CONCURRENT VALIDITY OF THE FITBIT FLEX PERSONAL ACTIVITY MONITOR AND ACTIGRAPH GT3X+ IN FREE-LIVING SETTING

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Nicklaus Alan Redenius

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Health, Nutrition and Exercise Sciences

December 2016

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Concurrent Validity of the Fitbit Flex Personal Activity Monitor and ActiGraph GT3X+ in Free-Living Setting

By

Nicklaus Alan Redenius

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Wonwoo Byun
Chair

Dr. Bryan Christensen

Dr. Mary Larson

Dr. Clayton Hilmert

Approved:

December 9, 2016
Date

Dr. Yeong Rhee
Department Chair

ABSTRACT

This study investigated the concurrent validity of two accelerometer-based physical activity (PA) monitors, the wrist-worn Fitbit Flex, and ActiGraph's hip-worn GT3X+. Specifically, we examined the relationship, differences, and level of agreement between Fitbit and GT3X+ sedentary behavior (SED) and moderate-to-vigorous PA (MVPA) estimates.

Sixty-seven adults (mean age: 47.1 ± 14.1, female: 73.1%) from North Dakota State University wore the Fitbit and GT3X+, and logged any sleep and non-wear time, for seven consecutive days in free-living conditions. GT3X+ estimates were calculated using ActiGraph Freedson, Troiano, and Freedson's VM3 cut-points. Fitbit estimates were calculated via Fitabase. Only data during waking hours where both PA monitors were worn were analyzed.

Fitbit and GTX+ estimates strongly correlated. Fitbit produced similar SED (mean difference = -35.83 minutes (min)/day), but significantly higher MVPA (mean differences = -59.7 – 77.41 min/day) to GT3X+. As the mean volume of MVPA increased, so did differences between Fitbit and GT3X+ estimates.

ACKNOWLEDGMENTS

With great enthusiasm and humility I wish to express my sincere gratitude to those that have supported me along the journey to completing this degree. First, I wish to thank my best friend and wife, Katie, for her patience, encouragement, and commitment through the many turns of this endeavor. Thank you to my family members who have invested time, money, and energy into my education, and actively shared in the process with me. Without their steadfast support, this accomplishment would not be possible.

Thank you to my close friends and professional colleagues who have been patiently understanding of my education goals and enabling me to pursue them.

Lastly, I want to thank my mentor, Dr. Wonwoo Byun for taking me on as his student, believing in my potential, and putting valuable challenges in front of me and guiding me through them. I also want to thank Sarah Greterman and Dr. Gary Liguori, who from the start, inspired me to pursue this degree. I also wish to thank the supportive faculty who have actively supported me throughout this process, my committee members Dr. Bryan Christensen, Dr. Mary Larson, and Dr. Clayton Hilmert, as well as Dr. Donna Terbizan, Dr. Margaret Fitzgerald, and the faculty of the Health, Nutrition, and Exercise Sciences Department. I am truly grateful for these committed educators who have given me so many rich experiences as a graduate student.

Thank you.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF ABBREVIATIONS viii

INTRODUCTION 1

REVIEW OF LITERATURE 7

METHODOLOGY 44

MANUSCRIPT 48

SUMMARY 63

REFERENCES 66

LIST OF TABLES

Table Page

1. Participant characteristics by occupational status. 53

2. Mean valid wear days and mean daily minutes of MVPA and SED. 54

3. Pearson Correlations between Fitbit and GT3X+ SED and MVPA estimates. 54

4. Mean differences between GT3X+ and Fitbit SED and MVPA estimates. 55

LIST OF FIGURES

Figure Page

1. Bland-Altman plots illustrating level of agreement between GT3X+ and Fitbit SED and PA classification estimates. Dashed lines show 95% limits of agreement (± 1.96 SD). 56

2. Bland-Altman plots illustrating level of agreement between GT3X+ MVPA and Fitbit MVPA classification estimates. A) Freedson cut-points B) Troiano cut-points. Dashed lines show 95% limits of agreement (± 1.96 SD). 57

LIST OF ABBREVIATIONS

ACLS ....Aerobics Center Longitudinal Study

BRFSS...Behavioral Risk Factor Surveillance System

DLW ....Doubly-labeled Water

EE..Energy Expenditure

IC...Indirect Calorimetry

LIPA..Light Intensity Physical Activity

LTPA...Leisure-time Physical Activity

MET ....Metabolic Equivalent

MVPA...Moderate-to-Vigorous Intensity Physical Activity

NHANES ....National Health and Nutrition Examination Survey

NHIS ....National Health Interview Survey

PA ....Physical Activity

SED...Sedentary Behavior

INTRODUCTION

Physical activity (PA) is essential for maximizing well-being and reducing the burden of chronic disease. Indeed, the many health benefits associated with PA are well documented.[1,2] Performing PA may lower resting blood pressure, improve fasting blood glucose levels, reduce depressive symptoms, and enhance quality of life.[3,4] In addition, engaging in regular PA may protect against disease and premature death.[5] Results from epidemiological studies demonstrated a clear dose-response relationship between PA and all-cause mortality. Findings from the Harvard Alumni and Aerobics Center Longitudinal studies showed that adults with higher levels of PA have a lower risk of pre-mature death.[6-8] Subsequent studies demonstrated this protective effect extends to disease morbidity. Prospective studies, such as the Nurses' Health study, provide evidence of PA's protective effect against pre-mature mortality in women. Adults engaging in little to no PA are at greater risk than active counterparts for chronic diseases such as CHD,[9] type 2 diabetes (T2D),[10] hypertension,[11] and certain cancers, such as breast cancer in women and colon cancer in both men and women.[12,13] These results join a wide body of evidence supporting the notion of a dose-response relationship between habitual PA and health benefits.[14-16]

As a result of accumulated evidence on relationships between PA and health, researchers and public health officials recognized the advantages of engaging in regular PA. In 1995, Centers for Disease Control and Prevention (CDC) and American College of Sports Medicine (ACSM) published PA recommendations emphasizing the value moderate intensity PA and not merely exercise which typically consisted of vigorous intensity PA aimed at improving fitness.[17] These recommendations proposed that all adults acquire at least 30 minutes of moderate intensity PA on most or all days of the week. The American Heart Association (AHA) and ACSM updated these recommendations in 2007, highlighting the significance of moderate-to-vigorous PA (MVPA). The update broadened the previous CDC/ACSM recommendations by endorsing a minimum weekly volume of MVPA achieved in bouts ≥ 10 minutes. This amount of weekly PA was considered the minimum needed to obtain health benefits from PA. However, adults were encouraged to exceed the minimum dose to realize further health benefits and protect against undesired weight gain.

Supported by previously published PA recommendations, in 2008, the Department of Health and Human Services published the first ever Physical Activity Guidelines for Americans.[18] The guidelines outlined the benefits associated with regular PA and established a weekly PA prescription for adults as well as guidelines for other populations such as adolescents and older adults. Similar to the previous recommendations, the current PA guidelines set forth a minimum weekly PA volume of "health-enhancing" PA, differentiating this from "baseline" PA which was simply PA limited to low-intensity activities of daily living. Adults could achieve the minimum health-enhancing PA through: 1) ≥150 minutes of moderate intensity PA per week, 2) ≥ 75 minutes vigorous intensity PA per week, or 3) accumulating any equivalent combination of MVPA, performed in bouts lasting ≥ 10 minutes. Individuals who do not accumulate this recommended level of weekly PA are considered "inactive", whereas PA beyond the recommended level is encouraged as it confers greater health benefits. To simply communicate this PA recommendation, a popular adage emerged stating some PA is good, more is better.[5]

Despite the substantial efforts to promote PA, most Americans do not adhere to these PA guidelines.[19,20] Data from surveillance systems such as the National Health and Nutrition Examination Survey (NHANES) or National Health Interview Study (NHIS) show less than half of the U.S. adult population are adequately active.[20,21] However, surveillance data has typically relied on subjective measures of PA (i.e., self-reported questionnaires) and limitations of self-report are evident including misinterpretations of the meaning of survey questions and social desirability of responses.[22,23] Given PA was measured by self-reported questionnaires in these surveillance systems, levels of PA in American adults can be significantly lower if objectively surveyed.[24,25] In fact, recent studies utilizing objective PA measures suggest less than 10 % of the US adult population actually achieve the minimum PA guidelines, whereas rates from self-report were near 60%.[20,26]

Several objective instruments are available for measuring PA. Doubly Labeled Water (DLW) method is widely accepted as the gold standard measure of physical activity energy expenditure (PAEE) in free-living conditions. However, DLW is expensive and cumbersome because it requires collection and processing of urine samples. Indirect calorimetry (IC) measures the ratio of oxygen and carbon dioxide consumption and production to provide estimates of energy expenditure (EE). IC has been used in laboratory settings, but recently developed portable models facilitate its use in field settings. Compared with DLW technique, portable IC is less invasive and provides measures across all dimensions of PA including frequency, intensity, duration, and energy expenditure. For these reasons, portable IC offers considerable promise as a criterion method for studies that validate various types of physical activity monitor in field-based settings.

Accelerometer-based PA monitors are small, lightweight, and inconspicuously worn, requiring minimal subject burden and can detail PA intensity, duration, and frequency. Accelerometers are particularly appealing for monitoring PA in free-living settings. Due to improvements in battery life and memory capacity, many accelerometers can provide continuous PA monitoring for multiple weeks. ActiGraph accelerometers are one of the most widely used to measure PA in free-living conditions.[27,28] The earlier models of ActiGraph accelerometer were uniaxial (1 vertical), thus only capable of detecting vertical accelerations.[29] However, the most recent model of ActiGraph accelerometer, GT3X+, detects movement in three planes (vertical, anteroposterior, and mediolateral) with well-established accuracy.[30-32] Due to its high validity and feasibility, GT3X+ was the method of choice for measuring PA in NHANES 2011 – 2014, which is one of the most representative surveillance system in the U.S.[33]

The emergence of accelerometer-based PA monitors has been well-received the growing availability of similar accelerometers in the consumer electronics market. Many commercially-marketed PA monitors are designed to provide immediate PA feedback, helping consumers monitor their own daily PA levels.[34] A recent study examined the accuracy of several most popular consumer-based PA monitors in measuring PAEE in a lab setting. Among eight different monitors, the Fitbit monitors were second and third most accurate, following only the Sensewear Armband, when the estimates of PAEE were compared with IC.[35] However, this study only assessed pre-determined activities, mostly common examples of leisure-time PA (LTPA), such as various walking or jogging. These activities were performed in-succession for only 5 minutes at a time in a laboratory setting. Though the Fitbit monitors appeared to produce similar PAEE estimates as IC, it is unknown whether this accuracy would persist in free-living conditions in which individuals perform unscripted activities.

Furthermore, few studies have examined the accuracy of wrist-worn PA monitors. One advantage of wrist-worn accelerometers is their ability to detect upper-body movement while still providing reasonably accurate estimates of PA in various populations. For example, the GT3X+, placed at the wrist, accurately predicted PA energy expenditure compared to IC in manual wheelchair-bound subjects.[36] Other studies have established the validity of wrist-worn accelerometers for estimating PA in other populations such as children[37] and pregnant women.[38] Additionally, wrist-worn accelerometers may appear less burdensome to users than other bodily placements, potentially increasing the feasibility for PA monitoring research.[39] To this point, there has been a rapid increase in the sales of wrist-worn accelerometer-based PA monitors, with the Fitbit Flex among the top selling models on the consumer market.[40] Nonetheless, the validity of this device for estimating PA is unknown.

Purpose of the Study

The purpose of this study was to determine the concurrent validity of the Fitbit Flex accelerometer in a free-living condition. Specific aims of this thesis were:

1. To determine the correlations of SED and MVPA measurements between Fitbit Flex and ActiGraph GT3X+.
2. To compare the mean differences in SED and MVPA estimates between the Fitbit Flex and ActiGraph GT3X+.
3. To examine the agreement of SED and MVPA estimates between the Fitbit Flex and ActiGraph GT3X+.

Scope (Delimitations)

Participants were student, faculty, and staff population at North Dakota State University. We used a previously validated accelerometer, ActiGraph GT3X+, as the criterion.

Significance

Several subjective and objective methods are available to measure PA. There are a number of well-known weaknesses with subjective assessment methods.[41] Recall bias and error may lead to over- or underestimation of PA. In addition, subjective methods may lack the accuracy to detail specific components PA such duration or intensity. Current practices in PA research favor objective measurement of PA variables.[28,30] Recent trends in consumer electronics suggest that laypersons are interested in personal monitors that provide user-friendly feedback of PA variables. This "data-driven" movement may present an opportunity to promote PA through self-monitoring.[42,43] However, the validity of these monitors is not well established. Therefore, results from this study will provide evidence for the accuracy and utility of new objective PA monitoring options for field-based studies.

Limitations of the Study

1. The accelerometers are not waterproof. Subjects were asked to remove them when participating in water sports or bathing.
2. Hip-worn accelerometers have limited ability to measure upper body movement.

Definition of Terms

1. Physical activity: Any bodily movement that is produced by skeletal muscle contractions, which significantly increases energy expenditure.[44]
2. Sedentary behavior (SED): Physical activity that is characterized by an absolute rate of energy expenditure ≤1.5 MET.[45]
3. Moderate-to-vigorous physical activity (MVPA): Physical activity characterized by an absolute rate of energy expenditure between 3.0 and 6.0 METs.[17]
4. Accelerometer: devices that detect and measure the velocity of position displacements with respect to reference axes over time.[46]
5. Metabolic equivalent (MET): An estimate of the rate of energy expenditure of physical activity. One MET represents a metabolic cost of 3.5 ml/kg/min.

REVIEW OF LITERATURE

Health Benefits of Physical Activity

Physical Activity (PA) is any skeletal muscle movement that requires energy expenditure (EE),[44] which includes many types of intentional and unplanned activities at a variety of intensities in occupational and leisure-time settings.[2] Performing PA improves plasma glucose and blood pressure regulation and may attenuate depressive symptoms.[3,47] In addition, PA may reduce the risk of premature mortality as well as morbidity of chronic diseases such as coronary heart disease (CHD),[48] type 2 diabetes,[49] and stroke.[50] In a landmark study, Morris and colleagues found that men with occupations requiring routine PA were less likely to die of CHD than their sedentary co-workers.[51] In the Harvard Alumni study, men who regularly engaged in leisure-time PA (LTPA) were also less likely to die from any cause compared to their least active peers.[6] Other large sample studies have also demonstrated that regularly active women are less likely to die from all-causes and cardiovascular diseases compared to inactive women.[52,53] Overall, research consistently demonstrates PA confers health benefits to both men and women, with more PA resulting in greater health benefits.[54]

Early PA research presented the notion of a dose-response relationship between PA and longevity. Results from the Harvard Alumni study revealed a graded inverse association between LTPA and all-cause mortality. Compared to inactive men reporting less than 500 kilocalories/week of LTPA, men expending more than 500 kilocalories/week through moderate-to-vigorous intensity PA (MVPA) experienced a 27% reduced risk of all-cause mortality, with no further benefits from higher volumes of PA (> 2500 kcal/week) on both all-cause and cardiovascular disease mortality.[6] A subsequent study with the same cohort examined the influence of PA intensity on all-cause mortality.[55] Self-report activities were classified into light, moderate, and vigorous intensity based on metabolic equivalent (MET) values established in the compendium of physical activities. The investigators used these MET values and time estimates to calculate the participants' weekly PAEE (kcal/week).[56] When considering different patterns of PA by intensity, only vigorous intensity PA upheld a consistent inverse association with all-cause mortality in men across EE categories. Light intensity PA was not associated with a protective benefit, whereas men experienced the greatest benefit from moderate intensity PA volumes between 760 – 1400 kcal/week, with additional PA yielding no additional benefit. Similarly, in the Aerobics Center Longitudinal Study (ACLS) cohort, "highly active" men (reporting some vigorous intensity PA) were less likely to die from any cause compared to moderately active (some moderate intensity but not vigorous intensity PA) or inactive men (no MVPA).[57]

An inverse dose-response relationship between cardiorespiratory fitness (CRF) and longevity was also reported in the ACLS. Researchers evaluated CRF of healthy men and women (ages 20 – 88 y.) by a maximal treadmill exercise test with duration of the test (in seconds) serving as the indicator of CRF. Age- and sex-specific quintiles of CRF were established, with the lowest fifth in each age and sex group were assigned into the low fitness category, with increasing fitness group constituting the second through fifth quintiles. The quintiles were further grouped into low (1st quintile), moderate (2nd and 3rd quintiles), and high (4th and 5th quintiles) fitness categories.[8,58] Subjects were followed up to 15 years to determine deaths.[8] Risk of death from any causes was three times higher in low fit men (RR = 3.44, CI 95% [2.05, 5.77]) and over four and one half times higher in women (RR = 4.65, 95% CI [2.22, 9.75]) in the low fit group (the lowest 20%, 1st quintile) compared to the highly fit group (the highest 20%, 5th quintile). The observed inverse relationship between CRF and risk of all-cause mortality persisted even after adjusting for other risk factors for CVD such as smoking, alcohol consumption, diabetes, and hypertension.[58] Although there was a U-shaped relationship between all-cause mortality and CRF in women who were known smokers, had high blood cholesterol, or reported health status as unhealthy, all-cause mortality was significantly lower for women with moderate CRF compared to low CRF. These results suggest low fitness is a considerable risk-factor all-cause mortality.

Evidence suggests PA not only protects against all-cause mortality, but also pre-mature mortality from specific causes such as stroke or certain cancers. In the Harvard Alumni study, researchers assessed self-reported LTPA in 11,130 male subjects at baseline and at 11-yr. follow-up. Weekly EE (kcal/week) was calculated and subjects were categorized into quintiles of EE. Compared to the least active group (< 1000 kcal/week), moderately active men, expending between 2000 – 2999 kcal/week, had nearly half the risk of death from stroke (RR = 0.54, 95% CI [0.38, 0.76]).[59] Similarly , risk for stroke is inversely associated with PA in women.[60] Researchers assessed LTPA in female subjects (N = 72,488) at baseline and every other year of an 8-year follow-up period. Subjects were then categorized into quintiles of weekly LTPA volume (MET h/week) based on self-reported PA. Relative risk of death from ischemic stroke was inversely associated with LTPA, even after adjusting for several covariates such as age, BMI, menopausal status, hypertension, and family history of heart attack.

