	0	0.033	0.009	0.522
7	0.046	0.049	0.137	0.054	0.173	0.003	0.255	0.002	0.007	0.011
8	0.043	0.136	0.028	0.005	NA	0.046	0.016	0.026	0.03	NA
9	NA	0.001	0.083	0.001	0.851	0.156	0.081	0.117	0.295	0.001
10	0.144	0.003	0.057	0.069	0.471	0	0.186	0.003	0.003	0.441
11	0.28	0.003	0.274	0.038	NA	0.156	0	0.134	0	0.061
12	0.123	0.084	0.028	0.085	0.791	0.186	0.546	0.128	0.052	NA
13	0.287	0.408	0.007	0.001	0.037	0.258	0.041	0.01	0.026	0.348
14	0.11	0.448	0.006	0.137	0.045	0.007	0.022	0	0.023	NA
15	0.331	0.389	0.084	0.035	NA	0.346	0.897	0.01	0.04	0.033

Note: NA = not applicable.

Table 4.10 summarizes the correlation between BMI and MVPA time by age, sex and race. The correlation 1 and -1 in Table 4.10 does not indicate high correlation, but only two samples in the corresponding category. There was no correlation between BMI and PA time by age, sex and race.

Table 4.10

Correlation between BMI and MVPA time

	Male					Female				
Race	1	2	3	4	5	1	2	3	4	5
6	-0.086	0.455	0.031	0.005	0.402	0.296	0.403	0.223	0.37	-0.668
7	-0.181	-0.114	-0.257	-0.052	0.203	-0.302	0.416	0.023	0.226	-0.077
8	-0.09	-0.46	-0.015	-0.08	-1	-0.019	0.051	0.114	0.344	0.504
9	1	-0.179	0.302	-0.151	-0.381	-0.42	0.226	0.13	-0.471	0.357
10	-0.349	0.136	-0.198	-0.243	0.644	-0.077	0.471	0.188	-0.001	-0.786
11	-0.55	0.225	0.401	0.052	-0.353	-0.768	0.055	0.295	-0.058	0.489
12	-0.451	0.589	-0.363	-0.063	-0.855	-0.254	0.532	0.424	-0.207	-1
13	-0.261	0.422	-0.336	-0.031	0.024	-0.721	-0.225	-0.053	-0.235	0.566
14	0.185	-0.638	0.145	0.156	0.232	0.318	-0.212	0.058	0.413	-0.475
15	0.485	-0.464	-0.214	-0.015	-1	0.559	-0.94	0.047	0.33	0.044

Discussion

This study examined the MVPA time of U.S. children and youth by using the NNYFS PA dataset. The algorithm developed in Study 1 and cross-validated in Study 2 was used to transform MIMS, in which form that PA data was released in the NNYFS dataset, into the MIMS-EQ-Count. This study addressed three Specific Aims, i.e., to determine the PA participation of U.S. children and youth, to determine the PA participation difference between age, sex, and racial/ethnic groups, and to examine the relationship between PA participation and obesity in U.S children and youth.

There are three major findings of this study:

1. Based on the wrist MIMS data, almost all children and youth satisfied the guidelines of 60 minutes of MVPA every day, which, again, was much higher than expected.

2. Within the category of six to 15-year-olds, as age increases, the time spent on daily MVPA showed a decreasing trend, in both boys and girls, which was consistent with the literature.
3. Within each specific age-gender group, MVPA time only had small differences between BMI categories, as well as in each age-gender-race groups.

Differing from the previous national studies about PA participation, the most surprising finding of this study was a much higher passing rating in terms of the percentages of children and youth met the PA recommendation. In the report card of PA in U.S. in 2018, only 24% of children and youth from six to seventeen-year-olds met the PA guideline of 60 min/day (Katzmarzyk et al., 2018). However, the data from NNYFS yielded to almost all children and youth meeting that 60-minute guideline, which, we believed, was likely caused by several limitations of NNYFS, specifically, how PA data were collected and released.

