November 2017

A COMPREHENSIVE VALIDATION OF ACTIVITY TRACKERS FOR ESTIMATING PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR IN FREE-LIVING SETTINGS

Albert R. Mendoza
University of Massachusetts Amherst

A COMPREHENSIVE VALIDATION OF ACTIVITY TRACKERS FOR ESTIMATING PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR IN FREE-LIVING SETTINGS

A Dissertation Presented

by

ALBERT R. MENDOZA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 2017

Department of Kinesiology

A COMPREHENSIVE VALIDATION OF ACTIVITY TRACKERS FOR ESTIMATING PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR IN FREE-LIVING SETTINGS

A Dissertation Presented

by

ALBERT R. MENDOZA

Approved as to style and content by:

____
Patty S. Freedson, Chair

____
Kate Lyden, Member

____
John Sirard, Member

____
John Staudenmayer, Member

____
Catrine Tudor-Locke, Member

____
Jane Kent, Department Chair
Department of Kinesiology

DEDICATION

To my wife and mother

ACKNOWLEDGMENTS

To my advisor, Dr. Patty S. Freedson. You are amazing! I can only hope to be half the person that you are – compassionate, selfless and a stellar researcher. You have taken the time to train and mentor me and I appreciate all that you have done for me. I will never forget the time that we talked in your office and I asked if you were okay with my plan to pursue a tenure-track faculty position, to which you replied, “all I want is for you to be happy.” Thank you for giving me the opportunity to live a happier life. I will continue my path to share my knowledge, create knew knowledge in the field of PA and Health and continue the legacy of Team Freedson.

To the rest of my Committee members. Dr. Kate Lyden. Also, a disciple of Patty. Thank you for your guidance, input and inspiration. You have always made time for me and I appreciate it. Your expertise and extensive knowledge of both private industry and the field of objective measurement of PA were paramount to the success of my dissertation project. You continue to challenge and inspire me. I love every minute of it and I look forward to a life filled with more. Dr. John Sirard. Another disciple of Patty. Thank you for all your expertise and advise. You have been a great mentor and I appreciate all that you have done for me. You always found the time to meet with me and provide valuable insight and feedback. Dr. John Staudenmayer. Thank you for taking time to train me in the art of R Programming. You have taught me so much about simplifying complex data and helping me to develop new scientific approaches to data. Also, thank you for Co-Sponsoring (with Patty) me on my NIH grant. There is no way that I could have done it without you. You are one of the most positive people that I know and you are one hell of an athlete. I still look forward to beating me up on the trails one day. Dr. Catrine Tudor-Locke. Thank you for your valuable input and for helping to make my project a success. You always found time to meet with me when I needed it and I appreciate it.

Dr. Barry Braun. You are responsible for me being here at Umass. I remember meeting you at ACSM in Baltimore, MD. Dr. Marialice Kern spoke so highly of you and was kind enough to introduce us. After meeting you, all I could think about was getting here to Umass Amherst to continue my academic career in Kinesiology. Thank you for mentoring me and being such a great friend. Also, thank you for introducing me to the Shutesbury Coffee Cake Club.

My former Physical Activity & Health Lab members. Dr. Amanda Hickey. You are like Wonder Woman. Thank you for all your support and friendship. I cannot imagine how it would have been without you here. We made a great team and I look forward to working with you in the future. Dr. Jeffer Sasaki. Thank you for all your mentorship and friendship. You are such a good person and a great researcher. I look forward to working with you in the future as well. ¡Viva Brazil! Dr. Dinesh John. You have been such a good friend and mentor to me. Thank you for sharing your expertise with me. You helped to strengthen my project and I appreciate it.

Dr. Sandy Peterson and the Northeast Alliance for the Graduate Education and the Professoriate (NEAGEP). Sandy, I want to thank you for your support throughout my entire PhD experience. You always had a way of knowing when I was in need. Be it funding, writing, academics and/or personal. You have always been there to assist. There is absolutely no way that I would have been successful without the support that you and NEAGEP provided me. NEAGEP works. I am a product of NEAGEP and I will continue to advocate for the NEAGEP program.

The National Institutes of Health (NIH): National Heart Lung and Blood Institute (NHLBI). I want to thank the NHLBI for supporting my dissertation project. Specifically, I was funded via the Ruth L. Kirschstein National Research Service Award Individual Predoctoral Fellowship to Promote Diversity in Health-Related Research (Parent F31 - Diversity). This is an invaluable funding mechanism, which aims to enhance the diversity of the health-related research workforce by supporting the research training of predoctoral students from population groups that have been shown to be underrepresented in the biomedical, behavioral, or clinical research workforce, including underrepresented racial and ethnic groups and those with disabilities. I look forward to continuing my relationship with the NIH and growing as an independent researcher in the field of PA and Health. Thank you, NIH and NHLBI for supporting me.

My family. My mom, Elena Rosemary Mendoza. Thank you for being my everything all the time. You raised us on your own. You worked hard and sacrificed all to give us a better life. You never complained or blamed anyone for our situation. Rather, you found a way to give us (Steve, Ramon and I) all that you could give. You did not have a college degree but were still a very successful business woman. Your work and life experiences ranged from the fields and the barrio to the boardroom and a neighborhood with sidewalks. You too, grew up without a father and found a way to survive. You taught me many life-skills that I will continue to use throughout my time on earth. My PhD is dedicated to you and Maricela (wife) and I am very sad that you are not with us anymore and as a result cannot enjoy this momentous occasion with us. All I

have ever wanted is to make you proud of me - at this moment, I hope that you are able. I love you and I miss you terribly. My brother, Ramon. You have always been an inspiration to me. As a scientist, you have been a great example of what it takes to be a successful researcher. As a brother, you have been a rock. I love you. To my brother, Steve. You have always taught me to be a good person and to find the good in others. You are loving, hard-working and accepting of all others. Thank you for being my big brother. To my extended family: Antonio, Elena, Mona, Frances "Panche" and Rachel Mendoza; Adolpho, Margaret, little Margaret, Sylvia, Frances, Adolpho Junior "Popo" and Danny Nava; Rosendo, Maria, Lydia and Velia Mendoza; Franciso, Rosie, Larry and Sharon Nava; Herlinda Nava; Pete and Ester Nava; Pete, Andy, Alfonso, Rudy Henry and Mike Nava; Eulalio and Sergio Frausto; Pauline Ramirez; Lisa Maciel; Raymond Mendoza; Manuel, Lola, Raymond (dad) and Josie Mendoza and many others. All came to California and began working in the fields to provide a better life themselves and others. Thank you for instilling in me the "Nava-Mendoza fire" that burns inside of me driving me to never give up. To be proud of who I am and where I came from and to never forget those who have sacrificed for me. I love you and hope that I am making you proud as well.

From San Francisco State University, Drs. Matt Lee and Marialice Kern. You have both been such an inspiration to me and have played an integral part in my collegiate success. Matt, I remember sitting in your Exercise Physiology Lecture thinking to myself, "this is what I want to do." Marialice, you have been such a wonderful mentor for me and I appreciate all that you have taught me. Thank you for supporting me and being such great friends. I look forward to spending more time with you.

Dr. Robert Urtecho. You started me on this path. You are the person that I wanted to become. Sitting in your 8:00 am (WHAT?) Anatomy lecture, listening to you and enjoying your enthusiasm for the discipline was the impetus to a positive change in my life's path. Thank you for being a friend and colleague during the best and worst of times.

Maurice White and Pamela Perry. I still think about all the lessons that I learned during out time together in Los Angeles, CA. Maurice, I still quote you. Thank you both for helping me to realize my potential and for not giving up on me.

Felton "Sonny", Marilyn, Kevin, Mickey, Craig and Kirk Prescott. Thank you for your love and support. You have always been good to me and I will always remember what you have done for me.

Edward, Mary, Brady and Trevor Pendleton. You are such a wonderful family. Thank you for being such great friends to my family and myself. Brady, I have always been envious (in a good way) of the bond that you had with your dad. I know that he is proud of you and the father that you have become. I love you, Brother.

Un-Il "Walter", Un-Chol "Scott", Chun, Sun, Young and Un Yi, Dolson Kwan, Gene Sablan, Joe Dawson, Mark, Jeff and Pat Scott and others. I credit you with helping me to realize my competitiveness and drive to succeed. I will always remember those Summers that we played sports (football, basketball, baseball, soccer, etc.) from sun up to sun down. We used to beat the hell out of each other and then go hang out and eat at one of our houses. I still remember like it was yesterday, how some of you used to help me finish my chores so that we could go out to play. You have been an inspiration to me and I will always hold you in a special place within my heart.

My New England running community and friends. The Shutesbury Coffee Cake Club, the Amity Hill Horror Running Club, Tuesday night and (sometimes) Saturday morning trail running group and Noonrun. Thank you for helping me to keep my head straight. I look forward to seeing you on the West Coast when you visit.

Rich and Rebekah Wood and my Restaurant Family (especially David O'Malley a.k.a O'Malley). Thank you for giving me the opportunity to develop the life-skills needed to be successful in graduate school and life in general. Indeed, I am built restaurant tough. Rich, you have always supported me, mentored me and given me the freedom to make plenty of mistakes and to learn from them. I love you all very much.

Finally, my beautiful wife, Maricela. I have always said that you are my hero. You have been my savior throughout this whole process. Though I know this is not what you signed up for, I appreciate the love and encouragement that you continue to provide me. You are the most brilliant, stunningly gorgeous, hard working, thoughtful, gracious and giving person that I know. It is because of you that I could accomplish this degree - I owe you my life. I guess that I will spend the rest of my life repaying you. And I am fine with it. I love you! As educators, I am excited for us and the unique opportunity that we have to imbue positive change in people's lives and making this world a better place than it was when we entered it.

I feel responsible for setting an example for others like me. I am a product many who have sacrificed so that I may be successful. I am humbled and welcome the challenges that undoubtedly lie ahead. ¡Si se Puede!

ABSTRACT

A COMPREHENSIVE VALIDATION OF ACTIVITY TRACKERS FOR ESTIMATING PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR IN FREE-LIVING SETTINGS

SEPTEMBER 2017

ALBERT R. MENDOZA, B.S., SAN FRANCISCO STATE UNIVERSITY

M.S., SAN FRANCISCO STATE UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Patty S. Freedson

The aim of study one of this dissertation was to compare consumer activity trackers (ATs) with the research-grade ActiGraph™ GT3X-BT accelerometer (AG) in estimating energy expenditure (EE) and steps during orbital shaking at different frequencies. To address this aim, we utilized an electronic orbital shaking protocol (twenty-four, 3-minute trials; 2-hour trials). For all comparisons, the AG served as the reference measure. In the 3-min protocol, we showed that on average, the NL-1000 pedometer (NL) produced the lowest error (-9 steps/3-min) at 0.9 Hz (corresponding to moderate intensity). The magnitude of the error for the NL was 14 steps/3-min at a 3.0 Hz frequency (corresponding to very vigorous intensity). For the 2-hr protocol, estimates from all others were equivocal, with some overestimating steps (bias range: 1,331 steps/2-hrs for the Misfit Shine to 1,921 steps/2-hrs for the Misfit Flash [MFF]). For estimated EE bias ranged from26.6 kcals/2-hrs for the MFF to 45.8 kcals/2-hrs for the Misfit Shine. For other ATs, steps were underestimated (bias range: -5,770 steps/2-hrs for the Garmin Vivofit [GV] to -570 steps/2-hrs for the NL). For EE, the bias ranged

from -436.8 kcals/2-hrs for the GV to -261.7 kcals/2-hrs for the Fitbit Flex [FBF]). This study provides evidence about the differences in prediction algorithms by device across a broad range of oscillation frequencies that corresponded to different PA intensity levels.

For study two, we sought determine the accuracy and precision of activity trackers (ATs) in estimating steps, EE, activity minutes and sedentary time compared to direct observation (DO)-derived measures (criterion measures) in free-living settings. We also validated commonly used research-grade devices (e.g. hip-worn AG (AGhip), wrist-worn AG (AGwrist). Thirty-two healthy men and women (50% female, 37.5% minority; mean ± SD: Age = 32.3 ± 13.3 years; BMI = 24.4 ± 3.3 kg·m-2) were directly observed while completing three, 2-hour visits on different days while wearing ten ATs, three research-grade devices and a biometric shirt. A validated DO system was used to derive criterion measures for activity and sedentary time (ST) outcomes. ATs were accurate with varying precision in estimating physical activity (PA) behaviors in free-living settings. Additionally, ATs and research-grade accelerometers performed similarly (e.g. more accurate in estimating steps and less accurate in estimating moderate-to-vigorous PA [MVPA] minutes). For all devices, step estimates were accurate and strongly correlated (r range: 0.91 for the Apple iWatch to 0.97 for the AGhip) with criterion measures but EE and MVPA estimates were less accurate and more variable (EE: r = 0.32 [GV] to r = 0.85 [AGhip]; MVPA: r = 0.2 [NL] to r = 0.75 [AGhip]). For ATs, estimates of sedentary time were the least accurate and weakly correlated (r=0.06 Fitbit One [FBO] and FBF) with criterion measures derived from DO. Implications from this study are that consumers and the research community using ATs such as Fitbit (FB) to track steps can be confident in estimating steps but less confident in estimating sedentary time. This study advances our understanding of the performance characteristics of ATs in free-living natural settings using a validated DO method to derive PA and ST measures. This work significantly advances the field of activity monitor validation that should set the standard for future work.

The aims of study three were: 1) to examine the ability of ATs to detect change in PA and ST in free-living settings and 2) to examine the ability of research-grade accelerometers to detect change in PA and ST in free-living settings. To address these aims, we used an innovative approach to analyze data from study two. We defined change as a visit-to-visit difference that was greater than the within-subject standard deviation of the criterion measure (estimated by a linear-mixed model). Confusion matrices were used to examine percent agreement between DO visit-to-visit change and device visit-to-visit change. Key findings were focused on the widely used FBO and FBF and research-grade devices. We showed that, there was similar agreement between the hip-worn FBO and FBF with AGhip and AGwrist in estimates of change in steps (89.1% FBO, 88.8% FBF and 88.3% AGwrist, 91.4% AGhip correct classification), EE (73.4% FBO, 70.6% FBF and 77.0% AGhip correct classification) and MVPA minutes (accept FBF) (79.7% FBO, 65.2% FBF and 71.2% AGwrist, 77.0% AGhip correct classification) with criterion measured change. However, change in ST was more difficult to detect for the FB and AGhip (46.8% FBO, 42.3% FBF, 53.1% AGhip and 72.7% AGwrist correct classification). This novel study provides evidence that as an alternative to research-grade accelerometers, researchers may employ FB to measure step accumulation pre- and post-intervention and have a satisfactory level of confidence in FB change detection.

This work significantly advances the field of activity monitor validation research and informs intervention practices that should set the standard for future work. This body of work provides the first comprehensive validation of ATs from highly controlled orbital shaker testing to directly-observed free-living settings. This translational research which has broad applications for using ATs for surveillance and intervention research and by the consumer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TERMS AND ABBREVIATIONS

AG – ActiGraph

AGhip – Hip-worn ActiGraph GT3X-BT

AGwrist – Wrist-worn ActiGraph GT3X-BT

AiW – Apple iWatch Sport

AT – Activity Tracker

Bias - Measurement bias is the average difference between predicted and criterion measures

DO – Direct Observation, criterion measure for activity type

EE – Energy expenditure

FBF – Fitbit Flex

FBO – Fitbit FBO

Guideline MVPA – The U.S. PA Guidelines defined MVPA. Activities where intensity is greater than 2.99 METS.

GV – Garmin Vivofit

HxSkin – Hexoskin biometric shirt

MB – Microsoft Band

MET – Metabolic equivalent defined as an oxygen consumption of 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$. This value approximates resting oxygen consumption

MFF – Misfit Flash

MFS – Misfit Shine

MVPA – Moderate-to-vigorous physical activity, typically defined as an absolute activity intensity raging from 3.0-6.0 METs.

Non-Guideline MVPA – Activity tracker proprietary estimates that are not explicitly defined.

PA – Physical activity

PL – Polar Loop

SB – Sedentary behavior

ST – Sedentary time

WP – Withings Pulse

CHAPTER 1
INTRODUCTION

Statement of the Problem

Lack of physical activity (PA) is strongly implicated in virtually all leading causes of chronic disease morbidity and mortality. To attenuate the prevalence of these preventable chronic diseases and promote health benefits, the U.S. Government recommends that Americans engage in at least 150 minutes of moderate PA per week, [1] increase daily expenditure approximately 150 kilocalories (kcals) per day (equivalent to about 1,000 kilocalories/week) [2] and/or accumulate at least 10,000 steps/day. [3] Additionally, sedentary behavior (SB) recommendations from Australia state that adults should minimize the amount of time spent in prolonged sitting and break up long periods of sitting. [4] Dissemination of these recommendations has led to a heightened awareness of the importance of PA monitoring as a strategy for chronic disease management. Tools such as wearable devices to track personal PA provide a mechanism to be more informed about activity behavior. As a result, consumer devices that track PA behavior are increasingly popular for researchers, the general public, and developers and manufacturers of activity trackers (ATs).

According to a recent report, the global wearable technology market will grow from over $30 billion in 2016 and should reach over $150 billion in 2026. [5] Activity trackers such as, the Fitbit (Fitbit Inc., San Francisco, CA) provide estimates of activity minutes, sedentary time (sitting), energy expenditure (EE) and steps. According to a recent report, Fitbit remained the leading brand in ATs in 2015, accounting for 79 percent of sales. [6] This expanding market for ATs is driven in part by lower cost, longer battery life, more memory (e.g. to store data for days or weeks. However, growth of the market and advances in consumer device technology far outpace our knowledge about the validity of such devices. This gap is of major concern since it is not clear if these devices provide accurate information. Therefore, to address this problem, it is essential to improve our understanding of the accuracy and precision of the activity output measures of consumer devices.

Our group [7-9] and several other research teams [10-14] have conducted research to improve our understanding of the accuracy and precision of research-grade activity monitors to estimate PA intensity (e.g. minutes of moderate-to-vigorous physical activity [MVPA]) and activity EE. The vast majority of this validation work has been performed in laboratory settings where specific activities are performed over pre-determined and fixed time intervals where EE is measured with portable metabolic measurement equipment. [7,9,15] This calibration work on research-quality activity monitors uses signals from the monitors to generate simple (e.g. linear regression) to complex (e.g. machine learning) algorithms to estimate activity intensity, activity type, and EE that are freely available to apply to data collected with these research-quality devices. In contrast, the consumer monitors use proprietary algorithms that provide users with estimates of steps, EE, activity minutes, sedentary time, and other related measures. The accuracy and precision of the AT output (e.g. steps, PA EE, minutes of activity) in free-living settings, is not well understood.

To date, most AT validation studies have been performed under controlled laboratory conditions. This is a reasonable first step, but to truly understand the accuracy and precision of consumer ATs, validation studies must be performed in free-living settings while people are engaged in natural behaviors. In addition, ATs are often implemented to monitor improvements in PA behaviors, thus, exploring and understanding the accuracy of ATs in detecting change in PA and ST [16] is needed.

Four output variables have been studied in the investigations that have tested the accuracy of ATs: number of steps, EE, activity minutes (moderate-to-vigorous activity), achievement of PA recommendations and sedentary time. The results of these studies are equivocal. Activity trackers under- or overestimate these measures with substantial between-subject variability. For step counts, seven studies showed, ATs overestimated steps in laboratory settings [17-23] and thirteen studies showed ATs underestimated steps. [17-20,23-31] In free-living settings, four studies showed that ATs overestimated steps and lack precision, [21,32-34] and two studies showed that ATs underestimated steps. [31,32] For EE, six studies showed ATs overestimated kcals, [18,25,30,35-37] and 12 studies showed that ATs underestimated kcals [18,24-27,30,35-40] with variable precision and are most accurate for during locomotion and in lab-setting testing conditions [18,25-27,30,36,38,39] compared with non-locomotive activities [18,26,35,36,38,40] and free-living settings. [31,32,37] For activity minutes, one study reported, ATs overestimated MVPA in free-living settings, [32] and two studies reported, ATs underestimated MVPA in free-living settings. [31,33] For sedentary time, only one study has shown, ATs overestimated sedentary time and lack precision in free-living settings. [41]

Based on this evidence, we sought to expand our understanding of the accuracy and precision of ATs in estimating steps, EE, activity minutes and sedentary time in free-living settings using a validated direct observation (DO) system as the criterion measure.[42] Previous free-living studies employed accelerometers as a surrogate for gold-standard criterion measures (e.g. DO, doubly labeled water) to assess PA. [32-34,43-46] Limitations in using accelerometers as criterion measure to assess PA in free-living settings include 1) the inability to validate compliance (e.g. wear-time, wear-location) and 2) substantial variability in prediction equations used to convert accelerometer data into meaningful PA outcomes (e.g. moderate intensity activity, metabolic equivalents [METs]). [47-49] The use of DO as a criterion measure in free-living settings address these limitations and will attenuate the sources of error inherent in previous free-living studies. The evidence from this novel study will inform consumers, researchers, clinicians and interventionists about the utility of ATs as intervention tools and potentially, assessment tools for research. This dissertation addressed three knowledge gaps in assessing activity tracker performance. The first study addressed differences in ATs outputs compared to research-grade accelerometers in a tightly controlled environment. The second study validated consumer and research-grade activity monitors in estimating PA and ST compared to criterion measured PA and ST in free-living settings. The third study examined the ability of ATs to detect change in PA and ST in free-living settings. We also examined this question for commonly used research-grade devices.

Aims of Dissertation Studies

Study One: A Comparison of Consumer Activity Tracker Accelerometer Output and a Research-Grade Accelerometer Output During Orbital Shaking

The ActiGraph (AG)(ActiGraph, LLC, Pensacola, FL) accelerometer provides an objective estimate of human PA and is used in many research and clinical applications. [50-52] Standardized electronic validation and reliability testing of the AG have been performed on the GT3X+, GT3X, GT1M, 7164 and 71256 models. In these studies, electronic devices such as wheels, [53,54] a table, [55] and orbital shaking [56-58] were employed in controlled laboratory settings. In general, ActiGraph accelerometers are valid and reliable during electronic oscillation testing. Validity and reliability are improved with the low frequency extension filter at lower frequency oscillations (e.g. ≤ 0.6 Hz) and plateau beyond its bandpass filter limit of 2.5 Hz. [54,56,59-61]

The benefits of electronic orbital shaker testing are that it allows us to: (1) expose activity trackers (ATs) to different oscillation frequencies to simulate different movement intensities and (2) vary oscillation frequencies to simulate variation in free-living whole body acceleration. Orbital shaker testing removes human variation. As a result, observed differences would be due to technological features of the devices – not impacted by human variation. The electronic orbital shaker informed us of how ATs perform under highly controlled conditions.

Recently, our lab employed an electronic orbital shaker to assess the validity of several consumer ATs compared to the AG GT3X+ accelerometer (unpublished observations). [62] We found that AT output was highly correlated with oscillation frequency (r range: 0.92 to 0.99). Activity trackers output variables increased as oscillation frequency increased (p range: < 0.001 to 0.04).

The objective of this study was to examine estimates of EE and steps from commercially available consumer ATs, compared to the research-grade GT3X-BT accelerometer using an electronic orbital shaker as the standardized motion detector.

Therefore, the first dissertation study addressed the limitations in the current literature by exposing ATs to known frequencies and durations and comparing their output to research-grade accelerometer output.

1. **Specific Aim:** To compare consumer ATs with the research-grade ActiGraph™ GT3X-BT (GT3X-BT) accelerometer in estimating energy expenditure (EE) and steps during orbital shaking at different frequencies.
    a. **Hypothesis:** Energy expenditure and step estimates from consumer ATs will be similar to the EE and step estimates of the research grade GT3X-BT accelerometer during standardized testing using an electronic orbital shaker.

Study Two: Validation of Consumer and Research-Grade Monitors in Free-Living Settings

We evaluated the performance of consumer ATs in free-living settings using DO as the criterion measure for steps, EE, MET-minutes and time spent in different intensities of activity. [63-65] Our lab has validated DO in estimating PA and ST [42,65] using indirect calorimetry as the criterion measure.

Several studies have validated ATs in free-living settings, however, none have employed DO as the criterion measure for steps, EE, activity minutes or sedentary time. Therefore, the aim of this study was to validate AT estimates of steps, PA and ST in free-living settings compared to criterion measures.

2. **Specific Aim**: to determine the accuracy and precision of ATs in estimating steps, EE, activity minutes and sedentary time compared to direct observation-derived measures (criterion measures) in free-living settings. We also validated commonly used research-grade devices.

Study Three: Activity Trackers Sensitive to Change in Physical Activity and Sedentary Behaviors in Free-Living Settings

Tools such as wearable devices to track personal physical activity (PA) provide a mechanism to be more informed about activity behavior. Consumer devices that track PA behavior are increasingly popular for consumers and for researchers, clinicians and of interest to National Institutes of Health[66] who recognize the value of using sensor-based wearable monitors to assess PA behaviors. Currently, there are at least 149 active or recruiting clinical trials funded by NIH that are employing consumer ATs to measure (estimate) change in PA behaviors such as energy expenditure (EE) and/or steps. [67]

The research and clinical communities have rapidly adapted ATs, however, their utility within these communities has yet to be realized. Moreover, unlike research-grade devices that have been utilized by the research and clinical communities in the past, ATs have yet to undergo rigorous testing in both laboratory and free-living settings. In particular, there is no evidence examining the effectiveness of ATs for detecting change in PA behaviors in free-living settings. This knowledge gap is of major concern since ATs are widely used to monitor change in PA behaviors. Therefore, the aim of this study was to examine the ability of ATs to detect change in PA and ST in free-living settings. We also examined this question with research-grade accelerometers.

From study 2, we calculated criterion measured and AT estimated visit-to-visit change in steps, EE, activity minutes and sedentary time. The objective of this exploratory study was to examine AT estimates of change in comparison to the criterion measure estimates of change.

3. **Specific Aim one**: To examine the ability of ATs to detect change in PA and ST in free-living settings.
4. **Specific Aim two**: To examine the ability of research-grade accelerometers to detect change in PA and ST in free-living settings.

Significance of Dissertation Studies

Americans suffer from preventable chronic diseases such as heart disease, stroke, obesity and type 2 diabetes mellitus. [68] Current recommendations for PA and ST seek to use positive changes in these behaviors to improve chronic disease morbidity and mortality. These recommendations have led to increased public awareness of the importance of engaging in daily PA and the negative consequences of not engaging in daily PA. Tools such as, ATs to monitor PA behaviors are emerging as a valuable mechanism to be more informed about PA and ST.

Technological advancements such as, improved battery life, affordability and personalized feedback capabilities have helped lead the general public and researchers to use ATs as a PA behavior measurement instrument. However, unlike previous measurement instruments used by researchers, ATs have not been scrupulously tested for the validity of the estimates they provide in the natural environment where they are used. Several lab-based validation studies have been performed comparing activity tracker PA estimates (e.g. EE) to criterion measured PA (e.g. indirect calorimetry). To our knowledge, no studies validating activity tracker PA estimates compared to DO-criterion measured PA in free-living settings have been performed.

In two recent projects, our lab successfully employed lab-based protocols to 1) validate an AT in estimating EE compared to indirect calorimetry and 2) a DO system in estimating PA and ST. We expanded and integrated these two protocols to include ten of the most popular ATs currently on the market and compare their estimates of PA and ST to DO measured PA and ST in free-living settings. This study setting and criterion measure are superior to lab-based and comparison measures (e.g. accelerometer-based).

We directly observed participants while wearing ATs in their natural environment, which allowed us to capture and quantify PA, and ST where activity type and duration were not regulated. We chose an ecologically valid study setting and criterion measures, which advanced our understanding of AT performance under conditions in which they are used.

This information is beneficial to both the general public and research community. Providing the evidence of the accuracy and precision of ATs in estimating PA and ST improves the general public and researchers ability to make evidence-based decisions regarding selection of devices for their specific needs. Activity tracker estimates of PA and ST have been validated in free-living settings using research-grade accelerometers. However, validating ATs employing DO as a criterion measure for PA and ST in free-living settings is unexplored. A comprehensive understanding of activity tracker PA and ST estimates and associated errors are important for the general public and researchers seeking to understand the dose-response relationships between activity, ST and health.

CHAPTER 2
REVIEW OF THE LITERATURE

Accurate measurement of physical activity (PA) and sedentary time (ST) is important to improve our understanding of the dose-response relationship between these lifestyle behaviors and risk of numerous chronic diseases. The U.S. relies on large-scale surveillance studies (e.g. National Health and Nutrition Examination Survey [NHANES], Women' Health Study [WHS]) to quantify, analyze and interpret PA and ST. In part these data are used to 1) inform the public, 2) update existing and/or design new public health policies, 3) publish PA and ST statistics and recommendations, and 4) evaluate trends in PA over time.

The NHANES and WHS began using accelerometers in 2003 [69] and 2011, [70] respectively. For several decades accelerometers have been employed to objectively measure PA and are currently the device of choice for researchers. Accelerometers have been well received by the research community, as they are relatively low burden on participants and researchers. Advancements in technology have led to increased memory capacity, reduction in size of the devices, and improved filtering capabilities. Advancements in software and firmware provide greater user autonomy (e.g. initialization/download options) so that accelerometer sensors are easy to use in lab and field-based settings.

Advancements in accelerometer sensor technology coupled with the lower costs of accelerometer sensors led to the development and marketing of consumer-grade activity trackers (ATs). However, unlike research-grade accelerometers, ATs have yet to undergo rigorous and comprehensive testing to understand the benefits and limitations of the PA and ST estimates they produce. The evidence of ATs' accuracy and precision in estimating PA and ST is limited and there is no evidence about how well these consumer devices detect activity changes in these behaviors. This knowledge gap is of major concern since these devices are often used to monitor PA and ST improvements (i.e. detect change). This review of the literature will describe the main areas of research that were addressed in this dissertation. First, describing what is known about the accuracy and precision in estimating ST and PA from research-grade accelerometers. Second, describing and analyzing what is known about consumer ATs regarding accuracy and precision. Lastly, presenting current knowledge of ATs and detection of change in PA behaviors.

Study One: A Comparison of Consumer Activity Tracker Accelerometer Output and a Research-Grade Accelerometer Output During Orbital Shaking

Calibration of Research-Grade Monitors

Since the early 1930s, accelerometers have been employed to assess PA parameters such as gait [71] and whole body movement. [72-74] Originally, accelerometers were used to estimate steps, [75] energy expenditure (EE) [76] and determining external mechanical work during locomotion. [73] These and other initial studies demonstrated the capacity of accelerometers to objectively estimate PA, giving rise to the first generation (in the1980's) of accelerometers, which were developed to estimate PA and EE.

Laboratory Studies

Calibration and validation. Accelerometer calibration and validation studies have been performed in laboratory and free-living settings. A strength of a laboratory setting is that it allows easy replication of experimental protocols.

There are limitations to this method for testing accelerometers. Evaluating accelerometers in estimating PA EE in a laboratory is scripted and structured in comparison to PA behavior in free-living settings where behaviors are random, sporadic, and variable. As a result, laboratory-based study results do not directly translate to study results obtained from free-living settings. Unlike laboratory settings, free-living settings allow researchers to capture and measure "real-world" PA behavior. Another strength is enhanced generalizability over laboratory-based studies.

Unit Calibration of Wearable Accelerometers: Machine Testing.

Unit calibration of accelerometers is performed by comparing the direct acceleration signals to a "gold standard." Typically, this is accomplished by spinning the accelerometer in an electronic oscillator with a known radius and frequency (RPM), intra- and inter-unit variability and can be determined and also one can verify that values are within the manufacturer's stated tolerance limits. [63]

Several groups have calibrated accelerometers using electronic methods. In 1987, Bassey et al.[75] employed an electronic turntable to test the stability, range (e.g. threshold accelerations) and reproducibility of the Yamasa Digiwalker when exposed to different acceleration frequencies. They reported an acceleration threshold below which the Digiwalker does not respond. Sensitivity increases linearly and rapidly (1-4 m/s) until reaching a plateau response. Next, Brage et al. (2003) [54] employed an electronic wheel to examine the intra- and inter-instrument reliability; influence of movement frequency and filtering on validity of the ActiGraph (AG) 7164 accelerometer when exposed to varying radii (22.0, 35.5, 49.0 mm), oscillation frequencies (Range 0.5-4.0 Hz) and oscillation frequency increments (0.25 Hz and 0.125 Hz). They reported large relative variability at very low and very high oscillation frequencies. Mean intra-instrument coefficient of variation, which is a measure of variability, was 4.4% for all units in all trials. Excluding two lowest frequencies, max intra-instrument coefficient of variation was 18%. Detection of changes in oscillation frequencies varied between units, with larger errors at the lowest oscillation frequencies, and for each frequency and across radius settings between unit correlation coefficients ranged from 0.92-1.00. Lastly, in 2008, Rothney, M.P. et al. [56] employed an electronic oscillator to characterize dynamic responses and inter-monitor and inter-generational variability of several AG accelerometer models (7164, 71256, and GT1M) when exposed to varying radii at a constant frequency (150 rpm) and varying frequencies with a fixed radius (46.6 mm). A linear relationship between counts and radius for all measured values, all generations were significantly different from each other at frequencies >160 rpm. For example, at the lowest frequencies the 7164 and 71256 responded similarly but GT1M required greater accelerations to detect changes, suggesting differences in sensitivity or filtering approach used in different models.

From these studies, and others, [54,56,59-61] it can be concluded that these accelerometers are valid and reliable based on electronic oscillation testing. Validity and reliability are higher if a low frequency extension filter is used at lower frequency oscillations (e.g. ≤ 0.6 Hz) and does not continue to increase beyond its bandpass filter limit of 2.5 Hz.

Value Calibration of Wearable Accelerometers: Human Studies

Value calibration of wearable accelerometers is described as converting accelerometer signals into estimates of EE, time spent in various intensity categories, and/or activity type while simultaneously collecting criterion data (e.g. indirect calorimetry). [63] Several accelerometers have been developed and calibrated for research on quantifying PA and EE. Examples include, the AG,[77-81] Tritrac,[82] Actical,[78,83] and the GENEA [55] accelerometers. Two laboratory based and two free-living setting calibration studies laid the groundwork for subsequent accelerometer research, discovery and development.

First, in 1983, Montoye et al. [76] examined if the waist-worn Caltrac accelerometer and the Large-Scale Integrated Motor Activity Monitors, 'LSI' mounted at the waist and wrist could estimate oxygen consumption during various activities, including locomotion. It was reported that the standard error of estimate for the EE algorithm used in the Caltrac was ± 6.6 ml/kg/min. Further, the reproducibility of the waist-worn Caltrac output during locomotion and various activities was high (r=0.94). [76] These findings demonstrated the ability of a waist-worn device to estimate EE during specific activities in a lab setting.

In a study of another accelerometer sensor, Freedson et al. [77] estimated PA intensity categories and EE from treadmill walking and running. The criterion measure was indirect calorimetry and the data revealed a linear relationship (r=0.88) between counts per minute from the accelerometer and EE (METs). A linear regression model was developed to predict point estimates of EE. Accelerometer count cut-points were also created to classify PA as light (< 3 METs), moderate (3-5.99 METs), vigorous (6-8.99 METs) and very vigorous ≤ 9 METs). Though this linear regression model was built from only controlled treadmill walking and running, the simple regression remains a primary tool to translate activity counts from an AG accelerometer into minutes of activity in different absolute intensity levels and EE.

Laboratory based calibration studies were a good first step, but to advance this knowledge base, accelerometers were also tested in free-living settings. Several free-living accelerometer studies have been conducted and this review will highlight two studies executed by Pfeiffer et al. [83] and Pate et al. [84] Both investigators sought to calibrate and cross-validate accelerometers in estimating PA for use with 3-5 year old children compared to indirect calorimetry (criterion). Both studies employed structured and unstructured sedentary (e.g. sitting), locomotive (e.g. overgound brisk walk) and non-locomotive (e.g. sports/play) activities while simultaneously collecting metabolic data. Pfeiffer et al. employed a right hip-worn Actical accelerometer and Pate et al. employed a right hip-worn AG 7164 accelerometer. It was reported that the Actical and AG counts strongly correlated with the criterion EE (r=0.89 and r=0.82, respectively). Cut-points for the Actical and AG were established for moderate intensity activity (20 mL/kg/min), 715 counts/15 seconds and 420 counts/15 seconds, respectively. Cut-points for vigorous intensity activity (30 mL/kg/min), were 1411 counts/15 seconds and 842 counts/15 seconds for the Actical and AG, respectively. Cross-validation of structured to unstructured activities revealed that both the Actical and AG 7164 accelerometers are valid and appropriate tools for measuring PA in young children. These data demonstrate the ability of accelerometers to estimate PA energy expenditure over a broad range of activities (e.g. sedentary to vigorous) in children, and underscores device-specific differences in absolute count values even though the EE is the same.

Many other accelerometer calibration studies that included a variety of sedentary/lifestyle activities, [8,78-80] locomotion [8,78,81,82,85] and/or sports, [8,78,85] have been published. The evidence indicates that an accelerometer worn on the hip or wrist is a good tool to estimate features about PA and EE in children and adults.

Algorithms to Quantify Physical Activity Behaviors

Originally, simple regression models to quantify PA intensity were constructed using accelerometer counts to generate cut-points, such as, sedentary, moderate and vigorous intensity using indirect calorimetry as the criterion measure of intensity. [77,52,78,84,86-91] This was an important first step, however, a single regression cannot accurately estimate EE across a wide range of activities and intensities. [92] For example, the Freedson model was derived from the count-EE relationship during treadmill walking and running. Thus, this model may under- or over-estimate EE for non-locomotive activities and/or free-living PA. In 2000, Hendelman et al.[79] applied a linear regression model developed from locomotion activities to a data set of locomotive and non-locomotive activities yielded a modest relationship between hip monitor counts and EE of r = 0.59. These results led researchers to include non-locomotion activities in model development [49,80,93,94] and use of additional accelerometers positioned at various wear-locations. [80,95-97] In general, the addition of output from several wear locations (e.g. hip- and wrist worn accelerometer) into models improved EE estimations. A two-regression model using the standard deviation of counts/min to identify the appropriate regression model to predict EE from accelerometers have also been developed. [98-101] These methods improved EE estimates across a wider range of activity types and intensities and led to advanced techniques (e.g. pattern recognition) for conducting accelerometer value calibration. Employing machine-learning pattern recognition techniques uses the activity counts [94,102-104] or raw acceleration patterns [13,105] within the accelerometer signal to estimate activity type and intensity. Signal features (e.g. time- and frequency-domain) are used to predict PA measures. For example, Staudenmayer et al. [9] developed pattern recognition methods to estimate PA energy expenditure and activity type during a wide range of activity intensities and activity types in a lab-based setting using the AG 7164. They reported that a neural network pattern recognition prediction of METs root mean squared error was 1.22 METs and correctly classified activity type 88.8% of the time. This method was an improvement over previous methods for estimating EE and activity type. Recently, Lyden et al. [102] broadened the scope of machine-learning by applying it to a free-living setting. The Sojourn-1 Axis (soj-1x) and Sojourn-3 Axis (soj-3x) were shown to be more accurate at estimates of MET-hours (soj-1x: % bias = 1.9 [-2.0 to 5.9], root-mean-squared error (RMSE) = 1.0 [0.6 to 1.3]; soj-3x: % bias = 3.4 [0.0 to 6.7], RMSE = 1.0 [0.6 to 1.5]) and activity minutes (soj-1x: % bias = 8.8 (sedentary), -18.5 (light), and -1.0 (MVPA); soj-3x: % bias = 0.5 (sedentary), -0.8 (light), and -1.0 (MVPA) compared to previous methods. These and other studies employing pattern recognition/ machine-leaning techniques such as, Hidden Markov Methods, [106] artificial neural networks (ANNs), [11-13,103,104,107] and support vector machines [108,109] are superior compared to simple linear regression modeling and offer the advantage of identifying activity type in addition to activity intensity.

Study Two: Validation of Consumer and Research-Grade Activity Monitors in Free-Living Settings

Validation of Research-Grade Monitors

The purpose of validating accelerometers against gold standard methods is to evaluate the accuracy and precision in estimating the specific outcome(s) such as steps, EE, activity intensity and activity type. Gold standard methods for EE include direct and indirect calorimetry, doubly labeled water (DLW) and DO. Direct observation is also the gold standard method for measuring steps. Validation studies of accelerometers are device, population, protocol and outcome specific. For example, the AG GT3X+ accelerometer has been shown to be valid in estimating minutes spent in MVPA during treadmill walking/running in a group of men and women ages 21 to 39 years may not be valid for minutes of MVPA in free-living older adults.

Laboratory Studies

Validation studies date back to the early 1980s when Montoye et al.[76] tested the Caltrac accelerometer for estimating EE compared to indirect calorimetry in a laboratory setting. In 1995, Melanson et al.[97] conducted a validation study of the Computer Science and Applications, Inc. (CSA) accelerometer in assessing PA during treadmill walking and running at varying grades compared to indirect calorimetry. The CSAs were worn on the hip, wrist and ankle. The most accurate prediction of EE was obtained when body mass and CSA ankle, hip, and wrist activity counts were used as predictors. This model predicted mean EE within 1%, but had a relatively large SEE of 0.85 kcals per min (11.4%). The CSA counts from any location were significantly correlated with EE (*r*=0.77-0.89). The main findings of Montoye et al. and Melanson et al., were that on average accelerometer(s), 1) estimated EE was highly correlated with speed and 2) underestimated EE during graded treadmill exercise. These results were confirmed by Nichols et al. [82] who validated the Tritrac accelerometer in estimating EE compared to indirect calorimetery during treadmill walking and running. The Tritrac was found to be highly correlated with speed ($r=0.97$, $p<0.0001$), the relationship between vector magnitude and EE across all speeds (1.9, 3.9, and 6.0 mph) was highly linear ($R^2 = 0.90$, SEE = 0.014 kcal/kg/min), and underestimated EE at 5% grade (Mean difference at 6.4km/h=-0.0107 kcal/kg/min). These studies and others [52,110,111] supported that generally, accelerometers correlated with criterion EE and activity type during locomotion.

The relationship between EE and counts during non-locomotive activities such as activities of daily living (ADLs) and cycling is less linear and more variable. For example, during non-weight bearing exercise a waist-worn accelerometer underestimates EE. Puyau et al. [86] employed a room calorimeter to validate accelerometers during locomotion, sedentary behaviors, ADLs and sport. Dissociation between EE and accelerometer counts was observed during weight lifting and stair climbing as well. Hickey at al.[112] compared step output from several different research-grade accelerometers during ST, locomotion and ADLs compared to manually counted steps (DO). The largest errors reported were during ADLs (mean difference range: -178 to 78 steps/5-minutes) and the highest accuracy was during rhythmic/ anterior-posterior movements (percent error range: 0.2 to 15.0%) compared to non-rhythmic movements (percent error range: 6.5 to 78.0%). These and other lab-based studies, [113-116] showed that in general during rhythmic locomotion accelerometers were valid in estimating EE and activity type, and that the relationship between EE and accelerometer counts is influenced by factors such as intensity and activity type.

Device Location

Device location influences estimates of PA and ST. There are differences in output from hip and wrist locations depending on activity type and intensity and environment (i.e. lab-based or free-living). Generally, hip-worn accelerometers underestimate EE during non-weight bearing activities (e.g. cycling) and graded locomotion (e.g. ascending stairs). The wrist-worn accelerometers tend to overestimate EE during overground walking, some ADLs (e.g. vacuuming) and SB (e.g. computer work). Trost el al.[105] employed the AG GT3X+ to develop an activity recognition algorithm and compared rates of activity classifiers trained on the raw triaxial acceleration signal collected from accelerometers worn on the wrist and hip. They reported that wrist-worn accelerometers misclassify upright, non-ambulatory activities with significant arm movement (e.g. sweeping the floor) compared to hip-worn accelerometers. Several other investigators also reported differences in hip and wrist output. For example, McMinn et al.[114] reported that both EE and steps were different between the hip and wrist worn AG (GT3X+) accelerometer during self-selected treadmill walking and running compared to indirect calorimetry and manually counted steps. For example, the mean difference between GT3X+ steps for hip and wrist locations for the medium and fast walk were 2 and 7, and 1 and 14 steps, respectively. The authors concluded that wrist-mounted device outputs were not comparable to waist-mounted outputs. Mahar et al. [117] examined output from hip and wrist worn GT3X+ from treadmill walking and running and 2-days of free-living time. They reported minutes of moderate (hip: 46 ± 21 min; wrist 143 ± 51 min; $r = .52$) and vigorous (hip: 4 ± 6; wrist 16 ± 14 min; $r = .83$) PA were higher ($p < .05$) for the wrist worn than for the hip worn monitors. Later, Hildebrand et al.[118] found significantly higher output from wrist monitors than hip observed for children and adults during treadmill and simulated free-living activities.

Free-Living Studies

Validation studies of accelerometers in estimating EE and activity type in free-living settings are integral to building a comprehensive knowledge base of accelerometer accuracy and precision. Free-living accelerometer validation studies have relied upon indirect calorimetry, DLW or DO as criterion measures for EE and each of these criterion measures have its limitations. Limitations include 1) indirect calorimetry is expensive and impedes numerous free-living activities, 2) DLW is expensive and only provides a measure of total EE and 3) DO is highly dependent on observer training and requires observer judgment of intensity. Because of these limitations, it is imperative that researchers choose the appropriate criterion measure for validating wearable accelerometers in free-living settings. For example, if quantification of MET-minutes of activity is required, DLW is not the appropriate criterion measure. However, DLW has been used extensively as a criterion measure of EE for validating wearable accelerometer estimates of EE in free-living settings. The DLW technique is an isotope-based method that measures the EE of subjects based on the difference in enrichments of 2 isotopes: hydrogen and oxygen. [119] In 1991, Heyman et al.[120] validated the Caltrac activity monitor compared to DLW in estimating EE in free-living young adult men over 10-days. They found that though the total EE estimates from the Caltrac were strongly correlated with DLW (r=0.87, p<0.05), it underestimated total EE by %22 (r=0.87, p<0.001). These results were not surprising as all of the participants had full time sedentary jobs and the Caltrac was worn on the waist. Thus, most upper body movements and associated energy costs performed while seated could not be detected. In an effort to capture upper body movement and more accurately estimate EE, devices such as the Sensewear Armband (BodyMedia L.L.C., Pittsburgh, PA) have been designed to be worn on the upper arm.

The Sensewear Armband is worn over the left tricep, and integrates motion data from a triaxial accelerometer along with several other physiological sensors (heat flux, skin temperature, and galvanic skin response). These data are applied to proprietary algorithms to estimate EE. Free-living validation studies of the Sensewear Armband compared to DLW have shown that in youth a total error 44 kcals/day and mean absolute percent error (MAPE)=11.7%, [121] in adults a total error 22 kcals/day and MAPE=8.3%,[122] and in older adults a total error -21.5 kcals/day and MAPE=8.0%.[123] The MAPE is a common metric used by researchers to allow for comparisons of error between monitors and should be approached with caution, as the MAPE provides no information pertaining to device bias or individual errors. Instead, the MAPE indicates the absolute, average group error. These are only a few examples to illustrate the use of DLW as a criterion measure for EE in free-living settings, and though considered a “gold standard” many limitations are noted. For example, the high cost of isotopes (e.g.$_{18}$O) sum to about $1,000 to 1,500 per subject and sophisticated equipment is required for analysis. More importantly, the DLW method does not allow for quantification of minutes of MVPA, PA bouts (i.e. ≥ 10 min) and steps. Of course, in science, the methods are driven by the question(s) being asked. Thus, in some cases DLW should be the criterion measure of choice, in other cases, alternative criterion measures such as DO should be the method of choice.

Direct Observation

Direct observation as the criterion in free-living validation studies of accelerometers in estimating PA energy expenditure date back to the mid 1980s. Klesges et al.[124] were the first to employ DO as a criterion measure to validate the Caltrac accelerometer in free-living adults (N=50) and preschoolers (N=30). Another aim was to compare the Caltrac to the then widely used, Large Scale Integrated Moving Activity Counter (LSI). The LSI houses a ball of mercury with a mercury switch that registers an internal counter when exposed to a 3% incline or decline. The Caltrac, uses a piezoelectric accelerometer that measures vertical dynamic changes in accelerations and converts them to voltages. Briefly, participants were observed for 1-hour using focal sampling (10-seconds observed, 10-seconds record) and activity type (e.g. sitting, walking, running) and intensity (e.g. minimal, moderate, extreme) were recorded. The DO training included rigorous quality control and reliability assessments to insure that the observational data were collected accurately. For example, a trained observer designation required inter-rater correlations of at least r = 0.90. By the end of observer training period inter-rater reliability was 97%. For adults, it was reported that, on average, the accelerometer was strongly correlated with DO for activity type (r=0.70, p<0.001) and intensity (r=0.76, p<0.001) and with the LSI (r=0.83, p<0.001). For preschoolers on average, the accelerometer was moderately correlated with DO for activity type (r=0.39, p<0.05) and the LSI (r=0.42, p<0.001) but weakly correlated with intensity (r=0.25). The inability of either device to estimate PA levels of preschoolers was attributed in part to not adequately detecting and quantifying "short burst" activities. These data provided the first evidence that the Caltrac accelerometer is a valid tool for estimating activity levels in adults but not preschoolers in free-living settings.

Recently, Lyden et al. [42] were the first to validate DO as a criterion in estimating PA and ST compared to indirect calorimetry. Briefly, participants were observed for three, 2-hour sessions in the laboratory while engaging in sedentary (reading, writing, computer use) and PA behaviors (walking, treadmill use, cycling) while simultaneously collecting metabolic data. Though this was a laboratory setting, participants' behaviors were designed to resemble the free-living nature of behaviors. Behaviors (activities) were observed and recorded by a trained researcher. A hand-held personal digital assistant (PDA) with custom software (The Observer, Noldus Inc., Wageningen, Netherlands) was used to record participant behavior (e.g. activity type and associated MET value). They reported that DO accurately and precisely estimated MET-hours [% bias (95% Cl) =-12.7% (-16.4, -7.3), ICC = 0.98], time in low intensity activity [% bias (95% Cl) = 2.1% (1.1, 3.2), ICC = 1.00] and time in moderate to vigorous intensity activity [% bias (95% Cl) -4.9% (-7.4, -2.5), ICC = 1.00]. This study provided the first evidence to support the use of DO as a criterion for PA and ST in free-living settings.

The same DO system validated by Lyden et al.[42] was employed as the criterion in several validation studies of accelerometers in free-living settings. First, Kozey et al.[65] employed DO to validate the activPAL (AP; Physical Activity Technologies, Glasgow, Scotland) and the AG (GT3X) in estimating ST in free-living settings. Participants were observed for two (1, normal behavior; 1, less sitting) 6-hour sessions while wearing the AP on mid-thigh of the right leg and the GT3X on the right hip. They noted that the correlation between the AP and DO was $R^2 = 0.94$, and the AG 100 counts per minute threshold for sedentary and DO sedentary minutes was $R^2 = 0.39$. Only the AP was able to detect reductions in sitting time. The ActiGraph 150-counts-per-minute threshold demonstrated the lowest bias (1.8%) of the ActiGraph cut-points. These data provided the evidence that the activPAL is a valid tool for quantifying ST and detecting change in ST in free-living settings. Next, Lyden et al.[125] extended this work by using DO to validate the AP and AG (GT3X) in estimating breaks from sedentary behaviors, absolute number of breaks and break-rate in free-living settings. Participants were observed for two, 10-hour conditions (1, normal daily activity; 1, reduced and intermittent sedentary time). The AP produced valid estimates of all ST measures and was sensitive to changes in break-rate between conditions (baseline: 5.1 [2.8 to 7.1] brks.sed-hr−1, treatment: 8.0 [5.8 to 10.2] brks.sed-hr−1). Additionally, the GT3X was not accurate in estimating break-rate or absolute number of breaks and were not sensitive to changes between conditions. These results further support the utility of expressing break-rate from sedentary time as a metric specifically relevant to free-living behavior, and that the AP is a valid tool for measuring ST in free-living environments. Lastly, Lyden et al.[102] compared artificial neural network (ANN) techniques “sojourn methods” estimates of active and ST from a waist-worn GT3X to DO in free-living settings. As previously discussed, both novel ANNs (soj-1x and soj-3x) improved the accuracy and precision in estimating free-living MET-hours, sedentary time, and time spent in light-intensity activity and MVPA compared to a previously developed [9] and validated [7] ANN method. Further, the soj-3x was found to be the superior method for differentiating ST from light-intensity activity. Together, these studies 1) serve as the foundation that DO is a valid criterion for estimating PA and sedentary time in free-living settings, and 2) further support and underscore the utility of wearable accelerometers' in estimating PA and sedentary time in free-living settings.

Summary

Both unit machine oscillation calibration studies and several human studies have shown that accelerometers are valid and reliable in estimating features of activity and SB. [53,55,57,75] In addition, the relationship between EE and counts per minute is not linear for a wide range of activity types and intensities. As a result, a single regression model will not sufficiently estimate EE and other measures of activity and ST. Also, accurate detection of high intensity activity accelerometer signals plateau (~6.2 to 7.4 mph, ~10,000 counts per minute). [59,81] There is a rich set of signal features within the accelerometer that is captured but not analyzed. The detailed features of the signal are used with pattern recognition techniques for estimating PA energy expenditure and activity type. These techniques improve estimates by utilizing more signal information, such as time and or frequency domain features.

The differences in output by monitor location is significant as large surveillance studies have transitioned from hip-worn to wrist-worn accelerometers in an effort to increase compliance. For example, the National Health and Nutrition Examination Survey (NHANES) 2011-2014 data collection cycle has implemented a new protocol where wrist-worn accelerometers are being used for objective assessment of PA. This protocol was initiated based on evidence of increased adherence to monitor use [126] and for the measurement of sleep. Unfortunately, compliance is increased at the expense of data analysis. To date, there is no uniform decision from the PA community of how to analyze the data. Recently, Staudenmayer et al. addressed this issue. Specifically, wrist-worn accelerometer data were used to build machine-learning and regression models that estimated 1) MET-hours, 2) time in different activity intensity categories (light, moderate, and vigorous), 3) the amount of time the wearer is sedentary or not, and 4) the amount of time the wearer is locomoting or not. The wrist models estimated METs with a RMSE of 1.21 METs, and classified: activity intensity 75% correct, sedentary time 96% correct, and locomotion time 99% correct. These methods offer a validated technique with which to analyze NHANES accelerometer data.

Over the past several decades significant advances have been made toward a comprehensive understanding of the strengths and limitations accelerometers' possess in estimating PA and sedentary time. The advances in objective measurements of PA and ST have provided a blueprint of how to 1) ask poignant scientific questions related to PA and health, 2) design and execute meaningful accelerometer experiments, 3) develop simple and complex methods to analyze accelerometer data, 4) identify limitations of accelerometer data and suggest methods to for improvement, and 5) provide evidence of the relationships of accelerometer-based activity and sedentary time estimates for quantification of dose of exposures of these behaviors and health outcomes

Activity Trackers: Introduction

In contrast to research-grade accelerometers, ATs have largely bypassed rigorous, scientific testing and proceeded directly to the consumer market.

Validation of four monitor output variables have been reported by studies that have tested the accuracy of activity trackers: step counts, EE, activity minutes (analogous to MVPA) and sedentary time. The following section of this literature review will summarize the current state of the evidence regarding the validity of ATs in estimating each of these output variables.

Producers of ATs have promoted self-monitoring PA and ST by providing output to users that allow self-tracking and inform users about successful achievement of U.S. PA recommendations and/or Australian SB recommendations. For example, Fitbit provides output in "active minutes." According to Fitbit, active minutes are defined by activities at or above about 3 METs. To satisfy the Center for Disease Control's "10 minutes at a time is fine" concept, minutes are only awarded after 10 minutes of continuous moderate-to-intense activity. [127] Given the importance and public awareness of meeting PA and SB recommendations, steps, EE, activity minutes and sedentary time are a critical metrics to provide users of wearable devices

Researchers are testing the relationship between AT output with criterion and/or comparison measures and the accuracy of ATs in estimating steps, EE, activity minutes and sedentary time compared to criterion/comparison measures. Preliminary results from this body of research reveal moderate to strong relationships between AT estimates of steps, EE and activity minutes in both lab and free-living settings (range: r= .52 to .99). However, many AT estimates of PA and ST lack accuracy and precision.

In this review, ATs will be identified by location of wear. For example, hip-worn AT. Tables 1 and 2 summarize the results from AT validation studies. Activity trackers with corresponding output and data extraction method and features are provided in Table 4 and Appendix E.