Risk of cancer mortality is generally inversely associated with PA, but, the strength of the evidence depends on the type of cancer and sex.[61] Evidence from the ACLS cohort suggests CRF protects against risk of death from all cancers in men.[62] Compared to the lowest CRF group, relative risk of death from cancer in the second lowest CRF category was 0.54 (95% CI 0.35 – 0.84). In male subjects, further increases in CRF conferred additional protective benefit against any cancer death. The same study also investigated this relationship in a cohort of women but the inverse relationship between risk for cancer mortality and CRF did not reach significance. However, PA has been shown to reduce the risk of death from breast cancer. Rockhill and colleagues observed a dose-response effect of MVPA on risk of breast cancer in a large cohort of women representing nearly 2 million person-years.[63] Compared to women achieving less than one hour of MVPA/week, those with more weekly MVPA saw a 12 – 18% reduced risk of breast cancer death over a 12-year period. McTiernan and colleagues reported similar findings in a prospective study of over 74,000 post-menopausal women.[64] Past lifetime and recent MVPA were both inversely associated with risk of breast cancer. As with the Rockhill et al. study, the magnitude of risk reduction was greatest for women reporting the most PA (more than 7 hours MVPA/week).

In men, the risk of lung cancer is inversely associated with PA. In a cohort of Norwegian men (n = 53, 242), subjects reporting usual moderate-to-vigorous LTPA of at least 4 hours/week experienced a 25% reduced risk of lung cancer compared to sedentary counterparts.[65] A further decrease in risk was observed in the most active men reporting engaging in at least 4 hours of weekly MVPA for the purpose of maintaining fitness or competitive sports (RR = 0.71 95% CI [0.52, 0.97]). Results from the Harvard Alumni study further support the notion of a dose-response effect from PA on lung cancer. In a group of 13,905 men (mean age = 60.2 ± 9.5 yr.), researchers observed a significant inverse trend in relative risk of lung cancer in subjects expending 1000 – 1999, 2000 -2999, and ≥3000 kcal/week in MVPA compared to those expending less than 1000 kcal/week.[13] Relative risks were 0.87 (95% CI 0.64 – 1.18), 0.76 (0.52 – 1.11), and 0.61 (0.41 – 0.89) (P for trend = .008) across categories of weekly MVPA EE ≥ 1000 kcal compared to the reference category (< 1000 kcal/week).[66] Even after controlling for covariates such as smoking status and smoking intensity, the dose-response trends remained significant.

Research has highlighted the potential health benefits of PA, yet much of, the U.S population remains inactive resulting in a substantial burden to public health. Chronic diseases constitute over half of the top 10 causes of death in the US.[67] Considering the fact that recent estimates suggest nearly 10% of all deaths in the US are at least partially attributable to physical inactivity, the need to promote PA cannot be overemphasized.[68]

Physical Activity Recommendations and Adherence

In 2008, the Department of Health and Human Services issued the first ever physical activity guidelines for Americans.[18] These guidelines, building on previously published PA recommendations,[1,2,17] provided a comprehensive PA reference for adults. In brief, they advise adults achieve a minimum weekly volume of 150 min of moderate intensity PA or 75 min of vigorous intensity PA or combination thereof. The current PA guidelines also indicate that bouts of PA should last no less than 10 minutes per session, PA should be distributed over multiple days per week, and adults should strive to avoid inactivity. The guidelines also provided broad classifications differentiating levels of weekly MVPA: highly active (>300 minutes of moderate-intensity PA or MVPA equivalent), sufficiently active (achieving 150 - 300 minutes of moderate-intensity PA or MVPA equivalent), insufficiently active (less than 150 minutes of MVPA equivalent) and inactive (no MVPA bouts ≥ 10 minutes). Evidence suggests achieving the minimum weekly dose of MVPA is associated with multiple health benefits including reduced risk for premature death, coronary heart disease, and type 2 diabetes.[18] Additional health benefits, such as weight loss or maintenance, are possible but adults may need to exceed the minimum MVPA threshold and possibly include dietary modifications.[18]

Despite the clear benefits of PA, years of surveillance data show the majority of U.S. adults do not achieve the recommended level of PA.[2,17,21,26] Large surveillance systems such as the Behavioral Risk Factor Surveillance System (BRFSS), the National Health and Nutrition Examination Study (NHANES), and the National Health Interview Survey (NHIS) have provided trends of PA participation in U.S. adults. The surveys collect PA data via self-report interviews. Based on 2001 and 2003 BRFSS data, the CDC estimated that about 54% of the U.S. adult population was insufficiently active (less than 150 minutes/week of moderate-intensity PA).[69] The 2012 NHIS data approximate 50% of the U.S. adult population meet the minimum aerobic PA recommendations, but only 21% met both aerobic and muscle strengthening recommendations.[70] Similarly, data from NHANES 2003 -2004 estimates 51% of the adult population met the recommendation of 150 minutes/week of moderate-intensity PA.[20] However, using accelerometer-determined data from a subgroup of the 2003 – 2004 NHANES population sample, Troiano and colleagues reported adherence rates below 5% for both adult males and females.[20]

Recent surveillance data have substantiated the discrepancies between self-report and objectively-determined population PA patterns. Tucker, Welk, and Beyler compared NHANES 2005 – 2006 self-report and accelerometer-determined PA estimates.[26] The authors compared adherence rates to the U.S. PA Guidelines using three separate PA analyses criteria. Individuals were categorized as performing "adequate" PA if they achieved 1) ≥ 150 min. MVPA / week 2) ≥ 150 min. / week moderate PA plus at least two bouts of vigorous intensity PA, and 3) accumulated ≥ 500 MET-minutes / week of MVPA. The first two analyses calculated MVPA when performed in bouts lasting 10 or more consecutive minutes. The third analysis considered total minutes of all MVPA regardless of bout duration. Using the first two criteria, estimates of adherence to the guidelines were 62.0 % for self-report and less than 10% for accelerometer-determined PA. Even using the more liberal PA criteria (sum of all minutes of MVPA regardless of bout duration), less than half the population achieved the minimum recommended weekly dose of MVPA.

Healthy People 2020 set a goal of increasing the proportion of the adult population achieving both the minimum PA guidelines as well as the "highly active" criteria of > 300 minutes per week of moderate or > 150 minutes per week of vigorous, or equivalent MVPA combination. Through the U.S. PA Guidelines, public health officials have sought to translate the evidence-based health benefits of PA into manageable, actionable steps. Recent reduction in the proportion of U.S. adults maintaining an inactive lifestyle may suggest the public is responding, yet PA levels of the U.S. adult population has remained relatively unchanged, with less than 50% meeting minimum PA Guidelines.[71] Many occupational settings have become increasingly sedentary.[72] In addition, the appeal of leisure-time sedentary activities may discourage PA.[73] Furthermore, recent population surveillance studies demonstrate that these figures based on self-report data may greatly overestimate actual population PA levels.[20,26] This national PA initiative may benefit from increased feasibility of objectively measuring PA to enhance population PA surveillance and improve PA promotion efforts.

Methods of Assessing Physical Activity

Subjective Physical Activity Measures: Self-report Questionnaires

Researchers have used a number of subjective and objective methods to assess PA. Large-scale surveillance systems and prospective studies have relied upon self-report methods, for assessing PA. The advantages of self-report PA measures include minimal subject burden and feasibility for administering to large subject samples. Several disadvantages associated with self-report PA include subject recall bias and question interpretation error, social desirability, large variances in PA outcomes between self-report instruments, as well as limited accuracy and validity.[41] Self-report instruments vary by the recall period, type of activity, means of administration, and PA outcomes measured. Thus, the validity and reliability of PA questionnaires needs to be established before use in research. The subsequent paragraphs describe characteristics of specific PA questionnaires.

The Stanford Seven-day Physical Activity Recall questionnaire (PAR) is an interviewer-administered self-report instrument. The PAR has respondents daily time spent sleeping and engaging in moderate, hard, and very hard activities over the past week. Researchers may calculate total weekly PA, minutes spent in PA intensities, as well as EE from PA. The reliability and validity of PAR has been assessed in a variety of populations and with several objective criteria, with reliability estimates ranging from $\rho = 0.08 – 0.99$.[74] Sallis and colleagues demonstrated the PAR had reasonable reliability for estimating hard and very hard PA ($\rho = 0.31$, $\rho = 0.61$, $p<.05$, respectively), but not for moderate PA.[75] Hayden-Wade and colleagues found participants tended to overestimate PA compared to accelerometer-determined (TriTrac) PA.[24] Correlations coefficients between accelerometer and both in-person and telephone-administered PAR PA estimates were similar ($r = .41$ and $r = .43$ [no p-value reported], respectively). Correlations between accelerometer and PAR PA estimates were considerably stronger for very hard PA compared to hard and moderate PA for both in-person ($r = 0.74, 0.43, 0.33$, respectively) and over-the-phone recall ($r = 0.78, 0.39, 0.26$, respectively). In other studies the PAR has demonstrated significant correlations with the Caltrac accelerometer as criterion for total and vigorous PA, whereas estimates of moderate activities have been equivocal.[76,77]

The Minnesota Leisure-time Physical Activity Questionnaire assesses usual LTPA during the past 12 months. An interviewer inquires on subjects' typical participation in sporting and recreational activities from a pre-determined list. Interviewers are trained to prompt respondents estimate PA duration and frequency over the course of a year.[78] These estimates are multiplied by a pre-assigned intensity value and calculated to estimate weekly PA levels or EE by PA intensity category. Total PA from the questionnaire have correlated moderately with accelerometer-determined PA as well as CRF (Vo2 max).[78] However, the instrument appears to lack sensitivity to both moderate and light intensity PA compared to 48-hr PA logs.[78] Likewise, in a group of Spanish women (n = 250, 18 – 61 yr.) self-report vigorous intensity PA estimates demonstrated strong correlation with duration of a treadmill maximal fitness test compared to moderate PA ($\rho = 0.51, 0.13, p < 0.05$, respectively), while light intensity PA did not significantly correlate with treadmill time.[79]

The International Physical Activity Questionnaire (IPAQ) is a self-report instrument with multiple short- and long-form versions. The IPAQ can be self-administered or via telephone correspondence and has been developed for use in multiple countries. Both long (31 items) and short-form (nine items) IPAQ provide estimates of weekly PA (min) minutes categorized by activity and intensity. EE estimates are calculated by multiplying minutes of PA by MET values assigned to each activity. Craig and colleagues examined the reliability and validity of multiple short and long-form IPAQ versions.[80] Repeatability were similar for both long and short-forms (pooled $\rho = .81$, [95% CI 0.79 – 0.82] $\rho = 0.76$ [95% CI 0.73 – 0.77], respectively). IPAQ validity was assessed using PA output from a hip-worn ActiGraph accelerometer (then operating as Computer Science and Applications, Inc. [CSA]) as the criterion. Total PA from both long and short-forms showed modest correlations with CSA-determined PA (pooled $\rho = 0.33$ 95% CI 0.26 – 0.39] and $\rho = 0.30$ [95%CI 0.23 – 0.36]).

Subsequent research on the IPAQ suggested that the instrument may not be sensitive to moderate PA. Hagstromer and colleagues compared previous 7-day PA estimates from IPAQ (long form) to accelerometer in 46 male and female subjects.[81] Questionnaire-determined vigorous intensity PA (VPA) significantly correlated with ActiGraph-determined VPA estimates ($\rho = 0.63$, $p<.001$). When considering MVPA, the correlation between subjective and objective instruments was attenuated, but remained significant ($\rho = 0.36$, $p<.001$). When only moderate intensity PA was examined, the relationship was no longer significant.

National surveys within the US also collect PA data through self-report. Most notably, the Behavioral Risk Factor Surveillance System (BRFSS) is the largest U.S. surveillance system and annually gathers state-specific, and nationally representative, outcomes on various risk factors related to both communicable and non-communicable disease. Though state-specific data is collected, nationally representative outcomes can be derived from BRFSS.[71] However, fundamental differences in assessing PA may affect its ability to accurately estimate national PA adherence compared to NHANES and NHIS. In 2001 - 2007, the BRFSS PA questions prompted responses regarding the weekly duration and frequency of both moderate and vigorous LTPA. Prior to 2001, and more recently (2008 – 2010), the BRFSS assessed non-occupational PA with one dichotomous item.[82] Comparing 2005 PA estimates from all three national surveillance systems, Carlson et al reported substantially larger estimates of the proportion of the US adult population meeting the Healthy People 2010 PA objectives according to BRFSS than NHIS or NHANES.[71] Further analysis demonstrated these potentially inflated estimates persisted from 2001 – 2007.

Compared to objective PA measures, researchers have reported varying validity of BRFSS questionnaires for assessing PA. Strath and colleagues compared 7-day PA outcomes between the 2001 BRFSS and a accelerometer plus heartrate monitor in 25 subjects.[83] The authors reported no significant differences in mean MVPA minutes/day between BRFSS and monitor output. In addition, the BRFSS demonstrated significant agreement with objective criterion for identifying subjects meeting the 2010 healthy people PA recommendations for moderate (≥30 minutes/day, ≥ 5 days/week) or vigorous intensity PA(≥20 minutes/day, ≥ 3 days/week) ($\kappa = 0.40$ and 0.58, respectively). However, results of this study are limited as subjects wore the ActiGraph (then operating as MTI) accelerometer on the wrist and thigh and the PA was assessed for a shorter period than typically calculated by the BRFSS.

Yore and colleagues compared BRFSS and objective measure PA outcomes in 55 subjects over 22 days.[84] Researchers examined subjects' self-report PA, using BRFSS 2001 survey, during first (baseline and second survey) and third weeks (final survey). Subjects also wore a pedometer and hip-mounted ActiGraph accelerometer for seven consecutive days during the second week. Correlations between BRFSS and ActiGraph suggest self-report is more closely related to objectively-measured, moderate versus vigorous intensity PA ($\rho = 0.17 – 0.26$; $\rho = 0.52 – 0.63$, respectively). Contrasting with the findings from Strath et. al, BRFSS estimates exhibited relatively poor agreement with ActiGraph in discriminating those achieving Healthy People 2010 PA objectives on both first and third surveys for moderate intensity PA ($\kappa = 0.31$, 95% CI [0.09, 0.53]), and particularly for vigorous intensity PA ($\kappa = 0.17$, 95% CI [0.01, 0.32]; $\kappa = 0.26$, 95% CI [0.10, 0.43], respectively).

The limitations of self-report are established. As illustrated in the aforementioned studies, compared to objective measures, self-report PA assessments typically have wide outcomes ranges within and between subjects. Variability may be largely due to recall error. Individuals tend to inaccurately recall PA less intense PA more so than vigorous intensity PA. In addition, lengthy recall periods may be particularly prone to shorter recall self-assessments. These may be particularly problematic for researchers trying to calculate PA outcomes since many adults spend a majority of waking hours in sedentary to moderate activities. In addition, PA at these intensity levels may accrue in sporadic bouts whereas vigorous activities are typically structured or planned. Thus, researchers have explored other means of quantifying PA.

Objective Physical Activity Measures: Indirect Calorimetry and Doubly-Labeled Water

Indirect calorimetry (IC) is a commonly used objective monitoring technique for assessing PA and estimating EE in both laboratory and field settings. Indirect calorimetry is not a direct measure of EE (e.g. heat loss), but rather, allows calculations of EE from direct measures of gas exchange. That is, IC measures the volumes of ventilated respiratory gases; the composition of these expired gases can be analyzed to determine the oxygen (O_2) and carbon dioxide (CO_2) content and calculate oxygen uptake ($\dot{V}O_2$) as well as CO_2 production ($\dot{V}CO_2$).[85] Because the metabolic cost of substrate utilization is known, approximations of EE from gas exchange measurements are possible.[85] Traditional approaches to IC were often limited to lab settings (i.e. hood ventilator techniques), or required cumbersome, time sensitive collections of expired air in bags for later analysis (i.e. Douglas bag).[85,86] Due to technological advancements, computerized IC units, such as metabolic carts, are now preferred to the Douglas bag technique as they allowed for continuous, real-time monitoring of respiratory gases. Barring errors in calibration, computerized IC output has been show to yield equivalent values to the Douglas bag for total gas expired or inspired, fractions of O_2 (F_eO_2 and CO_2 (F_eCO_2) in expired gases, as well as calculations for $\dot{V}O_2$, minute ventilation, and respiratory exchange ratio (RER).[87] Even with automated calibrations and calculations, metabolic carts were historically confined to laboratory settings due to limitations in portability.[87]

Recently, portable IC models have been developed to allow for mobile PA assessment. The Cosmed K4b^2 and Care Fusion Oxycon Mobile are two commonly used portable IC devices used in PA research. Whereas early versions of these portable IC were limited to measuring ventilation over discrete time periods, both K4B^2 and Oxycon Mobile models are able to yield breath-by-breath measurements of expired gases. With these capabilities, the K4b^2 has demonstrated equivalent $\dot{V}O_2$ estimates compared to Douglas bag during indoor stationary cycling at work intensities < 250 watts, and during outdoor running at speeds ranging from 8 – 22 km/h.[88,89] Similarly, Oxycon Mobile's $\dot{V}O_2$ measurements are comparable to DB during indoor cycling at submaximal and maximal efforts.[90] Oxycon Mobile has further demonstrated accurate and stable metabolic measurements compared to DB criterion in high-wind and humid environmental settings.[91]

Despite its accuracy and utility to serve as criterion reference for determining the validity of other PA monitoring methods,[35,92,93] portable IC instruments are not well-suited for large-scale, free-living PA monitoring. Software needed to process portable IC output is expensive and the units themselves may be deployed for limited periods due to battery life. Compared to other smaller portable PA monitors, portable IC units are not inconspicuous. These drawbacks reduce the feasibility of using IC to monitor PA of larger populations in free-living settings.