First, NNYFS used the wrist placement to collect the PA data children and youth, which might introduce noise. As was shown in the previous studies, data collected from wrists created much higher values than that from hips, even when results measure the same PA (John et al., 2019; Loprinzi & Smith, 2017). While the Count from hips has been widely studied and has built good relationships with VO_2, the criterion measure in terms of PA intensities, little has been done for the Count at wrist although it is well known that the wrist placement usually over-estimated PA measured (Loprinzi & Smith, 2017). In tracking free-living PA, since the participants might perform unique activities, a record or a journal is necessary for future analyses, either by using the compendium from Ainsworth et al. (2011)

or by transforming wrist data to hip data. However, this part of information, i.e., activity journal, was absent, making it such that sole wrist data is not capable of restoring enough PA information to be analyzed accurately.

Second, cut-off scores of the Count of wrist worn ActiGraph in free-living is also another limitation of the current study due to no cut-off scores of PA intensity was really established for the free-living situation. For example, this study used the cut-off scores from Chandler et al. (2015), which were derived from free-living situations in a summer camp. Other cut-off scores were also available but were derived from a few designed activities (Crouter et al., 2015; Ekblom et al.,2012). As is shown in Loprinzi and Smith (2017), even the same type of PA, e.g., running, the hip-wrist relationship could be different in comparison to the intensity.

Third, MIMS may set the noise-blocking threshold too low, allowing a great amount of noise from wrist to remain. As we know, a noise-blocking threshold is usually set up when an accelerometry-based PA device is developed. Without blocking or setting the threshold too low, little and meaningless movement will be recorded as some meaningful PA. John et al. (2019) summarized the values of the Count of hip-worn and wrist-worn ActiGraphs in multiple activities. In laundry, sweeping and loading the shelf, the Count on hip is about four to ten times higher than the Count on wrist. However, in walking, running, biking outdoor, going upstairs, the ratio was only one to two times. As a result, if two people have totally different lifestyles, such as distributions of time spending on household, exercise or sports, even if the Counts taken from their wrists are the same, their ground truth of EE might be quite distinct. It is expected that many hours recorded as MVPA by MIMS in NNYFS were likely from the sedentary, but active in hands movements. There is an urgent need to set and validate PA intensity cutoff scores for wrist-based PA monitors.

Finally, the commonly used research design to set PA intensity for accelerometry-based PA on the hip placement, in which a few low-, light- and high-intensity PA, e.g., sitting, walking and running, were employed, is likely not working for the wrist placement. This is because the noise from this kind of design will likely not be detected and removed. In the real-live free-living conditions, hands could move fast, e.g., when a child is playing computer or video game, but the activity itself will bring little physiological impact and health benefits to children and youth. More and typical sedentary, but hands-active activities, we call it "sedentary active behavior", should be included in the future studies when setting PA intensity cutoff scores for wrist-worn accelerometers.

Conclusion

While a few PA patterns, e.g., as age increases, children and youth tend to be less active, were confirmed using NNYFS PA data, the passing rates of meeting 60-min MVPA guideline were too high to be true. The flaws were likely from the wrist-placement employed, low noise blocking set by MIMS and not well-studied wrist-based PA intensity cutoff scores. There is an urgent need to address these issues in the future PA studies.

References

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: A second update of codes and MET values. *Medicine & Science in Sports & Exercise*, *43*(8), 1575–1581.

Bowser, J., Martinez-Donate, A. P., Carrel, A., Allen, D. B., & Moberg, D. P. (2016). Disparities in fitness and physical activity among children. *The Wisconsin Medical Journal*, *115*(5), 245–250.

Bülbül, S. (2020). Exercise in the treatment of childhood obesity. *Turkish Archives of Pediatrics*, *55*(1), 2–10.

Chandler, J. L., Brazendale, K., Beets, M. W., & Mealing, B. A. (2015). Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children. *Pediatric Obesity*, *11*(2), 120-127.

Crouter, S. E., Flynn, J. I., & Bassett Jr, D. R. (2015). Estimating physical activity in youth using a wrist accelerometer. *Medicine and Science in Sports and Exercise*, *47*(5), 944.

Dhar, P., & Robinson, C. (2016). Physical activity and childhood obesity. *Applied Economics Letters*, *23*(8), 584–587.

Ekblom, O., Nyberg, G., Bak, E. E., Ekelund, U., & Marcus, C. (2012). Validity and comparability of a wrist-worn accelerometer in children. *Journal of Physical Activity & Health*, *9*(3), 389-393.