Validation of Activity Trackers

Laboratory Studies

Steps

In general, ATs are accurate and precise in step estimates during locomotion. Not surprisingly, differences in step estimations between lab and free-living settings and, the hip and the wrist location have been reported.

To date, thirteen studies have validated ATs in estimating steps in lab-based settings. All but three of these studies employed DO as the criterion measure for steps. From these studies, two protocol trends have emerged: simulated free-living and locomotion only.

Simulated free-living. Simulated free-living protocols include long and short durations engaged in varying activity types and intensities, either or both self-selected and predetermined. A total of four studies have been published validating ATs in estimating steps in lab settings. All have employed DO (step counting) as the criterion measure for steps.

Chen et al. [20] validated wrist-worn ATs during locomotive and non-locomotive (e.g. ADL's) activities. They reported an absolute percent error (APE) ranging from

1.5% to 9.6% during treadmill walking and running. In addition, accuracy was improved at faster speeds (4.9 mph) for all the monitors (APE < 2.5%). Mean bias (±SD) for locomotive activities ranged from -13.5 (±2) to -35.4 (±2.4) steps. For non-locomotive activities, mean bias (±SD) ranged from 2.9 (±45.5) to -65.9 (±25.8) steps and significant differences between dominant and non-dominant were also reported. Mammen et al. [17] validated hip-, pocket- and collar-worn ATs during overground walking (20 steps), treadmill walking and running and while driving. They reported that all ATs estimated steps within ±5%, only one AT (pedometer) detected steps while driving, and statistically significant differences were found between the criterion and steps detected by two hip-worn ATs ($p<0.001$, $p<0.05$) at 1.2 mph and by a hip-worn AT at 1.8 mph ($p<0.05$). All ATs were accurate at normal walking speeds (2.7 and 3.7 mph). Nelson et al. [18] examined the accuracy of hip- and wrist-worn ATs in estimating steps for specific activities and activity categories. Results showed that for the household activity category, MAPE ranged from 54% to 79%. In contrast, for the ambulatory category, MAPE ranged from 3% to 6%. For walking and jogging, hip- and wrist worn ATs produced MAPEs of 2% to 3%, and 8% to 11%, respectively. For cycling, all ATs displayed large MAPEs ranging from 70% to 93%.

Differences in AT step estimates during non-locomotive activities have also been reported by Stackpool et al. [26] Employing a protocol that included self-selected walking and running and sports. They found that during locomotion, ATs were accurate within 10% of total steps, and collectively they averaged 4% underestimation. During sport activities, the errors in step counts were systematically less than the criterion measure, ranging from 3-24%, and averaging 18%.

In summary, on average, hip-worn ATs are more accurate and precise in estimating steps as compared to wrist-worn ATs during simulated free-living activities/behaviors. Especially during locomotion, hip-worn ATs produce errors that are within 5% of criterion measured, which is within the generally acceptable range of less than 5%. [128-131] As expected, this trend has been consistent throughout the literature.

Locomotion only. Twelve studies have validated ATs in estimating steps during locomotion. Of these, nine have employed DO as the criterion for steps and three have employed secondary measures as the step criterion. Regardless of which step criterion was employed, findings from all studies are in agreement.

In general, ATs are accurate and precise in estimating steps during locomotion. However, differences between hip- worn and wrist-worn ATs exist. Case et al. [29] evaluated the step count accuracy of hip- and wrist-worn ATs during treadmill walking at 3.0 mph for 500 and 1500 steps. Compared with DO, the relative difference in mean step count ranged from −0.3% to 1.0% for hip-worn ATs, and −22.7% to −1.5% for wrist-worn ATs. Storm et al. [22] tested the accuracy of hip- and wrist-worn ATs in estimating steps during indoor and outdoor walking and descending and ascending stairs. They reported step underestimations of -15±18 (MAPE=1.6±1) by hip-worn ATs and -253±156 (MAPE=24±14) by wrist-worn ATs. Several other groups have reported similar findings from wrist-worn AT estimates during locomotion. [21,22,24,25,30,31]

Diaz et al. [25] sought to validate hip- and wrist-worn ATs in estimating steps during treadmill walking and jogging. They found that the wrist-worn AT significantly underestimated steps. Mean differences ranged from -26 to -3 steps. No significant differences in step estimates were observed between the hip-worn AT and the criterion. Recently, Diaz et al. [30] employed the same treadmill protocol to validate hip-wrist- and bra-worn ATs. They reported that the wrist-worn AT significantly underestimated steps (mean bias: -11 steps) for all treadmill walking and running speeds (range: 1.9 to 5.2 mph), and produced large errors ranging from 16±28% to 2±6%. In contrast, percent error for step estimates for the hip- and bra-worn ATs were ≤3% for all walking and running speeds. In both studies, the step estimates from the hip-worn AT was more accurate (e.g. mean difference range: -0.7 to 2.0 steps) and precise (e.g. mean percent error = -0.8±2.0), as compared to the wrist-worn AT (e.g. mean difference range: -15.5 to 3.4 steps; mean percent error -4.0±15.2).

At slower walking speeds (e.g. ≤ 1.9 mph), hip-worn ATs have been shown to produce relative errors as great as 40% with large variation. For example, Beevi et al. [28] evaluate the step count accuracy of hip-worn ATs during slow walking at 0.6, 1.2, and 1.8 mph. All ATs underestimated steps. Mean biases (±SD) ranged from -37.5 (±16.1) to -52.0 (±26.6), and the error rate of all ATs decreased with the increase of speed.

These data provide evidence that in general: 1) hip-worn ATs are accurate and precise in estimating steps during locomotion that is ≥ 1.9 mph 2) wrist-worn ATs significantly underestimate steps during locomotion, 3) differences are market at slower speeds (e.g. 3.0 mph), and 4) hip-worn ATs estimates of steps are more precise (less variable) as compared to wrist-worn AT estimates of steps. ATs are less valid in estimating steps is in free-living settings as compared to lab settings.

Energy Expenditure

A total of twelve validation studies evaluating AT estimates of EE have been performed using one either a room calorimeter or breath-by-breath metabolic measurement systems as the criterion measure.

Criterion: room calorimetry. Two studies have evaluated AT estimates of EE compared to room calorimeter. Both study protocols included locomotion, lifestyle activities and ST; one included cycling. Dannecker et al. [40] tested the accuracy of a hip-worn AT during 4-hours of data collection. Briefly, participants performed a series of randomly assigned postures (e.g. sitting) and activities (e.g. treadmill walking) for 3-hours. The last hour of data collection consisted of self-selected free-living activities. They found that the hip-worn AT significantly underestimated EE by 143.2 kcal compared criterion kcals. The large underestimation may have resulted from activities with movement not detected by the hip-worn ATs such as cycling and computer work. Recently, Murakami et al. [37] extended this idea to include more time in the room calorimeter, more ATs and meals. They sought to validate ten ATs worn at various locations including the wrist, waist and pocket in estimating EE over 24-hours. For this study, participants completed a standardized protocol simulating normal daily life, which included 3 meals, deskwork, watching TV, housework, treadmill walking, and sleeping. Results showed that despite strong correlations with criterion measured kcals (rho range= .71 to .93). Three wrist-worn ATs significantly underestimated kcals ranging from -278 to -249 kcals. A waist- and a pocket-worn AT significantly overestimated kcals by 175 and 205 kcals respectively.

Criterion: indirect calorimetry. A total of ten studies have been published validating ATs estimate of EE compared to indirect calorimetry. From these studies, two protocol trends have emerged: simulated free-living and locomotion only.

Simulated free-living. Simulated free-living protocols include long and short durations engaged in varying activity types and intensities, either or both self-selected and predetermined. Lee et al. published the first large-scale (N=60) validation study of ATs in 2014. [35] The purpose of this study was to examine the validity of EE estimates from a variety of ATs (chest-, hip-and wrist-worn) under free-living conditions. To simulate free-living conditions, the protocol consisted of 13 different activities and SBs that were performed for 5-minutes each (3-minutes each for treadmill activities) for a total of 69-minutes. Total EE estimated from the ATs was compared to criterion EE. The results showed differences between ATs and AT location. The MAPEs for hip-worn ATs were 10.1% and 10.4%, wrist-worn ATs ranged from 12.2% to 23.5% and 12.8% for chest-worn ATs. Mean biases for hip-worn ATs were -26.0 and 13.2, wrist-worn ATs ranged from -85.5 to -6.7 and -23.1 kcals/69-minutes for chest-worn ATs. It was concluded that the majority of the ATs yielded reasonably accurate estimates of EE compared with the criterion values (i.e. within 10% – 15% error).

In 2016, Bai et al. [36] conducted a validation study of ATs during large time-blocks of activities. For this study, participants performed semi-structured periods (25 minutes each) of self-selected sedentary activity, aerobic exercise, and resistance exercise while wearing several wrist monitors for a total of 80-minutes. Mean absolute percent error (all activities) ranged from 16.8% to 30.4%. Mean biases (SD) ranged from -72.4 (87.2) to 42.3 (55.1) kcals/80-minutes.

Three studies have validated ATs in estimating EE during simulated free-living activities such as locomotion, sports, lifestyle and SBs. First, Sasaki et al. [38] validated a hip-worn AT in estimating EE compared to criterion measured EE and found that the mean bias across all activities was -4.5 ± 1.0 kcals/6-min. with 95% limits of agreement (LOA) ranging from -25.2 to 15.8 kcals/6-min. Also, the hip-worn AT significantly underestimated EE during household activities and graded locomotion. Differences in estimates by activity were also reported by Nelson et al. [132] whom examined the accuracy of hip- and wrist-worn ATs in estimating EE for specific activities and activity categories. They reported that all ATs predicted EE within 8% of criterion measured EE for sedentary activity but overestimated activity EE by 16%–40% during ambulatory activity. Similar to the findings of Sasaki et al., all ATs significantly underestimated EE for cycling by 37%– 59% (p=0.025–<0.001). Lastly, for all activity categories (sedentary, household, and ambulatory), all ATs displayed high MAPE (>10% of criterion) for EE estimation, ranging from 13% to 35%. In agreement with other studies, overall EE estimates may be interpreted differently if analyzed by activity type. Differences in hip-worn AT energy expenditure estimates during non-locomotive activities have also been reported by Stackpool etl al. [26] The protocol included locomotion and sports. They found that the hip-worn AT significantly underestimated EE during non-locomotive activities and no significant differences in EE during locomotive activities.

These data provide evidence that in general: 1) hip-worn ATs significantly underestimate EE during cycling, upper-body activities of daily living and inclined locomotion, 2) wrist-worn ATs significantly overestimate EE during locomotion and some sedentary activities, 3) differences are less striking if data are averaged across activities, and 4) hip-worn ATs estimates of EE are more precise (less variable) as compared to wrist-worn AT estimates of EE. Differences in AT estimates during simulated free-living activities and ST extend to locomotion only.

Locomotion only. Five studies have validated ATs in estimating EE during locomotion and the results are equivocal. For example, Diaz et al. [25] sought to validate hip- and wrist-worn ATs in estimating EE during treadmill walking and jogging. They found that the wrist-worn AT significantly overestimated EE during moderate (3.0 mph) and brisk (4.0 mph) walking by 52.4% and 33.3%, respectively. No significant differences in EE estimates were observed between the hip-worn AT and the criterion. Recently, Diaz et al. [30] employed the same treadmill protocol to validate hip-wrist- and bra-worn ATs. They reported that the wrist-worn AT significantly overestimated EE for all treadmill walking and running speeds (range: 1.9 to 5.2 mph), and produced large errors ranging from to 24.5±28.0% to 83.4±45.2%. In contrast, the hip-worn AT significantly underestimated EE during slow walking (1.9 mph) and the bra-worn AT outperformed the wrist-worn AT; errors ranged from 9 to 19%. In both studies, the estimates of EE from the hip-worn AT were more accurate (e.g. mean difference range: -0.8 to 0.4 kcals) and precise (e.g. mean percent error 5.15±0.97), as compared to the wrist-worn AT (e.g. mean difference range: -0.2 to 2.6 kcals; mean percent error

51.4±34.0). Different from these findings, Alsubheen et al. [24] validated a wrist-worn AT in estimating EE during self-selected walking at varying grades (0, 5 and 10%). They found that the wrist-worn AT significantly underestimated kcals by 29% (mean bias: -20.2 kcals) across all conditions. These findings were supported by Dondzila et al. [39] In this study, participants walked and ran at speeds ranging from 3.0 to 6.0 mph. Results showed that, the wrist-worn AT overestimated EE during walking (3.0 mph) and significantly ($p<0.05$) underestimated EE, overall. Lastly, Noah et al. [27] validated a hip-worn AT during flat and graded walking, running and stairs. They found that the hip-worn AT significantly ($p<0.001$) underestimated EE during inclined walking and stairs by an average of 40%. It was concluded that the hip-worn AT is valid for monitoring overground EE.

These data provide evidence that in general: 1) hip-worn ATs significantly underestimate EE during slow waling and inclined locomotion, 2) wrist-worn ATs significantly overestimate EE during locomotion and significantly underestimate EE during graded locomotion, 3) bra-worn ATs are less accurate and precise than hip-worn ATs but outperform wrist-worn ATs, and 4) hip-worn ATs estimates of EE are more precise (less variable) as compared to wrist-worn AT estimates of EE. Differences in hip- and wrist-worn AT estimates of EE have also been reported in free-living environments.

Free-Living Studies

Steps

Seven studies have validated ATs in estimating steps in free-living settings and the results are equivocal. All have employed research-grade accelerometers (i.e. secondary measures) as the step comparison measure. Several studies have shown that in free-living settings, wrist-worn ATs tend to significantly underestimate steps. [21,31,32] However, in cardiac patients significant step overestimations as great as 1,038 steps per day have been reported. [43]

The trend in significant underestimation is extended to hip-worn ATs as well. However, the accuracy and precision of step estimates from hip-worn ATs is superior to step estimates from wrist-worn ATs. Studies have reported mean absolute differences ranging from 6.3% to 7.4% for hip-worn ATs compared to 8.1% to 25.6% for wrist-worn ATs. [32] In contrast, studies have reported significant overestimation of steps from hip-worn a ATs (e.g. 7,477 steps/d).[34] Lastly, pocket-worn ATs show promise yet tend to overestimate (not significantly) steps even when compared to the thigh-worn ActivPAL. [21]

These data provide evidence that in general: 1) hip-worn ATs underestimate and overestimate steps, 2) wrist-worn ATs underestimate and overestimate steps, and 3) hip-worn ATs are more accurate and precise in estimating steps in free-living compared to wrist-worn ATs.

Energy Expenditure

Three studies have validated ATs in estimating EE in free-living settings, and each differs in criterion measure and duration. One study employed DLW, a criterion standard method for measuring total EE.

Murakami et al. [37] validated several ATs (hip-, wrist- and pocket-worn) in estimating EE during 15-days of free-living time compared to DLW. They found that all ATs underestimated total kcals. Mean biases ranged from -590.2 to -171.9 kcals/d (wrist-worn), -280.0 to -69.2 kcals/d (hip-worn) and -220.0 to -93.1 kcals/d (pocket-worn) compared to DLW. It was concluded that most ATs do not produce a valid measure of total EE. The authors speculated that underestimation might be due to periods of not wearing the devices. Ferguson et al. [32] validated several ATs (hip- and wrist-worn) in estimating EE compared to the BodyMedia SenseWear during 48-hours of free-living time. Similar to previous findings, all ATs underestimated total kcals and the wrist-worn ATs produced the greatest bias and least precision compared to hip-worn ATs. For total kcals, mean biases ranged from -533 to -475 (hip-worn) and from -898 to -479 (wrist-worn). The mean absolute differences ranged from 11.6% to 16.1% (hip-worn) and from to 15.6% to 28.8% (wrist-worn). It was concluded that hip-worn ATs outperformed wrist-worn ATs in estimating total EE in free-living settings. Sushames et al. [31] examined the validity of a wrist-worn AT in estimating EE compared to the hip-worn AG (GT3X+) accelerometer during an unspecified time ("several hours") in free-living settings. Data from the ActiGraph accelerometer were post-processed and EE was estimated via a previously validated equation [52] in ActiLife. In contrast to findings in previous free-living studies, the results showed that the wrist-worn AT recorded consistently higher estimated EE by 50% higher (808.1±282.9 kcals) compared to the AG (GT3X+)(538.9±194.0 kcals), with a mean bias (95% CI) of 269.2 (182.6, 355.8) kcals.

Activity Minutes

To date, four studies have reported on ATs in estimating activity minutes (e.g. active time), all were performed in free-living settings. In 2015, Gomersall et al. [33] compared active minutes from a hip-worn AT to a the hip-worn AG (GT3X-BT) (standard using Troiano cut-points) during 14-days free-living time. The AT was found to be strongly correlated rho=0.80 (0.73-0.85; $p<0.01$) with the GT3X-BT but underestimated MVPA by 18±9 minutes per day. Underestimations of MVPA ranged from -189 to -77 minutes per week, which may misclassify a person as not meeting PA recommendations and negatively influence their health. Applying the Freedson 1998 cut-points, [77] Ferguson et al. [32] reported only moderately-to-strong correlations of MVPA (r=0.52-0.91) between ATs and the AG (GT3X+). This study compared several hip- and wrist-worn ATs in estimating activity minutes to a hip-worn ActiGraph (standard measure) during 48-hours of free-living time. They observed large median absolute differences between the AT estimate of activity minutes and the ActiGraph ranging from 26% (wrist-worn) to 298% (hip-worn). For minutes of MVPA, mean biases ranged from 65.9 to 190.4 and -5.2 to 22.7 for hip- and wrist-worn ATs, respectively. Recently, Sushames et al. [31] examined the validity of a wrist-worn AT in estimating MVPA compared to the hip-worn GT3X+ during unspecified ("several hours") free-living time. Utilizing the Freedson 1998 cut-points, they found that the wrist-worn AT produced a mean bias of -35.4 minutes of MVPA. Lastly, Alharbi et al. [43] compared minutes of MVPA from a wrist-worn AT to a hip-worn GT3X-BT in cardiac rehabilitation patients during 4-days of free-living time. Significant correlations (r=.74) between the wrist-worn AT and AG were found for MVPA. However, the wrist-worn AT significantly overestimated MVPA by 10 minutes per day. It was further reported that the wrist-worn AT had high sensitivity (1.00 CI: 91.96, 100) and lower specificity (0.67 CI: 9.43, 99.16) in classifying participants who achieved ≥150 minutes of MVPA per week thereby meeting the recommended PA guidelines using the ActiGraph as the ground truth measure.

The results of these studies are equivocal. Two studies reported ATs underestimated MVPA in free-living settings,[31,33] one study reported ATs overestimated MVPA in free-living settings, [43] and one study reported ATs underestimated and overestimated MVPA in free-living settings. [32]

Sedentary Time

One study has validated an AT (wrist-worn) in estimating sedentary time compared to a hip-worn AG (GT3X+) during 14 days of free-living time. [41] The cut-points used for sedentary time were <100 CPM [133] and MVPA ≥2020 CPM. [90] Longest idle time from the AT was compared to ActiGraph estimates of longest sedentary bout. The results showed that the validity of the wrist-worn AT measure of sedentary time ("longest idle time") was poor. The differences between the AT and GT3X+ estimates of longest sedentary bout were biased, with larger differences when bouts were longer. The limits of agreement were unbiased but wide (mean difference ±88 minutes), varying by up to 150% of the mean estimate according to GT3X+. Though it did accurately classify more than 80% of the sample days as active or inactive based on the 10,000 steps criterion, days were frequently misclassified for meeting public health guidelines of 30 minutes/day of MVPA. The use of an ActiGraph to estimate sedentary time may not be the optimal secondary measurement. If fact, the authors recommended that future studies should consider using the activPAL (PAL Technologies Ltd, Glasgow, UK) device, a thigh-worn accelerometer/inclinometer that evaluates time spent sedentary based on posture rather than the cut-point method. Recently, our group has shown that the thigh-worn activPAL is superior to the hip-worn GT3X+ in estimating sedentary time. [65] Clearly, more studies are needed to validate ATs in estimating sedentary time.

Major Findings and Next Steps

Four output variables have been studied in the investigations that have tested the accuracy of ATs: number of steps, EE, to estimate calories, activity minutes (moderate-to-vigorous activity), achievement of PA recommendations, EE, to estimate calories and sedentary time. The results of these studies are equivocal. Activity trackers under- or overestimate these measures with substantial between-subject variability. For step counts, seven studies showed, ATs overestimated steps in laboratory settings[17-23] and thirteen studies showed ATs underestimated steps.[17-20,23-31] In free-living settings, four studies showed that ATs overestimated steps and lack precision,[21,32-34] and two studies showed that ATs underestimated steps.[31,32] For EE, six studies showed ATs overestimated kcals,[18,25,30,35-37] and 12 studies showed that ATs underestimated kcals [18,24-27,30,35-40] with variable precision and are most accurate for during locomotion and in lab-setting testing conditions[18,25-27,30,36,38,39] compared with non-locomotive activities [18,26,35,36,38,40] and free-living settings.[31,32,37] For activity minutes, one study reported, ATs overestimated MVPA in free-living settings,[32] and two studies reported, ATs underestimated MVPA in free-living settings. [31,33] For sedentary time, only one study has shown, ATs overestimated sedentary time and lack precision in free-living settings. [41]

Based on this evidence, we sought to expand our understanding of the accuracy and precision of ATs in estimating steps, EE, activity minutes and sedentary time in free-living settings using a validated DO system as the criterion measure.[42] Previous free-living studies employed accelerometers as a surrogate for gold-standard criterion measures (e.g. direct observation, DLW) to assess PA.[32-34,43-46] Limitations in using accelerometers as criterion measure to assess PA in free-living settings include 1) the inability to validate compliance (e.g. wear-time, wear-location) and 2) substantial variability in prediction equations used to convert accelerometer data into meaningful PA outcomes (e.g. moderate intensity activity, METs). [47-49] The use of DO as a criterion measure in free-living settings address these limitations and will attenuate the sources of error inherent in previous free-living studies. The evidence from this novel study will inform consumers, researchers, clinicians and interventionists about the utility of ATs as intervention tools and potentially, assessment tools for research.

Study Three: Activity Trackers are Sensitive to Change in Physical Activity and Sedentary Behaviors in Free-Living Settings

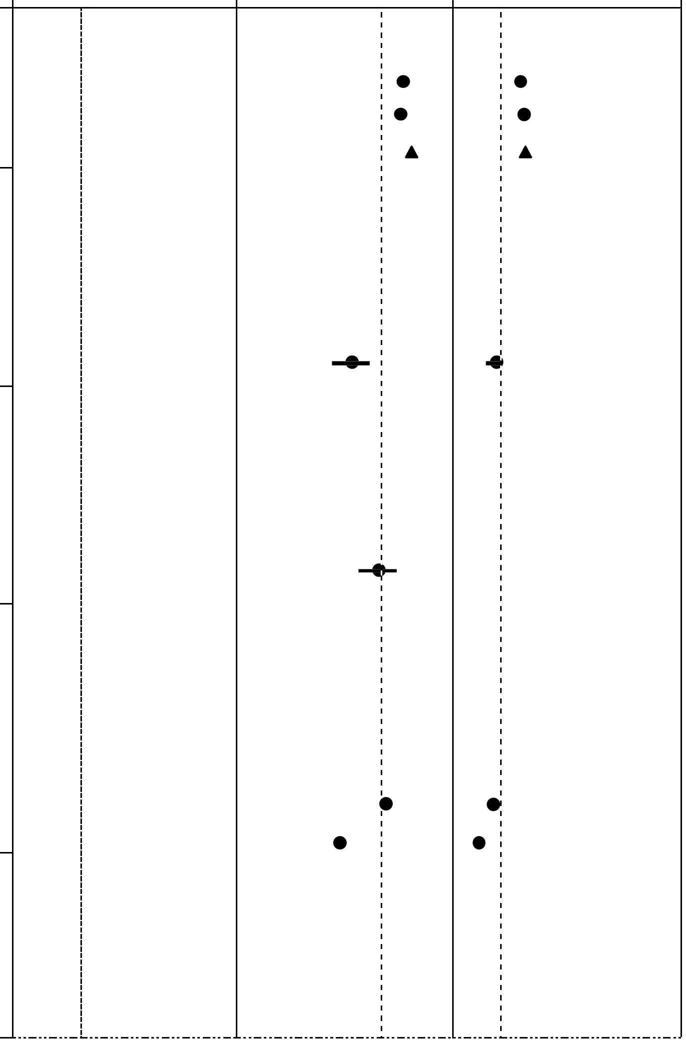
To date, no studies have investigated the ability of ATs to detect change in PA behaviors in free-living settings. Activity trackers are becoming increasingly popular with consumers, researchers and clinicians, and used as both PA exposures and PA outcomes. Examining the capacity of ATs to detect change in free-living PA behaviors is an important next step to broadening our understanding of these devices. Examining research-grade accelerometers’ in detecting change in PA behaviors in free-living settings is of equal importance.

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
Laboratory Based				
Bai 2015 [36]	N=52 (18-60 yr) **LOC:** TRD walking & running; **Sports**: Resistance EX; and **SB** Criterion: EE=IC **Results**: • FB-Wrist [a]			
Beevi 2016 [28]	N=14 (29.9±4.9 yr) **LOC:** flat TRD walking at 0.6, 1.2, and 1.8 mph Criterion: Steps=DO (100 steps) **Results**: ⊃⊂ • FB-Hip [d]			
Case 2015 [29]	N=14 (28.1±6.2 yr) **LOC:** flat TRD walking (3.0 mph;1500 steps) Criterion: Steps=DO **Results**: • FB-Hip [b] • FB-Wrist [b]			
Chen 2016 [20]	N=30 (21.5±2.0 yr) **LOC**: TRD walking, SS: overground walking w/load/stroller, stairs; **Lifestyle**: laundry;**SB** Criterion: Steps=DO **Results**: • FB-Wrist [d] ○ **LOC** ▪DOM ▪Non-DOM ○ **Non-LOC** ▪DOM ▪Non-DOM			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
Dannecker 2013 [40]	N=14 (28.1±6.2 yr). Duration: 4-hrs. **LOC**: TRD walking, stairs; **Sports**: cycle ergometer; **Lifestyle**: sweeping, standing; **SB** Criterion: EE=RC **Results**: • FB-Hip [c]			
Diaz 2015 [25]	N=23 (20-54 yr) **LOC:** flat TRD walking & running Criterion: EE=IC; steps=DO **Results**: • FB-Hip [d] • FB-Wrist [d]			
Diaz 2016 [30]	N=13♀ (32.0±9.2 yr) **LOC:** flat TRD walking & running Criterion: EE=IC; steps=DO **Results**: • FB-Hip [j] • FB-Wrist [j] • FB-Bra [j]			
Dondzila 2016 [39]	N=19♂ (24.6±3.1 yr) **LOC:** flat TRD walking & running Criterion: EE=IC **Results**: • FB-Wrist [i]			
Fortune 2014 [19]	N=12 (25-55 yr) **LOC:** SS: overground walking & jogging Criterion: Steps=DO **Results**: • FB-Hip [h] • FB-Ankle [h]			
Kooiman 2015 [21]	N=33 (39±13.1) **LOC:** flat TRD walking (2.9 mph; 30 min)			

Author	Research Study	Activity Minutes (- 0 +)	Energy Expenditure (- 0 +)	Steps (- 0 +)
	Criterion: Steps=Optogait system **Results**: ⊃⊂ • FB-Wrist [i] • FB-Pocket [i]			
Lee 2014 [35]	N=60 (18-43 yr) **LOC:** TRD walking & running, overground walking (20 steps); **Sports:** cycle ergometer, elliptical, Wii Tennis, basketball; **SB.** Criterion: EE=IC. **Results**: • FB-Hip [a]			
Mammen 2012 [17]	N=10 (23.0±1.2 yr) **LOC:** TRD walking & running, overground walking (20 steps); **TRANS:** driving. Criterion: Steps=DO **Results**: • FB-Hip [a] • FB-Pocket [a] • FB-Collar [a]			
Murakami 2016 [37]	N=19 (18-80 yr). Duration: 24-hrs. **LOC:** TRD walking; **Lifestyle**: eating, computer, TV, housework; **SB** Criterion: EE=RC **Results**: • FB-Wrist [d]			
Nelson 2016 [18]	N=30 (18-80 yr) **LOC:** SS: flat TRD walking & jogging, overground walking & jogging, stairs; **Sports:** cycle ergometer, **Lifestyle**: sweeping, dusting, laundry, bedding, gardening, standing; **SB** Criterion: EE=IC; Steps=DO			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	Results: • FB_Z-Hip [i] • FB_O-Hip [i] • FB-Wrist [i]			
Noah 2013 [27]	N=23 (26.6±7.5 yr) **LOC:** TRD walking & running Criterion: EE=IC; steps=Actical-Hip Accelerometer. **Results**: • FB-Hip [e g]			
Sasaki 2012 [38]	N=20 (24.1±4.5 yr) **LOC:** TRD walking & running, stairs; **Sports**: cycle ergometer, golf, tennis, basketball; **Lifestyle**; and **SB** Criterion: EE=IC **Results**: • FB-Hip [a]			
Stackpool 2014 [26]	N=20 (18-44 yr) **LOC:** TRD walking & running; **Sports**: elliptical, exercise/agility & ladder drills Criterion: EE=IC; steps=DO **Results**: • FB-Hip [e]: ○ **LOC** ○ **Non-LOC**			
Storm 2015 [22]	N=16 (28.87±2.65 yr) **LOC:** SS: indoor & outdoor overground walking, stairs Criterion: Steps=OPAL sensors (L and R shanks) **Results**:			



Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	• FB-Hip [i]			
Sushames 2016 [31]	N=25 (23.7±5.8 yr) **LOC:** SS: TRD walking (flat & incline) & jogging, stepping Criterion: Steps=DO **Results**: • FB-Wrist [a]			
Takacs 2013 [23]	N=30 (29.6±5.7 yr) **LOC:** TRD walking & running Criterion: steps=DO **Results**: • FB-Hip [f]			
Free-Living				
*Alharbi 2016 [43]	N=48 (65.6±6.9 yr) cardiac rehab. Duration: 4-days Criterion: • Activity minutes: GT3X-BT-Hip ○ Cut-points NQLS-Senor • Steps: GT3X-BT-Hip **Results**: • FB-Wrist [a]			
Ferguson 2015 [32]	N=21 (32.8±10.2 yr). Duration: 48- hours. Criterion: • Activity minutes: GT3X+-Hip ○ Freedson 1998 cut-points • EE: BS • Steps: GT3X+-Hip; BS-upper arm **Results**: • FB-Hip [a g]			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
Gomersall 2015 [33]	N=29 (39.6±11 yr). Duration: 14-days Criterion: • Activity minutes: GT3X-BT-Hip ○ Troiano 2008 cut-points • Steps: GT3X-BT-Hip **Results**: • FB-Hip [g]			
Kooiman 2015 [21]	N=56 (37.1±10.6 yr). Duration: 7.5 hrs. Criterion: Steps=ActivPAL **Results**: ⊃⊂ • FB-Wrist [i] • FB-Pocket [i]			
Murakami 2016 [37]	N=19 (18-80 yr). Duration: 15-days Criterion: EE=DLW **Results**: FB-Wrist [d]			
Sushames 2016 [31]	N=25 (23.7±5.8 yr). Duration: Not stated (<24 hrs). Criterion: • Activity minutes: GT3X+-Hip ○ Freedson 1998 cut-points • EE: GT3X+-Hip ○ Actilife v6.2 • Steps: GT3X+-Hip **Results**: FB-Wrist [a]			
Tully 2014 [34]	N=40 (43±12 yr). Duration: 7-days Criterion: • Steps: GT3X-Hip ○ Freedson 1998 cut-points			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	Results: • FB-Hip [h]			
		-25 +5029	-1216 +455	-4165 +14971

Table 1. Summary of current Fitbit (FB) validation studies

LOC, locomotion
DOM, dominant hand
SB, sedentary behavior
TRANS, transportation
TRD, treadmill
EX, exercise
FB, Fitbit
EE, energy expenditure
IC, indirect calorimetry
DLW, doubly-labeled water
2MWT, 2-minute walk test
BS, BodyMedia Sensewear armband
a, mean bias (95% Limits of Agreement)
b, mean step count (relative difference)
c, RMSE
d, mean difference (range)
e, mean (range)
f, % relative error
g, mean (SD)
h, median (SD)
i, mean absolute percent error (95% CI)
j, mean % error (SD)
RC, room calorimeter
AG, ActiGraph
MVPA, moderate-to-vigorous physical activity

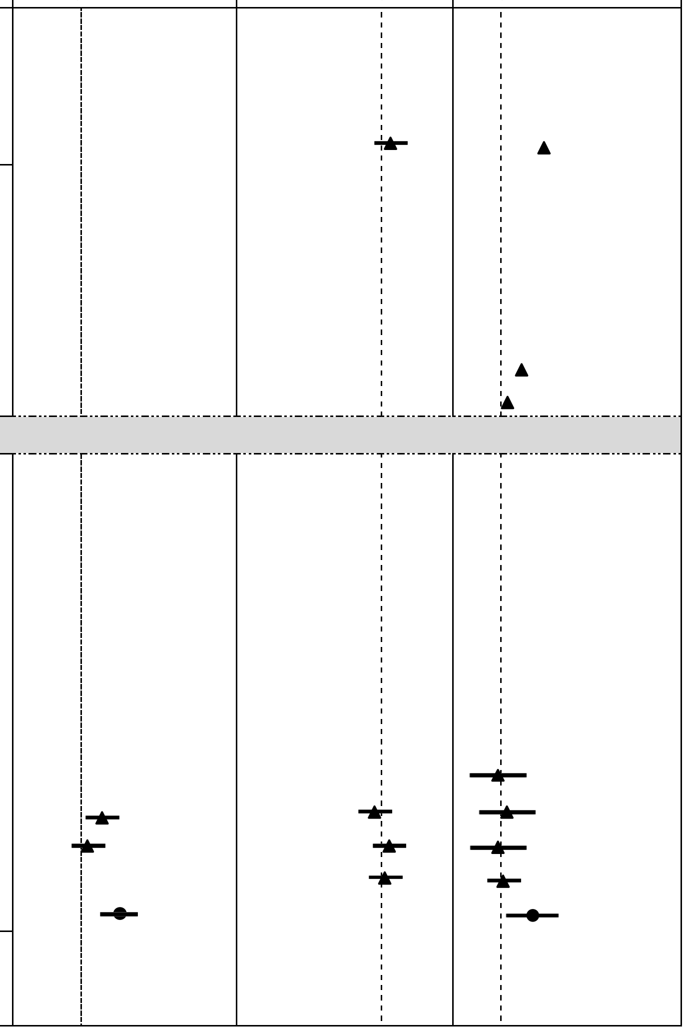
SS, self-selected
TBI, traumatic brain injury
NLQS, neighborhood quality of life study
● , hip-worn Fitbit
▲ , wrist-worn Fitbit
◆ , bra-worn Fitbit
○ , collar-worn Fitbit
⧗ , pocket-worn Fitbit
■ , ankle-worn device
* , special population
⊃⊂ , includes devices in both validation study tables.
Note. All studies were conducted on healthy populations (except for studies denoted with an *) and were approximately 50% female (except for studies denoted with an ♂= all male; ♀= all female).

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
Laboratory Based				
Alsubheen 2016 [24]	N=13 (40±11.9 yr) **LOC:** SS: flat and graded TRD walking Criterion: EE=IC; Steps=DO **Results**: GAR-Wrist [d]			
Bai 2015 [36]	N=52 (18-60 yr) **LOC:** TRD walking & running; **Sports**: Resistance EX; and **SB** Criterion: EE=IC **Results**: ⊃⊂ • MS-Wrist [a] • NFB-Wrist [a] • JU-Wrist [a]			
Beevi 2016 [28]	N=14 (29.9±4.9 yr) **LOC:** flat TRD walking at 0.6, 1.2, and 1.8 mph Criterion: Steps=DO (100 steps) **Results**: ⊃⊂ • YM-Hip [d] • OM-Hip [d]			
Case 2015 [29]	N=14 (28.1±6.2 yr) **LOC:** flat TRD walking (3.0 mph;1500 steps) Criterion: Steps=DO **Results**: ⊃⊂ • NFB-Wrist [b] • JB-Wrist [b] • YDW-Hip [b]			
Chen 2016 [20]	N=30 (21.5±2.0 yr) **LOC**: TRD walking, SS: overground walking w/load w/stroller, stairs; **Lifestyle**: laundry; **SB** Criterion: Steps=DO			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	Results: ⊃⊂ • GAR-Wrist [d] ○ **LOC** ▪DOM ▪Non-DOM ○ **Non-LOC** ▪DOM ▪Non-DOM • JB-Wrist [d] ○ **LOC** ▪DOM ▪Non-DOM ○ **Non-LOC** ▪DOM ▪Non-DOM			
Dondzila 2016 [39]	N=19 ♂ (24.6±3.1 yr) **LOC:** flat TRD walking & running Criterion: EE=IC **Results**: ⊃⊂ • JE-Ears [i] -19.1			
Fortune 2014 [19]	N=12 (25-55 yr) **LOC:** SS: overground walking & jogging Criterion: Steps=DO **Results**: ⊃⊂ • NFB-Wrist [h]			
Kooiman 2015 [21]	N=33 (39±13.1) **LOC:** flat TRD walking (2.9 mph; 30 min) Criterion: Steps=Optogait system **Results**: ⊃⊂ • JB-Wrist [i] • NFB-Wrist [i] • YDW-Waist [i]			

Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	• MS-Pocket [i] • OM-Pocket [i] • WP-Pocket [i]			
Lee 2014 [35]	N=60 (18-43 yr) **LOC:** TRD walking & running, overground walking (20 steps); **Sports:** cycle ergometer, elliptical, Wii Tennis, basketball; **SB** Criterion: EE=IC. **Results**: ⊃⊂ • DL-Chest [a] • B1-Wrist [a] • JB-Wrist [a] • NFB-Wrist [a]			
Murakami 2016 [37]	N=19 (18-80 yr). Duration: 24-hrs **LOC:** TRD walking; **Lifestyle**: eating, computer, TV, housework; **SB** Criterion: EE=RC **Results**: ⊃⊂ • WP-Wrist [d] • JB-Wrist [d] • GAR-Wrist [d] • SL-Waist [d] • PA-Waist [d] • EP-Wrist [d] • TAN-Pocket [d] • MS-Wrist [d] • OA-Waist [d] • OC-Pocket [d]			
Nelson 2016 [18]	N=30 (18-80 yr) **LOC:** SS: flat TRD walking & jogging, overground walking & jogging, stairs; **Sports:**			

Author	Research Study	Activity Minutes (- 0 +)	Energy Expenditure (- 0 +)	Steps (- 0 +)
	cycle ergometer, **Lifestyle**: sweeping, dusting, laundry, bedding, gardening, standing; **SB** Criterion: EE=IC; Steps=DO **Results**: ⊃⊂ • JB-Wrist [i]			
Storm 2015 [22]	N=16 (28.87±2.65 yr) **LOC:** SS: indoor & outdoor overground walking Criterion: Steps=OPAL sensors (L and R shanks) **Results**: ⊃⊂ • NFB-Wrist [i] • JB-Wrist [i]			
Free-Living				
Ferguson 2015 [32]	N=21 (32.8±10.2 yr). Duration: 48- hours. Criterion: • Activity minutes: GT3X+-Hip ○ Freedson 1998 cut-points • EE: BM • Steps: GT3X+-Hip; BS-upper arm **Results**: ⊃⊂ • NFB-Wrist [a g] • JB-Wrist [a g] • MS-Wrist [a g] • WP-Hip [a g] • SSP-Hip [a g]			
Kooiman 2015 [21]	N=56 (37.1±10.6 yr). Duration: 7.5 hrs. Criterion: Steps=ActivPAL **Results**: ⊃⊂			



Author	Research Study	Activity Minutes - 0 +	Energy Expenditure - 0 +	Steps - 0 +
	• JB-Wrist [i] • NFB-Wrist [i] • YDW-Waist [i] • MS-Pocket [i] • OM-Pocket [i] • WP-Pocket [i]			
Murakami 2016 [37]	N=19 (18-80 yr). Duration: 15-days Criterion: EE=DLW **Results**: ⊃⊂ • WP-Wrist [d] • JB-Wrist [d] • GAR-Wrist [d] • SL-Waist [d] • PA-Waist [d] • EP-Wrist [d] • TAN-Pocket [d] • MS-Wrist [d] • OA-Waist [d] • OC-Pocket [d]			

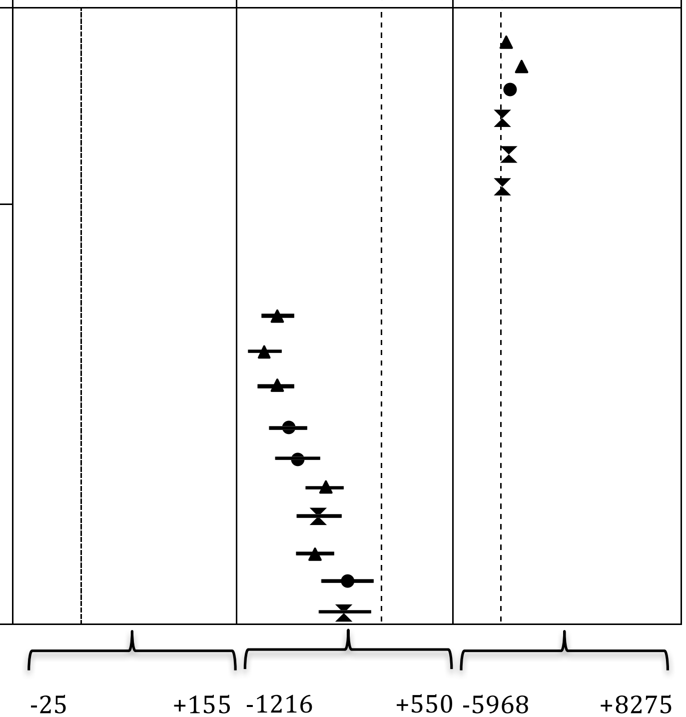


Table 2. Summary of current activity tracker validation studies; Fitbit excluded

LOC, locomotion
DOM, dominant hand
SB, sedentary behavior
TRANS, transportation
TRD, treadmill
EX, exercise
EE, energy expenditure

IC, indirect calorimetry
DLW, doubly-labeled water
BS, BodyMedia Sensewear armband
a, mean bias (95% Limits of Agreement)
b, mean step count (relative difference)
c, RMSE
d, mean difference (range)
e, mean (range)
f, % relative error
g, mean (SD)
h, median (SD)
i, mean absolute percent error (95%CI)
j, mean % error (SD)
GAR, Garmin
JB, Jawbone
JE, Jabra earbuds
NFB, Nike Fuel Band
MS, Misfit Shine
PA, Panasonic Actimarker
OC, Omron CaloriScan
OA, Omron Active Style Pro
EP, Epson Pulsense
SL, Suzuken Lifecorder
TAN, Tanita AM- 160
WP: Withings Pulse O_2
OM, Omron pedometer
YDW, Yamax Digi-Walker
SSP, Striiv smart pedometer
DL, DirectLife
B1, Basis 1
RC, room calorimeter

AG, ActiGraph
MVPA, moderate-to-vigorous physical activity
PED, pedometer
● , hip-worn device
▲ , wrist-worn device
⧗ , pocket-worn device
■ , ankle-worn device
□ , ear-worn device
✖, chest-worn device
⊃⊂ , includes devices in both validation study tables.
Note. All studies were conducted on healthy populations (except for studies denoted with an *) and were approximately 50% female.

CHAPTER 3
METHODS

Study One: A Comparison of Activity Tracker and ActiGraph™ GT3X-BT Accelerometers in Estimating Energy Expenditure and Steps During Orbital Shaking

Experimental Instrumentation and Procedures

Instrumentation

Research-grade accelerometer: Reference Standard. The ActiGraph (AG) GT3X-BT (GT3X-BT) accelerometer (ActiGraph™ LLC, Pensacola, Florida) served as the reference standard to which all ATs were compared. This device is a lightweight triaxial PA monitor (4.6cm x 3.3cm x 1.5cm, 19g) that measures acceleration ranging from -8 to +8 g's. The accelerometer output can be sampled at rates ranging from 30 to 100 Hz and is digitized by a twelve-bit analog-to-digital converter. The AG includes a micro-electro-mechanical system (MEMS) based accelerometer. The acceleration data are sampled by a 12-bit analog to digital converter and stored in a raw, non-filtered/accumulated format in g's. These data are stored directly in non-volatile flash memory. Raw data are collected at the selected sample rate (80 Hz for this study) and are post-processed in the ActiLife software. Users generate files containing any desired combination of parametric data (e.g. 1-sec epoch, 60-sec epoch) during the data processing step.

Activity Trackers. Activity trackers were chosen based on the following two criteria: (1) no known gravimeter within the device, and (2) researchers had at least two of the device. The rationale for not containing a gravimeter was that the electronic orbital oscillator does not apply vertical accelerations and as a result, a device that contains a gravimeter would produce inaccurate output. The rationale for at least two devices was to counterbalance each other in the electronic orbital oscillator. As a result, six different ATs were studied: 1) Fitbit Flex (FBF), 2) Fitbit One (FBO), 3) Garmin® Vivofit (GV), 4) Misfit Flash (MFF), 5) Misfit Shine (MFS) and 6) New Lifestyles NL-1000 pedometer (NL). See Appendix E for detailed specifications of each AT.

Electronic Orbital Shaker. The electronic orbital shaker (Advanced Orbital Shaker, Model 10000-2; VRW International, Radnor, PA) (Figure 1) produces controlled oscillations between 0.25 and 5.0 Hz. The electronic orbital shaker oscillates at various radii between 1.27 and 5.7 cm. Four trays (51 x 10 x 10 cm) are mounted on the base oscillating plate (60 x 60 cm) of the shaker. Each tray has four custom foam cushion slots that securely held the GT3X-BTs and ATs in place to eliminate device movement during electronic orbital shaking (Figure 2).

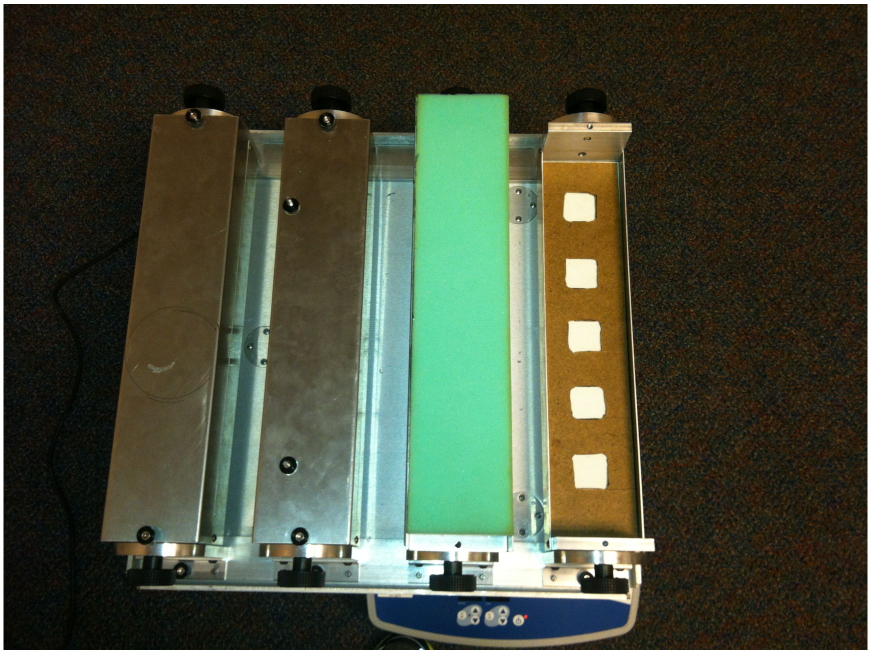


Figure 1. Electronic Orbital Shaker.

Procedures

Electronic Orbital shaker. The electronic orbital shaker (Figure 2) was used to perform motion testing. Two of each device were tested at the same time. All devices were placed in the custom foam cushion slots with their vertical plane perpendicular to the control panel of the electronic orbital shaker (figure 2).

The GT3X-BTs and ATs were oscillated for three: (1) twenty-four, 3-minute trials, and (2) 2-hour trials. Each 3-minute trial consisted of one monitor oscillation frequency increased from zero to 3.0 Hz in 0.1 Hz increments on a fixed radius [56,134] of 5.08 cm. The three 2-hour trials consisted of oscillation frequencies ranging from zero to 3.0 Hz., based on the American Time Use Survey, to simulate free-living whole body acceleration (e.g. variation). These frequencies simulate hip rotation ranging from no movement (e.g. sleep) to ambulation at speeds ranging between 1.5 and 16 miles per hour.[81]

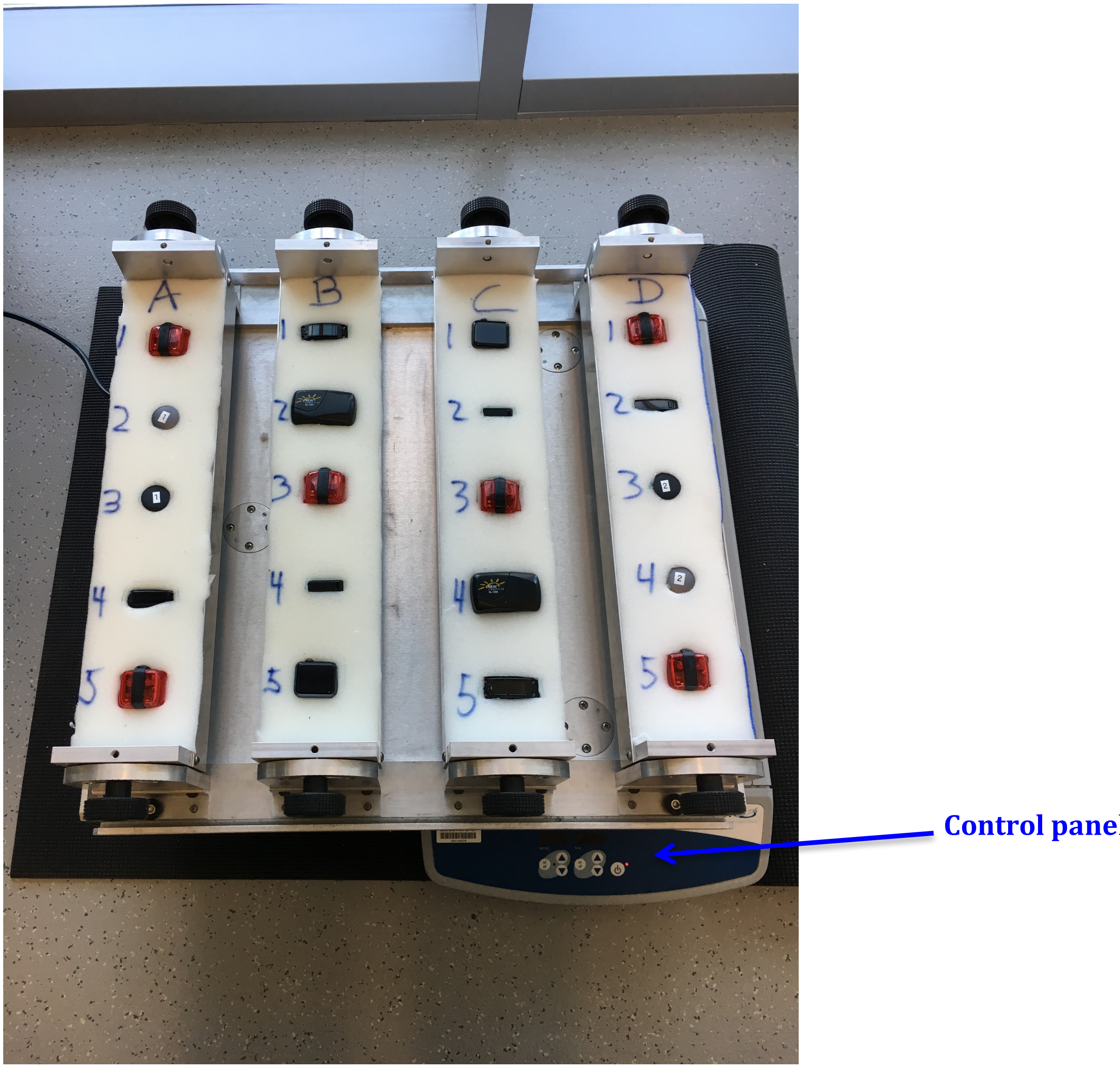


Figure 2. Electronic orbital shaker with devices in custom foam cushioned slots

Oscillation duration

The total oscillation duration for a given range of frequencies (range: 0.0, 0.25 to 3.0 Hz) reflected the American Time Use Survey (ATUS) percentages of time spent in selected activities (Figure 3), normalized for 2-hours (versus 24-hours) and conformed to 5-minute trials (e.g. 44.5-minutes to 45 minutes) (Figure 4). The ATUS describes the amount of time people spend doing various activities, such as paid work, childcare, volunteering, and socializing. According to the ATUS, in 2014, working people aged 25 to 54 years spent the majority of their weekdays sleeping (~32%) and working (~37%), with leisure and sport activities comprising 10% of daily activities (Figure 3).[135] Thus, the 2-hour oscillation trials reflected the ATUS percentages of time spent in each activity, normalized for 2-hours (versus 24-hours)(Figure 4). See table 3 for examples of activities and associated MET values.

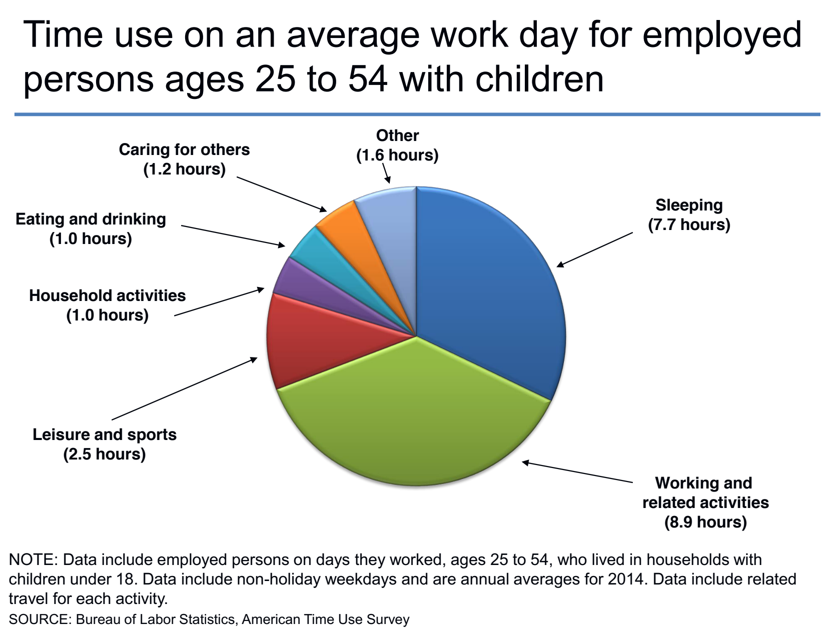


Figure 3. ATUS: Time use on an average workday for employed persons ages 25-54 in 2014

Total time (minutes) spent in each

activity category at a given frequency

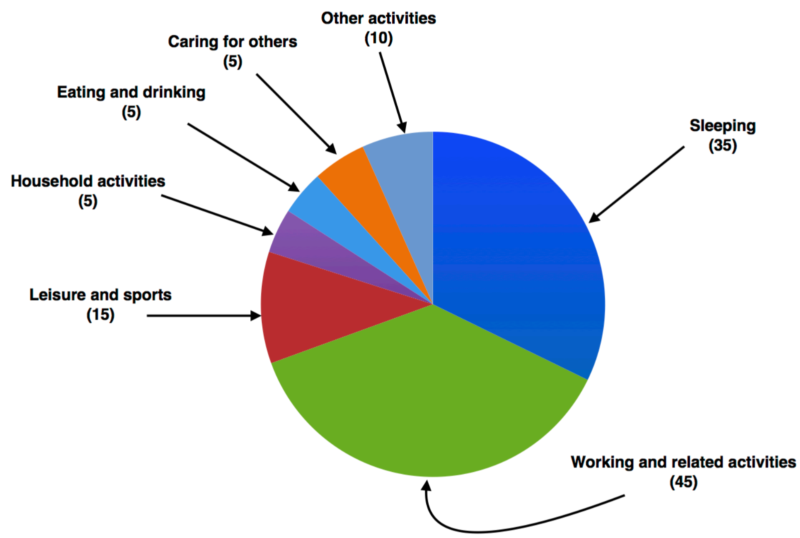


Figure 4. Time spent in each activity category at a given frequency (range: 0.0, 0.25 to 3.0 Hz) for 2-hour trials.