The doubly labeled water (DLW) method is considered the gold standard method for determining EE in free-living settings. The method dates back as early as the 1950s when used in animal studies,[94] yet it was Schoeler and colleagues who first fully utilized the technique to estimate EE in human subjects in the early 1980s.[95,96] The technique requires subjects to ingest baseline amounts of isotopes ^{2}H and ^{18}O and subsequent analysis of the labeled isotopes in urinary samples gathered at baseline (shortly after ingesting the isotope) and after a set period, typically 4 – 21 days.[97] The proportion of these elements remaining in the urine is representative of the amount of carbon dioxide expelled due to metabolic processes.[94] Equations have been formulated to determine the total amount of EE over a given period.[96,98] To determine the quantity of EE from PA each subject's resting metabolic rate and energy intake must be considered.[99] Therefore, with careful dietary monitoring and laboratory analysis of the residual isotope samples, EE estimates are possible.

The accuracy and minimal subject burden of the DLW technique makes it favorable for collecting data in free-living settings, but due to high cost and inability to determine precise daily patterns of PA, it is not widely utilized in large scale free-living PA research. Studies have reported mean group-level energy expenditure estimates from DLW falling within five percent (kcal/day) or less compared to IC-determined or food intake and body mass composition criterion.[95,96,100] Subsequently, DLW has served as a criterion to assess the validity of EE estimates from other PA assessment tools such as accelerometers, self-report instruments, and heartrate (HR) monitors.[27,99,101] However, DLW cannot detail the patterns of PA associated with these EE estimates. Additional PA assessment tools can overcome this primary disadvantage, but necessitate additional costs and subject compliance. In addition, the technique is expensive given the cost per use per subject of the ^{2}H and ^{18}O isotope solutions.

Objective Physical Activity Measures: Heartrate Monitors

Heartrate (HR) monitoring is a technique that has been widely used to objectively measure PA in both laboratory and field settings. It is well-established that HR increases linearly with oxygen uptake during PA, particularly PA incorporating whole-body movements.[102] Since most HR monitors can record HR data over a period of hours or days allowing calculations of PA duration, intensity, and frequency, they are a convenient way to monitor PA.

Despite the convenience and ease of deployment in field settings, HR monitoring carries limitations for objective PA monitoring. Several factors may impact individual heartrate with little impact oxygen consumption. Training status, age, ambient temperature, and psychological distress impact heartrate during rest.[103-105] An individual's hydration and training status, quantity of muscle mass recruited for movement, and body position all affect heartrate during PA.[106-109] Nonetheless, researchers have employed HR monitoring with some efficacy for assessing group-level PA.

Early studies attempted to predict total daily EE from HR output. Individual regression prediction equations were derived from comparing HR output to a criterion, such as IC, during PA and sedentary activities. Daily HR data were then recorded and mean HR data used to estimate daily total EE. Because the method did not account for sources of daily HR variability even at relatively similar levels of PA and sedentary behavior (SED), it proved unreliable for predicting EE.[106]

The development of the "heartrate flex" (HRFlex) method improved EE estimates from HR data. The technique involves individual calibration of heartrate data. Researchers determine a minimum heartrate threshold for each subject by comparing HR data to a criterion during a graded exercise protocol. The method asserts that below this HRFlex threshold a subject's EE is equivalent to resting EE, which is estimated through standard calculations of resting metabolic rate.[110] Above that threshold, HR is assumed to rise in a relatively linear fashion with increased oxygen consumption. Given thorough calibration, there may be little difference between HR-predicted and IC-determined EE.[92,111] Ceesay and colleagues demonstrated a high correlation between actual and HR-estimated EE in twenty subjects performing PA and various sedentary activities over a 21-hour period (r = 0.94, SE = 109 kcal/day). At the group level, mean EE estimates were minimal, with a mean nonsignificant difference of -1.2% between indirect calorimetry and HRFlex methods.[111]

Adequately accurate EE estimates are possible from HR data, but even individually calibrated predictions equations may exhibit wide variability, particularly within individuals. Livingstone and colleagues demonstrated wide individual variability in EE estimates using the HRFlex method compared to DLW in adults and adolescents. In the adult group (n = 14), mean EE estimates (kcal/day) from HR data were within two (+/- 17) percent (about 24 kcal/day) of DLW. However, individual differences in daily EE estimates between HR compared to DLW ranged from -22.2% to +51.2% underestimating EE by as much as 1149 kcal/day or overestimating as much as 1,618 kcal/day.[112] In a similar study of 36 adolescents (7 – 15 years old), Livingstone and colleagues demonstrated error rates between HR and DLW-determined EE (kcal/day) ranging from -16.7% to +18.8% or – 261 to + 330 kcal/day.[113]

Even though HR monitoring may produce reasonable group-level estimates of daily EE, it is vulnerable to daily sources of heartrate variability such as ambient temperature or type of PA performed. Researchers have conceded that the HR calibration methods work best for estimating EE based on the activities by which the calibration equations were produced.[106,111] Furthermore, individuals may spend most of their waking hours in sedentary or light-intensity activities resulting in a majority of HR output ranging near the HRFlex threshold where individually calibrated predictions equations are most likely to error.[111] In addition, the HRFlex technique requires time consuming individual calibration in order to be moderately effective. Such tedious calibration practices may not be feasible for large scale population monitoring. Some researchers have attempted to use alternative prediction models to estimate PA in larger groups. Though such methods may accurately rank subjects by PA level, there is wide variability between observed and predicted daily EE estimates, especially for individuals.[114,115] As such, without observation or time consuming individual calibration, it may not be feasible to accurately distinguish PA patterns and estimate EE from HR alone.

Objective Methods: Accelerometers

General Function

Accelerometers are instruments that are frequently employed as objective sensors for human activity monitoring. Early uses of accelerometer-based devices date back to the World War II era where they were mostly used in military aviation applications.[116] Today, accelerometers are widely manufactured and utilized in research and commercial applications for monitoring human movement. For example, accelerometers may be used in devices that used to provide biomechanical feedback to patients completing rehabilitation exercises;[117] monitor human sleep patterns;[118] quantify and promote daily PA.[119,120]

Technical Aspects and Functions of Accelerometer-based Activity Monitors

Accelerometer technology has advanced over time. Some of the early accelerometers, such the ActiGraph (formerly operating as Computer Science and Applications, Inc. (CSA) and later as Manufacturing Technology Incorporated (MTI)) 7164 model, contained piezoelectric cantilever beam sensors.[29] Industry improvements enabled new steady-state sensor technologies known as microelectromechanical system (MEMS), which featured greater memory capacity while still becoming smaller and more efficient.[121] These sensors operate using capacitive or piezoelectric sensors; accelerometers with capacitive sensors are more common than those with piezoelectric sensors.[46] The ActiGraph GT3X+, for example, is equipped with a capacitive accelerometer in which a reference mass is suspended by spring-like fixtures with two electrical plates opposing similar fixed plates. External forces can cause the mass to move resulting in a change in charge between the plates that is proportional to the intensity of the acceleration.[122,123]

Accelerometers operate by recording accelerations relative to the input force within distinct time intervals (also known as epochs).[46] In contrast to other motion sensors that utilized a switch-style sensors (e.g. pedometers), resulting in a dichotomous all-or-none output, accelerometer data reflect the frequency, duration, and intensity of all accelerations within a given range and direction.[124] Briefly, accelerometers that are most sensitive to inertial forces in a single plane of motion are described as uniaxial. Early approaches to accelerometer-based PA monitoring research maintained that vertical accelerations captured near the body's center of mass may best reflect the wide range of accelerations during ambulatory activities.[125,126] Thus, some of the earliest monitor PA monitors oriented their accelerometers to detect accelerations in the vertical plane.[127] The most common accelerometers are multiaxial, comprised of multiple sensing units oriented to two or more orthogonal planes of motion, namely the vertical (x), horizontal (y) and frontal (z) axes.[128] With multiaxial devices, acceleration data may be processed separately or in some combination reflecting the accelerations of two or more planes of motion. Some data suggests PA estimates from multiaxial devices is more accurate than those from uniaxial devices.[129,130] However, data processing may be more influential on PA estimates than the number of axis used to monitor PA.[27,28,30]

The initial recording of the acceleration signals is dependent on the accelerometer's parameters (e.g. dynamic range, sensitivity, and signal filtering) uniquely established by the manufacturer, though many devices designed for monitoring PA are similar in capacity.[131] Researchers may choose different monitors based on available resources, the intended outcome of interest or desired bodily placement.[124] It is beyond the scope of this review to compare the initial calibration differences between accelerometer-based PA monitors. Typically, they exhibit similar attributes, with battery and memory capacity allowing for multi-day deployment and data storage, dynamic ranges between 0.5 – 9.0 g, and frequency sampling up to 160 hz.[123,132,133] Most human movement falls within these ranges, with some very vigorous activity achieving over 12g.[134]

Application to Accelerometer-based Physical Activity Monitoring

Because accelerometer data can reflect the intensity, duration, and frequency of motion, researchers have been interested in obtaining PA estimates using accelerometer data. A multitude of uniaxial and multiaxial accelerometer devices are commercially available and designed to monitor PA. Single sensor monitors have been the most commonly used accelerometer-based devices in PA research, but others may utilize multiple sensing units placed simultaneously on the body (e.g. IDEEA), or incorporate accelerometers and other types of sensors such as inclinometer (e.g. ActivPAL) or skin sensors (e.g. BodyMedia SenseWear armband). Typical PA outcomes of interest include EE, minutes of PA spent in MET-based intensities, bouts of PA, and steps.

Bodily placement of single sensor accelerometers includes the wrist, upper arm, upper thigh, hip, and ankle.[125] Hip placement may be the best placement of a single accelerometer for detecting free-living activities, but accelerometers worn at the hip lack sensitivity to upper body movement as well as inclination changes (i.e. walking up steps).[93,125] Wrist-worn accelerometers are not uncommon, but there is little research establishing the validity of wrist placement for PA monitoring. Adding additional acceleration data from a wrist-worn accelerometer to hip-worn data may not substantially improve the accuracy of EE estimates over hip data alone, and would require additional resources or increased costs for research.[135] Despite its limitations, PA estimates from single sensor accelerometers placed at the hip have proven utility for objectively monitoring PA patterns and producing accurate estimates of PA.[136] For example, ActiGraph-determined mean EE estimates have been comparable to DLW-measured total EE over seven to 10-day periods.[101,137] Though ActiGraph-determined EE estimates exhibited wide individual variability, they produced the least variation compared to self-report, pedometer, and other accelerometer-determined EE estimates.

In laboratory settings, accelerometer data have demonstrated high correlations with gold standard PA measures such as directly observed steps and IC-determined EE during treadmill-based activities.[138,139] When applied in simulated free-living protocols incorporating activities across the spectrum of PA intensities, accelerometers produce acceptable estimates of group-level PA;[35] but they under- and overestimate PA outcomes for most individual activities.[140]

Data Processing Techniques: Linear Regression Equations and Other Methods

Accelerometers do not directly measure PA variables (e.g. steps, METs, energy expenditure). Rather, raw accelerometer data is a summary of accelerations (positive and negative) captured in a certain plane of motion.[126] Typically raw accelerations are represented by a unitless data point called counts, within a discrete period, referred to as an epoch.[46] Though counts do not have an inherent meaning, they can be analyzed post data collection and interpreted to estimate the PA variables of interest.[133,141] Researchers achieve these estimates through value calibration, a process where count output is analyzed with respect to output from a reference standard (e.g. indirect calorimetry) in order to derive PA estimates.[141] Count-based outcomes may vary depending on the device selected and post-collection data processing technique applied.[141]

Several linear regression prediction equations have been developed in laboratory settings to calculate count-based output into METs which can then be used to estimate EE or time spent in different intensities of PA compared to a criterion measure such as IC.[124] ActiGraph count output has demonstrated moderate to strong correlations with IC-determined METs ($r = 0.56 – 0.88$, $p < .01$).[135,142,143] Numerous MET prediction equations exist for the ActiGraph uniaxial accelerometer alone.[93] Most equations are based on treadmill and over-ground walking protocols with few equations derived from other recreational and domestic activities such as golfing, washing dishes, vacuuming, or playing with children.[135,142,143] Other MET prediction equations have been developed for ActiGraph's triaxial models.[144]

In an effort toward improving PA prediction outcomes, researchers have further explored advanced methods for processing accelerometer data. Hidden Markov, random forest, and artificial neural network models represent a range of advanced techniques introduced for translating accelerometer data into estimates of meaningful PA outcomes. Some evidence suggests that implementing one of these approaches may be improve PA estimates over single linear regression equations.[31,132,145] However, there is a lack of agreement on which approach is the best or standardized best practice for utilizing accelerometer data for predicting EE outcomes in free-living settings.[146,147] Researchers continue to favor these instruments for large-scale population monitoring and seek to refine methods of estimating PA outcomes from accelerometer data.[20,148] Due to the ease of application and ability to assess PA patterns at the group-level, linear regression equations are useful to predict PA outcomes from accelerometer data. In fact, it is commonly used in PA research today, and recently to evaluate nationally representative U.S. adult PA patterns.[20,26]

Researchers have been interested in incorporating additional data from other sensors to improve the monitoring and predictive capabilities of accelerometer-based devices. Findings suggest using data from pressure sensors or heartrate monitors with accelerometer data is more accurate than single sensor devices, but may face limitations.[149] In the early 1990s, Bouten and colleagues pioneered PA monitoring with a triaxial device that later called a Tracmor. The authors suggested that movement may better be assessed by combining measures from all three axis. They proposed that the relationship between accelerometer output and EE was would be improved using triaxial output represented by the integral sum of acceleration data from three orthogonal planes, known as vector magnitude (VM).[129] The authors concluded that data from all three orthogonal planes was superior to VT accelerations alone, but that using quadratic prediction equations was not superior to using traditional linear regression techniques. Energy expenditure estimates derived using the vector magnitude of accelerometer output demonstrated a strong correlation with IC-determined EE during sedentary and walking activities ($r = 0.95$, $p < 0.001$). Bouten et. al extended this approach in a one-week study of 30 subjects in free-living settings and identified an attenuated but significant correlation ($r = 0.53$, $p < 0.01$) between Tracmor VM and DLW-determined mean daily EE from waking metabolic expenditure. Energy expenditure estimates from the best single linear regression equation would accurately predict EE within 15% of actual EE for sedentary and walking activities. When walking alone was predicted best by the linear regression equation, predicted EE was within an average of 5% of measured EE. The highest correlation was between Tracmor output and mean EE from PA, or Physical Activity Level (PAL) ($r = 0.58$, $p < 0.001$). The authors concluded that, despite the lack of precise, individual PA estimates when employing linear regression analysis, the technique may be useful for ranking and grouping subjects based on overall PAL. Even when incorporating additional acceleration data from multiple axes, it appears that no one linear regression equation is able to accurately predict PA variables across all PA intensities.[93,149]

Chen and Sun investigated the predictive quality of a triaxial accelerometer using both linear and nonlinear equations over a range of activities and PA intensities over two 24-hour periods.[130] Their results suggest that nonlinear prediction equations may improve EE estimates compared to linear regression equations, particularly for estimating time spent in SED and light-intensity PA. When applying a generalized nonlinear model to accelerometer data for 125 participants, the predicted mean total EE/day was not significantly different from mean IC-measured total EE/day, whereas this outcome was significantly underestimated by the generalized linear prediction equation. Estimates of total daily EE in sedentary to vigorous intensities were calculated and compared to IC-measured EE within those intensities. Though the nonlinear model was the only model to accurately predict EE in SED and light intensity PA, the linear model was the only model that was not significantly different from IC for total EE in MVPA, and only on the day in which subjects performed structured PA (i.e. walking and stepping). For the "normal day," which only included self-selected PA, both nonlinear and linear equations underestimated estimates of EE in MVPA. Furthermore, the fact that both linear and nonlinear prediction equations were not significantly different from IC-determined EE during walking activities suggests that linear prediction equations are most accurate at estimating PA and EE when the data analyzed is similar to the activities used to formulate the value calibration. Though these findings suggest alternative prediction models, versus single linear models, may further improve estimates of PA outcomes from accelerometer data, these improvements might only be realized for estimating outcomes from certain types of PA or PA intensities. Since ActiGraph monitors are the most widely studied accelerometer-based devices in PA research, this review will provide further explain the validity of select linear prediction equations developed for ActiGraph monitors.[29]

Linear Regression Validity Studies

Freedson, Melanson, and Sirard developed one of the most widely utilized count-based cut-point criteria for the ActiGraph.[142] These researchers examined vertical axis counts of the ActiGraph 7164 compared to oxygen consumption during level treadmill walking at various speeds (4.8, 6.4, and 9.7 km/hour). ActiGraph output was sensitive to changes in oxygen consumption as treadmill speed increased, demonstrating a strong correlation with VO2 ($r = 0.88$). Researchers regressed the actual rate of EE (METs) on accelerometer output for each subject. A general regression prediction equation was derived and cross-validated on a separate sample ($n = 15$) with no significant difference in actual (IC-determined) and predicted (ActiGraph count-based) MET values. Count-based cut-points were then determined to express the intensity of PA based on ActiGraph output. The resulting count ranges were 0 – 1952, 1953 – 5724, 5725 – 9498, and > 9498 for light, moderate, vigorous, and very vigorous intensities, respectively.[150] Additionally, a separate regression prediction equation was developed to directly estimate EE from count-based data. This EE prediction equation demonstrated a very strong correlation with actual EE ($r = 0.93$) and produced non-significant mean differences of less than 0.50 kcal/min compared to IC in the cross-validated group. Certain drawbacks emerge when performing value calibration from such a limited number of activities and data points. That is, the accelerometer data were collected and summed for minute interval for only three activities that represented PA and did not include SED. Given the steady state nature of the protocol, one minute epochs may be adequate, but may not reflect subtle shifts EE possible from a variety of activities, particularly in free-living settings. Furthermore, the equation may lack sensitivity in distinguishing SED and light intensity PA.[149]