Fernandes, R. A., & Zanesco, A. (2010). Early physical activity promotes lower prevalence of chronic diseases in adulthood. *Hypertension Research*, *33*(9), 926–931.

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2020). Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. *The Lancet Child & Adolescent Health*, *4*(1), 23–35.

Hong, I., Coker-Bolt, P., Anderson, K. R., Lee, D., & Velozo, C. A. (2016). Relationship between physical activity and overweight and obesity in children: Findings from the 2012 National Health and Nutrition Examination Survey National Youth Fitness Survey. *American Journal of Occupational Therapy*, *70*(5), 7005180060p1–7005180060p8.

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition & Physical Activity*, *7*(1), 1–16.

John, D., Tang, Q., Albinali, F., & Intille, S. (2019). An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing. *Journal for the Measurement of Physical Behaviour*, *2*(4), 268–281.

Katzmarzyk, P. T., Denstel, K. D., Beals, K., Bolling, C., Wright, C., Crouter, S. E., McKenzie, T. L., Pate, R. R., Saelens, B. E., Staiano, A. E., Stanish, H. I., & Sisson, S. B. (2016). Results from the United States of America's 2016 report card on physical activity for children and youth. *Journal of Physical Activity and Health*, *13*(Suppl. 2), S307–S313.

Katzmarzyk, P. T., Denstel, K. D., Beals, K., Carlson, J., Crouter, S. E., McKenzie, T. L., Pate, R. R., Sisson, S. B., Staiano, A. E., Stanish, H., Ward, D. S., Whitt-Glover, M., & Wright, C. (2018). Results from the United States 2018 report card on physical activity for children and youth. *Journal of Physical Activity & Health, 15*(Suppl. 2), S422–S424.

Loprinzi, P. D., & Smith, B. (2017). Comparison between wrist-worn and waist-worn accelerometry. *Journal of Physical Activity and Health*, *14*(7), 539-545.

Lubans, D., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., Kelly, P., Smith, J., Raine, L., & Biddle, S. (2016). Physical activity for cognitive and mental health in youth: A systematic review of mechanisms. *Pediatrics*, *138*(3), e20161642.

Prentice-Dunn, H., & Prentice-Dunn, S. (2012). Physical activity, sedentary behavior, and childhood obesity: A review of cross-sectional studies. *Psychology, Health & Medicine*, *17*(3), 255–273.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147.

Saffer, H., Dave, D., Grossman, M., & Ann Leung, L. (2013). Racial, ethnic, and gender differences in physical activity. *Journal of Human Capital*, *7*(4), 378–410.

Safrit, M. J., & Wood, T. M. (1995). *Introduction to measurement in physical education and exercise science* (3rd ed.). St. Louis, MO: Mosby-Year Book, Inc.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, *40*(1), 181–188.

U.S. Department of Health and Human Services (2018). Active children and adolescents. In *Physical activity guidelines for Americans, 2nd edition* (pp.46–54). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

Wier, L. T., Ayers, G. W., Jackson, A. S., Rossum, A. C., Poston, W. C., & Foreyt, J. P. (2001). Determining the amount of physical activity needed for long-term weight control. *International Journal of Obesity*, *25*(5), 613–621.

World Health Organization (WHO; 2018). *Global action plan on physical activity 2018–2030: More active people for a healthier world*. https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf

World Health Organization (WHO; 2020). *Physical activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Chapter 5

Conclusion

Since PA assessment is very crucial for children and youth, this dissertation study focused on developing a statistical procedure to make a new national PA dataset useful and determining PA participation status of U.S. children and youth, as well as its relationship with childhood obesity. Specifically, this study aimed at developing a formula to transform MIMS, a new unit that used in 2012 NNYFS dataset for free-living PA, into the Count, an established unit which have been linked with criterion measures of PA, and then to assess the PA participation of U.S. children and youth using the NNYFS dataset.