Oscillation frequencies

Oscillation frequency ranges for each activity category were established by electronically oscillating six GT3X-BT accelerometers at 0.0 to 3.0 Hz in 0.1 Hz increments. Each 0.1 Hz. increment was oscillated for 3-minutes and the second minute of each trial was used to determine counts per minute at each frequency. Using the second minute ensured that the desired frequency was met. Figure 5 displays how oscillation frequency ranges were derived. Briefly, the GT3X-BT data were initialized to collect data at 80 Hz., with the low frequency extension for oscillation frequencies <0.7 Hz., post processed using ActiLife software (v 6.1.3) and aggregated into VM counts per minute. These data were scored in ActiLife using the Freedson VM3 cut-points. [52]

Lastly, the intensity categories and their associated frequencies were used to determine electronic oscillation trial: frequency, intensity and total time. Table 3 illustrates, performing household activities such as washing dishes produces MET values ranging from 1.5 (light) to 2.9 (light), which corresponds to oscillation frequencies ranging from 0.5 to 0.7 Hz.

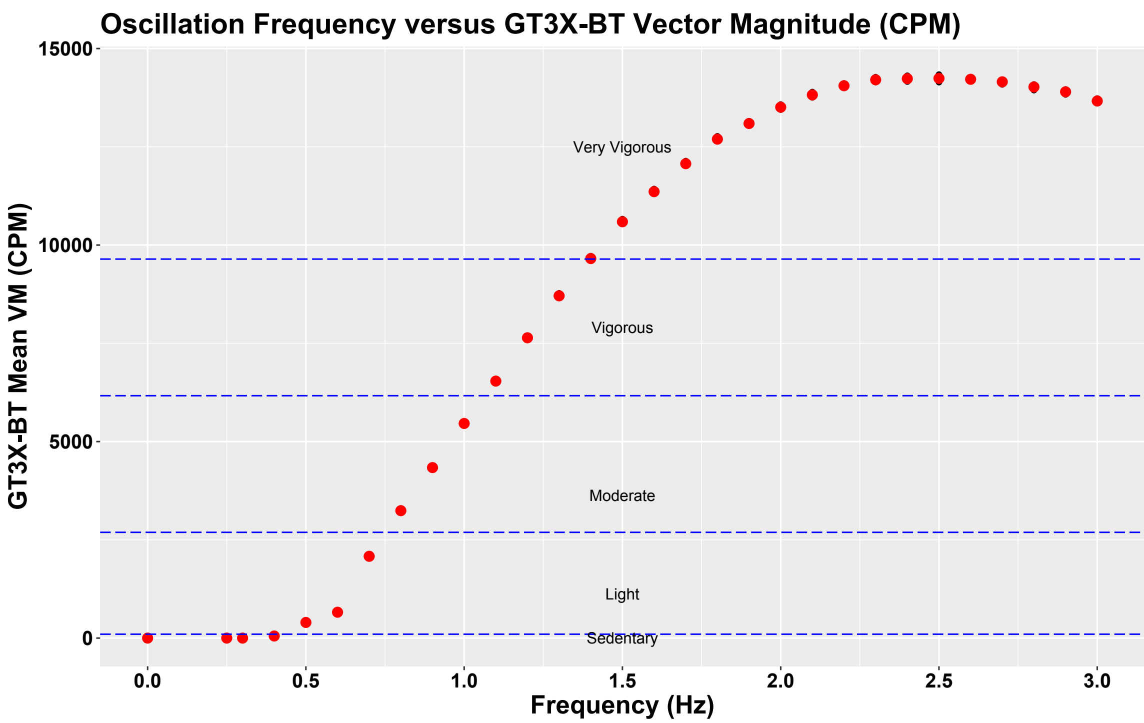


Figure 5. Determination of oscillation frequency ranges

During each 3-minute and 2-hour trial devices were oscillated at various frequencies to simulate different movement intensities. To simulate variation in free-living whole body acceleration, variation of the shaker oscillation frequencies occurred during testing (Table 3).

Frequency Range (Hz)	ActiGraph GT3X-BT (VMCPM)	Intensity	METs	Compendium of Physical Activities	
				Category	Activity
0.0 – 0.4	0 – 99	Sedentary	≤ 1.5	Inactivity	Sleeping; sitting quietly
0.5 – 0.7	100 – 2690	Light	1.5 – 2.9	Home Activity	Washing dishes; cooking
0.8 – 1.0	2691 – 6166	Moderate	3.0 – 5.9	Occupational	Farming, feeding cattle; massage therapist
1.1 – 1.3	6167 – 9642	Vigorous	6.0 – 8.9	Walking	Hiking, cross country; carrying load upstairs, general
1.4 – 3.0	≥ 9643	Very Vigorous	≥ 9.0	Sport	Track and field (e.g., steeplechase, hurdles)

Table 3. Electronic orbital shaker frequency ranges with corresponding: VMCPM, intensity categories, METs and activities
VMCPM, vector magnitude counts per minute, MET, metabolic equivalent

Data collection and processing. The GT3X-BTs were synced to the same laptop as the ATs and initialized in advance of data collection (sampling rate of 80Hz). These raw data were post processed into 1-second epochs/counts and steps via ActiLife v6.1.3 software.

Minute-by-minute EE (kcals) were estimated and summed for all 3-minute trials and for each 2-hour trial using the prediction equation previously developed by our group, [52] labeled the "Freedson VM3 (2011)" equation in the ActiLife software. The Freedson VM3 equation:

Kcals/min= $0.001064 \times VM + 0.087512(BM) - 5.500229$

Where,

VM = Vector Magnitude Combination (per minute) of all 3 axes

$(\sqrt{[(\text{Axis } 1)^2+(\text{Axis } 2)^2+(\text{Axis } 3)^2]})$

BM = Body Mass in kg

Weight was standardized for the GT3X-BTs and ATs. The low frequency extension (LFE) option was selected in the ActiLife software to detect lower amplitude movements. The LFE option lowers the baseband of the filter cut-off, expanding the bandwidth of the accumulated data. The LFE was selected to ensure acceleration detection at slower oscillation frequencies (e.g. 0.7 Hz).

Activity Trackers

Pre-3-minute oscillation trials and 2-hour oscillation trial. Thirty-minutes prior to the first 3-minute and the 2-hour oscillation trial, all ATs were initialized/synched using the same user profile (e.g. date of birth, gender, height and weight) and the same computer used to initialize the GT3X-BTs. Next, the GT3X-BT and activity trackers (FBF, FBO, MFF, MFS, GV and NL) were secured into their respective customized foam cushion slots within each tray of the electronic orbital shaker (Figure 2). Two of each device were tested in the electronic orbital shaker.

Immediately prior to each 3-minute oscillation trial and each 2-hour oscillation trial, researchers retrieved all Misfit data via the Misfit app (iPhone 6s) and recorded baseline energy EE and step values for the MFS and MFF, as neither device is equipped with a real-time display. The NL pedometers were set at 0 steps. The values for EE and steps from the Fitbit FBF, FBO and the GT3X-BT were retrieved and recorded pre-and post each 3-minute oscillation trial and each 2-hour oscillation trial. The start and stop time for each 3-minute oscillation trial and each 2-hour oscillation trial were synchronized with the time of the same laptop used for initialization/synching and downloading of all devices.

Data Processing and Statistical Evaluation

Data Processing

Following each 3-minute oscillation trial and each 2-hour oscillation trial total EE and steps for the: 1) MFF and MFS were downloaded via bluetooth and retrieved via the Misfit app (iPhone 6s), 2) FBF and FBO were synched/downloaded to the Fitbit Dashboard via Bluetooth and retrieved from the Fitabase website (described below), and 3) GV were retrieved from the real-time display. Total steps for the NL pedometer were retrieved from the real-time display. The GT3X-BT data were collected at 80 Hz, with the low frequency extension for oscillation frequencies <0.7 Hz, post processed using

ActiLife software (v 6.1.3) and aggregated into VM counts per minute. Total estimated kcals for each 3-minute oscillation trial and each 2-hour oscillation trial were calculated and summed employing the "Freedson VM3 (2011)" equation in Actilife (v 6.1.3). Total steps from the GT3X-BT were obtained by summing: 1) each 3-minute oscillation trial, and 2) each 2-hour oscillation trial.

Fitabase (Small Steps Labs, LLC. San Diego, Ca). All Fitbit data were exported using Fitabase. Fitabase is a research platform that accesses data from Internet connected consumer devices. Currently, Fitbit is the only consumer device company that utilizes Fitabase. The advantage of using this platform to acquire Fitbit data is that it provides minute-by-minute data for activity minutes (intensity), kcals, MET-minutes and steps in comparison to the Fitbit software and Dashboard which only provide total activity minutes (intensity), kcals and steps for the monitoring period.

Statistical evaluation

All data cleaning, processing and analysis were done using the open source *R statistical software package*, version 3.3.3 (www.r-project.org) and computing language R. [136]

Data Analysis. Three-minute oscillation trial and two-hour oscillation trials. Repeated measure random effects models were assessed main effects of device and frequency and the interaction of device x frequency on AT estimates of EE and steps compared to GT3X-BT estimates of EE and steps. Significance level was set at $\alpha = .05$.

Study Two: Validation Consumer and Research-Grade Activity Monitors in Free-Living Settings

Sample Size and Power

Using data from a previous (free-living) study,[42] we found a between subject standard deviation of 0.17 METs and a within subject standard deviation of 1.46 METs. The relatively larger within subject variability informed our decision to measure each subject multiple times. A sample of 32 subjects yielded at least 80% power to detect average MET differences of less than 0.45 METs per hour.

Recruitment, Eligibility

Thirty-two adults (16 females and 16 males) 18-59 years of age were recruited to participate in this study. Thirty-two participants yielded 192 hours of free-living data. Volunteers were from the Amherst, Massachusetts's area and were recruited using flyers and word of mouth. Volunteers were screened either in person (in the *Physical Activity and Health Laboratory (PAHL)*) or over the phone (from the *PAHL*) and were automatically excluded if they had any diagnosed cardiovascular, pulmonary, metabolic, joint, or chronic diseases, or limitation(s) in locomotion. If volunteers were considered eligible, they were invited to the *PAHL* for an informed consent visit.

Experimental Instrumentation and Procedures

Participants were fitted with a variety of activity monitors that were worn on the wrists, hips and ankle, and a biometric shirt. The devices included: (1) wrist-worn, GT3X-BT (AGwrist), Apple iWatch Sport (AiW), Fitbit Flex (FBF), Garmin Vivofit (GV), Microsoft Band (MB), Misfit Shine (MFS) and Polar Loop (PL); (2) hip-worn, GT3X-BT (AGhip), Fitbit One (FBO), Misfit Flash (MFF), New Lifestyles NL-1000 (NL) and Withings Pulse (WP); (3) ankle-worn, StepWatch (SW); and (4) Hexoskin Biometric shirt (HxSkin) (i.e. smart shirt).

The researchers video recorded participants for each of the 2-hr sessions while participants performed normal activities. If private time was required (i.e. going to the bathroom), we did not observe participants during these private time periods. At the end of the 2-hr recording period, the researchers removed the activity monitors and the participants removed the smart shirt.

Instrumentation

Research-grade accelerometer. The previously described, ActiGraph GT3X-BT (GT3X-BT) Accelerometer (ActiGraph™ LLC, Pensacola, Florida).

StepWatch™ (mōdus™ health llc, Washington, DC) monitor. The StepWatch monitor is worn at the ankle of the dominant leg. The StepWatch is a reliable [137] and accepted criterion measure for steps in healthy adults. [138] The StepWatch is a research and clinical tool for assessment of ambulatory function in free-living settings. It is an ankle- worn, microprocessor-controlled step counter, and detects steps for a wide variety of normal and abnormal gait styles and cadences. Step counts can be recorded every 3 to 60 seconds. The StepWatch has been cleared by the US government FDA as a class II device.

Activity Trackers. Nine different activity trackers were studied: 1) AiW, 2) FBF, 3) FBO, 3) GV, 4) MB, 5) MFF, 6) MFS, 7) PL, 8) WP and 9) NL. For the NL, participants' stride length was determined according to the manufacturers recommended method and programmed into the device. [139] See Appendix E for detailed specifications of each activity tracker.

Biometric Shirt. The market for emerging wearable categories including smart clothing is rapidly developing. According to a recent report, smart clothing shipments will grow from 140,000 units in 2013 to 10.2 million units by 2020. [140] The Hexoskin Biometric Shirt (Hexoskin) (Carré Technologies Inc., Montréal, Québec, Canada) is sustained, in large part, by its utility as a tool for the management of athletes' health, [141] remote medical monitoring for long-term space missions and space exploration [142] and objectively measuring clinical populations in research settings. [143]

The Hexoskin is a multi-parameter physiological recording system designed to monitor levels of physical activity and energy expenditure, which combines measurements of cardiac, ventilator, and hip-motion intensity (Figure 6).

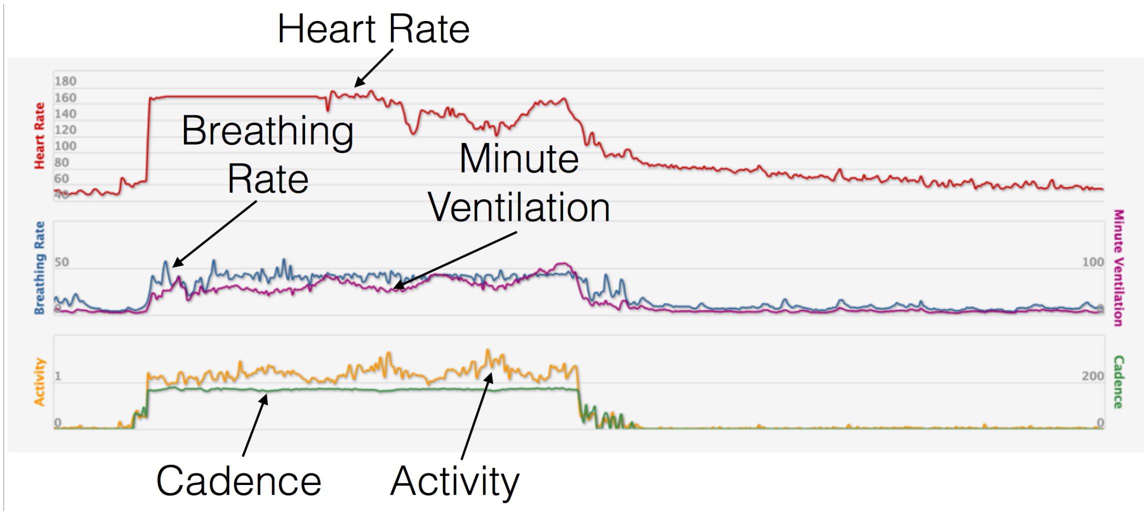


Figure 6. Hexoskin output from one observation session

Video Recording. We employed a GoPro Hero+ LCD (GOPRO, Inc. San Mateo, Ca) camera to record all observation sessions. The GoPro Hero+ LCD is a small, lightweight, waterproof camera that is capable of recording video at 1080 pixels and up to 70 frames per second. The GoPro app was used to password protect the GoPro Hero+ LCD camera via wifi. A 64 GB SanDisk micoSD™ memory card (SanDisk, Inc. Milpitas, Ca) was used to store the GoPro Hero+ LCD video files.

Noldus (Information Technology B.V: Wageningen, Netherlands). The Noldus Observer® XT is the software package for the collection, analysis, and presentation of observational data (Noldus Information Technology B.V: Wageningen, Netherlands). The Observer XT- Media Module was used in combination with The Observer XT® Base. This module allows for the playback of one video and the creation of video highlights (Figure 7).

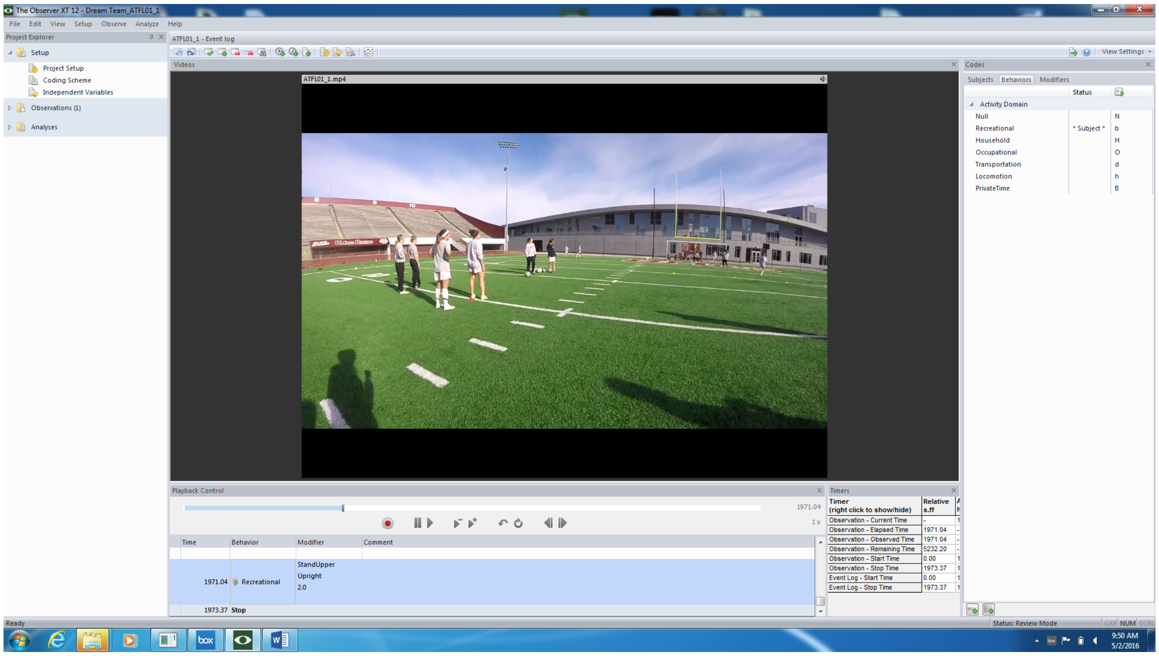


Figure 7. Noldus: The Observer XT

The Observer XT- External Data Module (software module to include physiological or other external data) supports the import and synchronization of data from a wide range of data acquisition systems. The system combines, synchronizes and analyzes accelerometer signals (e.g. ActiGraph GT3X-BT accelerations) with the behavioral data and video (Direct Observation). The Observer® XT has been developed to enable advanced analysis of multimodal data in relation to observational data (Figure 8).

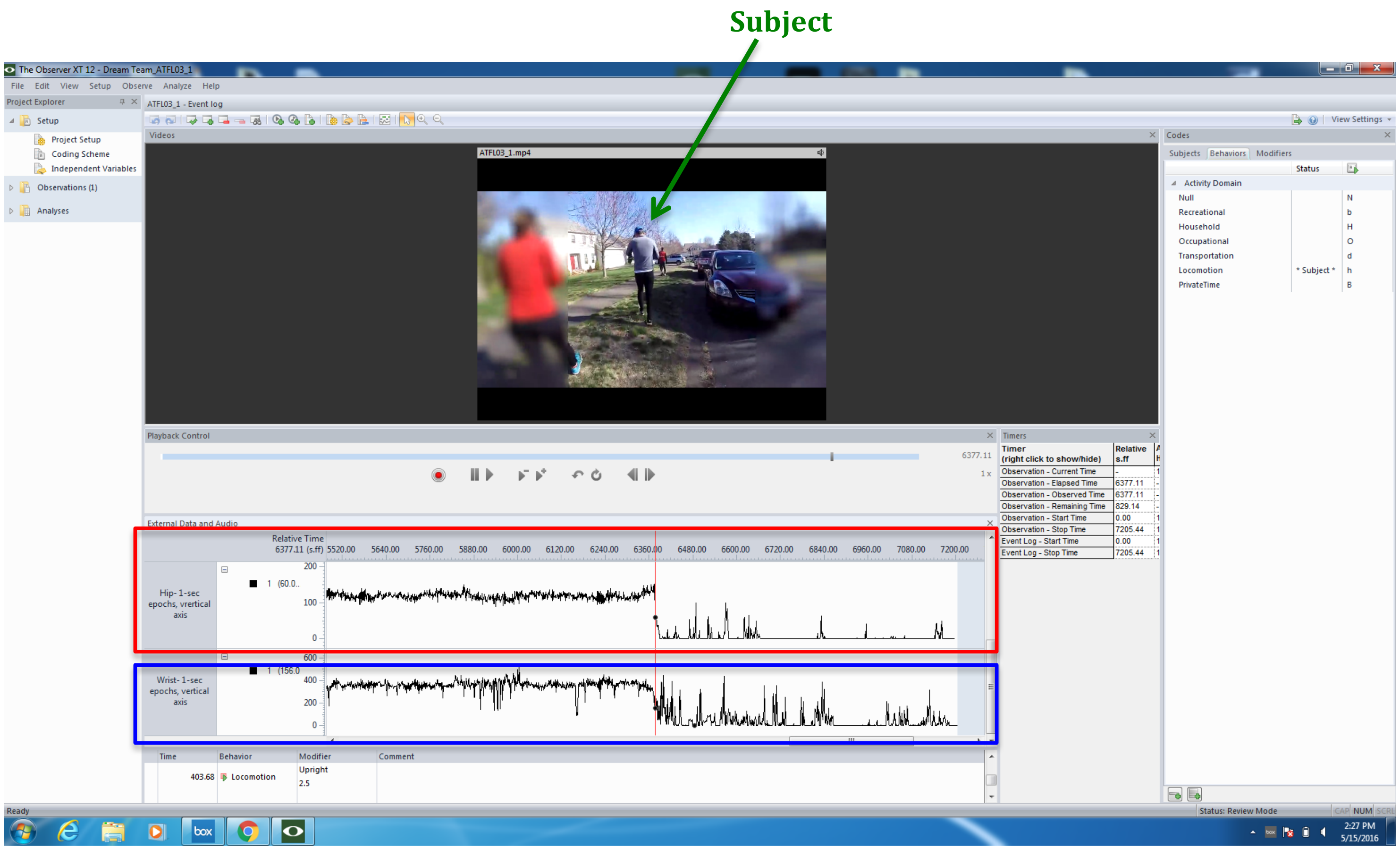


Figure 8. Screenshot from The Observer XT while following one subject
The Observer XT coding synched with imported GT3X-BT accelerometer data from the vertical axis at 1-second epochs for the hip (red box) and wrist (blue box) locations (the GT3X-BT is a triaxial (i.e. vertical axis, anterior-posterior axis and medial-lateral axis) accelerometer). The subject transitioned from a run to a walk and then stretching

Compendium of Physical Activities: Estimation of METs and Kcals from Direct Observation. The Compendium of Physical Activities was developed for use in epidemiologic studies to standardize the assignment of MET intensities in physical activity questionnaires. [144] A MET is defined as the ratio of the work metabolic rate to the resting metabolic rate. One MET is defined as 1 kcal/kg/hour and is roughly equivalent to the energy cost of sitting quietly. A MET also is defined, as oxygen uptake in ml/kg/min with one MET equal to the oxygen cost of sitting quietly, equivalent to 3.5 ml/kg/min. The Compendium has been used in studies worldwide to assign intensity units to physical activity questionnaires and to develop innovative ways to assess energy expenditure in physical activity studies.

Procedures

Visit 1- Informed Consent, Questionnaires, Measurement of Height and Weight

During the informed consent visit, a researcher explained the study and answered any questions. If the potential participant agreed to enroll as a subject, he/she signed the UMass Institutional Review Board approved informed consent document (ICD) (see Appendix B for approved ICD) and completed two questionnaires: 1) physical activity readiness (PAR-Q) and 2) physical activity status (NASA physical activity scale (PAS)) (Appendices C and D). For the PAS, participants were asked to choose a number which best describes their activity during the previous 30 days. Possible responses range from 0 to 7, with 0 corresponding to “avoided walking or exertion (e.g. always used the elevator, drove whenever possible instead of walking)”, and 7 corresponding to “ran more than 10 miles per week or spent over 3 hours per week in comparable physical activity”. Next, participants' height was measured using a standard floor stadiometer and weight was measured using a Tanita scale (DC-430) to the nearest 0.25 cm and 0.1 kg, respectively. Participants then provided demographic information (e.g. ethnicity) and were scheduled for three 2-hr data collection (observation) sessions.

Research-grade accelerometer. Participants were fitted with two ActiGraph GT3X-BT activity monitors. Both GT3X-BT monitors were synced to the same laptop and initialized in advance to collect data at a sampling rate of 80 Hz. They were positioned on the wrist and right hip of each participant. The wrist monitor was secured using a Velcro strap to the non-dominant wrist (positioned over the dorsal aspect of the wrist midway between the radial and ulnar styloid processes), and the hip monitor was secured using a belt at the iliac crest in line with the anterior axilla. The initialization and wrist wear location are consistent with the current National Health and Nutrition Examination Survey (NHANES) activity monitoring study protocol. [145]

StepWatch™ monitor. Participants were fitted with a SW monitor, fastened using a Velcro strap to the dominant ankle (positioned superior to the lateral malleolus). The SW was programmed to record at 3-second intervals, with sensitivity set to 13 and cadence set to 73, consistent with a previous study that our lab conducted.[112] Sensitivity, the magnitude of acceleration that the device qualifies as constituting a step, and cadence, how often the device searches for steps taken.

Activity Trackers. The device placement was counterbalanced across subjects. For example, the total number of devices worn was the same for all subjects but the order in which the devices were placed on subjects was different between subjects. The same placement positions within each participant across the 3 observation sessions (e.g. participant 2, Misfit Flash, left hip for all observation sessions) was used.

Direct Observation

Criterion: Direct Observation. Participants were met by a trained observer in their natural environment (e.g. home, place of work) and observed for approximately two consecutive hours. The GoPro video files were imported into the Behavior coding software The Observer® XT.

Focal sampling and duration coding (FSD) were used to record participant behavior (activity type, body posture, intensity and duration). The FSD method is one where every time a behavior changes (e.g. sitting to standing) the observer recorded the new activity type, body posture and intensity into The Observer XT program. Each entry of a behavior change was time stamped and the duration of each behavior bout was saved. During the two-hour observation time, participants could have "private time" when needed. Reasons for "private time" included behaviors such as using the restroom and changing clothes. During these activities, the observer did not video the participant and the camera was pointed to the ground and recorded as private time in Observer XT.

Direct observation observer training. Training involved research assistants learning how to identify and record activities described in the direct observation (DO) methods (Criterion DO below). The observer training objectives were to: (1) complete an extensive verbal and video training and testing, (2) learn strategies to avoid disrupting free-living behavior, and (3) to accurately record activity type and intensity, all prior to observing participants in a free-living environment. Observers completed DO training that includes review of a training manual, (2) 2-hours of training videos (videos include subjects in free-living settings engaging in a variety of free-living behaviors such as, activities of daily living, locomotion and exercise), and (3) DO practice sessions with the GoPro camera (minimum of 12-hrs of training). After the training, study observers completed a testing video that is ~10 minutes in duration and included different activities with various postures. Before data collection, researchers were required to correctly classify at least 90% (Cohen's kappa coefficient (k) ≥ 0.90) of the body positions, intensity levels, and duration of activities throughout the testing video.

Compendium of Physical Activities. The Compendium of Physical Activities was not developed to determine the precise energy cost of physical activity within individuals, but rather to provide a classification system that standardizes the MET intensities of physical activities used in survey research. The values in the Compendium do not estimate the energy cost of physical activity in individuals in ways that account for differences in body mass, adiposity, age, sex, efficiency of movement, and geographic and environmental conditions in which the activities are performed. [146] Therefore, the Compendium of Physical Activities was used in concert with the preceding coding scheme to estimate physical activity (MET level) in free-living environments. Study observers were extensively trained (previously described) in how to identify physical activity behaviors and their associated MET values within the Compendium of Physical Activities, before observing participants in a free-living environment.

Direct Observation Sessions: Visit 2, 3 and 4. Participants were met by a trained observer in their natural environment (e.g. home, place of work, school) and observed for approximately two consecutive hours. A GoPro video camera was used to record each observation session. Each of the 2-hr observation sessions were done at different times of the day (e.g. Session 1: morning; Session 2: afternoon; Session 3: evening), including one weekend day, in the participants' free-living settings (e.g. home, work, driving). If/when a participant drove; researchers either road with the participant or followed from a safe distance in a separate car. Two researchers were always present during the observation sessions, one videographer and one to take notes, support videographer and drive if needed. For these three visits, researchers initialized devices and met the participants in their free-living environment to be fitted with a variety of activity monitors that were worn on the wrists, hip and ankle, and a biometric shirt. Participants wore 7 monitors on the wrists (4 on one wrist and 3 on the other wrist), 5 monitors on the right and left hip, 1 monitor on the ankle of the dominant leg and 1 biometric shirt worn as an undergarment (Figure 9). The biometric shirt estimates energy expenditure and steps (Table 4).

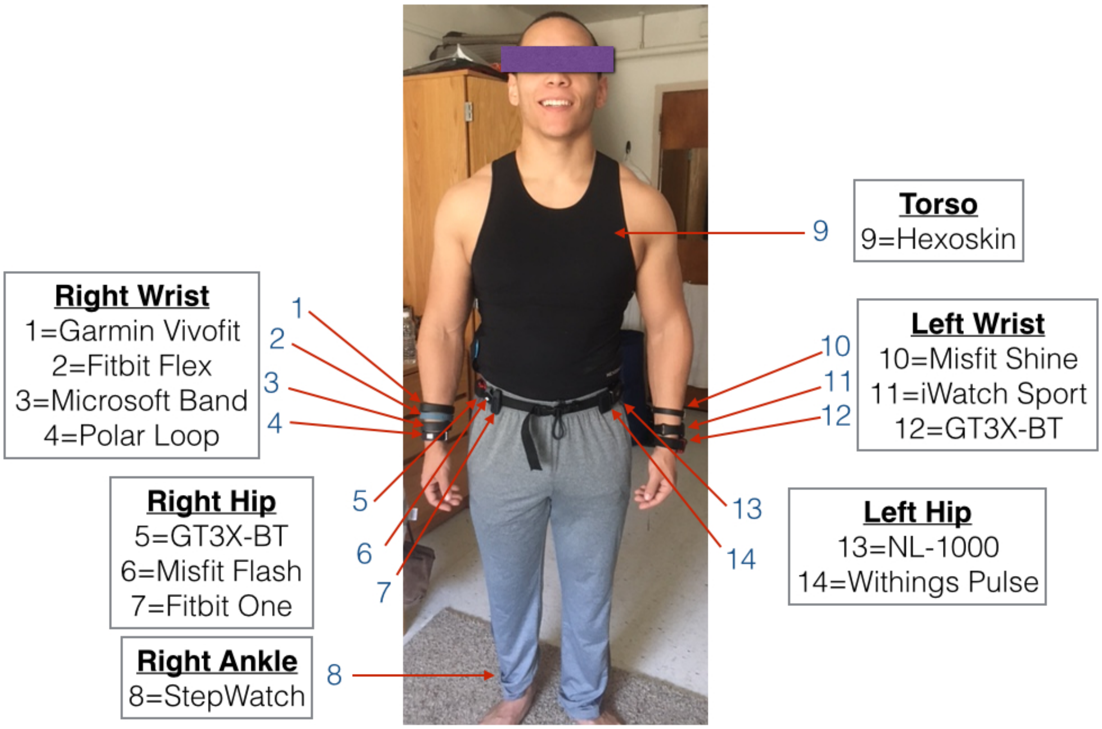


Figure 9. Participant equipped with all devices for observation session

The researchers video recorded participants for the 2-hr sessions while participants performed normal activities (including driving). Every attempt was made to avoid including faces in these video recordings. If any faces appear in the video we edited these shots to blur from the video recording. If private time was required (i.e. going to the bathroom), participants were not observed during these private time periods. At the end of the 2-hr recording period, the researchers recorded the data from the ATs via the real-time display or iPhone app, and removed the activity monitors. Finally, researchers returned to the lab and downloaded data from monitors and video recording.

Other considerations. We expected that some participant' normal activities may bring them to the Recreation Center for individualized workouts or group activity classes. We were sensitive to the privacy of our participants and those persons of the surrounding environment and as a result we had safeguards in place to ensure that privacy was preserved. For example, sound was not recorded, and identities (faces) of all individuals in the video were blurred, thus, individuals are not identifiable. If the participant took part in a group fitness class, we communicated with the instructor, informing her of the purpose of our study and gave a short (~ 60 seconds) explanation to the class and handed out study information (Appendix F).

Data Processing and Statistical Evaluation

Data Processing

All data cleaning, processing and analysis were done using the open source *R statistical software package* (www.r-project.org) and computing language R. [136]

Criterion: Direct observation. For an observation to be included in the analysis, the full 2-hour observation was continuous including private time. Behavior coded, as "private time" were eliminated from analysis.

Focal sampling and duration coding were used, with trained data collectors coding the real-time occurrence (i.e. The Observer XT Media Module synchronized with activity tracker data using The Observer XT External Data Module) of the eight activity categories, body positions, and intensities described below:

1. Lying: Individuals were flat on their backs (horizontal); sedentary (<1.5 METs).

2. Sitting: Individuals had some of their body weight supported by the buttocks or thighs. The upper body was not parallel to the ground. If they kneeled, they were coded based on the thigh position (i.e., if the thigh was parallel to the ground, sitting was selected).
3. Standing still: Individuals were standing with little or no contribution from the upper body. They were not carrying a load >1 kg. Standing still included talking with hand gestures, looking at something, or waiting in a line; sedentary (<1.5 METs).
4. Standing with upper body movement: Individuals were upright with some contribution from the upper body that causes an increase in energy expenditure (holding a load >1 kg, filing papers, or doing a task that required the arms). The purpose of the activity included the upper body; light (1.5 – 2.9 METs).
5. Standing/moving: Individuals were engaging in activities that were of light intensity (1.5 – 2.9 METs); e.g., walking at a speed <2.5 mph and not be carrying a load). These activities included movements around an office or a home but not for locomotion (e.g., traveling between one place and another).
6. Moving moderate: Individuals were engaging in activities (3.0 – 5.9 METs). Examples include walking >2.5 mph, gardening, vacuuming, and carrying a load.
7. Moving vigorous: Individuals were engaging in activities (6.0-8.9 METs). This typically involves purposeful exercise including jogging, walking briskly uphill, and sporting activities.

8. Moving very vigorous: Individuals were engaging in activities >9.0 METs. This typically involves purposeful exercise including running 6 mph, bicycling at 200 watts, and conditioning classes.

A log of the start and stop of each behavior recorded by the observer was exported to a csv file using a custom software and profile (Noldus: The Observer XT 12.5). These data were used to determine criterion measures of activity and inactivity including, MET-hours, MET-minutes (where, 1 MET = 1 kcal/kg/hr), kcals per hour (where, Kcals=METs x time x BW (kg) and time in categories of intensity. The Mifflin-St Jeor equation was used to estimate participant resting metabolic rate (RMR), [147] which has been shown to be valid and reliable in estimating RMR in adults. [148,149] The RMR was added to the EE estimates from the WP and summed to estimate total calories (e.g. exercise + resting).

Total EE was determined by summing/totaling the amount of time spent in all body positions from the DO coding system (e.g. total MET-minutes). METs were then converted to Kcals/minute as recommended by Ainsworth et al.[144]

Kcals= METS x time x body weight in kilograms

Criterion. DO steps were defined as each event when the foot was completely lifted off and lowered back to the ground. To determine criterion step count, steps were manually counted twice for each 2-hour video recording session, and averaged. If there was a 5% difference between total step counts, the video was analyzed a third time and the average of the two closest total step counts was used for analysis (% difference = ((Count 1 – Count 2)/ Count 1) x 100)

Research-grade accelerometers

ActiGraph GT3X-BT. Accelerometer data were downloaded to a laptop using the *ActiLife v6.1.1* software (ActiGraph, Pensacola, FL) and were later extracted to match the corresponding DO time blocks. These data were then be processed to derive total time spent in each activity type and intensity for each participant.

StepWatch™ monitor. StepWatch data were downloaded to the same laptop used for all devices via the StepWatch software (v3.4). Next, the StepWatch data from the observation session were exported and saved for analysis. Total steps were determined by summing/totaling the number of steps taken.

Activity Trackers

Steps, EE, Activity minutes (if provided by activity tracker) and sedentary time (if provided by AT) data from activity trackers were recorded at the beginning and at the end of each observation session. Total estimates were then calculated by subtracting the beginning values from the ending values.

Fitabase (Small Steps Labs, LLC. San Diego, Ca). All Fitbit data were exported using Fitabase. Fitabase is a research platform that acquires data from Internet connected consumer devices. Currently, Fitbit is the only consumer device company that utilizes Fitabase. The advantage of using this platform to acquire Fitbit data is that it provides minute-by-minute data for activity minutes (intensity), kcals, MET-minutes and steps in comparison to the Fitbit software and Dashboard which only provides totals for activity minutes (intensity), kcals, MET-minutes and steps for the monitoring period.

Biometric Shirt. First, the Hexoskin Biometric Shirt data was downloaded to the HxServices Dashboard. Next, an "Activity" was created in the myhexoskin website for the 2-hour observation session (data are time stamped) (Figure 10) and EE and steps were recorded.

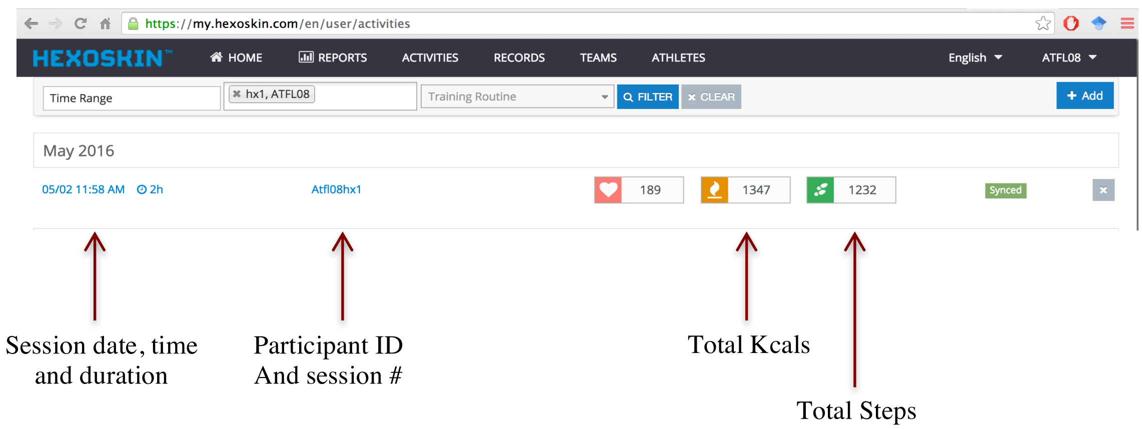


Figure 10. Hexoskin Biometric Shirt activity output

Device (Location)	Output	Data Extraction Method	
		Upload	Retrieval
Apple iWatch Sport (W)	EE, steps, active calories, min: exercise, total active time, stand hours	Bluetooth	Apple Activity App
GT3X-BT (W & H)	Steps, min: Sedentary, light, moderate, vigorous	USB cable	ActiLife
Fitbit Flex /One (W/H)	EE, steps, MET-min, min: sedentary, light, moderate, vigorous	Bluetooth	Fitabase
Garmin Vivofit (W)	EE, steps, active calories, %: sedentary, active, highly active	Bluetooth	Garmin Connect™ App
Hexoskin (T)	EE, steps	USB cable	Hexoskin dashboard
Microsoft Band (W)	EE, steps, active min	USB cable	MB dashboard
Misfit Flash/Shine (H/W)	EE, steps, active min: light, moderate, vigorous	Bluetooth	Misfit App
New Lifestyles NL-1000 (H)	Steps, MVPA min	RTD	RTD
The Observer XT (NA)	MET-hours, MET-min	The Observer XT	The Observer XT
Polar Loop (W)	EE, steps, time: lying, sitting, active, sitting, min: stand, walk, run	USB cable	Polar dashboard
StepWatch (A)	Steps	USB cable	StepWatch dashboard
Withings Pulse (H)	EE, steps	Bluetooth	Withings App

Table 4. Devices with corresponding output and data extraction method
H, hip; W, wrist: T, torso; A, Ankle; NA, not applicable; EE, energy expenditure; min, minutes; MVPA, moderate-to-vigorous physical activity; MB, Microsoft Band; RTD, real-time display.

Video files. Video files were edited (e.g. combined and participants de-identified) using CyberLink PowerDirector 13 Ultra (CyberLink LLC. Boyds, MD) video editing software (Figure 10).

Figure 11. De-identified observation session video

Statistical Evaluation

Statistical Analysis Plan

Accuracy

Bias. We used bias in units of minutes, kcals and steps (activity tracker estimates – criterion) and as a percentage [(mean difference between the activity tracker estimates and the criterion/ criterion x 100].

Precision

We used confidence intervals (CI) as measures of precision. If the upper and lower 95% confidence interval of the bias span 0, then the estimate was not considered significantly different from the criterion at $\alpha = .05$. Higher precision was indicated by higher correlations and smaller CI. Linear-mixed models were used to account for the correlation within subjects, as each subject provides more than one observation (one from each 2 hour session).

Study Three: Activity Trackers are Sensitive to Change in Physical Activity and Sedentary Behaviors in Free-Living Settings

Experimental Procedures

The data used in this study are from our previous study, "Validation of Consumer and Research-Grade Activity Monitors in Free-Living Settings."

Data Processing and Statistical Evaluation

The aims of this study were: 1) to examine the ability of ATs to detect change in PA and ST in free-living settings and 2) to examine the ability of research-grade accelerometer to detect change in PA and ST in free-living settings. Described below are the statistical methods to address this aim.

Data Processing

All data cleaning, processing and analysis were done using the open source *R statistical software package* (www.r-project.org) and computing language R. [136]

Statistical Evaluation

Direct observation provided criterion measures of change in steps, EE, activity minutes and sedentary time

	Visit 1		Visit 2		Visit-to-Visit Change	
	MFS	DO	MFS	DO	MFS	DO
Kcals/2-hrs	416	287	583	325	-167	-38

Table 5. Example of one subject's data for Misfit Shine estimated kcals and DO measured Kcals for visits 1 and 2.
MFS, Misfit Shine; DO, direct observation (criterion Kcals)

For both the criterion measure and the device estimates, we calculated the differences between the visits (i.e. visit 1 minus visit 2, visit 1 minus visit 3 and visit 2 minus visit 3) for estimated steps, EE, activity minutes, and or sedentary time. We then classified the criterion and device measured outcomes for visit-to-visit change into one of three categories: increase, no change or decrease where an increase or decrease was defined as a change that was greater than the within-subject standard deviation of the criterion measure (estimated by a linear-mixed model). Finally, confusion matrices were used to determine percent agreement between criterion visit-to-visit change and device visit-to-visit change. Table 6 illustrates a confusion matrix and percent agreement for DO visit-to-visit change and FBF visit-to-visit change for seven participants.

Percent Agreement = 100

		Fitbit Flex Changes		
		Decrease	No Change	Increase
Direct	Decrease	7		
Observation	No Change		3	
Changes	Increase			11

Table 6. Confusion matrix and percent agreement change in energy expenditure between sessions (session 1 – session 2, session 1 – session 3, session 2 – session 3) for seven participants

CHAPTER 4

STUDY ONE – A COMPARISON OF CONSUMER ACTIVITY TRACKER ACCELEROMETER OUTPUT AND A RESEARCH-GRADE ACCELEROMETER OUTPUT DURING ORBITAL SHAKING

Introduction

Electronic testing of research-grade accelerometers has provided valuable information about device performance during controlled accelerations at different frequencies. This information has been essential for researchers to gain a more comprehensive understanding of the strengths and weaknesses of accelerometers in highly controlled testing conditions using electronic testing systems. The ActiGraph (AG)(ActiGraph, LLC, Pensacola, FL) accelerometer provides an objective estimate of human physical activity (PA) and is used in many research studies and in clinical settings. [50-52] The ActiGraph GT3X-BT detects a wide range of accelerations and samples acceleration from 30 to 100 Hz. Standardized electronic validation and reliability testing of the AG has been performed using the GT3X+, GT3X, GT1M, 7164 and 71256. In these studies, electronic devices such as wheels, [53,54] a table, [55] and orbital shaking [56-58] were employed to examine accelerometer output at fixed frequencies. From these studies, we have advanced our understanding of differences in accelerometer counts, steps, and raw acceleration as a function of model, band-pass filter methods, sampling frequency, accelerometer type (piezoelectric versus solid state microelectromechanical systems (MEMS)) and firmware.

To date, our laboratory has published the only study that has applied electronic testing techniques to examine the accelerometer output of consumer activity trackers (ATs).[150] The benefits of electronic orbital shaker testing are that it allows us to: (1) expose ATs to different oscillation frequencies to simulate different movement intensities and (2) vary oscillation frequencies to simulate variation in free-living whole-body acceleration. The electronic orbital shaker will inform researchers of how ATs perform under highly controlled conditions. Orbital shaker testing removes human variation from the testing environment. As a result, observed differences would be due to technological features of the devices that are not impacted by human variation. Our lab employed an electronic orbital shaker to assess the data of several consumer ATs compared to the AG GT3X+ accelerometer (unpublished observations). [62] We found that AT data was highly correlated with oscillation frequency (*r* range: 0.92 to 0.99).

Electronic testing of ATs is a necessary first step in building a scientific knowledge base of these increasingly popular devices. Therefore, the purpose of the present study was to compare consumer ATs with the research-grade ActiGraph™ GT3X-BT accelerometer in estimating energy expenditure (EE) and steps during orbital shaking at different frequencies. We hypothesized that EE and step estimates from consumer ATs would be similar to the EE and step estimates of the research grade GT3X-BT accelerometer during standardized testing using an electronic orbital shaker.

<u>Methods</u>

Instrumentation

<u>Research-grade accelerometer: Reference Standard.</u> The ActiGraph GT3X-BT (GT3X-BT) accelerometer (ActiGraph™ LLC, Pensacola, Florida) served as the

reference standard to which all ATs were compared. This device is a lightweight triaxial PA monitor (4.6cm x 3.3cm x 1.5cm, 19g) that measures acceleration ranging from -8 to +8 g's. Data were collected at a sample rate of 80 Hz and post-processed in the ActiLife software version 6.1.3 to 60-second epochs.

<u>Activity Trackers</u>

Activity trackers were chosen based on three criteria: (1) no known gravimeter within the device, (2) researchers had at least two of the device, and (3) the device fit in the cushioned slots of the shaker. The rationale for excluding ATs that contain a gravimeter was that the electronic orbital oscillator does not produce vertical accelerations and as a result, a device that contains a gravimeter would produce inaccurate data, as it would detect no change in gravitational position. The rationale for at least two devices was to counterbalance each other in the electronic orbital oscillator. Six different ATs were studied: 1) Fitbit Flex (FBF)(Fitbit® Inc., San Francisco, California), 2) Fitbit One (FBO)(Fitbit® Inc., San Francisco, California), 3) Garmin® Vivofit (GV)(Garmin Ltd., Schaffhausen, Switzerland), 4) Misfit Flash (MFF)(Fossil Group, Inc.), 5) Misfit Shine (MFS)(Fossil Group, Inc.), and 6) New Lifestyles NL-1000 pedometer (NL)(New Lifestyles, Inc., Lee's Summit, Missouri). See table 7 for detailed specifications of each activity tracker.

Device	Fitbit Flex	Fitbit One	Garmin vívofit	New Lifestyles NL-1000	Misfit Flash	Misfit Shine
Cost	$39.95	$99.95	$99.99	$54.95	$19.99	$39.95
Wear location	Wrist	Clip on (multiple locations)	Wrist	Hip	Clip on (multiple locations)	Clip on (multiple locations)
Tracks Calories Burned	✓	✓	✓	✗	✓	✓
Tracks Active Time	✓	✓	✓	✓	✓	✓
Tracks Steps	✓	✓	✓	✓	✓	✓
Tracks Distance	✓	✓	✓	✓	✓	✓
Tracks Elevation/Stairs	✗	✓	✗	✗	✗	✗
Tracks Sleep	✓	✓	✓	✗	✓	✓
Tracks Heart Rate	✗	✗	✗	✗	✗	✗
Battery or Chargeable	Chargeable (every 5 days)	Chargeable (every 10+ days)	Battery (every 1+ years)	Battery (up to 18 months)	Battery (lasts up to 6 months)	Battery (lasts up to 6 months)
Uploading Data	Bluetooth	Bluetooth	Bluetooth	Real-time data	App	App
Tracker Display	LED progress indicator	Real-time data	Real-time data	Real-time data	LED progress indicator	LED progress indicator

Table 7. Features of consumer-based activity trackers
LED, Light-Emitting Diode; USB, Universal Serial Bus; App, application

<u>Electronic Orbital Shaker.</u> The electronic orbital shaker (Advanced Orbital Shaker, Model 10000-2; VRW International, Radnor, PA) (Figure 1) produces controlled oscillations between 0.25 and 5.0 Hz. The electronic orbital shaker oscillates at various radii between 1.27 and 5.7 cm. Four trays (51 x 10 x 10 cm) are mounted on the base oscillating plate (60 x 60 cm) of the shaker. Each tray has four custom foam cushion slots to securely hold the GT3X-BTs and ATs in place to eliminate device movement during electronic orbital shaking (Figure 12).

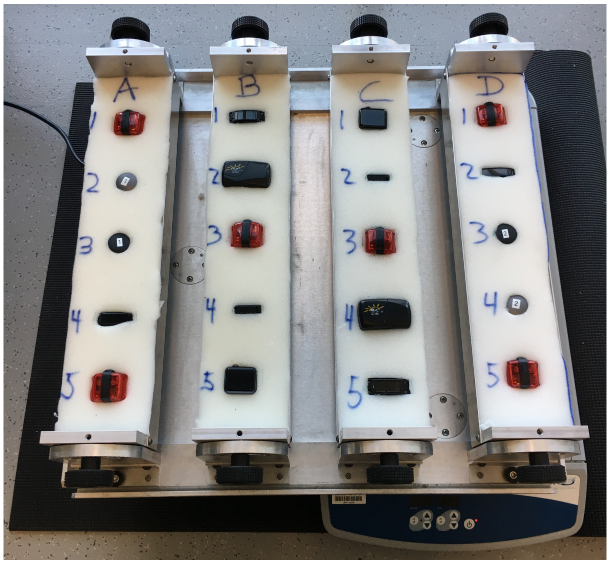


Figure 12. Electronic orbital shaker with devices in custom foam cushioned slots

Procedures

Electronic Orbital shaker. The electronic orbital shaker was used to perform motion testing. Two of each device were tested at the same time. All devices were placed in the custom foam cushion slots with their vertical plane perpendicular to the control panel of the electronic orbital shaker (figure 12).

The GT3X-BTs and ATs were oscillated using two protocols, (1) 3-minute trials, and (2) 2-hour trials. Each 3-minute trial consisted of one monitor oscillation frequency (e.g. 0.9 Hz). Oscillation frequencies were increased from zero, 0.25 to 3.0 Hz in 0.1 Hz increments for a total of 24-trials. The step-wise increase in frequency allowed researchers to test the effect of specific frequencies on device output. Each protocol (twenty-four, 3-min trials; 2-hour trial) was repeated three times. The 2-hour trials consisted of oscillation frequencies ranging from zero to 3.0 Hz., based on the American Time Use Survey (ATUS) percentages of time spent in selected activities, normalized for 2-hours, [135] to simulate free-living whole body acceleration (e.g. variation). These frequencies simulate hip rotation ranging from no movement (e.g. sleep) to ambulation at speeds ranging between 1.5 and 16 miles per hour. [81] Two-hour blocks were chosen as it was not feasible to test devices for 24-hours consecutively. All shaker oscillations were performed on a fixed radius [56,134] of 5.08 cm.

Oscillation Frequencies. Oscillation frequency ranges for each activity category were established by electronically oscillating six GT3X-BTs at 0.0, 0.25 to 3.0 Hz in 0.1 Hz increments and applying cutpoints. Each 0.1 Hz. increment was oscillated for 3-minutes and the second minute of each trial was used to determine counts per minute at each frequency. Using the second minute ensured that the desired frequency was achieved for the entire minute. The GT3X-BT data were collected at 80 Hz., with the low frequency extension for oscillation frequencies <0.7 Hz., post processed using ActiLife software (v 6.1.3) and aggregated into VM counts per minute. These data were scored in ActiLife using the Freedson VM3 cut-points. [52] Lastly, the intensity categories and their associated frequencies were used to determine the 2-hour electronic oscillation trial: frequency, intensity and total time.

Data Collection and Processing. The GT3X-BTs were synched to the same laptop as the ATs and initialized in advance of data collection (sampling rate of 80Hz). These raw data were post processed into 1-second epochs/counts and steps via ActiLife v6.1.3 software.

Minute-by-minute EE (kcals) was estimated and summed for all 3-minute trials and for each 2-hour trial using the prediction equation previously developed by our group, [52] labeled the "Freedson VM3 (2011)" equation in the ActiLife software. The Freedson VM3 equation has been validated in classifying PA intensity.[52]

The same user profile (e.g. weight in kg) was used for the GT3X-BTs and ATs. The low frequency extension (LFE) option was selected in the ActiLife software to detect lower amplitude movements. The LFE option lowers the baseband of the filter cut-off, expanding the bandwidth of the accumulated data. The LFE was selected to ensure acceleration detection at slower oscillation frequencies (e.g. 0.7 Hz).

<u>Activity Trackers</u>

Pre-3-minute oscillation trials and 2-hour oscillation trial. Thirty-minutes prior to the first 3-minute and the 2-hour oscillation trial, all activity trackers were initialized/synched using the same user profile (e.g. date of birth, gender, height and weight) and the same computer was used as was used to initialize the GT3X-BTs. Next, the GT3X-BT and ATs (FBF, FBO, MFF, MFS, GV and NL) were secured into their respective customized foam cushion slots within each tray of the electronic orbital shaker (Figure 12). Two of each device were tested in the electronic orbital shaker.

Immediately prior to each 3-minute oscillation trial and each 2-hour oscillation trial, all Misfit data were retrieved via the Misfit app (iPhone 6s) and baseline step values for the MFS and MFF were recorded as neither device is equipped with a real-time display. Steps from the GV were recorded from the real-time display. The NL pedometers were set at 0 steps. The values for EE and steps from the FBF, FBO and the GT3X-BT were retrieved and recorded pre-and post each 3-minute oscillation trial and each 2-hour oscillation trial. The start and stop time for each 3-minute oscillation trial and each 2-hour oscillation trial were synchronized with the time of the same laptop used for initialization/synching and downloading of all devices.

Data Processing. Following each 3-minute and 2-hour oscillation trial total steps for the: 1) MFF and MFS were downloaded via Bluetooth and retrieved via the Misfit app (iPhone 6s), 2) FBF and FBO were synched/downloaded to the Fitbit Dashboard via Bluetooth and retrieved from the Fitabase website (described below), and 3). Garmin Vivofit were retrieved from the real-time display. Total steps for the NL pedometer were retrieved from the real-time display. Total EE for the FBF and FBO were synched/downloaded to the Fitbit Dashboard via Bluetooth and retrieved from the Fitabase website (described below). The GT3X-BT data were collected at 80 Hz, with the low frequency extension for oscillation frequencies <0.7 Hz (3-minute oscillation trials only), post processed using ActiLife software (v 6.1.3) and aggregated into VM counts per minute. Total estimated kcals for each 3-minute and 2-hour oscillation trial were calculated and summed employing the "Freedson VM3 (2011)" equation in Actilife (v 6.1.3). Total steps from the GT3X-BT were obtained by summing: 1) each 3-minute oscillation trial, and 2) each 2-hour oscillation trial.

Fitabase (Small Steps Labs, LLC. San Diego, Ca). All Fitbit data were exported using Fitabase. Fitabase is a research platform that accesses data from Internet connected consumer devices. The advantage of using this platform to acquire Fitbit data is that it provides minute-by-minute data for activity minutes (intensity), kcals, MET-minutes and steps in comparison to the Fitbit software and Dashboard which only provide total activity minutes (intensity), kcals and steps for the monitoring period.