Whereas the Freedson ActiGraph cut-points were derived largely from treadmill walking and jogging, Swartz and colleagues calibrated ActiGraph vertical axis count output from a variety of household chores and recreational activities.[135] The participants wore ActiGraph monitors on the hip and wrist and a portable IC unit while performing several activities such as golfing (carrying or pulling clubs), playing with kids outdoors, sweeping, and walking while carrying a weighted item. Both, hip- and wrist-worn ActiGraph counts were significantly correlated with IC-determined MET ($r = 0.563$, $p < .001$ and $r =.181$, $p < 0.01$, respectively), but the addition of data from wrist-worn ActiGraph to hip-worn data did not substantially increase the correlation between ActiGraph counts and actual MET ($r = 0.586$, $p < 0.001$). The estimated EE (MET) were regressed on activity counts from hip-, wrist-, and hip and wrist-worn data. The regression equation with activity counts from the wrist alone explained only 3% of the variance in EE across the activities, whereas activity counts from the hip alone as well as hip and wrist explained over 30% of the variance, respectively. Given modest improvements in EE prediction by combining wrist and hip count data, it is questionable whether adding additional acceleration data from another site merits the additional investment of time and money. Accordingly, activity count output from the hip alone only to develop cut-point criteria for moderate and vigorous intensity PA (574 – 4945 counts/min. and > 4945 counts/min., respectively). Compared to the Freedson equation, the Swartz equation provided a lower threshold for moderate intensity PA but substantially higher x-intercept, essentially condensing the range for both light intensity PA and SED. In subsequent validation studies, the ActiGraph EE estimates using the Swartz equation tended to improve EE estimates for non-treadmill, moderate intensity activities, but overestimated lighter intensity activities and underestimated vigorous intensity PA.[93,151]

Several alternative prediction equations and cut-points have been proposed for ActiGraph count output. For example, ActiGraph accelerometer data were incorporated into the on-going NHANES population surveillance study in 2003. Researchers calibrated ActiGraph count output by analyzing a weighted average combination of previously established cut-point criteria for the ActiGraph 7164.[152] The resulting cut-points for are 2020 - 5998 counts/min and > 5998 counts/min for moderate and vigorous-intensity PA, respectively. Troiano et al. used these cut-points to patterns of daily MVPA in a subsample of the NHANES participants (n = 1828, ages 20 – 59 y.).[20] These cut-points have been applied in other studies of the 2003 – 2004 and 2005 – 2006 NHANES data to further describe the PA patterns of the U.S. adult population.[26,152] However, these cut-points are designed to estimate MVPA only, and cannot distinguish between light-intensity PA and SED.

Crouter and colleagues proposed that the use of two regression equations to analyze VT counts would reduce the likelihood of misclassifying PA by analyzing the pattern of count output (i.e. the coefficient of variation). The authors suggested this approach would help distinguish walking and running from other activities.[153] The coefficient of variation was calculated by analyzing six 10-second intervals within a 60-second epoch. Steady-state activities were those with an average coefficient of variation < 10 counts per 10-seconds throughout the minute interval; other activities comprised an average coefficient of variation of 10 or more counts per minute. In contrast to previous regression equations that adopted a linear prediction model, the authors determined that exponential and cubic models were most appropriate for the steady-state and lifestyle-type activity, respectively. In a cross-validation group, data were analyzed using the two-equation models as well as Freedson and Swartz linear models. The Crouter two-equation model EE estimates proved superior to Freedson and Swartz models; it accurately predicted EE for both individual activities and all activities combined. Yet the Swartz equation estimates were not significantly different to actual EE at the aggregate level. Despite the promise of the two-equation method, subsequent validation research in a free-living population demonstrated that the approach did not provide accurate EE estimates; it instead misclassified SED and light-intensity PA compared to IC-criterion.[154]

Triaxial accelerometers offer the potential for improved PA estimates through additional count data from the transverse and antero-posterior axes. Yet, when using similar data treatment approaches, the PA estimates may not be substantially different from those obtained from uniaxial accelerometer data. For example, previous research has demonstrated the strong comparability of VT axis output between several generations of uniaxial and dual-axial ActiGraph monitors.[29] Most recently researchers have demonstrated that ActiGraph's triaxial monitor, the GT3X, supports the comparability between generations of Actigraph monitors, but have proposed new MET prediction equations utilizing the triaxial VM count output.[31,144] Subsequent research comparing EE estimates from the Sasaki VM equation to estimates from the Freedson equation showed both equations produced estimates with large mean absolute percent errors (MAPE). Average MAPE for light, moderate, and vigorous intensity activities errors were lower for Sasaki equation (86.0%, 28.0%, and 21.0%, respectively) compared to Freedson equation estimates (95.4%, 47.7%, and 24.0%, respectively).[155] It should be noted that this protocol included many static (e.g. standing, typing at computer) and sedentary activities (e.g. reading a book while seated). Previous research suggests that special analysis should be applied to ActiGraph output for determining time spent in sedentary activities.[156,157] and these techniques were not applied in this study. Notwithstanding, these results suggest using additional count data from multiple axes may not substantially improve PA estimates over VT axis estimates when utilizing linear regression prediction techniques.

Further highlighting the variability between prediction equations and count-based cut-point ranges for PA intensity categories, Matthews and colleagues examined both laboratory- and field-based acceleration data to determine the likely upper and lower limits of activity counts for moderate intensity PA. The best prediction model yield limits 760 – 5724 counts/min for moderate intensity PA.[158] Other ActiGraph predictions equations have proposed moderate intensity limits as low as 191 counts/minute and as high as 7526 counts/minute. Researchers have suggested this reflects the variability in activities used to derive these prediction equations. As such, there is not currently a single regression prediction equation able to accurately predict PA across all intensity categories. However, as these select studies suggest, regression-based prediction equations are suitable for estimating MVPA from ActiGraph accelerometer at the group level despite presenting large variability at the individual level.

Unlike cut-points established for MVPA, there is less variation in thresholds for classifying SED. Self-report evidence suggests a majority of Americans may spend substantial amounts of waking hours in sedentary activities, necessitating accurate and objective measures of SED. Matthews et al. attempted to objectively estimate time spent in SED from 2003 – 2004 NHANES accelerometer data by applying a cut-point of < 100 count/min. to define SED.[148] However, there is a lack evidence supporting the validity of this cut-point compared to gold-standard measures or other activity count cut-point criteria. In fact, some evidence suggests an activity count threshold of 150 counts/minute improves SED estimates by being less vulnerable to spurious bodily movements during sedentary activities.[156] Nonetheless, research has shown

ActiGraph yields equivalent estimates of SED using the <100 counts/minute compared to ActivPAL-determined sedentary time, a standard criterion for SED in free-living settings, thus supporting the validity of this count-based threshold.[156,159]

Consumer-based Physical Activity Monitors

Industry Growth

Recent consumer electronic sales trends reflect an increasing public interest in PA monitoring.[40] Consumer-marketed activity trackers, or 'wearables', are advertised as electronic accessories, designed to be fashionable and functional while still providing feedback on various PA outcomes (e.g. steps, EE, stairs climbed), as well as, in some cases, estimates of sleep quality.[118] Some devices allow the user to track and compare themselves with others on the various outcomes via social media or other online platforms. Manufacturers largely rely on accelerometer technology in these devices, though certain models may incorporate other technology such as GPS or LED optical pulse meters. Recent market sales analysis show rapid industry growth for "wearable" technology products.[40,160] Fitness-related tracking devices are the most commonly reported in consumer data.[161] The Fitbit Company has lead this sector of the industry, boasting the largest consumer sales in 2014.[160,162] These products may present an advantage to researchers regarding subject compliance since the burden of wearing a device is lessened by the social and functional appeal of the devices.[42,163] However, to date, very few published studies exist which substantiate the claims that these devices can measure various PA outcomes.

Validity of Consumer-based Physical Activity Monitors

Lee and Welk compared EE estimates from several consumer-based PA monitors (e.g. Fitbit, Jawbone, Basis B1, BodyMedia FIT) to IC criterion, finding moderate to strong correlations (r = 0.32 - 0.81) and wide ranging MAPE(13 – 48%).[35] Fitbit devices outperformed all but one PA monitor (BodyMedia FIT) in this study. Similar comparison studies have also observed Fitbit devices may yield more accurate PA estimates (e.g. EE, steps) compared to several other consumer-marketed devices (e.g. Nike Fuel band, Polar loop, Misfit Shine).[35,164] Fitbit monitors have also been utilized as a PA intervention tool and PA monitor in unique populations (e.g. clinical COPD patients, transfemoral amputees).[43,120,165] Given notable consumer interest in Fitbit monitors and their current utilization in PA monitoring research, the remaining section will focus on reviewing existing literature on the validity of Fitbit monitors.

Validity of Fitbit Hip-Worn Monitors

The majority of evidence supporting the validity and reliability of Fitbit monitors comes from studies of its hip-worn models. Some studies explored the validity of the Fitbit step count feature against other objective criteria, namely direct observation, accelerometer, and pedometer. Hip-worn Fitbit step estimates have demonstrated strong agreement with directly observed step counts across a wide range of treadmill and self-paced walking speeds in both adult and elderly populations.[166-168] The lowest agreement reported between Fitbit and directly observed step counts (ICC (2,1) = 0.70) was found in a group of 30 elderly stroke survivors with a mean gait speed of 0.84 (± 0.34) m/s. Other studies have reported mean error percentages of 8% or less in apparently healthy adult samples.[166,167] Similarly, Fitbit step-counts have compared well to other PA monitors step estimates. Gusmer et. al reported strong mean correlations (0.97, $p < 0.01$ and $r = 0.99$, $p < 0.01$) between Fitbit and ActiGraph's GT1M step estimates during slow and brisk treadmill walking (mean = 1.11 and 1.36 m/s, respectively). Fitbit monitors have produced equivalent step estimates when compared to Yamax and Omron pedometers, but may overestimate mean steps per day compared to ActiGraph monitors.[169,170]

Evidence supports the reliability of Fitbit monitors for estimating EE and steps. In a study comparing the reliability of EE and step estimates from two hip-worn Fitbit models (Fitbit Ultra and Fitbit Tracker) compared to IC and ActiCal, both Fitbit models produced strong inter-instrument correlations for step estimates during walking and stepping activities (r = 0.88 – 0.99; r = 0.88 – 0.96, respectively).[171] Both Fitbit Tracker and Fitbit Ultra demonstrated stronger correlations with IC-criterion EE measures during treadmill walking and jogging (r = 0.56 – 0.70; r = 0.81 – 0.87, respectively) and attenuated correlations during stepping activity (r = 0.18; r = 0.58, respectively). Thus, while Fitbit monitor output appears to have high reliability within a certain model; different Fitbit models may not necessarily produce similar PA estimates during all activities.

Most studies of hip-worn Fitbit monitors have examined the validity of Fitbit's EE estimates compared to other PA monitors and IC-criterion. Group-level Fitbit EE estimates have demonstrated strong correlations with IC-determined EE (r = 0.78 - .81)[35,164] but have much wider range for individual activity correlations (r = 0.18 – 0.87).[171] Like other accelerometer-based monitors, Fitbit tend to over- and under-estimate the metabolic cost of individual activities, with mixed results for group-level estimates.[172]

In fact, some research suggests Fitbit EE may not be accurate without additional user input.[173] Danneker and colleagues found that Fitbit significantly underestimated total EE compared to IC, with a mean error rate of 28.7 % over the course of four hours.[173] After classifying the activities on Fitbit's proprietary online interface, the authors reported nonsignificant differences between Fitbit and IC estimates, with a mean error rate of 12.9%, suggesting hip-worn Fitbit monitors may not be suitable for monitoring PA in free-living settings without additional user input. In a similar study, EE estimates from Fitbit were not significantly different from IC during several treadmill-based and simulated free-living activities, with a small mean bias of less than 5 kcal/6 minutes reported.[172] The inconsistent results between the studies may stem from differences in the activity protocol chosen to simulate free-living activities. The study by Danneker et al. consisted of activities that are especially challenging to hip-worn monitors, including several sedentary activities with little to no lower-body movement, such as seated computer work, standing, or lying down. Of eight the physical activities, only two involved level walking. Though standing, sweeping, inclined walking and cycling are not necessarily uncommon to typical free-living setting, they may not well-represent the majority of daily non-sedentary activity, which constituted only a small portion of the short protocol. Thus, the accuracy of the Fitbit may depend not only on the type of PA, but on the proportion time spent in PA versus sedentary activities.

Nonetheless, research has shown Fitbit monitors provide reasonable group-level estimates of PA and may outperform competitor devices. In one study the Fitbit Zip and Fitbit One models demonstrated the lowest MAPE (10.1 and 10.4%, respectively) and were the only consumer-marketed devices, along with the Nike Fuel band, to fall within 10% equivalence of IC criterion.[35] When compared to all other consumer-marketed devices (except the body Media FIT), both Fitbit Zip and Fitbit One EE estimates demonstrated the highest correlations with IC-determined EE ($r = .807$, $r = 0.808$, $p < 0.01$, respectively). The authors note the Body Media FIT "…is a consumer version of a research-based, armband monitor known as the SenseWear armband."[35] These results support the author's findings in which the hip-worn Fitbit device had the second highest mean correlation (r = 0.70) and second lowest MAPE (14%) of eight consumer-marketed devices.[35] The study activities were performed in controlled (laboratory) settings but simulated lifestyle activities, including several seated activities with some upper-body movement but little lower-body movement; this suggests that at the group level, Fitbit monitors may be suitable for free-living PA surveillance.

Few studies have investigated the validity of Fitbit monitors for estimating PA outcomes in truly free-living settings. In fact, to our knowledge only one study investigated the accuracy of Fitbit monitors for estimating time spent in specific PA intensities with data from obtained in free-living settings. Rosenberger and colleagues examined the validity of PA classification estimates of a hip-worn Fitbit in a free-living setting.[170] The hip-worn ActiGraph GT3X+ served as the criterion for light and MVPA. At the individual level, Fitbit underestimated EE for light-intensity by an average of 64 minutes and overestimated EE MVPA by an average of 76 minutes per participant over a 24-hour period. When compared to ActivPAL criterion, the Fitbit also overestimated time spent in SED activity. However, Fitbit SED estimates were comparable to GT3X+, overestimating SED time by an average of 34 and 48 minutes per participant, respectively.[170] Group-level error were represented by MAPE and equivalence testing. Fitbit demonstrated the highest error (MAPE >60%) for MVPA (min / 24-hours), followed by light-intensity PA and SED (MAPE > 20% and >10%, respectively), compared to ActiGraph and ActivPAL criterion.[170] Previous research has shown a strong correlation between EE estimates from hip-worn Fitbit and ActiGraph monitors.[35] However, the presence of a strong relationship between EE estimates does not equate to strong agreement on PA classification. These data echo previous validity research showing that accelerometer-based PA estimates may be valid at the group-level, but no single, accelerometer-based monitor is perfectly-suited to quantify all types and intensities of PA.[93]

Validity of Fitbit Wrist-Worn Monitors

Even fewer studies have examined the validity of the Fitbit wrist-worn monitors. Previous researchers have explored whether accelerometer data from a wrist-worn monitor alone, or added to hip-worn data, improves the accuracy of accelerometer-based PA estimates.[135,175] Previous findings have been inconclusive, with some evidence indicating wrist-worn monitors may slightly improve PA estimates compared to hip-worn accelerometer output alone.[135] In one study comparing EE estimates from several consumer-marketed PA monitors, including the wrist-worn Fitbit Flex, to IC-determined EE from sedentary, aerobic exercise, and resistance training activities, the Fitbit Flex outperformed competitor monitors and produced the smallest mean bias in EE estimates compared to IC.[164] As with previous studies of Fitbit monitors, Fitbit EE output were highly correlated with IC-determined EE ($r = 0.78$, $p < 0.01$), second only to the Body Media Core armband. Though Fitbit failed to achieve 10% equivalency with IC, overall group-level estimates were closer to achieving equivalency than most other competitor monitors. Individual level error was highest for aerobic exercise and lowest for SED, with all mean absolute error percentages reported above 20% for activity specific domains.

Though the wrist-worn Fitbit appears to produce comparable, if not superior group-level PA estimates compared to other consumer-based PA monitors, activity specific PA estimates may differ substantially from other validated monitors used in PA research. In the aforementioned study, ActiGraph EE estimates achieved 10% equivalence testing with IC-determined EE. Mean individual EE estimate errors were greater for ActiGraph than Fitbit for sedentary and resistance training activities, but they were substantially lower than Fitbit for aerobic exercise.[164] Fitbit estimates were likely closer to actual EE during sedentary activities and resistance exercise because the hip-worn ActiGraph would not have been sensitive to the nuanced arm movements during SED, such as desk work or fidgeting during “active standing.” However, a majority of EE in this 80-minute protocol would have occurred during exercise bouts (aerobic and resistance exercise). Resistance training included mostly upper body- and some lower body-dominant activities that were self-selected by participants. It is reasonable to infer that subjects may have self-selected more upper-body dominant exercises during resistance training, thus, leading to improved EE estimates from Fitbit compared to ActiGraph. Since accelerometers only predict PA based on movement (accelerations), total EE attributed to this domain may have been less than aerobic exercise, and these estimates would not be able to account for additional EE used to move external loads or due to excess post-exercise oxygen. Furthermore, if subjects demonstrated various arm patterns during treadmill walking or jogging it might have attenuated Fitbit PA estimates during aerobic exercise, explaining the greater MAPE compared to ActiGraph, a device that has been validated for estimating EE during ambulatory activities. It is uncertain if the results of this study reflect the predictive validity of the wrist-worn Fitbit monitors in free-living settings.

Consumer-marketed PA monitors hold promise for multiple research applications, including population-level PA surveillance or integration into PA interventions. Yet, due to a lack of research in free-living settings, their utility is yet to be determined, particularly for wrist-worn monitors such as the Fitbit Flex.

Summary

Several health benefits are possible by regularly engaging in PA such as lowered risk of pre-mature death, improved quality of life, and improved physiological functioning.[4,58,176]

Despite public health documents that both disseminate these benefits and outline the quantity of PA needed to achieve them, the majority of adults in the United States to do not meet these weekly PA recommendations.[20] In addition to copious amounts of SED characterizing the leisure-time of a majority of Americans, self-report PA research suggests that many adults also overestimate the amount and intensity of their own PA.[177,178]

Objective PA measures may be beneficial to accurately monitoring and effectively promoting PA. Current gold standard objective PA tools, namely IC and DLW, are costly and are not easily deployable to large population in free-living settings. On the other hand, low-cost and inconspicuous objective monitors such as HR monitors or pedometers, provide limited PA information or are vulnerable to error from other physiological influences. However, improvements in accelerometer-based PA monitors show promise for objectively monitoring and estimating PA. Due to their small size, they can be worn inconspicuously by users in everyday settings. Increased memory capacity and battery efficiency without increases in monitor size allows them to be deployed for several days or weeks at a time. Weighing less than 50g, many accelerometer-based monitors used in research can monitor PA for a week or more without recharging.[46,123] Furthermore, modern accelerometer-based PA monitors do not require regular maintenance whereas previous accelerometer technology would lose calibration over time or with wide temperature shifts.[123]

There are clear limitations and weaknesses associated with using linear prediction equations and count-based cut-points to estimate PA from accelerometer data.[141] Because accelerometers record accelerations, they are unable to readily record EE and PA during weight-bearing activities that may or may not incorporate static muscle contraction. Similarly, accelerometers may not readily detect PA of body segments that are functioning relatively independent of the device attachment point. Cycling is a good example of an activity that is particularly challenging to hip-worn and wrist-worn monitors. Water-based activities like swimming are also challenging since most accelerometer devices are not submergible. Misclassification of individual activities is likely when utilizing regression prediction equations to estimate PA from accelerometer data. Yet, researchers with basic statistical analysis skills can easily apply the technique which have proven validity. Furthermore, current device software may even automate processing the data. Researchers may favor using multiple PA monitoring devices or using advanced statistical analysis, but these options may not always be feasible if additional time and resources are not available.

Detecting everyday movement patterns is paramount for assessing daily PA across a range of activities and intensities. Despite the limitations known for using accelerometers to estimate PA, they are objective instruments and able to monitor a wide range of activities with minimal cost and user burden; they are feasible for large scale free-living PA monitoring.

METHODOLOGY

The purpose of this study was to determine the concurrent validity of the Fitbit Flex accelerometer in a free-living condition. We hypothesized that estimates of time spent in MVPA and SED would not significantly differ between Fitbit and the ActiGraph GT3X+ (GT3X+).

Participants

A convenient sample of 67 adults from the North Dakota State University volunteered to participate in the study. Participants were recruited via University email listservs, posted fliers, and word-of-mouth. Participants were eligible to participate if they were a University student or employee, were at least 18 years of age or older and could attend one of the orientation meetings. North Dakota State University's Institutional Review Board approved the study and all participants voluntarily consented to participate in the study.

Instruments

Fitbit Flex

The Fitbit Flex (Fitbit Inc., San Francisco, CA), is an accelerometer-based physical activity monitor that is 3.2 cm long and weighs less than 15 grams (including wrist-band).[179] It features a MEMs tri-axial accelerometer which tracks movement in three planes (mediolateral, vertical, anteroposterior). The monitor continually acquires data and with onboard storage capacity for approximately seven days of data without syncing. Data is transferred via Bluetooth technology to the Fitbit application program interface (API) either through Fitbit's mobile app or a Bluetooth dongle connected to a computer. Users are able to access personal physical activity estimates via the Fitbit dashboard (Fitbit.com). The dashboard is a cloud-based interface, which provides real-time estimates and daily summaries of PA including steps, EE, ambulatory distance, and "active minutes."

The Fitbit dashboard provides limited resolution of PA estimates (approximately 15 minutes) and does not allow easy access to export PA data. Fitabase (Small Steps Labs, LLC, San Diego, CA) is a third-party online-based service marketed toward research applications of Fitbit devices. Fitabase has an access to Fitbit's API which manages and houses the data for Fitbit.com's online users. Fitabase subscribers are able to view and export data of specific PA variables at daily, hourly, and minute-by-minute resolutions. We utilized Fitabase services to acquire minute-by-minute SED and PA estimates for each participant.

ActiGraph GT3X+

ActiGraph (Pensacola, FL) currently offers multiple models of tri-axial accelerometer-based devices. The ActiGraph GT3X+ is a lightweight (19 g), tri-axial MEMs accelerometer-based monitor with a dynamic range of -/+ 6 G.[180] Users may choose sampling frequencies from 30 Hz to 100 Hz. We chose a sampling rate of 30hz (over one-minute epochs) as this range should adequately capture most accelerations due to human movement.[133] Accelerations are recorded and stored in raw format. Using ActiGraph's proprietary software, ActiLife, users can select data filtering options to obtain count-based device output. Users can then obtain SED and PA estimates through ActiLife by selecting from previously established count-based criteria or specifying data-treatment parameters.

Data from ActiGraph accelerometer were downloaded and scored using ActiLife version 6.11.4 (ActiGraph Corp., Pensacola, FL). Non-wear time and sleep time were defined using Choi (2011)[157] criteria and participants' sleep logs, and excluded from the analysis. Minute-by-minute activity counts from accelerometer were then scored using three different cut-points Freedson (1998), Troiano (2008), and Freedson VM3 (2011), and reduced as time spent in sedentary, light, and moderate-to-vigorous physical activity. Fitbit non-wear time were defined individually by removing sleep time and any non-wear time recorded on participant sleep logs. Fitbit data were then individually matched to each participant's GT3X+ wear time. For our comparisons, only data from waking hours where both devices were worn, as determined by ActiLife wear-time analysis and participant wear time logs, were considered.

Procedures

All participants reported to group orientation session where they were briefed on the purpose of the study and how to wear and use the PA monitors. Participants voluntarily consented to be in the study and completed a demographic questionnaire before beginning the protocol. Participants simultaneously wore the Fitbit and ActiGraph monitors for seven consecutive days. Participants were instructed to wear all devices during all waking and sleep hours except during bathing and recreational water activities (e.g. swimming). Participants wore a Fitbit monitor on the dorsal aspect of the non-dominant wrist, secured by a Fitbit wristband, like a watch. The ActiGraph GT3X+ monitor was worn on the dominant hip in-line with the midline of the thigh and the approximate peak of the iliac crest. The GT3X+ was secured onto subject's pant or belt by a belt clip, with a semi-elastic leash clipped to a different part of the pant for added security. Participants kept a log of any non-wear time during waking hours, extraordinary amounts of PA, and daily sleep times.

An email was sent to participants three days prior to the conclusion of the data period reminding them to charge the Fitbit once, preferably during the overnight hours, and continue wearing all devices during waking hours until the conclusion of the data period. At the conclusion of the data collection period, all devices and participant logs were personally retrieved by the investigators and data downloaded immediately.

Statistical analysis

ActiGraph-determined PA (Freedson, Troiano, and Freedson VM3 cut-point criteria) served as the criterion reference for all comparisons. Due to unequal sample size, we used the Welch's T-test to assess differences in daily PA and SED between the participant groups (Students vs. faculty and staff). Pearson correlations were used to determine the relationship between Fitbit and GT3X+. Repeated measures one-way analysis of variance (ANOVA) served to examine differences in estimates of SED and MVPA, comparing estimates from Fitbit with those from GT3X+ using three different cut-points (only two cut-points for SED estimates). Repeated measures ANOVA was preferred over multiple t-tests to reduce the likelihood of committing a Type-I error for SED and MVPA comparisons. Bland-Altman plots were used to illustrate any potential bias between GT3X+ and Fitbit daily estimates of SED and MVPA. All data analyses were conducted using IBM SPSS 24.0 for Windows (SPSS, Armonk, NY). Alpha level of 0.05 defined significance for all statistical analysis.

MANUSCRIPT

Surveillance of physical activity (PA) is vital for better understanding of the relationship between PA and specific health outcomes. Although limitations of self-report are evident, historically surveillance data have relied on subjective measures of PA such as self-reported questionnaires.[71,80,83] Accelerometers are particularly appealing for PA monitoring in free-living conditions; several accelerometer-based devices have been used in PA research applications.[28] ActiGraph accelerometer is the most widely used to measure PA in research and surveillance systems.[27,28] For example, due to its high validity and feasibility, the ActiGraph GT3X+ was the method of choice for measuring PA in NHANES 2011 – 2014, one of the most representative surveillance system in the U.S.[33]

Researchers have used the strong correlations between accelerometer and IC output to devise count-based prediction models for PA outcomes such as minutes of PA in MET categories or energy expenditure (EE).[135,142,143] Using regression equations with accelerometer counts (i.e. counts/60 second) as a predictor, several activity count cut-points have been developed to estimate the amount of time spent in different intensities of PA (e.g. minutes of vigorous PA). Among those cut-points, two developed by Freedson et al., and one developed by Troiano et al. are most widely utilized cut-points for estimating time spent in varying intensities of PA including sedentary, light, moderate, and vigorous PA.[99,137,142,149,155] However, PA estimates may significantly vary depending on cut-point criteria applied to accelerometer output, which primarily caused by inconsistency in monitor placement types and duration of activities used to calibrate the equation.[93,181,182] As such, there is no single cut-point criteria able to accurately classify accelerometer-based PA estimates across all intensity categories and activities.[93] Thus, studies investigating the validity of PA monitors with an accelerometer-based criterion may be limited by utilizing only one cut-point prediction model. Nonetheless, the Freedson cut-points have produced acceptable estimates of MVPA even when compared to more recent cutpoints.[93,149,181]

Fitbit is a leading manufacturer of accelerometer-based PA monitors sold in the consumer electronics.[160,162] Given the popularity and acceptability of these consumer-based PA monitors, there may be an opportunity to use them as a research tool. Few studies have examined the validity of Fitbit, but most used only hip-worn Fitbit models in controlled settings.[35,171,173] Fitbit step estimates have demonstrated strong agreement with directly observed step counts across a wide range of walking speeds in both adult and elderly populations.[166-168] Group-level Fitbit EE estimates have demonstrated strong correlations with IC-determined EE, but correlations were lower at individual-level. .[35,164,171] Previous research also shows Fitbit tend to over- and under-estimate the metabolic cost of individual activities, with mixed results for group-level estimates.[172] Nonetheless, Fitbit monitors provide reasonable group-level estimates of PA.[35]

Fewer studies have assessed the accuracy of Fitbit's wrist-worn PA monitor, the Fitbit Flex. One laboratory-based study showed the Fitbit Flex produced EE estimates highly correlating with IC-determined EE, but overestimated EE for specific activity domains (e.g. aerobic activity).[164] Other research has shown the hip-worn Fitbit PA classification estimates greatly differ with ActiGraph criterion for classifying MVPA.[170] Therefore, the purpose of this study was to examine the concurrent validity of the wrist-worn Fitbit Flex compared to the hip-worn ActiGraph GT3X+ utilizing three different cut-point criteria in a free-living condition.

Methods

Participants

A convenient sample of 67 participants (age: 47.1 ± 14.1 years, Female: 73.1%) was recruited from the North Dakota State University students, faculty and staff by email, posted fliers, and word-of-mouth. Participants who were under the age of 18, pregnant, physical disabled, or unable to engage in regular PA as recommended by a physician, were not eligible to be in the study. The North Dakota State University Institutional Review Board approved the study and all participants voluntarily provided consent to participate in the investigation.

Instruments

The Fitbit Flex (Fitbit, Inc., San Francisco, CA), is a physical activity monitor that is 3.2 cm long and weighs less than 15 grams (including wrist-band).[179] It features a tri-axial accelerometer and continually acquires data and with onboard storage capacity for approximately seven days of data without syncing. Data is transferred via Bluetooth technology to the Fitbit application program interface (API) either through Fitbit's mobile app or a Bluetooth dongle connected to a computer. Since the Fitbit dashboard provides limited resolution of PA estimates and does not allow users to export PA data without a premium subscription, we chose to access minute-by-minute data for the Fitbit Flex through an online-based third-party service (Small Steps Labs, LLC, San Diego, CA). Fitabase subscribers are able to view and export data of specific PA variables at daily, hourly, and minute-by-minute resolutions from the Fitbit API.

ActiGraph (ActiGraph Corp., Pensacola, FL) currently offers multiple models of tri-axial accelerometer-based devices. The ActiGraph GT3X+ is a lightweight (19 g), tri-axial accelerometer-based device with a dynamic range of -/+ 6 G.[180] Users may choose sampling frequencies from 30 Hz to 100 Hz. We chose a sampling rate of 30hz (with one-minute epochs) as this range should adequately capture most accelerations due to human movement.[133] Data from GT3X+ accelerometer were downloaded and scored using ActiLife version 6.11.4 (ActiGraph Corp., Pensacola, FL).

Participants kept a log of any non-wear time during waking hours and daily sleep times. Participants were also instructed to note any days that included extraordinary amounts of PA that may appear unusually high for their typical routine (i.e. running a half-marathon).

Procedures

Participants completed an orientation session and began the free-living protocol after voluntarily consenting to be in the study and completing a demographic questionnaire. Participants simultaneously wore the Fitbit and GT3X+ monitors for seven consecutive days during all waking and sleep hours except during bathing and recreational water activities (e.g. swimming). Participants wore a Fitbit monitor on the dorsal aspect of the non-dominant wrist, similar to a watch. The GT3X+ monitor was worn on the dominant hip in-line with the midline of the thigh and the approximate peak of the iliac crest. At the conclusion of the data collection period, all devices and participant logs were personally retrieved by the investigators and ActiGraph data downloaded immediately.

Data reduction

Non-wear time and sleep time were defined using Choi (2011)[157] criteria and participants' sleep logs respectively, and were excluded from the analysis. No participants noted any extraordinary PA during the protocol. Thus, all minute-by-minute activity counts during validated waking hours from GT3X+ accelerometer were then scored into daily time spent in sedentary, light, and moderate-to-vigorous physical activity (min/day) using three different cut-points Freedson (1998), Troiano (2008), and Freedson VM-3 (2011, MVPA only). Similarly, Fitbit wear time was validated by removing sleep and non-wear time from the participant sleep log. Fitbit minute-by-minute data were then time synced to corresponding GT3X+ validated wear time. Thus, only data from waking hours where both devices were worn, as determined by ActiLife wear-time analysis and participant wear time logs, were considered.

Statistical Analysis

Pearson correlations were used to determine the relationship between estimates from Fitbit and those from GT3X+. Due to unequal sample size, we used the Welch's T-test to assess differences in daily PA and SED between the groups (Students vs. faculty and staff). To avoid committing a Type-I error with SED and MVPA comparisons, repeated measures one-way analysis of variance (ANOVA) served to examine differences in SED and MVPA estimates, comparing Fitbit estimates and those from GTX+ using three different cut-points (only two cut-points used for SED comparisons). Significant overall ANOVA was followed by pair-wise comparisons using Bonferroni adjustment. Bland-Altman (BA) plots were used to illustrate any potential systematic bias between GT3X+ and Fitbit SED and MVPA estimates. All data analyses were conducted using IBM SPSS 24.0 for Windows (SPSS, Armonk, NY). Alpha level of 0.05 was set to define significance for all statistical analyses.

Results

Subject characteristics are summarized in Table 1. The sample was relatively homogenous, mostly female and non-Hispanic white. Participant ages ranged 20 – 70 years (y). Given the distinctive age range and occupational status represented in the sample, the mean daily minutes of PA and SED were presented separately for students and the faculty/staff (Table 2). Participants recorded an average of 5.9 valid wear days (14.9 hours/day) over the 7-day period indicating high compliance with the protocol. Participants spent the majority of waking hours in

SED and least amount of waking hours in MVPA. Between groups, students spent less time in SED and more time in MVPA compared to faculty and staff. The Welch's T-test results showed students achieved significantly more MVPA and significantly less SED than faculty and staff. However, the comparison between groups was not integral to the intended analysis. Thus, we combined data from the entire sample (n = 67) for the remainder of the analysis.

Table 1. Participant characteristics by occupational status.

	Students (N = 11)	**Faculty/Staff (N = 56)**	**Total (N = 67)**
Age (mean years ± SD*)	23.0 ± 1.4	45.4 ± 12.5	47.1 ± 14.1
Sex (%)			
Male	63.6	80.4	26.9
Female	36.4	19.6	73.1
Ethnicity (%)			
Non-Latino	100.0	100.0	100.0
Race (%)			
Asian	9.1	1.8	3.0
White	90.9	98.2	97.0
BMI (mean kg/cm^2 ± SD)	25.4 ± 2.1	26.0 ± 4.8	25.9 ± 4.4

*SD: standard of deviation

Table 2. Mean valid wear days and mean daily minutes of MVPA and SED.

	Students (n = 11)	**Faculty/Staff (n = 56)**	**Total (n = 67)**
Valid Wear Days (±SD*)	5.9 ± 1.4	5.8 ± 1.2	5.9 ± 1.2
SED[a] (min/day ±SD)			
Fitbit	586.1 ± 72.3	645.0 ± 92.5	635.3 ± 91.7
GT3X+ (Freedson)	560.4 ± 90.2	607.2 ± 90.7	599.5 ± 91.6
GT3X+ (Troiano)	560.4 ± 90.2	607.2 ± 90.7	599.5 ± 91.6
GT3X+ (VM3[†])	--	--	--
MVPA[b] (min/day ±SD)			
Fitbit	141.1 ± 23.4	100.5 ± 29.3	107.1 ± 32.1
GT3X+ (Freedson)	51.0 ± 16.3	25.5 ± 16.0	29.7 ± 18.6
GT3X+ (Troiano)	48.8 ± 15.8	24.3 ± 15.8	28.3 ± 18.2
GT3X+ (VM3)	74.3 ± 20.8	42.2 ± 21.7	47.4 ± 25.5

*SD: standard of deviation
[†]VM3: sedentary behavior estimates were not available from the Freedson VM3 cut-points.
[a]SED: sedentary behavior
[b]MVPA: moderate-to-vigorous physical activity

We found strong correlations for SED estimates ($r = .89$, $P < 0.01$) between GT3X+ and Fitbit (Table 3). For MVPA, the correlations between Fitbit and GT3X+ were moderately strong across the ActiGraph cut-points applied ($r = .66 - .77$, $P < 0.01$).

Table 3. Pearson Correlations between Fitbit and GT3X+ SED and MVPA estimates.

		Fitbit	**GT3X+ (Freedson)**	**GT3X+ (Troiano)**	**GT3X+ (VM3)**
SED[a]	Fitbit	1	.89*	.89*	--
	GT3X+ (Freedson)		1	1.00*	--
	GT3X+ (Troiano)			1	--
MVPA[b]	Fitbit	1	.67*	.66*	.77*
	GT3X+ (Freedson)		1	.99*	.88*
	GT3X+ (Troiano)			1	.87*
	GT3X+ (VM3)				1

*p < .01
[a]SED: sedentary behavior
[b]MVPA: moderate-to-vigorous physical activity

Results of one-way ANOVA revealed significant differences between Fitbit and GT3X+ criteria for MVPA [$F(3, 264) = 159.28$, $p < .01$] and SED [$F(2, 198) = 3.41$, $p < .05$]. Significant differences remained only for MVPA pair-wise comparisons (Table 4). Fitbit significantly overestimated MVPA compared to all GT3X+ criteria by notably wide margins. The mean differences in MVPA estimates between Fitbit and GT3X+ were 60 (VM3), 77 (Freedson), and 79 min/day (Troiano), respectively. There were no significant differences in daily SED estimates between Fitbit and GT3X+ Freedson and Troiano cut-point criteria.

Table 4. Mean differences between GT3X+ and Fitbit SED and MVPA estimates.

Intensity	Comparison	Mean Difference (min/day± SD)	95% CI
SED[a]	Freedson - Fitbit	-35.83 (42.74)	[-74.05; 2.40]
	Troiano - Fitbit	-35.83 (42.74)	[-74.05; 2.40]
MVPA[b]	Freedson - Fitbit	-77.41 (23.94)*	[-88.44; -66.39]
	Troiano - Fitbit	-78.83 (24.34)*	[-89.86; -67.80]
	VM3 - Fitbit	-59.70 (20.64)*	[-70.73; 48.67]

*p < .0167
[a]SED: sedentary behavior
[b]MVPA: moderate-to-vigorous physical activity

BA plots revealed that there is not an apparent bias in the agreement for SED estimates between the two devices (Figure 1). However, for MVPA, BA plots suggest Fitbit increasingly overestimates MVPA compared to GT3X+ as mean volume of MVPA increases (Figure 2).

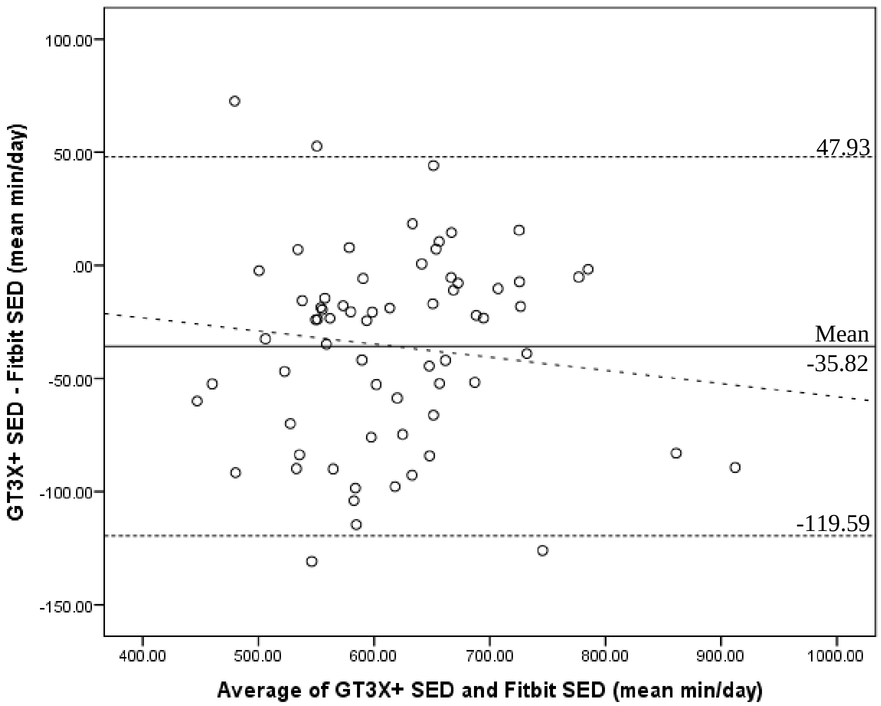


Figure 1. Bland-Altman plots illustrating level of agreement between GT3X+ and Fitbit SED and PA classification estimates. Dashed lines show 95% limits of agreement (± 1.96 SD).

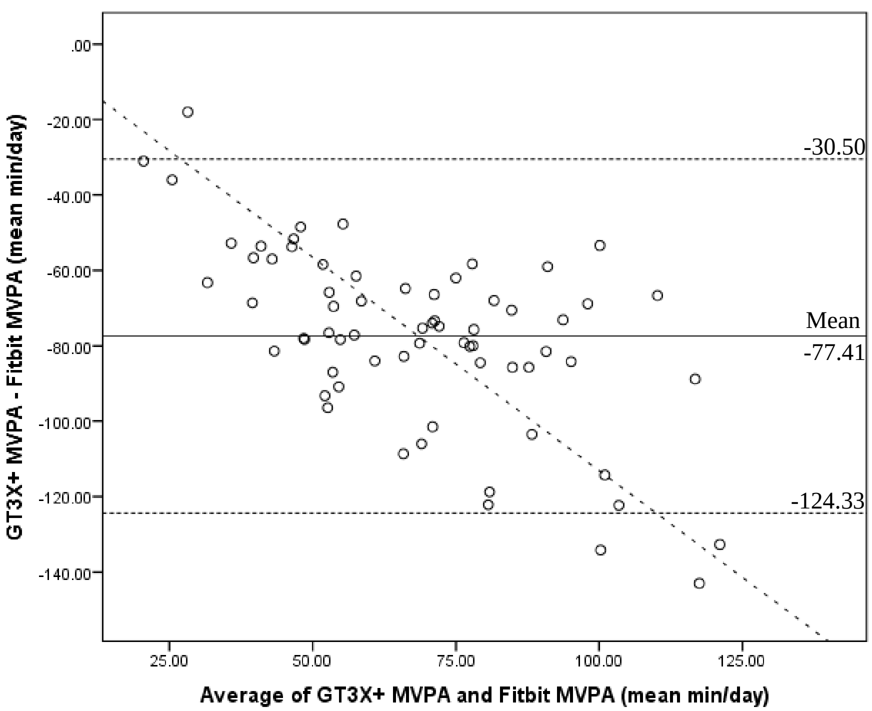
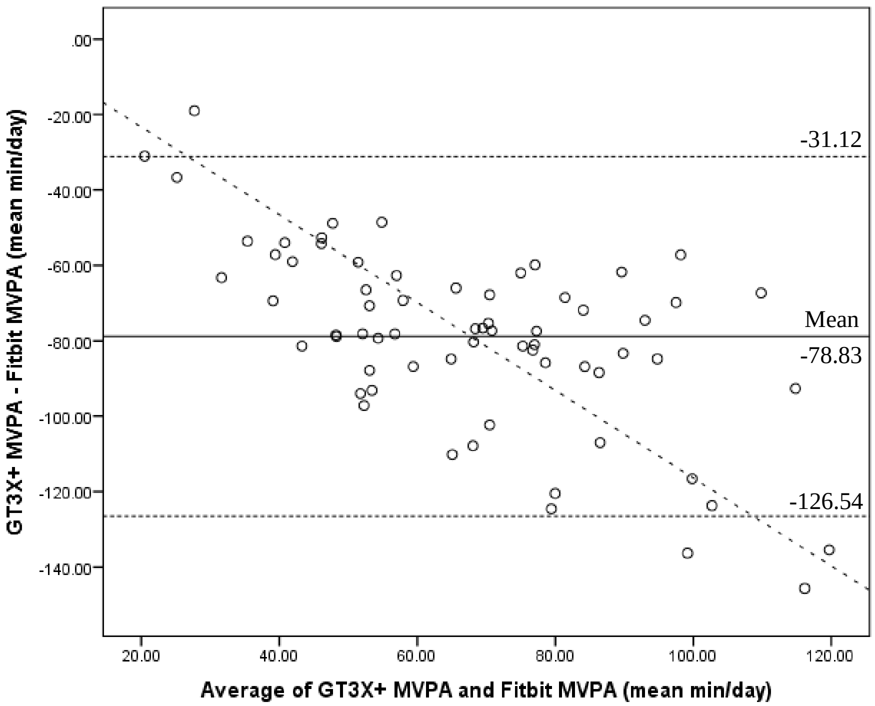
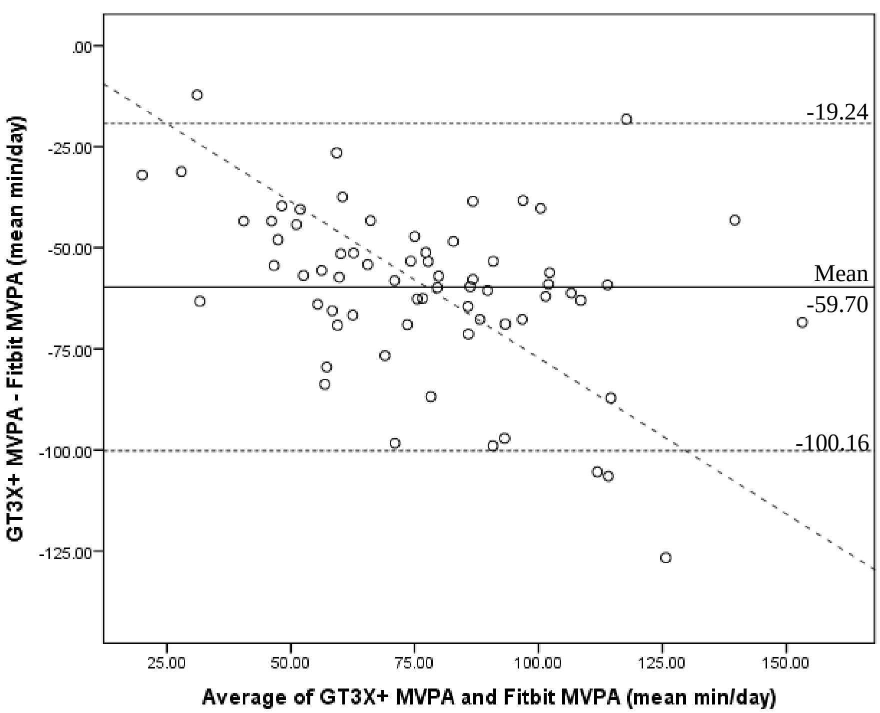


Figure 2. Bland-Altman plots illustrating level of agreement between GT3X+ MVPA and Fitbit MVPA classification estimates. A) Freedson cut-points. Dashed lines show 95% limits of agreement (± 1.96 SD).



B)



C)

Figure 2. Bland-Altman plots illustrating level of agreement between GT3X+ MVPA and Fitbit MVPA classification estimates (continued). B) Troiano cut-points. C) VM3 cut-points. Dashed lines show 95% limits of agreement (± 1.96 SD).

Discussion

This study sought to compare the accuracy of the Fitbit Flex PA monitor against a previously validated accelerometer, the ActiGraph GT3X+, for classifying PA intensity in free-living settings. Our results demonstrated moderate to strong significant relationships between the Fitbit and GT3X+ monitors for SED and PA estimates. Though Fitbit tended to overestimate SED compared to GT3X+, these differences were not statistically significant. However, Fitbit MVPA estimates significantly differed from all GT3X+ estimates. The observed differences show there were greater discrepancies between Fitbit-determined MVPA estimates and GT3X+ cut-point criteria developed from single axis regression equations (i.e. Freedson and Troiano cut-points). However, regardless of GTX+ cut-points used, Fitbit overestimated mean daily MVPA by nearly an hour or more. Furthermore, BA plots showed these differences increased as volume of MVPA increased, suggesting that Fitbit may systematically overestimate MVPA compared to GT3X+.

Previous research has shown strong correlations for EE, step, and MVPA estimates between hip-worn Fitbit models and ActiGraph GT3X+.[35,171,183,184] Similarly, Fitbit Flex and GT3X+ MVPA estimates have strongly correlated in studies of young adult and elderly populations, with the latter study also reporting moderate correlations for LPA.[185,186] Our results show the MVPA correlations between Fitbit and GT3X+ estimates fall between that of these two studies. Differences in methodologies may partially explain these differences. Sushames et al., used a protocol lasting less than 24 hours, with a mix of scripted PA and free-living activity.[184] Alharbi and colleagues investigated free-living activity over a 4-day period with elderly subjects in a clinical population.[185] Our study collected free-living data over a longer period and with a more diverse age-range of healthy adults. Thus, the longer study protocol in our investigation may better represent the correlations between Fitbit and GT3X+ in free-living conditions.

That Alharbi et al. found stronger MVPA correlations than ours may be explained by differences in participant characteristics. The authors reported lower correlations between Fitbit and GT3X+ MVPA estimates for females versus males.[185] Whereas their study sample was mostly male, a majority of our participants were female. A similar investigation to ours found that over a 14-day study period, male participants spent greater time in MVPA than females, though the latter achieved more vigorous PA as proportion of total daily MVPA.[187] Our study shows greater discrepancies between Fitbit and GT3X+ for vigorous PA than moderate PA. These results may explain the lower correlation for MVPA found in our study comprised mostly of female participants. As other researchers have suggested, future studies should consider investigating the impact of participant sex on differences in PA estimates between PA monitors.[188]

Few studies have assessed the Fitbit Flex classification estimates for SED, particularly in free-living settings. Compared to IC criterion Fitbit Flex underestimates SED EE where as GT3X+ overestimated SED EE, though these findings were derived from a short protocol including only 20 minutes of SED.[164] A different study with an extensive free-living protocol showed Fitbit Flex overestimated the mean daily proportion of time spent in SED by 23% compared to GT3X+.[187] Previous research has shown accelerometers placed at the hip demonstrate less count variability than wrist and ankle placement over a wide range of sedentary and physical activities.[175,189] It may be that a hip-worn device may be better suited to capturing SED than a wrist-worn device. Indeed, recent evidence suggests a hip-worn Fitbit may yield similar SED estimates as GT3X+ in free-living settings, whereas SED estimates from a wrist-worn Fitbit are overestimated.[170] Nonetheless, our results show that, overall, Fitbit-determined SED estimates were not significantly greater than GT3X+-determined SED.

It may be that neither Fitbit Flex nor the ActiGraph GT3X+ are ideal monitors for estimating SED in free-living settings. To the point, defining count-based criteria for SED is inconsistent and may be operationalized to include variables such as posture, a variable not captured by the Fitbit Flex.[190] The GT3X+ has the low-frequency option, allowing the user to increase the monitor's sensitivity to movement by lowering the frequency threshold for recording accelerations. However, based on current evidence, it is unclear whether researchers should enable the low-frequency extension feature when initializing the ActiGraph if the goal is to specifically monitor SED.[156,191]

Though the Freedson and VM3 cut-points were derived from accelerometer output using different numbers of axis (i.e. vertical axis only versus vector magnitude of three axes), research has demonstrated that the equations perform similarly compared to IC-criterion.[31,155] Furthermore, though there is some evidence linear regression prediction equations from triaxial output may be superior to vertical axis output alone, the magnitude of the differences may be small, and these differences have not been tested between ActiGraph's triaxial and single-axis linear regression equations in free-living settings.[149] In our study, the MVPA estimates were significantly different between Freedson and Troiano cut-points compared to VM3 cut-points. However Fitbit MVPA estimates were consistently significantly higher than any GT3X+ estimates. Thus, we simply refer to the Freedson cut-points when discussing the comparisons of the Fitbit to GT3X+ MVPA estimates.

Our results show the Fitbit Flex and GT3X+ produce very different estimates of MVPA. Specifically, Fitbit Flex overestimated mean daily MVPA by nearly an hour, or more, compared to the ActiGraph GT3X+. The hip-worn Fitbit One has overestimated MVPA compared to GT3X+, with researchers reporting mean absolute percent errors of over 60%.[170] In our study Fitbit discrepancies may be exaggerated further due to the wrist placement of the Fitbit Flex. Indeed, Rosenberger and colleagues found that a hip-worn accelerometer has nearly twice the sensitivity in capturing MVPA than did a wrist-worn accelerometer.[175]

Recently Nelson and colleagues found Fitbit Flex overestimated the metabolic cost of walking (3.3 – 4.6 METs) and jogging (7.0-7.9 METs) activities compared to IC criterion.[188] However, the activities were only performed for five minutes. In our study, participants averaged nearly 30 minutes of ActiGraph-determined MVPA per day. Thus, we might expect the magnitude of the discrepancy between ActiGraph- and Fitbit-determined MVPA to be much greater. In support of this explanation in our analysis of the BA plots of Fitbit and GT3X+ MVPA, each data plot was below zero, indicating the Fitbit flex overestimated mean daily MVPA for each participant. We also observed a negative slope for the fit line, suggesting that this discrepancy tends to increase as total mean daily MVPA volume increases. Other research has found similar systematic bias for Fitbit Flex step estimates, but not for EE estimates.[164,185,186]

The strengths of this investigation include the length of the free-living protocol, the wide age range represented in the participant sample, and high number of valid wear days. Only one previous study has investigated the wrist-worn Fitbit in a protocol lasting at least seven days and that study only included 19 subjects between ages 19 – 37.[187] In addition, our investigation included a wrist-worn consumer-based accelerometer-based monitor, which are more popular than hip worn models and may potentially increase compliance in future research studies. Lastly, our investigation utilized the ActiGraph accelerometer-based monitor as a criterion measure using three validated prediction equations. Previous investigations have typically only compared the validity of consumer-based monitors to ActiGraph using one cut-point criteria.[35,170,184]

Certain limitations of this study must be considered when interpreting our results. Fitbit does not currently have a wear time validation mechanism *per se,* though other researchers have applied typical validation approaches to minute-by-minute Fitbit data where 60 consecutive minutes of no PA during waking hours are assumed to be non-weartime.[187] Thus, it is not possible to truly know if such occurrences are due to non-wear time or extensive SED. Limitations of using ActiGraph for assessing SED have been reported; however, previous research has shown acceptable estimates of SED compared to ActivPAL and IC criterion.[159,188] Both Fitbit Flex and GT3X+ are not completely water proof. Thus, we were unable to capture activities such as swimming or bathing for this analysis.

In conclusion, our data suggest that though the Fitbit Flex SED estimates are not significantly different from GT3X+, the monitors do not produce equivalent estimates of MVPA. In particular, the Fitbit Flex overestimates MVPA compared to GT3X+ regardless of the use of different cut-points. On-going population surveillance will benefit from improved objective monitoring options that will maximize subject compliance and data accuracy. Improving the accuracy of MVPA monitoring is paramount to increasing population adherence to the Physical Activity Guidelines for Americans. Consumer-based PA monitors, such as the Fitbit Flex, show promise for promoting PA adherence to the general public by allowing individuals to self-monitor daily PA. However, if the Fitbit Flex overestimates MVPA, this may reduce the likelihood that an individual would meet or exceed the minimum recommended MVPA. Further research is needed to investigate the accuracy and precision of Fitbit Flex PA classification estimates in free-living settings.

SUMMARY

This study investigated the concurrent validity of two accelerometer-based PA monitors in a free-living condition. Specifically, we compared the SED and PA estimates of wrist-worn consumer-based PA monitor, the Fitbit Flex, to the hip-worn ActiGraph PA monitor, GT3X+. NDSU students, faculty, and staff participants wore the monitors simultaneously for seven consecutive days. Aside from maintaining their day-to-day lifestyle, participants kept a daily log indicating sleep and wake times as well as any large amounts of physical activity outside of their normal routine (i.e. running a half marathon).

To determine Fitbit and GT3X+ SED and PA estimates, we first validated wear time during waking hours. Non-wear time for ActiGraph was determined by using both a previously validated non-wear time algorithm (Choi criteria)[157] and participants' logs indicating sleep and non-wear time. Similarly, Fitbit wear time was determined by subtracting sleep and non-wear time recorded in participant logs from Fitabase data. GT3X+ and Fitbit validated wear-time output were time stamp matched, minute-by-minute, so that only data from periods of simultaneous wear time were considered for analysis. GT3X+ daily SED and PA estimates were then calculated using three previously validated cut-point criteria (i.e. Freedson, Troiano, and Freedson's VM3 cut-points).[20,142,144] Fitbit SED and PA estimates were determined using Fitabase, a third party service which generates data reports for specific Fitbit PA monitors from the Fitbit API. Fitabase subscribers are able to obtain daily, hourly, and minute-by-minute totals of steps, EE, as well as time spent in SED, LIPA, and MVPA.

When comparing estimates between GT3X+ and Fitbit, we observed strong correlations for SED estimates ($r = .891$, $P < 0.01$), but only moderately correlated for MVPA estimates ($r = .658 - .766$, $P < .01$). Repeated one-way ANOVA showed Fitbit and GT3X+ SED and MVPA estimates were significantly different. After applying post hoc adjustments, differences in SED estimates were no longer significant, but MVPA estimates remained significant regardless of which ActiGraph cut-points were applied. Compared to ActiGraph Freedson, Troiano, and VM3 cut-points, Fitbit overestimated mean daily MVPA by 77, 79, and 60 min/day, respectively. We did not observe any apparent bias between Fitbit and GT3X+ SED estimates (mean difference = 35.83). However, BA plots illustrated that Fitbit consistently overestimated MVPA compared to GT3X, and these differences increased as mean daily volume of MVPA increased.

Few studies have compared Fitbit and GT3X+ SED PA estimates in free-living settings. Recent evidence shows the Fitbit Flex overestimates MVPA and SED over seven days compared to GT3X+.[187] Likewise, our results show large discrepancies between the Fitbit Flex and GT3X+ MVPA estimates, and inconclusive differences in SED estimates. Taken together, our results suggest Fitbit Flex may not be a suitable measure of MVPA in free-living conditions. Currently researchers (and users) are unable to access raw count data from Fitbit Flex. Access to this data would allow researchers to directly compare count-based output between the monitors and further our understanding of the relationship between the Fitbit Flex and GT3X+ PA estimates.

Though not a primary aim of this study, we found significant differences in daily SED and MVPA between students and professional employees. Our sample included 11 students as well as 56 faculty and staff from the North Dakota State University campus. It is interesting to note that, on average, students achieved about 30 minutes more MVPA and 51 minutes less SED than faculty and staff. It is possible that the different occupations lend themselves to different amounts of PA, particularly upper-body movement that may account for these differences. Age could possibly explain these differences, as PA generally tends to decrease as age increases.[20] Yet, overall this sample was relatively active, with both students and professional University employees attaining an average of 48 minutes of MVPA per day. This daily average would be enough to achieve the recommended amount of weekly MVPA. Thus, it is possible that both student and professional employees regularly engaged in leisure-time PA, but that students also were less sedentary as a function of their occupation. Previous Fitbit validation studies in free-living conditions have been limited to small convenience samples while laboratory studies have assessed scripted PA from common leisure-time and household activities (e.g. walking, resistance training exercises, or folding laundry).[164,187,188] Future studies should consider investigating larger samples to include representation from numerous occupations to further examine the validity of the Fitbit Flex PA estimates across diverse occupational patterns of adult PA.

In conclusion, though the Fitbit Flex is popular consumer-based physical activity monitor, there is little research validating its PA classification estimates in a free-living condition. This study demonstrated that Fitbit Flex and GT3X+ produce similar SED but substantially different MVPA estimates in free-living conditions. If it is to be used as tool for PA surveillance tool, further research is needed to determine the validity of the Fitbit Flex PA estimates, especially by including larger samples representing multiple occupational and age groups to capture diverse patterns of PA.

REFERENCES

1. Surgeon General's report on physical activity and health. From the Centers for Disease Control and Prevention. *JAMA.* 1996;276(7):522. doi:10.1001/jama.1996.03540070018010
2. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circ.* 2007;116(9):1081-93.
3. Hardman A. Acute Responses to Physical Activity and Exercise. In: Bouchard C, Blair SN, Haskell WL, eds. *Physical Activity and Health*. Champaign, IL: Human Kinetics; 2007:67-82.
4. Weuve J, Kang JH, Manson JE, Breteler MM, Ware JH, Grodstein F. Physical activity, including walking, and cognitive function in older women. *JAMA.* 2004;292(12):1454-61.
5. Lee IM. Dose-response relation between physical activity and fitness: even a little is good; more is better. *JAMA.* Vol 297. 2007:2137-39.
6. Paffenbarger RS, Jr., Hyde RT, Wing AL, Hsieh CC. Physical activity, all-cause mortality, and longevity of college alumni. *N Engl J Med.* 1986;314(10):605-13.
7. McAuley PA, Artero EG, Sui X, et al. The obesity paradox, cardiorespiratory fitness, and coronary heart disease. *Mayo Clin Proc.* 2012;87(5):443-51.
8. Blair SN, Kohl HW, 3rd, Paffenbarger RS, Jr., Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA.* 1989;262(17):2395-401.

9. Leon AS, Connett J, Jacobs DR, Jr., Rauramaa R. Leisure-time physical activity levels and risk of coronary heart disease and death. The Multiple Risk Factor Intervention Trial. *JAMA.* 1987;258(17):2388-95.

10. Siegel LC, Sesso HD, Bowman TS, Lee IM, Manson JE, Gaziano JM. Physical activity, body mass index, and diabetes risk in men: A prospective study. *Am J Med.* 2009;122(12):1115-21.

11. Katzmarzyk PT, Janssen I. The economic costs associated with physical inactivity and obesity in Canada: an update. *Can J Appl Physiol.* 2004;29(1):90-115.

12. Lee IM, Rexrode KM, Cook NR, Hennekens CH, Burin JE. Physical activity and breast cancer risk: The Women's Health Study (United States). *Cancer Causes Control.* 2001;12(2):137-45.

13. Lee IM. Physical activity and cancer prevention--data from epidemiologic studies. *Med Sci Sports Exerc.* 2003;35(11):1823-27.

14. Blair SN, Kohl HW, Barlow CE. Physical activity, physical fitness, and all-cause mortality in women: do women need to be active? *J Am Coll Nutr.* 1993;12(4):368-71.

15. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc.* 2001;33(6 Suppl):S379-99; discussion S419-20.

16. Antero Kesaniemi Y, Danforth EJ, Jensen MD, Kopelman PG, Lefebvre P, Reeder BA. Dose-response issues concerning physical activity and health: an evidence-based symposium. *Med Sci Sports Exerc.* 2001;33(6)(Supplement):S351-S358.

17. Pate RR, Pratt M, Blair SN, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA.* 1995;273(5):402-7.

18. US Department of Health and Human Services U. 2008 Physical Activity Guidelines for Americans. In: USDHHS, ed. Washington DC. 2008.

19. Hagstromer M, Ainsworth BE, Oja P, Sjostrom M. Comparison of a subjective and an objective measure of physical activity in a population sample. *J Phys Act Health.* 2010;7(4):541-50.

20. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181-8.

21. Carlson SA, Fulton JE, Schoenborn CA, Loustalot F. Trend and prevalence estimates based on the 2008 Physical Activity Guidelines for Americans. *Am J Prev Med.* 2010;39(4):305-13.

22. Oyeyemi AL, Umar M, Oguche F, Aliyu SU, Oyeyemi AY. Accelerometer-determined physical activity and its comparison with the international physical activity questionnaire in a sample of nigerian adults. *PLoS ONE.* 2014;9(1):e87233. doi:10.1371/journal.pone.0087233.

23. Ainsworth BE, Caspersen CJ, Matthews CE, Masse LC, Baranowski T, Zhu W. Recommendations to improve the accuracy of estimates of physical activity derived from self report. *J Phys Act Health.* 2012;9 Suppl 1:S76-84.

24. Hayden-Wade HA, Coleman KJ, Sallis JF, Armstrong C. Validation of the telephone and in-person interview versions of the 7-day PAR. *Med Sci Sports Exerc.* 2003;35(5):801-9.

25. Matthews CE, Ainsworth BE, Hanby C, et al. Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc.* 2005;37(6):986-94.

26. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S. adults: Compliance with the Physical Activity Guidelines for Americans. *Am J Prev Med.* 2011;40(4):454-61.

27. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity (Silver Spring).* 2007;15(10):2371-79.

28. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: new insights and validation studies. *Obes Rev.* 2013;14(6):451-62.

29. John D, Tyo B, Bassett DR. Comparison of Four ActiGraph accelerometers during walking and running. *Med Sci Sports Exerc.* 2010;42(2):368-74.

30. Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act.* 2012;9:84. doi:10.1186/1479-5868-9-84.

31. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med.* 2013;34(11):975-82.

32. Kaminsky L, Ozemek C. A comparison of the Actigraph GT1M and GT3X accelerometers under standardized and free-living conditions. *Physiol Meas.* 2012;33:1869-76.

33. NCHS. *National Health and Nutrition Examination Survey (NHANES) Physical Activity Monitor (PAM) Procedures Manual.* Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention;2011.

34. Sieverdes JC, Sui X, Hand GA, et al. Cardiometabolic results from an armband-based weight loss trial. *Diabetes Metab Syndr Obes.* 2011;4:187-94.

35. Lee J, Kim Y, Welk G. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc.* 2014;46(9):1840-48.

36. Nightingale TE, Walhim JP, Thompson D, Bilzon JL. Predicting physical activity energy expenditure in manual wheelchair users. *Med Sci Sports Exerc.* 2014;46(9):1849-58.

37. Phillips LR, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *J Sci Med Sport.* 2013;16(2):124-28.

38. van Hees VT, Renstrom F, Wright A, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS One.* 2011;6(7):e22922. doi:10.1371/journal.pone.0022922.

39. Hassani M, Kivimaki M, Elbaz A, Shipley M, Singh-Manoux A, Sabia S. Non-consent to a wrist-worn accelerometer in older adults: the role of socio-demographic, behavioural and health factors. *PLoS One.* 2014;9(10):e110816. doi:10.1371/journal.pone.0110816.

40. Danova T. Just 3.3 million fitness trackers were sold in the US last year. *Business Insider.* 2014. http://www.businessinsider.com/33-million-fitness-trackers-were-sold-in-the-us-in-the-past-year-2014-5. Accessed 10/2/2014.

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

42. Wolf G. The Data-Driven Life. *The New York Times.* 2010. http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=all&_r=0#. Accessed 10/2/2014.

43. Vooijs M, Alpay LL, Snoeck-Stroband JB, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. *Interact J Med Res.* 2014;3(4):e14. doi:10.2196/ijmr.3056.

44. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep.* 1985;100(2):126-31.

45. Bouchard C, Blair S, Haskell W. Why Study Physical Activity and Health? In: Bouchard C, Blair S, Haskell W, eds. *Physical Activity and Health*. Champaign, IL: Human Kinetics; 2007:3-21.

46. Yang C-C, Hsu Y-L. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors.* 2010;10(8):7772-88.

47. Sieverdes JC, Ray BM, Sui X, et al. Association between leisure time physical activity and depressive symptoms in men. *Med Sci Sports Exerc.* 2012;44(2):260-5.

48. Sesso HD, Paffenbarger RS, Jr., Lee IM. Physical activity and coronary heart disease in men: The Harvard Alumni Health Study. *Circ.* 2000;102(9):975-80.

49. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS, Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med.* 1991;325(3):147-52.

50. Gillum RF, Mussolino ME, Ingram DD. Physical activity and stroke incidence in women and men: The NHANES I Epidemiologic Follow-up Study. *Am J Epidemiol.* 1996;143(9):860-9.

51. Morris JN, Heady JA, Raffle PA, Roberts CG, Parks JW. Coronary heart-disease and physical activity of work. *Lancet.* 1953;265(6795):1053-7.

52. Rockhill B, Willett WC, Manson JE, et al. Physical activity and mortality: a prospective study among women. *Am J Public Health.* 2001;91(4):578-83.

53. Sherman SE, D'Agostino RB, Cobb JL, Kannel WB. Physical activity and mortality in women in the Framingham Heart Study. *Am Heart J.* Vol 128. 1994:879-84.

54. Lee IM, Skerrett P. Physical activity and all-cause mortality: what is the dose-response relation? *Med Sci Sports Exerc.* 2001;33(6):S459-S471.

55. Lee IM, Paffenbarger RS, Jr. Associations of light, moderate, and vigorous intensity physical activity with longevity. The Harvard Alumni Health Study. *Am J Epidemiol.* 2000;151(3):293-9.

56. Ainsworth BE, Haskell WL. Compendium of physical activities: Classification of energy costs of human physical activities. *J Am Diet Assoc.* 1993;93(6):71-80.

57. Paffenbarger RS, Jr., Hyde RT, Wing AL, Lee IM, Jung DL, Kampert JB. The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *N Engl J Med.* 1993;328(8):538-45.

58. Blair SN, Kampert JB, Kohl HW, 3rd, et al. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA.* 1996;276(3):205-10.

59. Lee IM, Paffenbarger RS, Jr. Physical activity and stroke incidence: the Harvard Alumni Health Study. *Stroke.* 1998;29(10):2049-54.

60. Hu FB, Stampfer MJ, Colditz GA, et al. Physical activity and risk of stroke in women. *JAMA.* 2000;283:2961-7.

61. Thune I, Furberg AS. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. *Med Sci Sports Exerc.* 2001;33(6 Suppl):S530-550; discussion S609-510.

62. Kampert JB, Blair SN, Barlow CE, Kohl HW, 3rd. Physical activity, physical fitness, and all-cause and cancer mortality: a prospective study of men and women. *Ann Epidemiol.* 1996;6:452-7.

63. Rockhill B, Willett WC, Hunter DJ, Manson JE, Hankinson SE, Colditz GA. A prospective study of recreational physical activity and breast cancer risk. *Arch Intern Med.* 1999;159(19):2290-6.

64. McTiernan A, Kooperberg C, White E, et al. Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women's Health Initiative Cohort Study. *JAMA.* 2003;290:1331-6.

65. Thune I, Lund E. The influence of physical activity on lung-cancer risk: A prospective study of 81,516 men and women. *Int J Cancer.* 1997;70:57-62.

66. Lee IM, Sesso HD, Paffenbarger RS, Jr. Physical activity and risk of lung cancer. *Int J Epidemiol.* 1999;28(4):620-5.

67. CDC. Deaths and Mortality. 2010; http://www.cdc.gov/nchs/fastats/deaths.htm. Accessed April 8, 2015.

68. Danaei G, Ding EL, Mozaffarian D, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med.* 2009;6(4):e1000058. doi:10.1371/journal.pmed.1000058.

69. Adult participation in recommended levels of physical activity--United States, 2001 and 2003. *MMWR Morb Mortal Wkly Rep.* 2005;54:1208-12.

70. Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. *Vital Health Stat* 2014;20(260):1-161.

71. Carlson SA, Densmore D, Fulton JE, Yore MM, Kohl HW, 3rd. Differences in physical activity prevalence and trends from 3 U.S. surveillance systems: NHIS, NHANES, and BRFSS. *J Phys Act Health.* 2009;6 Suppl 1:S18-27.

72. Church TS, Thomas DM, Tudor-Locke C, et al. Trends over 5 Decades in U.S. occupation-related physical activity and their associations with obesity. *PLoS ONE.* 2011;6(5):e19657. doi:10.1371/journal.pone.0019657.

73. Salmon J, Owen N, Crawford D, Bauman A, Sallis JF. Physical activity and sedentary behavior: a population-based study of barriers, enjoyment, and preference. *Health Psychol.* 2003;22(2):178-88.

74. Sarkin J, Campbell, J., Gross, L., Roby, J., Bazzo, S., Sallis, J., Calfas, K. Seven-Day Physical Activity Recall. *Med Sci Sports Exerc.* 1997;29(6):89-103.

75. Sallis JF, Haskell WL, Wood PD, et al. Physical activity assessment methodology in the Five-City Project. *Am J Epidemiol.* 1985;121(1):91-106.

76. Rauh MJ, Hovell MF, Hofstetter CR, Sallis JF, Gleghorn A. Reliability and validity of self-reported physical activity in Latinos. *Int J Epidemiol.* 1992;21(5):966-71.

77. Jacobs DR, Jr., Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc.* 1993;25(1):81-91.

78. Richardson MT, Leon AS, Jacobs DR, Jr., Ainsworth BE, Serfass R. Comprehensive evaluation of the Minnesota Leisure Time Physical Activity Questionnaire. *J Clin Epidemiol.* 1994;47:271-81.

79. Elosua R, Garcia M, Aguilar A, et al. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish women. *Med Sci Sports Exerc.* 2000;32(8):1431-7.

80. Craig CL, Marshall AL, Sjostrom M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35(8):1381-95.

81. Hagstromer M, Oja P, Sjostrom M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr.* 2006;9(6):755-62.

82. Centers for Disease Control and Prevention (CDC). *Behavioral Risk Factor Surveillance System Survey Questionnaire*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2015.

83. Strath SJ, Bassett DRJ, Ham SA, Swartz AM. Assessment of physical activity by telephone interview versus objective monitoring. *Med Sci Sports Exerc.* 2003;35(12):2112-8.

84. Yore MM, Ham SA, Ainsworth BE, et al. Reliability and Validity of the Instrument Used in BRFSS to Assess Physical Activity. *Med Sci Sports Exerc.* 2007;39(8):1267-74.

85. Ferrannini E. The theoretical bases of indirect calorimetry: a review. *Metabolism.*1988;37:287-301.

86. Shephard R. A critical examination of the Douglas bag technique. *J Physiol.* 1955;127(3):515-24.

87. Bassett DR, Howley ET, Thompson DL, et al. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. *J Appl Physiol.* 2001;91(1):218-24.

88. Parr BB, Strath SJ, Bassett DR, Jr., Howley ET. Validation of the Cosmed K4b2 portable metabolic measurement system. *Med Sci Sports Exerc.* 2001;33(5)Suppl(1):S300. doi:10.1097/00005768-200105001-01691

89. Ross R, Fudge B, Gibson A, Ojiambo R, Pitsiladis YF. K4b2 validity and reliability during fast outdoor running: 781: Board #2 3:15 PM - 5:15 PM. *Med Sci Sports Exerc.* 2011;43(5) Suppl(1):80. doi:10.1249/01.MSS.0000402913.38341.ef

90. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. *Eur J Appl Physiol.* 2010;109(2):159-71.

91. Salier Eriksson J, Rosdahl H, Schantz P. Validity of the Oxycon Mobile metabolic system under field measuring conditions. *Eur J Appl Physiol.* 2012;112(1):345-55.

92. Crouter SE, Albright C, Bassett DR, Jr. Accuracy of polar S410 heart rate monitor to estimate energy cost of exercise. *Med Sci Sports Exerc.* 2004;36(8):1433-9.

93. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol.* 2006;98(6):601-12.

94. Speakman JR. The history and theory of the doubly labeled water technique. *Am J Clin Nutr.* 1998;68(4):932S-938S.

95. Schoeller DA, van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol Respir Environ Exerc Physiol.* 1982;53(4):955-9.

96. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jequier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *Am J Physiol.* 1986;250(5 Pt 2):R823-830.

97. Schoeller DA, Racette SB. A review of field techniques for the assessment of energy expenditure. *J Nutr.* 1990;120(Suppl 11):1492-5.

98. Speakman JR, Nair KS, Goran MI. Revised equations for calculating CO2 production from doubly labeled water in humans. *Am J Physiol.* 1993;264(6 Pt 1):E912-917.

99. Leenders NY, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. *Med Sci Sports Exerc.* 2001;33(7):1233-40.

100. Schoeller DA, Webb P. Five-day comparison of the doubly labeled water method with respiratory gas exchange. *Am J Clin Nutr.* 1984;40(1):153-8.

101. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc.* 2011;43(5):867-76.

102. Bradfield RB. A technique for determination of usual daily energy expenditure in the field. *Am J Clin Nutr.* 1971;24(9):1148-54.

103. LaPorte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep.* 1985;100(2):131-46.

104. Fohr T, Tolvanen A, Myllymaki T, et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. *J Occup Med Toxicol.* 2015;10:39. doi:10.1186/s12995-015-0081-6.

105. Tulppo MP, Hautala AJ, Mäkikallio TH, et al. Effects of aerobic training on heart rate dynamics in sedentary subjects. *J Appl Physiol.* 2003;95(1):364-72.

106. Christensen CC, Frey HM, Foenstelien E, Aadland E, Refsum HE. A critical evaluation of energy expenditure estimates based on individual O2 consumption/heart rate curves and average daily heart rate. *Am J Clin Nutr.* 1983;37(3):468-72.

107. Vokac Z, Bell H, Bautz-Holter E, Rodahl K. Oxygen uptake/heart rate relationship in leg and arm exercise, sitting and standing. *J Appl Physiol.* 1975;39(1):54-9.

108. Armstrong LE, Maresh CM, Gabaree CV, et al. Thermal and circulatory responses during exercise: effects of hypohydration, dehydration, and water intake. *J Appl Physiol (1985).* 1997;82(6):2028-35.

109. Strath SJ, Swartz AM, Bassett DRJ, O'Brien WL, King GA, Ainsworth BE. Evaluation of heart rate as a method for assessing moderate intensity physical activity. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S465-S470.

110. Wareham NJ, Hennings SJ, Prentice AM, Day NE. Feasibility of heart-rate monitoring to estimate total level and pattern of energy expenditure in a population-based epidemiological study: the Ely Young Cohort Feasibility Study 1994-5. *Br J Nutr.* 1997;78:889-900.

111. Ceesay SM, Prentice AM, Day KC, et al. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br J Nutr.* 1989;61(2):175-86.

112. Livingstone MB, Prentice AM, Coward WA, et al. Simultaneous measurement of free-living energy expenditure by the doubly labeled water method and heart-rate monitoring. *Am J Clin Nutr.* 1990;52(1):59-65.

113. Livingstone MB, Coward WA, Prentice AM, et al. Daily energy expenditure in free-living children: comparison of heart-rate monitoring with the doubly labeled water (2H2(18)O) method. *Am J Clin Nutr.* 1992;56(2):343-52.

114. Rennie KL, Hennings SJ, Mitchell JO, Wareham NJ. Estimating energy expenditure by heart-rate monitoring without individual calibration. *Med Sci Sports Exerc.* 2001;33(6):939-45.

115. Schrack JA, Zipunnikov V, Goldsmith J, Bandeen-Roche K, Crainiceanu CM, Ferrucci L. Estimating energy expenditure from heart rate in older adults: a case for calibration. *PLoS ONE.* 2014;9(4):e93520. doi:10.1371/journal.pone.0093520.

116. Walter PL. The History of the Accelerometer: 1920s-1996 - Prologue and Epilogue, 2006. J *Sound Vib.* 2007;41(1):84-92.

117. Gamecho B, Silva H, Guerreiro J, Gardeazabal L, Abascal J. A context-aware application to increase elderly users compliance with physical rehabilitation exercises at home via animatronic biofeedback. *J Med Syst.* 2015;39(11):135. doi:10.1007/s10916-015-0296-1.

118. Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. *Sleep Breath.* 2012;16(3):913-7.

119. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *J Med Internet Res.* 2014;16(8):e192. doi:10.2196/jmir.3469.

120. Thorndike AN, Mills S, Sonnenberg L, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. *PLoS One.* 2014;9:e100251. doi:10.1371/journal.pone.0100251.

121. Li-Peng W, Wolf RA, Jr., Yu W, et al. Design, fabrication, and measurement of high-sensitivity piezoelectric microelectromechanical systems accelerometers. *J Microelectromech Syst.* 2003;12(4):433-9.

122. Chen KY, Janz KF, Zhu W, Brychta RJ. Redefining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc.* 2012;44(1 Suppl 1):S13-23.

123. Lowe SA, Olaighin G. Monitoring human health behaviour in one's living environment: a technological review. *Med Eng Phys.* 2014;36(2):147-68.

124. Ridgers ND, Fairclough S. Assessing free-living physical activity using accelerometry: Practical issues for researchers and practitioners. *Eur J Sport Sci.* 2011;11(3):205-13.

125. Cleland I, Kikhia B, Nugent C, et al. Optimal placement of accelerometers for the detection of everyday activities. *Sensors (Basel).* 2013;13(7):9183-200.

126. Bouten CV, Sauren AA, Verduin M, Janssen JD. Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Med Biol Eng Comput.* 1997;35(1):50-6.

127. Wong TC, Webster JG, Montoye HJ, Washburn R. Portable accelerometer device for measuring human energy expenditure. *IEEE Trans Biomed Eng.* 1981;28(6):467-71.

128. Khusainov R, Azzi D, Achumba IE, Bersch SD. Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations. *Sensors (Basel).* 2013;13(10):12852-902.

129. Bouten CV, Westerterp KR, Verduin M, Janssen JD. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Med Sci Sports Exerc.* 1994;26(12):1516-23.

130. Chen KY, Sun M. Improving energy expenditure estimation by using a triaxial accelerometer. *J Appl Physiol.* 1997;83(6):2112-22.

131. Bassett DR, Jr., Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc.* 2012;44(1 Suppl 1):S32-38.

132. John D, Sasaki J, Staudenmayer J, Mavilia M, Freedson PS. Comparison of raw acceleration from the GENEA and ActiGraph GT3X+ activity monitors. *Sensors (Basel).* 2013;13(11):14754-63.

133. Chen KY, Bassett DR, Jr. The technology of accelerometry-based activity monitors: current and future. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S490-500.

134. Woodward MI, Cunningham JL. Skeletal accelerations measured during different exercises. *Proc Inst Mech Eng H.* 1993;207(2):79-85.

135. Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S450-6.

136. Welk GJ, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI Actigraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity (Silver Spring).* 2007;15(4):918-28.

137. Leenders NY, Sherman WM, Nagaraja HN. Energy expenditure estimated by accelerometry and doubly labeled water: do they agree? *Med Sci Sports Exerc.* 2006;38(12):2165-72.

138. Esliger DW, Probert A, Connor Gorber S, Bryan S, Laviolette M, Tremblay MS. Validity of the Actical accelerometer step-count function. *Med Sci Sports Exerc.* 2007;39(7):1200-4.

139. Kelly LA, McMillan DG, Anderson A, Fippinger M, Fillerup G, Rider J. Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. *BMC Med Phys.* 2013;13(1):5. doi:10.1186/1756-6649-13-5.

140. Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S471-480.

141. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc.* 2005;37(11 Supplement):S501-S511.

142. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998;30(5):777-81.

143. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S442-449.

144. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport.* 2011;14(5):411-6.

145. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR, Jr. Refined two-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc.* 2010;42(5):1029-37.

146. Matthews CE, Hagstromer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. *Med Sci Sports Exerc.* 2012;44(1 Suppl 1):S68-S76.

147. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc.* 2012;44(1 Suppl 1):S1-4.

148. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. *Am J Epidemiol.* 2008;167(7):875-81.

149. Rothney MP, Schaefer EV, Neumann MM, Choi L, Chen KY. Validity of physical activity intensity predictions by ActiGraph, Actical, and RT3 accelerometers. *Obesity (Silver Spring).* 2008;16(8):1946-52.

150. Dinger MK, Behrens TK. Accelerometer-determined physical activity of free-living college students. *Med Sci Sports Exerc.* 2006;38(4):774-9.

151. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol.* 2011;111(2):187-201.

152. Metzger JS, Catellier DJ, Evenson KR, Treuth MS, Rosamond WD, Siega-Riz AM. Patterns of objectively measured physical activity in the United States. *Med Sci Sports Exerc.* 2008;40(4):630-8.

153. Crouter SE, Clowers KG, Bassett DR, Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol (1985).* 2006;100(4):1324-31.

154. Rothney M, Brychta R, Meade N, Chen K, Buchowski M. Validation of the ActiGraph two–regression model for predicting energy expenditure. *Med Sci Sports Exerc.* 2010;42(9):1785-92.

155. Kim Y, Welk GJ. Criterion validity of competing accelerometry-based activity monitoring devices. *Med Sci Sports Exerc.* 2015;47(11):2456-63.

156. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson P. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc.* 2011;43(8):1561-7.

157. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc.* 2011;43(2):357-64.

158. Matthews CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S512-522.

159. Matthews CE, Keadle SK, Sampson J, et al. Validation of a previous-day recall measure of active and sedentary behaviors. *Med Sci Sports Exerc.* 2013;45(8):1629-38.

160. Allied Business Intelligence I. https://www.abiresearch.com/press/q1-2014-wearable-shipments-activity-trackers-outso. Accessed October, 2014.

161. The Harris poll rolls out results of survey on wearable tech. *Professional Services Close-Up.* 2013. http://search.proquest.com/docview/1458832815?accountid=6766. Accessed March 14, 2015.

162. Comstock J. Report: Fitbit and Samsung currently leading wearables market. *Mobi Health News.* 2015. http://mobihealthnews.com/40640/report-samsung-and-fitbit-currently-leading-wearables-markets/. Accessed March 14, 2015.

163. Wang JB, Cadmus-Bertram LA, Natarajan L, et al. Wearable sensor/device (Fitbit One) and sms text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. *Telemed J E Health.* 2015;21(10):782-92.

164. Bai Y, Welk GJ, Nam YH, et al. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sports Exerc*. 2015;48(1):151-8.

165. Albert MV, Deeny S, McCarthy C, Valentin J, Jayaraman A. Monitoring Daily Function in Persons With Transfemoral Amputations Using a Commercial Activity Monitor: A Feasibility Study. *Pm R*. 2014;6(12):1120-7.

166. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *J Sci Med Sport*. 2014;17(5):496-500.

167. Fortune E, Lugade V, Morrow M, Kaufman K. Validity of using tri-axial accelerometers to measure human movement - Part II: Step counts at a wide range of gait velocities. *Med Eng Phys*. 2014;36(6):659-69.

168. Fulk GD, Combs SA, Danks KA, Nirider CD, Raja B, Reisman DS. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. *Phys Ther*. 2014;94(2):222-9.

169. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Research Notes. 2014;7:952. doi:10.1186/1756-0500-7-952.

170. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Med Sci Sports Exerc*. 2015;48(3):457-65.

171. Noah JA, Spierer DK, Gu J, Bronner S. Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. *J Med Eng Technol*. 2013;37(7):456-62.

172. Sasaki J, Hickey A, Mavilia M, et al. Validation of the Fitbit Wireless Activity Tracker® for prediction of energy expenditure. *J Phys Act Health*. 2015;12(2):149-54.

173. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*. 2013;45(11):2105-12.

174. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc*. 2013;45(5):964-75.

175. Chakravarty EF, Hubert HB, Lingala VB, Fries JF. Reduced disability and mortality among aging runners: a 21-year longitudinal study. *Arch Intern Med*. 2008;168(15):1638-46.

176. Schuna JM, Jr., Johnson WD, Tudor-Locke C. Adult self-reported and objectively monitored physical activity and sedentary behavior: NHANES 2005-2006. *Int J Behav Nutr Phys Act*. 2013;10:126. doi:10.1186/1479-5868-10-126.

177. Haskell WL. Physical activity by self-report: a brief history and future issues. *J Phys Act Health*. 2012;9 Suppl 1:S5-10.

178. Fitbit Inc. Fitbit Flex Product Manual. 2015:1-25.

179. ActiGraph Corp. GT3X + and wGT3X+ Device Manual. 2015:1-15.

180. Ham SA, Reis JP, Strath SJ, Dubose KD, Ainsworth BE. Discrepancies between methods of identifying objectively determined physical activity. *Med Sci Sports Exerc*. 2007;39(1):52-8.

181. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc*. 2005;37(11 Suppl):S582-588.

182. Gusmer R, Bosch T, Watkins A, Ostrem J, Dengel D. Comparison of FitBit® Ultra to ActiGraph™ GT1M for assessment of physical activity in young adults during treadmill walking. *Open Sports Med J*. 2014;8:11-5.

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

184. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. *Eur J Prev Cardiol*. 2016;23(14):1476-85.

185. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One*. 2016;11(9):e0161224. doi:10.1371/journal.pone.0161224.

186. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. *JMIR Mhealth Uhealth*. 2016;4(3):e110. doi:10.2196/mhealth.6281.

187. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. *Med Sci Sports Exerc*. 2016;48(8):1619-28.

188. Ozemek C, Kirschner MM, Wilkerson BS, Byun W, Kaminsky LA. Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiol Meas*. 2014;35(2):129-38.

189. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. *Exerc Sport Sci Rev*. 2010;38(3):105-13.

190. Carr LJ, Mahar MT. Accuracy of intensity and inclinometer output of three activity monitors for identification of sedentary behavior and light-intensity activity. *J Obes.* 2012;2012:460271. doi:10.1155/2012/460271.