This dissertation study consists of three sub-studies:

1. Study 1 linked MIMS to the Count of ActiGraph by using the equipercentile test equating method and yielded an equating table for transformation. The estimated the Count is called “MIMS-EQ-Count.”
2. Study 2 cross-validated the transformation by taking the consideration of EE data available. Specifically, Study 2 proved that MIMS-EQ-Count and the Count have the same relationship with VO_2, indicating that MIMS-EQ-Count could be a valid substitute of the Count.
3. Finally, Study 3 took the advantages of Study 1 and 2 and transformed the MIMS data from NNYFS into MIMS-EQ-Count and made PA assessment of U.S. children and youth. Study 3 found that the time in MVPA decreases as age increases, but no clear

differences were observed between sex, as well as races and BMI categories. This result goes against the conclusions of previous studies that obese children and youth are less active (Guerra et al., 2006). Perhaps the most surprising finding of this study is that U.S. children and youth were rather physically active and the rates to meet the PA recommendation, i.e., 60 minutes per day (U.S. Department of Health and Human Services, 2018), were all close to or at 100%! This is clear too high to be true, which is why the original title of Chapter 4 "Are American children and youth physically active?" was changed to the current one "Are American children and youth really physically active?".

This "too high to be true" finding is likely caused by how the PA data were collected and processed in NNYFS. Specifically, the data in NNYFS were collected by asking children wearing ActiGraph on their wrists and processed and released not in the Count of ActiGraph, but on MIMS, an open platform for accelerometer data. Together, these two practices introduced two uncertainties at the same time, i.e., MIMS instead of the Count, and wrist-worn instead of hip-worn accelerometer, making the dataset not able to link with the previous PA intensity studies.

While the uncertainty of MIMS has been solved by equating MIMS and the Count in Studies 1 and 2, the uncertainty of wrist remained, although Study 3 used wrist cutoff scores to evaluate intensities. John et al. (2019) made the comparison between wrist-worn and hip-worn data in terms of both the Count and MIMS in multiple specific free-living activities, covering occupational, household, exercises and sport activities. The ratio of the Count between wrist and hip are quite different across the activities: For household or occupational activities, e.g., laundry and loading the shelf, the Count on wrist is 5 to 10 times higher than the Count on hip. On the other hand, exercises and sports, e.g., walking, biking, frisbee, running, the Count on wrist is about 1 to 2 times of the Count on hip, sometimes even lower than the Count on hip. Considering hip-measured Count has been proved by calorimeter to be valid measure of intensity, the wrist-measured Count could be also valid if and only if the ratio between wrist and hip could be fixed. However, the NNYFS dataset only provided "free-living" wrist data but did not specify the detailed activity type through all 7 days, causing impossibility to calculate hip MIMS, or the Count, data from the wrist data. Therefore, the amount useful information from this dataset regarding to PA participation is very limited.

Another limitation of this study should also be acknowledged, i.e., the data (Liu et al., in press) that were used in Studies 1 and 2 to derive and cross-validate the MIMS-Count relationship has several limitations. First, walking and running with different speed were the only activities included in the study. More activities could be helpful to expand the range of the raw acceleration. Second, only obsessed children and youth were included in the study. Wider range of BMI could also make the range of movement types broader. Finally, the age range was small, only from 10 to 17 yr. old.

While the findings of Study 3 were questionable and disappointing, the method and findings of Studies 1 and 2 should make a meaningful contribution to the PA assessment literature, specifically on how to set PA intensity cutoff scores for a new device. In the past whenever a new PA monitor was developed, PA intensity cutoff scores have to be set for it. To do so, the device has to link with a criterion measure, e.g., calorimeter (Butte et al., 2014; Evenson et al., 2008; Freedson et al., 1998; Freedson et al., 2005; Puyau et al., 2002; Santos-Lozano et al., 2013; Treuth et al., 2004), as well as link with some health outcomes. As a result, this kind of calibration is often expensive and time consuming. As an example, it took years' effort make ActiGraph and its Count a meaningful measurement tool. Because of the huge effort and cost invested, the developer of the device usually holds the algorithm of setting cutoff scores secret. The consequence of this practice is that the scales and values among commonly used devices are not exchangeable.

To break this barrier, John et al. (2019) developed an open platform through which all accelerometry-based data can be set on the same scale. In fact, PA data of NNYFS were processed by the platform and released in MIMS. However, since no PA intensity cutoff scores were set for MIMS, the data had limited meaning in evaluating PA since we may know the more MIMS, the more PA, but we will not know how much a person spend time on sedentary behavior, light, moderate and vigorous PA. To set the cutoff scores for MIMS, we have to find a group of participants and calibrate MIMS with VO_2 and derive the cutoff scores accordingly, which again is expensive and time consuming.