Statistical evaluation

All data cleaning, processing and analysis were done using the open source *R statistical software package*, version 3.3.3 (www.r-project.org) and computing language R. [136]

Data Analysis. To evaluate AT estimates, we used two statistical tools: bias (mean difference between the estimate and the reference) provides information about the accuracy of the estimate and if the upper and lower 95% confidence intervals of the bias span 0, then the estimate is not significantly different from the reference at $\alpha = .05$. Linear mixed effects models assessed main effects of device and frequency and random effects of trial on activity tracker estimates of EE and steps compared to GT3X-BT estimates of EE and steps. Significance level was set at $\alpha = .05$.

Results

Figure 13 shows steps per 3-minutes during electronic oscillation. The NL was not significantly different from the GT3X-BT beginning at 0.9 Hz (corresponding to moderate intensity PA). The largest difference between the NL and the GT3X-BT was 142 steps/3-min at 0.8 Hz (corresponding to moderate intensity PA). All other AT step estimates were significantly different than GT3X-BT steps. For the MFF, the largest difference was 285 steps/3-min at 1.5 Hz (corresponding to very vigorous intensity PA). However, these differences were smaller beginning at 2.4 Hz (44 steps/3-min) (corresponding to very vigorous intensity PA). For the MFS, the largest difference was

102 steps/3-min at 1.5 Hz (corresponding to very vigorous intensity PA). However, these differences were smaller beginning at 2.4 Hz (44 steps/3-min) (corresponding to very vigorous intensity PA). For the FBO, the largest difference was -310 steps/3-min at 2.4 Hz (corresponding to very vigorous intensity PA). However, these differences were smaller beginning at 2.5 Hz, with the smallest difference at 2.6 Hz (-264 steps/3-min) (corresponding to very vigorous intensity PA). For the FBF, the largest difference was -385 steps/3-min at 2.4 Hz (corresponding to very vigorous intensity PA). However, these differences were smaller beginning at 2.5 Hz (corresponding to very vigorous intensity PA), with the smallest difference at 2.6 Hz (-317 steps/3-min) (corresponding to very vigorous intensity PA). The GV detected no steps at all frequencies tested.

Figure 14 shows energy expenditure per 3-minutes during electronic oscillation. Energy expenditure estimates from both the FBO and FBF were significantly different than GT3X-BT estimates of EE. For the FBO, the largest difference was -35 kcals/3-min at 2.4 Hz. However, these differences were smaller beginning at 2.5 Hz, with the smallest difference at 2.6 Hz (-34 kcals/3-min). For the FBF, the largest difference was -39 kcals/3-min at 2.3 Hz. However, these differences were smaller beginning at 2.5 Hz, with the smallest difference at 2.9 Hz (-31 kcals/3-min).

Figure 15 shows steps per 2-hours during electronic oscillation. Average steps for the GT3X-BT were, 5831, 5178 and 6301 steps/2hr. for trials 1, 2 and 3, respectively. On average, the FBO and the GV significantly underestimated steps for all trials. These underestimations ranged from -6200 to -4200 steps/2-hrs. On average, the FBF underestimated steps with two trials significantly different than the GT3X-BT. The NL significantly underestimated one trial compared to the GT3X-BT. In contrast, the MFS and MFF significantly overestimated steps for two trials compared to the GT3X-BT. These overestimations ranged from 50 to 2,200 steps per 2-hrs.

Figure 16 shows energy expenditure per 2-hours during electronic oscillation. Average kcals for the GT3X-BT were, 601, 508 and 681 for trials 1, 2 and 3, respectively. The FBF, FBO and the GV significantly underestimated kcals for all trials. These underestimations ranged from -580 to -65 kcals/2-hrs. In contrast, the MFS and MFF significantly overestimated kcals for two trials and significantly underestimated kcals for one trial compared to the GT3X-BT. Average overestimations ranged from 105 to 190 kcals/2-hrs. Average underestimations ranged from -160 to -170 kcals per 2-hrs.

Figure 17 illustrates the relationship between hertz and acceleration. Note that the relationship between hertz and counts differs, as the relationship is curvilinear starting at 2.5 Hz (see Figure 5).

Discussion

Currently, no published studies have examined ATs during electronic shaker testing. Therefore, in this discussion interpretation it is necessary to compare our results to human studies. There is evidence that electronic oscillation of the GT3X simulates hip rotations similar to ambulation at speeds ranging between 1.5 and 16 miles per hour. [81] Therefore, to provide meaning, and context to the present study's findings, the following discussion will present evidence from validation studies comparing Fitbit, Garmin, Misfit and NL-1000 estimates of steps and/or EE to criterion measures during lab-based treadmill walking and running.

The purpose of the present study was to compare consumer ATs with the ActiGraph™ GT3X-BT accelerometer in estimating EE and steps during orbital shaking at different frequencies. To address this question, two protocols employing an electronic orbital shaker were developed: a 3-minute trial at specific frequencies and a 2-hr trial at various frequencies.

Our main findings from the 3-minute protocol were that for steps, the NL was not significantly different from the GT3X-BT beginning at 0.9 Hz and held constant through 3.0 Hz. Previously, our group exposed GT3X-BTs to electronic oscillation frequencies from zero to 3.0 Hz in 0.1 Hz increments and applied the widely used Freedson (VM3) cut-points to categorize frequencies into corresponding intensity levels. We found that 0.9 Hz elicits GT3X-BT VM counts corresponding to moderate intensity PA. Additionally, 0.6 to 0.8 Hz corresponds to the change from light to moderate intensity PA, suggesting, that the NL may be less sensitive to sedentary to light PA as compared to moderate, vigorous and very vigorous PA. All other AT step estimates were significantly different than GT3X-BT steps. Another finding was step estimates from devices of the same company displayed similar trends. For example, both Misfits produced the largest and smallest errors at 1.5 and 2.4 Hz, respectively. The Fitbits, produced the largest and smallest errors at 2.4 and 2.6 Hz, respectively. In contrast, the GV detected no steps. According to the Garmin website, the Vivofit only possesses one sensor: an accelerometer.[151] However, we posit that this device utilizes a gravimeter, which continuously identifies true vertical axis. The electronic orbital shaker oscillates in the horizontal versus the vertical plane. These findings elucidate technological differences between ATs. For example, ATs employ triaxial accelerometers in concert with user information, band-pass filters, firmware, and proprietary algorithms to estimate PA behaviors, such as, steps. For researchers, this "black-box" method of deriving PA behavior estimates remains a challenge.

Main findings from the 3-minute protocol were that for EE, both the FBO and FBF were significantly different that GT3X-BT estimates of EE. The largest and smallest differences ranged from 2.4 to 2.9 Hz. This frequency range corresponds to very vigorous intensity PA. This suggests that estimates of EE from the FBO and FBF may not be comparable to EE estimates from the GT3X-BT derived via the Freedson EE equation. Moreover, these findings strengthen the argument that proprietary algorithms may be a primary cause of observed differences in AT data compared to GT3X-BT data.

Our main findings from the 2-hour protocol were that for steps, the NL produced the smallest bias (bias for all trials = -570 steps/per 2-hrs), and two of three trials were not significantly different than our reference measure (GT3X-BT). Our findings of the relationship between NL steps and GT3X-BT steps are consistent with the those of Abel et al. [152] Briefly, 59 participants performed treadmill walking and running at speeds ranging from 2.2 to 4.0 mph while wearing the NL on the waist. They reported that, the NL and AG 7164 yielded the most accurate step counts at a range of walking speeds in individuals with different physical characteristics. Next, we found that, the FBO and the GV significantly underestimated steps for all trials. Bias for all trials was -5120 and -

5770 for the FBO and GV, respectively. Additionally, the FBF significantly underestimated steps for two of the three trials. Bias for all trials was -1651 steps/2-hrs. Lab-based validation studies have provided evidence that, in general, FBF and FBO underestimate steps with varying precision compared to criterion measures during treadmill walking and running. [20,25,26,28-30] Our finding that GV significantly underestimated steps, is supported by Chen et al. [20] who showed that the GV significantly underestimated steps compared to DO during treadmill walking and running at speeds ranging from 2.0 to 5.0 mph ($p<.05$). Lastly, we found, both Misfits significantly overestimated steps for two trials. Bias across all trials was 1,921 and 1,332 steps/2-hrs, for the MFF and MFS, respectively. Two studies have examined the MFS during treadmill walking and running and results are equivocal. First, Kooiman et al. [21] examined the MFS during treadmill walking at 2.0 mph for 30-minutes compared to criterion steps (Optogait system) and reported a bias (SD) of -6(43) steps. Chen et al. [20] validated the MFS during treadmill walking and running (speed range: 2.0 to 5.0 mph) The MFS significantly underestimated steps at all speeds compared to criterion steps ($p<.05$), however, accuracy improved at higher speeds. These data provide evidence that, the NL produced the smallest error compared to the GT3X-BT during 2-hours of electronic orbital shaking at frequencies ranging from zero to 3.0 Hz, which highlights potential issues with comparing step estimates from ATs.

Main findings from the 2-hour protocol were that for EE, the FBF, FBO and the GV significantly underestimated kcals for all trials (range: -580 to -65 kcals/2-hrs) compared to the GT3X-BT. Several validation studies support and refute this finding. For example, Price et al. [153] examined FBO (hip-worn) and GV (wrist-worn) EE estimates during treadmill walking (1.5, 2.8 and 4.0 mph) and running (5.0, 6.4 and 7.4 mph) compare to indirect calorimetry. They reported, EE estimates from the FBO and GV correlated significantly ($p< 0.01$; $r= 0.702$; 0.854) with criterion across all gait speeds (1.5 - 7.4 mph). However, EE estimations of single speeds were overestimated by the FBO and underestimated by the GV. Further, EE estimations of single speeds were overestimated by the FBO and underestimated by the GV. One reason for these differences may have been the result of device location. Specifically, the FBO was hip-worn and the GV was wrist-worn. Our laboratory compared AG GT3X+ wrist and hip accelerations (g's) at different locomotion speeds (unpublished). We found a significant difference between the slopes (m) (speed vs vector magnitude (VM)) for the hip, (m = 0.052 [95% CI: 0.033, 0.103] compared to the wrist, m = 0.195 [95% CI: 0.160, 0.230], $p<0.001$), and concluded that the pattern of change is different and more variable between subjects for the wrist VM. The FBF has also been shown to significantly overestimate EE during treadmill walking and running compared to indirect calorimetry ($p<.05$). [25] The Fitbit One has demonstrated both significant overestimation- [26,30] and underestimation [27] of EE during treadmill locomotion ($p<.05$). For these studies, participant populations and protocols were similar. E.g. healthy adult, age range: 19 – 41 years and treadmill locomotion, respectively. Lastly, we found that the MFS and MFF significantly overestimated kcals (range: 105 to 190 kcals/2-hrs) for two trials and significantly underestimated kcals (range: -160 to -170 kcals/2-hrs.) for one trial compared to the GT3X-BT. Currently, no studies have examined EE estimations from Misfits during treadmill locomotion, only.

Energy expenditure estimates from the FBO, FBF, MFS, MFF and GV during electronic orbital shaking and EE estimates from human, treadmill studies illustrate the need for further investigation into possible origins of device differences.

This study has several strengths. First, electronic orbital shaker testing removes human variation. As a result, we are confident that observed differences are due to technological features of the devices and not impacted by human variation. Second, ATs were tested over a wide range of frequencies, which allowed identification of exact frequencies where differences were present. This information may be valuable to both consumers and manufacturers, depending on their needs. For example, a consumer may seek to use a device that can detect steps while walking at a moderate pace. A manufacturer may choose to adjust filters and/or algorithms to allow step detection at lower oscillation frequencies (e.g. 0.7 Hz). Lastly, we employed a widely used, valid, and reliable, research-grade accelerometer as our reference.

This study has several limitations. First, the ATs that were tested in the present study are made to be worn by people. It is possible that these devices possess algorithms, and/or filters to detect, and remove artificial human movement (i.e. electronic orbital shaking). Thus, AT data would differ from our reference. Another limitation is that step estimates and EE estimates from Freedson VM3 equation were developed via human-testing. Though studies have provided evidence that sensor output is often calibrated during standardized activities such as walking on a treadmill, [154] applying the same algorithm to electronic oscillations may be inappropriate. Lastly, we did not include estimated EE from all ATs. Currently, Fitbit is the only AT company that provides minute-by-minute data via the research platform Fitabase. For all other ATs, the exact time between 3-minute trials could not be determined. As a result, it was not possible to compare EE estimates from these ATs to our reference for the 3-minute protocol.

In conclusion, this study provides the first evidence of AT estimates of steps and EE compared to the GT3X-BT during electronic shaking. Our main findings were that, on average, the NL produced the smallest error. All other ATs were equivocal, with some overestimating steps or EE, and others underestimating steps or EE compared to the GT3X-BT. This study is a first step toward a more comprehensive understanding of AT estimates of steps and EE during electronic shaker. More research is needed to identify specific causes for these differences so to improve the accuracy and precision in AT estimates of steps and energy expenditure.

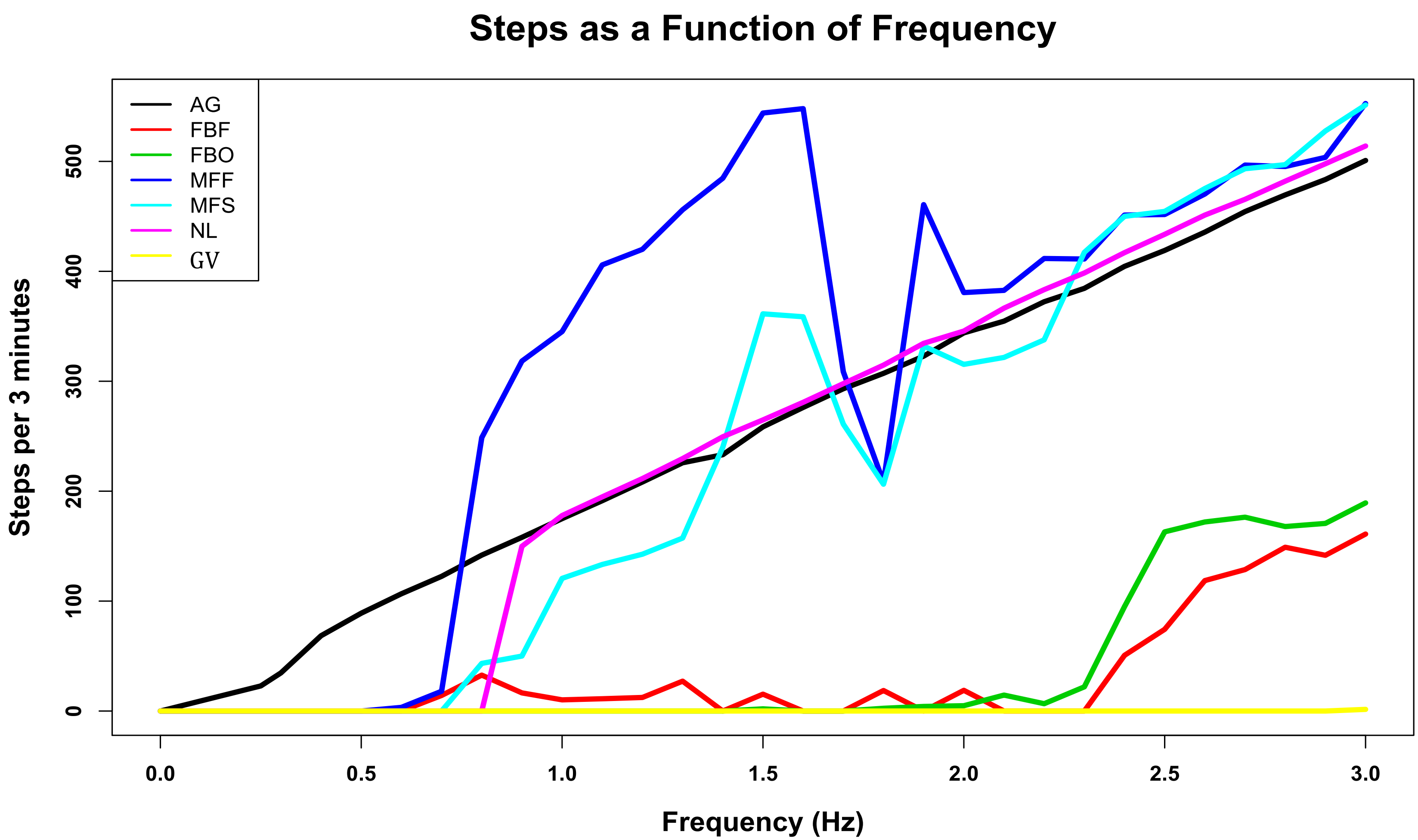


Figure 13. Steps per 3-minutes during electronic oscillation
AG, ActiGraph wGT3X-BT; FBF, Fitbit Flex; FBO, Fitbit One; MFF, Misfit Flash; MFS, Misfit Shine; NL, NL-1000 pedometer; Vivofit, Garmin Vivofit.

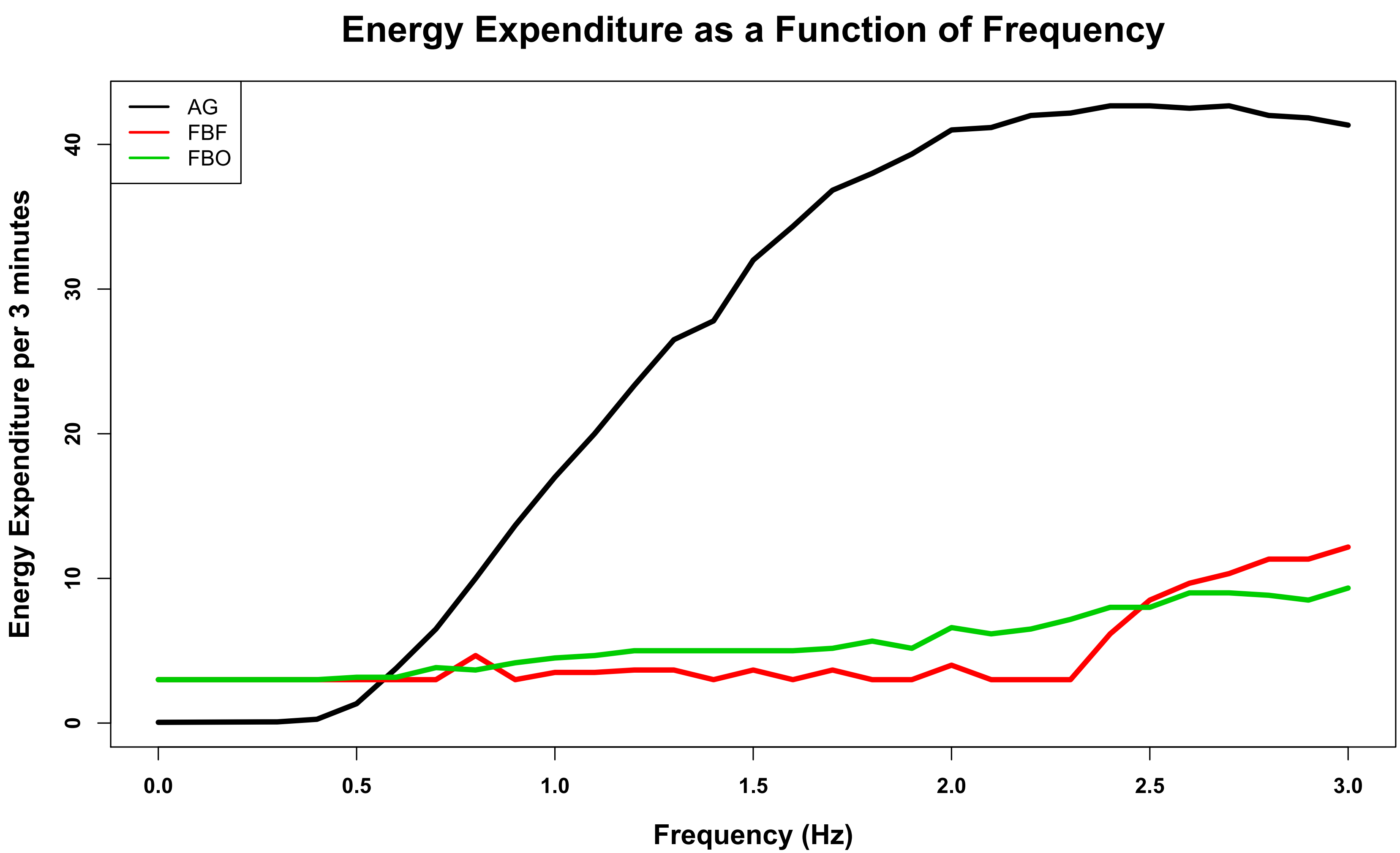


Figure 14. Energy expenditure per 3-minutes during electronic oscillation
AG, ActiGraph wGT3X-BT; FBF, Fitbit Flex; FBO, Fitbit One

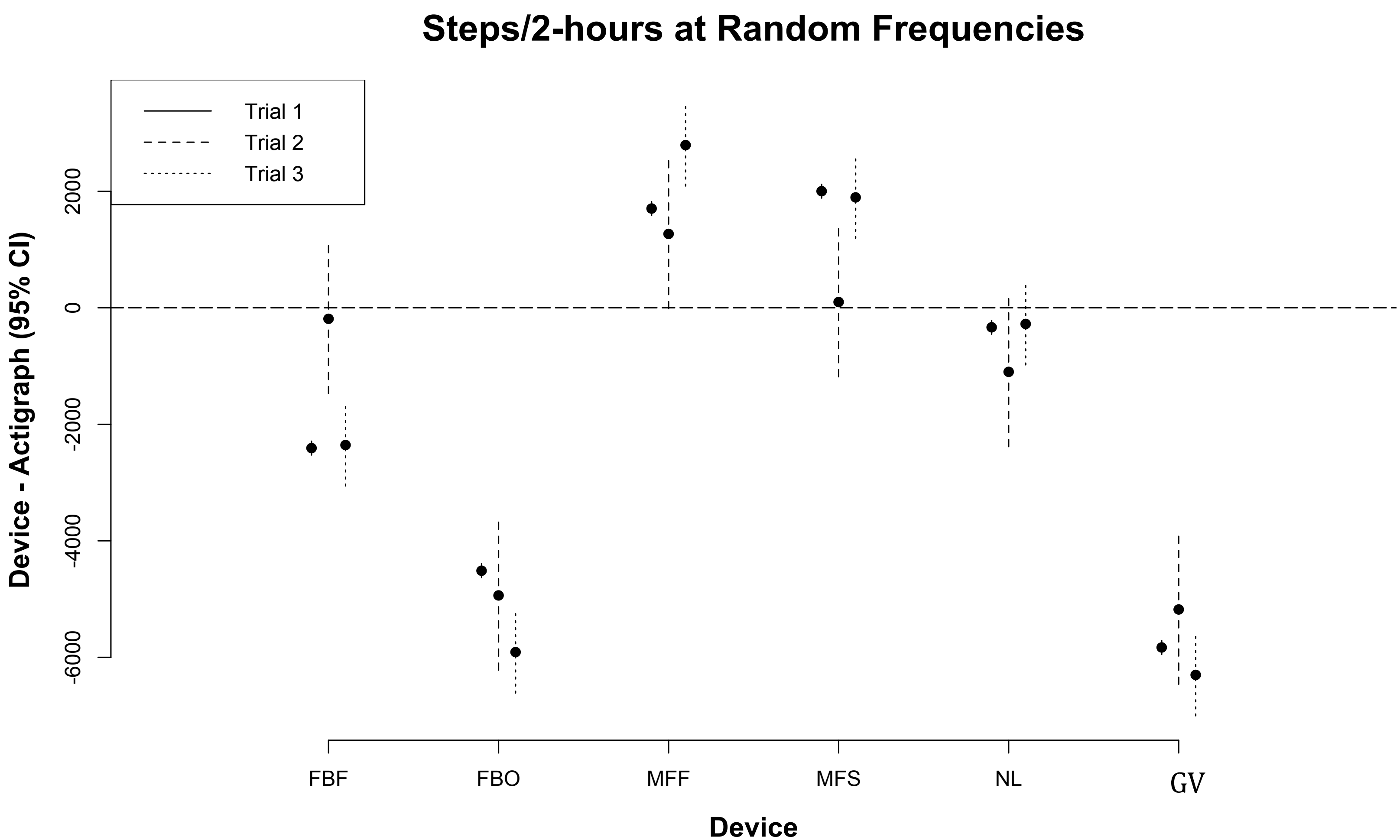


Figure 15. Steps per 2-hours during electronic oscillation
Data presented as mean and 95% confidence intervals
FBF, Fitbit Flex; FBO, Fitbit One; MFF, Misfit Flash; MFS, Misfit Shine; NL, NL-1000 pedometer: GV, Garmin Vivofit.

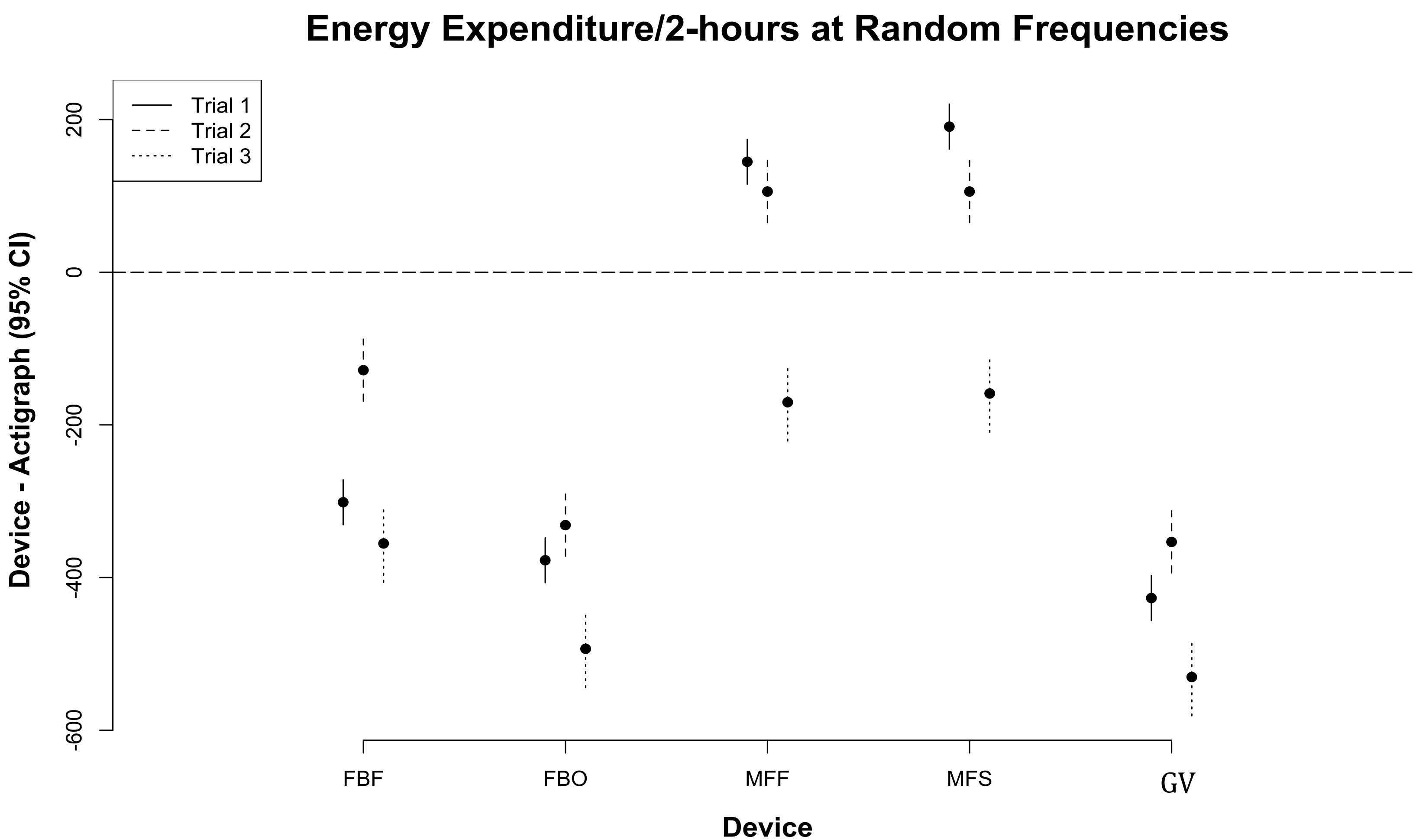


Figure 16. Energy expenditure per 2-hours during electronic oscillation
Data presented as mean and 95% confidence intervals
FBF, Fitbit Flex; FBO, Fitbit One; MFF, Misfit Flash; MFS, Misfit Shine; NL, NL-1000 pedometer: GV, Garmin Vivofit.

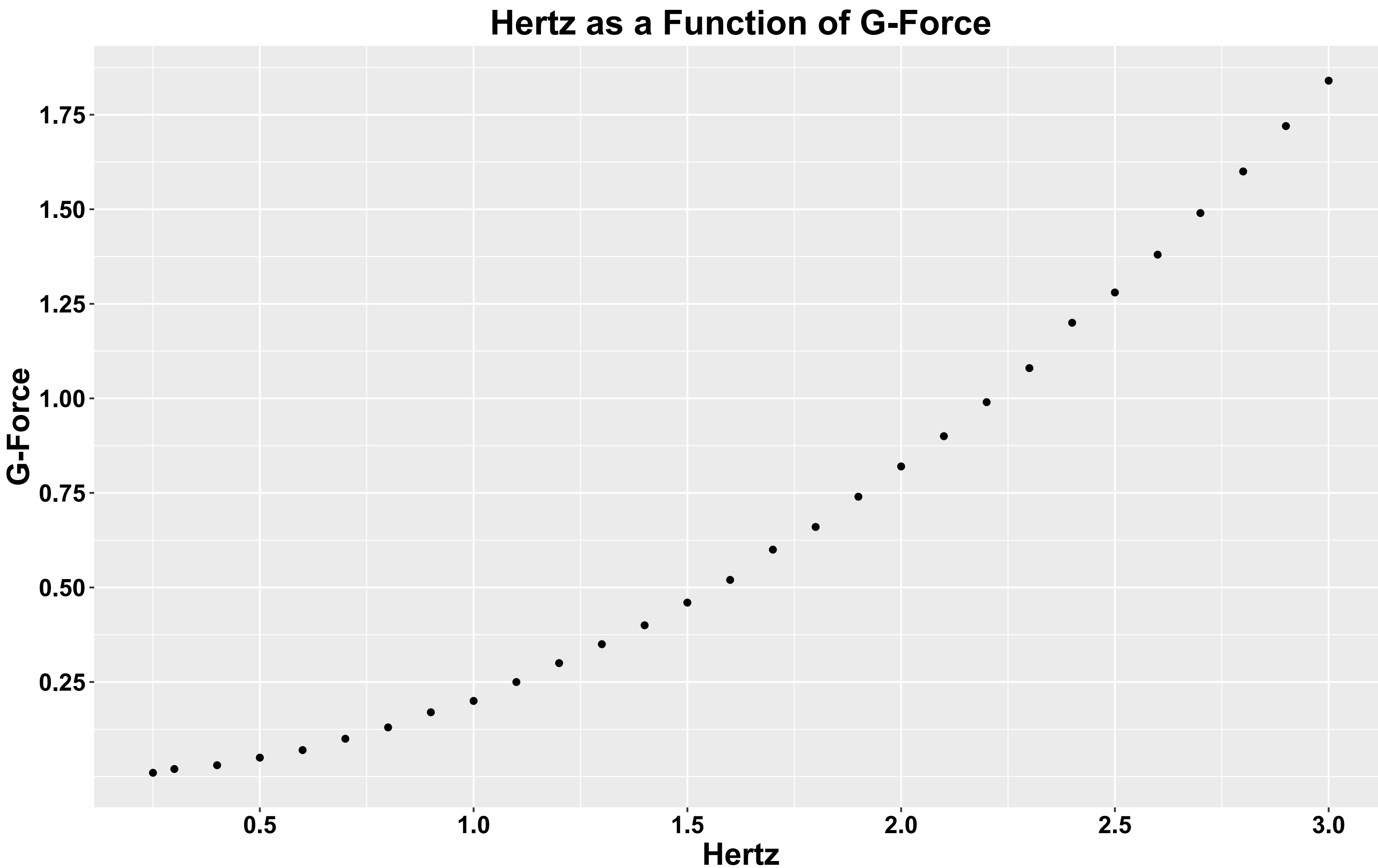


Figure 17. Hertz as a function of acceleration (g's)

CHAPTER 5

STUDY TWO – VALIDATION OF CONSUMER AND RESEARCH-GRADE ACTIVITY MONITORS IN FREE-LIVING SETTINGS

Introduction

Lack of physical activity (PA) is strongly implicated in virtually all leading causes of chronic disease morbidity and mortality. To attenuate the prevalence of these preventable chronic diseases and promote health benefits, the U.S. Government recommends that Americans accumulate at least 10,000 steps/day,[3] increase daily expenditure approximately 150 kilocalories (kcals) per day (equivalent to about 1,000 kilocalories/week)[2] and/or engage in at least 150 minutes of moderate PA, or 75 minutes of vigorous PA, or an equivalent of combined moderate-to-vigorous PA (MVPA) per week.[1] Although there are currently no federal guidelines for sedentary behavior (SB) from the United States, SB recommendations from Australia state that adults should minimize the amount of time spent in prolonged sitting and break up long periods of sitting. [4] Dissemination of these recommendations has led to a heightened awareness of the importance and value of PA monitoring as a strategy for chronic disease management. Tools such as wearable devices to track personal PA provide a mechanism to be more informed about activity behavior. As a result, consumer devices that track PA behavior are increasingly popular for researchers, the general public, and developers and manufacturers of ATs.

According to a recent report, the global wearable technology market will grow from over $30 billion in 2016 and should reach over $150 billion in 2026.[5] Activity trackers such as the Fitbit (FB) (Fitbit Inc., San Francisco, CA) provide estimates of steps, energy expenditure (EE), activity minutes and sedentary time (sitting). According to reports, Fitbit remained the leading brand in ATs in 2015, accounting for 79 percent of sales.[6] This expanding market for ATs is driven in part by lower cost, longer battery life and more memory (e.g. to store data for days or weeks). However, growth of the market and advances in consumer device technology far outpace our knowledge about the validity of such devices. This gap is of major concern since it is not clear if these devices provide accurate information. Therefore, to address this problem, it is essential to improve our understanding of the accuracy and precision of the activity output measures of consumer devices. Several studies to validate ATs have been conducted in lab-based settings. Lab-based activity protocols range from fixed time treadmill and overground walking and running to SBs and simulated free-living activities (e.g. vacuuming, computer work). From these studies, we have advanced our knowledge of the accuracy and precision of ATs in estimating physical activity (PA) and ST in laboratory settings. However, there is limited knowledge of how ATs perform outside of a laboratory setting (i.e. free-living environment) where these devices are used by consumers. Addressing this knowledge gap is essential to a comprehensive understanding of the validity of ATs for estimating PA and ST.

To date, few studies have validated ATs in free-living settings. Of these studies, none have employed direct observation (DO) as the criterion measure for steps, EE, activity minutes or sedentary time. The objective of the present study was to determine the accuracy and precision of ATs in estimating steps, EE, activity minutes and sedentary time compared to direct observation-derived measures (criterion measures) in free-living settings. We also validated commonly used research-grade devices.

Methods

Eligible participants were 18-59 years of age and were in good physical health. All participants sign an informed consent document approved by the University of Massachusetts Institutional Review Board.

Participants wore a variety of activity monitors on the wrists, hips or ankle, and a biometric shirt, while carrying out their daily activities in the wild (free-living environments) for three, 2-hour sessions. Participants were videotaped (i.e. direct observation) for all sessions. The video data were imported and processed in a custom behavioral analysis program previously validated.[42]

Instrumentation

Research-grade accelerometer. ActiGraph GT3X-BT (AG) Accelerometer (ActiGraph™ LLC, Pensacola, Florida). This device is a lightweight triaxial PA monitor (4.6cm x 3.3cm x 1.5cm, 19g) that measures acceleration ranging in magnitude from -8 to +8 g's. The accelerometer output has a sampling output range of 30 to 100 Hz and is digitized by a twelve-bit analog-to-digital convertor.

Research-grade step counter. StepWatch™ (SW) (Mōdus™ Health llc, Washington, DC) monitor. The SW is an ankle- worn device (dominant leg) that contains a microprocessor-controlled step counter, and detects steps. Step counts can be recorded every 3 to 60 seconds.

Activity Trackers

Nine different activity trackers were studied: 1) Apple iWatch Sport (AiW) 2) Fitbit Flex (FBF), 3) Fitbit One (FBO), 3) Garmin® Vivofit (GV), 4) Microsoft Band (MB), 5) Misfit Flash (MFF), 6) Misfit Shine (MFS), 7) Polar Loop (PL), 8) Withings Pulse (WP) and 9) New Lifestyles NL-1000 pedometer (NL). Participants' stride length was determined according to the manufacturers recommended method and programmed into the devices requiring this input. [139] (See Tables 8 and 9 for detailed specifications of each AT)

Biometric Shirt. The Hexoskin Biometric Shirt (HxSkin) (Carré Technologies Inc., Montréal, Québec, Canada) is a multi-parameter physiological recording system designed to monitor levels of PA and EE, which combines measurements of cardiac, ventilation, and trunk acceleration.

Video Recording and Direct Observation. We employed a GoPro Hero+ LCD (GOPRO, Inc. San Mateo, Ca) camera to record all observation sessions. The Noldus Observer® XT (Information Technology B.V: Wageningen, Netherlands) is the software package for the collection, analysis, and presentation of observational data

Procedures

Participant height was measured to the nearest 0.25 cm using a standard floor stadiometer and weight was measured using a Tanita scale (DC-430) to the nearest 0.1 kg.

Research-Grade Accelerometer. Participants were fitted with two AG activity monitors. Both AG monitors were synced to the same laptop and initialized in advance to collect data at a sampling rate of 80 Hz. The wrist monitor (AGwrist) was secured using a Velcro strap to the non-dominant wrist (positioned over the dorsal aspect of the wrist midway between the radial and ulnar styloid processes), and the hip monitor (AGhip) was secured using a belt at the iliac crest in line with the anterior axilla. The initialization and wrist wear location are consistent with the current National Health and Nutrition Examination Survey (NHANES) activity monitoring study protocol. [145]

StepWatch™. Participants were fitted with a SW monitor which was fastened using a Velcro strap to the dominant ankle (positioned superior to the lateral malleolus). The SW was programmed to record at 3-second intervals, with sensitivity (the magnitude of acceleration that the device qualifies as constituting a step) set to 13 and cadence (how often the device searches for steps taken) set to 73, consistent with a previous study. [112] The cadence setting is the length of time (cadence settings x 0.01 s) after a step is taken during which a subsequent step cannot be counted and sensitivity setting is the threshold acceleration that must be exceeded to register a step. [155]

Activity Trackers. The device placement was counterbalanced across subjects. For example, the total number of devices worn was the same for all subjects but the order in which the devices were placed on subjects was different between subjects. The same placement positions within each participant across the 3 observation sessions was used. (e.g. participant 2, MFF, left hip for all observation sessions)

Direct Observation: criterion

The DO method employed in the present study was previously described by Lyden et al.[42,102] Briefly, participants were met by a trained observer in their natural environment (e.g. home, place of work, school) and observed for approximately two consecutive hours. A GoPro video camera was used to record each observation session. The GoPro video files were imported into the behavioral coding software (Noldus Observer XT). Focal sampling and duration coding (FSD) were used to record participant behavior (activity type, body posture, intensity and duration). The FSD method is one where every time a behavior changes (e.g. sitting to standing) the observer recorded the new activity type from a coding scheme of general categories from the MET value (from the Compendium of Physical Activities[144]) associated with that category. Each entry of a behavior change was time stamped and the duration of each behavior occurrence was saved. Steps were manually counted from the video. For a detailed description of the procedures used to train researchers and the development of a comparable DO technique see Kozey-Keadle et al.[65]

Direct observation observer training

Training involved research assistants learning how to identify and record activities described in the DO methods. After the training, study observers completed a testing video (~10 min) that included different activities with various postures. Before data collection, researchers were required to correctly classify at least 90% (Cohen's kappa coefficient (k) ≥ 0.90) of the body positions, intensity levels, and duration of activities throughout the testing video.

Direct Observation Sessions

Participants were met by a trained observer in their natural environment (e.g. home, place of work, school) and were observed for approximately two consecutive hours. Each of the 2-hr observation sessions was done at different times of the day (e.g. Session 1: morning; Session 2: afternoon; Session 3: evening), including one weekend day, in the participants' free-living settings (e.g. home, work, driving). For these three visits, researchers initialized devices and met the participants in their free-living environment to be fitted with a variety of activity monitors that were worn on the wrists, hip and ankle, and a biometric shirt. At the end of the 2-hr recording period, researchers recorded the data from the ATs via the real-time display or iPhone app, and downloaded data from monitors and video recording (see table 9 for detailed device initializing and download).

Data Processing and Statistical Evaluation

Data Processing

All data cleaning, processing and analysis were done using the open source *R statistical software package* (www.r-project.org) and computing language R. [136]

Criterion: Direct observation

A log of the start and stop of each behavior recorded by the observer was exported to a csv file using a custom software and profile (Noldus: The Observer XT 12.5). These data were used to determine criterion measures of activity and inactivity including, MET-minutes, MET-hours (where, 1 MET = 1 kcal/kg/hr), kcals per hour (where, Kcals=METs x time x BW [kg]) and time in categories of intensity. The Mifflin-St Jeor equation was used to estimate participant resting metabolic rate (RMR), [147] RMR was added to the EE estimates from the WP monitor only and summed to estimate total calories (e.g. exercise + resting).

Total EE was determined by summing/totaling the amount of time spent in all body positions from the DO coding system (e.g. total MET-minutes). METs were then converted to Kcals/minute (Kcals= METs x time x body weight in kilograms).

Criterion. DO steps were defined as each event when the foot was completely lifted off and lowered back to the ground. To determine criterion step count, steps were manually counted twice for each 2-hour video recording session, and averaged. If there was a greater than 5% difference between total step counts, the video was analyzed a third time and the average of the two closest total step counts was used for analysis. Total steps were determined by summing/totaling the number of steps manually counted/2-hour session. Two trials required a third measure (2.1%).

Research-grade accelerometers

ActiGraph GT3X-BT (AG). Accelerometer data were downloaded to a laptop using the ActiGraph *ActiLife v6.1.1* software and were later extracted to match the corresponding DO time blocks. These data were then processed to derive total time spent in each activity type (cutpoint/method: AGhip;[77] AGwrist[156]), intensity (cutpoint/method: AGhip;[77] AGwrist[156]), EE (method: AGhip[52]) and steps for each participant. The previous ActiGraph model GT3X+ has been shown to be a valid measure of both step count compared with observation[111,157,158] and MVPA compared to indirect calorimetry.[77]

StepWatch™ (SW). StepWatch data were downloaded to the same laptop used for all devices via the SW software (v3.4). The SW data from the observation session were then exported and saved for analysis. Total steps was determined by summing/totaling the number of steps taken.

Activity Trackers

Steps, EE, activity minutes (if provided by AT) and sedentary time (if provided by AT) data from ATs were recorded at the beginning and at the end of each observation session. Total estimates were then calculated by subtracting the start values from the end values.

Fitabase (Small Steps Labs, LLC. San Diego, Ca). All FB data were exported using Fitabase, a research platform that acquires data from Internet connected consumer devices. Currently, FB is the only consumer device company that utilizes Fitabase. The advantage of using this platform to acquire FB data is that it provides minute-by-minute data for steps, kcals, MET-minutes and activity minutes (intensity) in comparison to the FB software and Dashboard which only provides totals for steps, kcals and intensity for the monitoring period.

Biometric Shirt. The HxSkin data were downloaded to the HxServices Dashboard. Next, an "Activity" was created in the myhexoskin website for the 2-hour observation session (data are time stamped). Then energy expenditure and step estimates were recorded for each 2-hour observation session.

Statistical Evaluation

Correlations

We used the Pearson correlation coefficient to determine the strength of the relationship between criterion measured and device estimated steps, EE, activity minutes and ST.

Accuracy

We used bias in units of steps, kcals and minutes (AT estimates – criterion) and as a percentage [(mean difference between the AT estimates and the criterion/ criterion x 100]. The percentage bias is useful because, for instance, a 10% bias of 15,000 steps/2-hrs could be applied to an observation time of 2-hrs (a 2-hr overestimate), compared to bias of +1,500 steps/2-hrs.

Precision

We used confidence intervals (CI) as measures of precision. If the upper and lower 95% confidence interval of the bias span 0, then the estimate was not considered significantly different from the criterion at $\alpha = .05$. Higher precision was indicated by higher correlations and smaller CI.

Linear mixed models were used to compare the accuracy and precision of the steps, EE, activity minutes and ST estimates from the devices.

Results

Table 11 illustrates participant' characteristics. Thirty-two healthy adults (50% female, 37.5% minority). Average age (yrs.) and BMI (kg*m-2) were 32.3 and 24.4, respectively.

Table 12 summarizes participant visits by day of week and time block. Morning (time from arising from bed for the day until lunchtime [or 12:00 PM if no lunch]), afternoon (period during lunch [or 12:00 PM] until dinner [or 6:00 PM if no dinner]) and evening (time after dinner until getting into bed for the night) visits totaled 29, 34 and 33, respectively. Weekday and weekend visits totaled 62 and 34, respectively.

Table 13 summary statistics (in minutes) of top eight activity categories that participants engaged in during 2-hr visits. Activity categories are based on the Compendium of Physical Activities.[146] Means ranged from 18.1 minutes

(Transportation) to 90.5 minutes (Occupational). Minimums ranged from 1.0 minute (Walking) to 55.0 minutes (Running). Maximums ranged from 12.0 minutes (Walking) to 120.0 minutes (Conditioning Exercise, Home Activities and Occupational). As a percent of 2-hours, means were 15.1% minutes, 15.3% minutes, 18.8% minutes, 47.6% minutes, 54.0% minutes, 59.2% minutes, 62.8% minutes and 75.4% minutes for transportation, walking, self-care, home activities, miscellaneous, running, conditioning exercise and occupational, respectively.

Relationship between criterion measured and device estimated steps: Figures 18 - 21 show correlations between DO measured steps and device estimated steps. Correlations ranged from r= 0.86 (FBF) to r = 0.97 (AGhip, NL).

Differences between criterion measured and device estimated steps: Figure 22 shows bias of ATs, AGhip, AGwrist, and SW step estimates compared to DO measured steps. Average steps for DO was 2,623/2-hours. The SW and PL were not significantly different from DO. Average steps were -119 (CI:-439,201) and -57 (CI:-291,175) steps/2-hours for the SW and PL, respectively. All other devices significantly underestimated steps compared to DO. For several devices, underestimations ranged from -753 to -524 steps/2-hours. The FBF, WP, FBO, MFS, HxSkin, AGhip and MB underestimates were -753 (CI:-1,144,-362.8), -725 (CI:-887,-564), -647 (CI:-869,-425), -628 (CI:-816,-440), -586 (CI:-768,-403), -558 (CI:-699,-417), and -524 (CI:-689,-358) steps/2-hours, respectively. For other devices, underestimations ranged from -437 to -285 steps/2-hours. In this range, the NL, MFF, AGwrist, GV, and AiW underestimations were -437 (CI:-581,-292), -435 (CI:-621,-250), -379 (CI: -717,-41), -341 (CI:-525,-156), and -285 (CI:-559,-11) steps/2-hours, respectively.

Percent differences between criterion measured and device estimated steps: Figure 23 shows percent bias of ATs, AGhip, AGwrist, and SW step estimates compared to DO measured steps. The SW and PL were not significantly different from DO. Percent average steps were -4.5% (%CI:-16.7,7.6) and -2.1% (%CI:-11.1,6.7) steps/2-hours for the SW and PL, respectively. All other devices significantly underestimated steps compared to DO. For several devices, percent underestimations ranged from -28,7% to -19.9% steps/2-hours. In this range, the FBF, WP, FBO, MFS, HxSkin, AGhip and MB percent underestimates were -28.7% (%CI:43.6,-13.8), -27.6% (%CI:-33.8,-21.5), -24.6% (%CI:-33.1,-16.2), -23.9% (%CI:-31.1,-16.7), -22.3% (%CI:-29.3,-15.4), -21.2% (%CI:-26.6,-15.8), and -19.9% (%CI:-26.3,-13.6) steps/2-hours, respectively. For other devices, percent underestimations ranged from -16.6 to -10.8% steps/2-hours. In this range, the NL, MFF, AGwrist, GV, and AiW percent underestimations were -16.6% (%CI:-22.1,-11.1), -16.6% (%CI:-23.6,-9.5), -14.4% (%CI:-27.3,-1.5), -13.0% (%CI:-20.0,-5.9), and -10.8% (%CI:-21.3,-0.4) steps/2-hours, respectively.

Relationship between criterion measured and device estimated kcals Figures 24 -26 show correlations between DO measured kcals and device estimated kcals. Correlations ranged from r = 0.32 (GV) to r = 0.85 (AGhip)

Differences between criterion measured and device estimated kcals: Figure 27 shows bias of AT and AGhip kcal estimates compared to DO measured kcals. Average kcals for DO was 329/2-hours. The PL, MFF and MFS were not significantly different than DO. Bias for the PL, MFF and MFS were -7.0 (CI:-37.0,22.8), 6.9 (CI:-36.6,50.4), and 8.3 (CI:-47.1,63.9) kcals/2-hours, respectively. The MB, WP, FBO, FBF, GV, AiW and AGhip significantly underestimated kcals compared to DO. Underestimates for the MB, WP, FBO, FBF, GV, AiW and AGhip were -121.8 (CI:-163.7,-79.9), -107.7 (CI:-136.1, -79.4), -90.6 (CI:-120.7,-60.5), -85.3 (CI:-123.8,-46.7), -71.4 (CI:-127.5,-15.3), -60.2 (CI:-93.9,-26.5), and -48.8 (CI:-75.3,-22.3) kcals/2-hours, respectively. In contrast, the HxSkin significantly overestimated kcals compared to DO. Average overestimation was 119.3 (CI:52.2,186.3) kcals/2-hours.

Percent differences between criterion measured and device estimated kcals: Figure 28 shows percent bias of AT and AGhip kcal estimates compared to DO measured kcals. Average kcals for DO was 329/2-hours. The PL, MFF and MFS were not significantly different than DO. Percent bias for the PL, MFF and MFS were -2.1% (%CI:-11.2,6.9), 2.0% (%CI:-11.1,15.3), and 2.5% (%CI:-14.3,19.4) kcals/2-hours, respectively. The MB, WP, FBO, FBF, GV, AiW and AGhip significantly underestimated kcals compared to DO. Percent underestimates for the MB, WP, FBO, FBF, GV, AiW and AGhip were -36.9% (%CI:-49.7,-24.2), -32.7% (%CI:-41.3,-24.1), -27.5% (%CI:-36.6,-18.3), -25.8% (%CI:-37.5,-14.2), -21.6% (%CI:-38.7,-4.6), -18.2% (%CI:-28,5,-8.0), and -14.8% (%CI:-22.8,-6.7) kcals/2-hours, respectively. In contrast, the HxSkin significantly overestimated kcals compared to DO. Average percent overestimation was 36.2% (%CI:15.8,56.5) kcals/2-hours.

Relationship between criterion measured and device estimated MVPA minutes: Figures 29 and 30 show correlations between DO measured MVPA minutes and device estimated MVPA minutes. The correlations between DO MVPA minutes and FBF, AGwrist, FBO and AGhip were r = 0.54, 0.70, 0.71 and 0.75, respectively.

Differences between criterion measured and device estimated MVPA minutes: Figure 31 shows bias of MVPA minutes per 2-hours for the AGhip, AGwrist, FBO and FBF compared to DO MVPA minutes. Average MVPA for DO was 27 minutes/2-hours. The AGhip and FBO significantly underestimated MVPA minutes by -11.8 (CI:-15.5,-8.1) and -5.4 (CI:-9.9,-0.9) /2-hours, respectively. In contrast, the AGwrist significantly overestimated MVPA by 6.9 minutes (CI:2.5,11.4) /2-hours. The FBF was not significantly different from DO MVPA minutes. On average, the FBF underestimated MVPA by -3.5 (CI:-9.6,2.4) minutes/2-hours.

Percent differences between criterion measured and device estimated MVPA minutes: Figure 32 shows percent bias of MVPA minutes per 2-hours for the AGhip, AGwrist, FBO and FBF compared to DO MVPA minutes. The AGhip and FBO significantly underestimated MVPA minutes by -43% (%CI:-57.3,-29.9) and -20% (%CI:-36.7,-3.3) /2-hours, respectively. In contrast, the AGwrist significantly overestimated MVPA minutes by 25% (%CI:9.2,42.1) /2-hours. The FBF was not significantly different from DO MVPA minutes. On average, the FBF underestimated MVPA minutes by -13% (%CI:-35.6,9.1) minutes/2-hours.

Relationship between criterion measured and device estimated output analogous to MVPA minutes: Figures 33 and 34 show correlations between DO MVPA minutes and MVPA minutes from the NL, AiW, MFF, MFS, and PL. The correlations ranged from to r = 0.20 (NL) to r = 0.64 (MFF). Correlations for the PL, MFS, and the AiW were r= 0.40, 0.56, and 0.57, respectively

Differences between criterion measured and device estimated output analogous to MVPA minutes: Figure 35 shows bias of MVPA minutes per 2-hours for the AiW, PL, NL, MFS, and MFF compared to DO MVPA minutes. All ATs significantly underestimated MVPA minutes. On average, underestimations ranged from -17 to -13 minutes/2-hours. Confidence intervals ranged from -24 to -9 minutes per 2-hours. The PL and NL estimates resulted in the widest CIs of approximately 14 minutes/2-hours, respectively.

Percent differences between criterion measured and device estimated output analogous to MVPA minutes: Figure 36 shows percent bias of MVPA minutes per 2-hours for the AiW, PL, NL, MFS, and MFF compared to DO MVPA minutes. All ATs significantly underestimated MVPA minutes. On average, percent underestimations ranged from -64 to -48% MVPA minutes per 2-hours. Confidence intervals ranged from -88 to -36% MVPA minutes per 2-hours. The PL and NL estimates resulted in the widest CIs of approximately 48% and 50% MVPA minutes/2-hours, respectively.

Relationship between criterion measured and device estimated sedentary time: Figure 37 shows correlations between DO sedentary minutes and AT estimates of sedentary minutes. Correlations for the FBO, FBF AGhip and AGwrist were r = 0.06, -0.06, 0.59 and 0.77, respectively.

Differences between criterion measured and device estimated sedentary time: Figure 38 shows the bias of AT and AGhip sedentary time estimates compared to DO sedentary time. All devices significantly overestimated sedentary time compared to DO. Overestimates for the Fitbit One, FBF and AGhip were 14.3 (CI:2.8,25.8), 20.9 (CI:9.3,32,5), and 52.0 (CI:43.6,60.4) sedentary minutes/2-hours, respectively.

Percent differences between criterion measured and device estimated sedentary time: Figure 39 shows the percent bias of AT and AGhip sedentary time estimates compared to DO sedentary time. All devices significantly overestimated sedentary time compared to DO. Overestimates for the FBO, AGwrist, FBF and AGhip were 34% (%CI:7.7,61.5), 47% (%CI:31.3,64.0), 50% (%CI:23.0,77.1), and 118% (%CI:101.1,135.5) sedentary minutes/2-hours, respectively.

Discussion

The purpose of the present study was to validate ATs and research-grade accelerometers in free-living settings to estimate steps, EE, activity minutes and sedentary time using DO as the criterion method. In general, all devices accurately estimated steps and the estimates were highly correlated with DO. Estimates of EE, MVPA minutes were less accurate and more variable across devices and correlations between the estimated measure and the measure derived from DO ranged from weak (r=0.20) to moderate (r=0.75). Devices were the least accurate in estimating sedentary time, although one method[156] was more correlated with DO (AGwrist r=0.77) (Table 14).

Activity Trackers

To date, several studies have validated ATs in estimating steps, EE, MVPA minutes and sedentary time and the results are equivocal. Activity trackers have been reported to significantly under- and- overestimate PA and ST. Most studies have used AG accelerometer measures as the reference. The results from our study indicate that this is not an appropriate reference measure, given the differences we observed between AG measures of PA and ST in comparison to the measures derived from DO.

Steps

Our main findings were that ATs produced accurate step estimates and were highly correlated to criterion measures. Current research has compared steps from consumer ATs and research-grade devices in free-living settings. Some studies reported that ATs overestimated steps, [21,32-34,159-161] while others reported that ATs underestimated steps. [21,31,32,159] Differences in step results from previous studies compared to this study may be related to the use of different reference measures. We reported that the AGhip significantly underestimates steps by -558 (CI:-699,-417)/2-hours. Therefore, in most of the studies that used the AG as the reference tool would yield results that indicate an overestimation of steps by the ATs. For example, Ferguson et al[32]. reported that the WP significantly underestimated daily steps by -632 compared to hip-worn ActiGraph

GT3X+ steps (reference measure). Applying our AGhip findings (significant underestimation [-558 steps/2-hrs]} to Ferguson's may impact their findings (e.g. the WP overestimates daily steps).

Energy Expenditure

We found that AT estimates of EE were less accurate than step estimates, and highly correlated with criterion measures. Similar findings have been reported.[37] The current research comparing the EE estimates of ATs in free-living settings has primarily used research-grade devices as the reference measure, with one study employing doubly-labeled water (DLW).[37] For EE, two studies showed ATs overestimated kcals,[31,37] and three studies showed that ATs underestimated kcals[32,162,163] with variable precision in free-living settings.[31,32,37,162,163] We compared EE data recorded by ATs to EE data estimated from DO. Because the DO system used has been validated as a criterion for free-living PA and ST,[42] our study improves upon the current literature.

MVPA Minutes/ Sedentary Time

The U.S. PA Guidelines (Guideline) define MVPA as activities where intensity is greater than 2.99 METS. Currently, devices from one AT company (Fitbit) provide MET values, retrievable via the research platform Fitabase. Accordingly, these data are directly comparable to our criterion measure. Although ATs from other companies provide proprietary estimates of PA intensity, they do not explicitly define MVPA (i.e. Non-Guideline; Table 10).

In general, we found AT estimates of MVPA minutes (Guideline and Non-Guideline) were less accurate than step and EE estimates, and were moderately to weakly correlated with criterion measures. To date, one validation study examining ATs in free-living settings on adults employed Fitabase to retrieve MET-minutes of activity.[164] They reported, that the FBF significantly overestimated daily MET rate (mean difference 0.7, SD 0.09, METs/day, P<.001), proportion of time in moderate (mean difference 3.0%, SD 11.0%, per day, P<.001) and vigorous PA (mean difference 3.0%, SD 1.0%, per day, P<.001 compared to the AG GT3X. Several studies have examined estimates of Non-Guideline MVPA minutes from ATs compared to accelerometer derived MVPA minutes in free-living adults. Two studies reported ATs underestimated MVPA minutes,[31,33] three studies reported ATs overestimated MVPA minutes, [43,161,165,166] and one study reported ATs underestimated and overestimated MVPA minutes in free-living settings.[32]

We reported that ATs were the least accurate at estimates of sedentary time (overestimated) and were weakly correlated with criterion measures. Underestimations[164] and overestimations[41,166] of sedentary time by ATs have been previously reported. Two studies[164,41] used the AG accelerometer as the comparison measure and applied different count cutpoints to define sedentary time. One study used the activPAL as the comparison measure.[166] The contrasting findings of previously reported sedentary time to ours may be reflective of reference measures used to compare ATs: accelerometers versus DO. Direct observation is considered superior to accelerometers in estimating PA behaviors[167] as it provides instant, visual information regarding activity type, posture and context- aspects that govern PA intensity- and may influence device output. For example, we coded all seated activity (e.g. seated typing) as sedentary, while a wrist-worn AT may detect hand/wrist accelerations as steps. We are confident that employing DO provides us with a true measure of free-living behaviors.

Research-grade accelerometers

We found that step estimates from research-grade accelerometers were accurate and highly correlated with DO (range: r=0.91 (SW) to 0.97 (AGhip)). Despite statistical differences (e.g. over-or-underestimations), the AGhip and AGwrist were highly correlated with DO (AGhip range: r=0.56 (sedentary time) to 0.97 (steps); AGwrist range: r=0.70 (MVPA) to 0.95 (steps). Similar to our findings, the SW has been shown to accurately estimate steps in simulated free-living laboratory investigations.[112] Accept for sedentary time, AGwrist estimates were greater than AGhip estimates. Several studies have reported the wrist location produces greater output as compared to the hip location, in free-living settings.[168,169] For MVPA, the AGwrist and AGhip were less accurate (than for steps) (percent bias: 25.6% and -43.6, respectively) but remained moderately correlated with DO (r=0.70 and 0.75, respectively). This may be the result of differences in wear-location and/or methods used to derive MVPA minutes (e.g. raw accelerations for AGwrist compared to counts per minute for AGhip). This is in contrast to the study by Murakami et al.[37] who reported that AGhip underestimated EE (bias -534.9 kcals/d) compared to doubly-labeled water with strong correlations (r=0.80) during 15-days of free-living time. Similar to previous studies AGhip tended to underestimate[102] and AGwrist overestimated MVPA minutes[169] compared to reference measures (DO and AGhip, respectively). AGhip and AGwrist overestimated sedentary time (percent bias: 118.3 and 47.5%, respectively). Previous studies are in agreement with our AGwrist findings[156] but disagree with our AGhip findings.[65] It is possible that the underestimation of sedentary time in the present study is due to differences in data processing techniques to estimate sedentary time. For example, Kozey-Keadle et al.[65]

used AG counts per minute to categorize ST compared to Staudenmayer et al. who used AG 15-second epochs from raw-accelerations and machine-learning techniques to categorize ST.

Activity trackers compared to Research-grade accelerometers

We showed that ATs perform similar to research-grade accelerometers is estimating steps and EE. However, ATs are less precise and accurate in estimating MVPA minutes and sedentary time. These differences are likely the result contrasting methods used to define MVPA minutes and sedentary time. For example, we defined MVPA and ST using posture and intensity, whereas ATs rely entirely on proprietary algorithms. For the AGs, MVPA and ST were defined using counts/min (AGhip) and 15-second epochs (from raw-accelerations) and random forest (machine-learning technique) (AGwrist).

There is a growing movement toward using ATs as a measurement tool in PA intervention trials. There are many clinical trials underway that are employing ATs to estimate PA and ST exposures and outcomes. Our findings suggest that ATs are accurate in estimating PA behaviors such as steps in free-living settings. In fact, step accuracy was similar between ATs and research-grade accelerometers. Though more research validating ATs in free-living settings compared to DO is needed, it is reasonable to employ these devices to estimate measures of steps, EE and MVPA minutes. On the other hand, our results indicate that the accuracy and precision of ATs in estimating ST is less certain.

Given the widespread use of ATs, we have an opportunity to engage the public and industry leaders who sell these devices in conversations about their experiences in using these devices with the goal to improve the user experience to enhance long-term compliance and adapting a more active lifestyle. Evidence presented in this study support the accuracy and serve as an anchor for these conversations.

Strengths and Limitations

The primary strengths of this study were the use of a validated DO method[42] to derive criterion measures for PA and sedentary behavior measures and conducting this study in the natural environment Most previous free-living studies employed accelerometers as a surrogate for gold-standard criterion measures (e.g. DO, doubly labeled water) to assess steps, EE, activity minutes and/or ST. [32-34,43-46] Another strength was the wide range of activities, intensities of activities and the duration of activities performed naturally by participants. Activities ranged from sleeping to trail running. Intensities ranged from 1.0 to 12.0 METs, and duration of specific activities ranged from seconds to hours. All provided a unique opportunity to capture a rich dataset critical to proper scrutiny of AT estimates of PA and ST. Additional strengths included having observations conducted at all times of day ranging from 5:00 am to after 11:00 pm and conducted in multiple settings as diverse as a Zumba class to a nightclub (Tables 12 and 13).

This study also has limitations. We employed a validated DO system that used the Compendium of Physical Activities to apply MET values to activities. The values in the Compendium do not estimate the energy cost of PA in individuals in ways that account for differences in body mass, adiposity, age, sex, efficiency of movement, and environmental conditions in which the activities are performed.[146] Therefore, it is possible that activities were misclassified by intensity category, which may have resulted in inaccuracies of activity minutes, sedentary time and energy expenditure. The observation duration within each trial was another limitation. We observed participants for three, 2-hour time frames. Compared to previous free-living AT validation studies, the time frame for observation is short. In general, ATs are designed to be worn during waking hours and while sleeping. Thus, our findings may not be a true representation of whole day behavior. However, we do have a broad range of activities ranging from light to very vigorous activity and a balanced distribution of time of day and day of week that participants were observed (Tables 12 and 13).

In conclusion, this study provides evidence that ATs are accurate with varying precision in estimating steps, EE and activity minutes. Sedentary time estimates from ATs were less accurate. Further, AT and research-grade accelerometers performed similarly (e.g. both were more accurate in estimating steps and less accurate in estimation MVPA minutes). This work significantly advances the field of activity monitor validation that should set the standard for future work.

Device	Apple iWatch Sport	Fitbit Flex	Fitbit One	Garmin Vivofit	New Lifestyles NL-1000	Microsoft Band	Misfit Flash	Misfit Shine	Polar loop	Withings Pulse
Cost	$350.99	$79.95	$99.95	$99.99	$54.95	$199.99	$29.99	$69.99	$109.95	$119.95
Wear location	Wrist	Wrist	Clip on (multiple locations)	Wrist	Hip	Wrist	Clip on (multiple locations)	Clip on (multiple locations)	Wrist	Clip on and wrist band
Tracks Calories Burned	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Active Time	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Steps	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Distance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Elevation/Stairs	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓
Tracks Sleep	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Heart Rate	✓	✗	✗	✗	✗	✓	✗	✗	✓	✓
Battery or Chargeable	Chargeable (every 18 hours)	Chargeable (every 5 days)	Chargeable (every 10+ days)	Battery (every 1+ years)	Battery (up to 18 months)	Chargeable (every 48 hours)	Battery (lasts up to 6 months)	Battery (lasts up to 6 months)	Chargeable (up to 6 days)	Chargeable (every 2 days)
Uploading Data	Bluetooth	Bluetooth	Bluetooth	Bluetooth	Real-time data	USB			USB	Bluetooth
Tracker Display	Real-time data	LED progress indicator	Real-time data	Real-time data	Real-time data	Real-time data	LED progress indicator	LED progress indicator	Real-time data	Real-time data

Table 8. Features of consumer-based activity trackers

LED, Light-Emitting Diode; USB, Universal Serial Bus

Device (Location)	Output	Data Extraction Method	
		Upload	Retrieval
Apple iWatch Sport (W)	EE, steps, active calories, min: exercise, total active time, stand hours	Bluetooth	Apple Activity App
GT3X-BT (W & H)	Steps, min: Sedentary, light, moderate, vigorous	USB cable	ActiLife
Fitbit Flex /One (W/H)	EE, steps, MET-min, min: sedentary, light, moderate, vigorous	Bluetooth	Fitabase
Garmin Vivofit (W)	EE, steps, active calories, %: sedentary, active, highly active	Bluetooth	Garmin Connect™ App
Hexoskin (T)	EE, steps	USB cable	Hexoskin dashboard
Microsoft Band (W)	EE, steps, active min	USB cable	MB dashboard
Misfit Flash/Shine (H/W)	EE, steps, active min: light, moderate, vigorous	Bluetooth	Misfit App
New Lifestyles NL-1000 (H)	Steps, MVPA min	RTD	RTD
The Observer XT (NA)	MET-hours, MET-min	The Observer XT	The Observer XT
Polar Loop (W)	EE, steps, time: lying, sitting, active, sitting, min: stand, walk, run	USB cable	Polar dashboard
StepWatch (A)	Steps	USB cable	StepWatch dashboard
Withings Pulse (H)	EE, steps	Bluetooth	Withings App

Table 9. Devices with corresponding output and data extraction method

H, hip; W, wrist: T, torso; A, Ankle; NA, not applicable; EE, energy expenditure; min, minutes; MVPA, moderate-to-vigorous physical activity; MB, Microsoft Band; RTD, real-time display

Device	Output	Definition
Apple iWatch	Exercise minutes	Anything above a brisk walk is classed as exercise. Every full minute of movement equaling or exceeding the intensity of a brisk walk counts towards daily Exercise goal (30 min).
Fitbit Flex/One	Active minutes	Activities at or above about 3 METs. Minutes are only awarded after 10 minutes of continuous moderate-to-intense activity.
Misfit Flash/Shine	Light-, moderate-, vigorous- minutes	No definitions provided.
NL-1000	MVPA	Moderate-to-vigorous physical activity (MVPA) time accumulation.
Polar Loop	WALK and JOG	Medium and high intensity activity, respectively.

Table 10. Activity tracker intensity outputs and definitions

Characteristic	Mean (SD)
Age (yrs.)	32.3 (13.3)
BMI (kg*m-2)	24.4 (3.3)
N (%)	
Female	16 (50)
Minority	12 (37.5)

Table 11. Participant characteristics (N = 32)
SD, standard deviation; BMI, body mass index

	Weekday	Weekend Day	Total Visits
Morning	18	11	29
Afternoon	20	14	34
Evening	24	9	33
Total Visits	62	34	96

Table 12. Summary of visits by day of week and time block

Morning, the time from arising from bed for the day until lunchtime (or 12:00 PM if no lunch); Afternoon, the period during lunch (or 12:00 PM) until dinner (or 6:00 PM if no dinner); Evening, the time after dinner until getting into bed for the night

	Mean	SD	Minimum	Maximum
Conditioning Exercise	75.4	33.5	5.0	120.0
Home Activities	57.1	43.5	1.0	120.0
Miscellaneous	64.8	43.4	2.5	120.0
Occupational	90.5	38.0	5.0	120.0
Running	71.0	17.7	55.0	90.0
Self-Care	22.5	24.6	1.0	66.0
Transportation	18.1	12.7	5.0	45.0
Walking	18.3	23.7	1.0	12.0

Table 13. Summary statistics (in minutes) of top eight activity categories that participants engaged in during 2-hr visits
Activity categories are based on the Compendium of Physical Activities
SD, standard deviation; NA, not applicable

		Steps	EE (kcals)	MVPA (min)	SED (min)
Criterion (avg)		2,623	329.0	27.0	43.0
Device					
AGhip	Accuracy (%)	-579‡ (-22.0%)	-48.8‡ (-14.8%)	-11.8‡ (-43.6%)	50.6‡ (118.3%)
	Precision	-718,-439	-75.3,-22.3	-15.5, -8.3	43.2,57.9
	Correlation	0.97	0.85	0.75	0.59
AGwrist	Accuracy (%)	-379‡ (-14.4%)	NA	6.9‡ (25.6%)	20.3‡ (47.5%)
	Precision	-717,-40	NA	2.5, 11.4	13.3,27.3
	Correlation	0.95	NA	0.70	0.77
SW	Accuracy (%)	-180 (-6.8%)	NA	NA	NA
	Precision	-421,60	NA	NA	NA
	Correlation	0.92	NA	NA	NA
AiW	Accuracy (%)	-285‡ (-10%)	-60.2‡ (-18.2%)	-16.8‡ (-62.0%)	NA
	Precision	-559,-11	-93.9,-26.5	-21.8,-11.7	NA
	Correlation	0.91	0.75	0.57	NA
FBF	Accuracy (%)	-753‡ (-28.7%)	-85.3‡ (-25.8%)	-.35 (-13.2%)	21.4‡ (50.0%)
	Precision	-1,144,-362	-123.8,-46.7	-9.6,2.4	9.8,33.0
	Correlation	0.83	0.71	0.54	-0.06
FBO	Accuracy (%)	-647‡ (-24.6%)	-90.6‡ (-27.5%)	-5.4‡ (-20.0%)	14.8‡ (34.6%)
	Precision	-869,-425	-120.7,-60.5	-9.9,-0.9	3.3,26.3
	Correlation	0.96	0.76	0.71	0.06
GV	Accuracy (%)	-341‡ (-13.0%)	-71.4‡ (-21.6%)	NA	NA
	Precision	-525,-156	-127.5,-15.3	NA	NA
	Correlation	0.95	0.32	NA	NA
HxSkin	Accuracy (%)	-586‡ (-22.3%)	119.3‡ (36.2%)	NA	NA
	Precision	-768,-403	52.2, 186.3	NA	NA
	Correlation	0.96	0.67	NA	NA

		Steps	EE (kcals)	MVPA (min)	SED (min)
Criterion (avg)		2,623	329.0	27.0	43.0
Device					
MB	Accuracy (%)	-524‡ (-19.9%)	-121.8‡ (36.9%)	NA	NA
	Precision	-689,-358	-163.7,-79.9	NA	NA
	Correlation	0.96	0.41	NA	NA
MFF	Accuracy (%)	-435‡ (-16.6%)	6.9 (2.0%)	-13.1‡ (-48.4%)	NA
	Precision	-621,-250	-36.6,50.4	-17.7,-8.5	NA
	Correlation	0.96	0.75	0.64	NA
MFS	Accuracy (%)	-628‡ (-23.9%)	8.3 (2.5%)	-15.7‡ (-57.9%)	NA
	Precision	-816,-440	-47.1,63.9	-20.5,-10.9	NA
	Correlation	0.96	0.71	0.56	NA
NL	Accuracy (%)	-437‡ (-16.6%)	NA	-16.6‡ (-61.2%)	NA
	Precision	-581,-292	NA	-24.4,-9.7	NA
	Correlation	0.97	NA	0.20	NA
PL	Accuracy (%)	-57 (-2.1%)	-7.0 (-2.1%)	-17.4‡ (-64.4%)	NA
	Precision	-291,175	-37.0,22.8	-23.9,-10.8	NA
	Correlation	0.95	0.8	0.40	NA
WP	Accuracy (%)	-725‡ (-27.6%)	-107.7‡ (-32.7%)	NA	NA
	Precision	-887,-564	-136.1, -79.4	NA	NA
	Correlation	0.96	0.77	NA	NA

Table 14. Summary of device accuracy, percent accuracy, precision and correlations in estimating steps, energy expenditure, MVPA and sedentary minutes compared to criterion measures

MVPA, moderate-to-vigorous physical activity; SED, sedentary; EE, energy expenditure; avg, average; AGhip, hip-worn GT3X-BT; AGwrist, wrist-worn GT3X-BT; NA, not applicable.

‡, significantly different than criterion ($p<0.05$).

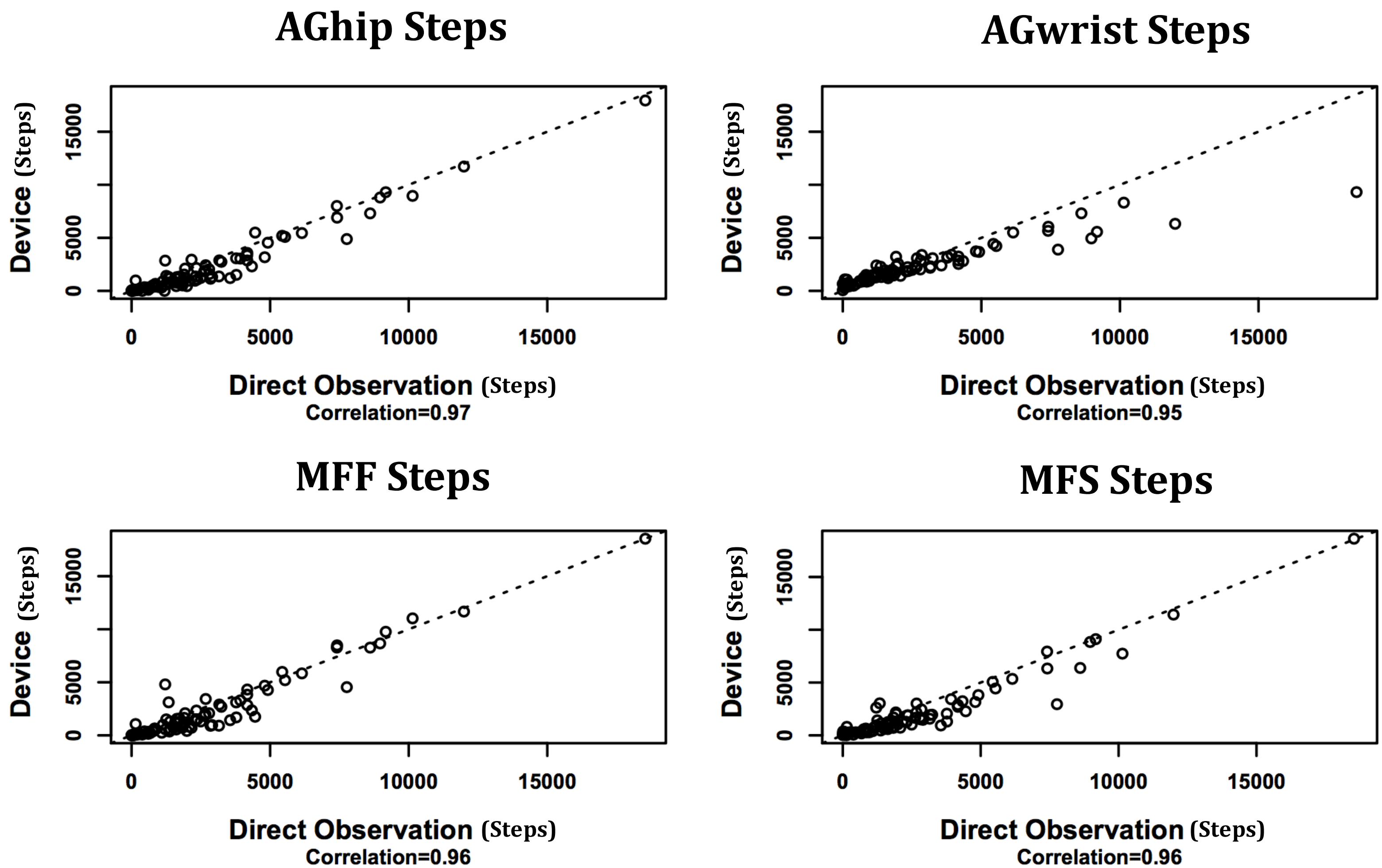


Figure 18. Relationship between criterion steps and hip- and- wrist-worn ActiGraph, Misfit Flash and Misfit Shine estimated steps

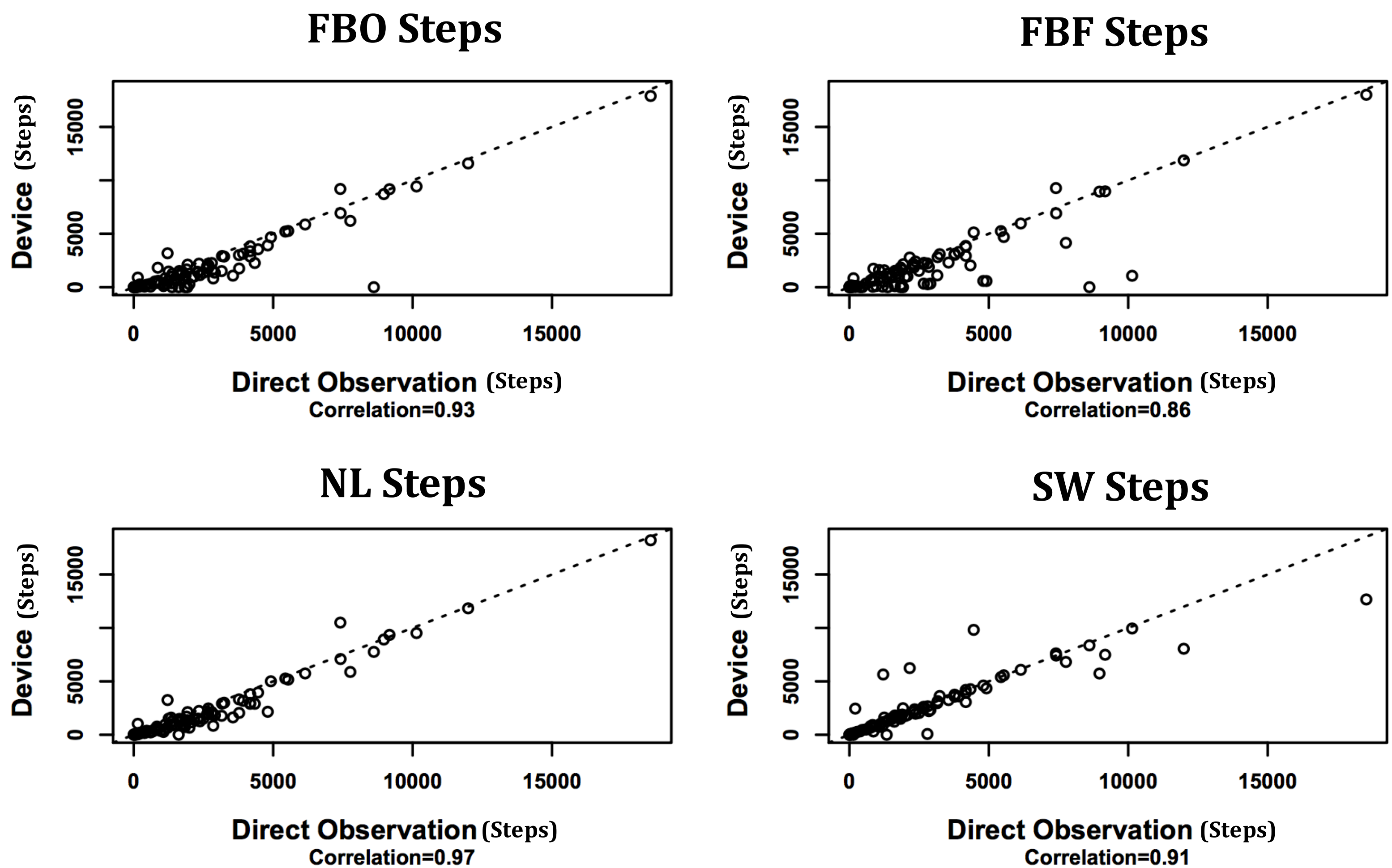


Figure 19. Relationship between criterion steps and Fitbit One, Fitbit Flex, NL-1000 and StepWatch estimated steps

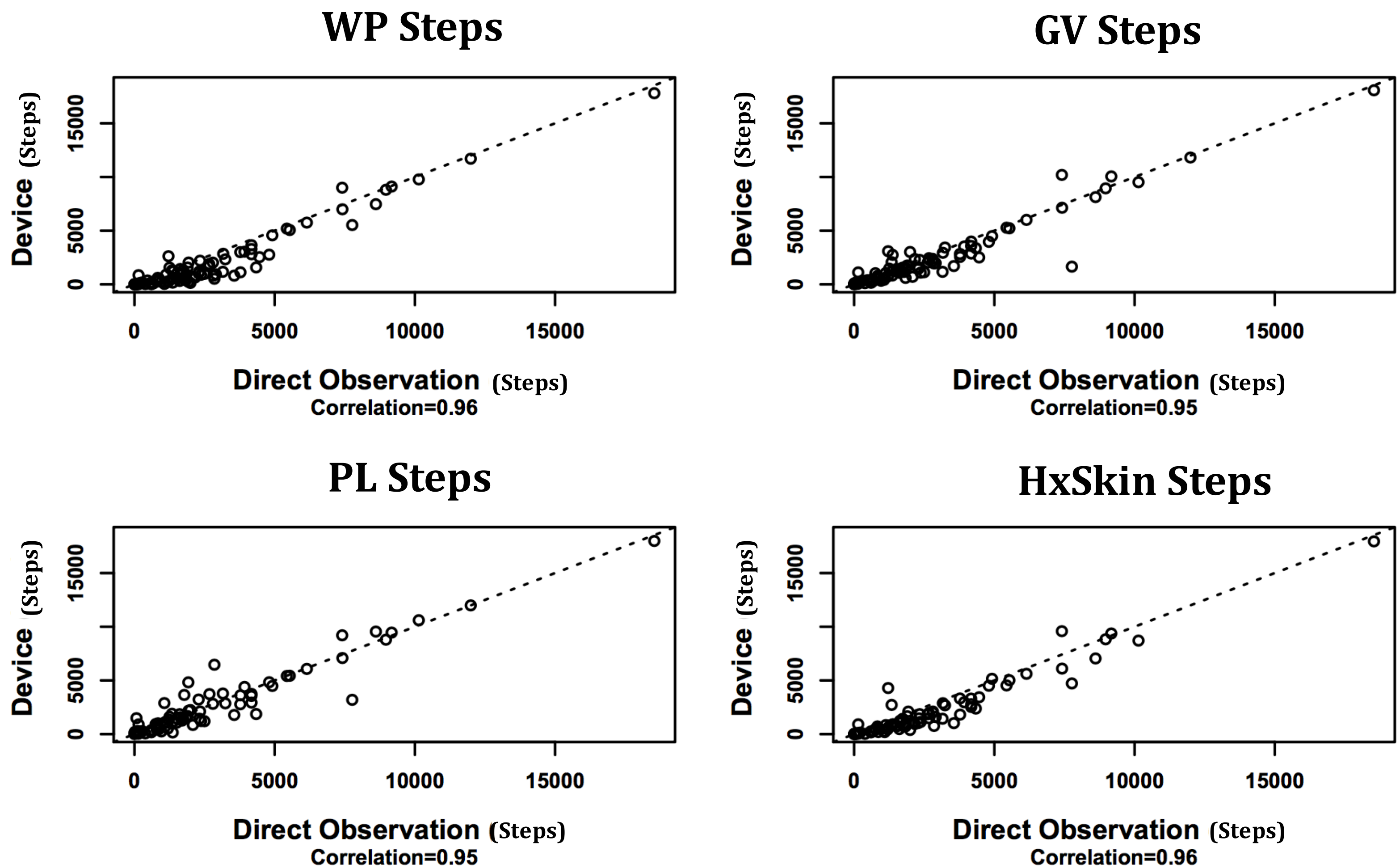


Figure 20. Relationship between criterion steps and Withings Pulse, Garmin Vivofit, Polar Loop and Hexoskin estimated steps

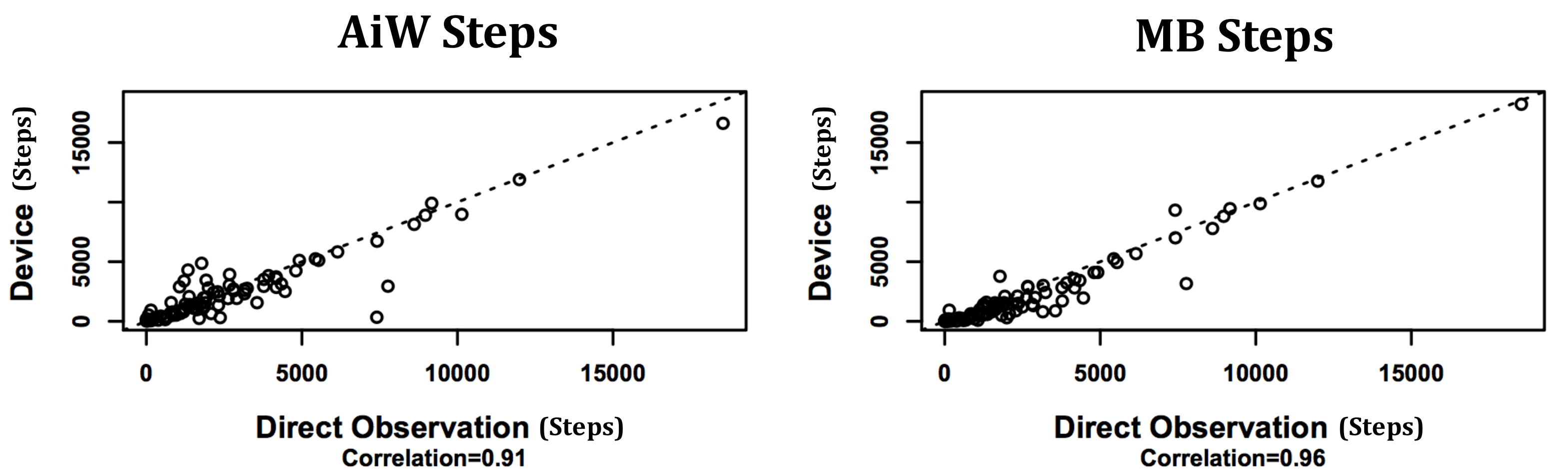


Figure 21. Relationship between criterion steps and Apple iWatch and Microsoft Band estimated steps

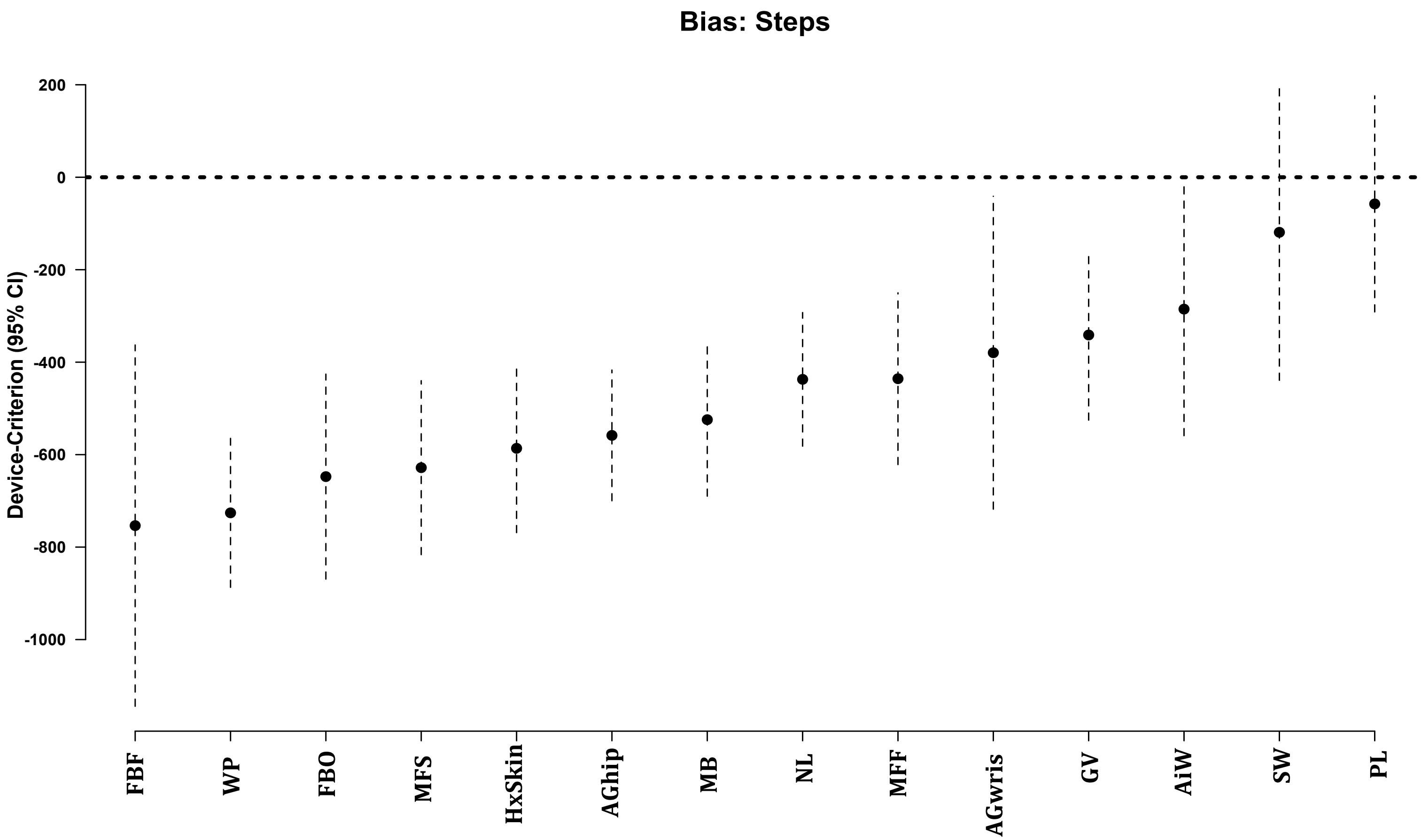


Figure 22. Bias for Fitbit Flex (FBF), Withings Pulse (WP), Fitbit One (FBO), Misfit Shine (MFS), Hexoskin (HxSkin), hip-worn ActiGraph (AGhip), Microsoft Band (MB), NL-1000 (NL), Misfit Flash (MFF), wrist-worn ActiGraph (AGwrist), Garmin Vivofit (GV), Apple iWatch (AiW), StepWatch (SW) and Polar Loop (PL), step estimates compared to criterion steps

Data presented as mean and 95% confidence intervals

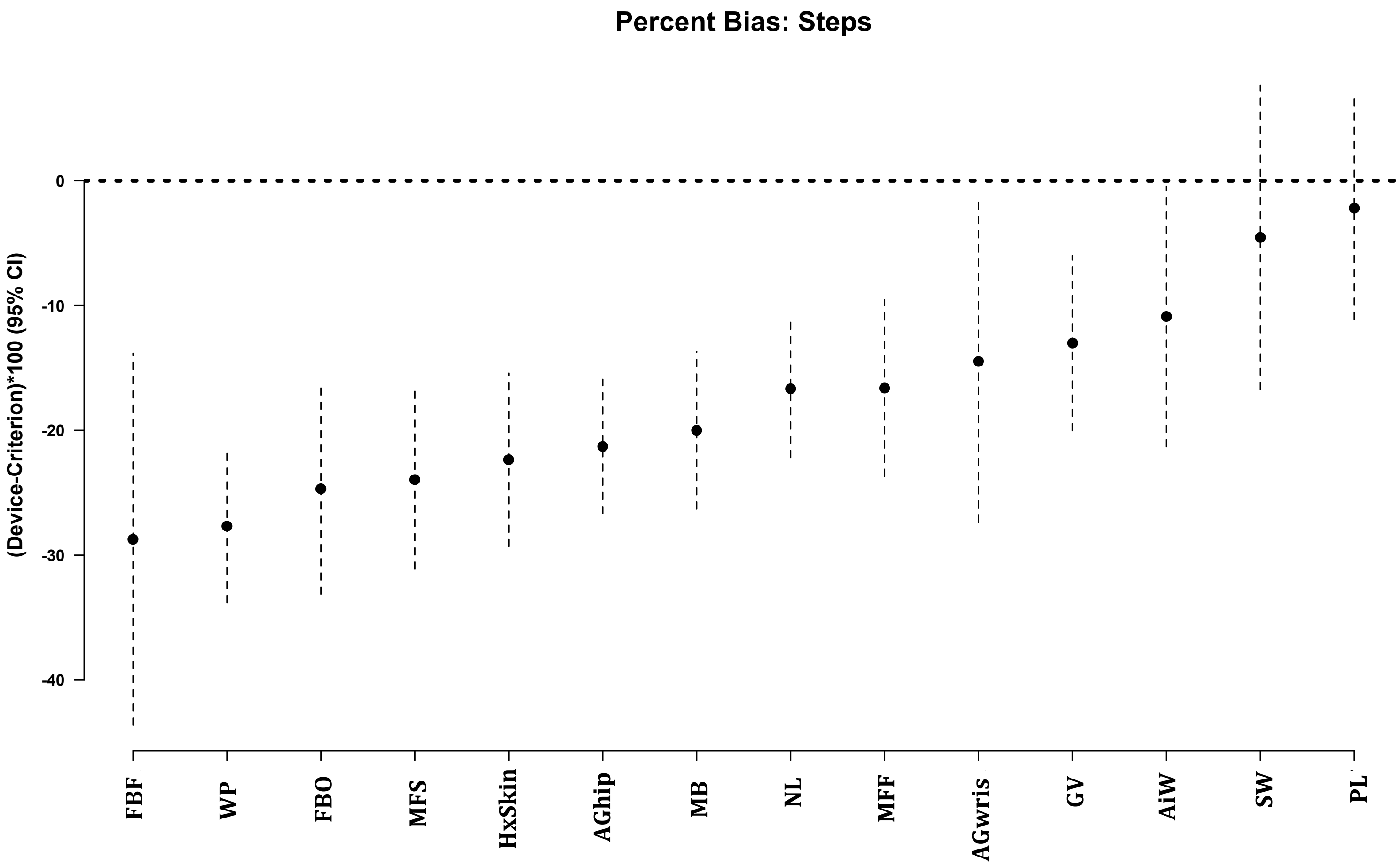


Figure 23. Percent bias Fitbit Flex (FBF), Withings Pulse (WP), Fitbit One (FBO), Misfit Shine (MFS), Hexoskin (HxSkin), hip-worn ActiGraph (AGhip), Microsoft Band (MB), NL-1000 (NL), Misfit Flash (MFF), wrist-worn ActiGraph (AGwrist), Garmin Vivofit (GV), Apple iWatch (AiW), StepWatch (SW) and Polar Loop (PL), step estimates compared to criterion steps

Data presented as mean and 95% confidence intervals

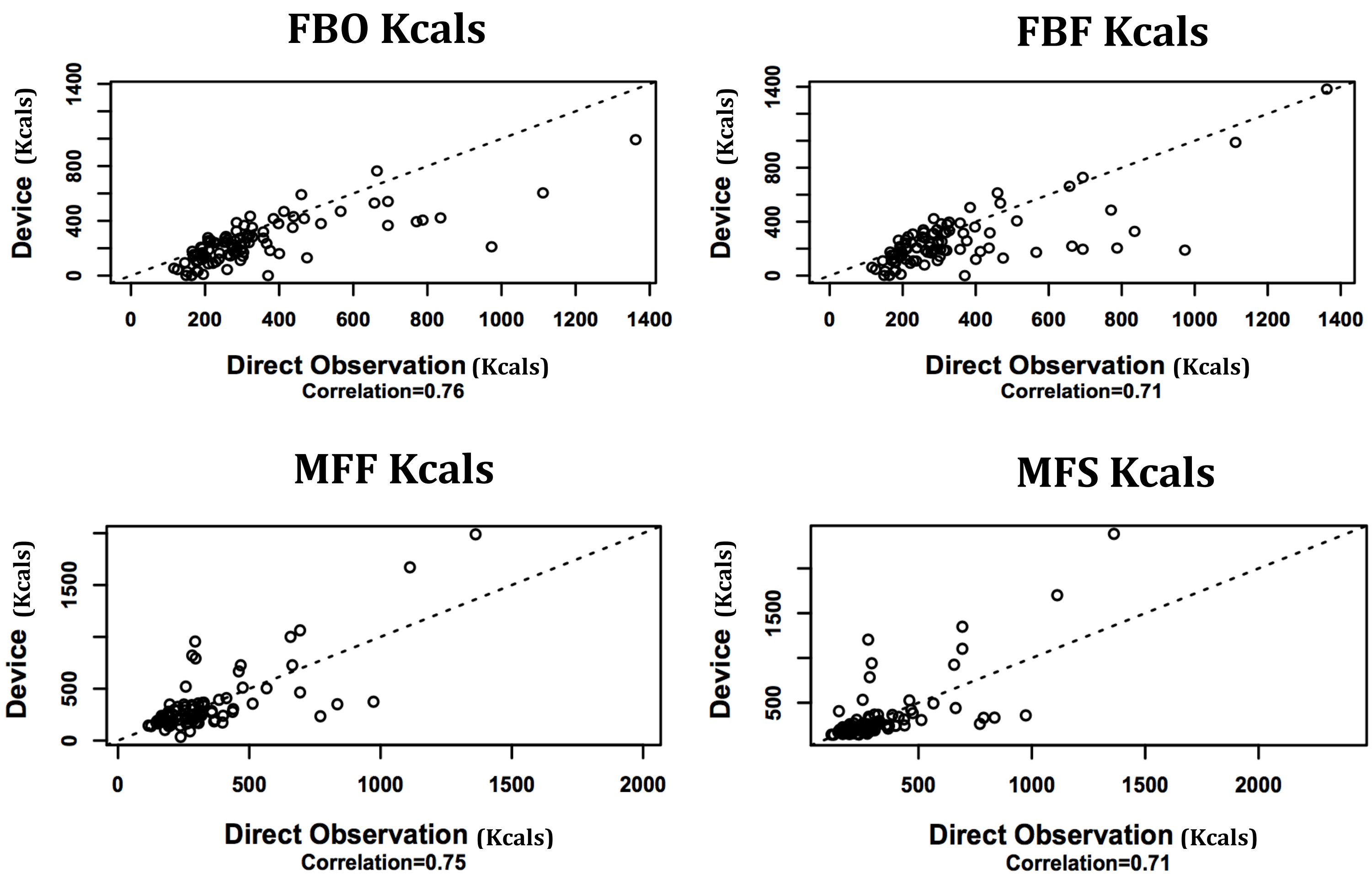


Figure 24. Relationship between criterion energy expenditure and Fitbit One (FBO), Fitbit Flex (FBF), Misfit Flash (MFF) and Misfit Shine (MFS) estimated energy expenditure

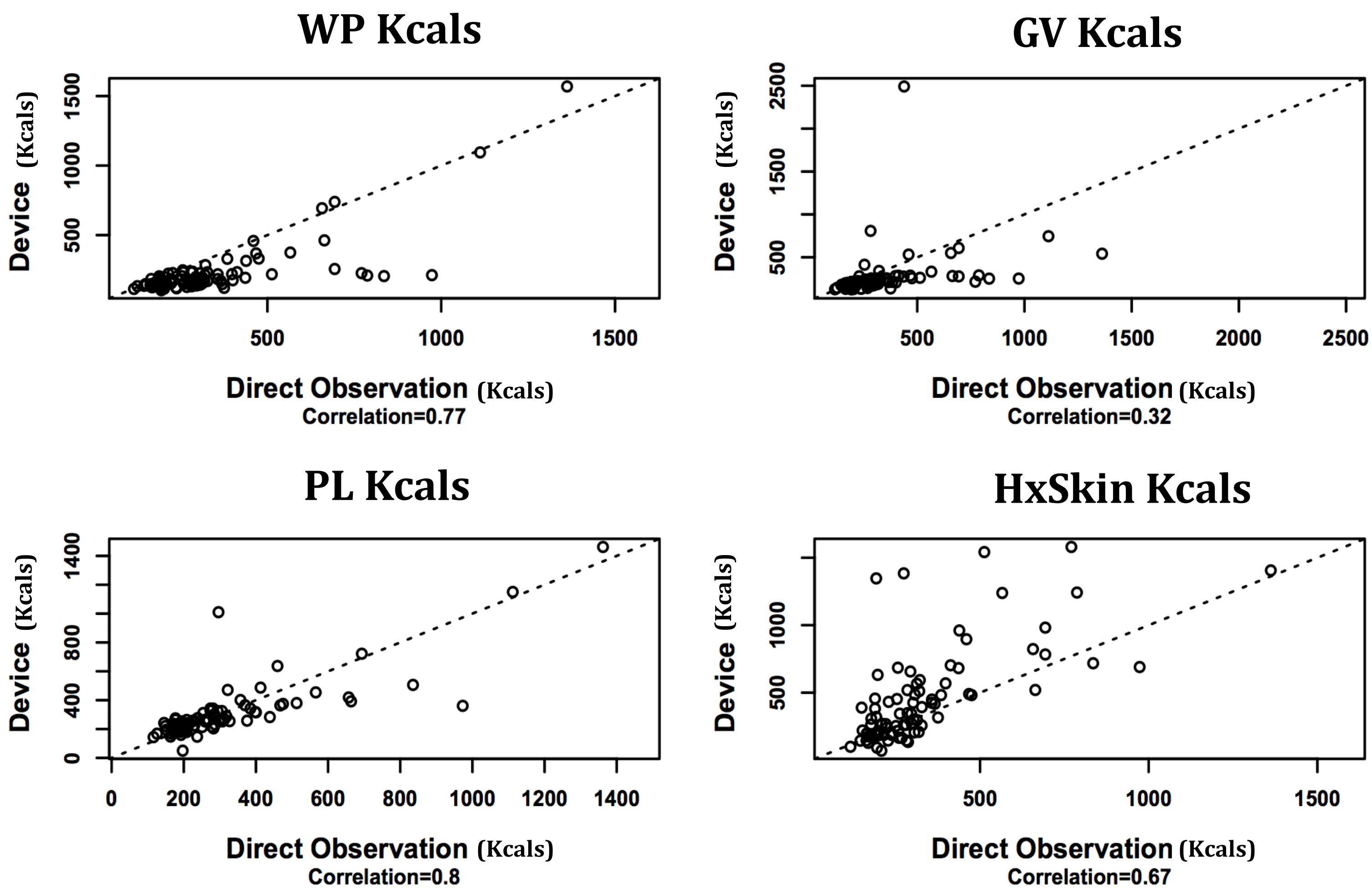


Figure 25. Relationship between criterion energy expenditure and Withings Pulse (WP), Garmin Vivofit (GV), Polar Loop (PL) and Hexoskin HxSkin) estimated energy expenditure

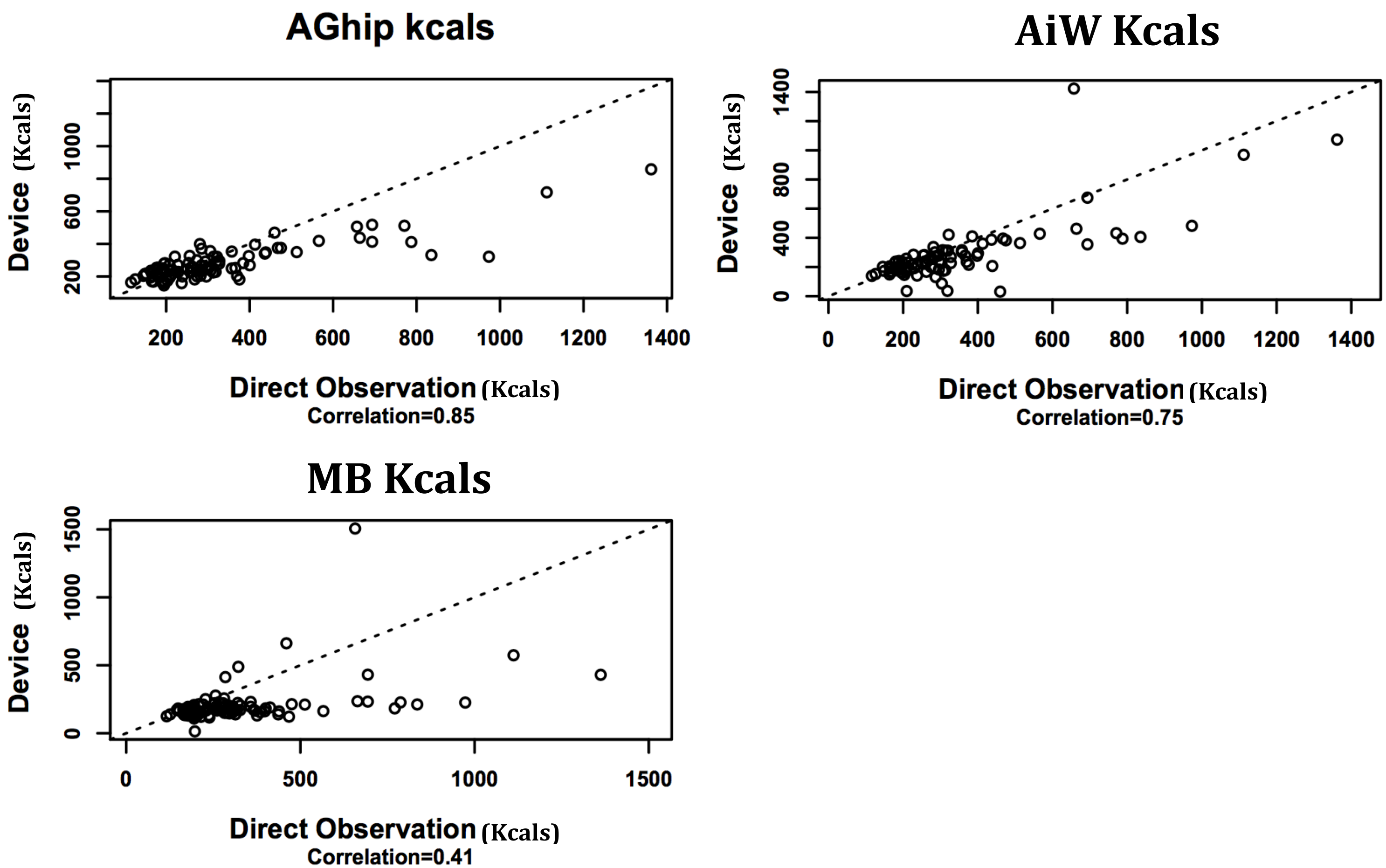


Figure 26. Relationship between criterion energy expenditure and hip-worn ActiGraph (AGhip), Apple iWatch (AiW) and Microsoft Band (MB) estimated energy expenditure

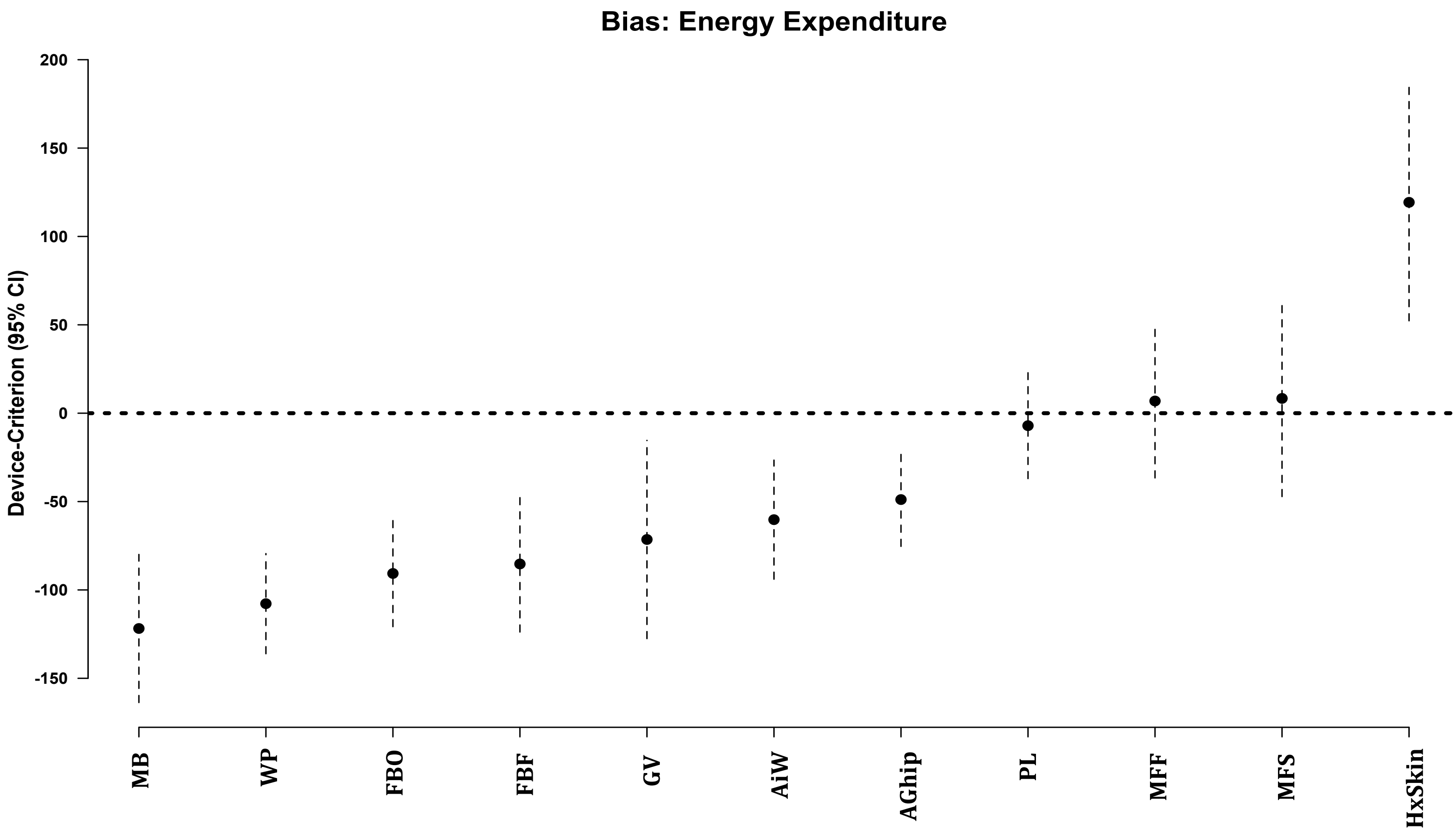


Figure 27. Bias from Microsoft Band (MB), Withings Pulse (WP), Fitbit One (FBO), Fitbit Flex (FBF), Garmin Vivofit (GV), Apple iWatch (AiW), hip-worn ActiGraph (AGhip), Polar Loop (PL), Misfit Flash (MFF), Misfit Shine (MFS) and Hexoskin (HxSkin) energy expenditure estimates compared to criterion energy expenditure

Data presented as mean and 95% confidence intervals

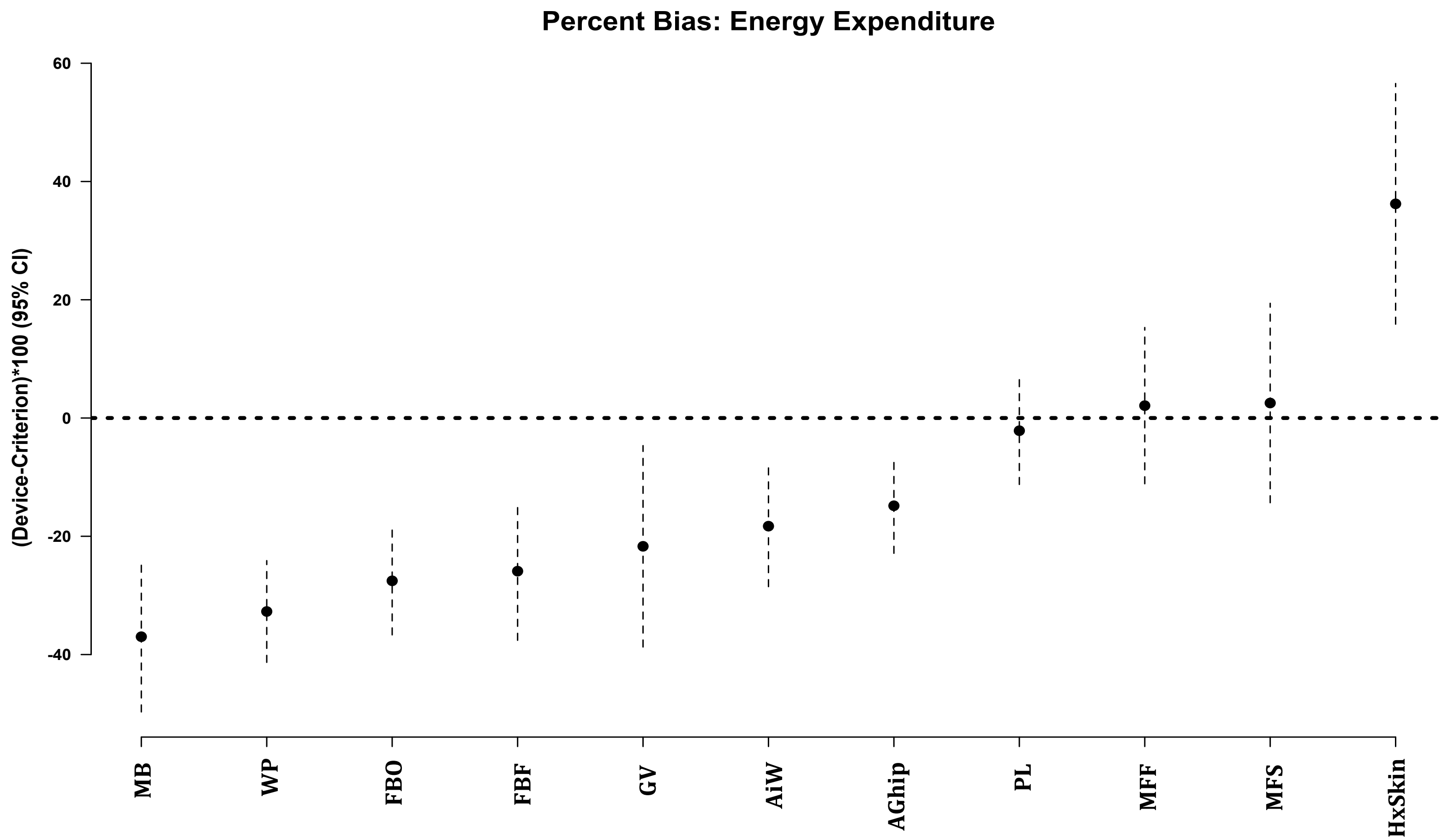


Figure 28. Percent bias from Microsoft Band (MB), Withings Pulse (WP), Fitbit One (FBO), Fitbit Flex (FBF), Garmin Vivofit (GV), Apple iWatch (AiW), hip-worn ActiGraph (AGhip), Polar Loop (PL), Misfit Flash (MFF), Misfit Shine (MFS) and Hexoskin (HxSkin) energy expenditure estimates compared to criterion energy expenditure
Data presented as mean and 95% confidence intervals

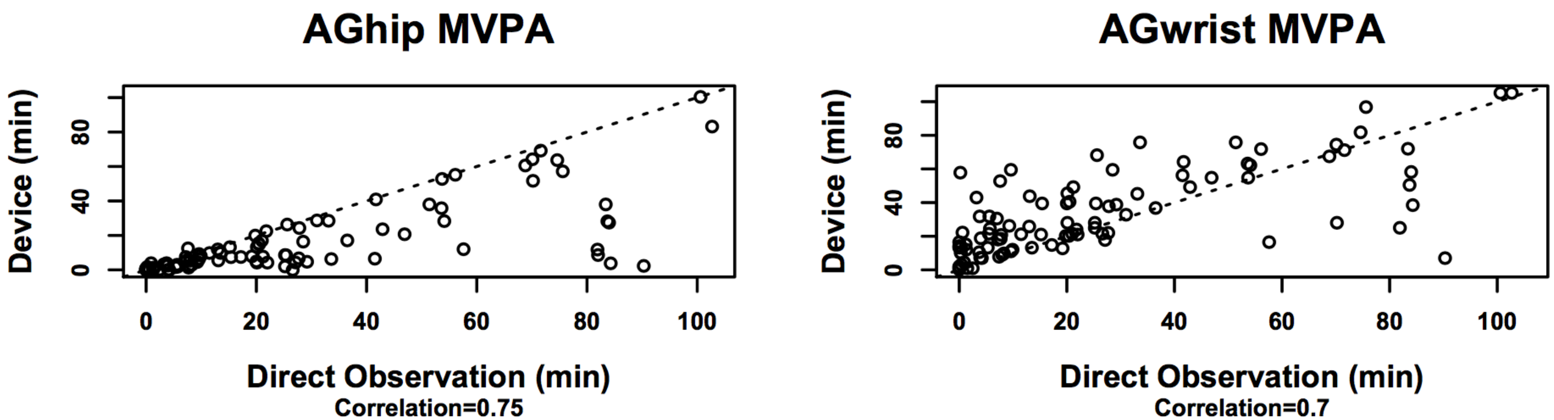


Figure 29. Relationship between criterion MVPA minutes and hip- and- wrist-worn ActiGraph (AGhip, AGwrist) estimated MVPA minutes

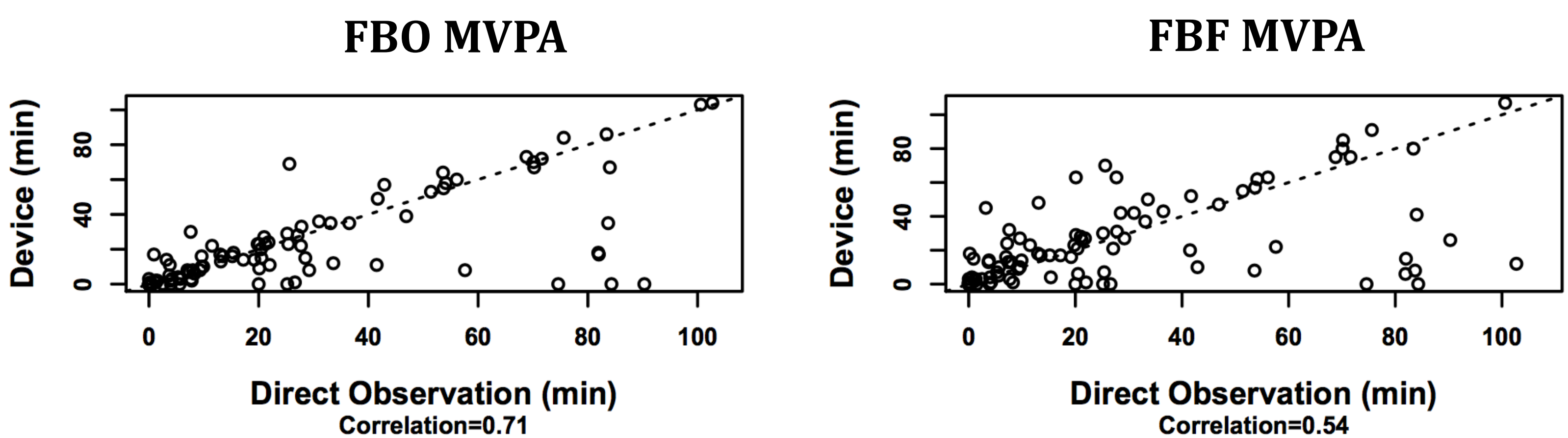


Figure 30. Relationship between Criterion MVPA minutes and Fitbit One (FBO) and Fitbit Flex (FBF) estimated MVPA minutes

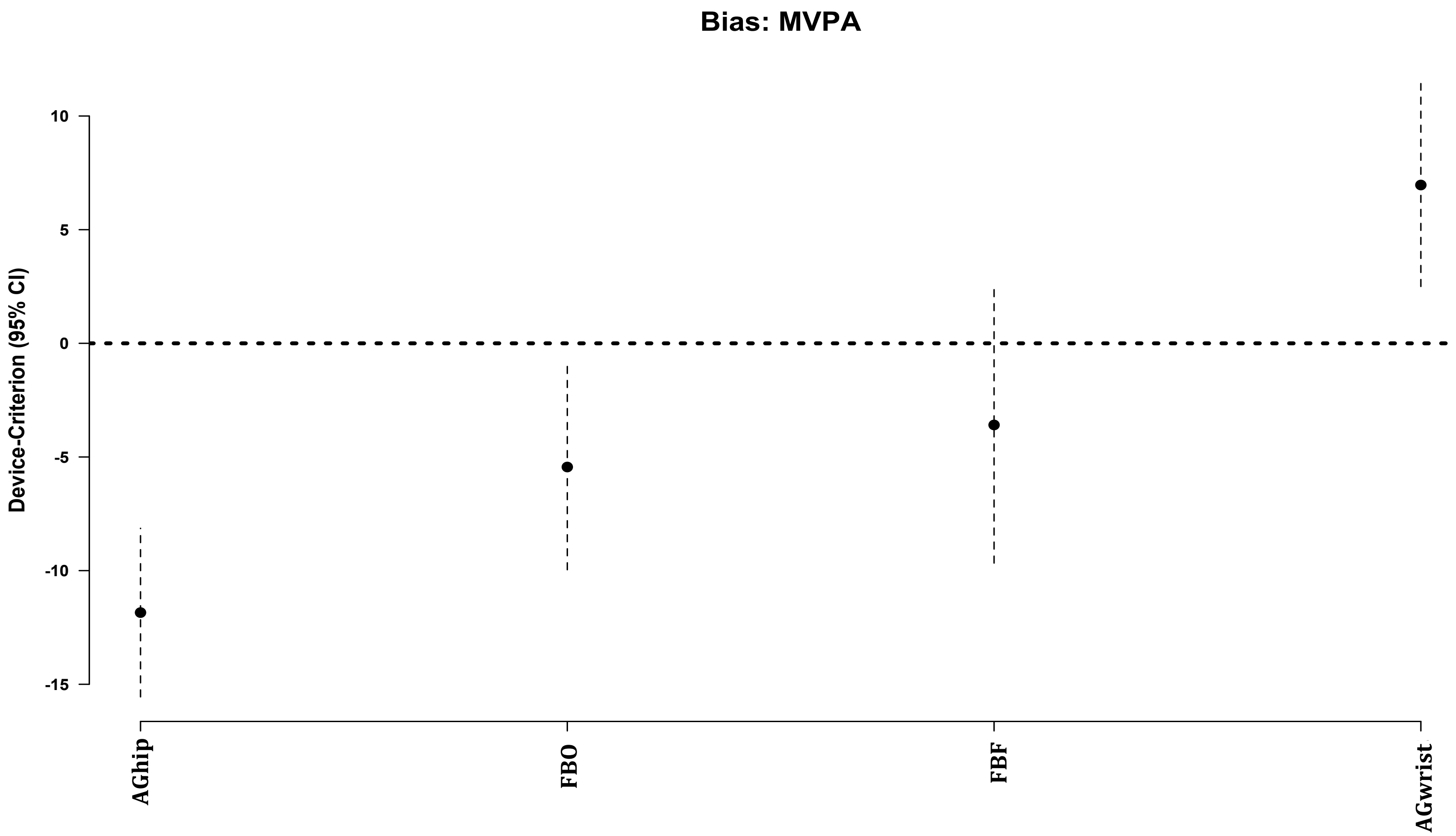


Figure 31. Bias from hip- and wrist-worn ActiGraph (AGhip, AGwrist), Fitbit One (FBO) and Fitbit Flex (FBF) MVPA minutes estimates compared to criterion MVPA minutes
Data presented as mean and 95% confidence intervals

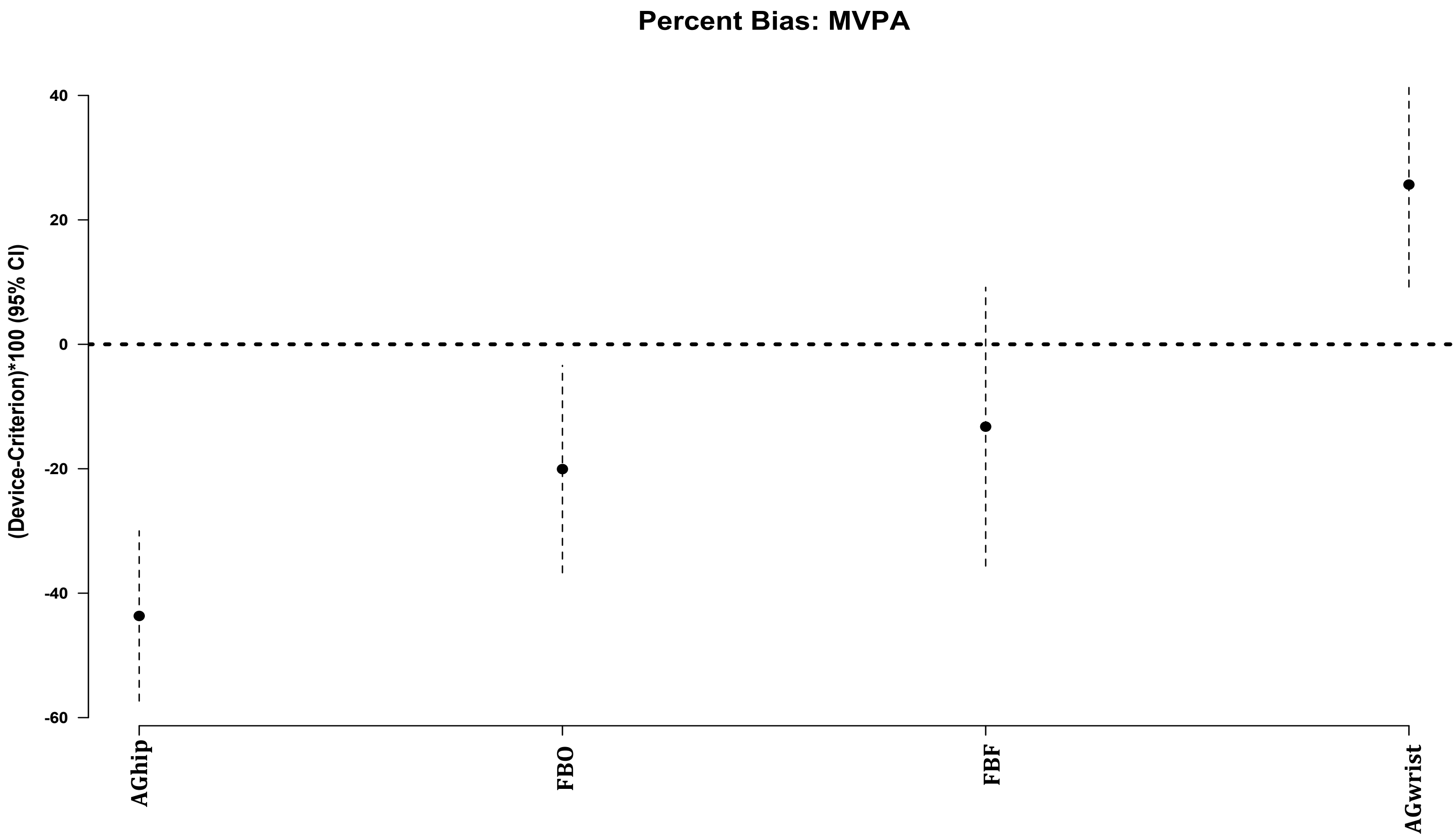


Figure 32. Percent bias from hip- and wrist-worn ActiGraph (AGhip, AGwrist), Fitbit One (FBO) and Fitbit Flex (FBF) MVPA minutes estimates compared to criterion MVPA minutes
Data presented as mean and 95% confidence intervals

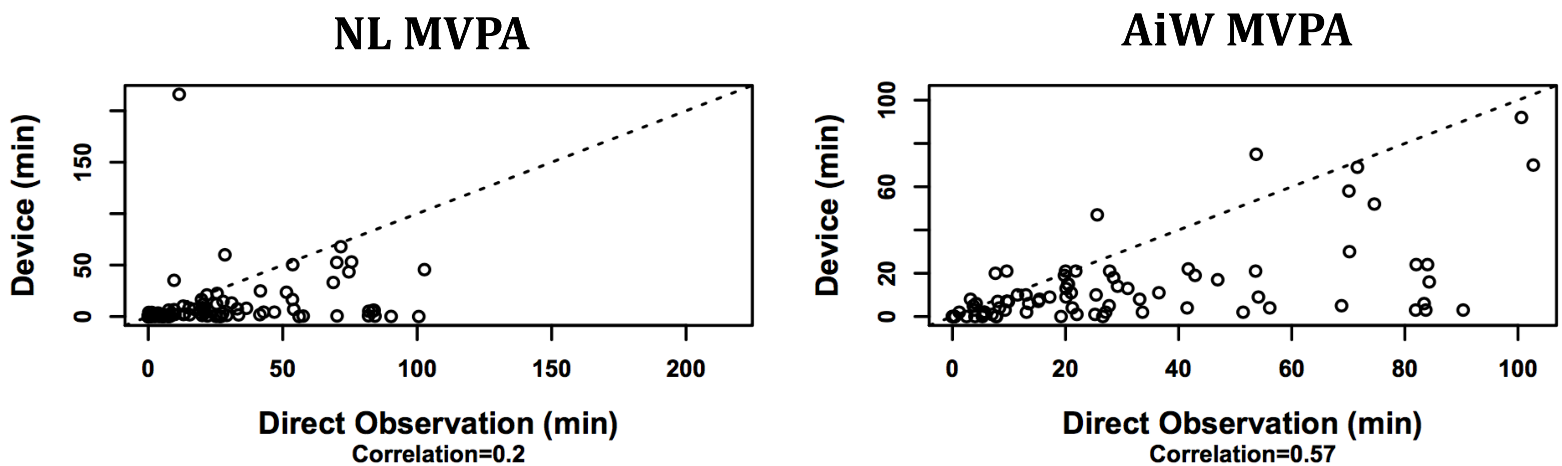


Figure 33. Relationship between criterion MVPA minutes and NL-1000 (NL) and Apple iWatch (AiW)estimated MVPA minutes

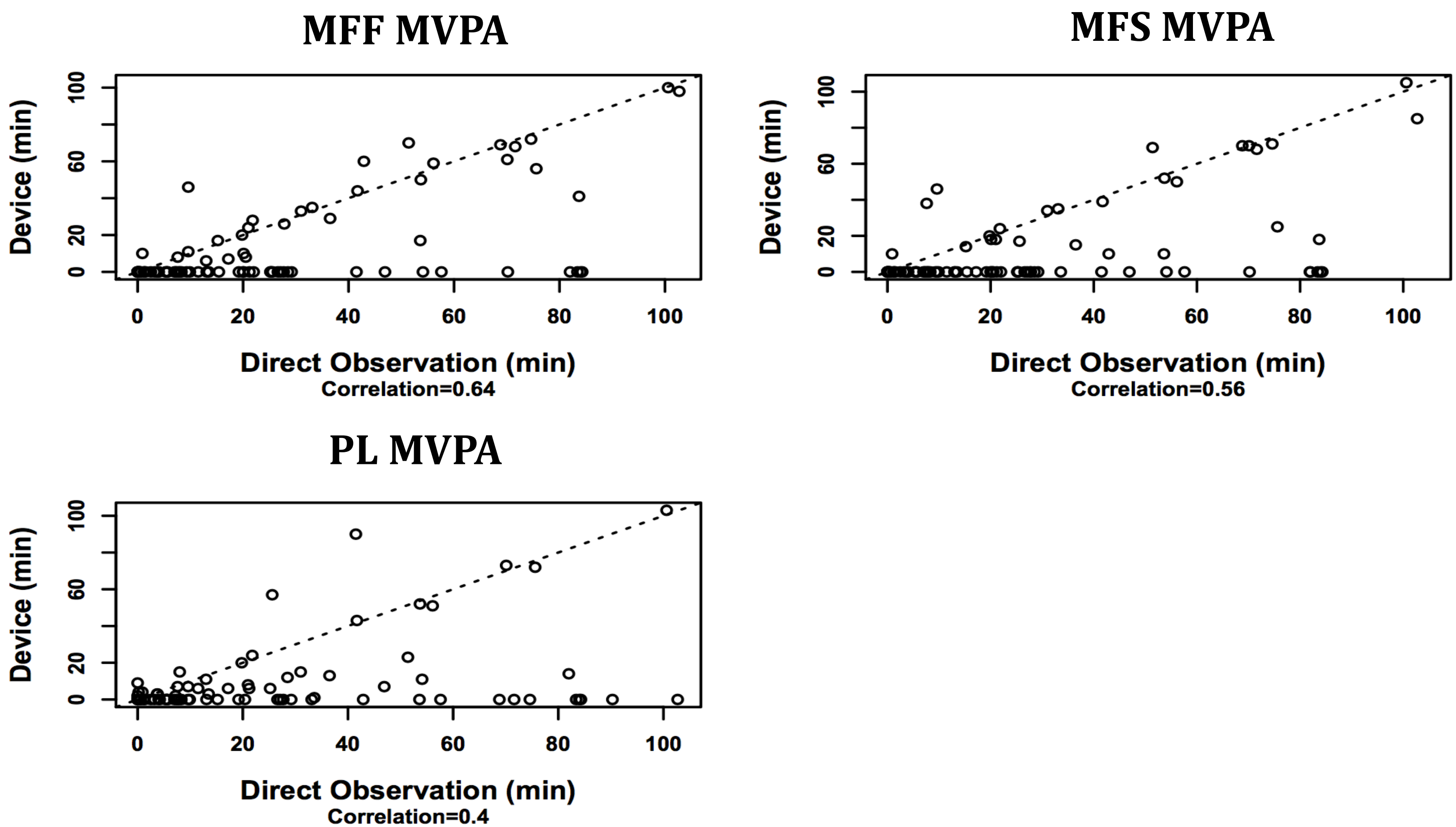


Figure 34. Relationship between criterion MVPA minutes and Misfit Flash (MFF), Misfit Shine (MFS) and Polar Loop (PL) estimated MVPA minutes

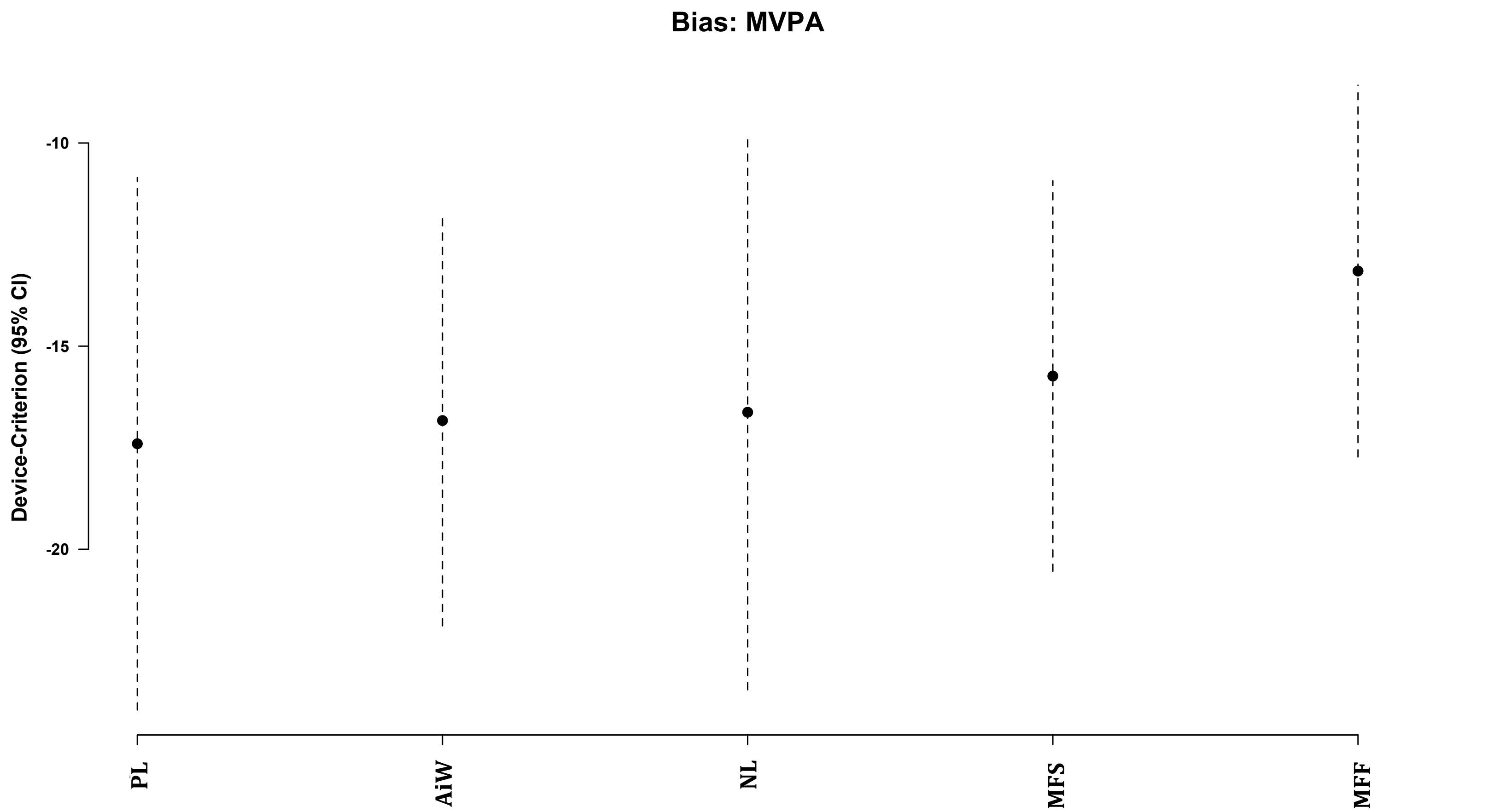


Figure 35. Bias from Apple iWatch (AiW), Polar Loop (PL), NL-1000 (NL), Misfit Shine (MFS) and Misfit Flash (MFF) MVPA minutes estimates compared to criterion MVPA minutes
Data presented as mean and 95% confidence intervals

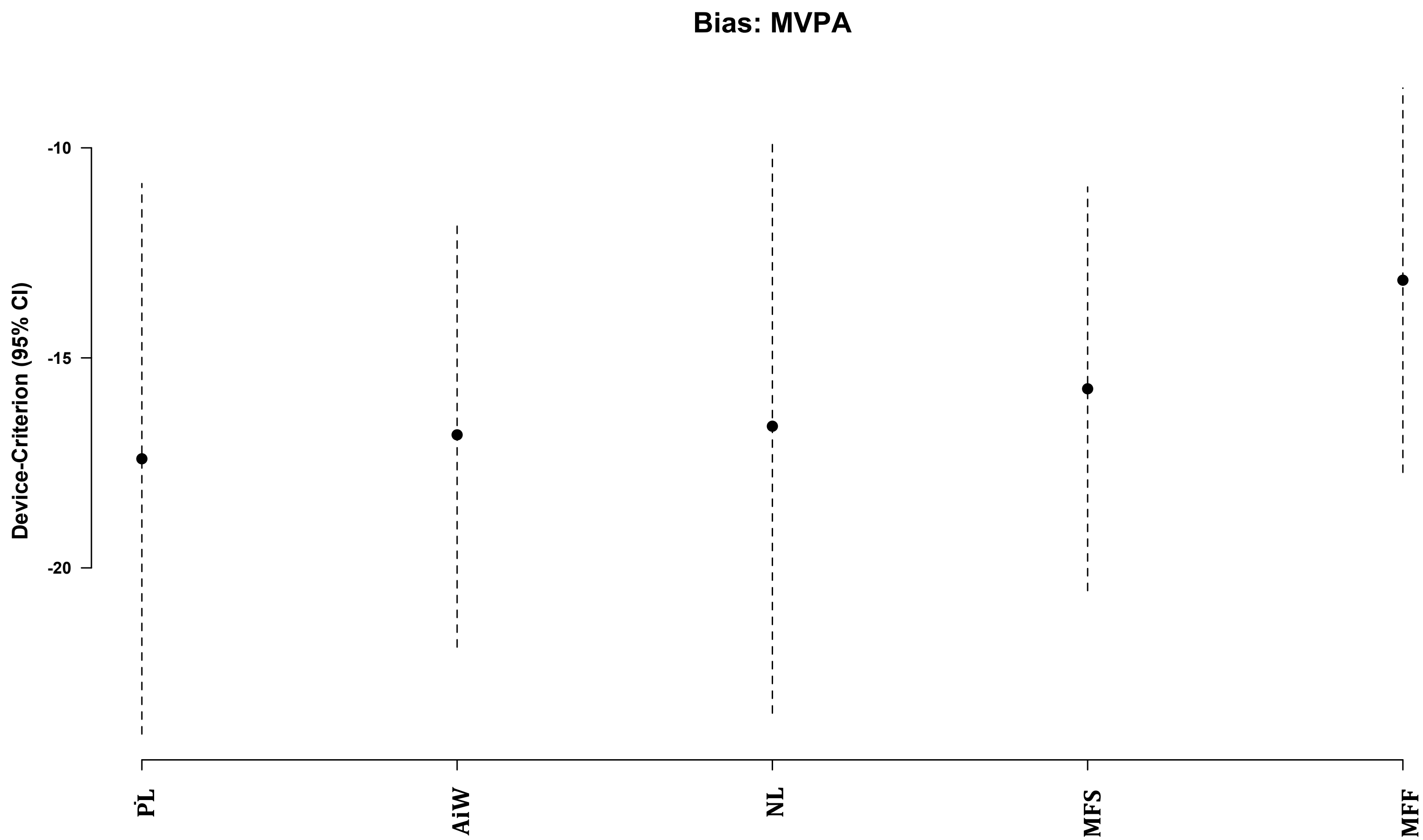


Figure 36. Percent bias from Apple iWatch (AiW), Polar Loop (PL), NL-1000 (NL), Misfit Shine (MFS) and Misfit Flash (MFF) MVPA minutes estimates compared to criterion MVPA minutes
Data presented as mean and 95% confidence intervals

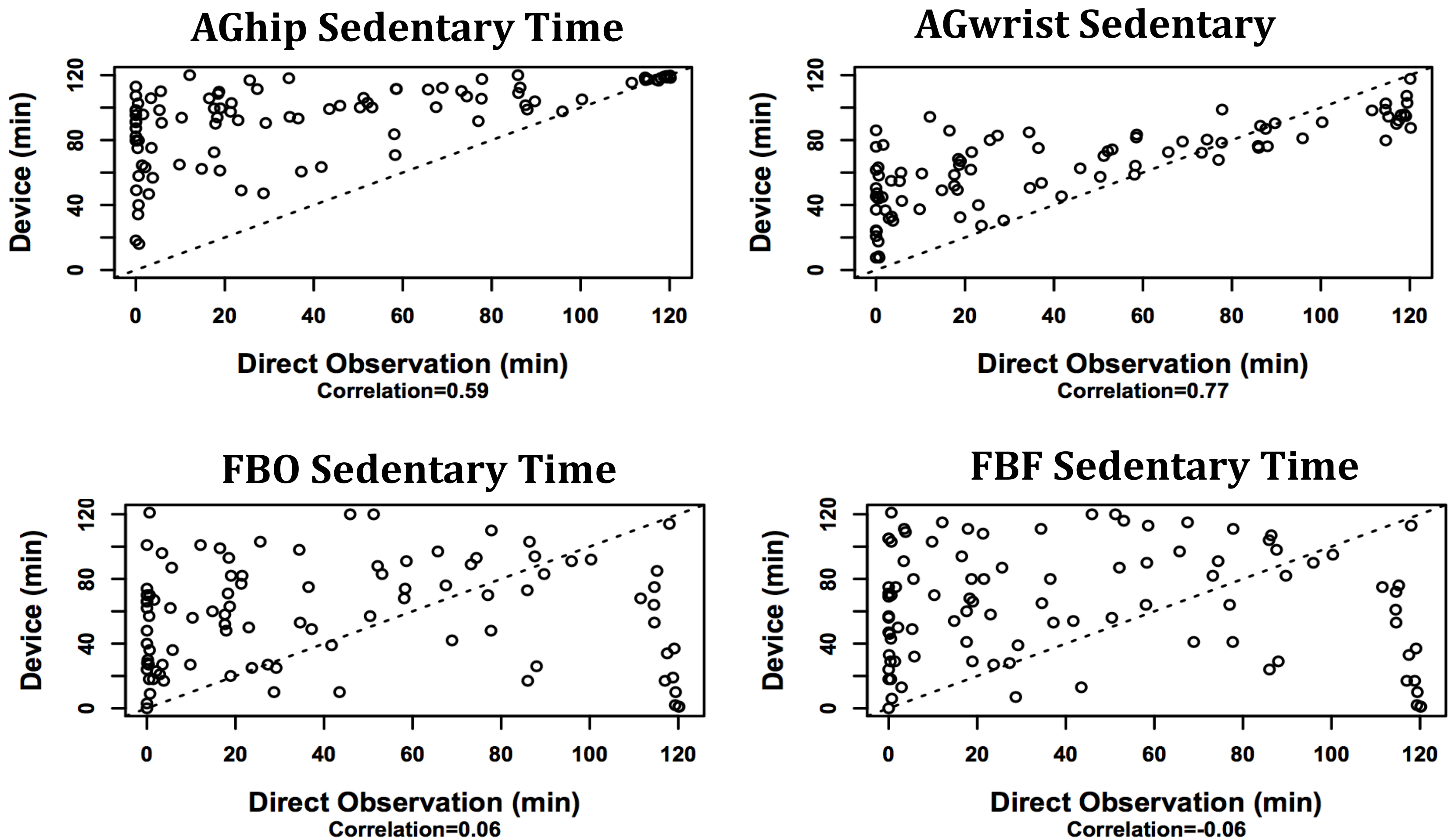


Figure 37. Relationship between criterion sedentary minutes and Fitbit One (FBO), Fitbit Flex (FBF) and hip- and- wrist-worn ActiGraph (AGhip, AGwrist) estimated sedentary minutes

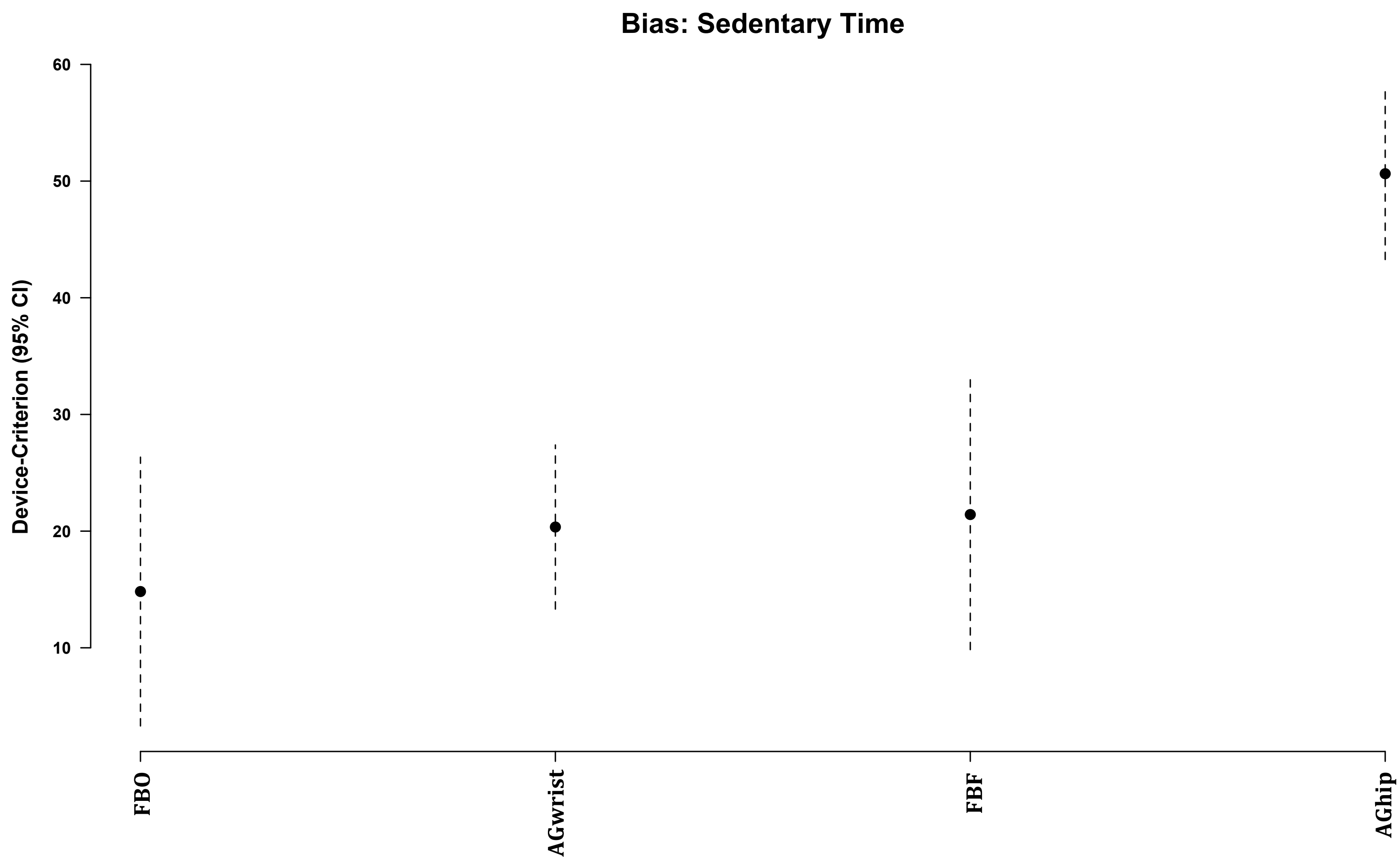


Figure 38. Bias from Fitbit One (FBO), wrist-worn ActiGraph (AGwrist), Fitbit Flex (FBF) and hip-worn ActiGraph (AGhip) sedentary minutes estimates compared to criterion sedentary minutes
Data presented as mean and 95% confidence intervals

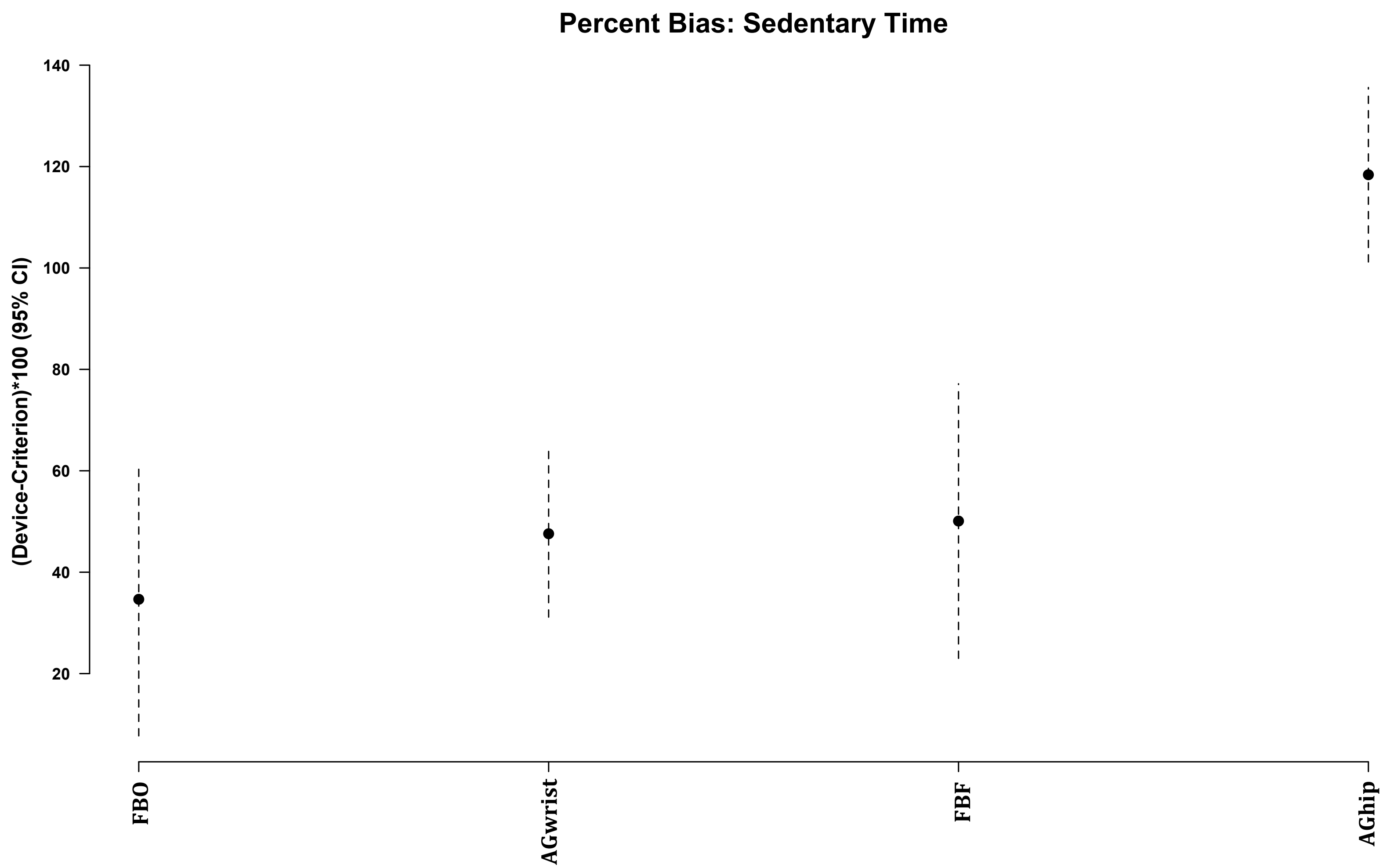


Figure 39. Percent bias from Fitbit One (FBO), wrist-worn ActiGraph (AGwrist), Fitbit Flex (FBF) and hip-worn ActiGraph (AGhip) sedentary minutes estimates compared to criterion sedentary minutes
Data presented as mean and 95% confidence intervals

CHAPTER 6
STUDY THREE - ACTIVITY TRACKERS ARE SENSITIVIE TO CHANGE IN PHYSICAL ACTIVITY AND SEDENTARY BEHAVIORS IN FREE-LIVING SETTINGS

Introduction

Tools such as wearable devices to track personal physical activity (PA) provide a mechanism to be more informed about activity behavior. Consumer devices that track PA behavior are increasingly popular for consumers, researchers, clinicians and of interest to the National Institutes of Health[66] who recognize the value of using sensor-based wearable monitors to assess PA behaviors. Consumers are using these devices to monitor and track personal PA. In many cases, clinicians and researchers are using consumer devices to track change in PA behavior[44] and to assess PA exposure[170] and outcomes.[165] Currently, there are at least 149 active or recruiting clinical trials funded by NIH that are employing consumer ATs to measure (estimate) change in PA behaviors such as energy expenditure (EE) and/or steps. [67]

The research and clinical communities have rapidly adapted ATs, however, their utility within these communities has yet to be realized. Moreover, unlike research-grade devices that have been utilized by the research and clinical communities in the past, ATs have yet to undergo rigorous testing in both laboratory and free-living settings. In particular, there is no evidence examining the effectiveness of ATs for detecting change in PA behaviors in free-living settings. This knowledge gap is of major concern since ATs are widely used to monitor change in PA behaviors. Therefore, the aims of the present study were to: 1) examine the ability of ATs to detect change in PA and ST in free-living settings and 2) examine the ability of research-grade accelerometers to detect change in PA and ST in free-living settings.

Methods

Procedures

The data used in the present study are from our previous study, "Validation of Consumer and Research-Grade Activity Monitors in Free-Living Settings." Briefly, thirty-two healthy men and women (50% female, 37.5% minority; mean ± SD: Age = 32.3 ± 13.3 years; BMI = 24.4 ± 3.3 kg·m-2) were directly observed while completing three, 2-hour visits on different days. During these visits, participants wore several different ATs, research-grade devices and a biometric shirt. At the end of each visit, data from all devices were recorded and processed for analysis. For comparison, a validated DO system (The Observer XT) was used to compute criterion measures for activity and sedentary time outcomes. (See Tables 12 and 13 for detailed description of devices).

Data Processing and Statistical Evaluation

For both the criterion measure and the device estimates, we calculated the differences between the visits (i.e. visit 1 minus visit 2, visit 1 minus visit 3 and visit 2 minus visit 3) for estimated steps, EE (Except AGwrist. No validated EE estimates from a wrist-worn AG), activity minutes, and or sedentary time. We then classified the criterion and device measured outcomes for visit-to-visit change into one of three categories: increase, no change or decrease where an increase or decrease was defined as a change that was greater than the within-subject standard deviation of the criterion measure (estimated by a linear-mixed model). Finally, confusion matrices were used to determine percent agreement between criterion visit-to-visit change and device visit-to-visit change.

All data cleaning, processing and analysis were performed using the open source *R statistical software package* (www.r-project.org) and computing language R. [136]

Results

Table18 shows percent agreement between criterion measured visit-to-visit change and device estimated visit-to-visit change for each output metric. Correct classification of steps ranged from 79.2% (AiW) to 93.3% (MFS and WP). Correct classification of energy EE ranged from 71.2% (MB) to 82.1% (AiW). Correct classification of moderate-to-vigorous physical activity (MVPA) change in minutes ranged from 77.6% (FBF) to 74.7% (AGwrist). Non-Guideline MVPA minutes (previously described in study 2, see Table 15 for definitions) ranged from 58.4 (PL) to 73.8% (MFF). Correct classification of sedentary time change ranged from 43.4 % (FBF) to 53.1% (AGhip).

Figures 40 to 43 illustrate criterion measured visit-to-visit change with FBO (A) and/or FBF (B) visit-to-visit change, for steps, EE, MVPA minutes and sedentary time. Correct classifications ranged from 46.8% (sedentary minutes) to 89.1% (steps) and from 43.4% (sedentary minutes) to 88.3% (steps) for the FBO and FBF, respectively.

Figure 44 illustrates criterion measured visit-to-visit change versus SW visit-to-visit change, for steps (91.1% correct classification).

Figures 46 to 48 illustrate criterion measured visit-to-visit change versus AGhip (A) and/or AGwrist (B) visit-to-visit change, for steps, EE, MVPA minutes and sedentary time. Correct classifications ranged from 53.1% (sedentary minutes) to 91.4% (steps) and from 53.1% (sedentary minutes) to 88.3% (steps) for the AGhip and AGwrist, respectively.

Discussion

The purpose of the present study was to examine the ability of ATs to detect change in PA and ST during free-living time. This discussion will highlight key findings from the FBF and FBO since this AT is the one most widely used in intervention research and by consumers. Discussion will also include an analysis of the change classification results for the AG hip, AGwrist and SW research-grade devices (Table 18 presents a summary of the percent agreement between criterion visit-to-visit change and device estimated visit-to-visit change for all the outcome measures).

All ATs detected change in PA with varying levels of agreement with criterion change (see Appendices G – J). For example, percent agreement for ATs that provided estimates of Non-Guideline MVPA minutes (previously described in study 2, see Table 15 for definitions) ranged from 55.8% (PL) to 71.4% (MFF). Fitbit Flex and FBO percent agreement for steps, EE and Guideline MVPA minutes (previously described in study 2) were 65% or greater. Agreement was lowest for sedentary time (46.8% FBO, 42.3% FBF) (Figures 39-42).

Fitbit and AG change estimates for PA and ST were similar to criterion measures with a few exceptions. For the hip location, percent agreement for Guideline MVPA minutes and sedentary time was approximately 8% and 12% higher for the FBO versus the AGhip, respectively. For the wrist location, percent agreement for Guideline MVPA minutes was approximately 8% higher for the FBF versus the AGwrist. Additionally, percent agreement for sedentary time was approximately 30% higher for the AGwrist versus the FBF. These findings suggest that the hip-worn FBO and FBF may be suitable alternative devices to research-grade devices for detecting change in free-living steps, EE and Guideline MVPA minutes. However, accurately estimating change in sedentary time will require further refinement in prediction models for this behavior.

Currently, there are at least 117 active clinical trials employing the FB as either an outcome or an exposure measure of PA, and the rate of adoption of this tool is rapidly increasing.[67] To date, the ability of these measurement tools to detect increases, decreases or no change in PA behaviors is largely unknown. The evidence from the present study is the first study addressing this issue. The findings of this study support the use of this device to detect and monitor changes in free-living steps, EE and Guideline MVPA minutes. The accuracy of the FB for detecting and monitoring change in ST is not sufficient and thus changes in ST may be harder to accurately assess from the FB and AG.

All research-grade devices (AGhip, AGwrist and SW) detected change in PA (see Figures 43-46). The highest percent agreement was for steps (88.3% AGwrist, 91.1% SW, 91.4% AGhip classification), followed by EE (77.0% AGhip classification), MVPA minutes (71.2% AGwrist, 77.0% AGhip classification) and sedentary time (53.1% AGhip, 72.7% AGwrist classification). Hip-and ankle-worn research-grade devices have been examined and been shown to detect change in steps and activity minutes in lab-based settings.[96,171,172] We were unable to examine changes in EE from the AGwrist since there are no widely accepted validated algorithms to estimate EE from this wrist-worn AG. There are no validated EE estimates from a wrist-worn AG.

This study has several strengths. The primary strength was the study design. These data are from our validation study of AT in free-living settings, where we employed a validated direct observation (DO) system as the criterion measure.[42] Our use of DO to derive criterion measures of PA and ST is a major advance in this line of research since previous free-living studies employed accelerometer estimates of activity and ST as a substitute for gold-standard criterion measures to assess PA.[32-34,43-46] Second, an ecological study setting allowed us to examine AT performance while participants wore them in their natural environment which has high research-translation value. Third, the use of within-subject standard deviation (SD) of the criterion measures allowed us to use an evidence-based behavior cut-point of the minimum outcome level to define change. For example, based on our data, if an activity intervention observed a ±3,000 step/2hr change, the FBF could detect this change.

This study also has limitations. We employed a validated DO system that uses the Compendium of Physical Activities to apply MET values to activities. The values in the Compendium do not estimate the energy cost of physical activity in individuals that account for differences in body mass, adiposity, age, sex, efficiency of movement, and environmental conditions in which the activities are performed.[146] Therefore, it is possible that activities were misclassified by intensity category, which may have resulted in inaccuracies of criterion EE, activity minutes and sedentary time.. The trial duration was another limitation. We observed participants for three, 2-hour time frames versus a whole-day, thus, our findings may not represent change in whole day behavior.

In summary, the present study is a major advance beyond traditional validation studies in the lab and simulated free-living studies where activities are performed over fixed time and activity menu driven fixed time and activity studies. This study used a novel protocol that is truly free-living, which is relevant to real-life applications. Thus far, no studies have examined ATs ability to detect change in PA and ST in free-living settings where these devices are used. Our findings suggest that in general, there is similar agreement between the hip-worn FBO and FBF with hip- and- wrist-worn AGs in estimates of change in steps, EE (accept AGwrist) and MVPA minutes (except FBF) with criterion measured change. However, change in ST was more difficult to detect for the FB and AGhip.

Device	Apple iWatch Sport	Fitbit Flex	Fitbit One	Garmin Vívofit	New Lifestyles NL-1000	Microsoft Band	Misfit Flash	Misfit Shine	Polar loop	Withings Pulse
Cost	$350.99	$79.95	$99.95	$99.99	$54.95	$199.99	$29.99	$69.99	$109.95	$119.95
Wear location	Wrist	Wrist	Clip on (multiple locations)	Wrist	Hip	Wrist	Clip on (multiple locations)	Clip on (multiple locations)	Wrist	Clip on and wrist band
Tracks Calories Burned	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Active Time	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Steps	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Distance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Elevation/Stairs	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓
Tracks Sleep	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Heart Rate	✓	✗	✗	✗	✗	✓	✗	✗	✓	✓
Battery or Chargeable	Chargeable (every 18 hours)	Chargeable (every 5 days)	Chargeable (every 10+ days)	Battery (every 1+ years)	Battery (up to 18 months)	Chargeable (every 48 hours)	Battery (lasts up to 6 months)	Battery (lasts up to 6 months)	Chargeable (up to 6 days)	Chargeable (every 2 days)
Uploading Data	Bluetooth	Bluetooth	Bluetooth	Bluetooth	Real-time data	USB			USB	Bluetooth
Tracker Display	Real-time data	LED progress indicator	Real-time data	Real-time data	Real-time data	Real-time data	LED progress indicator	LED progress indicator	Real-time data	Real-time data

Table 15. Features of consumer-based activity trackers

LED, Light-Emitting Diode; USB, Universal Serial Bus

Device (Location)	Output	Data Extraction Method	
		Upload	**Retrieval**
Apple iWatch Sport (W)	EE, steps, active calories, min: exercise, total active time, stand hours	Bluetooth	Apple Activity App
GT3X-BT (W & H)	Steps, min: Sedentary, light, moderate, vigorous	USB cable	ActiLife
Fitbit Flex /One (W/H)	EE, steps, MET-min, min: sedentary, light, moderate, vigorous	Bluetooth	Fitabase
Garmin Vivofit (W)	EE, steps, active calories, %: sedentary, active, highly active	Bluetooth	Garmin Connect™ App
Hexoskin (T)	EE, steps	USB cable	Hexoskin dashboard
Microsoft Band (W)	EE, steps, active min	USB cable	MB dashboard
Misfit Flash/Shine (H/W)	EE, steps, active min: light, moderate, vigorous	Bluetooth	Misfit App
New Lifestyles NL-1000 (H)	Steps, MVPA min	RTD	RTD
The Observer XT (NA)	MET-hours, MET-min	The Observer XT	The Observer XT
Polar Loop (W)	EE, steps, time: lying, sitting, active, sitting, min: stand, walk, run	USB cable	Polar dashboard
StepWatch (A)	Steps	USB cable	StepWatch dashboard
Withings Pulse (H)	EE, steps	Bluetooth	Withings App

Table 16. Device output and data extraction methods

H, hip; W, wrist: T, torso; A, Ankle; NA, not applicable; EE, energy expenditure; min, minutes; MVPA, moderate-to-vigorous physical activity; MB, Microsoft Band; RTD, real-time display

Device	Output	Definition
Apple iWatch	Exercise minutes	Anything above a brisk walk is classed as exercise. Every full minute of movement equaling or exceeding the intensity of a brisk walk counts towards daily Exercise goal (30 min).
Fitbit Flex/One	Active minutes	Activities at or above about 3 METs. Minutes are only awarded after 10 minutes of continuous moderate-to-intense activity.
Misfit Flash/Shine	Light-, moderate-, vigorous- minutes	No definitions provided.
NL-1000	MVPA	Moderate-to-vigorous physical activity (MVPA) time accumulation.
Polar Loop	WALK and JOG	Medium and high intensity activity, respectively.

Table 17. Activity tracker intensity outputs and definitions

		Device – Percent Agreement (%)													
	CRIT w/i Sub SD	AGhip	AGwrist	SW	AiW	FBF	FBO	GV	HxSkin	MB	MFF	MFS	NL	PL	WP
Metric															
Steps	±2,809	91.4	88.3	91.1	79.2	88.8	89.1	88.0	89.4	89.1	82.5	93.3	91.4	84.9	93.3
EE (kcals)	±213.0	77.0	NA	NA	82.1	72.8	76.5	72.3	78.2	71.2	77.2	77.7	NA	74.0	78.2
MVPA (min)	±28.0	77.0	71.2	NA	67.1ǂ	65.2	79.7	NA	NA	NA	71.4ǂ	64.1ǂ	63.5ǂ	55.8ǂ	NA
SED (min)	±41.0	53.1	72.7	NA	NA	42.3	46.8	NA	NA	NA	NA	NA	NA	NA	NA

Table 18. Percent agreement between criterion measured visit-to-visit change and device estimated visit-to-visit change for each output metric

CRIT, criterion; Sub, subject: SD, standard deviation; AGhip, hip-worn GT3X-BT; AGwrist, wrist-worn GT3X-BT; SW, StepWatch; AiW, Apple iWatch; FBF, Fitbit Flex; FBO, Fitbit One; GV, Garmin Vivofit; HxSkin, Hexoskin; MB, Microsoft Band; MFF, Misfit Flash; MFS, Misfit Shine; NL, New Lifestyles-1000; PL, Polar Loop; WP, Withings Pulse; MVPA, moderate-to-vigorous physical activity; SED, sedentary; min, minutes; EE, energy expenditure; kcals, calories; NA, not applicable;

ǂ, Non-Guideline MVPA minutes.

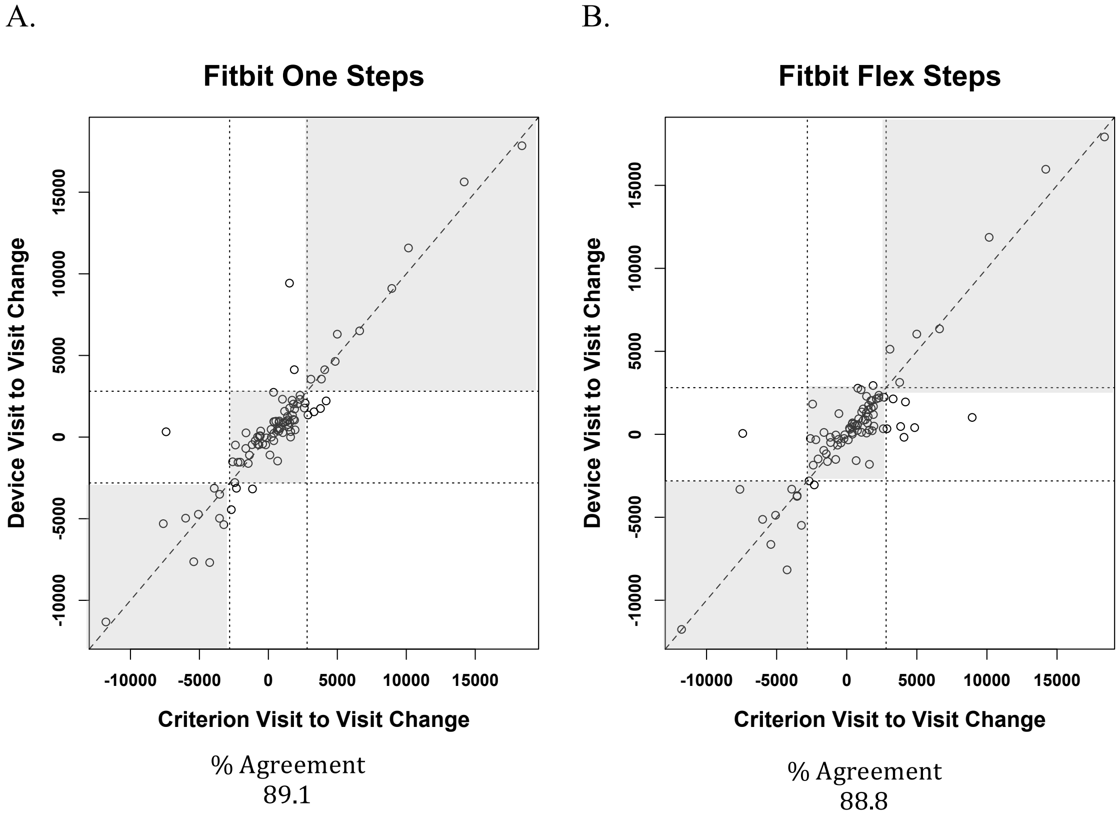


Figure 40. Steps: criterion measured visit-to-visit change and Fitbit One (A) Fitbit Flex (B) visit-to-visit change

The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

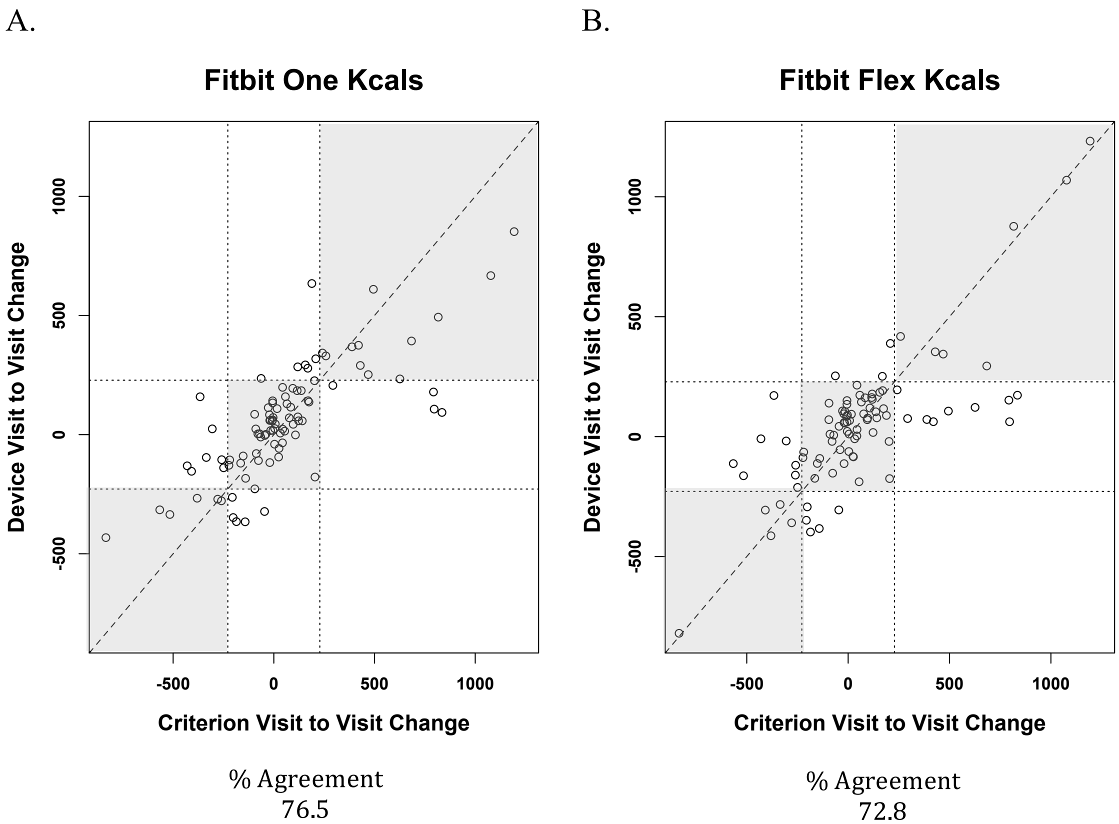
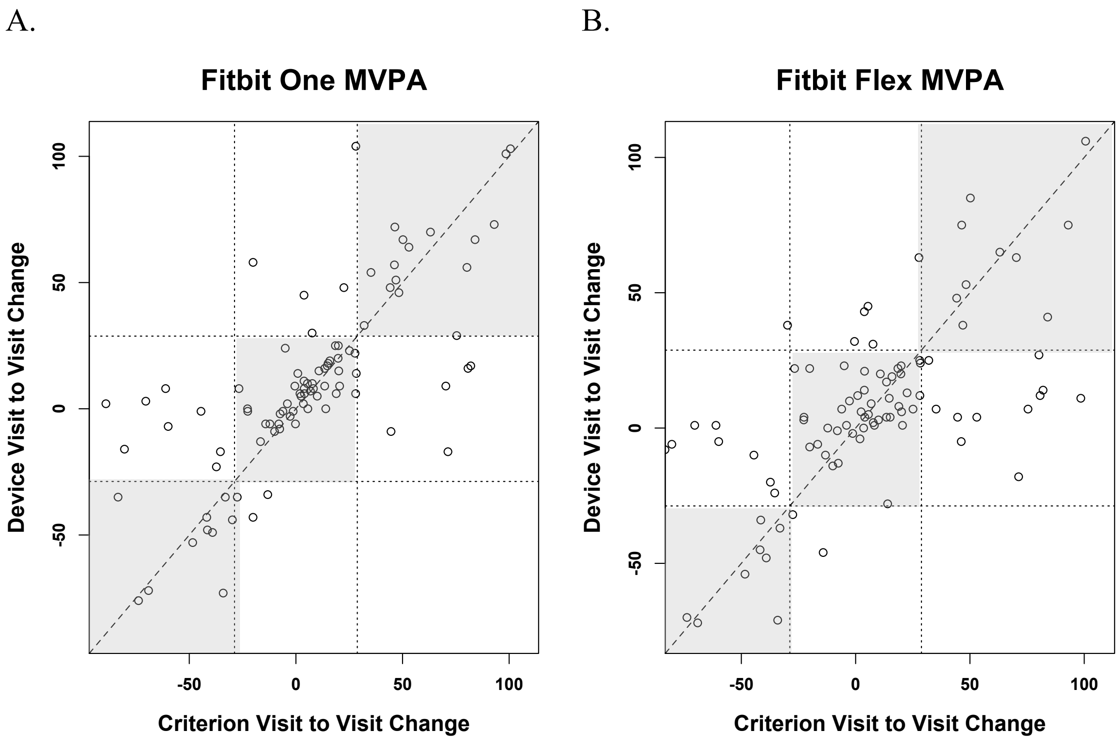


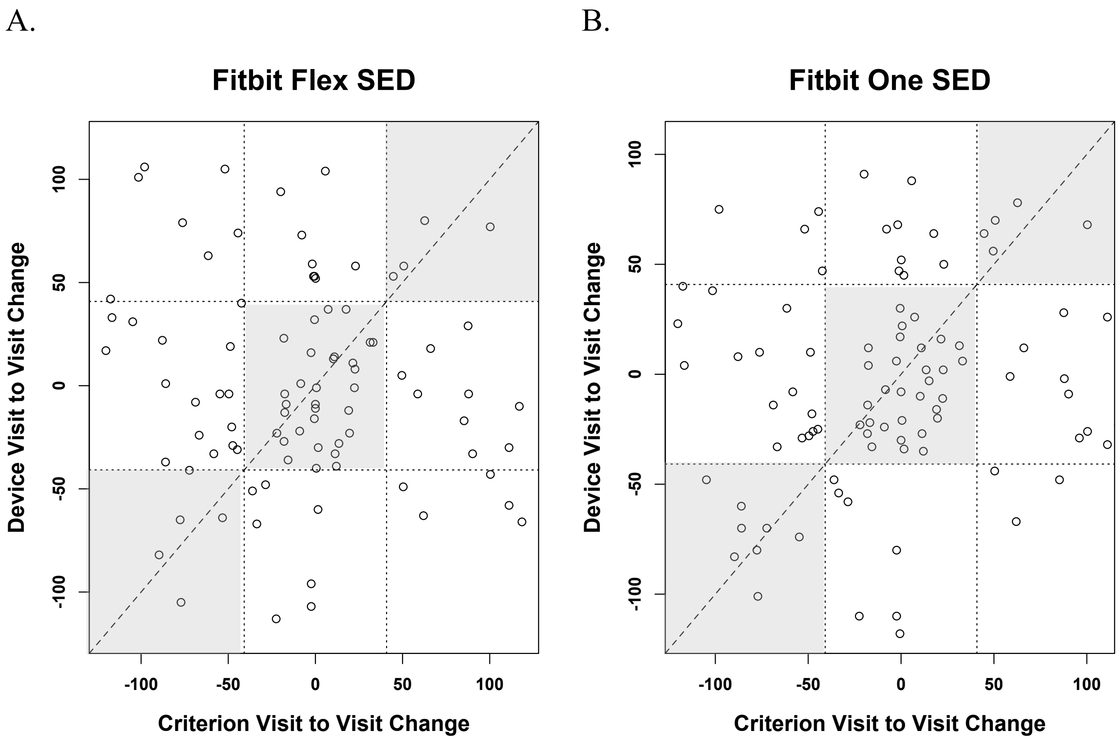
Figure 41. Energy expenditure: criterion measured visit-to-visit change and Fitbit One (A) Fitbit Flex (B) visit-to-visit change
The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.



% Agreement
65.2

% Agreement
79.7

Figure 42. Moderate-to-vigorous physical (MVPA): criterion measured visit-to-visit change and Fitbit One (A) and Fitbit Flex (B) visit-to-visit change
The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.



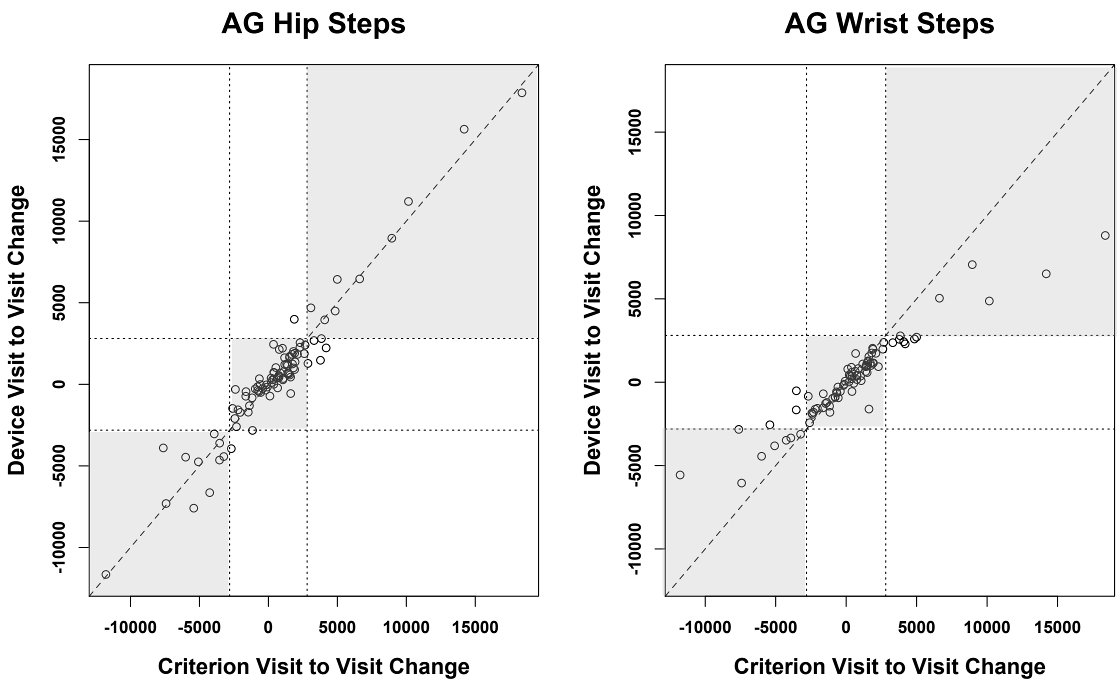
% Agreement
42.3

% Agreement
46.8

Figure 43. Sedentary time: criterion measured visit-to-visit change and Fitbit Flex (A) and Fitbit One (B) visit-to-visit change

The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

A. B.

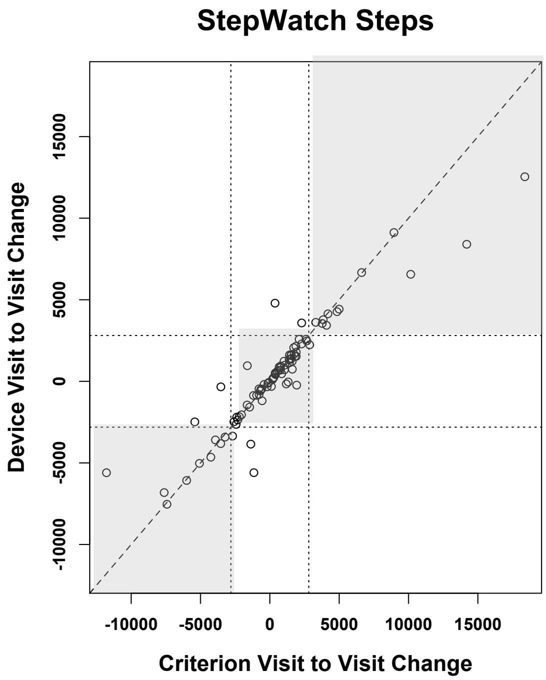


% Agreement
91.4

% Agreement
88.3

Figure 44. Steps: criterion measured visit-to-visit change and ActiGraph hip (A) ActiGraph wrist (B) visit-to-visit change

The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.



% Agreement
91.1

Figure 45. Steps: criterion measured visit-to-visit change and StepWatch visit-to-visit change
The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

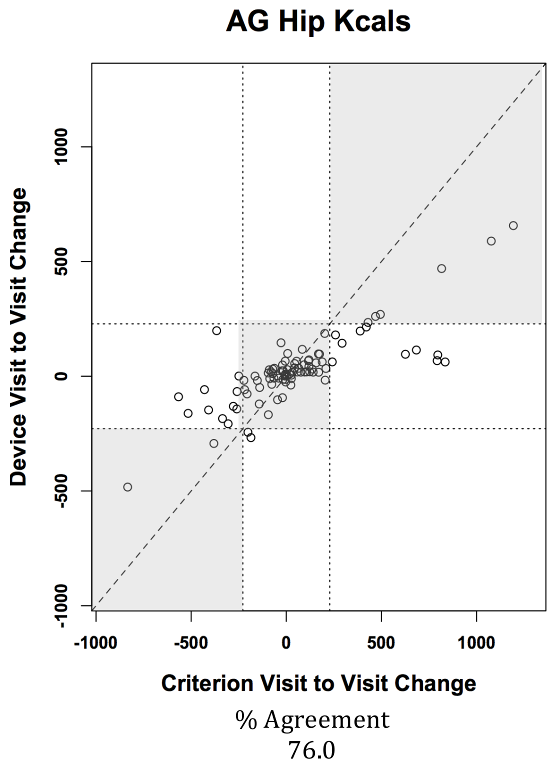
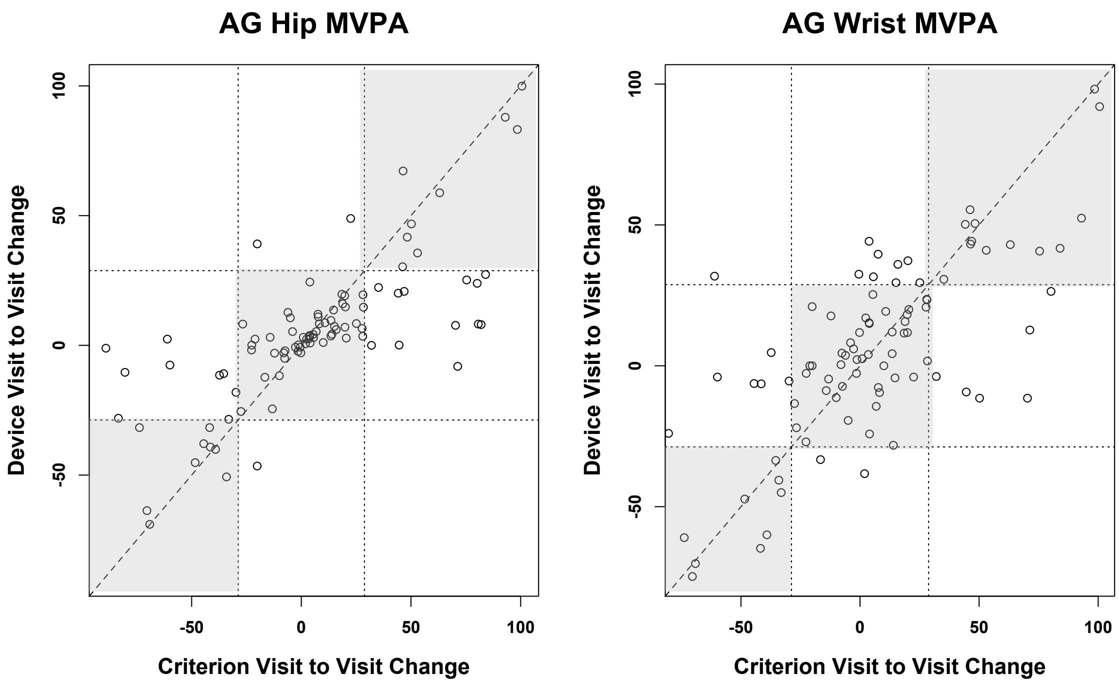


Figure 46. Energy expenditure: criterion measured visit-to-visit change and ActiGraph hip visit-to-visit change

The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

A. B.



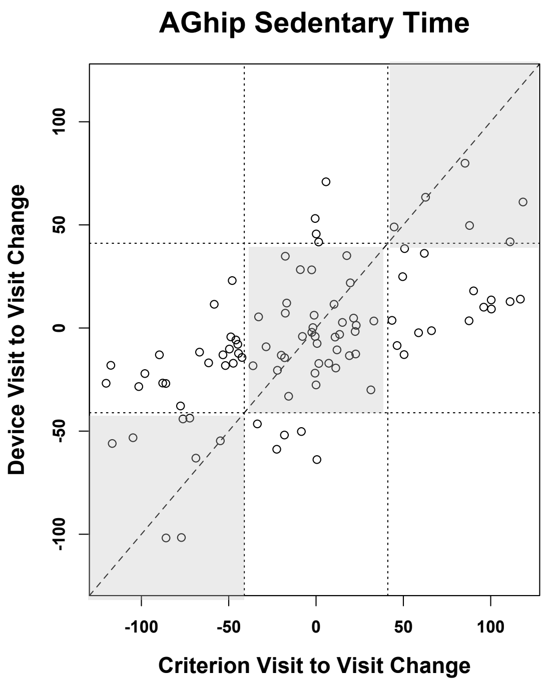
% Agreement 77.0

% Agreement 71.2

Figure 47. Moderate-to-vigorous physical (MVPA): criterion measured visit-to-visit change and ActiGraph hip (A) and ActiGraph wrist (B) visit-to-visit change

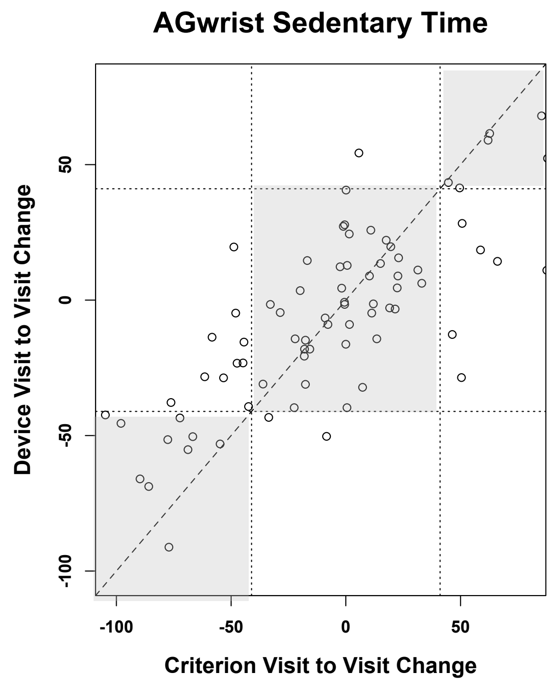
The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

A.



% Agreement
53.1

B.



% Agreement
72.7

Figure 48. Sedentary time: criterion measured visit-to-visit change and ActiGraph hip (A) and ActiGraph wrist (B) visit-to-visit change

The *open circles* are visit-to-visit change, *dotted lines* are the criterion measured within-subject standard deviation, the *dashed line* is the line of identity, and the shaded areas illustrate agreement.

CHAPTER 7
OVERALL SUMMARY AND CONCLUSIONS

The overall goal of this dissertation was to develop a comprehensive understanding of AT estimates of PA and SB using innovative methods to address critical knowledge gaps in the field of PA and health.

Study One

This study was the first to examine AT performance under highly controlled conditions using an electronic orbital shaking protocol. We showed that, on average, the NL produced the smallest error and detected steps similar to our reference (AG) at a 0.9 Hz (corresponding to moderate intensity) and maintained this small error up to a 3.0 Hz (corresponding to very vigorous intensity). Estimates from all other ATs were equivocal, with some overestimating steps or EE, and others underestimating steps or EE compared to the AG. Isolating devices from external influences allowed us to glean valuable insight. There is strong evidence of differences in prediction algorithms by device. Shaking devices across a wide range of frequencies in short increments allowed us to understand how the behavior of the output from the ATs changed across different oscillation frequencies. We were able to also associate oscillation frequencies with intensity levels to provide PA context.

Study Two

In study two, we compared consumer ATs and research-grade activity monitors with DO in free-living settings. Estimates of PA and ST from three research-grade accelerometers and 11 activity monitors during 192 DO-hours were analyzed. The innovation of study two was rooted in the DO criterion measure. We are the first to provide evidence of AT estimates of PA and SB in free-living settings compared to DO. This unique dataset revealed that ATs are accurate with varying precision in estimating PA behaviors in free-living settings. Additionally, ATs and research-grade accelerometers perform similarly (e.g. more accurate in estimating steps and less accurate in estimating MVPA minutes [Table 13]). For all devices, step estimates were accurate and strongly correlated but EE and MVPA estimates were less accurate and more variable but at least moderately correlated. For ATs, estimates of sedentary time were the least accurate and weakly correlated with criterion measures. These findings may stem from the fact that typically, acceleration signals (e.g. vertical accelerations) are used to detect steps, however, ATs use proprietary prediction equations to estimate EE, MVPA and sedentary time. These methods work for some individuals and for others they do not. Implications from this novel study are that consumers and the research community using ATs such as Fitbit, to track steps can be confident in their estimates of PA but less confident in estimating sedentary time. This study advances our understanding of the performance characteristics of ATs in free-living natural settings using a validated DO method to derive PA and SB measures.

Study Three

To date, more than 230 clinical trials have used Fitbit to measure PA behaviors as an outcome and/or exposure,[67] for example, daily step accumulation pre and post PA behavior intervention. Until the current study, no evidence existed of ATs ability to detect change in PA behaviors in free-living settings. Study three was pioneering as it addressed this knowledge gap by examining the ability of ATs to detect change in PA and SB in free-living settings. Our findings suggested that in general, there is similar agreement between the hip-worn FBO and FBF with hip- and- wrist-worn AGs in estimates of change in PA behaviors with criterion measured change. However, change in SB was more difficult to detect for the FBO, FBF and AGhip. Results from study two suggest that the reason for the poorer performance to detect SB change is related to the large bias and inaccuracies of these devices in estimating SB. Results from this innovative study have significant implications regarding the deployment of ATs to estimate PA and SB exposure and outcome measures. We have advanced the field by translating our findings from study two into real-world applications. For example, as an alternative to research-grade accelerometers, researchers may employ FB to measure step accumulation pre- and post-intervention and have confidence in FB step estimates. If the goal of the intervention was to increase steps/2-hrs beyond baseline, average 2-hr step count from the FB should be able to detect that change. Our findings are applicable to activity monitor users worldwide and should be used to disseminate a positive public health message. For example, using activity monitors to promote increasing PA and decreasing ST to produce positive health outcomes.

.

Strengths

Study one was the first to employ electronic orbital shaker testing over a wide range of frequencies to examine AT estimates of steps and EE compared to a widely used research-grade accelerometer. The orbital shaker methods remove the subject to subject variation. As a result, we are confident that observed differences are due to technological features of the devices and are not a function of human variation.

For studies two and three we employed a validated DO system as the criterion measure.[42] An ecologically valid study setting allowed us to examine AT performance while participants wore these devices in their natural environment. Other strengths were the wide range of activities (sleep to trail running), intensities (1.0 to 12.0 METs), activity duration (seconds to hours) and the range of settings and times used for data collection.

In study three, we used the within-subject standard deviation (SD) of the criterion measures to define change, which allowed us to use an evidence-based behavior cut-point of the minimum outcome level to define change.

Limitations

A limitation was that we used the EE estimates from Freedson VM3 equation which was developed via human-testing.[52] Though studies have provided evidence that sensor output is often calibrated during standardized activities such as walking on a treadmill, [154] applying the same algorithm to electronic oscillations may be inappropriate.

In studies two and three the DO procedures used to derive PA and ST measures are not ideal. We employed a validated DO system that uses the Compendium of Physical Activities to apply MET values to activities. The values in the Compendium do not estimate the energy cost of PA in individuals in ways that account for differences in body mass, adiposity, age, sex, efficiency of movement, and environmental conditions in which the activities are performed.[146] Therefore, it is possible that activities were misclassified by intensity category, which may have resulted in inaccuracies of activity minutes, sedentary time and EE. The study trial duration was another limitation for study two. We observed participants for three, 2-hour time frames versus whole-day. Thus, our findings may not be a true representation of whole day behavior. Misclassifications may have impacted study three in at least two ways: (1) the magnitude and direction of visit-to-visit change and, (2) within-subject SD of visit-to-visit change.

Significance and Future Directions

Each study in the present dissertation provides new evidence of wearable monitor estimates of PA and ST. Study one, shows how electronic orbital shaking affects device output. Ultimately, these data provided clear evidence of differences in algorithm by device. The evidence from study two offers a major new contribution to the field of PA measurement. We reported how AT estimates of PA and ST performed under free-living settings. In study three we employed analytic procedures that defines translational research. Our evidence examining the detection of change in PA and ST provides direct meaning and value in using these devices for research, clinical applications and the individual consumer.

Collectively, these studies provide the foundation to building a more comprehensive understanding of the performance characteristics of consumer and research-grade monitors. This is the first evidence detailing how these devices behave in highly controlled and free-living settings. The study designs and data should become the foundation for future work in this field and can be used as evidence for best practices in activity monitor validation studies.

APPENDICES

APPENDIX A

CERTIFICATION OF HUMAN SUBJECTS APPROVAL

University of Massachusetts Amherst
108 Research Administration Bldg.
70 Butterfield Terrace
Amherst, MA 01003-9242

Research Compliance
Human Research Protection Office (HRPO)
Telephone: (413) 545-3428
FAX: (413) 577-1728

Certification of Human Subjects Approval

Date: May 11, 2015
To: Albert Mendoza, Kinesiology
Other Investigator: Patty Freedson, Kinesiology
From: Lynnette Leidy Sievert, Chair, UMASS IRB

Protocol Title:Validation of activity trackers in estimating energy expenditure, activity minutes and steps in free-living settings
Protocol ID: 2015-2492
Review Type:EXPEDITED - NEW
Paragraph ID: 4,6
Approval Date: 05/11/2015
Expiration Date:05/10/2016
OGCA #:115-0883

This study has been reviewed and approved by the University of Massachusetts Amherst IRB, Federal Wide Assurance # 00003909. Approval is granted with the understanding that investigator(s) are responsible for:

Modifications - All changes to the study (e.g. protocol, recruitment materials, consent form, additional key personnel), must be submitted for approval in e-protocol before instituting the changes. New personnel must have completed CITI training.

Consent forms - A copy of the approved, validated, consent form (with the IRB stamp) must be used to consent each subject. Investigators must retain copies of signed consent documents for six (6) years after close of the grant, or three (3) years if unfunded.

Adverse Event Reporting - Adverse events occurring in the course of the protocol must be reported in e-protocol as soon as possible, but no later than five (5) working days.

Continuing Review - Studies that received Full Board or Expedited approval must be reviewed three weeks prior to expiration, or six weeks for Full Board. Renewal Reports are submitted through e-protocol.

Completion Reports - Notify the IRB when your study is complete by submitting a Final Report Form in e-protocol.

Consent form (when applicable) will be stamped and sent in a separate e-mail. Use only IRB approved copies of the consent forms, questionnaires, letters, advertisements etc. in your research.

Please contact the Human Research Protection Office if you have any further questions. Best wishes for a successful project.

APPENDIX B

INFORMED CONSENT DOCUMENT – STUDY TWO & THREE

Consent Form for Participation in a Research Study
University of Massachusetts Amherst

Researcher(s):	Albert Mendoza, M.S., Principal Investigator, Dr. Patty Freedson, Amanda Hickey, M.S.
Study Title:	Validation of Activity Trackers in Estimating Energy Expenditure, Activity Minutes and Steps in Free-living Settings
Funding Agency:	NIH: National Heart, Lung, and Blood Institute

1. WHAT IS THIS FORM?

This form is called the Informed Consent. It will give you information about the study so you can make an informed decision about participation in this research. This information will outline everything you will need to do to participate and any known risks, discomfort, or inconveniences that may occur during your participation in this research. Feel free to ask questions at any time. If you decide to participate in this study, you will be asked to sign this form and initial each page. You will be given a copy for your records.

2. WHO IS ELIGIBLE TO PARTICIPATE?

To participate in this study:
(1) You must be between 18 and 59 years of age, and women must not be pregnant.
(2) You must be in good physical health (no diagnosed cardiovascular, metabolic, joint, or chronic diseases).
(3) You must be able to do normal daily activities and are not limited by musculoskeletal problems that would impair your ability to be normally active.
(4) You must be willing to comply with the study protocol described below.

3. WHAT IS THE PURPOSE OF THIS STUDY?

The purpose of this study is to determine the accuracy and precision of activity trackers in estimating energy expenditure, activity minutes and steps in free-living settings. Researchers will also examine whether activity trackers can detect changes in your activity behavior (e.g. classify you as inactive or active) in free-living settings.

4. WHERE WILL THE STUDY TAKE PLACE AND HOW LONG WILL IT LAST?

The study will consist of four visits.
Visit One: Informed consent visit will take place in the Physical Activity and Health Lab on the UMass Amherst campus (~30 min).
Visits 2, 3 and 4: Observations sessions: Each of the observation visits will last 2-hrs and 15 min: 10 min to put on monitors, 2-hrs of observation while you perform normal daily activities, 5 min to remove monitors.
Your total time commitment while participating in this study will be approximately 7.25 hours over a 3 week time period. After the fourth visit your participation in the study will be complete.

Initials: ______

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 04/12/2016 **Protocol #:** 2015-2492
Valid Through: 05/10/2017
IRB Signature:

5. WHAT WILL I BE ASKED TO DO?
If you agree to participate in this study during the first visit you will be asked to initial the bottom of each page of this informed consent and sign the last page. During the first visit you will fill out a Physical Activity Status (PAS) Questionnaire and a Physical Activity Readiness Questionnaire (PAR-Q), which will ask questions about how active you are and if you have any injuries or health impairments that prevent you from physical activity. Based on your answers to the questionnaire you may not qualify for this study. Then you will also fill out a School of Public Health Compensation Form and a W9 form. Researchers will record your height and weight and demographic information like your date of birth and ethnicity (**Visit One).** You may skip any question you feel uncomfortable answering. You will then be scheduled for 3, 2-hr and 15 min testing sessions including one weekend testing session **(Visits 2, 3 and 4).** The first visit will be approximately 30 minutes.

Visits 2, 3 and 4. Each of the 2-hr 15 min observation sessions (scheduled at the end of Visit 1) will be done at different times of the day (e.g. Session one: morning; Session 2: afternoon; Session 3: evening), in your free-living settings (e.g. home, work, driving). If/when you drive, researchers will follow from a safe distance in a separate car. Two researchers (or 1 researcher and 1 research assistant) will always be present during the observation sessions, and for female participants, at least one of the 2 researchers will be female. For these three visits, you will come to the Physical Activity and Health Laboratory to be fitted with a variety of activity monitors that will be worn on the upper arm, wrists, hip and ankle, and a smart shirt. You will wear 1 monitor on the upper arm, 8 monitors on the wrists (4 on the right wrist and 4 on the left wrist), 2 monitors on the right hip, and 1 smart shirt to be worn as an undergarment. The smart shirt estimates energy expenditure and steps, as well as respiration (how much you breathe) and heart-rate. You will then leave the lab with two researchers who will stay with you for the 2-hr session. The researchers will be video recording you for each of the entire 2-hr sessions while you carry out your normal activities (including driving). Every attempt will be made to avoid including your head in these video recordings. If your head does appear in the video we will edit these shots to blur or eliminate your head from the video recording. If private time is required (i.e. going to the bathroom), we will not observe you during these private time periods. At the end of the 2-hr recording period, the researchers will remove the activity monitors, you will remove the smart shirt in private wherever you feel most comfortable (e.g. restroom or secured room) and your testing session is complete. Your data from these monitors will be downloaded to computer. Visits 2, 3 and 4 will each take approximately 2 hours and 15 minutes.

6. WHAT ARE MY BENEFITS OF BEING IN THIS STUDY?
You may not directly benefit from this research; however we hope that your participation in the study may provide valuable information about the accuracy and precision of activity monitors in free-living settings and will provide evidence about how the consumer monitors compare to one another (using direct observation/video recording as the criterion [truth] measure) for estimating energy expenditure, activity minutes, steps and sedentary time. Once all data are collected and analyzed, we will provide you with your results that will describe the number of minutes you were active and sedentary during the observation periods and your estimated activity energy expenditure and steps during the observation periods.

7. WHAT ARE MY RISKS OF BEING IN THIS STUDY?
All possible attempts will be made to minimize any risks. The risks are minimal and are simply risks that occur carrying out your normal daily activities. We will not ask you to do anything out

of the ordinary pattern of what you would typically do. You may be uncomfortable with the video recording but we will make every attempt to record your body movements without your head being recorded. In the event your head is recorded we will use our software to blur or edit out your face from all video recordings when the recordings are downloaded into the computer. You may notice that you are wearing several devices on your upper arm, wrists, hip, and ankle, and a smart shirt. There is a very minimal risk that a device or devices you are wearing become uncomfortable or cause you discomfort. You are free to remove any device and/or the smart shirt if you feel that causes you a problem during the observation sessions.

8. HOW WILL MY PERSONAL INFORMATION BE PROTECTED?
The information obtained in this study will be regarded as privileged and confidential. If the results of this study are published in a scientific journal or presented at a scientific meeting, your name will not be used. All records, including questionnaire data, activity monitor data, and video data will be identified only with a numerical ID. Activity monitor data will be stored on a password protected PC and a password protected portable hard drive (portable hard drive will be stored in a locked file cabinet). Video data will be downloaded into a PC and portable hard drive (hard drive will be stored in locked file cabinet). The video data will then be deleted from the camera. All efforts will be made to not capture your head in the video recordings. In the event that head data are contained in the video file, we will either blur the face images or edit these images out of the video recording. After we code the data from the PC files, we will delete this record and will only keep the video record stored on the hard drive in the event we have to go back to review and verify coding.

9. WILL I RECEIVE ANY PAYMENT FOR TAKING PART IN THE STUDY?
Payment will be sent as a check by mail to the address you provide in 6-8 wks. You will receive partial payment if you decide to leave the study at any point. For each 2-hr observation period completed, you will receive $25.00 (maximum will be $75.00 for completion of all 3, 2-hr observation sessions). If you complete at least 1-hr of any observation period, you will receive $12.50. After completing all visits you will be paid $75.00 total.

10. WHAT IF I HAVE QUESTIONS?
You are encouraged to ask any questions, voice any concerns or doubts regarding the study at any time. Investigators will attempt to answer all questions to the best of their ability. The investigators fully intend to conduct the study with your best interest, safety, and comfort in mind. Mr. Mendoza can be reached at 413.545.1583 or by cell at 415.297.9327, Professor Freedson can be reached at 413.545.2620 and Ms. Hickey may be reached at 413.545.1583. If you would like to discuss your rights as a participant in a research study or wish to speak with someone not directly involved with the study you may contact the Human Research Protection Office at humansubjects@ora.umass.edu.

11. CAN I STOP BEING IN THE STUDY?
Participation in this study is completely voluntary. You may withdraw consent at any time in writing or by telephone (413.545.1583) and discontinue participation in the study without prejudice to you or your medical care at UMass Amherst.

Initials: _____

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 04/12/2016 Protocol #: 2015-2492
Valid Through: 05/10/2017
IRB Signature: Nancy C. Swett

12.WHAT IF I AM INJURED?
In the unlikely event of an injury resulting directly from participation in this study, investigators will assist you in every way to insure that you receive proper medical attention. The University of Massachusetts does not have a program to compensate subjects for injury or complications related to human subjects research but the study personnel will assist you in getting treatment. It also should be understood that by your agreement to participate in this study, you are not waiving any of your legal rights.

13. SUBJECT STATEMENT OF VOLUNTARY CONSENT
I confirm that this document has explained the purpose of the research, the study procedures that I will undergo and the possible risks and discomforts as well as the benefits I may experience. I have read and I understand the consent form. Therefore, I agree to participate in this study.

Recall, that the video will not include your face.

> **______I agree that segments of the recordings made of my participation in this research may be used for conference presentations, as well as education and training of future researchers/practitioners.**

> **______I agree to have my recordings archived for future research in the field of Kinesiology.**

> **______I do not agree to have my recordings archived for future research in the field of Kinesiology.**

> **______I do not agree to allow segments of recordings of my participation in this research to be used for conference presentations or education and training purposes.**

______________________	___________________	_________
Participant Signature:	Print Name:	Date:

By signing below I indicate that the participant has read and, to the best of my knowledge, understands the details contained in this document and has been given a copy.

______________________	___________________	_________
Signature of Person Obtaining Consent	Print Name:	Date:

The investigators will retain the original copy of this document for their records. You will be given a copy of the document if you would like one.

APPENDIX C

PHYSICAL ACTIVITY READINESS QUESTIONNAIRE

PHYSICAL ACTIVITY READINESS QUESTIONNAIRE (PAR-Q)

Please read the following questions carefully and answer each one honestly: check YES or NO.

YES	NO	
☐	☐	1. Has your doctor ever said that you have a heart condition <u>and</u> that you should only do physical activity recommended by a doctor?
☐	☐	2. Do you feel pain in your chest when you do physical activity?
☐	☐	3. In the past month, have you had chest pain when you were not doing physical activity?
☐	☐	4. Do you lose your balance because of dizziness or do you ever lose consciousness?
☐	☐	5. Do you have a bone or joint problem that could be made worse by a change in your physical activity?
☐	☐	6. Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?
☐	☐	7. Do you know of <u>any other reason</u> why you should not do physical activity?

PAR-Q (Thomas, Reading, & Shephard, 1992)

APPENDIX D

PHYSICAL ACTIVITY STATUS QUESTIONNAIRE

Participant ID: _________ **Date:** _____________

Physical Activity Status

Using the descriptions below, record the highest number (0 to 7) which best describes your general activity level during the **previous month.** If you did more than section 1, then move on to section 2, and so on. You want to pick the highest number in this list to represent your activity level.

Section 1: Did not participate regularly in programmed recreational sport or heavy physical activity.

0 Avoided walking or exertion, e.g. always used the elevator, drove whenever possible instead of walking.

1 Walked for pleasure, routinely used the stairs, occasionally exercised sufficiently to cause heavy breathing or perspiration.

Section 2: Participated regularly in recreation or work requiring modest physical activity, such as golf, horseback riding, calisthenics, gymnastics, table tennis, bowling, weight lifting, yard work.

2 10 to 60 minutes per week.

3 Over 1 hour per week.

Section 3: Participated regularly in heavy physical exercise such as running or jogging, swimming, cycling, rowing, skipping rope, running in place or engaged in vigorous aerobic activity type of exercise such as tennis, basketball, or handball.

4 Ran less than 1 mile per week or spent less than 30 minutes per week in comparable physical activity.

5 Ran 1 to 5 miles per week or spent 30 to 60 minutes per week in comparable physical activity.

6 Ran 5 to 10 miles per week or spent 1 to 3 hours per week in comparable physical activity.

7 Ran more than 10 miles per week or spent over 3 hours per week in comparable physical activity.

Physical Activity Status during the previous month (highest score): ______

APPENDIX E

FEATURES OF CONSUMER-BASED ACTIVITY TRACKERS

Device	Apple iWatch Sport	Fitbit Flex	Fitbit One	Garmin Vívofit	New Lifestyles NL-1000	Microsoft Band	Misfit Flash	Misfit Shine	Polar loop	Withings Pulse
Cost	$350.99	$79.95	$99.95	$99.99	$54.95	$199.99	$29.99	$69.99	$109.95	$119.95
Wear location	Wrist	Wrist	Clip on (multiple locations)	Wrist	Hip	Wrist	Clip on (multiple locations)	Clip on (multiple locations)	Wrist	Clip on and wrist band
Tracks Calories Burned	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Active Time	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Steps	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Distance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tracks Elevation/Stairs	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓
Tracks Sleep	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Tracks Heart Rate	✓	✗	✗	✗	✗	✓	✗	✗	✓	✓
Battery or Chargeable	Chargeable (every 18 hours)	Chargeable (every 5 days)	Chargeable (every 10+ days)	Battery (every 1+ years)	Battery (up to 18 months)	Chargeable (every 48 hours)	Battery (lasts up to 6 months)	Battery (lasts up to 6 months)	Chargeable (up to 6 days)	Chargeable (every 2 days)
Uploading Data	Bluetooth	Bluetooth	Bluetooth	Bluetooth	Real-time data	USB			USB	Bluetooth
Tracker Display	Real-time data	LED progress indicator	Real-time data	Real-time data	Real-time data	Real-time data	LED progress indicator	LED progress indicator	Real-time data	Real-time data

LED, Light-Emitting Diode; USB, Universal Serial Bus

APPENDIX F

STUDY INFORMATION SHEET

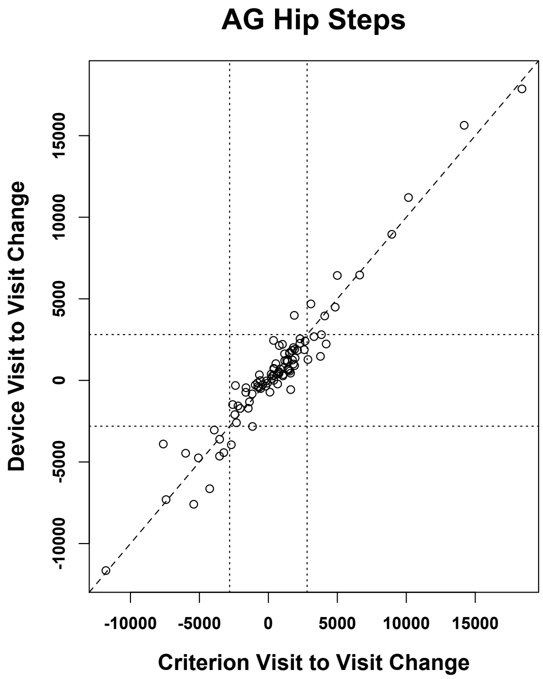
ACTIVITY TRACKER

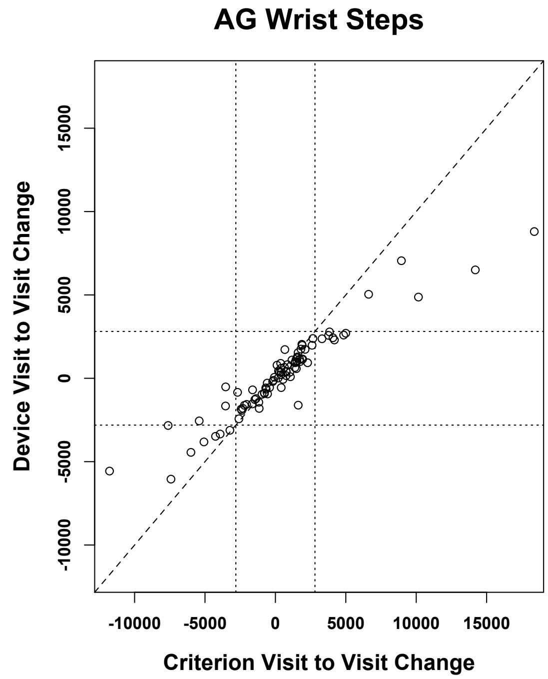
VALIDATION STUDY

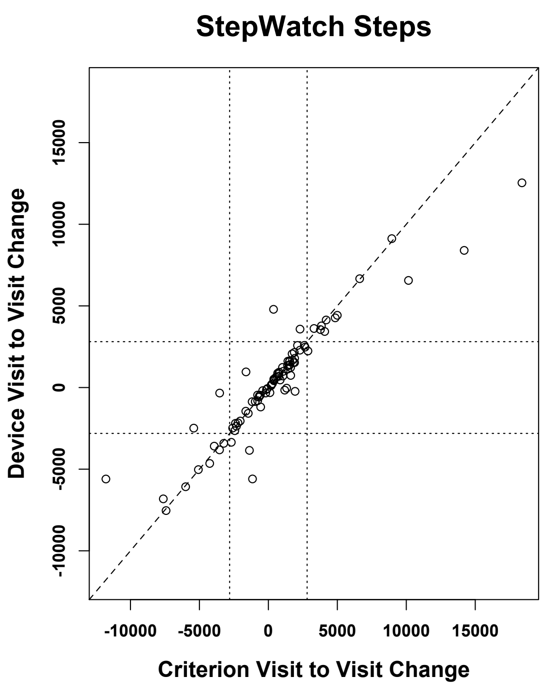
- The Physical Activity and Health Lab is conducting a study to test the accuracy of consumer activity trackers in estimating how active people are
- Participants are directly observed (recorded with a GoPro) while engaging in their daily activities while wearing several activity trackers
- We are sensitive to participants privacy and those persons of the surrounding environment
- To ensure privacy preservation:
  - Sound is not recorded
  - Identities (faces) of all individuals in the video will be blurred
  - Individuals will not be identifiable
- This study protocol has been approved by the UMass Amherst Human Subjects Board (IRB)
- If further clarification is needed please contact:
  - Dr. Patty Freedson at 413-545-2620 or psf@kin.umass.edu
  - Albert Mendoza at 413-545-1583 or amendoza@kin.umass.edu

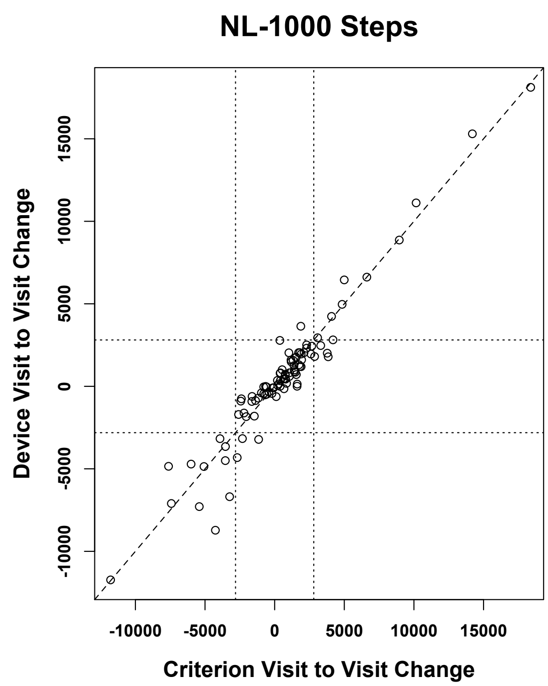
APPENDIX G

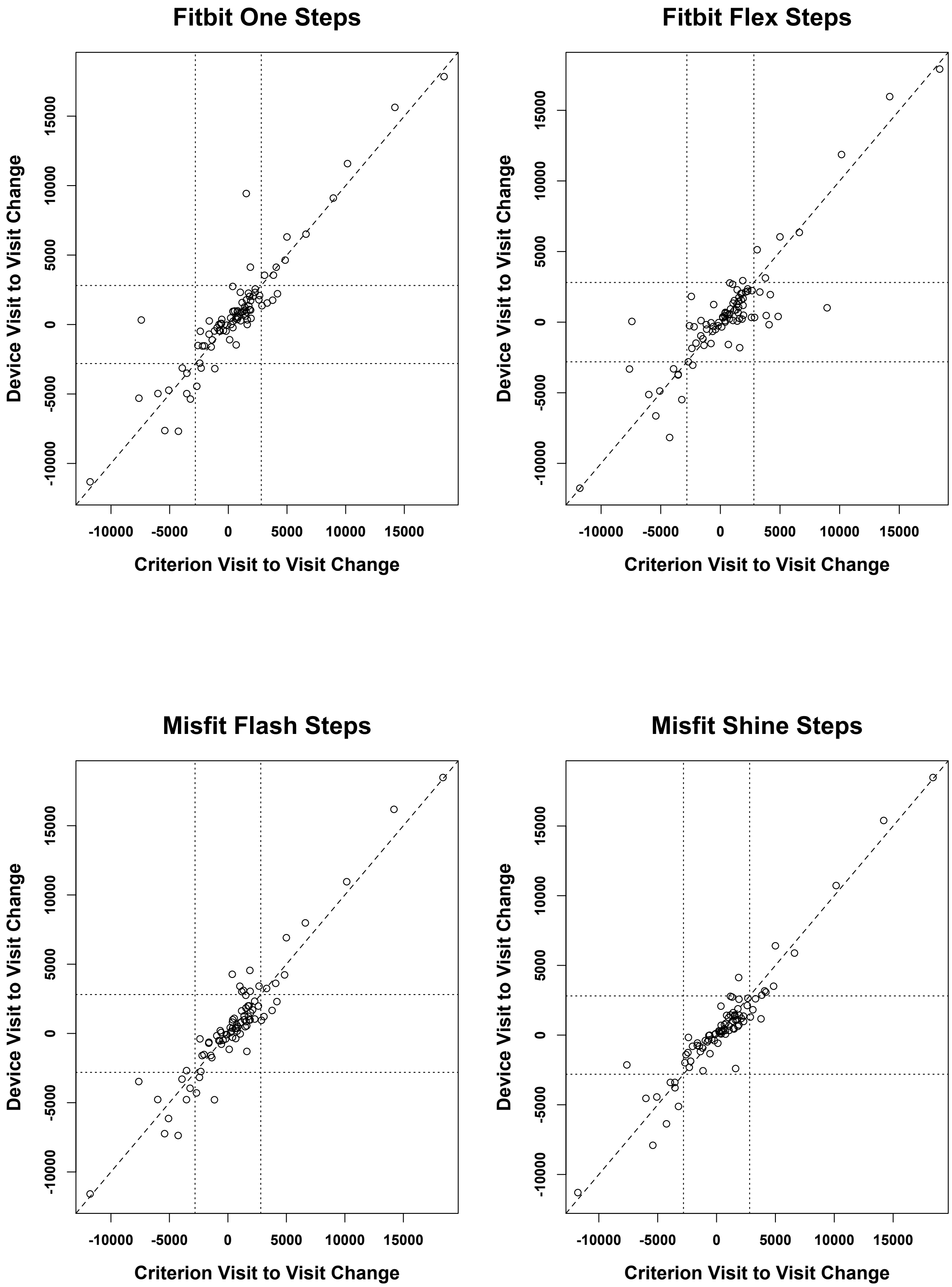
STEPS: CRITERION MEASURED VISIT-TO-VISIT CHANGE WITH DEVICE ESTIMATED VISIT-TO-VISIT CHANGE


AG Hip Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000 -5000 0 5000 10000 15000

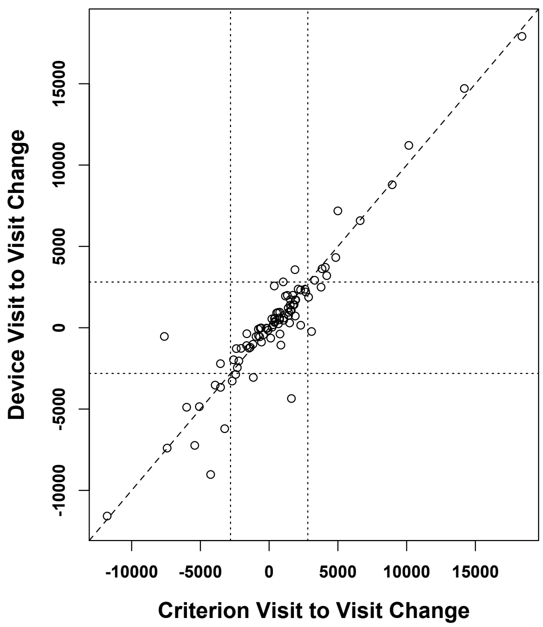

AG Wrist Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000 -5000 0 5000 10000 15000


StepWatch Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000 -5000 0 5000 10000 15000

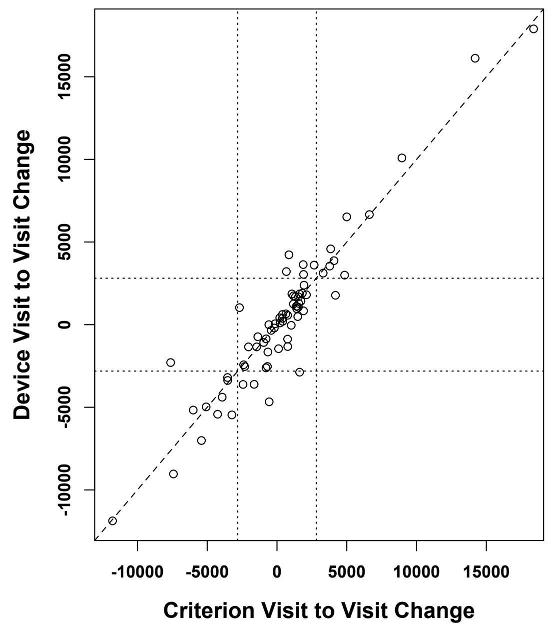

NL-1000 Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000 -5000 0 5000 10000 15000


Fitbit One Steps
Fitbit Flex Steps
Misfit Flash Steps
Misfit Shine Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000
-5000
0
5000
10000
15000

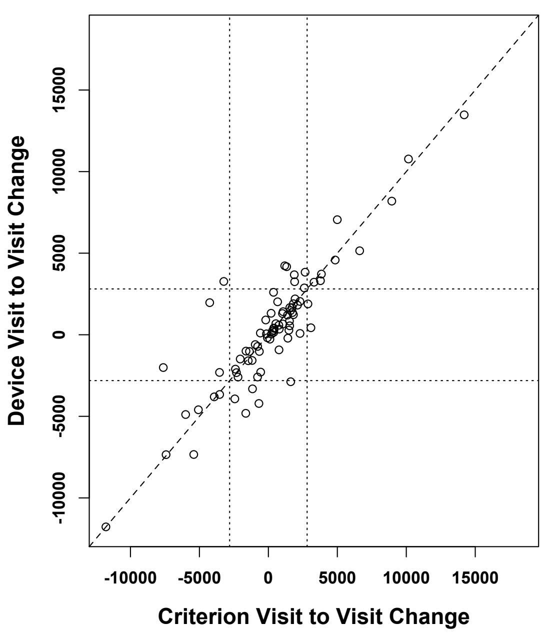
Garmin Vivofit Steps



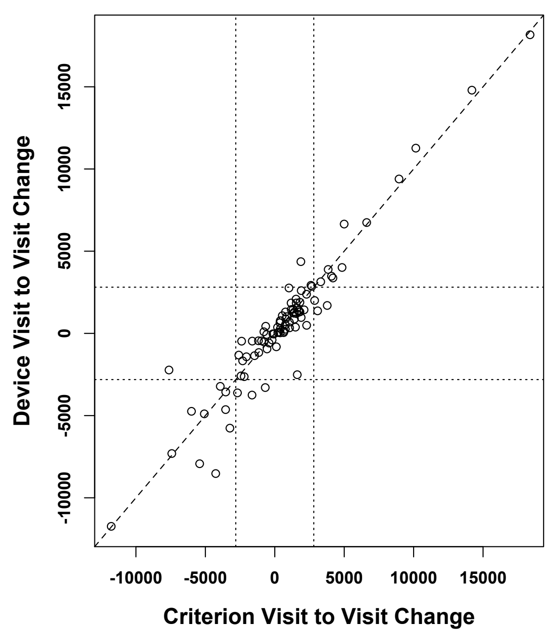
Polar Loop Steps

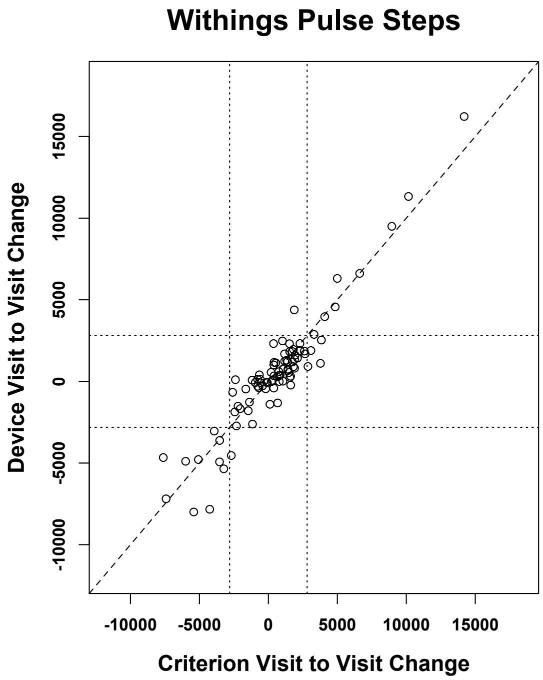


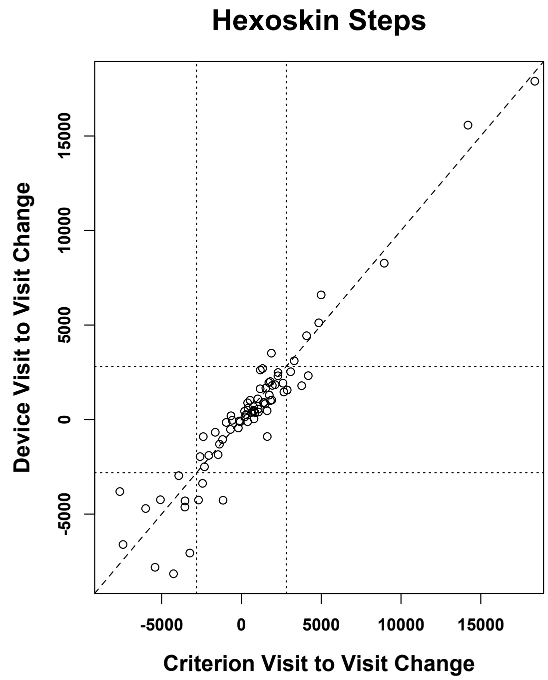
Apple iWatch Steps



Microsoft Band Steps

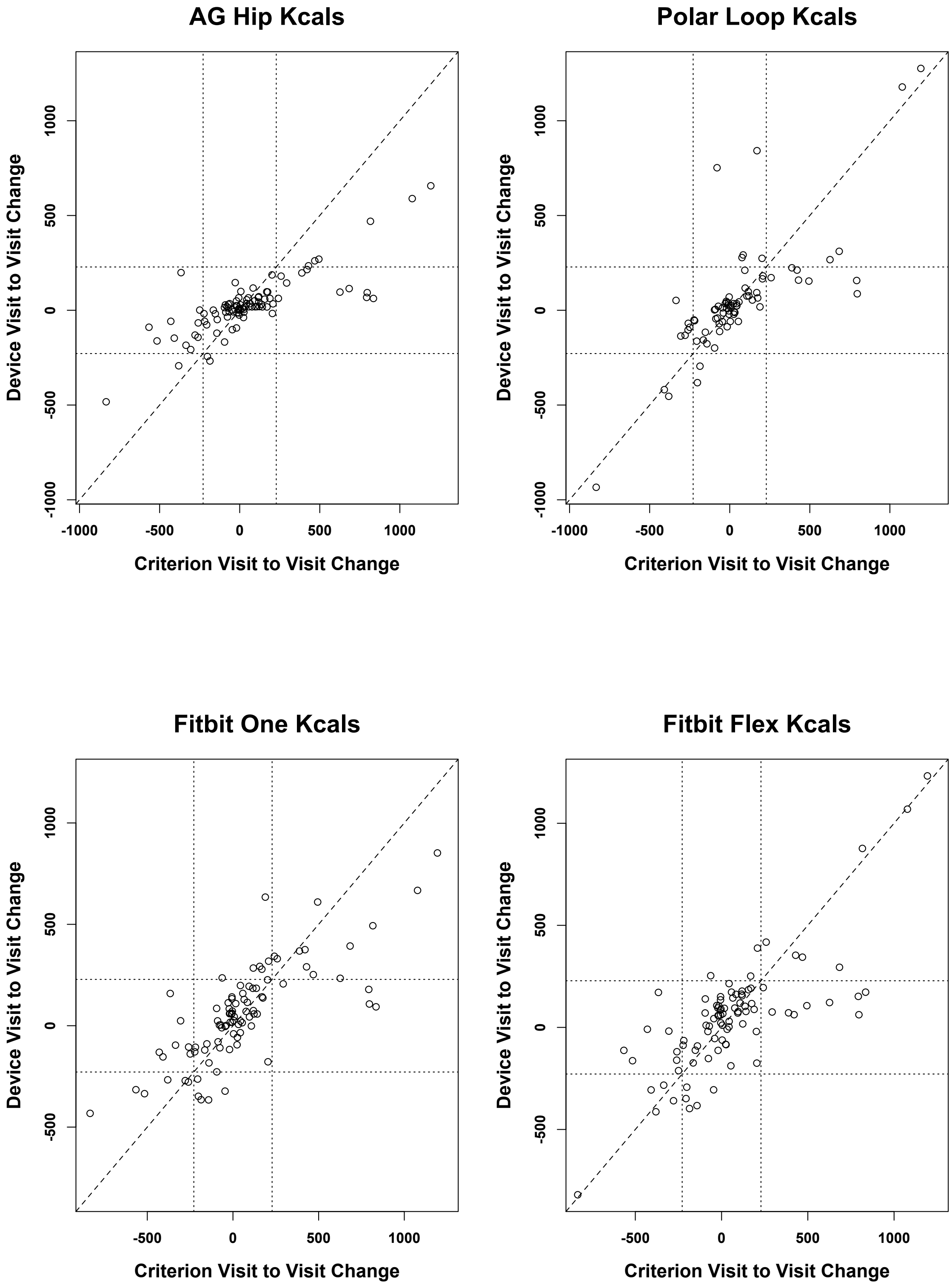


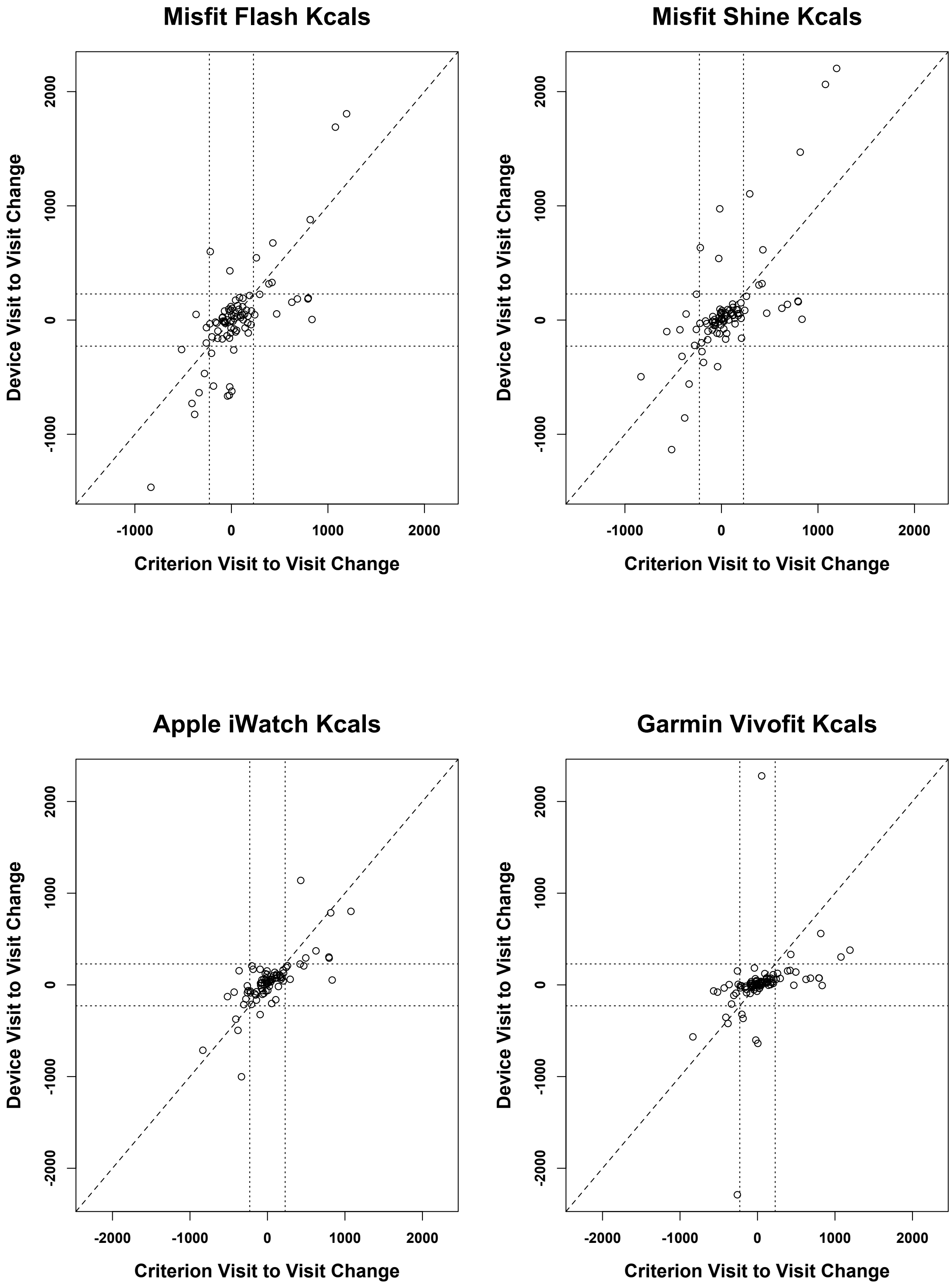

Withings Pulse Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-10000
-5000
0
5000
10000
15000

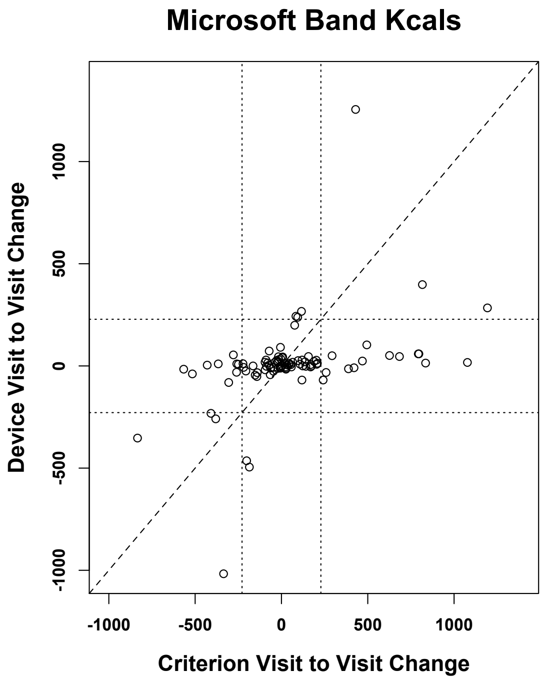

Hexoskin Steps
Device Visit to Visit Change
Criterion Visit to Visit Change
-5000
0
5000
10000
15000

APPENDIX H

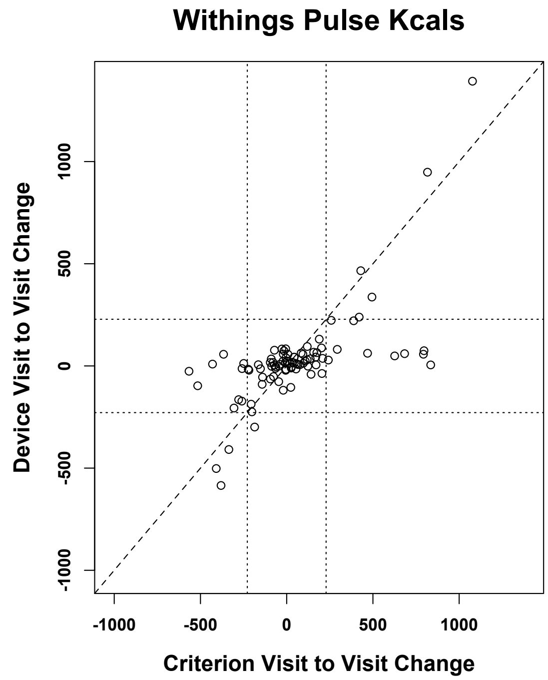
ENERGY EXPENDITURE: CRITERION MEASURED VISIT-TO-VISIT CHANGE WITH DEVICE ESTIMATED VISIT-TO-VISIT CHANGE


AG Hip Kcals
Polar Loop Kcals
Fitbit One Kcals
Fitbit Flex Kcals
Device Visit to Visit Change
Criterion Visit to Visit Change
-1000
-500
0
500
1000

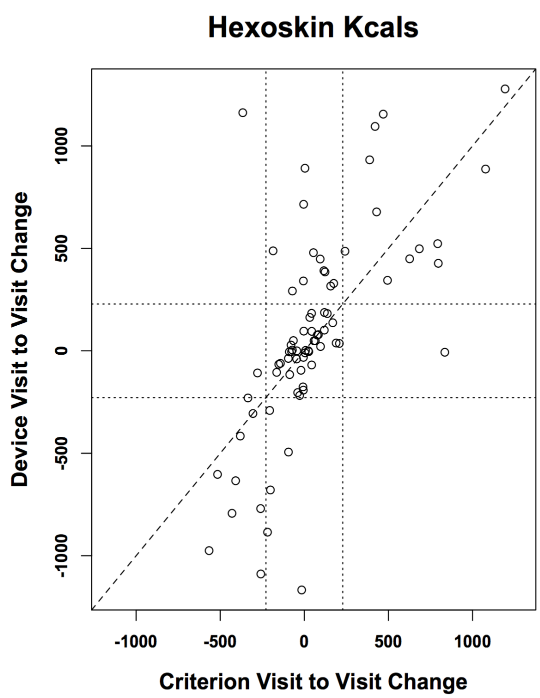

Misfit Flash Kcals
Misfit Shine Kcals
Apple iWatch Kcals
Garmin Vivofit Kcals
Device Visit to Visit Change
Criterion Visit to Visit Change
-2000
-1000
0
1000
2000



Microsoft Band Kcals
Device Visit to Visit Change
Criterion Visit to Visit Change
-1000
-500
0
500
1000



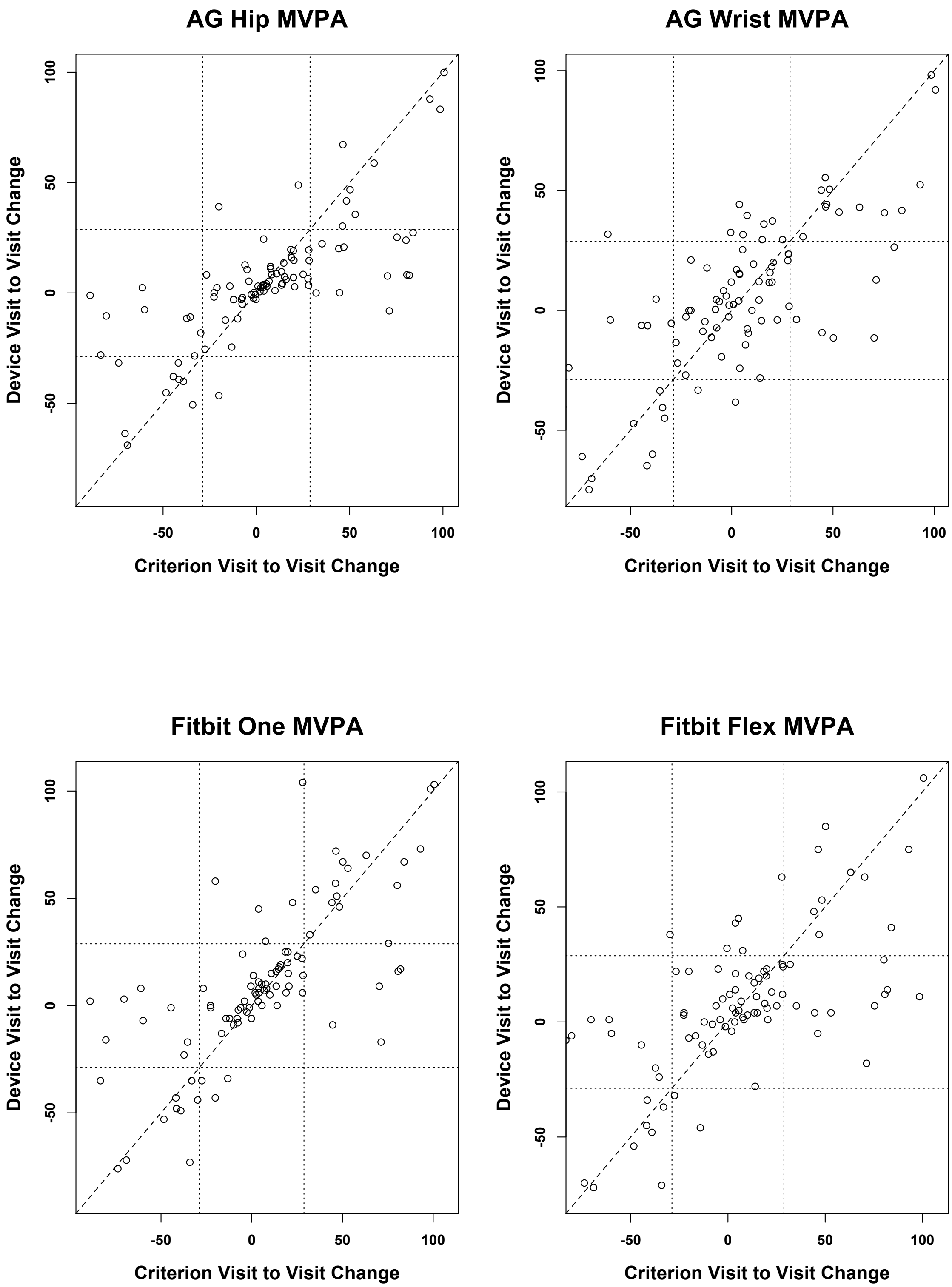
Withings Pulse Kcals
Device Visit to Visit Change
Criterion Visit to Visit Change
-1000
-500
0
500
1000

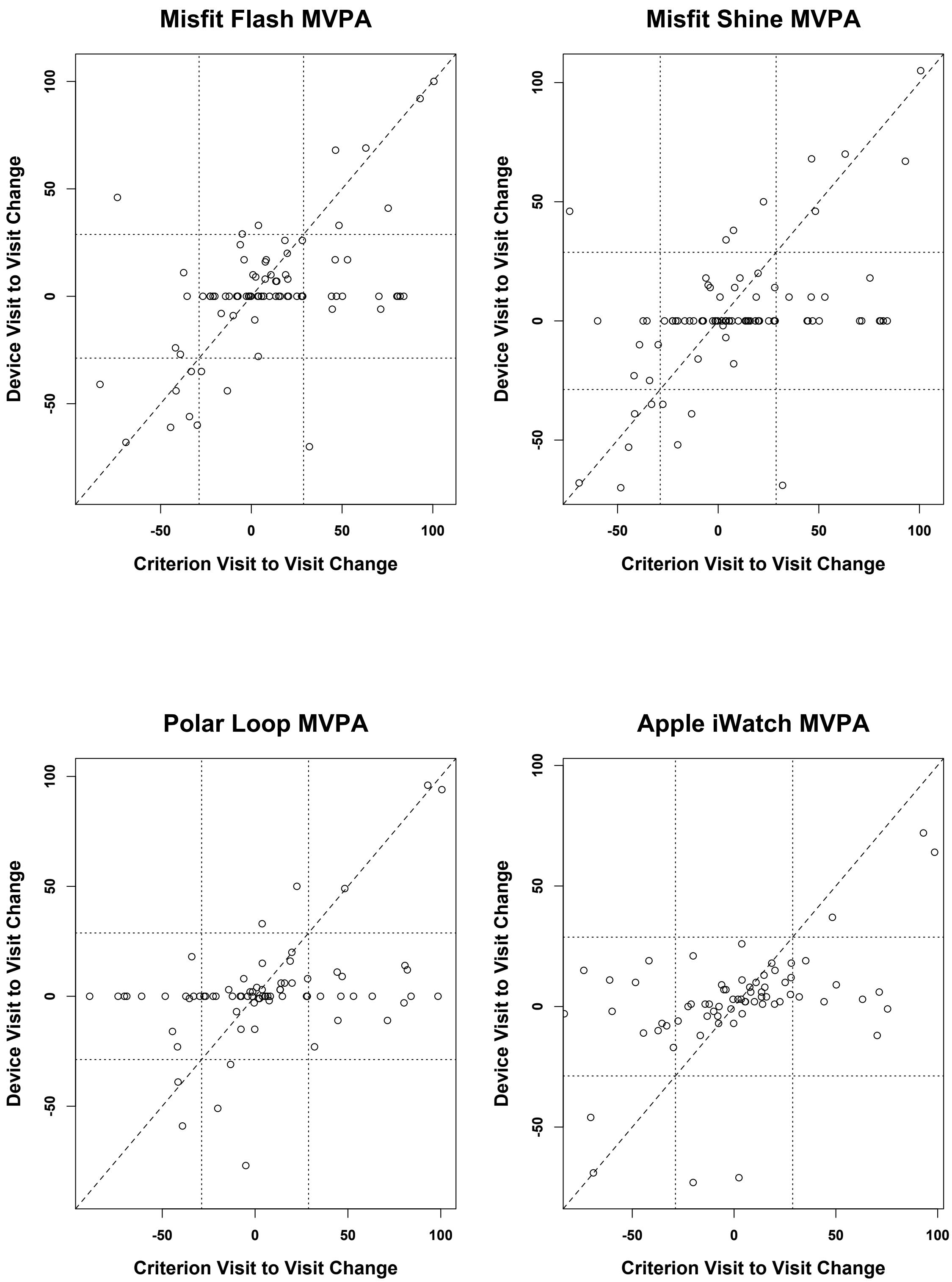


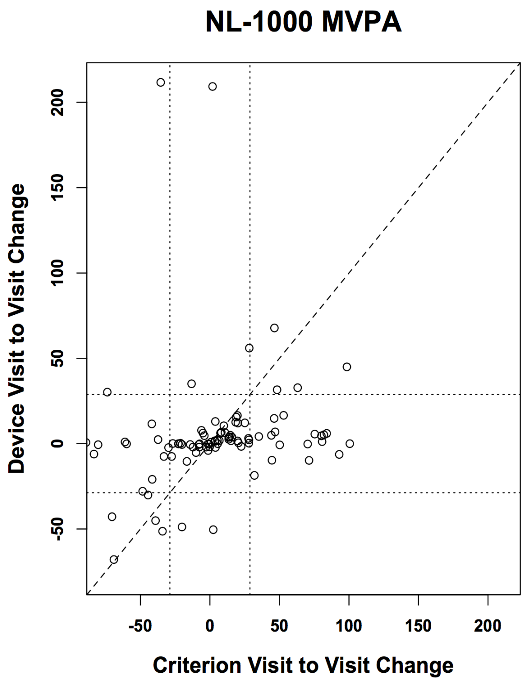
Hexoskin Kcals
Device Visit to Visit Change
Criterion Visit to Visit Change
-1000
-500
0
500
1000

APPENDIX I

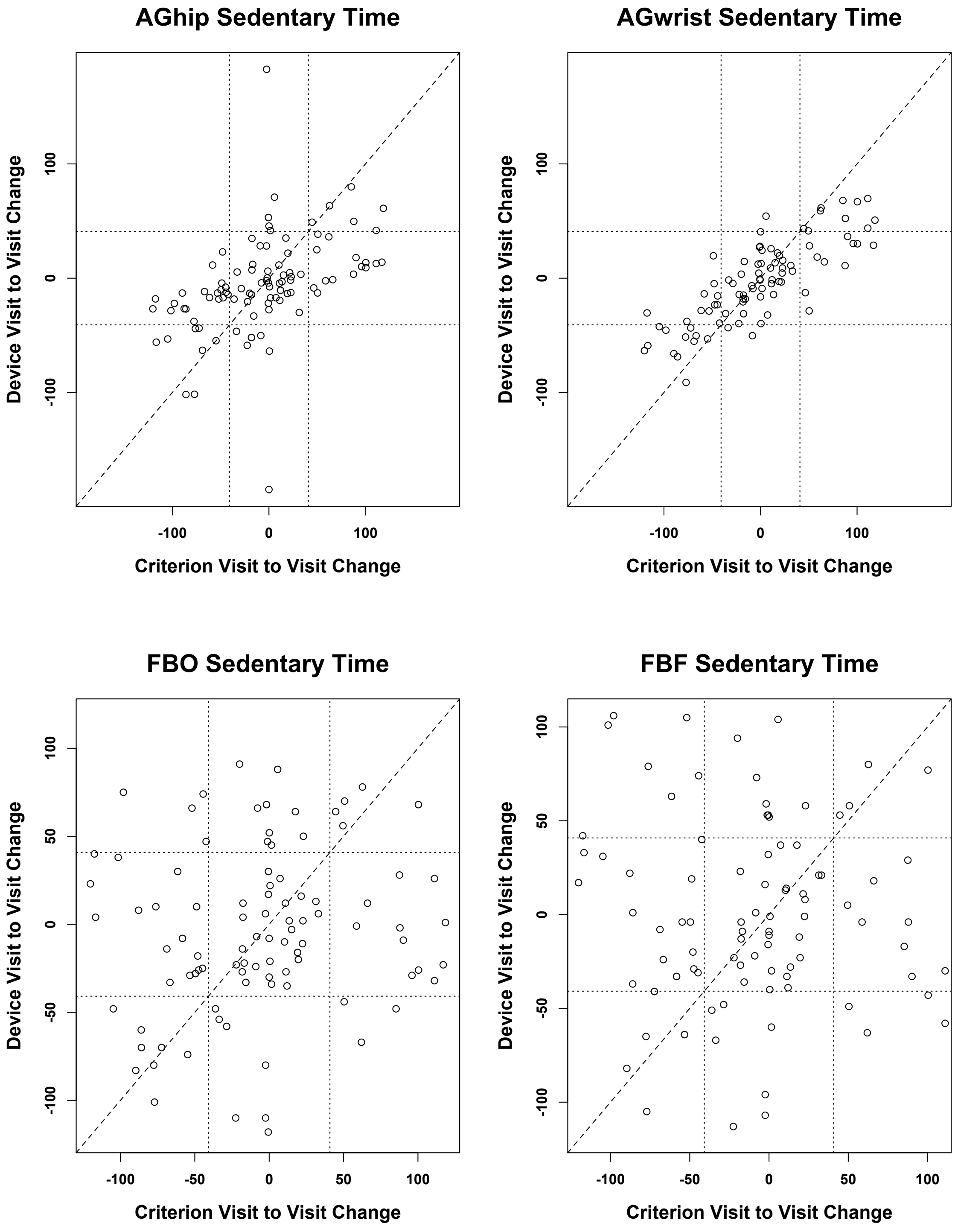
MODERATE-TO-VIGOROUS PHYSICAL ACTIVITY (MVPA): CRITERION MEASURED VISIT-TO-VISIT CHANGE WITH DEVICE ESTIMATED VISIT-TO-VISIT CHANGE


AG Hip MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100
AG Wrist MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100
Fitbit One MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100
Fitbit Flex MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100


Misfit Flash MVPA
Misfit Shine MVPA
Polar Loop MVPA
Apple iWatch MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100


NL-1000 MVPA
Device Visit to Visit Change
Criterion Visit to Visit Change
-50
0
50
100
150
200

APPENDIX J SEDENTARY MINUTES: CRITERION MEASURED VISIT-TO-VISIT CHANGE WITH DEVICE ESTIMATED VISIT-TO-VISIT CHANGE


AGhip Sedentary Time
AGwrist Sedentary Time
FBO Sedentary Time
FBF Sedentary Time
Device Visit to Visit Change
Criterion Visit to Visit Change
-100
0
100
-50
50

REFERENCES

1. Services UDoHaH. 2008 Physical Activity Guidelines for Americans. 2008; http://www.health.gov/paguidelines/. Accessed August 2, 2015.

2. United States. Department of Health aHS. *Physical activity and health: a report of the Surgeon General.* DIANE Publishing; 1996.

3. Tudor-Locke C, Bassett DR, Jr. How many steps/day are enough? Preliminary pedometer indices for public health. *Sports Med.* 2004;34(1):1-8.

4. Australian Government DoH. Australia's Physical Activity & Sedentary Behaviour Guidelines for Adults (18-64 years). In: Australian Government DoH, ed2014:1-8.

5. Hayward J, Chansin G, Zervos H. Wearable Technology 2017-2027: Markets, Players, Forecasts. 2017; http://www.idtechex.com/research/reports/wearable-technology-2017-2027-markets-players-forecasts-000536.asp. Accessed July 1, 2017.

6. Year-Over-Year Wearables Spending Doubles, According to NPD [press release]. Port Washington, NY: The NPD Group, Inc, February 1 2016.

7. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *Journal of Applied Physiology.* 2011;111(6):1804-1812.

8. John D, Liu S, Sasaki JE, et al. Calibrating a novel multi-sensor physical activity measurement system. *Physiological measurement.* 2011;32(9):1473-1489.

9. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of applied physiology.* 2009;107(4):1300-1307.

10. Albinali F, Intille SS, Haskell W, Rosenberger M. Using wearable activity type detection to improve physical activity energy expenditure estimation. *Ubicomp 2010: Proceedings of the 2010 ACM Conference on Ubiquitous Computing.* 2010:311-320.

11. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. *Medicine and science in sports and exercise.* 2011;43(1):101-107.

12. Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society.* 2008;12(1):20-26.

13. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *Journal of applied physiology.* 2007;103(4):1419-1427.

14. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. *Medicine and science in sports and exercise.* 2012;44(4):742-748.

15. Chen KY, Acra SA, Majchrzak K, et al. Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes technology & therapeutics.* 2003;5(6):1023-1033.

16. Bankoski A, Harris TB, McClain JJ, et al. Sedentary activity associated with metabolic syndrome independent of physical activity. *Diabetes care.* 2011;34(2):497-503.

17. Mammen G, Gardiner S, Senthinathan A, McClemont L, Stone M, G F. Is this bit fit? Measuring the quality of the FitBit step-counter. *Health & Fitness Journal of Canada.* 2012;5(4).

18. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. *Medicine and science in sports and exercise.* 2016;48(8):1619-1628.

19. 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. *Medical engineering & physics.* 2014;36(6):659-669.

20. Chen MD, Kuo CC, Pellegrini CA, Hsu MJ. Accuracy of wristband activity monitors during ambulation and activities. *Medicine and science in sports and exercise.* 2016.

21. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation.* 2015;7:24.

22. Storm FA, Heller BW, Mazza C. Step detection and activity recognition accuracy of seven physical activity monitors. *PloS one.* 2015;10(3):e0118723.

23. Takacs J, Pollock CL, Guenther JR, Bahar M, Napier C, Hunt MA. Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of science and medicine in sport / Sports Medicine Australia.* 2014;17(5):496-500.

24. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. *Journal of medical engineering & technology.* 2016;40(6):298-306.

25. Diaz KM, Krupka DJ, Chang MJ, et al. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology.* 2015.

26. Stackpool CM, Porcari JP, Mikat RP, Gillette C, Foster C. The accuracy of various activity trackers in estimating steps taken and energy expenditure. *Journal of Fitness Research* 2014;3(3):32-48.

27. Noah AJ, 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. *Journal of medical engineering & technology.* 2013;37(7):456-462.

28. Beevi FH, Miranda J, Pedersen CF, Wagner S. An evaluation of commercial pedometers for monitoring slow walking speed populations. *Telemedicine Journal and E-Health.* 2016;22(5):441-449.

29. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *Journal of the American Medical Association.* 2015;313(6):625-626.

30. Diaz KM, Krupka DJ, Chang MJ, et al. Validation of the Fitbit One((R)) for physical activity measurement at an upper torso attachment site. *BMC Research Notes.* 2016;9(1):213.

31. 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.

32. 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. *International Journal of Behavioral Nutrition and Physical Activity.* 2015;12(1):42.

33. Gomersall Sjaan, Ng Norman, Pavey Toby, Wendy B. Free-living validation of consumer-based activity trackers as measures of physical activity and sedentary behaviour- Jawbone UP and Fitbit One. International Conference on Ambulatory Monitoring of Physical Activity and Movement (ICAMPAM); June 10-12, 2015; Limerick, Ireland

34. Tully MA, McBride C, Heron L, RF H. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Research Notes.* 2014;7(1).

35. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Medicine and science in sports and exercise.* 2014;46(9):1840-1848.

36. Bai Y, Welk GJ, Nam YH, et al. Comparison of consumer and research monitors under semistructured settings. *Medicine and science in sports and exercise.* 2015;48(1):151-158.

37. Murakami H, Kawakami R, Nakae S, et al. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. *Journal of American Medical Association Internal Medicine.* 2016.

38. Sasaki JE, Hickey A, Mavilia M, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. *Journal of Physical Activity and Health.* 2014.

39. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. *Journal of medical engineering & technology.* 2016;40(6):325-329.

40. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Medicine and science in sports and exercise.* 2013;45(11):2105-2112.

41. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. *Journal of medical Internet research.* 2016;18(9):e239.

42. Lyden K, Petruski N, Staudenmayer J, Freedson P. Direct observation is a valid criterion for estimating physical activity and sedentary behavior. *Journal of physical activity & health.* 2014;11(4):860-863.

43. 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. *European Journal of Preventive Cardiology.* 2016.

44. 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. *Telemedicine Journal and E-Health.* 2015;21(10):782-792.

45. Paul SS, Tiedemann A, Hassett LM, et al. Validity of theFitbit activity tracker for measuring steps in community-dwelling older adults. *BMJ Open Sport & Exercise Medicine.* 2015;1(1):e000013.

46. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. *Journals of Gerontology Series A, Biological Sciences and Medical Sciences.* 2016.

47. Matthew CE. Calibration of accelerometer output for adults. *Medicine and science in sports and exercise.* 2005;37(11 Suppl):S512-522.

48. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation.* 2013;128(20):2259-2279.

49. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *European Journal of Applied Physiology.* 2011;111(2):187-201.

50. Bassett DR, Troiano RP, McClain JJ, Wolff DL. Accelerometer-based physical activity: total volume per day and standardized measures. *Medicine and science in sports and exercise.* 2015;47(4):833-838.

51. Tracy DJ, Xu Z, Choi L, Acra S, Chen KY, Buchowski MS. Separating bedtime rest from activity using waist or wrist-worn accelerometers in youth. *PloS one.* 2014;9(4):e92512.

52. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *Journal of science and medicine in sport / Sports Medicine Australia.* 2011;14(5):411-416.

53. Ried-Larsen M, Brond JC, Brage S, et al. Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *International Journal of Behavioral Nutrition and Physical Activity.* 2012;9:113.

54. Brage S, Wedderkopp N, Franks PW, Andersen LB, Froberg K. Reliability and Validity of the Computer Science and Applications Accelerometer in a Mechanical Setting. *Measurement in physical education and exercise science.* 2003;7(2):101-119.

55. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. *Medicine and science in sports and exercise.* 2011;43(6):1085-1093.

56. Rothney MP, Apker GA, Song Y, Chen KY. Comparing the performance of three generations of ActiGraph accelerometers. *Journal of applied physiology.* 2008;105(4):1091-1097.

57. John D, Sasaki J, Staudenmayer J, Mavilia M, Freedson PS. Comparison of raw acceleration from the GENEA and ActiGraph GT3X+ activity monitors. *Sensors.* 2013;13(11):14754-14763.

58. John D, Sasaki J, Hickey A, Mavilia M, Freedson PS. ActiGraph activity monitors: "the firmware effect". *Medicine and science in sports and exercise.* 2014;46(4):834-839.

59. Fudge BW, Wilson J, Easton C, et al. Estimation of oxygen uptake during fast running using accelerometry and heart rate. *Medicine and science in sports and exercise.* 2007;39(1):192-198.

60. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. *Medicine and science in sports and exercise.* 2010;42(2):368-374.

61. Cain KL, Conway TL, Adams MA, Husak LE, Sallis JF. Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. *International Journal of Behavioral Nutrition and Physical Activity.* 2013;10:51.

62. Sirard JR, Masteller B, Mendoza AR, Hickey A, Freedson PS. Comprehensive laboratory- and field-based validation of youth-oriented activity trackers: The Physical Activity Tracker Testing in Youth (P.A.T.T.Y.) Study. . International Journal of Behavioral Nutrition and Physical Activity (In Review): University of Massachussetts Amherst; 2015.

63. Bassett DR, Jr., Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Medicine and science in sports and exercise.* 2012;44(1 Suppl 1):S32-38.

64. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and science in sports and exercise.* 2012;44(1 Suppl 1):S1-4.

65. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. *Medicine and science in sports and exercise.* 2011;43(8):1561-1567.

66. (NIH) NIoH. NIH invests almost $32 million to increase utility of biomedical research data. *NIH News and Events* 2014; http://www.nih.gov/news/health/oct2014/od-09.htm. Accessed August 1, 2015.

67. U.S. National Institutes of Health HHS. ClinicalTrials.gov. 2017; https://clinicaltrials.gov/ct2/results?term=fitbit&pg=1. Accessed July 5, 2017.

68. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *The Lancet.* 2014;384(9937):45-52.

69. Herrmann SD. Physical activity measurement. *Physical Activity and Public Health Practice*. Boca Raton, FL: CRC Press/Taylor & Francis, ©2012.; 2012:179-194.

70. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *British journal of sports medicine.* 2014;48(3):197-201.

71. Liberson WT, Holmquest HJ, Halls A. Accelerographic study of gait. *Archives of physical medicine and rehabilitation.* 1962;43:547-551.

72. Cavagna G, Saibene F, Margaria R. A three-directional accelerometer for analyzing body movements. *Journal of applied physiology.* 1961;16:191.

73. Cavagna G, Saibene F, R M. External work in walking. *Journal of applied physiology.* 1963;18:1-9.

74. Morris JR. Accelerometry--a technique for the measurement of human body movements. *Journal of Biomechanics.* 1973;6(6):729-736.

75. Bassey EJ, Dallosso HM, Fentem PH, Irving JM, Patrick JM. Validation of a simple mechanical accelerometer (pedometer) for the estimation of walking activity. *European Journal of Applied Physiology and Occupational Physiology.* 1987;56(3):323-330.

76. Montoye HJ, Washburn R, Servais S, Ertl A, Webster JG, Nagle FJ. Estimation of energy expenditure by a portable accelerometer. *Medicine and science in sports and exercise.* 1983;15(5):403-407.

77. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and science in sports and exercise.* 1998;30(5):777-781.

78. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences.* 2008;26(14):1557-1565.

79. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and science in sports and exercise.* 2000;32(9 Suppl):S442-449.

80. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and science in sports and exercise.* 2000;32(9):S450-S456.

81. John D, Miller R, Kozey-Keadle S, Caldwell G, Freedson P. Biomechanical examination of the 'plateau phenomenon' in ActiGraph vertical activity counts. *Physiological measurement.* 2012;33(2):219-230.

82. Nichols JF, Morgan CG, Sarkin JA, Sallis JF, Calfas KJ. Validity, reliability, and calibration of the Tritrac accelerometer as a measure of physical activity. *Medicine and science in sports and exercise.* 1999;31(6):908-912.

83. Pfeiffer KA, McIver KL, Dowda M, Almeida MJ, Pate RR. Validation and calibration of the Actical accelerometer in preschool children. *Medicine and science in sports and exercise.* 2006;38(1):152-157.

84. Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. *Obesity (Silver Spring).* 2006;14(11):2000-2006.

85. Rowlands AV, Stiles VH. Accelerometer counts and raw acceleration output in relation to mechanical loading. *Journal of Biomechanics.* 2012;45(3):448-454.

86. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obesity Research.* 2002;10(3):150-157.

87. Treuth MS, Schmitz K, Catellier DJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. *Medicine and science in sports and exercise.* 2004;36(7):1259-1266.

88. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. *Medicine and science in sports and exercise.* 2005;37(Supplement):S523-S530.

89. Mattocks C, Leary S, Ness A, et al. Calibration of an accelerometer during free-living activities in children. *International Journal of Pediatric Obesity.* 2007;2(4):218-226.

90. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise.* 2008;40(1):181-188.

91. Crouter SE, DellaValle DM, Haas JD, Frongillo EA, Bassett DR. Validity of ActiGraph 2-regression model, Matthews cut-points, and NHANES cut-points for assessing free-living physical activity. *Journal of physical activity & health.* 2013;10(4):504-514.

92. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine and science in sports and exercise.* 2005;37(Supplement):S501-S511.

93. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology.* 2006;98(6):601-612.

94. 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):1946-1952.

95. Heil DP. Predicting activity energy expenditure using the Actical activity monitor. *Research Quarterly for Exercise and Sport.* 2006;77(1):64-80.

96. Leenders NY, Nelson TE, Sherman WM. Ability of different physical activity monitors to detect movement during treadmill walking. *International journal of sports medicine.* 2003;24(1):43-50.

97. Melanson EL, Jr., Freedson PS. Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. *Medicine and science in sports and exercise.* 1995;27(6):934-940.

98. Crouter SE, Clowers KG, Bassett DR, Jr. A novel method for using accelerometer data to predict energy expenditure. *Journal of applied physiology.* 2006;100(4):1324-1331.

99. Crouter SE, Bassett DR, Jr. A new 2-regression model for the Actical accelerometer. *British journal of sports medicine.* 2008;42(3):217-224.

100. Crouter SE, Dellavalle DM, Horton M, Haas JD, Frongillo EA, Bassett DR, Jr. Validity of the Actical for estimating free-living physical activity. *European Journal of Applied Physiology.* 2011;111(7):1381-1389.

101. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR, Jr. Refined two-regression model for the ActiGraph accelerometer. *Medicine and science in sports and exercise.* 2010;42(5):1029-1037.

102. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. *Medicine and science in sports and exercise.* 2014;46(2):386-397.

103. Mannini A, Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors.* 2010;10(2):1154-1175.

104. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. *Physiological measurement.* 2009;30(4):R1-33.

105. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiological measurement.* 2014;35(11):2183-2189.

106. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and science in sports and exercise.* 2006;38(9):1626-1634.

107. Ellingson LD, Schwabacher IJ, Kim Y, Welk GJ, Cook DB. Validity of an integrative method for processing physical activity data. *Medicine and science in sports and exercise.* 2016.

108. Liu S, Gao RX, John D, Staudenmayer JW, Freedson PS. Multisensor data fusion for physical activity assessment. *Transactions on Biomedical Engineering.* 2012;59(3):687-696.

109. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENEA: optimum sampling frequency and number of axes. *Medicine and science in sports and exercise.* 2012;44(11):2228-2234.

110. Brage S, Wedderkopp N, Franks PW, Andersen LB, Froberg K. Reexamination of validity and reliability of the CSA monitor in walking and running. *Medicine and science in sports and exercise.* 2003;35(8):1447-1454.

111. Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. *Medicine and science in sports and exercise.* 2007;39(4):716-727.

112. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of activity monitor step detection Is related to movement patterns. *Journal of Physical Activity and Health.* 2016;13(2):145-153.

113. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. *International journal of sports medicine.* 2013;34(11):975-982.

114. McMinn D, Acharya R, Rowe DA, Gray SR, Allan JL. Measuring activity energy expenditure- accuracy of the GT3X+ and Actiheart monitors. *International Journal of Exercise Science.* 2013;6(3):217-229.

115. Santos-Lozano A, Torres-Luque G, Marin PJ, Ruiz JR, Lucia A, Garatachea N. Intermonitor variability of GT3X accelerometer. *International journal of sports medicine.* 2012;33(12):994-999.

116. Whitcher L, Papadopoulos C. Accelerometer derived activity counts and oxygen consumption between young and older individuals. *Journal of aging research.* 2014;2014:184693.

117. Mahar TF, Rowe DA, Mahar MT. Comparison of ActiGraph hip worn and wrist worn activity monitors for assessment of physical activity. *Medicine and science in sports and exercise.* 2013;45(5):322.

118. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and science in sports and exercise.* 2014;46(9):1816-1824.

119. Speakman JR. The history and theory of the doubly labeled water technique. *The American journal of clinical nutrition.* 1998;68(4):932S-938S.

120. Heyman MB, Fuss P, Young VR, Evans WJ, Roberts SB. Prediction of total energy expenditure using the Caltrac activity monitor. *International journal of obesity.* 1991;15 (supplement 1).

121. Calabro MA, Stewart JM, Welk GJ. Validation of pattern-recognition monitors in children using doubly labeled water. *Medicine and science in sports and exercise.* 2013;45(7):1313-1322.

122. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Medicine and science in sports and exercise.* 2010;42(11):2134-2140.

123. Calabro MA, Lee JM, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. *International Journal of Behavioral Nutrition and Physical Activity.* 2014;11:119.

124. Klesges RC, Klesges LM, Swenson AM, Pheley AM. A validation of two motion sensors in the prediction of child and adult physical activity levels. *American journal of epidemiology.* 1985;122(3):400-410.

125. Lyden K, Kozey Keadle SL, Staudenmayer JW, Freedson PS. Validity of two wearable monitors to estimate breaks from sedentary time. *Medicine and science in sports and exercise.* 2012;44(11):2243-2252.

126. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *British journal of sports medicine.* 2014;48(13):1019-1023.

127. Inc. F. What are active minutes? How do I earn active minutes? 2016; Definition of active minutes and how they are accumulated. Available at: https://help.fitbit.com/articles/en_US/Help_article/1379. Accessed September 15, 2016.

128. Fruin ML, Rankin JW. Validity of a multi-sensor armband in estimating rest and exercise energy expenditure. *Medicine and science in sports and exercise.* 2004;36(6):1063-1069.

129. Schneider PL, Crouter SE, Lukajic O, Bassett DR, Jr. Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Medicine and science in sports and exercise.* 2003;35(10):1779-1784.

130. Steeves JA, Tyo BM, Connolly CP, Gregory DA, Stark NA, Bassett DR. Validity and reliability of the Omron HJ-303 tri-axial accelerometer-based pedometer. *Journal of Physical Activity and Health.* 2011;8(7):1014-1020.

131. Ryan CG, Grant PM, Tigbe WW, Granat MH. The validity and reliability of a novel activity monitor as a measure of walking. *British journal of sports medicine.* 2006;40(9):779-784.

132. Nelson CP, Hamby SE, Saleheen D, et al. Genetically determined height and coronary artery disease. *New England Journal of Medicine.* 2015;372(17):1608-1618.

133. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. *American journal of epidemiology.* 2008;167(7):875-881.

134. Orendurff MS, Segal AD, Klute GK, Berge JS, Rohr ES, Kadel NJ. The effect of walking speed on center of mass displacement. *The Journal of Rehabilitation Research and Development.* 2004;41(6):829.

135. Statistics USDoLBoL. Time use on an average work day for employed persons ages 25 to 54 with children. *American Time Use Survey* 2014; http://www.bls.gov/tus/charts/home.htm. Accessed April 30, 2016, 2016.

136. RC T. R: A Language and Environment for Statistical. *Vienna, Austria: R Foundation for Statistical Computing.* 2013.

137. Mudge S, Taylor D, Chang O, Wong R. Test-retest reliability of the StepWatch activity monitor outputs in healthy adults. *Journal of Physical Activity and Health.* 2010;7(5):671-676.

138. Foster RC, Lanningham-Foster LM, Manohar C, et al. Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Preventive medicine.* 2005;41(3-4):778-783.

139. New Lifestyles I. NEW LIFESTYLES NL-1000 Activity Monitor user's guide & record book. 2005.

140. Tractica. The Wearable Devices Market is Poised for Expansion into Smart Clothing and Body Sensors. 2015; https://www.tractica.com/newsroom/press-releases/the-wearable-devices-market-is-poised-for-expansion-into-smart-clothing-and-body-sensors/. Accessed February 25, 2016.

141. Analog Devices I. IoT Solution Gives Sports Teams a Competitive Edge by Optimizing Athlete and Team Performance. 2016; http://investor.analog.com/releasedetail.cfm?releaseid=970186. Accessed May 13, 2016.

142. (NASA) NAaSA. NASA'S HERA MISSION IS WEARING HEXOSKIN TECH. 2016; https://www.nasa.gov/hrp/research/analogs/hera. Accessed May 13, 2016.

143. Banerjee T, Anantharam P, Romine WL, Lawhorne L, Sheth A. Evaluating a Potential Commercial Tool for Healthcare Application for People with Dementia. Write State University; 2015.

144. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of physical activities: a second update of codes and MET values. *Medicine and science in sports and exercise.* 2011;43(8):1575-1581.

145. CDC. National Health and Nutrition Examination Survey (NHANES): Physical Activity Monitor (PAM) Procedures Manual. 2011.

146. Ainsworth BE, Haskell WL, Herrmann SD, et al. The Compendium of Physical Activities Tracking Guide. 2011; https://sites.google.com/site/compendiumofphysicalactivities/. Accessed April 23, 2014.

147. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. *The American journal of clinical nutrition.* 1990;51(2):241-247.

148. Frankenfield DC, Rowe WA, Smith JS, Cooney RN. Validation of several established equations for resting metabolic rate in obese and nonobese people. *Journal of the American Dietetic Association.* 2003;103(9):1152-1159.

149. Frankenfield D, Roth-Yousey L, Compher C. Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: a systematic review. *Journal of the American Dietetic Association.* 2005;105(5):775-789.

150. Sirard JR, Masteller B, Freedson PS, Mendoza A, Hickey A. Youth oriented activity trackers: comprehensive laboratory- and field-based validation. *Journal of medical Internet research.* 2017;19(7):e250.

151. Ltd. G. Garmin Vivofit Specs. 2017; https://buy.garmin.com/en-US/US/p/143405. Accessed August 2, 2017.

152. Abel MG, Peritore N, Shapiro R, Mullineaux DR, Rodriguez K, Hannon JC. A comprehensive evaluation of motion sensor step-counting error. *Applied Physiology, Nutrition, and Metabolism.* 2011;36(1):166-170.

153. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *Journal of medical engineering & technology.* 2017;41(3):208-215.

154. Boerema ST, van Velsen L, Schaake L, Tonis TM, Hermens HJ. Optimal sensor placement for measuring physical activity with a 3D accelerometer. *Sensors.* 2014;14(2):3188-3206.

155. Coleman KL, Smith DG, Boone DA, Joseph AW, del Aguila MA. Step activity monitor: long-term, continuous recording of ambulatory function. *Journal of Rehabilitation Research and Development.* 1999;36(1):8-18.

156. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *Journal of applied physiology.* 2015;119(4):396-403.

157. Le Masurier GC, Tudor-Locke C. Comparison of pedometer and accelerometer accuracy under controlled conditions. *Medicine and science in sports and exercise.* 2003;35(5):867-871.

158. Le Masurier GC, Lee SM, Tudor-Locke C. Motion sensor accuracy under controlled and free-living conditions. *Medicine and science in sports and exercise.* 2004:905-910.

159. An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? *European Journal of Sport Science.* 2017;17(3):360-368.

160. Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F. Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. *Acta Gymnica.* 2016.

161. Reid RER, Insogna JA, Carver TE, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of science and medicine in sport / Sports Medicine Australia.* 2017;20(6):578-582.

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

163. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent validity of wearable activity trackers under free-living conditions. *Journal of Strength and Conditioning Research.* 2017;31(4):1097-1106.

164. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: A comparative study in free-living conditions. *Journal of Medical Internet Research mHealth and uHealth.* 2016;4(3):e110.

165. Van Blarigan EL, Kenfield SA, Tantum L, Cadmus-Bertram LA, Carroll PR, Chan JM. The Fitbit One physical activity tracker in men with prostate cancer: Validation study. *Journal of Medical Internet Research Cancer.* 2017;3(1):e5.

166. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine and science in sports and exercise.* 2016;48(3):457-465.

167. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sports Medicine.* 2001;31(6):439-454.

168. Tudor-Locke C, Barreira TV, Schuna JM, Jr. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine and science in sports and exercise.* 2015;47(4):839-842.

169. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed using hip- and wrist-worn accelerometers. *Gait & Posture.* 2016;44:23-28.

170. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. *American journal of preventive medicine.* 2015.

171. Montoye AH, Pfeiffer KA, Suton D, Trost SG. Evaluating the responsiveness of accelerometry to detect change in physical activity. *Measurement in Physical Education and Exercise Science.* 2014;18(4):273-285.

172. Whye L, Ng C, Jenkins S, Hill K. Accuracy and responsiveness of the StepWatch activity monitor and ActivPAL in patients with COPD when walking with and without a rollator. *Journal of Disability and Rehabilitation.* 2012;34(15).