Instead, we took the advantage of test equating in this dissertation study and linked MIMS to the well-studied Count and directly use the cutoff scores that have been developed and validated by previous studies. The findings of Studies 1 and 2 verified and supported our effort. With established relationship between the Counts and MIMS, any accelerometry-based data can be transformed as MIMS-EQ-Count and use the cutoff scores developed and validate for the Count.

The finding of Study 2 illustrated some additional benefits of the conversion relationship developed, i.e., as long as the data can be transformed as MIMS-EQ-Count, the established relationship between the Count and other physiological and health variables could be used, e.g., doubly labeled water measured PA EE (Plasqui et al., 2013; Plasqui & Westerterp, 2007), sedentary time and moderate-vigorous PA (MVPA) time (Migueles et al., 2021), sleeping time (Monk et al., 1999) and to correlate PA measures with cardiovascular disease risks (Alhassan & Robinson, 2008).

To make the conversion relationship more applicable, more types of activities, participants with different body composition status and age groups could be engaged into the calibration studies so that the data with a great of variety can be collected to enhance the conversion relationship.

Another urgently needed future work is to find means to transform wrist data to hip data, thereafter the wrist data from national studies, such as NNYFS, could be used to extract more PA information. Since the relationship between hip and wrist varies given different types of activities, e.g., walking, biking, laundry, sweeping, one issue to be noticed is that “free living” is not a specific type of movement, but an environment where any types of movements could happen. The studies need to narrow down to specific activity types.

In conclusion, this dissertation study developed and validate the conversion relationship between MIMS and the Count, which could help to put any portable accelerometer devices on the same scale of the Count and be able to use the cutoff scores and other useful information developed for the Count. The problems of the PA data in NNYFS were explained and the needs of future researches and directions were outlined.

References

Alhassan, S., & Robinson, T. N. (2008). Objectively measured physical activity and cardiovascular disease risk factors in African American girls. *Ethnicity & Disease*, *18*(4), 421–426.

Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of energy expenditure and physical activity in preschoolers. *Medicine and Science in Sports and Exercise*, *46*(6), 1216–1226.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences*, *26*(14), 1557–1565.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise*, *30*(5), 777–781.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise*, *37*(11), S523–S530.

Guerra, S., Teixeira-Pinto, A., Ribeiro, J. C., Ascensão, A., Magalhães, J., Andersen, L. B., Duarte, J. A., & Mota, J. (2006). Relationship between physical activity and obesity in children and adolescents. *Journal of Sports Medicine and Physical Fitness*, *46*(1), 79–83.

John, D., Tang, Q., Albinali, F., & Intille, S. (2019). An Open-Source Monitor-Independent Movement Summary for Accelerometer Data Processing. *Journal for the Measurement of Physical Behaviour*, *2*(4), 268–281.

Migueles, J. H., Cadenas-Sanchez, C., Alcantara, J., Leal-Martín, J., Mañas, A., Ara, I., Glin, N. W., & Shiroma, E. J. (2021). Calibration and cross-validation of accelerometer cut-points to classify sedentary time and physical activity from hip and non-dominant and dominant wrists in older adults. *Sensors*, *21*(10), 3326.

Monk, T. H., Buysse, D. J., & Rose, L. R. (1999). Wrist ActiGraphic measures of sleep in space. *Sleep*, *22*(7), 948–954.

Plasqui, G., Bonomi, A. G., & Westerterp, K. R. (2013). Daily physical activity assessment with accelerometers: New insights and validation studies. *Obesity Reviews*, *14*(6), 451–462.

Plasqui, G., & Westerterp, K. R. (2007). Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity*, *15*(10), 2371–2379.

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, *10*(3), 150–157.

Santos-Lozano, A., Santin-Medeiros, F., Cardon, G., Torres-Luque, G., Bailon, R., Bergmeir, C., Ruiz, J, R., Lucia, A., & Garatachea, N. (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine, 34*(11), 975–982.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J., Going, S., Norman, J. E., & Pate, R. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine & Science in Sports & Exercise*, *36*(7), 1259–1266.

U.S. Department of Health and Human Services (2018). Active children and adolescents. In *Physical activity guidelines for Americans, 2nd edition* (pp.46–54). Washington, DC: U.S. Department of Health and Human Services. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf