View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks@UMass Amherst

University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Masters Theses

Dissertations and Theses

December 2020

Validity Parameters for Step Counting Wearable Technologies During Treadmill Walking in Young People 6-20 Years of Age

Zachary Gould

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2

Part of the Public Health Commons

Recommended Citation

Gould, Zachary, "Validity Parameters for Step Counting Wearable Technologies During Treadmill Walking in Young People 6-20 Years of Age" (2020). *Masters Theses*. 966.
https://scholarworks.umass.edu/masters_theses_2/966

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

VALIDITY PARAMETERS FOR STEP COUNTING WEARABLE TECHNOLOGIES DURING TREADMILL WALKING IN YOUNG PEOPLE 6-20 YEARS OF AGE

A Thesis Presented

by

ZACHARY R. GOULD

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

MASTER OF SCIENCE

September 2020

Kinesiology

VALIDITY PARAMETERS FOR STEP COUNTING WEARABLE TECHNOLOGIES DURING TREADMILL WALKING IN YOUNG PEOPLE 6-20 YEARS OF AGE

A Thesis Presented

by

ZACHARY R. GOULD

Approved as to style and content by:

____

Catrine Tudor-Locke, Chair

____

John Sirard, Member

____

Amanda Paluch, Member

____

Jane Kent, Department Chair

Department of Kinesiology

ACKNOWLEDGEMENTS

I want to give my sincerest thanks to everyone who has helped me achieve this goal of successfully completing this thesis project and, further, this Kinesiology graduate program. To my advisor, Dr. Catrine Tudor-Locke, thank you so much for taking a chance on me and extending an invitation to be involved in your research, originally as a lab manager, and then as a graduate student. I am forever indebted to you for providing me this opportunity to learn and grow as a researcher, and as a young professional under your guidance. To my committee, Dr. John Sirard and Dr. Amanda Paluch, thank you for all of your support, your guidance, and your patience, helping me navigate this process, both with this project, and beyond. To Dr. John Staudenmayer, thank you for your time, your knowledge, and willingness to help thoroughly comprehend the proper statistics necessary to appropriately address this project, even with the remote learning style of the Spring 2020 semester. To the rest of the Kinesiology Department faculty, thank you for your expertise, and genuine interest in our learning, with all of your courses, lectures, and seminars. To the original “Team Tudor-Locke”, with it’s now transcontinental roots, thank you for your continued guidance, support, and mentorship, helping me develop as a curious student, thoughtful researcher, and prepared, independent young adult. To the rest of the Physical Activity and Health lab and Kin graduate students, your friendship and memories made together, both in and out of the classroom, is something so special that I will carry with me forever. And lastly, to my family, Mom, Dad, and Leslie, you guys are my rock, and have unwaveringly supported me every single moment of this journey, and beyond. You have held me up during my lowest times, and raised me even higher celebrating my accomplishments, and for that I feel so lucky. I would not be able to accomplish this without the help of each and every one of you, so I just want to say thank you.

ABSTRACT

VALIDITY PARAMETERS FOR STEP COUNTING WEARABLE TECHNOLOGIES DURING TREADMILL WALKING IN YOUNG PEOPLE 6-20 YEARS OF AGE

SEPTEMBER 2020

ZACHARY R. GOULD, B.S. UNIVERSITY OF MASSACHUSETTS AMHERST

M.S. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Catrine Tudor-Locke

Introduction: Wearable technologies play an important contemporary role in the measurement of physical activity (PA) and promotion of human health across the lifespan, including for young people (i.e., children, adolescents, and young adults). As new objective wearable technologies continue to develop, standardized approaches to documenting validation parameters (i.e., measures of accuracy, precision, and bias) are needed to ensure confidence and comparability in step-defined PA. **Purpose:** To produce validity parameters for step counting wearable technologies during treadmill walking in young people 6-20 years of age **Methods:** 120 participants completed 5-minute treadmill bouts from13.4 to 134.1 $m \cdot min^{-1}$. Participants wore eight technologies (two at the arm/wrist, four at the waist, one on the thigh, and one on the ankle) while steps were directly observed. Speed, wear location, and age -specific measures of accuracy (mean absolute percent error; MAPE), precision (correlation coefficient, standard deviation; SD, coefficient of variation; CoV), and bias (percent error; PE) were computed and cataloged. **Results:** Speed and wear location had a significant effect on accuracy and bias measures for wearable technologies ($p<0.0001$), but not for precision measures ($p=0.24$, $p=0.06$). Age did not have an effect on accuracy, precision, or bias measures ($p=0.21$-0.50, $p=0.84$, $p=0.21$-0.50). **Conclusion:** While the analyses indicate the significance of speed and wear location on wearable technology performance, the useful and comprehensive

validity reference values cataloged herein will help optimize measurement of PA in youth. Future research should continue to rigorously validate new wearable technologies as they are developed, and also extend these standardized reference values developed in the laboratory to the free-living environment.

TABLE OF CONTENTS

Page

AKNOWLEDGEMENTS…………………………………………………………………..iii
ABSTRACT……………………………………………………………………………...iv
LIST OF TABLES viii
LIST OF FIGURES ix
LIST OF ABBREVIATIONS xi
CHAPTER
1. INTRODUCTION 1
1.1 Physical Activity and Health 1
1.1.1 Measurement of Physical Activity (PA) 3
1.1.2 Wearable Technologies 8
1.1.3 Step Counting 11
1.1.4 Step Counting in Youth 14
1.1.5 Proper Assessment Methods for Validity 15
1.1.6 Accuracy 16
1.1.7 Precision 17
1.1.8 Bias 20
1.1.9 Threats to Validity 22
1.2 Summary 25
1.3 Purpose of the Study 26
1.4 Aims and Hypotheses 27
2. LITERATURE REVIEW 28
2.1. Overview 28
2.2 Objective Activity Monitoring in Youth 28
2.2.1 Pedometers 30
2.2.2 Accelerometers 35
2.2.3 Summary of Literature Review 37
3. METHODS 39
3.1 Computing and Comparing Wearable Technology Step Counts in Youth 39
3.2 Participants 40
3.3 Protocol 41
3.4 Measures 41
3.4.1 Participant Characteristics 41

3.4.2 Treadmill-based Variables 42
3.5 Statistical Analysis 42
3.5.1 Descriptive Statistics 43
3.5.2 Inferential Analysis 44
4. RESULTS 47
4.1 Participant characteristics 47
4.2 Accuracy: MAPE 47
4.2.1 MAPE by speed 47
4.2.2 MAPE by wear location 49
4.2.3 MAPE by age group 50
4.3 Precision: Correlation 50
4.3.1 Correlation by speed 50
4.3.2 Correlation by wear location 52
4.3.3 Correlation by age group 53
4.4 Bias: PE 53
4.4.1 PE by speed 53
4.4.2 PE by wear location 54
4.4.3 PE by age group 55
5. DISCUSSION 57
5.1 Effect of speed: 58
5.2 Effect of wear location: 60
5.3 Effect of age: 64
5.4 Validity metrics 64
5.5 Rank order 65
5.7 Limitations: 69
6. CONCLUSION 72
REFERENCES 115

LIST OF TABLES

Table Page

Table 1: Step counting treadmill validation studies among wearable technologies in youth………………………..…………………………………………………………73

Table 2: Participant characteristics…………………..…………………………………..75

Table 3: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies……………………………………………………………….…………..76

Table 4: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for children (6-12 years)…………………………………………….………78

Table 5: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for adolescents (13-17 years)……………………………………………….….80

Table 6: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for young adults (18-20 years)………………………………………………..…...82

Table 7: Precision: Correlation coefficient *(r)* of the tested wearable technologies by speed………………………………………………………………………………..84

Table 8: Bias: Percent error (PE) of the wearable technologies by speed……...………..85

Table 9: Bias: Percent error (PE) of the wearable technologies by speed for children (6-12 years)…………………………………………………………………………...…...87

Table 10: Percent error (PE) of the wearable technologies by speed for adolescents (13-17 years)………………………………………………………………………………….89

Table 11: Percent error (PE) of the wearable technologies by speed for young adults (18-20 years)………………………………………………………………………………….91

Table 12: Validity parameters (accuracy, precision, and bias) by wear location averaged across all wearable technologies…………………………………………………………93

Table 13: Summary validity parameters (accuracy, precision, and bias) by age averaged across all wearable technologies…………………………………………………………95

LIST OF FIGURES

Figures Page
Figure 1: MAPE of wearable technologies across walking speeds……………….………..97

Figure 2: SD of MAPE of wearable technologies across walking speeds……………..…97

Figure 3: MAPE of wearable technologies across walking speeds for children (6-12 years)………………………………………………………………………………….…….98

Figure 4: SD of MAPE of wearable technologies across walking speeds for children (6-12 years)…...……………………………………………………......……………...98

Figure 5: MAPE of wearable technologies across walking speeds for adolescents (13-17 years)………………..…………………………………………………………………99

Figure 6: SD of MAPE of wearable technologies across walking speeds for adolescents (13-17 years)……………………………………………………………….………….99

Figure 7: MAPE of wearable technologies across walking speeds for young adults (18-20 years)………………………………………………………………………...100

Figure 8: SD of MAPE of wearable technologies across walking speeds for young adults (18-20 years)…………………………………………………………………………100

Figure 9: Correlations of each wearable technology across all walking bouts…………101

Figure 10: PE of wearable technologies across walking speeds…………………….….102

Figure 11: SD of PE of wearable technologies across walking speeds…….……….102

Figure 12: PE of wearable technologies across walking speeds for children (6-12 years)……………………………………………………………………………….……...103

Figure 13: SD of PE of wearable technologies across walking speeds for children (6-12 years)……………………………………………………………………………....103

Figure 14: PE of wearable technologies across walking speeds for adolescents (13-17 years)………………………………………………………………………….…...104

Figure 15: SD of PE of wearable technologies across walking speeds for adolescents (13-17 years)………………………………………………………………...........104

Figure 16: PE of wearable technologies across walking speeds for young adults (18-20 years)……………………………………………………………………………..…..105

Figure 17: SD of PE of wearable technologies across walking speeds for young adults (18-20 years) ...……………………………………………………………………....105

Figure 18: The effect of speed on the accuracy of wearable technology's step counting ability…………………………………………………………………………………...106

Figure 19: The effect of speed on precision of wearable technologies step counting ability…………………………………………………………………………………...107

Figure 20: The effect of speed on the bias of wearable technology's step counting ability…………………………………………………………………………………...108

Figure 21: The effect of wear location on the overall accuracy of wearable technology's step counting abilities……………………………………………………………………..109

Figure 22: The effect of wear location on overall the precision of wearable technology's step counting abilities……………………………………………………………………..110

Figure 23: The effect of wear location on the overall bias of wearable technology's step counting abilities…………………………………………………………..……………111

Figure 24: The effect of age on the accuracy of wearable technology's step counting ability………………………………………………………………………….………..112

Figure 25: The effect of age on overall precision of wearable technology step counting ability…………………………………………………………………………………...113

Figure 26: The effect of age on the bias of wearable technology's step counting ability…………………………………………………………………………………...114

LIST OF ABBREVIATIONS

PA- Physical activity

PAGAC- Physical Activity Guidelines Advisory Committee

MVPA- Moderate-to-vigorous physical activity

METs- Metabolic equivalents

miles·hour^{-1} - miles per hour

m·min^{-1}- meters per minute

O_2- Oxygen

CO_2- Carbon Dioxide

VO_2- Volume of oxygen consumed

VCO_2- Volume of carbon dioxide produced

BMI- Body mass index

EMA- Ecological momentary assessment

CTA- Consumer Technology Association

MAPE- Mean absolute percent error

RMSE- Root mean square error

PE- Percent error

SD- Standard Deviation

CoV- Coefficient of variation

ICC- Intraclass correlation coefficient

CI- Confidence Interval

CHAPTER 1

INTRODUCTION

1.1 Physical Activity and Health

The numerous health benefits of regular physical activity (PA) across the lifespan have been well documented.[1] The second edition of the federal Physical Activity Guidelines for Americans released in 2018 compiled updated scientific evidence to establish guidelines for minimum levels of PA associated with positive health outcomes. These guidelines were developed by the Physical Activity Guidelines Advisory Committee (PAGAC), a group of esteemed researchers and experts who comprehensively reviewed the current research relevant to PA and health and provided detailed direction for disease prevention and health promotion relevant to people 3 years of age or older.[2] Specifically, the focus was on providing updated and clear information regarding recommended levels of moderate to vigorous intensity PA (MVPA) associated with reduced risk of premature mortality and development of chronic diseases such as heart disease, cancer, and diabetes. The PAGAC reports the synthesis of current PA research (specific to aerobic, muscle strengthening, bone strengthening, balance and flexibility activities) according to age-defined populations: youth (children and adolescents; 3-17 years of age), adults (18-64 years of age), and older adults (65 years of age and older).

The guidelines relevant to aerobic-type activities (like brisk walking or running) are formatted in terms of dose conveyed as frequency, accumulated time, and intensity. Specifically, they recommend 150 min·week^{-1} of moderate intensity or 75 min·week^{-1} of vigorous intensity PA for adults. The intensity of PA referred to in the guidelines is based on an objective and absolute definition which uses calorimetry to measure rate of oxygen consumption in METs (metabolic equivalents). One MET is equivalent to the baseline amount of oxygen your body requires at rest while awake. Oxygen consumption measured at 3.0-5.9 METs is considered to be moderate intensity, while ≥ 6.0 METS is vigorous intensity. For example, brisk walking at about 2.5 miles$\cdot$hour^{-1} (67.0m$\cdot$min^{-1}) requires 3.0 METs of energy to be expended, which classifies that activity as moderate intensity.[3] Further, increasing walking speed to 4.5 miles$\cdot$hour^{-1} (120.7 m$\cdot$min^{-1}) increases its intensity to 7.0 METs, classifying it as vigorous.[3]

While these dose guidelines are expressed specifically for adults, the PA guidelines for youth (3-17 years) differ somewhat based on recommended accumulated time. Specifically, diverse health benefits can be achieved by accumulating a daily total of 60 minutes of MVPA.[2] In youth samples, the risk of chronic disease and mortality are difficult to determine due to their prolonged development timeline, however, there is still a strong evidence-based association between PA and musculoskeletal health, cardiovascular health, adiposity levels, and blood pressure levels in youth.[4] For example, a 2014 study by Janz et al.[5] reported that girls and boys between the ages of five to 17 years who accumulated high levels of accelerometer-determined MVPA displayed improved bone health. The most active children (≥85 min $\cdot$day^{-1} of MVPA) realized significant benefits across multiple markers of bone health compared to their least active peers (≤85 min $\cdot$day^{-1} of MVPA). Anderson et al.[6] conducted a cross-sectional study of accelerometer-determined MVPA and cardiovascular risk factors with 1,732 children (52.8% female) aged nine to 15 years.[5] They reported that the most active group of children (167 minutes $\cdot$day^{-1} of MVPA) showed improved levels of glucose, insulin, cholesterol, and triglycerides compared to less active groups (≤116 min $\cdot$day^{-1} of MVPA) suggesting that increased levels of MVPA may lead to decreased cardiovascular disease risk factors in youth. Further, Janz et al.[7] conducted a longitudinal study with 333 children (55.6% female, ages 5, 8, and 11 years) and reported that those with higher levels of accelerometer-assessed MVPA at age five (51 vs. nine minutes for boys, 45 vs. 11 minutes for girls) had significantly lower fat mass during follow up visits at age eight and 11 ($p<0.05$). Lastly, a 2013 longitudinal study by Knowles et al.[8] demonstrated strong evidence supporting an association between higher levels of MVPA and lower blood pressure. In the Knowles et al.[8] study, 427 children (49.6% female) were assessed via accelerometry at six years of age and then again at eight years of age. For every 15 minutes of increased MVPA documented between these ages, there was a 0.55 mm Hg decrease in diastolic blood pressure ($P=0.03$).

The extent to which PA can modify health outcomes is potentially profound. However, PA can be assessed in numerous ways, depending on the outcome of interest. The lack of consensus in regards to the best practices for capturing PA is problematic. There is a need for standardized, accurate, precise and unbiased measures of PA that are also acceptable and interpretable to the widest possible audience. These are needed to assure confidence in any future expressions of dose associated with health in youth and to fully realize the potential for PA-related health throughout the life course.

1.1.1 Measurement of Physical Activity (PA)

There are several ways to estimate PA behavior including indirect calorimetry, doubly labelled water, heart rate monitoring, self-reported behavior (e.g., questionnaires, diaries, or logs), ecological momentary assessment (EMA), direct observation, and using wearable technologies such as accelerometers and pedometers. Each of these approaches has their own unique advantages and disadvantages. To start, indirect calorimetry estimates PA intensity by measuring the volume of chemical energy of the pulmonary gas exchanged during PA.[9] Specifically, oxygen (O_2) consumption during inhalation and carbon dioxide (CO_2) production (as the byproduct during exhalation) are tracked during activity as the volume of oxygen consumed (VO_2) and carbon dioxide produced (VCO_2), respectively. A mask is worn covering the nose and mouth with tubes attached near the mouth which are then connected to computers to quantify the relevant gases contained in inhaled and exhaled air. VO_2 and VCO_2 per unit of time is captured in order to calculate energy expenditure and convert it to METs where the ultimate unit of measure is ml O_2/kg body weight/minute. While it is considered to be an accurate and valid assessment (in comparison to direct measurement of total energy expenditure via body heat production assessed through direct calorimetry), the equipment necessary for the gas analysis (mask, computer equipment, software access) creates challenges for free-living PA measurements. Direct calorimetry is conducted in a calibrated, insulated room or chamber and measures the total amount of energy used via heat production emanated from the body through digestion of foods, mechanical energy for movement, etc. During PA, the body naturally produces heat through the mechanical energy required to perform tasks.[10] Both indirect calorimetry and direct calorimetry align with measures of heat produced during PA.[10]

Doubly labelled water is another way to assess energy expenditure, providing an estimate of overall PA level. A dose of radio-labelled isotope ($^2H_2{}^{18}O$) is ingested and the rates of O_2 consumption and CO_2 production produced from the washout of the isotope is measured over five to 14 days to estimate energy expenditure.[11] The doubly labelled water technique estimates within 8% of total daily energy expenditure when compared to body heat production measured through direct calorimetry.[12] However, the doubly labelled water technique is expensive, the isotopes are difficult to obtain, and controlling for the energy lost through the digestion of food via strict dietary records is especially difficult for children.[11] Further, the required washout period is prolonged (i.e., over the course of several days to measure total energy expenditure), making it impossible to assess energy expenditure specific to different activities performed in real time.[11]

Another way of estimating energy expenditure during PA is with heart rate monitoring. Specifically, electrodes embedded in chest straps are used to measure the cardiac electrical activity that represents the heart rate (e.g., beats·min^{-1}). An elevated heart rate is interpreted as an increased PA intensity.[13] Eston, Rowlands, and Ingledew[14] studied 30 children (50% female, age=8.2-10.8 years) who wore a chest heart rate monitor and walked and ran on a treadmill at four different speeds (walking: 2.5, 3.7 mph, running: 5.0, 6.2 mph). The heart rate monitor displayed a linear trend (R^2=0.78) with oxygen consumption across all treadmill speeds.[14] However, during lighter intensity activities or sedentary behavior there may be more variability in heart rate which may not be as indicative of PA behavior.[11] Specifically, environmental or psychological factors may have a greater effect than PA on heart rate.[11] To be clear, heart rate during sedentary or light intensity behaviors is generally quite low, however high levels of stress or anxiety, level of fitness, type of muscular contraction, hydration levels, and other environmental factors may influence heart rate.[15] Such externally shaped heart rate changes are not indicative of PA intensity.

Subjective measures of PA, including questionnaires, diaries, logs, and EMA rely on the participants themselves to recall/report their perceived levels of PA participation.

Questionnaires require responses to a range of queries regarding frequency (e.g., per day, per week, per month, or per year), intensity (e.g., perceived light, moderate or vigorous), duration (e.g., minutes, hours, days, weeks), type (e.g., walking, running, dancing, etc.), and/or domain (e.g., occupation, leisure, transportation, or household). Specific questions may focus on PA behavior in the past 24 hours, past week, or past few months.[16] For example, self-reported PA assessments could ask the individual how often they ran in the past seven days, and then assign an intensity score based on how difficult the activity was in regards to their own perceptions, providing information on type, intensity, duration, and frequency of PA behaviors. Similar to questionnaires, PA diaries ask participants to enter frequency, intensity, duration, type, and domain specific details periodically, preferably in almost real time.[16] Such information can be recorded intermittently throughout the day, for a single day or for multiple weeks. This approach is similar to using PA logs, which are simpler self-reported PA assessment instruments that typically include a checklist of activities for participants to select from, as well as a corresponding intensity score.[14]

Additionally, PA behavior can be captured through EMA. EMA relies on the recent advancements of mobile and sensor technologies (i.e. smartphones) to capture real time self-reported information on a variety of PA behavior contexts.[17] For example, a free-living PA behavior can be performed, and the use of EMA through smartphones can query the individual gathering self-reported information on the contexts of the behaviors, including type, intensity, emotional states, beliefs, attitudes, and perceptions. Data collected over the course of longer periods of time, across a large number of people, can perhaps start to give insight to determinants of PA as well. Further, because the PA behaviors are recalled more proximal to the actual performance of the PA, EMA reduces the amount of recall errors and biases, as the information is more immediate in memory.

Overall, subjective measures can be more easily used for large population studies as they can be administered via mail or electronically, and produce insightful, and highly descriptive information relevant to perceived PA behavior.[18] However, depending on how frequently participants are queried, self-report may be misconstrued and may not correctly represent all PA behaviors. For example, a self-report measure asking participants how often they participated in PA several months ago may hinder their ability to correctly recall the actual PA performed because of the inherent difficulty of accurately recalling distant behaviors (i.e., recall bias).[16] Specific to youth samples, a 1985 study by Wallace, Mackenzie, and Nader[19] investigated 7-day self-reported PA in eleven 11-13 year old boys and reported that the ability to recall PA progressively declines from day 1 to day 7. PA of eleven boys 11-13 years of age enrolled at a summer camp was documented by a camp counselor every 15 minutes (not including sleep) for a full week. Following the documentation period, the participants were asked to recall their PA behaviors during the past week. The researchers found that participants could only recall 46% of their PA over the week, with increased misreporting for greater recall periods. More specifically, there was greater error when recalling PA from seven days prior compared to 1 day prior to the questionnaire date. Further, participants may be tempted to record PA levels to align with what they think the researcher may want (i.e., social desirability bias).[16] Lastly, self-reported PA data may be skewed or incorrect due to a lack of understanding of actual intensity or duration.[11] For example, someone who played in a soccer game may report that they were physically active for two hours at a vigorous intensity, when in fact they were mainly standing on the field and performing only intermittent bursts of running sporadically over the course of the two hour match. Such issues raise concerns about accuracy and precision of subjective measures of PA.[20]

Direct observation is a common assessment method for PA measurement in children.[21] It entails researchers watching in real time, or video recording and subsequently watching, individuals participating in PA in order to document the observed behaviors.[21] Detailed information on mode, duration, frequency, and intensity can be recorded by the observer, providing rich contextual information for further evaluations. However, direct observation is associated with extensive researcher burden, as it requires a prolonged effort and the necessary focus to follow and track an individual's PA behavior. Therefore, direct observation is typically only used for short time intervals. [10,21]

1.1.2 Wearable Technologies

In recent years many researchers have turned to using wearable technologies (body-worn activity monitors such as pedometers and accelerometers) to capture the direct and objective bodily movements associated with PA behaviors.[21] Wearable technologies can be relatively low cost (pedometers $15-$30, accelerometers $175-$450) and provide estimates of PA directly expressed as step counts and/or time spent at different intensities.[22] Waist-worn pedometers incorporate simpler lever-based and battery-charged electronic circuitry-based technology that delivers a direct output of step counts detected from vertical oscillation of the hips during normal ambulation.[23] There are two main types of pedometers: spring-levered and piezoelectric.[23] Spring-levered pedometers rely on an internal, horizontally spring-suspended lever-arm that moves up and down with the movement produced from each step.[21] Specifically, the movement of the lever-arm opens and closes an electrical circuit, which is then registered as a step. Piezoelectric pedometers also respond to vertical accelerations that occur during ambulation, however, they generate and record electronic sine waves with each peak representing the force generated during a movement event, for example, a step.[22] The peaks of these waves are then internally converted into the total displayed step count.[22]

Accelerometers measure PA by recording the natural accelerations produced from the body's movement in gravitational units. Typically, these accelerations are sampled at frequencies of 40-100 Hz, meaning they can record 40 to 100 samples of acceleration data per second.[18] These electronic signals can then be converted into different time frames such as 1 second, 60 seconds, or one hour, depending on the desired resolution.[18] For example, if depicting PA behaviors on a minute to minute basis is desired, the accelerometer's electronic signals are processed in 60 second epochs (i.e., time intervals) to give acceleration-based PA data in per minute units. While pedometers are typically uniaxial (meaning they specifically only detect movement in the vertical plane), accelerometers can be biaxial or even triaxial meaning that they can simultaneously measure accelerations in multiple planes (anteroposterior, mediolateral, and vertical), for example, during unconventional movements like a cartwheel or contemporary dance.[24] Similar to pedometers, accelerometers were originally designed to be worn on the waist, however, increasingly, they can also be worn on the wrist or ankle too.[18] Using proprietary (i.e., specific to that model of accelerometer and owned by its manufacturer) algorithms, accelerations are typically distilled as unitless "activity counts," a metric that represents an implied combination of frequency and magnitude of the movement. Activity counts can then be used to estimate intensity levels (light, moderate, and vigorous), based on previously calibrated cutpoints. For example, in 2005, Freedson and colleagues[25] established that, specific to youth populations, frequencies 100- 2,200 activity counts·min^{-1} were associated with light PA, 2,200-4,135 activity counts·min^{-1} with moderate intensity PA, and ≥ 4,136 activity counts·min^{-1} with vigorous intensity PA. Since accelerometers are initialized in correspondence to a computer, they provide time-stamped data. This allows for a deeper analysis that can include PA patterns across various time points throughout the day. For example, time-stamped data can facilitate investigation into PA levels in the morning between 8:00-10:00 AM if this level of specificity is required. Further, the increased PA details that can be derived from accelerometers can be manipulated to provide information on frequency, duration, in addition to estimated intensity as described above.[15,23] Alternatively, the detected acceleration signal can be processed to output as step counts, and more specifically, stepping rates or cadence (i.e., steps·min^{-1}).[26]

Step-based metrics (steps·day^{-1}, steps·min^{-1}) extrapolated from wearable technologies allow for clear communication about PA that can be widely used and easily interpreted by researchers, clinicians, and consumers. While other measurements of PA, such as those described above, may provide additional details of PA (i.e., oxygen consumption from indirect calorimetry, type or domain of PA from self-report) relevant to specific research questions, their output may not be as easily understood by the general public for whom PA guidelines are ultimately intended. For example, a public health recommendation based upon quantifying oxygen consumption in reference to METs or energy expenditure assumes that members of the general public can understand and are personally capable of estimating and interpreting their own oxygen consumption levels necessary for accruing health benefits. On the other hand, step counts are more easily quantifiable without the use of overly expensive or complicated equipment. The simplistic and quantitative nature of step counts to measure PA provides a clear, easily interpretable measurement that can help relay comprehensible public health messages regarding PA behaviors.[22] Thus, it can be used as a simple yet useful way for people to quantify their PA, part of the larger movement noted as "the quantified self" where increasingly people are interested in gathering digitized data on various aspects of their daily life.[27]

1.1.3 Step Counting

A step is a fundamental mechanical unit underlying human bipedal locomotion, and at least for a healthy ambulatory population, humans naturally accumulate steps throughout the day as their basic mode of mobility and personal transportation. Step counting dates back to ancient Roman times when it served as a method for estimating distance traveled. In fact, the English word "mile" stems from the Latin origin "milia passuum" which means "one thousand paces" or two thousand steps.[22] Leonardo da Vinci is believed to be the first to design a mechanical step counter. This assertion was based on a drawing that depicted a lever-based mechanism that was intended to be tied to the thigh to capture the pendular movement generated during walking.[28] Subsequently, Thomas Jefferson used a lever-based step counter made from "one of the best watch makers in Paris" to measure the number of steps between landmarks.[22] Using this pedometer, Jefferson noted that there were 2066.5 steps in an English mile.[29] Contemporary step counters are frequently used to validate walking distance from PA questionnaires, conduct overall population PA surveillance, and promote PA with walking interventions.[30]

Steps counts can also be tracked and compared to emerging indices associated with positive health outcomes. For example, in 1965, the Japanese popularized 10,000 daily steps as an index associated with maintaining optimal health.[31] Recently, the PAGAC advocated for the implementation of more step-based research in regards to PA.[32] For example, the report claims that step counting technologies "are useful prescription tools for health care providers and trainers. Step counts blend well with public health messages encouraging the use of stairs rather than elevators, walking in airports rather than taking the train or shuttle, or parking at a distance from the final destination." [32] Although the potential for step counting was made clear in the PAGAC report, the committee also stated that more research on step counts and health outcomes such as all-cause mortality, cardiovascular disease and type 2 diabetes is needed before step-based guidelines can be clearly expressed.[32]

A 2007 systematic review by Bravata et al[33] of 26 pedometer-based interventions including 2,767 participants reported that those who increased their daily PA by >2,400 $steps \cdot day^{-1}$ realized an average decrease in systolic blood pressure of 3.8 mm Hg ($P<0.001$) and a BMI decrease of 0.4 $kg \cdot (m^2)^{-1}$($P = .03$). Richardson et al.[34] conducted a systematic review of nine studies representing 307 participants and reported that pedometer-based interventions were associated with weight loss (0.05 $kg \cdot week^{-1}$) and with greater effects evident with longer duration intervention. A separate review by Kang et al. [35] identified 32 studies including 2,570 participants using pedometers to increase PA levels. They concluded that pedometer-based interventions could increase daily PA by ≥2000 $steps \cdot day^{-1}$.[35] Additionally, Kang et al.[35] noted that this relationship was modulated by age, showing differences between older adults, adults, and youth.[35]

Specifically, youth displayed the strongest effect size (R^2=0.78), meaning that step counting via pedometer use can have a profound impact on increasing PA levels in youth. This was further documented in a systematic review by Lubans, Phillip, and Tudor-Locke[36] investigating the effectiveness of pedometers in promoting PA among youth. Of the 14 studies included in the review including 1,396 participants, 12 studies demonstrated increases in PA when using a pedometer.[36] Collectively, these review articles highlight that step counting is an effective way to positively impact PA levels and potentially realize a wide variety of health benefits, even in younger populations.

While steps·day^{-1} appears to be a feasible metric for quantifying overall volume of PA, research has shown that the stepping rate, or cadence (i.e., steps·min^{-1}) at which steps are accumulated can be used to interpret intensity.[37] A 2018 narrative review[38] focused on 'How fast is fast enough?' provided evidence to support cadence-based metrics as a proxy indicator for intensity. Specifically, it concluded that, at least for adults, cadences ≥ 100 steps·min^{-1} were strongly and consistently associated with absolutely defined moderate intensity PA (3 METs). A 2018 study by Tudor-Locke et al. [39] extended the research on this cadence-intensity relationship by studying 120 children (50% female) and young adults between the ages of 6-20 years of age. The sample included four males and four females from each age year who walked on a level treadmill for 5-minutes at 13.4-134.0 m·min^{-1} in 13.4 m·min^{-1} increments. In youth 6-8 years of age, moderate intensity was achieved on average at 128.4 steps·min^{-1}. This threshold gradually decreased as age increased. Specifically, 9-11 year olds reached moderate intensity at, on average, 116.5 steps·min^{-1}, 12-14 year olds at 106.6 steps·min^{-1}, 15-17 year olds at 101.3 steps·min^{-1}, and 18-20 year olds at 87.3 steps·min^{-1}. These findings illustrate that step-based research previously established in adult samples cannot be directly generalized to youth, as walking patterns and PA behaviors may differ over the course of the developmental lifespan.

1.1.4 Step Counting in Youth

As described above, the potential impact of step counting on PA-related health outcomes across the lifespan and including youth populations is notable. Prior to realizing this potential, however, there is a need to articulate and adopt standardized assessment methods for wearable technologies. Peake, Kerr, and Sullivan[40] conducted a critical review of wearable technologies and concluded that rigorous, standardized validation has been impeded by both the rapid pace of device development and an increasingly wide range of proprietary PA produced outputs including steps, heart rate, and various indicators of intensity. Herein, step counting is the focus as a feasible, interpretable metric that has recently been advocated by the PAGAC as a metric with great potential for communicating healthful PA doses.[32]

While the intended use of wearable technologies remains the same for people regardless of age, as indicated above, the generalizability of the results from previous step counter validation studies in adults cannot be translated directly to children.[41] As a further example, a 2011 study by Aloba et al.[42] identified that youth do not necessarily walk in a manner that is identical to adults' gait patterns. Using a motion capture system, the researchers found differences in walking patterns in youth (n=10, aged 5-9 years) and adults (n=10, aged 19-32 years) during walking and running conditions (i.e., walking in place, walking as fast as possible, running in place and running as fast as possible). More specifically, youth walk with significantly different step time, cycle time, cycle frequency, and cadence.[42] The researchers defined step time as the timed duration from when a foot lifts off the ground to when it subsequently touches the ground. Cycle time or stride time is the time between two consecutive steps, or the time it takes for one foot to complete a single cycle (two steps).[42] Further Aloba et al.[42] defined cycle frequency or stride frequency as the number of strides taken per unit time (i.e., cadence; steps·min^{-1}). Youth walked with an average step time of 0.33 seconds compared to 0.43 seconds for the adults. Similarly, cycle times were 1.05 seconds for youth and 1.26 seconds for adults, however, youth had higher cycle frequencies (1.15 seconds) compared to adults (0.90 seconds). Lastly, throughout the walking and running conditions of the study, youth had an average cadence of 203 steps·min^{-1} whereas adults had an average cadence of 166 steps·min^{-1}. Such spatial and temporal differences may limit the ability to directly translate the validation findings of step counting wearable technologies from adults to children.[43] It is important to note, however, that the sample size used for as comparison in that study is rather small, with arbitrary age ranges to define youth, and therefore may limit the ability to generalize these results across the youth age span.

1.1.5 Proper Assessment Methods for Validity

In the context of step counting, validity is defined as the degree to which the tested wearable technology's step counts compare to the criterion measure of directly observed steps. Standardized indicators of accuracy, precision, and bias are necessary to measure and convey validity and ultimately to facilitate comparability of different types of wearable technologies.

The Consumer Technology Association (CTA) released standards in 2016 to recommend certain performance levels to guide validity assessment of step counting wearable technologies.[44] The CTA suggested that under controlled laboratory conditions, researchers, device manufacturers, and consumers should use a hand tally counter to record the number of steps observed during treadmill ambulation, and then use the result as a criterion standard to determine and compare accuracies between different wearable technologies. Steps are best manually counted in real time and recorded using a video recorder for redundancy purposes and verification as needed to ensure the correct number of steps were indeed counted.[11]

Accuracy was defined by the CTA as the difference in number of steps recorded using the wearable technologies as compared to the directly observed steps. Precision is defined as the assessment of random error, and is commonly referred to as variance or variability.[45] Random error is that which is due to chance, or just naturally occurring. Measures of bias are useful for further assessing the level of agreement and systematic errors (over or underestimations) between wearable technologies and criterion measures.

1.1.6 Accuracy

The CTA recommended that the accuracy of wearable technologies be determined via mean absolute percentage error (MAPE) and optimal results should be ≤10% MAPE. MAPE is calculated as:

$$\text{MAPE}=\left(\frac{1}{n}\sum\left|\frac{Device-Criterion}{Criterion}\right|\right)\times 100$$

To illustrate, if a person walks 100 steps during a bout of walking, and their wearable technology records 95 steps, that device has a calculated MAPE of 5%. As MAPE moves closer to zero, confidence in device accuracy is increased. No rationale was presented for the CTA-recommended accuracy threshold for a MAPE of ≤10%. Further, there have been other proposed MAPE thresholds for step counting accuracy, such as <5%[46] or

<3%,[47] however, there is not yet enough evidence accumulated to support any of these indices as standardized accuracy thresholds to benchmark performance.

There are many strengths associated with using MAPE as a measure of accuracy. First, it is a simple index that employs a criterion measure of observed (true) steps. Further, reporting accuracy as a percentage scales the error to the total number of steps taken. For example, if a device detects 20 steps but the observed number of steps taken is really 25 steps, the difference is five steps and the MAPE is 25%. In comparison, if another device detects 95 steps out of 100 observed steps, the difference is five steps. However, the calculated MAPE is only 5%, indicating that the accuracy of the second device is better than the first one. Lastly, the absolute nature of the calculation eliminates the aggregate effects of over- or underestimation. For example, if in a single walking bout, a person accumulates 100 steps, but the wearable technology detects 115 steps, the result is interpreted as a 15-step overestimation. If in the subsequent bout, the same person walks 100 steps again, but the wearable technology records 85 steps, this result represents a 15-step underestimation. Without using the absolute value calculations of step differences, the average of these two walking bouts would erroneously suggest that this device had zero error (i.e., performed "perfectly"). However, by using the absolute value of the difference, or the MAPE, it is clear that there is an overall 15% error (in addition to the wildly fluctuating over and underestimations).

1.1.7 Precision

With respect to step counting wearable technologies, precision (i.e., random error) can be represented visually or digitally as the spread or distribution of the step counting error. For example, if during four different bouts of walking a specific wearable technology records -5 steps, +25 steps, +40 steps, and -35 steps relative to an observed criterion standard, the variance, distribution, or spread of the step count estimates appears to be random and not precise, as the step count estimates fluctuate across a range of 75 steps. The less distributed the error, the greater precision of the wearable technology.

In contrast to measures of accuracy or bias, precision does not directly compare the difference in step counts from wearable technology to directly observed steps. Rather, it is an assessment of natural variability of the difference between directly observed steps and those measured through wearable technology.[45] Common ways to assess and present precision include calculating standard deviation (SD), coefficients of variation (CoV), and correlation coefficients. With respect to step counting in wearable technologies, there is no consensus recommendation for an acceptable level of SD, CoV, or correlation coefficients to indicate sufficiently precise step counts.

Standard deviation is calculated as follows[45] :

$$SD = \sqrt{\frac{1}{n}\sum_{j-1}^{n}(E_j - E)^2}$$

Where E_j represents the step counts from the wearable technology, E is the directly observed steps, j is an individual sample, and n is the total sample.[45] The calculation of SD represents the variance, spread, or distribution of the step estimates obtained from the average difference of steps within each wearable technology across all participants. Since SD describes variance about a mean difference, it is a way of describing the precision of the data. For example, with MAPE described earlier, each MAPE value represents a mean value calculated from multiple data points (i.e., differences in step counts from a wearable technology across multiple participants). The calculation of the associated SD allows for evaluation of the precision of each wearable technology across the total sample. Larger SD values indicate less precision (more variance), and smaller SD values indicate more precise measures (less variance).

The CoV is similar to SD, just scaled and expressed as a percentage of the mean.[43] The formula for CoV can be defined as[45]:

$$CoV = \left(\frac{SD}{Mean}\right) * 100$$

Where SD is the variance of the step count from a wearable technology, and the mean represents the average step count from that device. For example, if a wearable technology has an SD of 10 steps within each trial, and the average step counts across multiple trials is 100 steps, the CoV is 10%. Lower CoV calculations indicate a more precise wearable technology step count measurement. Since it is an extension of SD, the inference derived from CoV is similar to that of SD, giving an estimate of variance from an averaged data point (i.e., MAPE of step counts from wearable technologies across multiple participants).

Lastly, calculating a correlation coefficient is another way to describe the precision of step counting wearable technologies by assessing their consistency when compared to directly observed steps.[43] This utilizes the XY scatter plot to depict the linear relationship between the independent (directly observed steps) and dependent (device-recorded steps) variables.[48] The variance, or spread of the data, illustrated in the scatter plot indicates the precision of the measures. A digital value for precision can be quantified by fitting a line to the plot which indicates the strength of the relationship between step counts determined by both wearable technologies and the criterion standard. Specific to step counting wearable technologies (where no negative relationship is realistically expected), the correlation coefficient is digitally represented as an *r* value ranging between 0 to +1, where r = +1 is interpreted as a strong, positive, tightly fitted, and linear relationship, and one where there is limited variance in the estimates.[48] To be clear, an correlation coefficient of +1 would indicate that the wearable technology consistently records the same number of steps that are directly observed. Values less than +1 indicate a relatively weaker relationship, and a correlation coefficient of 0 indicates no relationship at all, displaying a large variance in the data and indicating less precise measures.[48]

To illustrate the application of correlation coefficients in regards to step counting wearable technologies, Beets, Patton, and Edwards[43] calculated the index for four different wearable technologies at 40, 54, 67, 80, and 94 $m \cdot min^{-1}$ and reported a range from 0.225 to 0.999, with increasing *r* values at faster speeds. These findings indicate that at slower speeds, wearable technologies displayed reduced precision. However, at speeds 80 $m \cdot min^{-1}$ and greater, the technologies' *r* values improved to > 0.93, indicating greater precision.[43] In addition to the correlation coefficients calculated at each speed, 95% CIs were also reported. For example, for one wearable technology tested at the slowest speed of 40 $m \cdot min^{-1}$ the correlation coefficient was 0.460. In comparison, at the fastest speed of 94 $m \cdot min^{-1}$ the comparable values were 0.999 indicating a more precise (higher *r*) measure.

1.1.8 Bias

As previously mentioned, bias refers to ways of assessing systematic over- or underestimation of steps from the technologies when compared to the criterion measure of directly observed steps.[45] Overestimation refers to when devices record more steps than the actual number of steps taken, and underestimation refers to when devices record fewer steps than the actual amount taken. Common ways to assess measurement bias with respect to wearable technologies include calculating overall percentage error (PE) and drawing Bland-Altman plots. Similar to the calculation of MAPE, PE compares step counts from wearable technologies to the directly observed steps, and is reported as a percentage difference for scaling purposes. For example, Dueker, Gauderman, and McConnell[49] calculate PE with respect to bias of wearable devices as:

$$PE = \left(\frac{Wearable\ technology\ steps - Observed\ steps}{Observed\ steps}\right) * 100$$

By dividing the difference of the two measures by the observed (true) steps, the result is a scaled index that explains the difference, regardless of the total number of steps taken. As an illustrative example, if someone walks 100 steps but their wearable technology records 95 steps, the difference is five steps and the PE is -5%. Further, if someone else walks 1,000 steps but their wearable technology recorded 950 steps, there is a 50-step difference. This is seemingly larger than the 5-step difference from the first example, however, due to the scaling nature of PE (dividing by the observed steps), the PE is still -5%. The other benefit of PE is that it defines the direction of the error. For example, an underestimation of steps is indicated by a negative PE value, while a positive PE indicates an overestimation.

Bland-Altman plots are generally used to visualize data and interpret agreement between two methods. With respect to step counting, these plots represent the visually plotted agreement between the criterion and device-based estimates.[50] The plot itself is an XY scatter plot, where the X axis represents the average step counts detected by both direct observation and the wearable technology.[50] The Y axis then displays the difference between the two step counts. Each point on the scatter plot represents a single contrast between these two estimates.[50] Another benefit of Bland-Altman plots is that the graphical depiction also illustrates the 95% Limits of Agreement (LOA). This 95% limit of agreement is a range that represents results that are within ±1.96 standard deviations (described below) of the mean difference. Data points that fall outside of the 95% LOA represent outliers, or random error. However, Bland-Altman plots are graphical representations and therefore are not easily translatable when comparing studies. Digital (tabularized) representations of bias allow for more direct comparison for standardization purposes.

Lastly, a 95% confidence interval (CI) can also be calculated for group means (MAPE, correlation coefficients, PE)to assess the confidence in the estimate; that is, a determination that the finding is confidently reliable. A 95% CI provides a range of values, both above and below the single group average that contains the true value. The narrower the range, the more reliable the finding and therefore the greater the confidence in the precision of the estimate.

1.1.9 Threats to Validity

While variations in step counting performance are apparent when comparing devices across different models and manufacturers, there are certain external factors or threats to validity that may influence performance of wearable technologies. Among these, the two main sources of threats to validity are speed and wear location (e.g., wrist vs waist).[11] At slower walking speeds, each step may not produce enough force to be accurately recorded.[22] Tudor-Locke et al.[51] recommends that technologies should be sensitive enough to detect forces of ≥0.35 Gs (gravitational units), a threshold corresponding to the forces typically produced during normal ambulation. During slow ambulation (i.e., <53.6 m·min^{-1}), the vertical accelerations are less pronounced (<0.35Gs) during each step and therefore may not generate sufficient force to be detected as a step by all wearable technologies, resulting in an undercounting of steps.[52] The forces are not easily assessed, so speed is a reasonable proxy indicator of such force levels. Bassett et al.[22] suggests that most wearable technologies will record only 50-75% of the actual number of steps taken at 26.8 m·min^{-1}.

Wear location can further influence validity of step counting wearable technologies. Such devices were traditionally manufactured to be worn at the waist (belt, or waistband) to be most sensitive to the vertical accelerations produced during normal walking.[18] Commercial applications have more recently produced ankle, thigh, and wrist mounted technologies.

Ankle worn technologies capture steps taken by measuring the displacement of the lower limb created during ambulation. The Step Watch 3 is an ankle worn technology that detects step counts almost identical to those manually counted, producing a 99.7% accuracy even at treadmill speeds of 26.8 m·min^{-1}.[53] Ankle worn wearable technologies are closest to the foot, where there is the greatest level of displacement and force created during stepping This thereby increases the device's ability to capture each step taken, including those associated with slow movements, as the close proximity to the movement of the foot and eminent forces from foot striking during walking allow for more agreeable recordings of the actual number of steps taken. As previously discussed, slow walking speeds are known to diminish the validity of step counting estimates of waist worn technologies.

Currently, there is only one thigh mounted technology available for use, the activPAL monitor. Similar to waist mounted technologies, the activPAL responds to vertical accelerations produced during normal walking. At 26.8 $m \cdot min^{-1}$, the activPAL logged only a 3.5% underestimation of actual steps taken.[54] The decreased step counting error at slower speeds may be attributed to the fact that the technology is placed directly on the actively moving leg and therefore is more sensitive to even slow and low force ambulatory movements.

A wave of commercial wrist worn step counting wearable technologies are now available including offerings from Fitbit, Apple, Garmin, Polar, Samsung, ActiGraph, among others. From a scientific perspective, wrist worn devices were first trialed as a means of increasing overall compliance in the form of increased wear time.[55] Commercially, wrist worn devices also appeal to consumer interest in conscious consumption.[56]

The step counting abilities of wrist worn devices are similar to those worn at other locations in the sense that they are all intended to detect vertical accelerations created during walking (i.e., the up-and-down motion of the body during ambulation) and record "steps" based on the frequency and magnitude of these accelerations. Chen et al.[24] evaluated three commercially available wearable technologies in 30 participants (50% female; 21.5 ± 2.0 years of age) walking on a treadmill at 53.6, 80.5, 107.3, 134.1 $m \cdot min^{-1}$ for five minutes for each bout. Wrist worn technologies produced by a variety of different manufacturers displayed a range of 9.6-1.5% PE for detecting steps when compared to direct observation, with decreasing error evident during faster speeds. Further, when comparing waist versus wrist wear locations of technologies from the same manufacturer (ActiGraph), wrist worn technologies consistently recorded fewer steps than waist worn technologies across all speeds of walking on a treadmill.[57] In this 2015 study by Tudor-Locke, Barreira, and Schuna,[57] 15 participants (67% female; 27.5 ± 2.5 years of age) walked on a level treadmill across a range of speeds from 13.4-187.8 mph increasing in 13.4 $m \cdot min^{-1}$ increments. Across all walking speeds, wrist worn technologies produced outputs that were significantly different from directly observed steps, whereas waist worn technologies were only significantly different over a more limited speed range from 13.4-53.6 $m \cdot min^{-1}$.

Wrist worn technologies are subject to erroneously detecting non-stepping movements as steps throughout the day.[22] For example, when standing and folding laundry, devices may record "steps" from the movement produced, when in reality no steps were taken.[22] Examples of this, specific to the youth population, could be playing the drums, or folding laundry. The arms/wrists are in motion, producing forces that may lead to the technologies recording a "step" when no actual steps were taken Tudor-Locke, Barreira, and Schuna[57] also evaluated step counts from the same brand of wearable technology concurrently worn at the wrist and waist. They found that when compared across free-living conditions, wrist worn accelerometers recorded consistently higher average step counts (2500-8700 steps more) compared to waist worn monitors.[57] The difference is likely attributable to erroneously detected non-steps accumulated throughout the day.

1.2 Summary

As new objective wearable technologies continue to develop, standardized approaches to documenting validation parameters (i.e., measures of accuracy, precision, and bias) are needed to ensure confidence in step-defined PA. This validation effort should be extended to youth samples for improved objective activity monitoring methods and for a better understanding of PA and health related outcomes. To be clear, standardized communication about validation parameters would facilitate direct comparisons between research studies and populations and allow for clear communication about dose response relationships between PA (i.e., steps·day^{-1}, steps·min^{-1}) and associated health outcomes. New lines of wearable technologies are frequently developed and made commercially available. Yet validation studies are necessary if scientists and a range of practitioners are to optimally use such devices in research and/or clinical settings. Direct consumers also rely on their wearable technologies to provide them with dependable personal health-related data.[22]

1.3 Purpose of the Study

To conduct a secondary analyses of data previously collected with multiple wearable technologies (worn on arm/wrist, waist, thigh, and ankle).[36] The analyses will compute and compare age, wear location, and speed-specific measures of accuracy (MAPE), precision (correlation coefficient, SD, CoV), and bias (PE) across these multiple wearable technologies. The original data source is based upon 120 children (6-12 years), adolescents (13-17 years) and young adults (18-20 years) who wore several wearable technologies and completed a multiple speed treadmill protocol that also used a criterion standard of directly observed steps.[36] This effort provides a valuable source of reference material to better facilitate evaluation and comparisons of step counting wearable technologies.

1.4 Aims and Hypotheses

Aim 1 Catalog and evaluate digital indices of accuracy, precision and bias of steps for a variety of wearable technologies across a range of speeds.

H1: Accuracy, precision, and bias of steps will improve for all technologies as speed increases.

Aim 2: Catalog and evaluate digital indices of accuracy, precision and bias of steps for different wear locations (ankle, thigh. waist, and arm/wrist) of wearable technologies across a range of speeds.

H2: There will be differences in accuracy precision and bias of steps based on wear location (ankle>thigh>waist> arm/wrist*).*

Aim 3: Catalog and evaluate digital indices of accuracy, precision and bias of steps for children (6-12 years), adolescents (13-17 years) and adults (18-20 years) of wearable technologies across a range of speeds.

H3: There will be no difference in accuracy, precision, or bias of steps across each of the age categories.

CHAPTER 2

LITERATURE REVIEW

2.1. Overview

Wearable technologies are used widely in PA research to investigate behaviors across different populations, yet there have only been a few youth-specific studies that have evaluated the step counting features of these technologies. The purpose of this systematically focused literature review was to identify treadmill validation studies of step counting wearable technologies conducted with youth (specifically, those 6-18 years of age). This was done in an effort to catalog testing protocols (including treadmill speed, wear location of devices, and age group) and specifically how researchers have analyzed and reported accuracy, precision and/or bias of step counting wearable technologies as compared to directly observed steps.

2.2 Objective Activity Monitoring in Youth

This literature review was updated on PubMed as of December 13, 2019. After trialing and refining numerous arrangements, a final Boolean search string [(steps OR "steps per minute*" OR "steps/min*") AND (treadmill) AND (child* OR youth OR adoles*) AND (valid* OR reliab*)] was used to specify the most relevant articles to the overall purpose as stated above. Additionally, filters of timespan (inception or earliest PubMed records identified up to present times), species ("human") and language ("English") were applied. This electronic search strategy produced 42 individual original research articles, ranging in publication year from 1977-2019. Titles and abstracts were subsequently screened to confirm inclusion of youth samples (6-18 years of age) and wearable technology validation processes that included step counting. Eliminated articles at this stage of the review included six focused only on adult samples and 18 that did not consider step counting amongst evaluated outputs. The full texts of the 18 remaining articles were subsequently reviewed for inclusion criteria based on: 1) including a sample of youth (6-18 years of age), 2) reporting step counts as a metric of interest, 3) using directly observed (or video recorded) steps as a validation criterion measure, and 4) including bouts of walking or running on a level treadmill. Of these 18 articles, four were excluded because they did not detect step counts, six because they focused on populations outside of the intended age range, and one for not using a treadmill-based protocol. The reference sections of the remaining six articles were reviewed and a single additional article was identified that met all inclusion criteria. Ultimately, seven original, independent research articles were included in the review that follows.

As per search design, the purpose of each identified study was similar in nature, namely, to investigate the validity, and more specifically, the accuracy, precision, and/or bias of wearable technologies during treadmill walking in youth samples. Table 1 presents abstracted information including: 1) first author and year of publication, 2) sample characteristics (sample size, % female, age range or other age descriptive information as provided), 3) protocol characteristics (number, duration, and speed of bouts), 4) wearable technologies used, and 5) statistical methods and results used to evaluate and report accuracy, precision and/or bias. For ease of comparison, all descriptive numerical values have been rounded to a single decimal point and all speed values have been converted to metric measures (specifically, $m \cdot min^{-1}$). Any apparent reporting discrepancies between studies represent original differences that could not be otherwise reconciled.

2.2.1 Pedometers

Beets, Patton, and Edwards[43] evaluated step counting accuracy of four pedometers in 20 children (50% female) 5-11 years of age. All devices tested were worn at the waist and included the Walk4Life 2505 (WL; Plainfield, IL), Digiwalker SW-200 (DW200; Yamax Corp., Japan), Sun TrekLINQ (SUN; Arvada, CO) and Digiwalker SW-701 (DW701; Yamax Corp., Japan). The treadmill protocol consisted of five 2-minute bouts at 40, 54, 67, 80, and 94 $m \cdot min^{-1}$. Intraclass correlation coefficients (ICC) and MAPE were calculated for each device at each speed. There was a similar trend observed across all technologies where ICCs improved with increasing speeds, ranging from 0.225 (SUN) at the slowest speed to 0.999 (DW200) at the fastest speed. As mentioned in Chapter 1, ICC values closer to one represent a stronger linear relationship with directly observed steps, and therefore are considered to be indicators of a more precise technology. All wearable technologies assessed displayed ICC >0.9 for speeds ≥80 $m \cdot min^{-1}$.

The MAPE values were within 5% error for three of the devices at speeds ≥67 $m \cdot min^{-1}$ while the SUN model only reached this threshold at speeds ≥80 $m \cdot min^{-1}$. This result indicated that the step counts from the tested wearable technologies were within 5% of the directly observed steps. Although not directly quantified, as speed increased to 67-94 $m \cdot min^{-1}$ the MAPE decreased, indicating increasing accuracy of step estimates. At the slowest walking speed, all technologies were greater than 20% MAPE. At the fastest speed however, MAPE improved to within 5% of directly observed steps for all technologies. As previously indicated, at the slower speeds (40 and 54 $m \cdot min^{-1}$), the lower ICCs and higher MAPE values apparent for the tested technology indicated greater error in step measurement.

A similar validation study was conducted by Trapp et al.[58] of the Accusplit AH120M9 (Accusplit, Inc. Livermore, CA, USA) with built in memory, and the Yamax SW-700, both of which are a spring-levered pedometers. Researchers evaluated the step counting accuracy compared to directly observed steps using MAPE in 45 children (51.1% female, 10.7 ± 0.8 years) at slow (40m·min^{-1}), moderate (67 m·min^{-1}) and fast (91 m·min^{-1}) walking. Participants were outfitted with both devices worn on the right hip and walked on a treadmill at the three set speeds for three minutes each bout. The Accusplit displayed a MAPE of 46.9%, 15.9 % and 8.6% for the increasing speeds respectively. Further, the Yamax SW-700 showed a similar trend of decreasing MAPE values 44.1%, 14.0%, and 8.9%, respectively, with increasing speeds. Both devices undercounted step counts, especially at the slowest speed (40 m·min^{-1}), indicated by the relatively larger MAPE values (44.1-46.9%). Although MAPE is absolute in nature, meaning it is only expressed as a positive value, the evident magnitude of the error at slower speeds was due to undercounting of steps from the wearable technologies compared to directly observed steps. The Accusplit undercounted by an average of 124.4±52.2, 52.0±43.9, and 29.7±45.5 steps across the three speeds, respectively. Similarly, the Yamax SW-200 undercounted by an average of 116.3±53.4, 44.58±38.8, and 31.2±46.3 steps, across the slowest to fastest speeds, respectively.

Trapp et al.[58] also used Bland-Altman plots and 95% LOA to evaluate whether step count estimates from both devices agreed with directly observed steps at 40 m·min^{-1}, 67 m·min^{-1}, 91 m·min^{-1}. At the slowest speed of 40 m·min^{-1}, both technologies were not in agreement with directly observed steps, as no difference (i.e., 0 steps) fell outside of the 95% LOA, whereas at 67 m·min^{-1} and 91 m·min^{-1}, both technologies were considered in agreement based on their results being contained within the bands designated by the 95% LOA.

Ramírez-Marrero et al.[59] also investigated the Digiwalker SW-200 accuracy in treadmill walking with a sample of 31 African-American children (45.2% female, age=8.8 ± 1.4 years). The participants walked at three different speeds (59 $m \cdot min^{-1}$, 70 $m \cdot min^{-1}$, and 91 $m \cdot min^{-1}$) for two minutes each bout. Bouts were repeated 3 times at each speed, for a total of nine walking trials, and all were performed on the same day. Participants concurrently wore the two pedometers clipped to the waistline of their shorts, one on the midline of their thigh, and the second at the midline of their torso. Step counts from each wear location and each bout were compared to directly observed steps. The SW-200 at the midline of the thigh captured 85% ($p \leq 0.05$), 95%, and 99% of the directly observed steps for each of the respective speeds (slowest to fastest). The SW-200 at the midline of the torso captured 102%, 105%, and 102% of the directly observed steps at each of the three speeds, respectively. Consistently, the SW-200 overestimated the number of steps from each walking bout compared to steps that were directly observed. Additionally, when the same pedometer was worn at the midline of the thigh it consistently underreported steps, although it did improve accuracy with increased speeds. However, at slower speeds (59 $m \cdot min^{-1}$), the SW-200 significantly undercounted the actual number of steps taken when worn at the manufacturer's suggested wear location. This study replicates similar observed trends in increased accuracy of step counting wearable technologies with increases in walking speed.

Mitre et al.[60] investigated the accuracy of two commonly used pedometers (the Omron HJ-105 (Kyoto, Japan) and the Yamax Digi-Walker SW-200) at four different 5-

minute walking treadmill speeds (13.4 m·min^{-1} 26.8 m·min^{-1} 40.2 m·min^{-1} and 53.6 m·min^{-1}) with 27 children (51.9% female; age = 11 ± 1 years), categorized as normal weight (n=16, 62.5% female, BMI percentile=41±23%) and obese (n=11, 36.34% female, BMI percentile=92±4%) children.[60] Across all speeds, both pedometers displayed error compared to directly observed steps (measured using PE) that improved with increasing speeds. Specifically, at 13.4 m·min^{-1}, both pedometers displayed almost -100% error (i.e., they did not record any steps at this slowest speed), improving only to approximately -60% PE at 53.6 m·min^{-1}. This study also replicates similar trends shown by Ramírez-Marrero et al.[59] of increased accuracy from wearable technologies, including SW-200, with increases in walking speed.

Dueker, Gauderman, and McConnell[49] used PE and MAPE to describe step counting validity of a piezoelectric pedometer, the SportBrain iStep X1 (manufacturer information not reported), and a spring-levered pedometer, the Digiwalker SW-701. The study sample included 17 children (58.8% female, age =12.8 ± 1.7 years). Participants were outfitted with the SportBrain iStep X1 and Digiwalker SW-701 pedometer on the right waist and walked or ran on the treadmill at 53.6 m·min^{-1}, 80.5 m·min^{-1}, 107.3 m·min^{-1}, and 134.1 m·min^{-1} for 3-minute bouts. Steps were directly observed for the criterion measure, and PE and MAPE were calculated for each bout. Observed step count errors were almost entirely due to undercounting, leading to a negative PE, but positive MAPE. As described in Chapter 1, PE indicates the direction of the error and therefore a negative PE corresponds to an undercounting of steps when compared to directly observed steps. On the other hand, MAPE represents absolute errors, specifically, the magnitude of the error, but not the direction. Therefore, MAPE is always positive and never negative. At each of the four increasing speeds, the SportBrain displayed a -1.3%, -0.5%, -0.2% and -3.9% PE respectively. Concurrently, the DigiWalker displayed a -19.3%, -8.3%, -11.8%, and -4.3% PE at each of the four speeds respectively. These results indicate that at all speeds, the piezoelectric SportBrain iStep X1 showed a reduced PE compared to the spring-levered pedometer, the Digiwalker SW-701 pedometer.

Aminian and Hinckson[61] investigated the validity of the Yamax DigiWalker SW-200 and NL-2000 (NewLifestyles Inc., Lees Summit Missouri) measuring ambulatory activity in children. The study included a sample of 25 children (68% female; age= 9.9 ± 0.3 years) who were outfitted with the SW-200 and NL-2000 attached on the right and left side of the waistband, respectively. They walked on a treadmill for 2-minute bouts at 50 $m \cdot min^{-1}$, 66 $m \cdot min^{-1}$, 93 $m \cdot min^{-1}$, and 133 $m \cdot min^{-1}$. PE values were used to assess the bias of the device step count measures at slow (50 $m \cdot min^{-1}$) and fast (93 $m \cdot min^{-1}$) walking, as well as running (133 $m \cdot min^{-1}$). The SW-200 underreported steps (-4%) during slow walking, yet over reported steps during fast walking and running (1% and 2%, respectively). The NL-2000 reported a similar pattern, underestimating by -11% at slow walking and over estimating by 2% and 1% during fast walking and running, respectively.

Rosenkranz, Rosenkranz, and Weber[62] also assessed the Yamax DigiWalker SW-200 step counting function in 19 children (36.8% female; age=9.6 ± 1.2 years). Participants completed four 3-minute bouts on a treadmill at 40.2 $m \cdot min^{-1}$, 80.5 $m \cdot min^{-1}$, 120.7 $m \cdot min^{-1}$, and 161 $m \cdot min^{-1}$. There was no rest between speeds as these were continuously increased after each 3-minute bout. As a result, PE was used to assess bias across the entire walking protocol (averaged across bouts). The SW-200 undercounted by 4% during the full treadmill activity.

The pedometer studies reviewed above were purposely limited (via selected search strategies) to controlled laboratory (i.e., treadmill) settings that also used direct observation of steps taken as the criterion standard. Seven identified studies focused on a limited variety of pedometers (Walk4Life 2505, Digiwalker SW-200, Sun TrekLINQ, Digiwalker SW-701, Accusplit AH120M9, Yamax SW-700, Omron HJ-105, and SportBrain iStep X1) and tested small sample sizes (ranging from 17-45 participants) representing an age range of 5-17 years. Treadmill protocols varied in number of bouts (three to five bouts) and their duration (two minutes to five minutes), whether rest was provided (six studies) or not (one study), and whether (six studies) or not (one study) they included both walking and running bouts. Four studies included measures of accuracy (specifically, MAPE), four included measures of precision (ICC), and three included measures of bias (PE and Bland-Altman Plots). None included all three. Where reported, MAPE values ranged from 46.9% to 0.2%. PE ranged from -100% to -0.2%. ICC values ranged from 0.225 to 0.999. SD values were reported for the group means of MAPE, and PE when reported. CoVs were never reported.

2.2.2 Accelerometers

Aminian and Hinckson[61] also incorporated an ActivPAL, a thigh worn accelerometer, into the previously described protocol. ICCs were reported for all conditions. The authors reported a perfect correlation (ICC=1) between the ActivPAL and the directly observed steps in slow (50 $m \cdot min^{-1}$) and normal (66 $m \cdot min^{-1}$) paced walking. However, relatively low correlations (ICC=0.21 and ICC=0.34) between device and criterion were observed with fast walking (93 m·min^{-1}) and running (133 m·min^{-1}), respectively. Further, PE values were calculated to assess the bias of tested technologies at fast (93 m·min^{-1}) walking, as well as running (133 m·min^{-1}). The ActivPAL overestimated steps by 8% and 26% during both of these conditions, respectively. Further, although not directly quantified, the authors stated that the ActivPAL performed accurately during slow (50 m·min^{-1}) and normal (66 m·min^{-1}) walking speeds.

Rosenkranz, Rosenkranz, and Weber[62] also evaluated the validity of the Actical (Respironics Inc. Murrysville, PA) accelerometer in the same treadmill protocol described above that evaluated the accuracy of the Yamax DigiWalker SW-200 in 19 children (36.8% female; age=9.6 ± 1.2 years). Two accelerometers were initialized with 30 second epochs. One was worn on the waist and one on the ankle. The waist worn Actical undercounted steps (PE= -11%) across the full range of treadmill speeds. Additionally, the ankle worn Actical undercounted steps by 17% across all treadmill speeds. Further, both placements correlated (ICC) strongly with directly observed steps. Specifically, across all speeds, the waist worn Acticals displayed an ICC=0.927 while the ankle worn Acticals demonstrated an ICC= 0.854. Further, when presented as cadence instead of total accumulated steps, the authors reported less agreement (although not specifically quantified) between waist-worn Acticals and directly observed step at lower cadences. This finding was illustrated via Bland-Altman plots and 95% LOA. The plots showed that at a cadence ≤100 steps·min^{-1} there was a greater range of differences in step counts from waist worn Acticals as compared to directly observed steps. However, at cadence >100 steps·min^{-1} the range in differences of step counts decreased. This was indicated by a wider LOA at lower cadences that narrowed as cadence increased. This observed trend was reversed with ankle worn devices. Specifically, the ankle worn Acticals displayed a narrower LOA with directly observed steps at cadences ≤ 140 steps·min^{-1} and a wider LOA at higher cadences (>140 steps·min^{-1}). To be very clear, the Actical was biased (displayed greater systematic error) at higher cadences.

Similar to the pedometer studies, the accelerometer studies reviewed herein were purposely limited to those selected using search strategies limited to controlled laboratory (i.e., treadmill) settings. Two identified studies focused on a small number of wearable technologies (ActivPAL and Actical) and tested with limited sample sizes (ranging from 19 to 25 participants) and representing narrow age ranges (from seven to 11 years of age). The treadmill protocols each included four speeds ranging from 40.2 m·min^{-1} to 160.9 m·min^{-1}. The duration of bouts were two minutes and three minutes for each study, respectively. Both studies also included walking and running bouts. Neither study included measures of accuracy, however both included measures of precision (ICCs), and only one included an indicator of bias, and only a graphical one (i.e., a Bland-Altman plot). Where reported, ICC values ranged from 0.21 to 1.0. MAPE, PE, SD, and CoVs were not reported.

2.2.3 Summary of Literature Review

This literature review identified that study protocols tested wearable technologies across a broad range of speeds, and, for the most part, documented improved accuracy and precision as speeds increased. However, bias did not always display a linear improvement with increased speed as there were instances where bias was increased from slow walking to moderate walking, and then decreased into fast walking, and running. Further, one study showed greater bias at increased cadences, indicating a technology-specific inconsistent relationship with speed. This literature review also identified a number of information gaps and opportunities for further study. It is apparent that a larger sample size and a more comprehensively aged youth sample is needed to validate step counting wearable technologies across the developmental age span. A greater number of wearable technologies should be concurrently evaluated against the criterion standard and these should also represent diverse wear locations, specifically, ankle, thigh, waist, and wrist. Lastly, there is a need for a standardized validation approach to analyses that includes consistent and digital expressions of accuracy, precision, and bias. Although the CTA has suggested that step counting wearable technologies should produce a MAPE of <10%,[41] this guidance is not empirically guided and there are few other reference data available to guide interpretation of these validity parameters.

The analysis leverages the present knowledge base and enhances the field by providing standardized digital validation parameters (accuracy, precision, and bias) based upon a large sample of youth to young adults (6-20 years of age) who wore multiple wearable technologies attached at different wear locations and were tested across a broad range of treadmill speeds (0.5-5.0 mph). Together, these missing pieces of information will be combined into a highly useful catalog of reference values that could be used to set empirically-based performance expectations for step counting wearable technologies.

CHAPTER 3

METHODS

3.1 Computing and Comparing Wearable Technology Step Counts in Youth

This study represented a secondary analysis of the NIH funded R21 CADENCE-Kids: Cadence and intensity in Children and Adolescents dataset.[39] The primary aim for the original study was to establish cadence values associated with absolutely defined moderate intensity.[39]

The purpose of this study was to conduct a secondary analysis of these data previously collected with multiple wearable technologies (worn on ankle, thigh, waist, wrist).[36] This secondary analysis computed and compared speed, wear location, and age-specific measures of accuracy (MAPE), precision (correlation coefficients, SD, CoV), and bias (PE) across multiple step counting wearable technologies. The data source is based upon 120 children (6-12 years), adolescents (13-17 years) and young adults (18-20 years) who wore several wearable technologies and completed a multiple speed treadmill protocol that also used a criterion standard of directly observed steps.[36] This secondary analysis considers only walking bouts. Only 8.5% of total bouts (89/1,042 bouts) consisted of running, occurring individually-based and thus during different terminal bouts across the range of tested treadmill speeds. These running bouts were distributed across different speeds as follows: 80.5 $m\cdot min^{-1}$ (n=3), 93.9 $m\cdot min^{-1}$ (n=6), 107.2 $m\cdot min^{-1}$ (n=20), 120.7 $m\cdot min^{-1}$ (n=41), 134.1 (n=19). Due to the difference in gait patterns associated with running vs. walking, and overall lack of a robust sample size with running at standardized speeds, running bouts were excluded.

Participants wore nine step counting wearable technologies across multiple wear locations (ankle, thigh, waist, wrist and upper arm). Although the GENEActiv (Activinsights, Cambs, UK) wearable technology was used in the original study, there are no currently published and validated algorithms for generating step outputs. For this reason, the data from the GENEActiv was omitted. The remaining eight step counting wearable technologies (and their respective wear locations) evaluated in these secondary analyses are presented in the Tables 4 through 14.

3.2 Participants

Recruitment and data collection were originally designed to be evenly distributed across sexes (60 boys, 60 girls) and ages (four boys and four girls for each age year). Additionally, the sample was designed to be heterogeneous across race/ethnicity and BMI-defined obesity classifications. All study procedures were reviewed and approved by the Pennington Biomedical Institutional Review Board, where the original data collection took place. Approval was also permitted by the University of Massachusetts Amherst Institutional Review Board for secondary analyses of de-identified data.

In accordance with the ambulatory nature of the original study, exclusion criteria for participation included use of wheelchairs and other assistive walking devices and all other impairments that could affect normal ambulation. Other exclusion criteria included hospitalization for mental illness within five years of enrollment, any condition or medication that would affect the heart rate response to exercise testing, pregnancy, and presence of a pacemaker or any other implanted medical device, including the use of metal joint replacements.

3.3 Protocol

This secondary analysis focused on the data collected specifically during treadmill walking, a portion of the larger CADENCE-Kids study. Participants completed a series of five-minute walking bouts on a level (0% grade) treadmill, which started at 13.4 $m \cdot min^{-1}$ and increased by 13.4 $m \cdot min^{-1}$ increments up to 134.1 $m \cdot min^{-1}$. Each walking bout was separated by at least two minutes of standing rest to facilitate determination of bout-specific step counts. Directly observed steps were obtained using a hand tally counter for each bout, and then averaged across the total sample for each respective speed.

Some of the devices (ActiGraph, ActivPAL, Actical, StepWatch, and SenseWear) do not display immediate step count feedback, however, data were time stamped and downloaded following the protocol so that step data could be harvested. Specifically, the time stamped step count data were matched up to the exact digital timing of the protocol to inform records of bout-specific step counts for these specific wearable technologies. The NL-1000 and Yamax SW200 provided immediately available step count data for each bout. These data were recorded during scheduled rest periods between bouts. The protocol was terminated following the bout when the participant naturally chose to jog or run, or if they decided not to continue at any time point.

3.4 Measures

3.4.1 Participant Characteristics

Participant's biological sex, race/ethnicity, and age by year were self-reported. A series of anthropometric measurements that included standing height, seated height, weight, and BMI calculations were then obtained. Standing and seated height were recorded to the closest 0.1 cm using a wall-mounted stadiometer (Harpenden model;

Holtain Ltd., Crosswell, Crymych, Pembrokeshire, UK). Two measures were taken for each and averaged. A third measure was taken if the difference between the first two was >0.5 cm apart. The seated height was measured using a set stool height, where participants would sit with their back against the stadiometer and with their legs hanging unweighted. This procedure allowed for a calculation of the torso height by subtracting the set stool height from the total measurement of the seated height. Leg length was subsequently determined by subtracting seated height from standing height. Participant's weight was calculated (without socks and shoes) using a digital scale (Tanita SC-240; Tanita corporation, Tokyo, Japan). Again, this measure was taken twice and averaged, and a third measurement was performed if the difference between the two measurements was >0.5 kg. BMI was calculated using the standard equation of taking the measured weight and dividing it by the squared value of height (kg/m^2).

3.4.2 Treadmill-based Variables

Locomotor steps were directly observed during the treadmill protocol and recorded manually using a hand tally counter. A video camera was also directed at the participant's feet and recorded movements for redundancy purposes (i.e., to ensure the correct number of steps were recorded). The speed ($m \cdot min^{-1}$) of each bout was determined using the treadmill's digital output which was continuously affirmed using a tachometer.

3.5 Statistical Analysis

Statistical analysis was calculated using R-Studio (version 3.0.2, R Foundation for Statistical Computing, Vienna, Austria) and Microsoft Excel (2010). Statistical significance was set at $\alpha \leq 0.05$.

3.5.1 Descriptive Statistics

Sex of participants was presented as frequencies (% girls and boys). Race/ethnicity of participants was also reported as frequencies (% African-American, Caucasian, or other). The distribution of continuous data including age, height, leg length, weight, and BMI was first plotted to assess normality and then presented as means±SD, if appropriate.

Each treadmill bout was defined by the 5-minute period at each respective treadmill speed (13.4 $m \cdot min^{-1}$ to 134.1 $m \cdot min^{-1}$). The treadmill-based variables consisted of step counts averaged for each wearable technology at each respective speed. For example, each wearable technology had step count data available for 13.4, 26.8, 40.2…134.1 $m \cdot min^{-1}$ for each participant, where applicable. To be clear, not every participant reached the highest speed of 134.1 $m \cdot min^{-1}$ (due to the above-mentioned reasons), therefore the number of participants with available data diminished as speed increases. Also as mentioned above, there was a need to identify and separate data from those individuals who walked vs. ran during their final bout as the gait patterns are not the same, and therefore running bouts were removed from the analysis.

The accuracy (MAPE) and bias (PE) values for each wearable technology (including wear location information) are presented as averages with associated SD with CoV precision values across the sample relative to each walking speed, and age group (defined above). Correlation coefficients (*r*; precision) were averaged across the sample and are reported as a single value across total bouts, as well as slow, normal, and fast walking speed bouts.

3.5.2 Inferential Analysis

The inferential analyses are based on the difference between wearable technologies and directly observed steps. The inferential analyses used to test each Aims' associated hypothesis included building several regression models with repeated measures and their corresponding 95% CIs to test for significance. Respective to each Aim, factors (terms of each regression model; what was being evaluated) were assigned to the regression model to test for significance as described in each hypothesis (repeated below for clarity purposes). From there, the regression models were used to calculate significance (α) values for each factor. Additionally, 95% CIs allowed for comparison between factors. If 95% CIs overlapped (meaning that part of the range of values in one factor's CI overlapped with that of another factor's 95% CI) then it was assumed that they are not significantly different. However, if CIs did not overlap, then it was assumed that the two factors were significantly different from each other. The framing of this approach respective to each Aim is presented in more detail below. Because the regression models were based on group means, SD and CoV were purposely left out of the inferential analyses.

Aim 1 Catalog and evaluate digital indices of accuracy, precision and bias of steps for a variety of wearable technologies across a range of speeds.

H1: Accuracy, precision, and bias of steps will improve for all technologies as speed increases.

To address this hypothesis, accuracy, precision, and bias specific catalogs depicting technology performance across the range of walking speeds were created. Additionally, eight categories of regression models (for the eight wearable technologies included) with repeated measures and 95% CIs were built. Based on the cataloging of validation indices for each wearable technology there was also three models (one for MAPE, one for correlation coefficients, and one for PE) for each wearable technology, where the models' factor was speed (across the 10 increasing speeds from 13.4 to 134.1 m·min^{-1}). In total, 17 regression models were built. These represent the eight regression models respective to each of wearable technologies for MAPE and PE, and one regression model for correlation coefficients averaged across the tested wearable technologies at slow, normal, and fast walking). To be clear, it was anticipated that the three identified validity parameters for each tested wearable technology would significantly improve with increased speed.

Aim 2: Catalog and evaluate digital indices of accuracy, precision and bias of steps for different wear locations (ankle, thigh, waist, and arm/wrist) of wearable technologies across slow, normal, and fast walking speeds.

H2: There will be differences in accuracy precision and bias of steps based on wear location (ankle>thigh>waist> arm/wrist).

Indices of accuracy, precision, and bias were summarized and compared for each of the four wear locations (arm/wrist, waist, thigh, and ankle). The wear location group means were averaged across the four wear locations (arm/wrist, waist, thigh, and ankle), and reported respective to each bout. To address this Aim, wear location specific catalogs of accuracy, precision and bias, depicting technology performance, were created. Further, six regression models with repeated measures were built for MAPE and PE at slow (13.4, 26.8, 40.2 and 53.6 m·min^{-1}), normal (67.0, 80.5, 93.9, 107.2 m·min^{-1}), and fast (120.7, 134.1 m·min^{-1}) speeds (two validity parameters x three speed categories) where wear location was the factor in each of the models. For correlation coefficients, one model was built comparing wear locations specific correlations averaged across total walking bouts. It was anticipated that there would be a difference in accuracy, precision, and bias across slow, normal, and fast walking speeds based on wear location.

Aim 3: Catalog and evaluate digital indices of accuracy, precision and bias of steps for children (6-12 years), adolescents (13-17 years) and adults (18-20 years) of wearable technologies across slow, normal, and fast walking speeds.

H3: There will be no difference in accuracy, precision, or bias of steps across each of the age categories.

Again, accuracy, precision, and bias catalogs of device performance in regards to specific age groups were created. Regression models with repeated measures for accuracy, precision, and bias were built for children (6-12 years), adolescents (13-17 years), and young adults (18-20 years) at slow, normal, and fast walking speeds. The group means were averaged across all wearable technologies for children, adolescents, and young adults and reported respective to each bout. Again, six models were built for MAPE and PE (two validity parameters x three speed categories) to facilitate direct comparison of the accuracy and precision between each of the three age groups, represented as the model factors, across the three categories of speed. For correlation coefficients, one model was built comparing age group specific correlations averaged across total walking bouts. To be clear, it was predicted that there would be no difference in accuracy, precision, or bias between age groups for any of the three speed categories.

CHAPTER 4

RESULTS

4.1 Participant characteristics

The analytic sample consisted of 121 participants (49.6% female, age=13.0±4.2 years). The mean (±SD) values for measured height, leg length and weight were 155.3 ± 16.9 cm, 73.8 ± 9.7 cm and 55.8 ± 22.2 kg, respectively. The calculated BMI was 22.4 ± 6.3 kg/m^2. Race/ethnicity was comprised of 35.5% African-American, 62.0% Caucasian, and 2.5% Other. The details of these characteristics specific to children (n=56), adolescents (n=41) and young adults (n=24) are presented in Table 2.

As described in the Methods section, the number (n) of participants differed across bouts due to differential testing termination decisions. Additionally, due to occasional malfunction of the different technologies used, step-based data were only available for 117 people for at least one bout. Further, as speed increased, the number of participants completing each successive bout decreased from a high of 117 participants (who completed the initial 13.4 m·min^{-1} bout) to a low of 16 participants (who completed the final possible 134.1 m·min^{-1} bout). Of note, no children (i.e., those 6-12 years of age) completed the final possible bout of walking at 134.1 m·min-1. The full accounting of participant numbers for each walking bout are detailed in Table 3, and further described by age group in Tables 4-6

4.2 Accuracy: MAPE

4.2.1 MAPE by speed

With respect to Aim 1, the MAPE±SD (with corresponding CoV values) for each of the wearable technologies tested across the speed range for the total sample and by children, adolescent, and young adult age categories are detailed in Tables 3-6, respectively. Additionally, the MAPE and SD across the full range of speeds are presented visually in Figures 1 and 2. The MAPE values across all wearable technologies ranged between 98.7%±1.5% (Sensewear at 13.4 $m \cdot min^{-1}$) to 1.1%±1.8% (Actical at 93.9 $m \cdot min^{-1}$). Those worn at the arm/wrist displayed greater MAPE than those worn at the waist, thigh, or ankle across all speeds, with differences ranging between 79.3% to 47.2% error. Additionally, on average, the MAPE was greatest across the slow walking speeds (50.1±35.5%), compared to normal (15.9±21.7%) and fast (18.9±22.2%) walking speeds.

For the inferential analyses respective to Aim 1, the regression models built for each of the eight tested wearable technologies indicated that there was a significant effect of speed ($p<0.0001$) on MAPE values. The details of this are illustrated in Figure 18. Further, this figure illustrates that the significance is driven by improvements in MAPE during slower speeds (13.4 $m \cdot min^{-1}$ to 53.6 $m \cdot min^{-1}$). This is evident from the overall reduction in MAPE and lack of overlap of the 95% CI relevant to the mean MAPE values across these speeds. For example, MAPE values were significantly greater at 13.4 $m \cdot min^{-1}$ than they were at 26.8 $m \cdot min^{-1}$ for seven of the eight (87.5%) tested wearable technologies, with differences between contiguous bouts ranging from a 50.6% difference (ActivPAL) to a 1.9% difference (Sensewear; the only non-significant difference) indicating significantly more accurate performance at the faster speed. However, across the normal and fast walking speeds (67.0-134.1 $m \cdot min^{-1}$), MAPE values leveled out, as there were only four instances (of 40 total cases; 10%) of significant differences evident between contiguous bouts across the tested wearable technologies.

4.2.2 MAPE by wear location

With respect to Aim 2, MAPE relative to wear location (presented in Table 12) ranged from a high of 93.8% (arm/wrist at 13.4 $m \cdot min^{-1}$) to a low of 2.6% (waist at 120.7 and 134 $m \cdot min^{-1}$). MAPE values ranged across the four wear locations (respective to each speed) from 79.3% to 47.2%. Consistently across each speed, the arm/wrist worn technologies displayed the largest MAPE values. The SD for these wear location-respective MAPE values ranged from a high of ±37.1% (observed at the arm/wrist at 67.0 $m \cdot min^{-1}$) to a low of ±6.1% (observed at the waist at 134.1 $m \cdot min^{-1}$).

The regression models built specific to Aim 2 indicated that wear location had a significant effect on the MAPE values associated with measuring steps ($p<0.0001$) across slow, normal, and fast walking speeds. MAPE values ranged from 75.8% (arm/wrist) to 6.7% (ankle) at slow walking speeds. At normal walking speeds, MAPE values ranged from 51.6% (arm/wrist) to 4.1% (thigh). At fast walking speeds, the MAPE values ranged from 54.6% (arm/wrist) to 2.8% (hip). Significant differences between MAPE values across wear locations at slow, normal, and fast walking speeds were evident from interpreting overlap of associated 95% CIs. More specifically, at slow walking speeds, none of the 95% CIs for MAPE values associated with each of the wear locations overlapped, confirming the statistically significant differences in these indicators of accuracy. At normal walking speeds, however, the 95% CIs for the MAPE values generated for the hip, thigh, and ankle wear locations did overlap, indicating that they were not significantly different from each other. At these same normal walking speeds, the 95% CIs for MAPE values for the arm/wrist wear location did not overlap, indicating that this accuracy indicator was significantly different. Lastly, at fast walking speeds, none of the 95% CIs for the MAPE values associated with each of the wear locations overlapped, again indicating significant differences between wear locations. These details are further illustrated in Figure 21.

4.2.3 MAPE by age group

With respect to Aim 3, the MAPE values respective to each age group are reported in Table 13. These specific MAPE values ranged from 81.2% (adolescents at 13.4 m·min^{-1}) to 14.4% (young adults at 93.9 m·min^{-1}). The difference in group mean values was never greater than 8% (ranging between bouts from 7.9-1.1%). Further, the SD ranged from ±37.7% (adolescents at 26.8 m·min^{-1}) to ±25.1% (adolescents at 134.1 m·min^{-1}).

The regression models built specific to Aim 3 indicated that age did not appear to have a significant effect on the step counting accuracy of wearable technologies at slow, normal, or fast walking speeds (p=0.21-0.50). MAPE values ranged from 51.4% (adolescents) to 49.8% (young adults) at slow walking speeds. During normal walking, MAPE values ranged from 16.9% (children) to 15.1% (young adults). And lastly, the MAPE values ranged from 24.1% (children) to 17.7% (young adults) at fast walking. None of the 95% CIs for the MAPE values associated with the three age groups overlapped, indicating that there were no age-associated significant differences. These details are further illustrated in Figure 24.

4.3 Precision: Correlation

4.3.1 Correlation by speed

In regards to Aim 1, the correlation coefficients (*r*) depicting the strength of the linear relationship between directly observed steps and steps measured from wearable technologies are reported in Table 7. Overall correlation coefficients, as well as correlation coefficients for slow, normal, and fast walking speeds are reported for each wearable technology. Further, the strength of the relationship for each wearable technology is depicted in Figure 9. Overall correlation coefficients ranged from r=0.90 (NL-1000) to r=0.39 (Sensewear). Six of the eight tested wearable technologies (75%) had overall correlation coefficients of r>0.80. In further detail, at slow walking speeds, correlation coefficients ranged from r=0.88 (StepWatch) to r=0.31 (Sensewear). At normal walking speeds, correlation coefficients ranged between r=0.96 (Actical) to r=0.05 (ActiGraph wrist), where five of the eight technologies (62.5%) actually had a weaker correlation with directly observed steps during normal walking speeds than during slow walking, with reductions ranging from r=-0.60 (ActiGraph wrist) to r=-0.04 (SW-200). Of the three wearable technologies that exhibited improved correlation during normal walking, improvements ranged from +0.26 (Actical) to +0.06 (ActiGraph waist). Lastly, at fast walking speeds, correlations ranged from r=0.94 (SW-200) to r=-0.24 (StepWatch). This negative correlation indicates that as the number of directly observed steps increased, the number of steps detected from the wearable technology (i.e. StepWatch) decreased. From normal to fast walking, there were decreased correlation values with directly observed steps for six of the eight (75%) tested wearable technologies. These changes ranged from r=-0.7 (StepWatch) to r=-0.08 (Actical). Improvements were +0.27 and +0.08, respectively, for the two wearable technologies (SW-200 and NL-1000) that exhibited improved correlation from normal to fast walking.

The regression modeling respective to Aim 1 indicated that speed did not have an effect (p=0.24) on the correlation of wearable technology-derived steps to directly observed steps. The mean correlation coefficients for slow, normal, and fast walking, across wearable technologies were r=0.68, 0.54, and 0.37, respectively. Due to the smaller sample size, and the reduced power of the correlation regression models, the 95% CIs built around these speed-based group means were relatively wide. This led to an overlap of 95% CIs, indicating that while correlation coefficients were qualitatively different across speeds (r=0.68, 0.54, and 0.37 at slow, normal, and fast walking), there were no significant correlation differences across wearable technologies between speed groups. These details are depicted in Figure 19.

4.3.2 Correlation by wear location

Respective to Aim 2, correlation coefficients of step counts respective to wrist (r=0.66 for ActiGraph wrist) and arm (r=0.39 for Sensewear) worn technologies displayed the weakest correlations for any of the tested wear locations. Respective to waist worn technologies, correlation coefficients ranged from r=0.90 (NL-1000) to r=0.86 (ActiGraph waist), which was the highest of any of the wear locations. The thigh worn technology (ActivPAL) had a correlation coefficient of r=0.86, and the ankle worn technology (StepWatch) had a correlation coefficient of r=0.82. Group mean correlation coefficients respective to each wear location are presented in Table 12.

The regression modeling respective to Aim 2 indicated that wear location did not have a significant effect (p=0.056) on the correlation between directly observed steps and those measured by the tested wearable technologies. These details are further illustrated in Figure 22. Correlation coefficients of the arm/wrist, waist, thigh, and ankle-worn technologies were r=0.53, 0.89, 0.86, and 0.82, respectively. The only significant difference between locations was the waist and arm/wrist. There was no overlap apparent between waist (1.00-0.74) and arm/wrist (0.72-0.34) worn technologies indicating significant differences in technology precision between these specific wear locations. However, there is consistent overlap of the 95% CIs group means when comparing all other wear locations, which is indicative of no significant differences.

4.3.3 Correlation by age group

Lastly, addressing Aim 3, correlation coefficients respective to age group are detailed in Table 13. Correlation coefficients for children ranged between r=0.89 (NL-1000) and r=0.38 (Sensewear). For adolescents, correlation coefficients ranged from r=0.94 (SW-200) to r=0.40 (Sensewear). For young adults, correlation coefficients ranged from r=0.95 (SW-200) to r=0.43 (Sensewear). Differences in correlation coefficients specific to each tested wearable technology and respective to each age group ranged from a 0.15 difference (ActivPAL) between adolescents and young adults to a 0.03 difference (ActiGraph wrist) between adolescents and young adults.

The regression models respective to Aim 3 indicated that age group did not have a significant effect (p=0.84) on the correlation coefficients comparing wearable technology derived steps and directly observed steps. The details of this comparison are illustrated in Figure 25. Overall correlation coefficients for children, adolescents, and young adults were r=0.77, 0.81, and 0.82, respectively. Further, the 95% CIs surrounding the group means consistently overlapped, indicating no significant differences in wearable technology precision between age groups.

4.4 Bias: PE

4.4.1 PE by speed

Respective to Aim 1, PE±SD (with corresponding CoV values) for each of the wearable technologies tested across the speed range for the total sample, children, adolescents, and young adults are cataloged in Tables 8-11 and illustrated in Figures 10-17. PE values ranged from -98.7% (waist worn ActiGraph at 13.4 m·min^{-1}) to 2.0% (Actical at 134.1 m·min^{-1}). The PE was greater for arm/wrist-worn technologies compared to those worn at the waist, thigh or ankle, with differences ranging between -83.0% and -48.6%. Further, PE was greatest at slow walking speeds (-49.3±36.3%) compared to normal (-15.0±22.2%) and fast (-16.9±23.0%) walking speeds. The tested wearable technologies consistently undercounted steps taken, apparent from negative PE values in 90% of cases (72 bouts out of the 80 total bouts measured).

Regression modeling specific to Aim 1 indicated that speed had a significant effect ($p<0.0001$) on PE values across all wearable technologies. This was primarily driven by significant improvements in PE evident between bouts at slow walking speeds. These details are illustrated in Figure 20, where significant reductions in PE occurred at slower walking speeds, but then leveled out during normal and fast walking. For example, PE values for 87.5% (7/8) of the tested wearable technologies were significantly different between speeds of 13.4 1 m·min^{-1} to 26.8 1 m·min^{-1}, where the associated 95% CIs between bouts did not overlap. However, during normal and fast walking speeds, 97.5% (78/80) of cases did not show significant differences (i.e., the 95% Cis did overlap) between contiguous bout speeds.

4.4.2 PE by wear location

The overall PE values according to wear location (Aim 2) are cataloged in Table 12. PE values ranged from -93.8% (arm/wrist at 13.4 m·min^{-1}) to 2.6% (waist at 134.1

$m \cdot min^{-1}$). Further, the least amount of bias was apparent from a PE value of -0.6% PE for the ankle at 26.8 $m \cdot min^{-1}$. The difference in PE relative to the different wear locations ranged from -83.0% to -48.6%. The greatest PE was associated with the arm/wrist wear location at all speeds, averaging -61.3±15.5% error, whereas PE values from the waist, thigh, and ankle were -13.5±23.1%. SD values associated with these PE ranged from ±6.6% (waist at 134.1 $m \cdot min^{-1}$) to ±37.3% (arm/wrist at 80.5 $m \cdot min^{-1}$).

Regression modeling specific to wear location, indicated that there was a significant effect of wear location on PE values during slow, normal, and fast walking speeds ($p<0.0001$). PE ranged from -55.9% (arm/wrist) to -4.6% (ankle) at slow walking speeds. At normal walking speeds, PE ranged from -51.4% (arm/wrist) to -2.7% (thigh). Lastly, at fast walking speeds, PE ranged from -54.2% (arm/wrist) to -0.7% (waist). The regression model built respective to this aim indicated that there were significant differences in PE across all four wear locations at slow and fast walking speeds, evident from no overlap of the 95% CIs for any of the wear locations. However, at normal walking speeds, there were no significant differences (i.e., there was an overlap of 95% CIs) in PE between waist, thigh, and ankle worn technologies, whereas the arm/wrist was significantly different (no overlap) from the others in terms of PE. These details are further illustrated in Figure 23.

4.4.3 PE by age group

With respect to Aim 3, PE for each age group are cataloged in Table 13. Group mean values ranged from -80.9% (adolescents at 13.4 $m \cdot min^{-1}$) to -13.5% (adolescents at 93.9 $m \cdot min^{-1}$). The difference in group means for each bout ranged from -7.1% to -0.3%

between age groups. Additionally, the SD ranged from ±40.0% (young adults at 26.8 $m \cdot min^{-1}$) to ±25.8% (adolescents at 134.1 $m \cdot min^{-1}$).

The regression modeling specific to Aim 3 indicated that age did not have a significant effect on of the tested wearable technologies' PE values at slow, normal, or fast walking speeds (p=0.21-0.50). PE ranged from -50.5% (adolescents) to -48.8% (young adults) at slow walking speeds. During normal walking, PE ranged from -15.8% (children) to -14.2% (young adults). And lastly, the PE values ranged from -22.4% (children) to -15.7% (young adults) at fast walking speeds. The regression models produced 95% CIs associated with PE that consistently overlapped for children, adolescents, and young adults across walking speeds. This was a clear indication that there were no age-associated significant differences in PE across speeds. The details of these results are further illustrated in Figure 26.

CHAPTER 5

DISCUSSION

The results of this secondary analysis enhance the field by providing a highly detailed and standardized digital (numeric; quantitatively defined) catalog of validation parameters (accuracy, precision, and bias) based upon 117 children, adolescents, and young adults (6-20 years of age), which is larger than any previously published validation study evaluating step counting wearable technology and conducted with youth. Further, these participants wore multiple wearable technologies attached at different wear locations and were tested across treadmill speeds ranging from 13.4 to 134.1 $m \cdot min^{-1}$ which is a broader speed range than any other protocol identified in the literature review (Chapter 2). This digital catalog can be used by researchers to gain an understanding of expected wearable technology performance for capturing steps when designing future PA studies for youth. For example, when planning PA surveillance or interventions for youth, this catalog can be used by researchers to support decisions regarding selection of wearable technologies in terms of preferred validity attributes. Further, this catalog can be used for reference values for device manufacturers and researchers to compare performance of past, current, and future wearable technologies. For example, device manufacturers now have a set of validity parameters that can be referenced to support development of new technologies in an effort to assure standardized measurement performance. Further, if researchers seek to analyze previously collected walking behavior data captured via one of the wearable technologies included as part of this analysis, they will be able to confidently interpret results by referencing its accuracy, precision, and bias values as cataloged herein. The assembly of this catalog is an imperative first step in the effort to harmonize the validation of step-counting wearable technology, and provides a framework for standardized analysis and presentation of validation metrics moving forward. For example, future validation studies should seek to include comprehensive walking speeds, as well as associated validity metrics, to ensure comparable rigorous evaluation as new wearable technologies are developed. This analysis also provides insight into which factors do (speed and wear location) and do not (age) affect these validity metrics.

5.1 Effect of speed:

It was clear that speed had a significant effect on accuracy and bias indicators associated with the tested step-counting wearable technologies. Specifically, the calculated alpha values for accuracy and bias were both $p<0.001$. That said, the effect of speed on accuracy and bias was primarily driven by greater error evident at slower speeds, with considerable improvements (MAPE: 34.2%, PE: 34.4%) apparent during normal walking speeds. Further, although not restricted to youth samples, Bassett et al.[22] suggested that wearable technologies only record 50-75% (50-25% MAPE) of the actual number of steps taken at walking speeds $\leq 26.8\ m \cdot min^{-1}$. The data included herein showed comparable performance to this suggestion, apparent from the finding that group mean values across wearable technologies at speeds $\leq 26.8\ m \cdot min^{-1}$ missed only 30.5% of actual steps taken (within the 50-25% MAPE). However, there were no significant differences in accuracy or bias values for any of the tested wearable technologies across increasing speeds during normal and fast walking.

Less force is generated during each step taken at slower walking speeds. Manufacturers of wearable technologies must make decisions about the trade-off between sensitivity and specificity of measurement to avoid misinterpreting signal noise as a true step, which may influence the accuracy and bias of wearable technology performance. In turn, low force accelerations that are true steps associated with very slow walking may be miscategorized.[60] During normal and faster walking speeds, there are greater forces generated with each step, resulting in > 0.35 Gs of force, a sensitivity threshold suggested by some device manufacturers for detecting steps.[51] This trend of improved MAPE and PE values with increased speed ($\geq$ 67.0 m·min^{-1}) is consistent with previous literature.[40, 45, 55-59] For example, Trapp et al.[58] reported improving MAPE from slow (40m·min^{-1}) to fast (91 m·min^{-1}) walking for the Accusplit (from 46.9% to .6%) and the Yamax SW-200 (from 44.1% to 8.9%) pedometers.

Although accuracy and bias were influenced by speed, precision indicators (i.e., correlation coefficients) were not similarly affected (p=0.24). A methodological explanation for why there is a significant effect on accuracy and bias, but not precision, is the apparent difference in power in the regression analyses. For example, the regression models' correlation coefficients were averaged across the eight wearable technologies for each of the speeds represented within the broader categories of slow, normal, and fast walking. This situation created only 32 data points averaged for slow and normal walking (eight wearable technologies by four speeds) and 16 data points averaged for fast walking (eight wearable technologies by two speed). In contrast, there were more data points at slow (n=3696), normal (n=3104), and fast (n=552), and therefore has greater power for the MAPE and PE regression models, as values are averaged across the total sample (8 technologies by number of walking bouts respective to each speed). This decrease in power related to precision analyses also increased the width of the associated 95% CIs, which in turn increased the likelihood of apparent overlap between speed groups, an indication of no significant differences. Further, there was a decrease in correlation apparent with increased speed. In order for correlation coefficients to be meaningful, there needs to be a wide range of values for comparison. For example, at slow walking speeds, there was a greater range of values of steps taken by participants, which would produce a more meaningful correlation when comparing to wearable technology measured steps. However, as speeds increased during normal and fast walking speeds, the range of steps taken narrowed. This narrowed range of steps taken produced correlation coefficients that could be misleading. Even though the correlation coefficient was reduced at normal and fast walking, the reduction was likely an artifact of the limited range of steps taken at those speeds.

5.2 Effect of wear location:

Wear location also had a significant effect on the accuracy and bias ($p<0.0001$) of wearable technologies step count measures. Consistently, wearable technologies worn at the arm/wrist performed significantly worse than those worn on the waist, thigh, or ankle in terms of accuracy and bias indicators. Specifically, when comparing arm/wrist worn technologies to those worn at the waist, thigh, and ankle, differences in MAPE and PE values between ranged from 79.3% to 47.2% and -83.0% to -48.6%, respectively. During normal walking speeds, waist, thigh, and ankle-worn technologies were all comparable, and within 5% error for both MAPE and PE values. This level of performance fell well within the 10% MAPE error recommended by the CTA standard,[44] with none of the three wear locations (waist, thigh, and ankle) displaying a significant advantage over the others. Valid step counting requires that wearable technologies be most sensitive to accelerations and gravitational forces occurring with each step taken during ambulation. For example, technologies worn at the ankle and thigh undergo greater displacement during ambulation due to their proximity to the foot. Specifically, Sandroff et al.[53] reported a 0.03% MAPE at 26.8 $m \cdot min^{-1}$ for the Step Watch in adults with Multiple Sclerosis. Although the accuracy of the Step Watch was worse (MAPE=4.4%) at the same speed for the sample herein, the Step Watch was the still the most accurate wearable technology tested at this speed. Technologies worn at the waist reflect the sinusoidal movements of the center of mass during ambulation. Waist worn technologies analyzed herein ranged from 5.9 to 2.7% MAPE during normal walking, below the 10% threshold for acceptable performance. These wear locations of the waist, thigh, and ankle produced superior step counting estimates when compared to arm/wrist wear locations.

Acceleration/force signals at the arms or wrist are not sufficient (low or slow) to be effectively detected consistently. For example, during slow walking, individuals may not move their arms as much. Although youth samples were not exclusively examined, Tudor-Locke et al.[54] provided a similar explanation of such step counting differences between waist and wrist worn technologies during treadmill walking. In that 2015 study, wrist worn technologies produced outputs that were significantly different from directly observed steps at treadmill speeds ranging from 13.4 to 187.8 $m \cdot min^{-1}$, where waist worn technologies displayed no significant differences at speeds >53.6 $m \cdot min^{-1}$. Chen et al.[24] suggested that wrist worn technologies produced by a variety of different manufacturers displayed a range of 9.6 to 1.5% MAPE for detecting steps when compared to direct observation during treadmill walking from 54 to 134 $m \cdot min^{-1}$. Again, although the results reported here are restricted to youth population, the best level of accuracy (MAPE) of the two arm/wrist worn technologies was 47.1% (ActiGraph at the wrist from13.4 to 134 m·min^{-1}).

Where direct comparison from the literature review (Chapter 2) was possible (i.e. replicative wearable technologies, wear locations, and/or reported validity metrics) the associated accuracy and bias values calculated were consistent. For example, Aminian et al.[61] reported MAPE values for the SW-200 and ActivPAL in youth at speeds of 67.0, 93.9, and 134.1 m·min^{-1}. The associated speed-based MAPE values determined from Aminian et al.[61] were 3.9%, 1.3% and 2.1% (SW-200) and 0.0%, 8.0%, 25.0% (ActivPAL), respectively. At the same speeds, results reported herein for the SW-200 were comparable at 6.5±12.7%, 2.5±10.4%, and 1.9±4.7% at the increasing speeds, respectively. Further, analysis herein showed that the ActivPAL had comparable MAPE values at 3.2±8.9%, 3.7±9.9% but displayed a lower MAPE of 11±24.7% at 67.0, 93.9 and 134.1 m·min^{-1}. The MAPE values reported from Aminian et al.[61] are comparable in the sense that all values fall within one SD value calculated for the MAPE value for both technologies herein.

At fast walking speeds, accuracy and bias appeared to worsen for the arm/wrist, thigh, and ankle worn technologies, while that for the waist worn technologies improved. All four of the arm/wrist, thigh, and ankle worn technologies displayed increased MAPE (range= 21.0% to 3.0% change) and PE (range= -18.0% to -3.0% change) values, and decreased correlation coefficients (range= -0.70 to -0.03 change) from normal to fast walking, indicating decreased performance at higher speeds. This trend was inconsistent with that reported in three previous studies[49,61,62] that included speeds ≥120.7 m·min^{-1}. Of the four waist worn technologies included as part of these earlier studies,[49,61,62] two technologies displayed increasing error ranging from 0.8% to 0.7% between normal to fast walking speeds, while two showed decreasing error between 7.5% and 0.2%. It is obviously difficult to solely attribute performance to wear location as there is apparent inconsistency in performance between different brands/models of technologies more specifically. For example, one of the studies listed above, Dueker et al.,[49] utilized the SportBrain iStep X1 and Digiwalker SW-701, neither of which were included as part of the analysis herein. Therefore, while the PE values of the SportBrain iStep X1 and Digiwalker SW-701 at 134.1 $m \cdot min^{-1}$ (-3.9% and -4.3%, respectively) are comparable to the PE values of the four waist worn technologies included (2.0±4.6%[Actical] to -0.3±10.6% [ActiGraph]). The PE values of the SportBrain iStep X1 and Digiwalker SW-701 are within 1 SD of the waist worn technologies analyzed herein.

Another clear finding was that there was no effect of wear location on precision measures (correlation coefficients), which is inconsistent with the effect of wear location on accuracy and bias. Again, this inconsistency of significance across validity metrics is due to reduced power in the regression models. The details of this divergence in power in regression models respective to each validity metric are described above with the effect of speed on precision compared to the effect of speed on accuracy or bias. Further, there were only two technologies worn at the arm/wrist, one at the thigh, and one at the ankle. This further reduces the amount of data points included in regression models for correlation coefficients, and in turn produces wider 95% CIs. Wider 95% CIs increases the likelihood of overlap, which would indicate no significant differences.

Future studies should continue to explore the effect of wear location on technology performance, and include more wearable technologies at each location to further analyze the effect of location, and not merely the effect of the technology. For example, with only one wearable technology included at the thigh and the ankle, it is difficult to parse out the discrepancies in step detection error that was due to the location, rather than due more directly to the technology itself.

5.3 Effect of age:

No previous studies have directly analyzed the effect of age on performance of wearable technologies so this is a landmark study. Despite the comprehensive age span and associated developmental stage represented in the original data sample, there was no apparent effect of age on the step counting ability of the tested wearable technologies at slow, normal, or fast walking (p=0.21 to 0.50). As discussed in the Results section, MAPE and PE values never differed more than 8% between age groups, and respective to correlation coefficients, values never differed more than r=0.06. In turn, the 95% CIs for all three age groups overlapped at slow, normal, and fast walking speeds, indicating no significant differences due to age. One explanation for this is that beyond six years of age, there are no notable differences in walking patterns compared to those associated with adults.[63] PA studies and interventions designed to capture steps in youth do not need to consider age groups for optimizing measurement methods. However, since postural and locomotor control may differ in very young children[63] future studies should seek to include populations below the ages of six years to extend this catalog of validity parameters associated with various wearable technologies.

5.4 Validity metrics

In determining wearable technology performance, accuracy and bias measures give direct indication of wearable technology error. MAPE and PE are scaled measures

(%) and defined by the amount of error determined relative to directly observed steps. These metrics provide useful detail regarding the magnitude and direction of the error, and are more informative than precision measures relative to understanding performance of step-counting wearable technologies. Correlation coefficients (indicative of precision) are not as telling as accuracy and bias measures, as correlation coefficients are unitless values that indicate the strength of the linear relationship between wearable technology measured and directly observed steps. The strength of the linear relationship is not representative of the magnitude of error in respect to directly observed steps. To be clear, a wearable technology could consistently (i.e., precisely) undercount the same amount of steps respective to directly observed steps, and therefore, display a high correlation coefficient, despite consistent error (more apparent from inexact MAPE and PE values). Further, correlation coefficients require a wide range of steps taken in order to provide meaningful data. As described above, when broken down into groups for each respective aim (by speed, wear location, and age), the power is limited and therefore is not as informative about relative performance of a wearable technology. In contrast, MAPE and PE provide a more definitive depiction of device performance in regard to directly observed steps, by clear and more easily comparable magnitude and/or direction of error.

5.5 Rank order

As noted above, MAPE and PE are the most informative validity metrics useful for attempting to rank order wearable technology performance in regards to directly observed steps. Assessing the degree and direction of error, as opposed to the consistency (i.e. precision) of technology measurements, is more useful when assessing which technology is superior. Specifically, CTA[44] standards recommend MAPE be used as the metric of choice for evaluating technology performance over the range of normal walking speeds. While the <10% MAPE recommendation reported from the CTA[44] is not clearly based on data, the reference values produced herein can actually provide such data to inform more evidence-based thresholds. Five of the tested wearable technologies actually performed at <5% MAPE over the range of walking speeds.

With that being said, researchers should anticipate and consider the population of interest when selecting appropriate wearable technologies. For example, if the population of interest typically walks at slow walking speeds, the best technology for slow walking is different than it is for normal or fast walking. More specifically, at slow walking, the rank order of best technologies in descending order based on MAPE values is: StepWatch (6.6%), ActivPAL (19.4%), Actical (46.0%), SW-200 (49.0%), NL-1000 (55.0%), ActiGraph at the wrist (61.8%), ActiGraph at the waist (73.8%), and Sensewear (89.2%). While the CTA[44] recommended threshold of 10% is intended for normal walking speeds (67.0 to 107.2 $m \cdot min^{-1}$) one technology (StepWatch) fell below this threshold for acceptable error requirement for validation during slow walking speeds. Further, this rank order for slow walking speeds is the same for PE, where the respective values were StepWatch (-4.5%), ActivPAL (-17.8%), Actical (-45.3%), SW-200 (-48.6%), NL-1000 (-53.9%), ActiGraph at the wrist (-61.8%), ActiGraph at the waist (-73.7%), and Sensewear (-89.1%). No specific recommendation for wearable technology performance thresholds respective to PE exist, so these assembled reference values are informative. However, if researchers anticipate that their population of interest typically walk at normal speeds, the rank order of best technologies for capturing steps changes. Under normal walking speeds, the descending rank order of technologies based on

MAPE values was: Actical (1.4%), NL-1000 (2.9%), SW-200 (4.0%), ActivPAL (4.1%), StepWatch (4.7%), ActiGraph at the waist (7.4%), ActiGraph at the wrist (40.6%), and the Sensewear (62.3%). Here, six of the eight technologies met the CTA[44] recommended threshold for acceptable technology error, while the two worn at the arm/wrist were >10% (ActiGraph at the wrist, and the arm-worn Sensewear). Five (with the exception of the ActiGraph at the waist) actually performed at a more rigorous level of <5% MAPE across this normal speed range. Again, this order at normal walking speeds did not change with respect to PE, where the rank order with corresponding PE values was: Actical (0.0%), NL-1000 (-2.0%), SW-200 (-2.2%), ActivPAL (-2.6%), StepWatch (-3.5%), ActiGraph at the waist (-6.8%), ActiGraph at the wrist (-40.6%), and the Sensewear (-62.0%). Lastly, if a researcher is interested in investigating fast walking, again the rank order of technologies changes compared to slow or normal speeds. In regards to MAPE, the rank order at fast walking speeds was: SW-200 (1.6%), NL-1000 (1.7%), Actical (2.2%), ActiGraph at the waist (5.0%), ActivPAL (10.5%), StepWatch (22.1%), ActiGraph at the wrist (47.9%), and Sensewear (60.2%). The four technologies that met the < 10% recommended threshold, were the four waist worn technologies. The rank order did change slightly with respect to PE, where the NL-1000 displayed the least amount of error (0.5%) followed by the SW-200 (1.0%). This difference is due to the aggregation of over and underestimation when calculating bias. For example, if the Actical overestimates steps by 5% for one participant, but undercounts steps by 5% for another participant, the aggregated PE would be 0.0%.

The increased error apparent for StepWatch step detection at higher speeds was not expected. This increased error at higher walking speeds, however, could be explained in part by the fact that, as part of the set protocol, the StepWatch was programmed (i.e. initialized; calibrated) using the default settings that combine cadence and sensitivity settings to optimize step detection during normal walking (capturing 98% to 100% of all steps taken from 27 to 107 $m \cdot min^{-1}$).[64] However, with these default settings and at faster walking speeds, steps may be taken so frequently that when a single step is recorded subsequent steps may occur too quickly to also be recorded.[64] The results reported herein for the StepWatch error at the fastest walking speeds of 134.1 $m \cdot min^{-1}$ using the default cadence and sensitivity settings is comparable to that reported by Toth et al.[64] As part of their protocol, the StepWatch recorded an MAPE of 24.5% of steps at 134.1 $m \cdot min^{-1}$, comparable to the 23.5% MAPE reported herein. Although the StepWatch is know to be sensitive for recording steps using the default settings at slow and normal walking, if the population and/or circumstance of interest is known entail faster walking speeds, researchers may wish to consider altering the cadence and sensitivity settings to optimize step detection.

5.6 Strengths:

There are a number of strengths of this secondary analysis that tie back to the original data source. First, the methods of the original study used direct observation and video back up recordings to ensure an accurate criterion measure was used. This is consistent with the CTA recommendations for creating performance criteria data for step counting wearable technologies.[44] Secondly, the original protocol covered a broad range of walking speeds from 13.4 $m \cdot min^{-1}$ to 134.1 $m \cdot min^{-1}$. This comprehensive coverage of walking speeds extends beyond the data derived from previous studies that used a narrower range of walking speeds. For example, Rosenkranz and Rosenkranz[62] used a protocol that only studied four speeds ranging between 40.2 $m \cdot min^{-1}$ to 161 $m \cdot min^{-1}$. The broader and greater number of tested speeds afforded by the original study supplying the current analyses provided an opportunity to evaluate the effects of speed on accuracy, bias, and precision, and the incremental changes as speed increased. To be clear, this secondary analysis was not intended to recommend an optimal treadmill protocol for evaluating wearable technologies. However, the finer resolution of the incremental changes in treadmill speed offered by the original protocol allowed for enhanced inspection of more gradual performance changes of the tested wearable technologies across increasing speeds. This comprehensive view of accuracy, precision, and bias allows researchers to have a broader, more robust understanding of wearable technology performance, beyond just a single metric. Again, this is the first study in a young population to report all three metrics. Further, the study allowed for extensive collection of wearable technologies, representing diverse wear locations. Additionally, the use of a level treadmill with set speeds allowed for a standardized assessment at the specified walking speeds. Finally, the dataset used for this study included a large sample size (n=117), evenly distributed by age year, allowing for a robust evaluation of the potential effects of age on the three validity parameters (there were none).

5.7 Limitations:

While the analyses identified knowledge gaps in our understanding of validity of step-counting wearable technologies, there are several limitations that must be acknowledged. First, this secondary analysis included the evaluation of only eight step counting wearable technologies. While this was a more extensive array of technologies than evaluated in previously published validation studies,[43, 49,58-62] there are still many other wearable technologies available that were not included in this original data set. Some of the originally tested wearable technologies are also now obsolete (e.g., Sensewear). Publication of these specific validity-related values are still important for users of data previously collected with such devices, however, and also to enable robust comparisons between different types of wearable technologies, past, present, and into the future. That said, validation parameters that are assembled herein will be generalizable to only the named wearable technologies. Additionally, there was only one wearable technology worn on the ankle, one on the thigh, and two on the arm/wrist in the original study. In contrast, there were four wearable technologies located on the waist. When grouping mean values based on wear location, those locations with multiple wearable technologies may be affected by averaging mean validity parameters, as compared to locations where only a single wearable technology was used. Future studies will be able to extend this assembled validity parameter catalog to add additional location-based data. Another limitation to this secondary analysis is that the size of the available analytical sample diminished as speed increased in the original study. To be very clear, based on the very broad age range of the sample (6-20 years of age) and the termination criteria set in place for the original treadmill protocol, not all participants completed every bout, and therefore at higher speeds there were fewer data points. Again, this may have affected the estimates of central tendency and distribution of the listed validation parameters as there were fewer raw data points available. This situation may mean that individual data points may have had a greater effect on the overall group mean, compared to speeds with a larger and more disperse sample of data points. Future studies may determine that it is necessary to recruit samples capable of continuing for the same number of walking and running bouts. Another limitation was that not everyone ended the protocol with a complete running bout, and because the participants broadly ranged in age and stature, the exact speed when individuals walked vs. ran was variable. For that reason, only walking bouts were considered as part of this analysis. Future studies should extend these validity parameters to running by implementing protocols that will more consistently elicit a running bout, including specifically recruiting those capable of providing sufficient running data. Lastly, it is important to note that the step-based validation of these wearable technologies is not generalizable to the free-living condition. It has been shown that the step count parameters gathered from wearable technologies display greater error in a free-living setting compared to a highly controlled, laboratory-based setting (e.g., performed using a set treadmill protocol). The sporadic, and non-uniform patterns of children's movement may further limit the generalizability of these step count data when quantifying free-living step-based PA behavior.[20]

CHAPTER 6.

CONCLUSION

In conclusion, the results reported and discussed herein are an important first step to harmonizing the validation efforts of wearable technologies' step counting abilities. This effort provides comprehensive validity parameters for step counting wearable technologies observed during treadmill walking in young people 6-20 years of age. Further, it bridges the knowledge derived from fragmented validation efforts focused on adult populations by now extending to a large sample that included children, adolescent, and young adults. Cataloging expected validation parameters for step counting wearable technologies in young samples is important to inform and facilitate empirically-based reference values needed to evaluate accuracy, precision, and bias of new technologies. It was clear that speed and wear location had a significant effect on the accuracy and bias of wearable technologies step counting ability, but not the precision. However, age grouping does not influence wearable technology performance with regards to any of these three validity metrics. Future research should continue to rigorously validate new wearable technologies as they are developed, and also extend this standardized laboratory-based evaluation to the free-living environment.

Table 1: Step counting treadmill validation studies among wearable technologies in youth

Pedometer Studies						
Reference	**Sample**	**Protocol (duration and speeds)**	**Wearable Technologies**	**Accuracy**	**Precision**	**Bias**
Beets, Patton and Edwards 2004	20 children (50% females), 5 to 11 years	2 min@ 40, 54, 67, 80, and 94 $m \cdot min^{-1}$	Walk4Life 2505 (WL), Digiwalker SW-200 (DW200), Sun TrekLINQ (SUN) and Digiwalker SW-701 (DW701)	MAPE: Only visualized in graph	ICC WL:0.516 to 0.992 DW200: 0.727 to 0.998 SUN: 0.330 to 0.971 DW701:0.722 to 0.993	Not reported
Trapp et al. 2012	45 children (51.1% females), age 10.67 ± 0.77 years	3 min @ 40, 67, 91 $m \cdot min^{-1}$	Accuspllit AH120, DigiWalker SW-700	MAPE AH120= 46.9 to 8.6% SW-700= 44.1 to 8.9%	Not reported	Bland- Altman Plots
Ramírez-Marrero et al. 2002	31 children (45.2% females) 7 to 12 years	2 min @ 59, 70, 91 $m \cdot min^{-1}$	Digiwalker SW-200 (hip;belt), Digiwalker SW-200 (midline; pouch)	MAPE Hip= 1 to 15% Midline= 2 to 5%	Not reported	Not reported
Mitre et al. 2009	27 children (51.9% females) 8 to 12 years	5 min @ 13.4, 26.8, 40.2, 53.6 $m \cdot min^{-1}$	Omron HJ-105, DigiWalker SW-200	Not reported	Not reported	PE: Omron HJ-105 and DigiWalker SW-200: 100% to 60% (not specified) Bland-Altman Plots

Dueker, Gauderman, and McConnell, 2012	17 children (58.8% females), aged 10 to 17 years	5 min @ 53.6, 80.5, 107.3, 134.1 m·min^{-1}	SportBrain iStep X1, Digiwalker SW-701	MAPE	Not reported	PE: iStep X1= -1.3 to -3.9% SW-701= -19.3 to -4.3%
Aminian and Hinckson, 2012	25 children (68% females) 9.9 ± 03 years	2 min @ 50, 66, 93, 133 m·min^{-1}	Yamax DigiWalker SW-200 and NL-2000	Not reported	Not reported	PE: SW-200: -4% to 1% NL-2000: -11% to 1%
Rosenkranz, Rosenkranz, and Weber, 2011	19 children (36.8% female) aged 7–11 years	3 min @ 40.2, 80.5, 120.7, 161 m·min^{-1}	Digiwalker SW-200	Not reported	Not reported	PE: -4% (not specified by speeds)
Accelerometer Studies						
Aminian and Hinckson, 2012	25 children (68% females) 9.9 ± 03 years of age	2 min @ 50, 66, 93, 133 m·min^{-1}	ActivPAL	Not reported	ICC: ActivPAL:0.21 to 1.00	Not reported
Rosenkranz, Rosenkranz, and Weber, 2011	19 children (36.8% female) aged 7–11years	3 min @ 40.2, 80.5, 120.7, 161 m·min^{-1}	Actical (waist), Actical (ankle)	Not reported	ICC: Waist: 0.927 Ankle:0.854 (not specified by speeds)	PE: Waist: -11% Ankle: -17% (not specified by speeds)

Table 2: Participant characteristics

Variable		Total sample (N=121)	Children (N=56)	Adolescents (N=41)	Young Adults (N=24)
Sex (M/F)		49.6% F	48.2% F	51.2% F	50.0% F
Race/Ethnicity	African-American	35.5%	39.3%	34.1%	29.2%
	Caucasian	62.0%	57.1%	63.4%	70.8%
	Other	2.5%	0.6%	2.5%	0.0%
Age (years)		13.0 ± 4.2	9.1 ± 2.0	14.9 ± 1.4	19.0 ± 0.8
Height (cm)		155.3 ± 16.9	141.4 ± 12.8	165.6 ± 8.7	170.0 ± 8.9
Leg Length (cm)		73.8 ± 9.7	67.0 ± 8.3	79.4 ± 6.6	80.4 ± 5.4
Weight (kg)		55.8 ± 22.2	40.9 ± 14.8	68.7 ± 21.9	68.4 ± 14.6
BMI (kg/m^2)		22.4 ± 6.3	20.0 ± 5.1	24.9 ± 7.1	23.7 ± 5.1

Notes: All continuous values are reported as mean ± SD. Categorical data is reported as frequencies (%)

Table 3: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies

		Slow				Normal				Fast	
	Speed (m·min^{-1}) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
ActiGraph (Wrist)	MAPE±SD	88.8 ± 11.1	70.4 ± 18.4	46.9 ± 18.7	38.4 ± 22.0	38.0 ± 22.8	38.1 ± 22.3	41.1 ± 20.1	45.3 ± 18.7	48.8 ± 14.1	47.1 ± 7.8
	CoV	0.13	0.26	0.38	0.57	0.60	0.59	0.48	0.41	0.29	0.17
ActiGraph (Waist)	MAPE±SD	98.5 ± 4.1	92.7 ± 8.1	69.6 ± 18.2	34.6 ± 19.3	11.5 ± 15.6	6.7 ± 14.0	5.2 ± 12.7	6.5 ± 13.6	5.4 ± 16.1	4.6 ± 9.5
	CoV	0.04	0.09	0.26	0.56	1.36	2.09	2.42	2.10	3.01	2.08
ActivPAL (Thigh)	MAPE±SD	60.2 ± 28.2	10.9 ± 14.7	3.2 ± 8.7	3.5 ± 9.0	3.2 ± 8.9	3.3 ± 9.2	3.7 ± 9.9	6.0 ± 14.7	9.8 ± 23.1	11.1 ± 24.7
	CoV	0.47	1.35	2.74	2.60	2.80	2.79	2.67	2.43	2.36	2.22
Actical (Waist)	MAPE±SD	94.9 ± 10.5%	61.4 ± 26.5%	23.7 ± 16.0%	4.0 ± 6.8%	1.5 ± 4.5%	1.6 ± 3.4	1.1 ± 1.8%	1.4 ± 1.7	2.1 ± 3.8	2.4 ± 4.4
	CoV	0.11	0.43	0.67	1.71	2.92	2.11	1.58	1.20	1.82	1.86
StepWatch (Ankle)	MAPE±SD	14.9 ± 19.8	4.4 ± 13.1	3.4 ± 13.2	3.6 ± 13.8	3.2 ± 13.4	2.9 ± 10.4	4.1 ± 12.0	8.7 ± 15.7	20.7 ± 18.0	23.5 ± 7.9
	CoV	1.33	3.01	3.91	3.79	4.27	3.59	2.90	1.81	0.87	0.33
Sensewear (Arm)	MAPE±SD	98.7 ± 1.5	96.8 ± 6.1	87.4 ± 22.9	74.0 ± 35.1	66.9 ± 42.5	62.1 ± 44.1	61.0 ± 44.8	59.0 ± 43.0	61.0 ± 42.3	59.4 ± 39.4
	CoV	0.02	0.06	0.26	0.47	0.64	0.71	0.73	0.73	0.69	0.66
SW200 (Waist)	MAPE±SD	84.9 ± 12.9	63.6 ± 26.2	32.8 ± 25.1	14.8 ± 16.8	6.5 ± 12.7	2.9 ± 11.4	2.5 ± 10.4	2.9 ± 11.1	1.3 ± 1.5	1.9 ± 4.7
	CoV	0.15	0.41	0.77	1.14	1.95	3.92	4.13	3.78	1.18	2.42

NL-1000 (Waist)	MAPE±SD	91.4 ± 10.3	78.8 ± 21.4	36.6 ± 21.4	13.0 ± 11.8	4.4 ± 10.8	3.9 ± 14.0	2.0 ± 6.3	1.4 ± 2.2	1.7 ± 3.4	1.7 ± 4.5
	CoV	0.11	0.27	0.58	0.90	2.48	3.59	3.18	1.57	1.99	2.67

Notes: All MAPE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 4: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for children (6-12 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (53)	**26.8** (53)	**40.2** (51)	**53.6** (49)	**67.0** (47)	**80.5** (42)	**93.9** (34)	**107.2** (24)	**120.7** (10)	**134.1** (0)
ActiGraph (Wrist)	MAPE±SD	83.5 ± 11.2	67.5 ± 15.5	52.0 ± 14.9	43.0 ± 18.9	40.9 ±20.3	39.6 ± 20.1	43.8± 17.7	48.3± 14.1	48.5 ± 3.9	## ± ##
	CoV	0.13	0.23	0.29	0.44	0.49	0.51	0.40	0.29	0.08	#
ActiGraph (Waist)	MAPE±SD	97.2± 5.7	90.8 ± 7.7	68.3± 15.7	36.8 ± 16.0	12.7 ± 11.6	6.5 ± 8.3	4.3 ± 5.7	6.1 ± 6.8	10.9 ± 19.6	## ± ##
	CoV	0.06	0.09	0.23	0.43	0.91	1.27	1.32	1.11	1.80	#
ActivPAL (Thigh)	MAPE±SD	57.3 ± 26.3	10.5 ± 13.3	2.8 ± 8.2	3.3 ± 8.5	2.6 ± 7.8	3.3 ± 9.3	3.6 ± 9.5	9.0 ± 13.4	11.8 ± 15.8	## ± ##
	CoV	0.46	1.26	2.91	2.58	2.96	2.78	2.67	1.48	1.34	#
Actical (Waist)	MAPE±SD	90.6± 13.9	65.1 ± 21.9	23.2 ± 14.9	3.8 ± 4.8	0.9 ± 0.9	1.7 ± 3.9	0.9± 0.8	1.7 ± 2.0	4.6 ± 7.8	## ± ##
	CoV	0.15	0.34	0.64	1.27	0.95	2.21	0.87	1.18	1.71	#
StepWatch (Ankle)	MAPE±SD	16.0 ± 19.5	4.9 ± 13.8	3.7 ± 14.0	4.5 ± 15.3	3.4 ± 14.5	3.1 ± 6.0	6.3 ± 9.9	15.2± 15.1	28.9 ± 10.5	## ± ##
	CoV	1.22	2.80	3.80	3.37	4.21	1.96	1.58	0.99	0.36	#
Sensewear (Arm)	MAPE±SD	98.4 ± 2.1	95.4 ± 7.4	86.4 ± 20.9	75.7 ± 33.3	67.2 ± 41.3	61.8± 42.9	64.0 ± 42.8	61.2 ± 40.6	82.6 ± 29.0	## ± ##
	CoV	0.02	0.08	0.24	0.44	0.62	0.69	0.67	0.66	0.35	#
SW200 (Waist)	MAPE±SD	81.7 ± 16.0	58.3 ± 28.1	31.0 ± 26.0	14.7± 21.3	7.8 ± 17.4	7.1 ± 17.3	4.5± 17.0	5.6± 20.0	1.7 ± 1.4	## ± ##
	CoV	0.20	0.48	0.84	1.45	2.23	2.42	3.81	3.58	0.82	#

NL-1000 (Waist)	MAPE±SD	89.3 ± 13.2	72.9 ± 24.1	36.9 ± 17.2	10.6 ± 8.4	2.5 ± 3.5	4.8 ± 15.7	2.9 ± 9.6	1.2 ± 1.3	4.0 ± 7.1	## ± ##
	CoV	0.15	0.33	0.47	0.79	1.40	3.27	3.27	1.05	1.80	#

Notes: All MAPE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed. No child completed the final bout of 134.1 $m \cdot min^{-1}$, and therefor is represented by # in the last column.

Table 5: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for adolescents (13-17 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (40)	**26.8** (40)	**40.2** (40)	**53.6** (40)	**67.0** (39)	**80.5** (39)	**93.9** (36)	**107.2** (34)	**120.7** (25)	**134.1** (6)
ActiGraph (Wrist)	MAPE±SD	94.0 ± 7.1	74.0 ± 19.1	49.1± 20.4	36.1± 24.3	36.1 ± 24.8	36.5 ± 24.4	41.1± 20.7	47.3 ± 19.0	50.8 ± 11.4	44.1 ± 12.2
	CoV	0.08	0.26	0.42	0.67	0.69	0.67	0.50	0.40	0.22	0.28
ActiGraph (Waist)	MAPE±SD	99.6 ± 0.9	93.9 ± 8.5	70.4 ± 18.6	33.5± 20.7	9.1± 14.2	5.4 ± 12.6	4.0 ± 8.0	5.6 ± 9.1	1.6 ± 1.7	6.9± 14.4
	CoV	0.01	0.09	0.26	0.62	1.56	2.32	2.00	1.61	1.05	2.10
ActivPAL (Thigh)	MAPE±SD	61.5 ± 28.8	12.4 ± 18.2	4.6± 11.2	4.8 ± 11.6	5.0 ± 12.1	4.6 ± 11.6	5.4 ± 12.8	7.2 ± 19.3	15.2 ± 31.2	7.0 ± 15.0
	CoV	0.47	1.47	2.44	2.43	2.41	2.54	2.36	2.68	2.05	2.15
Actical (Waist)	MAPE±SD	98.4 ± 4.0	64.7 ± 26.6	27.2 ± 15.3	4.9 ± 9.5	2.3 ± 7.2	1.4 ± 3.4	0.9 ± 0.8	1.0 ± 0.8	1.2± 0.7	0.9 ± 0.6
	CoV	0.04	0.41	0.56	1.94	3.15	2.48	0.91	0.76	0.61	0.66
StepWatch (Ankle)	MAPE±SD	14.9 ± 17.9	1.7 ± 1.7	1.5 ± 2.2	1.4 ± 2.6	1.5± 2.8	1.0 ± 1.2	0.9 ± 0.8	2.9 ± 6.6	19.2 ± 14.4	20.3 ± 9.3
	CoV	1.20	1.02	1.50	1.91	1.86	1.19	0.87	2.25	0.75	0.46
Sensewear (Arm)	MAPE±SD	98.9 ± 0.7	97.6 ± 5.7	89.6 ± 22.3	77.2± 34.0	72.7 ± 40.8	65.9 ± 44.1	62.0 ± 45.8	60.7 ± 43.1	58.2 ± 42.7	51.3 ± 41.7
	CoV	0.01	0.06	0.25	0.44	0.56	0.67	0.74	0.71	0.73	0.81
SW200 (Waist)	MAPE±SD	88.3 ± 6.8	66.1 ± 23.5	34.7 ± 21.8	14.5 ± 11.8	4.9 ± 6.8	2.3 ± 3.1	1.0 ± 1.0	1.8 ± 3.0	1.4 ± 1.7	1.0 ± 0.5

	CoV	0.08	0.36	0.63	0.82	1.39	1.37	1.08	1.66	1.27	0.48
NL-1000 (Waist)	MAPE±SD	93.7 ± 3.1	82.9 ± 17.9	38.2 ± 24.0	16.1 ± 13.9	4.0 ± 7.0	1.7 ± 3.3	1.2 ± 3.0	1.3 ± 2.2	1.3 ± 1.3	0.7 ± 0.7
	CoV	0.03	0.22	0.63	0.86	1.73	1.92	2.47	1.71	1.03	1.03

Notes: All MAPE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 6: Accuracy: Mean absolute percent error (MAPE) of tested wearable technologies by speed for young adults (18-20 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (24)	**26.8** (24)	**40.2** (24)	**53.6** (24)	**67.0** (24)	**80.5** (23)	**93.9** (23)	**107.2** (23)	**120.7** (18)	**134.1** (10)
ActiGraph (Wrist)	MAPE±SD	92.1 ± 11.5	70.9 ± 22.4	45.3 ± 22.6	32.7 ± 22.9	33.7 ± 24.7	37.9 ± 23.3	38.5 ± 22.6	39.4 ± 21.7	46.1 ± 20.0	48.7 ± 2.9
	CoV	0.12	0.32	0.50	0.70	0.73	0.61	0.59	0.55	0.43	0.06
ActiGraph (Waist)	MAPE±SD	99.4 ± 2.1	94.5 ± 7.9	70.9 ± 22.9	31.7± 22.9	12.4 ± 23.1	9.0 ± 22.4	8.6 ± 22.5	8.1 ± 22.2	7.4 ± 23.3	3.2 ± 5.4
	CoV	0.02	0.08	0.32	0.72	1.86	2.48	2.61	2.75	3.13	1.71
ActivPAL (Thigh)	MAPE±SD	64.5 ± 31.5	9.3± 11.4	1.5 ± 2.5	1.6 ± 2.6	1.2 ± 1.9	1.1 ± 1.1	1.3 ± 2.3	1.2 ± 1.8	1.2 ± 0.8	13.6 ± 29.6
	CoV	0.49	1.23	1.63	1.62	1.54	0.98	1.81	1.51	0.66	2.17
Actical (Waist)	MAPE±SD	98.6 ± 4.4	47.7 ± 31.8	19.3 ± 18.7	2.9 ± 4.7	1.4 ± 2.3	1.8 ± 2.4	1.9 ± 3.3	1.8 ± 2.2	1.9 ± 2.4	3.2 ± 5.4
	CoV	0.04	0.67	0.97	1.65	1.62	1.38	1.75	1.26	1.25	1.68
StepWatch (Ankle)	MAPE±SD	12.5 ± 23.8	7.5 ± 20.2	5.9 ± 20.3	5.6 ± 20.2	5.2 ± 20.2	5.8 ± 20.6	6.1 ± 20.7	10.4 ± 22.2	18.3 ± 24.3	25.5 ± 6.6
	CoV	1.91	2.67	3.46	3.63	3.92	3.55	3.39	2.13	1.33	0.26
Sensewear (Arm)	MAPE±SD	99.2 ± 0.7	98.6 ± 1.2	85.7 ± 28.2	65.4 ± 40.4	58.0 ± 47.0	56.1 ± 47.3	55.2 ± 47.5	54.3 ± 46.8	52.7 ± 45.7	64.2 ± 39.4
	CoV	0.01	0.01	0.33	0.62	0.81	0.84	0.86	0.86	0.87	0.61
SW200 (Waist)	MAPE±SD	86.3 ± 11.5	71.3 ± 24.4	33.5 ± 28.8	15.4 ± 14.0	6.4 ± 8.2	1.6 ± 2.4	2.1 ± 3.6	1.9 ± 2.8	1.0 ± 1.4	2.5 ± 5.9

	CoV	0.13	0.34	0.86	0.91	1.28	1.52	1.69	1.48	1.35	2.37
NL-1000 (Waist)	MAPE±SD	92.4 ± 10.3	84.8 ± 17.2	33.3 ± 25.0	12.9 ± 13.1	8.4 ± 20.5	5.9 ± 20.6	1.8 ± 3.4	1.8 ± 2.9	1.0 ± 1.3	2.3 ± 5.7
	CoV	0.11	0.20	0.75	1.01	2.44	3.47	1.93	1.63	1.25	2.48

Notes: All MAPE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 7: Precision: Correlation coefficient (*r*) of the tested wearable technologies by speed

			Slow				Normal				Fast	
	Speed (m·min^{-1}) (N)	**Total**	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
ActiGraph (Wrist)	***r***	0.67	0.65				0.05				0.02	
ActiGraph (Waist)	***r***	0.86	0.60				0.65				0.44	
ActivPAL (Thigh)	***r***	0.86	0.86				0.65				0.12	
Actical (Waist)	***r***	0.89	0.70				0.96				0.88	
StepWatch (Ankle)	***r***	0.82	0.88				0.46				-0.24	
Sensewear (Arm)	***r***	0.39	0.31				0.08				-0.07	
SW200 (Waist)	***r***	0.89	0.73				0.68				0.95	
NL-1000 (Waist)	***r***	0.90	0.73				0.81				0.89	

Notes: All correlation coefficients are presented as values to two decimal places. the total number of walking bouts for each technology. N represents the number of participants in the analytical sample, as not every participant completed every speed. Values closer to 1.00 indicate stronger linear fit to directly observed steps. The total correlation coefficients represents strength of relationship across total walking bouts. Correlation coefficient of slow, normal, and fast walking are restricted to walking bouts of those respective speeds.

Table 8: Bias: Percent error (PE) of the wearable technologies by speed

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
ActiGraph (Wrist)	PE±SD	-88.8 ± 11.1	-70.4 ± 18.4	-49.6 ± 18.7	-38.4 ± 22.0	-37.6 ± 23.0	-38.1 ± 22.3	-41.1 ± 20.1	-45.3 ± 18.7	-48.8 ± 14.1	-47.0 ± 7.8
	CoV	-0.1	-0.3	-0.4	-0.6	-0.6	-0.6	-0.5	-0.4	-0.3	-0.2
ActiGraph (Waist)	PE±SD	-98.5± 4.1	-92.7 ± 8.1	-69.6 ± 18.2	-34.3 ± 19.8	-11.1 ± 15.8	-5.7 ± 14.4	-4.7 ± 12.9	-5.7 ± 13.9	-4.0 ± 16.5	-0.3± 10.6
	CoV	0.0	-0.1	-0.3	-0.6	-1.4	-2.5	-2.8	-2.4	-4.2	-31.5
ActivPAL (Thigh)	PE±SD	-59.5 ± 29.7	-8.9 ± 16.0	-1.6 ± 9.1	-1.2 ± 9.5	-1.8 ± 9.3	-1.7± 9.7	-2.1 ± 10.4	-5.0 ± 15.1	-8.5 ± 23.6	-7.2 ± 26.2
	CoV	-0.5	-1.8	-5.7	-8.3	-5.1	-5.8	-5.0	-3.0	-2.8	-3.7
Actical (Waist)	PE±SD	-94.9 ± 10.5	-64.1 ± 26.6	-22.5 ± 17.8	-2.4 ± 7.5	-0.4 ± 4.7	0.1 ± 3.8	0.2 ± 2.1	0.3 ± 2.2	0.3 ± 4.3	2.0 ± 4.6
	CoV	-0.1	-0.4	-0.8	-3.1	-11.1	32.8	13.3	7.5	15.2	2.3
StepWatch (Ankle)	PE±SD	-14.6 ± 20.0	-0.6 ± 13.8	-1.5 ± 13.5	-1.3 ± 14.2	-1.8 ± 13.6	-1.1 ± 10.8	-3.0 ± 12.4	-8.0 ± 16.1	-20.4 ± 18.4	-23.5 ± 7.9
	CoV	-1.4	-22.2	-9.3	-11.1	-7.5	-9.8	-4.1	-2.0	-0.9	-0.3
Sensewear (Arm)	PE±SD	-98.7 ± 1.5	-96.8 ± 6.1	-87.2 ± 23.5	-73.7 ± 35.8	-67.0 ± 42.6	-61.4 ± 45.0	-60.9 ± 45.1	-58.6 ± 43.6	-60.4 ± 43.1	-56.2± 44.1
	CoV	0.0	-0.1	-0.3	-0.5	-0.6	-0.7	-0.7	-0.7	-0.7	-0.8
SW200 (Waist)	PE±SD	-84.9 ± 12.9	-63.6 ± 26.2	-32.0 ± 26.1	-13.8 ± 17.7	-5.0 ± 13.3	-2.3± 11.9	-0.8 ± 10.7	-0.8 ± 11.5	0.2 ± 2.0	1.8 ±4.7

	CoV	-0.2	-0.4	-0.8	-1.3	-2.7	-5.3	-12.7	-15.2	8.4	2.6
NL-1000 (Waist)	PE±SD	-91.4 ± 10.3	-78.6 ± 21.9	-34.9 ± 24.1	-10.7 ± 14.0	-3.4 ± 11.1	-2.8 ± 14.3	-1.3 ± 6.5	-0.7 ± 2.5	-0.3 ± 3.8	1.4 ± 4.6
	CoV	-0.1	-0.3	-0.7	-1.3	-3.2	-5.2	-5.1	-3.6	-11.2	3.3

Notes: All PE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 9: Bias: Percent error (PE) of the wearable technologies by speed for children (6-12 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (53)	**26.8** (53)	**40.2** (51)	**53.6** (49)	**67.0** (47)	**80.5** (42)	**93.9** (34)	**107.2** (24)	**120.7** (10)	**134.1** (0)
ActiGraph (Wrist)	PE±SD	-83.5 ± 11.2	-67.5 ± 15.5	-52.0 ± 14.9	-43.0 ± 18.9	-40.9 ± 20.3	-39.6 ± 20.1	-43.8 ± 17.7	-48.3 ± 14.1	-48.5 ± 3.9	## ± ##
	CoV	-0.1	-0.2	-0.3	-0.4	-0.5	-0.5	-0.4	-0.3	-0.1	#
ActiGraph (Waist)	PE±SD	-97.2 ± 5.7	-90.8 ± 7.7	-68.3 ± 15.7	-36.1 ± 17.6	-12.7 ± 11.7	-5.1 ± 9.3	-3.8 ± 6.0	-5.4 ± 7.4	-7.7 ± 21.3	## ± ##
	CoV	-0.1	-0.1	-0.2	-0.5	-0.9	-1.8	-1.6	-1.4	-2.8	#
ActivPAL (Thigh)	PE±SD	-57.3 ± 26.3	-9.6 ± 13.9	-1.4 ± 8.5	-0.3 ± 9.2	-1.4 ± 8.1	-1.2 ± 9.8	-2.4 ± 9.9	-8.1 ± 14.0	-10.1 ± 17.0	## ± ##
	CoV	-0.5	-1.4	-6.0	-26.7	-5.7	-7.8	-4.0	-1.7	-1.7	#
Actical (Waist)	PE±SD	-90.6 ± 13.9	-65.1 ± 21.9	-23.2 ± 14.9	-2.0 ± 5.8	-0.2 ± 1.3	0.5 ± 4.2	0.4 ± 1.2	0.4± 2.6	-0.7± 9.1	## ± ##
	CoV	-0.2	-0.3	-0.6	-2.9	-5.8	8.2	2.9	6.8	-13.6	#
StepWatch (Ankle)	PE±SD	-15.9 ± 19.6	-1.6 ± 14.6	-1.7 ± 14.4	-0.9 ± 16.0	-2.2 ± 14.7	-0.4 ± 6.8	-5.1 ± 10.6	-14.7 ± 15.5	-28.9 ± 10.5	## ± ##
	CoV	-1.2	-9.3	-8.3	-18.7	-6.7	-18.0	-2.1	-1.1	-0.4	#
Sensewear (Arm)	PE±SD	-98.4 ± 2.1	-95.4 ± 7.4	-86.4± 20.9	-75.7 ± 33.3	-67.1 ± 41.1	-60.5 ± 44.6	-63.8 ± 43.1	-60.5 ± 41.7	-82.4 ± 29.7	## ± ##
	CoV	0.0	-0.1	-0.2	-0.4	-0.6	-0.7	-0.7	-0.7	-0.4	#
SW200 (Waist)	PE±SD	-81.7 ± 16.0	-58.3 ± 28.1	-31.0 ± 26.0	-14.5 ± 21.5	-6.8 ± 17.9	-4.9 ± 18.0	-2.3 ± 17.4	-3.4 ± 20.5	1.0 ± 1.9	## ± ##

	CoV	-0.2	-0.5	-0.8	-1.5	-2.6	-3.7	-7.4	-6.0	1.9	#
NL-1000 (Waist)	PE±SD	-89.3 ± 13.2	-72.5 ± 25.2	-36.9 ± 17.2	-10.1 ± 8.9	-2.0 ± 3.8	-3.2 ± 16.1	-1.6 ± 9.9	-0.2 ± 1.7	-1.9 ± 8.0	## ± ##
	CoV	-0.1	-0.3	-0.5	-0.9	-1.9	-5.1	-6.2	-10.7	-4.3	#

Notes: All PE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 10: Percent error (PE) of the wearable technologies by speed for adolescents (13-17 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (40)	**26.8** (40)	**40.2** (40)	**53.6** (40)	**67.0** (39)	**80.5** (39)	**93.9** (36)	**107.2** (34)	**120.7** (25)	**134.1** (6)
ActiGraph (Wrist)	PE±SD	-94.0 ± 7.1	-74.0 ± 19.1	-49.1 ± 20.4	-36.1 ± 24.3	-36.1 ± 24.9	-36.5 ± 24.4	-41.1 ± 20.7	-47.3 ± 19.0	-50.8 ± 11.4	-44.1 ± 12.2
	CoV	-0.08	-0.26	-0.42	-0.67	-0.69	-0.67	-0.50	-0.40	-0.22	-0.28
ActiGraph (Waist)	PE±SD	-99.6 ± 0.9	-93.9 ± 8.5	-70.4 ± 18.6	-33.5 ± 20.7	-8.6 ± 14.5	-4.8 ± 12.9	-3.3± 8.3	-4.8 ± 9.5	-0.6 ± 2.3	-5.2 ± 15.2
	CoV	-0.01	-0.09	-0.26	-0.62	-1.68	-2.66	-2.52	-1.96	-3.93	-2.92
ActivPAL (Thigh)	PE±SD	-59.3 ± 33.2	-8.0 ± 20.5	-2.7 ± 11.8	-2.8 ± 12.2	-3.3 ± 12.6	-3.2 ± 12.0	-2.5 ± 13.6	-6.3 ± 19.6	-14.3 ± 31.6	-5.6 ± 15.7
	CoV	-0.56	-2.57	-4.37	-4.32	-3.83	-3.81	-5.35	-3.11	-2.21	-2.83
Actical (Waist)	PE±SD	-98.4 ± 4.0	-64.7 ± 26.6	-25.1 ± 18.5	-3.6 ± 10.1	-1.2 ± 7.5	-0.4 ± 3.7	0.4 ± 1.1	0.3 ± 1.2	0.7 ± 1.2	0.9 ± 0.6
	CoV	-0.04	-0.41	-0.74	-2.79	-6.38	-10.06	2.72	3.98	1.63	0.66
StepWatch (Ankle)	PE±SD	-14.6 ± 18.1	0.9 ± 2.2	0.3 ± 2.6	0.0 ± 3.0	-0.1 ± 3.2	0.1 ± 1.6	0.4 ± 1.1	-2.2 ± 6.9	-18.9 ± 14.7	-20.3 ± 9.3
	CoV	-1.24	2.44	8.50	-67.60	-54.47	12.01	3.03	-3.18	-0.78	-0.46
Sensewear (Arm)	PE±SD	-98.9 ± 0.7	-97.6 ± 5.7	-89.6 ± 22.3	-77.2 ± 34.0	-72.6 ± 41.1	-65.6 ± 44.5	-61.8 ± 46.1	-60.4 ± 43.5	-57.8 ± 43.4	-50.3 ± 43.2
	CoV	-0.01	-0.06	-0.25	-0.44	-0.57	-0.68	-0.75	-0.72	-0.75	-0.86
SW200 (Waist)	PE±SD	-88.3 ± 6.8	-66.1 ± 23.5	-33.0 ± 24.3	-13.1 ± 13.4	-3.5 ±7.7	-0.3 ± 3.9	0.5 ± 1.4	0.5 ± 3.4	0.2 ± 2.2	0.8 ± 0.8
	CoV	-0.08	-0.36	-0.74	-1.02	-2.16	-15.07	2.91	6.99	8.81	1.05

NL-1000 (Waist)	PE±SD	-93.7 ± 3.1	-82.9 ± 17.9	-36.0 ± 27.3	-11.8 ± 17.8	-2.6 ± 7.7	-1.0 ± 3.6	-0.9 ± 3.2	-0.7 ± 2.5	0.4 ± 1.8	0.5 ± 0.5
	CoV	-0.03	-0.22	-0.76	-1.51	-3.00	-3.56	-3.58	-3.38	4.91	1.49

Notes: All PE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 11: Percent error (PE) of the wearable technologies by speed for young adults (18-20 years)

		Slow				Normal				Fast	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (24)	**26.8** (24)	**40.2** (24)	**53.6** (24)	**67.0** (24)	**80.5** (23)	**93.9** (23)	**107.2** (23)	**120.7** (18)	**134.1** (10)
ActiGraph (Wrist)	PE±SD	-92.1 ± 11.5	-70.9 ± 22.4	-45.3 ± 22.6	-32.7 ± 22.9	-33.7 ± 24.7	-37.9 ± 23.3	-38.5 ± 22.6	-39.4 ± 21.7	-46.1 ± 20.0	-48.7 ± 2.9
	CoV	-0.1	-0.3	-0.5	-0.7	-0.7	-0.6	-0.6	-0.5	-0.4	-0.1
ActiGraph (Waist)	PE±SD	-99.4 ± 2.1	-94.5 ± 7.9	-70.9 ± 22.9	-31.7 ± 22.9	-12.0 ± 23.3	-8.2 ± 22.7	-8.1 ± 22.6	-7.3 ± 22.5	-6.6 ± 23.5	2.6 ± 5.8
	CoV	0.0	-0.1	-0.3	-0.7	-1.9	-2.8	-2.8	-3.1	-3.6	2.2
ActivPAL (Thigh)	PE±SD	-64.5 ± 31.5	-8.7 ± 11.9	-0.1 ± 2.9	0.0 ± 3.0	-0.2 ± 2.3	0.0 ± 1.6	-0.7 ± 2.6	0.0 ± 2.2	0.6 ± 1.3	-8.2 ± 31.7
	CoV	-1.4	-23.6	-279.0	-11.9	38.5	-3.5	-1230.5	2.2	-3.9	-1.4
Actical (Waist)	PE±SD	-98.6 ± 4.4	-47.5± 32.1	-16.6 ± 21.2	-1.2 ± 5.4	0.4 ± 2.7	0.2 ± 3.0	-0.6 ± 3.7	0.2 ± 2.8	0.2 ± 3.1	2.7 ± 5.8
	CoV	0.0	-0.7	-1.3	-4.4	6.8	15.0	-6.3	16.2	16.5	2.2
StepWatch (Ankle)	PE±SD	-11.9 ± 24.1	-1.1 ± 21.6	-3.8 ± 20.8	-4.2 ± 20.6	-4.0 ± 20.5	-4.5 ± 21.0	-5.1 ± 21.0	-9.5 ± 22.6	-17.8 ± 24.8	-25.5 ± 6.6
	CoV	-2.0	-19.5	-5.5	-4.9	-5.1	-4.6	-4.1	-2.4	-1.4	-0.3
Sensewear (Arm)	PE±SD	-99.2 ± 0.7	-98.6 ± 1.2	-85.0 ± 30.3	-63.9 ± 42.8	-57.8 ± 47.3	-55.7 ± 47.7	-55.1 ± 47.7	-53.9 ± 47.2	-51.8 ± 46.7	-59.8 ± 46.6
	CoV	0.0	0.0	-0.4	-0.7	-0.8	-0.9	-0.9	-0.9	-0.9	-0.8
SW200(Waist)	PE±SD	-86.3 ± 11.5	-71.3 ± 24.4	-32.6 ± 29.9	-13.3 ± 16.1	-3.7 ± 9.8	-0.7 ± 2.8	-0.7 ± 4.1	0.2 ± 3.4	-0.2 ± 1.7	2.4 ± 6.0
	CoV	-0.1	-0.3	-0.9	-1.2	-2.7	-3.9	-6.1	21.0	-7.8	2.5

NL-1000 (Waist)	PE±SD	-92.4 ± 10.3	-84.8 ± 17.2	-28.7 ± 30.4	-10.0 ± 15.6	-7.7 ± 20.8	-5.1 ± 20.9	-1.4 ± 3.6	-1.2 ± 3.2	-0.5 ± 1.6	1.9 ± 5.9
	CoV	-0.1	-0.2	-1.1	-1.6	-2.7	-4.1	-2.6	-2.6	-3.5	3.1

Notes: All PE values are presented as mean % ± SD %. CoVs are presented as percentages. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Table 12: Validity parameters (accuracy, precision, and bias) by wear location averaged across all wearable technologies

Ankle (StepWatch)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
Accuracy:	**MAPE±SD**	14.9 ± 19.8	4.4 ± 13.1	3.4 ± 13.2	3.6 ± 13.8	3.1 ± 13.4	2.9 ± 10.4	4.1 ± 12.0	8.7 ± 15.7	20.7 ± 18.0	23.5 ± 7.9
Precision:	***r***	0.82 (0.54, 1.00)									
Bias:	**PE±SD**	-14.6 ± 20.0	-0.6 ± 13.8	-1.5 ± 13.5	-1.3 ± 14.2	-1.8 ± 13.6	-1.1 ± 10.8	-3.0 ± 12.4	-8.0 ± 16.1	-20.4 ± 18.4	-23.5 ± 7.9
Thigh (ActivPAL)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
Accuracy:	**MAPE ±SD**	60.2 ± 28.2	10.9 ± 14.7	3.2 ± 8.7	3.5 ± 9.0	3.2 ± 8.9	3.3 ± 9.2	3.7 ± 9.9	6.0 ± 14.7	9.8 ± 23.1	11.1 ± 24.7
Precision:	***r***	0.86 (0.59, 1.00)									
Bias:	**PE±SD**	-59.5 ± 29.7	-8.9 ± 16.0	-1.6 ± 9.1	-1.2 ± 9.5	-1.8 ± 9.3	-1.7 ± 9.7	-2.1 ± 10.4	-5.0 ± 15.1	-8.5 ± 23.6	-7.2 ± 26.2
Waist (ActiGraph, Actical, SW200, NL-1000)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)

Accuracy:	**MAPE±SD**	92.4 ± 11.2	74.1 ± 25.2	40.7 ± 26.8	16.6 ± 18.3	5.9 ± 12.1	4.1 ± 11.6	2.7 ± 8.9	3.1 ± 9.1	2.6 ± 8.6	2.6 ± 6.1
Precision:	***r***	0.89 (0.75, 1.00)									
Bias:	**PE±SD**	-92.4 ± 11.2	-74.1 ± 25.3	-39.7 ± 28.2	-15.3 ± 19.4	-5.0 ± 12.5	-2.6 ± 12.0	-1.7 ± 9.2	-1.7 ± 9.4	-0.9 ± 8.9	1.2 ± 6.6
Wrist/Arm (ActiGraph, SenseWear)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
Accuracy:	**MAPE±SD**	93.8 ± 9.4	83.6 ± 19.0	68.5 ± 28.1	56.2 ± 34.3	52.4 ± 37.1	50.1 ± 36.8	51.2 ± 36.1	52.2 ± 33.8	54.9 ± 32.0	53.2 ± 28.7
Precision:	***r***	0.53 (0.34, 0.72									
Bias:	**PE±SD**	-93.8 ± 9.4	-83.6 ± 19.0	-68.4 ± 28.3	-56.1 ± 34.5	-52.3 ± 37.2	-49.7 ± 37.3	-51.1 ± 36.2	-52.0 ± 34.1	-54.6 ± 32.5	-51.6 ± 31.5

Notes: All MAPE and PE values are presented as mean % ± SD %. CoVs are presented as percentages. All correlation coefficients are presented as *r* values to two decimal places with 95% CIs ranging above and below the point value. 95% CIs will also be reported to two decimal places. N represents the number of participants in the analytical sample, as not every participant completed every speed.

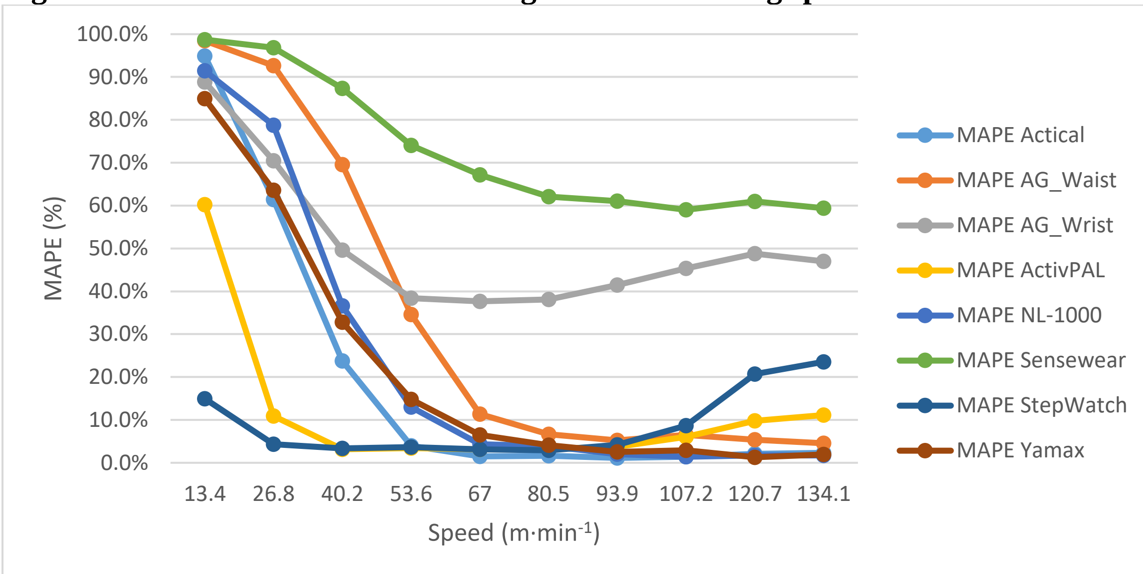
Table 13: Summary validity parameters (accuracy, precision, and bias) by age averaged across all wearable technologies

Children (6-12 years)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
Accuracy:	**MAPE±SD**	76.7 ± 30.0	58.2 ± 36.2	38.0 ± 32.6	24.1 ± 30.1	17.3 ± 29.3	16.0 ± 28.3	16.3 ± 28.9	18.5 ± 28.1	24.1 ± 30.2	## ± ##
Precision:	***r***	0.76 (0.63,0.90)									
Bias:	**PE±SD**	-76.7 ± 30.0	-57.6 ± 37.1	-37.6 ± 33.1	-22.8 ± 31.0	-16.7 ± 29.6	-14.3 ± 29.2	-15.3 ± 29.4	-17.5 ± 28.8	-22.4 ± 31.5	## ± ##
Adolescents (13-17 years)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)
Accuracy:	**MAPE±SD**	81.2 ± 30.0	61.7 ± 37.7	39.4 ± 33.6	23.6 ± 29.9	17.0 ± 30.0	14.9 ± 29.1	14.6 ± 28.7	16.0± 28.8	18.6 ± 29.3	16.5 ± 25.1
Precision:	***r***	0.81 (0.68, 0.94)									
Bias:	**PE±SD**	-80.9 ± 31.2	-60.8 ± 39.1	-38.2 ± 35.0	-22.3 ± 30.9	-16.0 ± 30.5	-14.0 ± 29.6	-13.5 ± 29.2	-15.1 ± 29.3	-17.6 ± 29.9	-15.4 ± 25.8
Young Adults (18-20 years)											
		Slow				**Normal**				**Fast**	
	Speed ($m \cdot min^{-1}$) (N)	**13.4** (117)	**26.8** (117)	**40.2** (115)	**53.6** (113)	**67.0** (110)	**80.5** (104)	**93.9** (93)	**107.2** (81)	**120.7** (53)	**134.1** (16)

Accuracy:	**MAPE±SD**	80.6 ± 31.9	60.6 ± 38.7	36.9 ± 35.5	21.0 ± 28.8	15.8 ± 29.4	14.9 ± 29.5	14.4 ± 28.7	14.9 ± 28.4	16.2 ± 28.9	20.4 ± 28.4
Precision:	***r***	0.82 (0.68, 0.95)									
Bias:	**PE±SD**	-80.6 ± 32.1	-59.7 ± 40.0	-35.4 ± 37.1	-19.6 ± 29.8	-14.8 ± 29.9	-14.8 ± 29.9	-13.8 ± 29.1	-13.9 ± 28.9	-15.3 ± 29.4	-16.6 ± 30.9

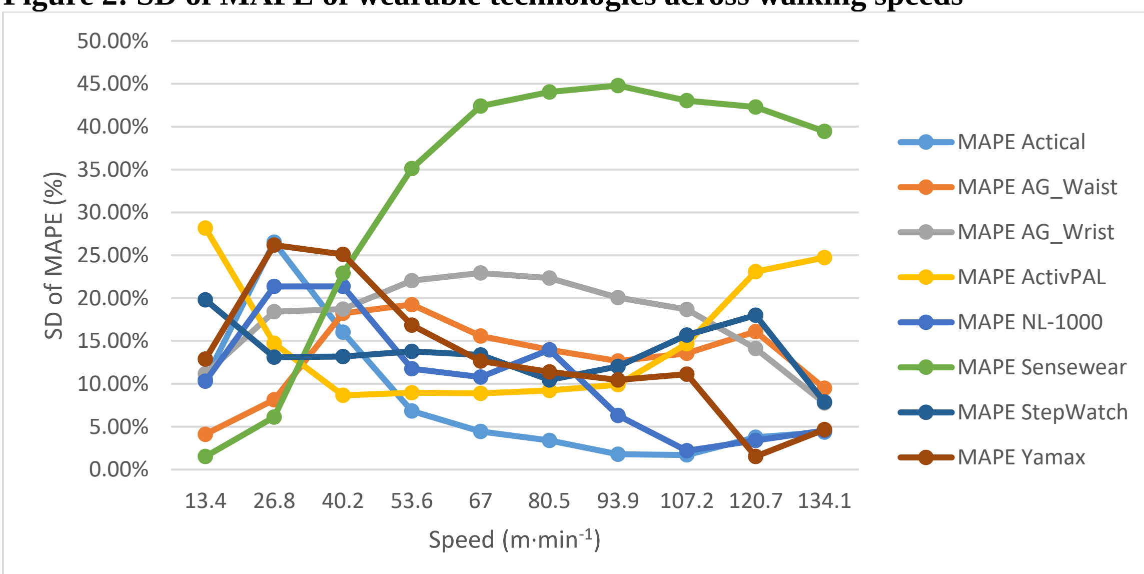
Notes: All MAPE and PE values are presented as mean % ± SD %. CoVs are presented as percentages. All correlation coefficients are presented as *r* values to two decimal places with 95% CIs ranging above and below the point value. 95% CIs will also be reported to two decimal places. N represents the number of participants in the analytical sample, as not every participant completed every speed.

Figure 1: MAPE of wearable technologies across walking speeds



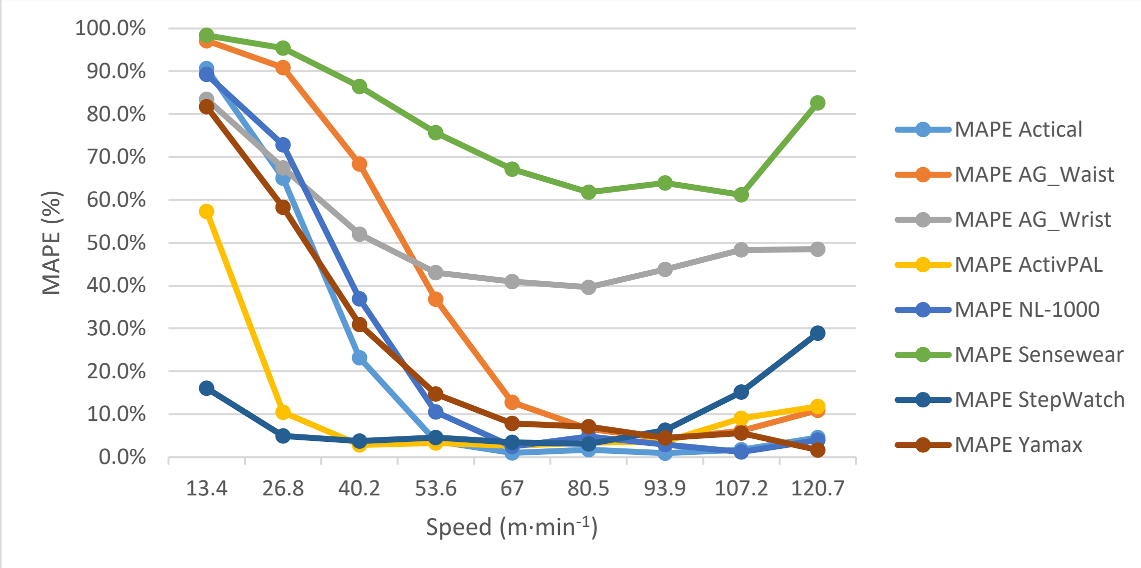
Notes: MAPE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decreases in MAPE indicate improved accuracy.

Figure 2: SD of MAPE of wearable technologies across walking speeds



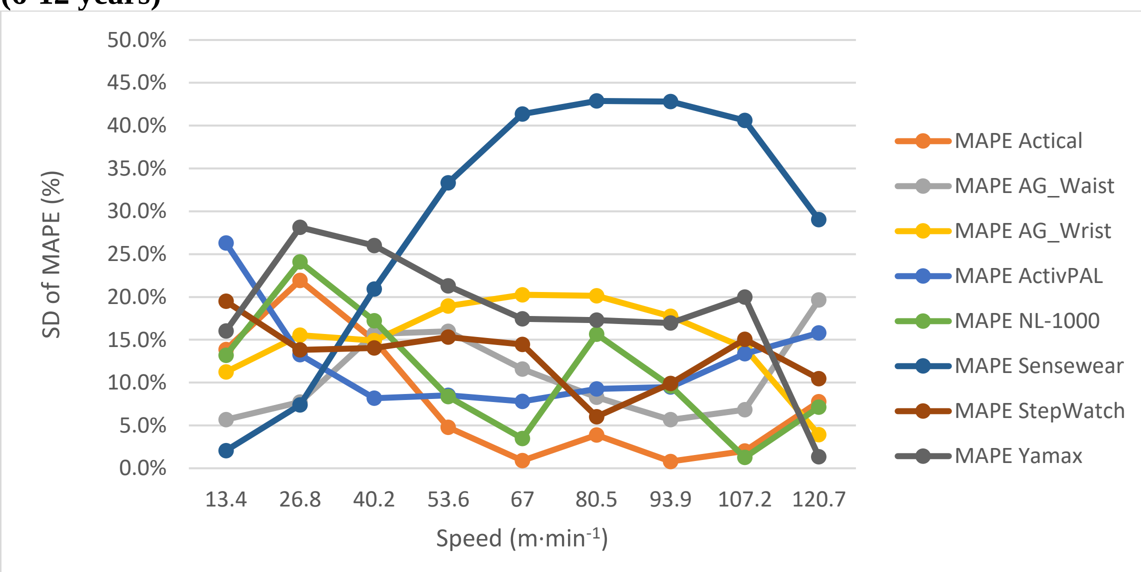
Notes: SD of respective MAPE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 3: MAPE of wearable technologies across walking speeds for children (6-12 years)



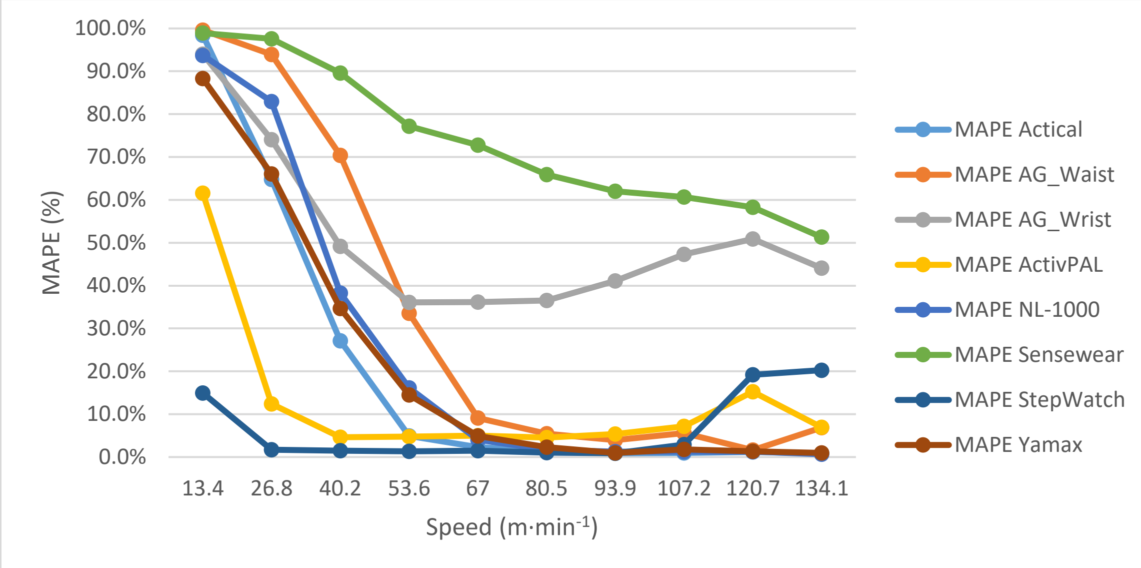
Notes: MAPE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Speed (m·min^{-1}) is on the X-axis and MAPE (%) is on the Y-axis. Decreases in MAPE indicate improved accuracy.

Figure 4: SD of MAPE of wearable technologies across walking speeds for children (6-12 years)



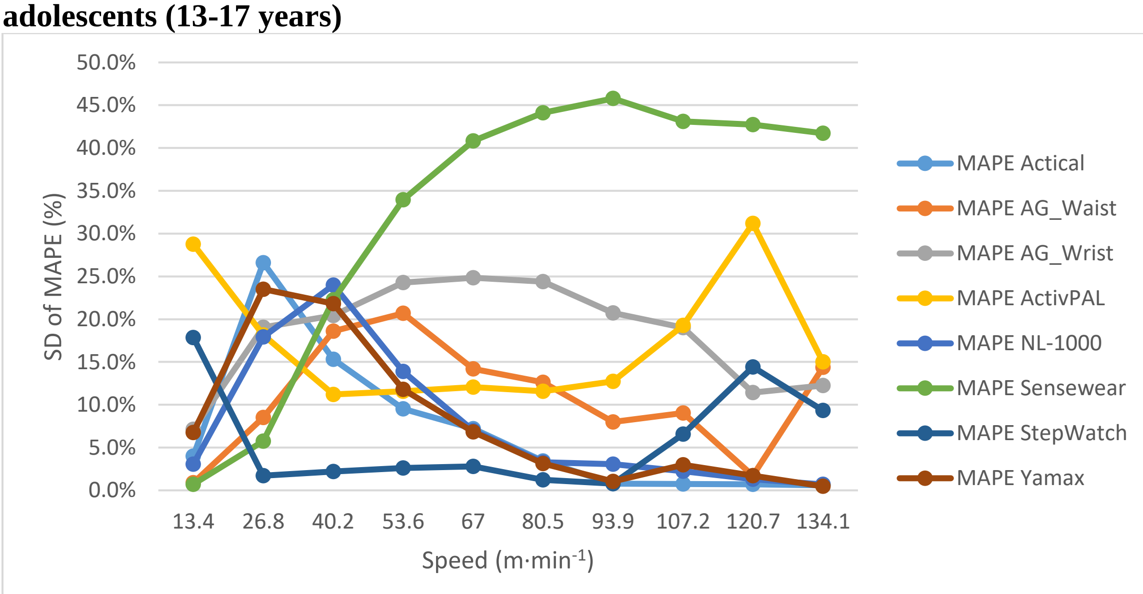
Notes: SD of respective MAPE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 5: MAPE of wearable technologies across walking speeds for adolescents (13-17 years)



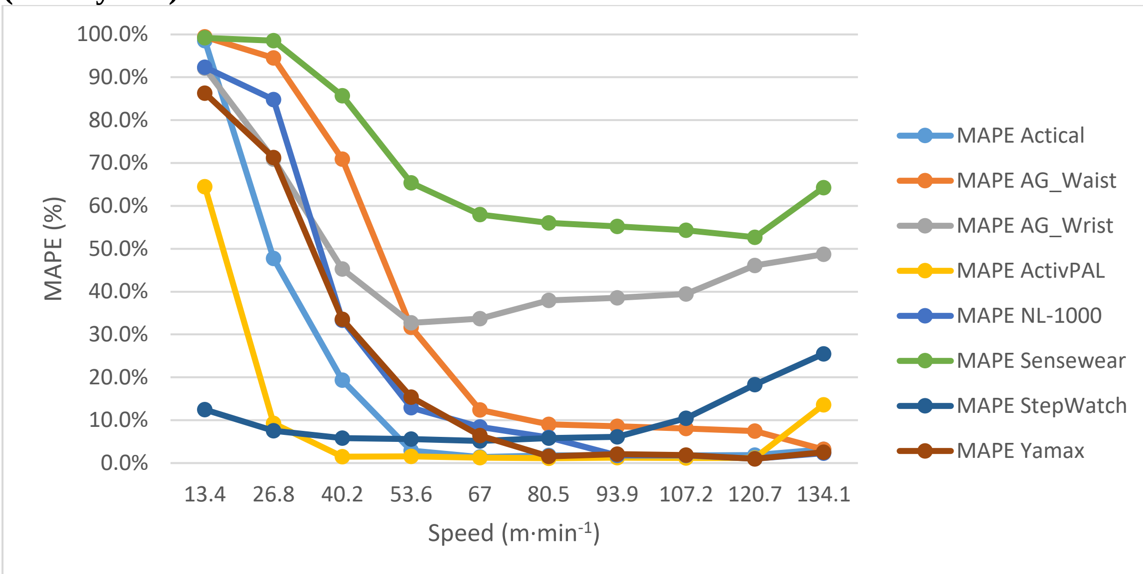
Notes: MAPE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decreases in MAPE indicate improved accuracy.

Figure 6: SD of MAPE of wearable technologies across walking speeds for adolescents (13-17 years)



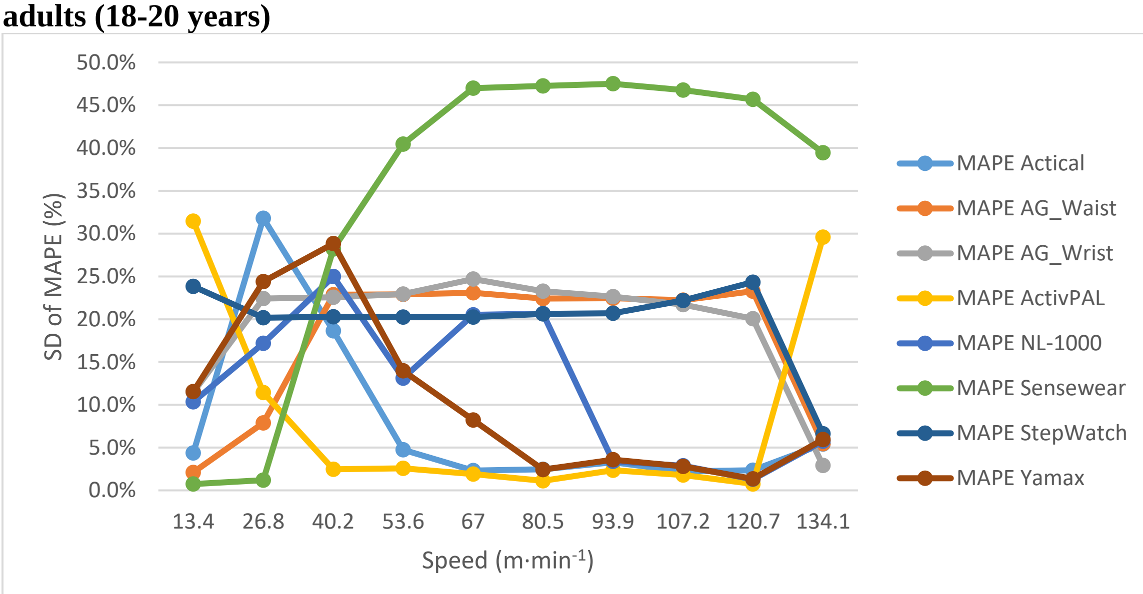
Notes: SD of respective MAPE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 7: MAPE of wearable technologies across walking speeds for young adults (18-20 years)



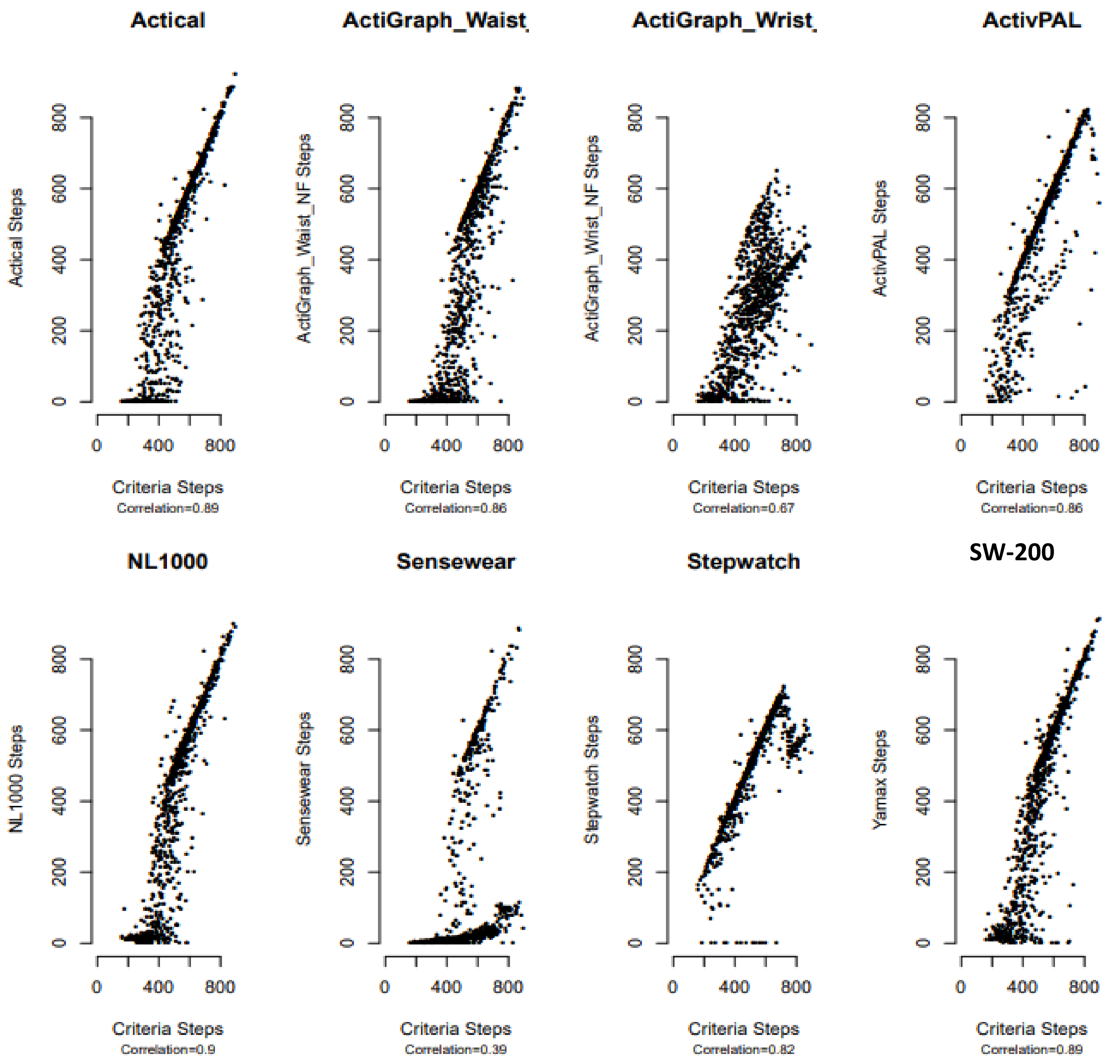
Notes: MAPE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decreases in MAPE indicate improved accuracy.

Figure 8: SD of MAPE of wearable technologies across walking speeds for young adults (18-20 years)



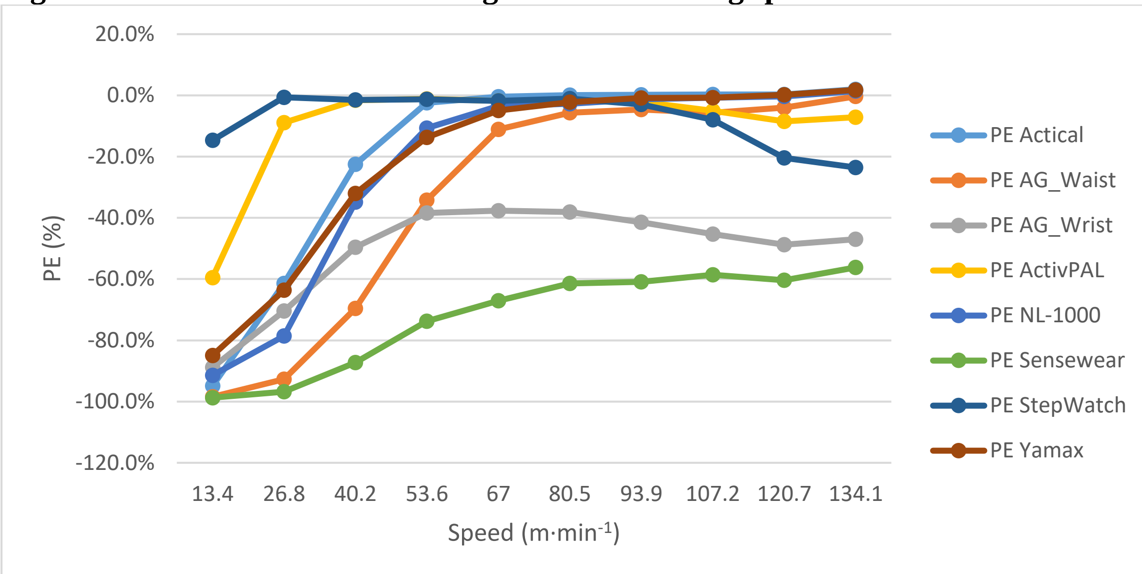
Notes: SD of respective MAPE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 9: Correlations of each wearable technology across all walking bouts



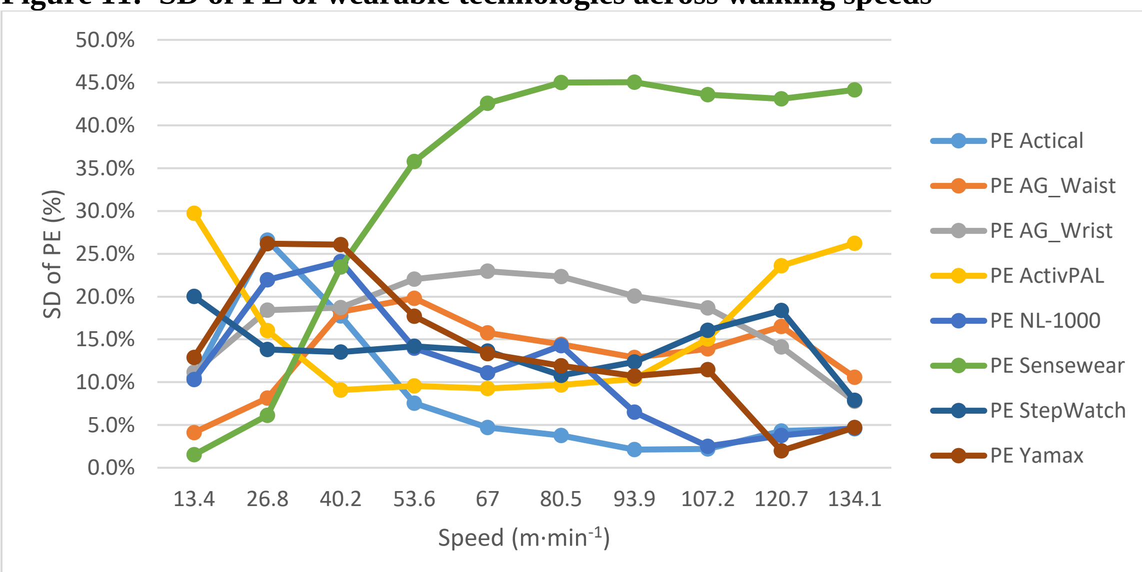
Notes: The X-Y scatterplot is representative of how tightly the wearable technology step counts hold to a linear relationship to directly observed steps across all walking bouts. X-axis is directly observed steps, and Y axis is wearable technology steps. Correlation coefficients closer to 1.0 indicate tighter relationship (more precise) to directly observed steps. Actical= 0.89, ActiGraph Waist= 0.86, ActiGraph Wrist= 0.67, ActivPAL= 0.86, NL-1000=0.90, Sensewear= 0.39, StepWatch= 0.82, Yamax SW -200= 0.89

Figure 10: PE of wearable technologies across walking speeds



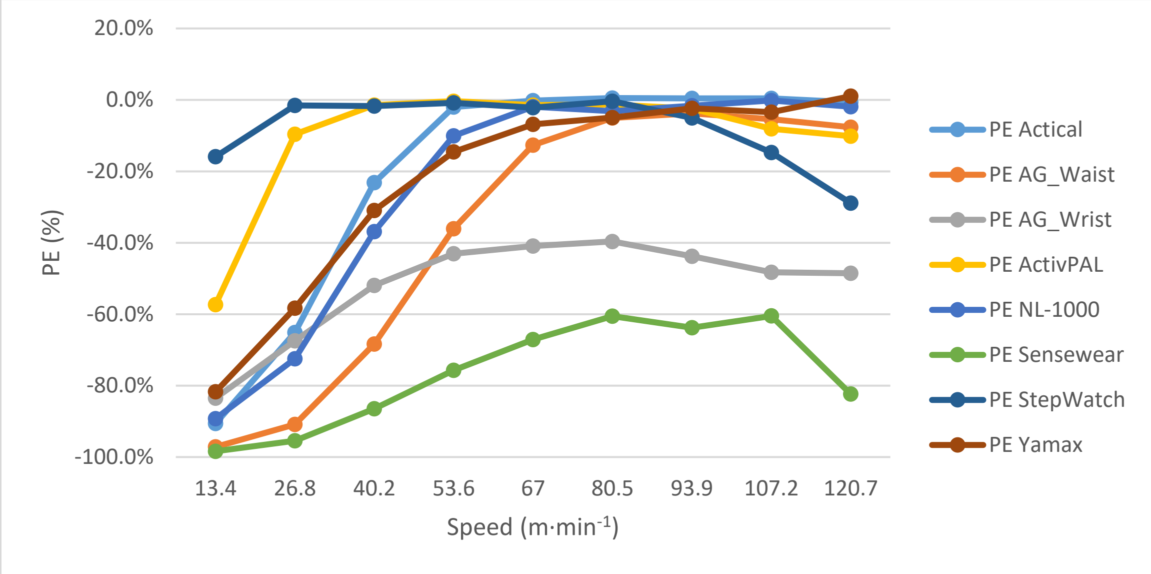
Notes: PE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. PE values closer of 0% indicate improved bias.

Figure 11: SD of PE of wearable technologies across walking speeds



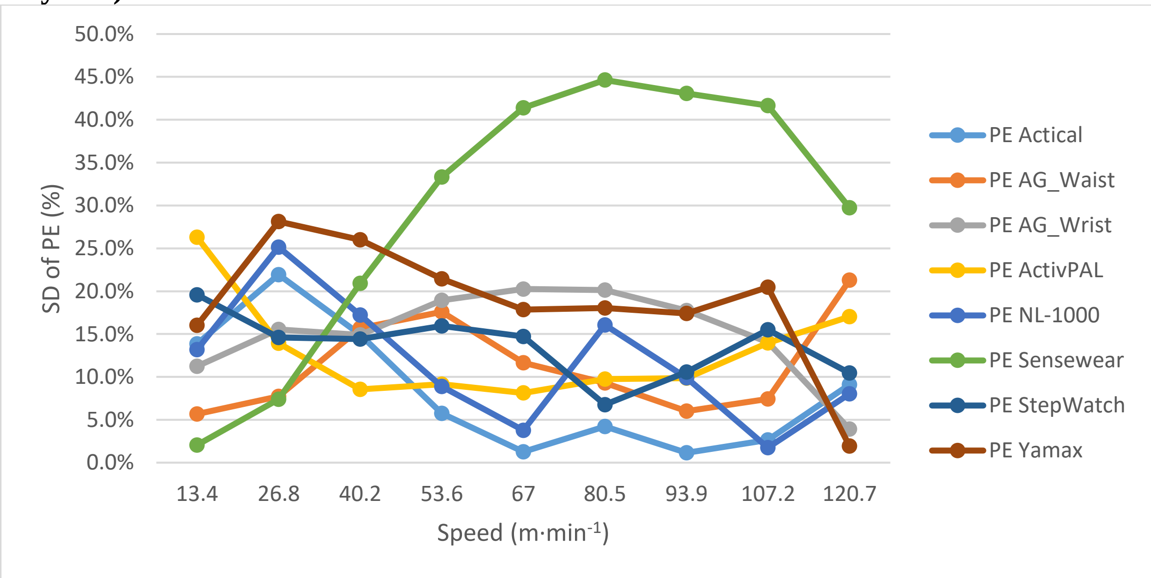
Notes: SD of respective PE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 12: PE of wearable technologies across walking speeds for children (6-12 years)



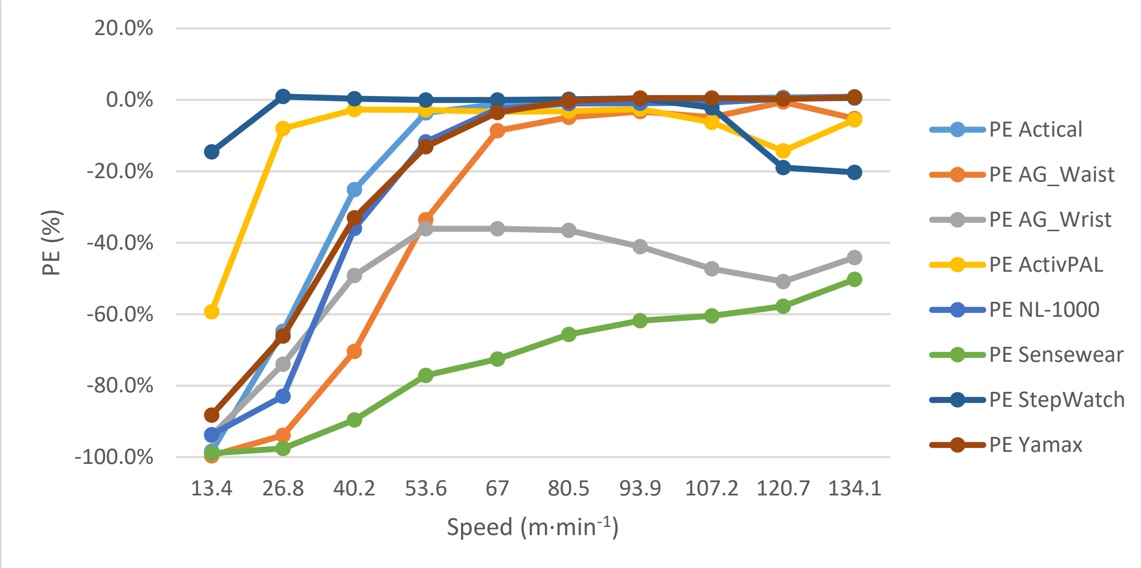
Notes: PE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. PE values closer of 0% indicate improved bias.

Figure 13: SD of PE of wearable technologies across walking speeds for children (6-12 years)



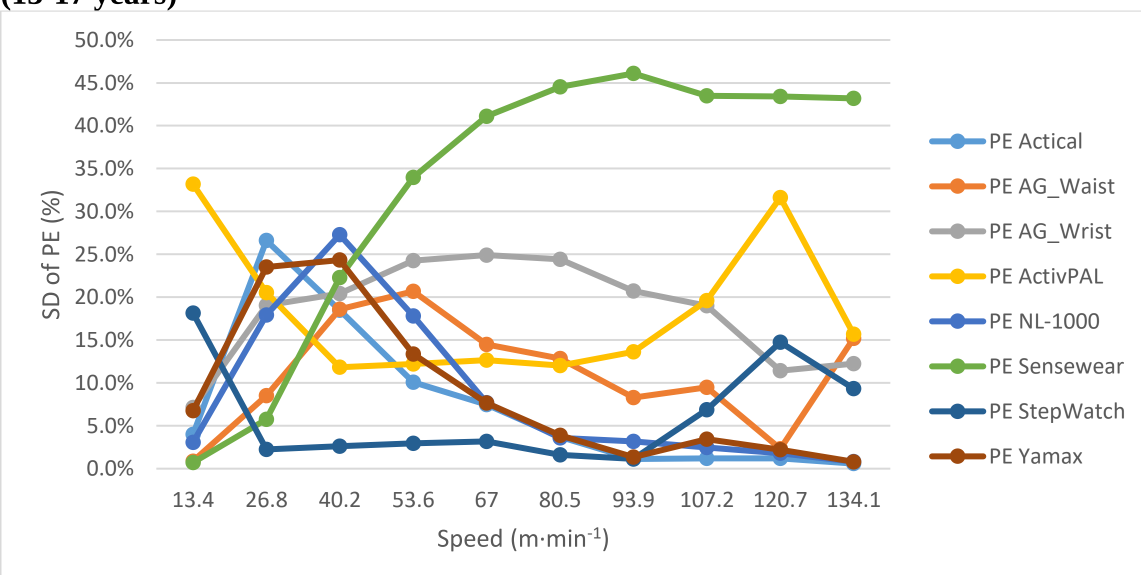
Notes: SD of respective PE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 14: PE of wearable technologies across walking speeds for adolescents (13-17 years)



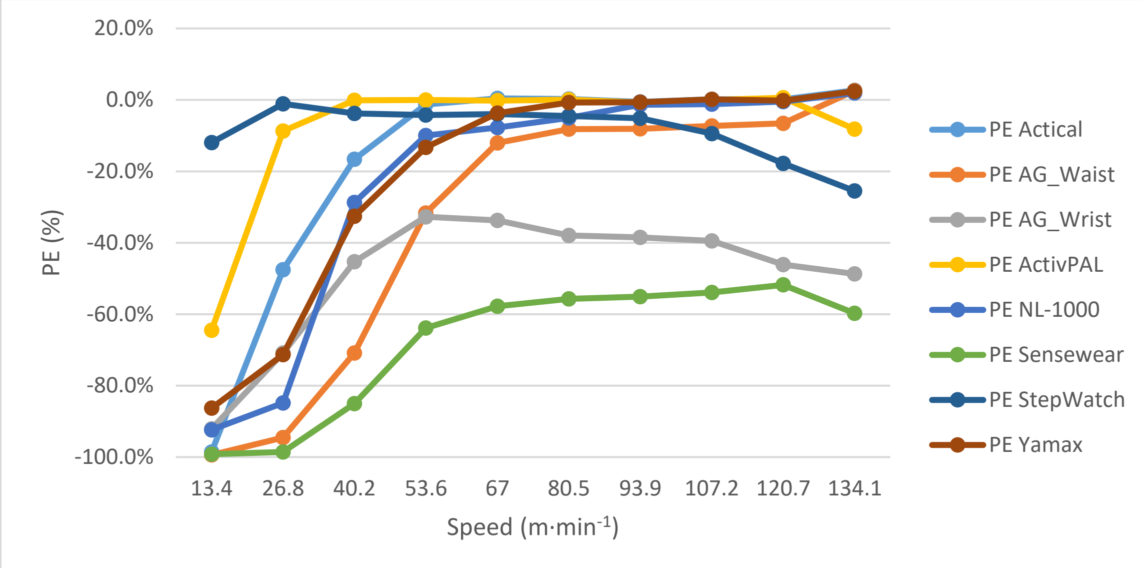
Notes: PE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. PE values closer of 0% indicate improved bias.

Figure 15: SD of PE of wearable technologies across walking speeds for adolescents (13-17 years)



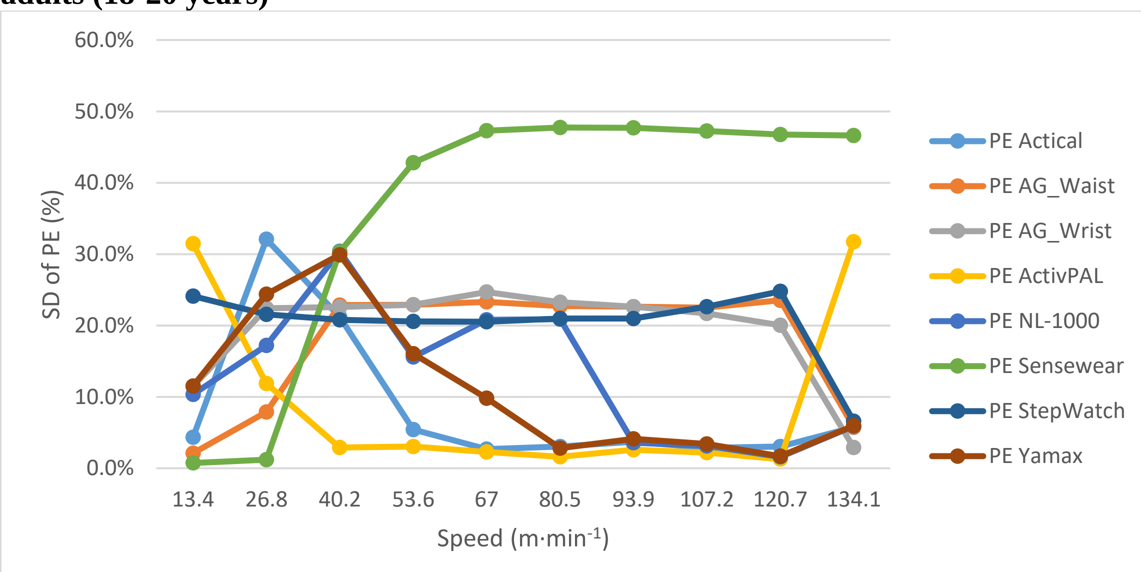
Notes: SD of respective PE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 16: PE of wearable technologies across walking speeds for young adults (18-20 years)



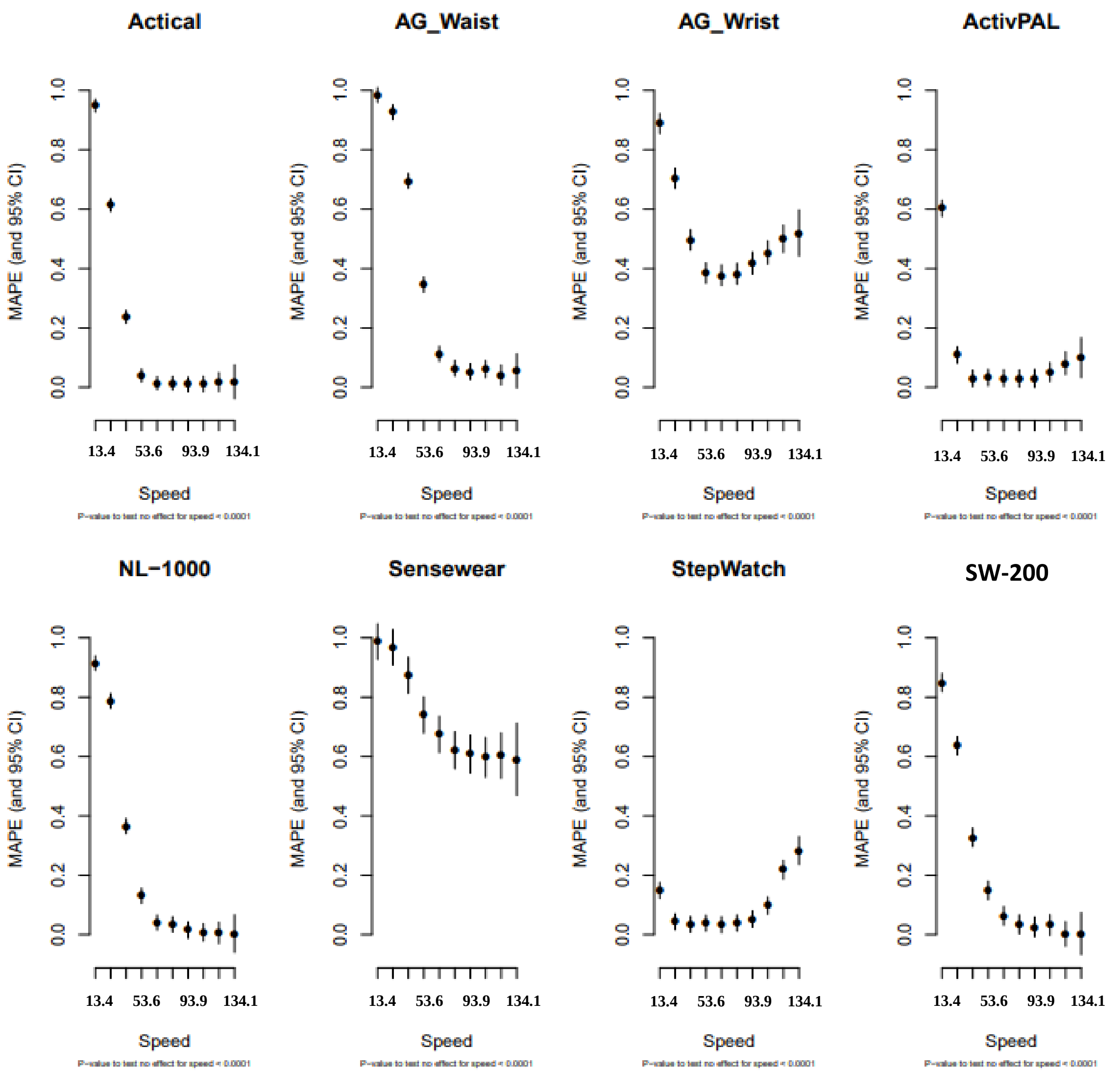
Notes: PE values of each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. PE values closer of 0% indicate improved bias.

Figure 17: SD of PE of wearable technologies across walking speeds for young adults (18-20 years)



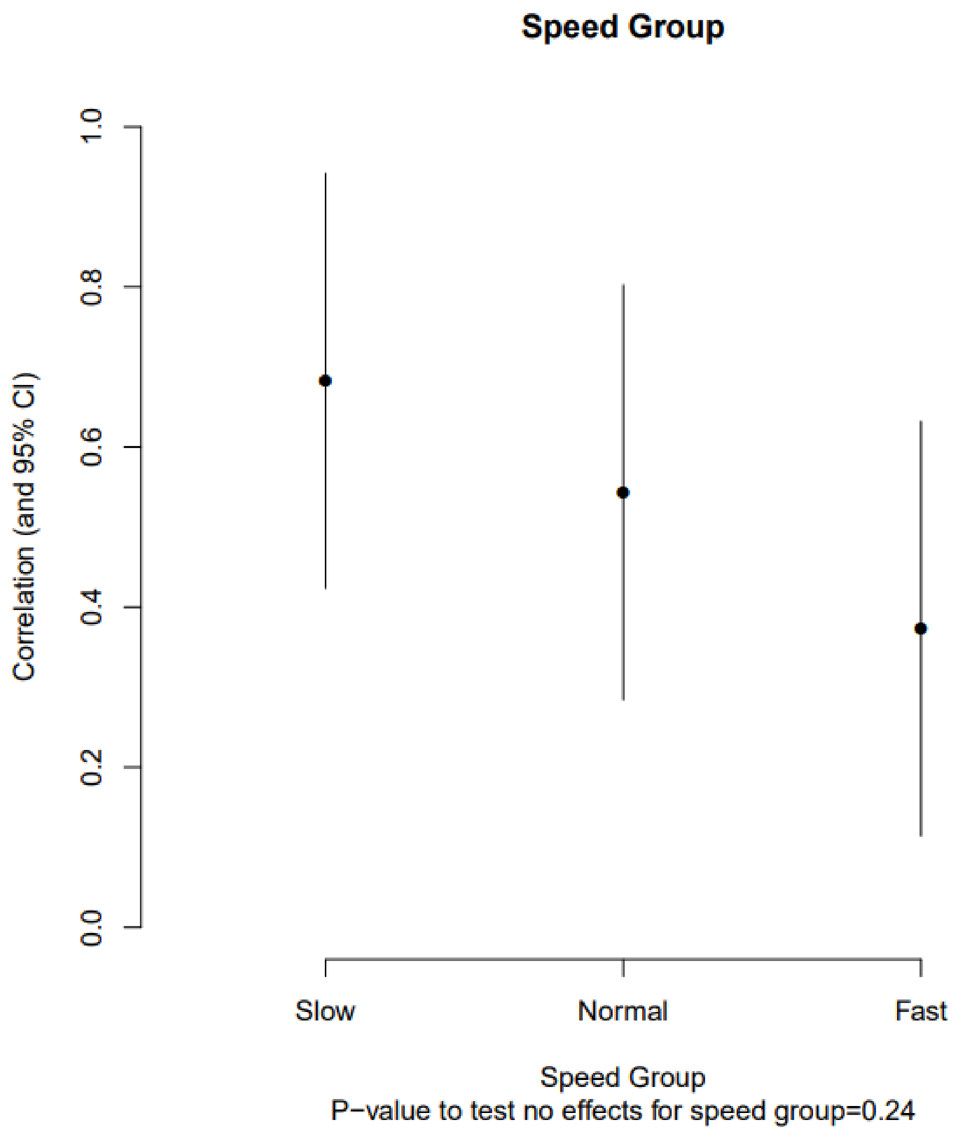
Notes: SD of respective PE values for each technology across the individual walking speeds. Each line is representative of grouped average for a single technology. Decrease in SD indicates more precise measures.

Figure 18: The effect of speed on the accuracy of wearable technology's step counting ability



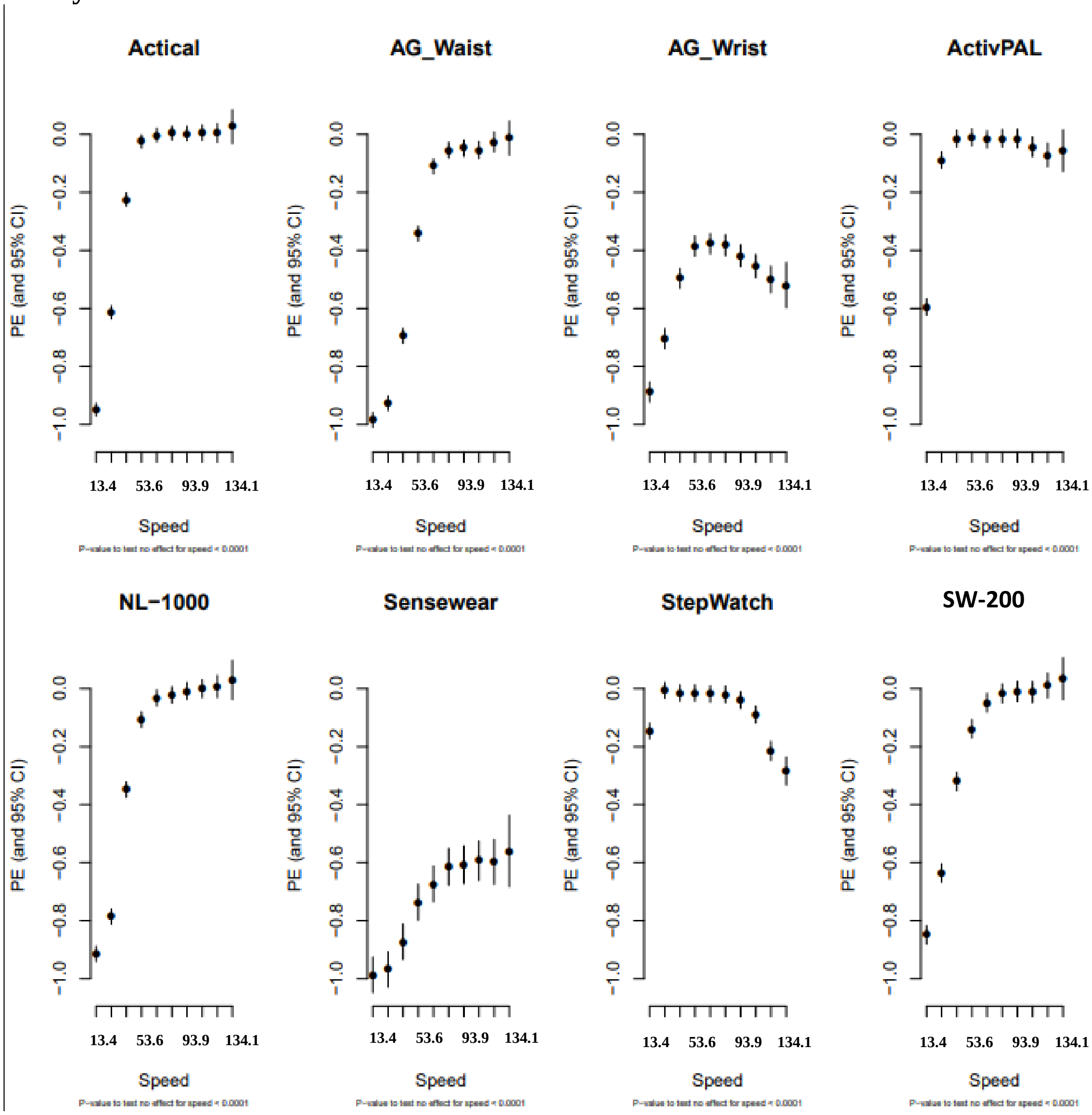
Notes: MAPE (%) and corresponding 95% CIs respective to each technology are plotted across speeds. Each point represents grouped averages of MAPE values, with 95% CIs extending above and below that point estimate. 95% CIs that do not overlap indicate significant differences, while those that do overlap indicate no significant differences. Speed had a significant effect on all eight technologies (p<0.0001).

Figure 19: The effect of speed on precision of wearable technologies step counting ability



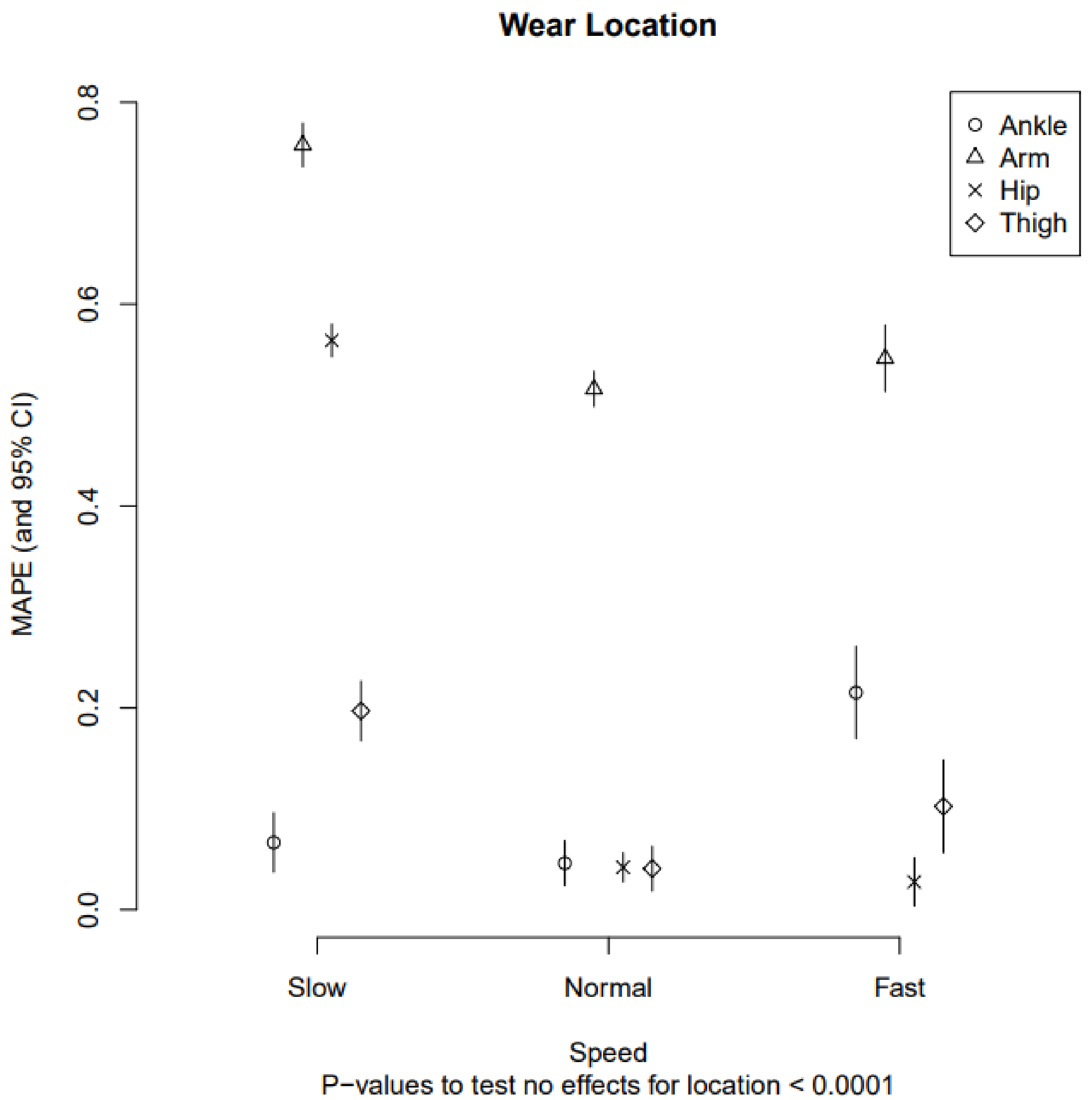
Notes: Correlation coefficients and corresponding 95% CIs for slow normal and fast walking. Correlation coefficients across each technology were averaged across each speed range. X axis refers to the speed group and Y axis is the correlation coefficient. Correlation coefficients closer to 1.0 indicate increased precision. 95% CIs overlap across speeds, indicating no significant difference.

Figure 20: The effect of speed on the bias of wearable technology's step counting ability



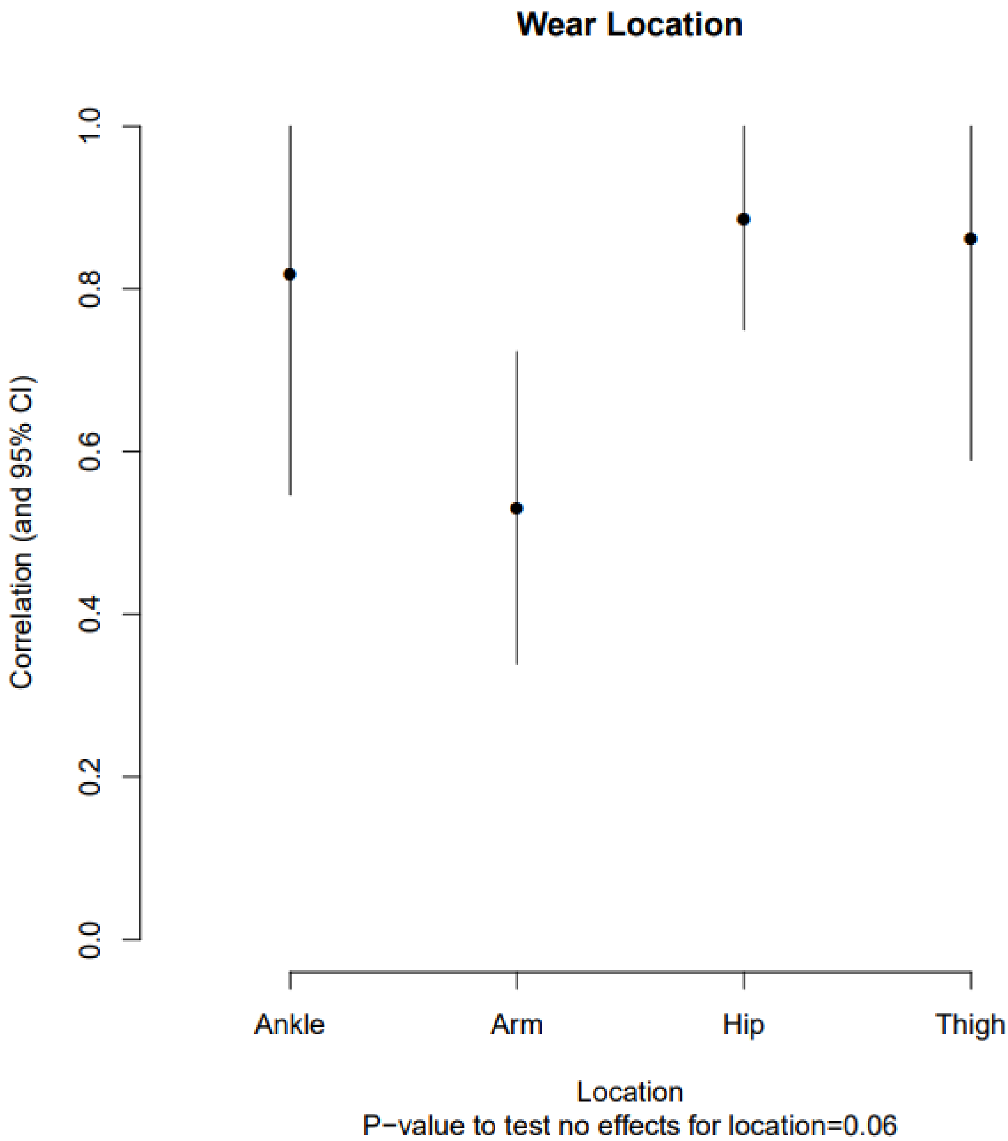
Notes: PE (%) and corresponding 95% CIs respective to each technology are plotted across speeds. Each point represents grouped averages of PE values, with 95% CIs extending above and below that point estimate. 95% CIs that do not overlap indicate significant differences, while those that do overlap indicate no significant differences. Speed had a significant effect on all seight technologies (p<0.0001).

Figure 21: The effect of wear location on the overall accuracy of wearable technology's step counting abilities



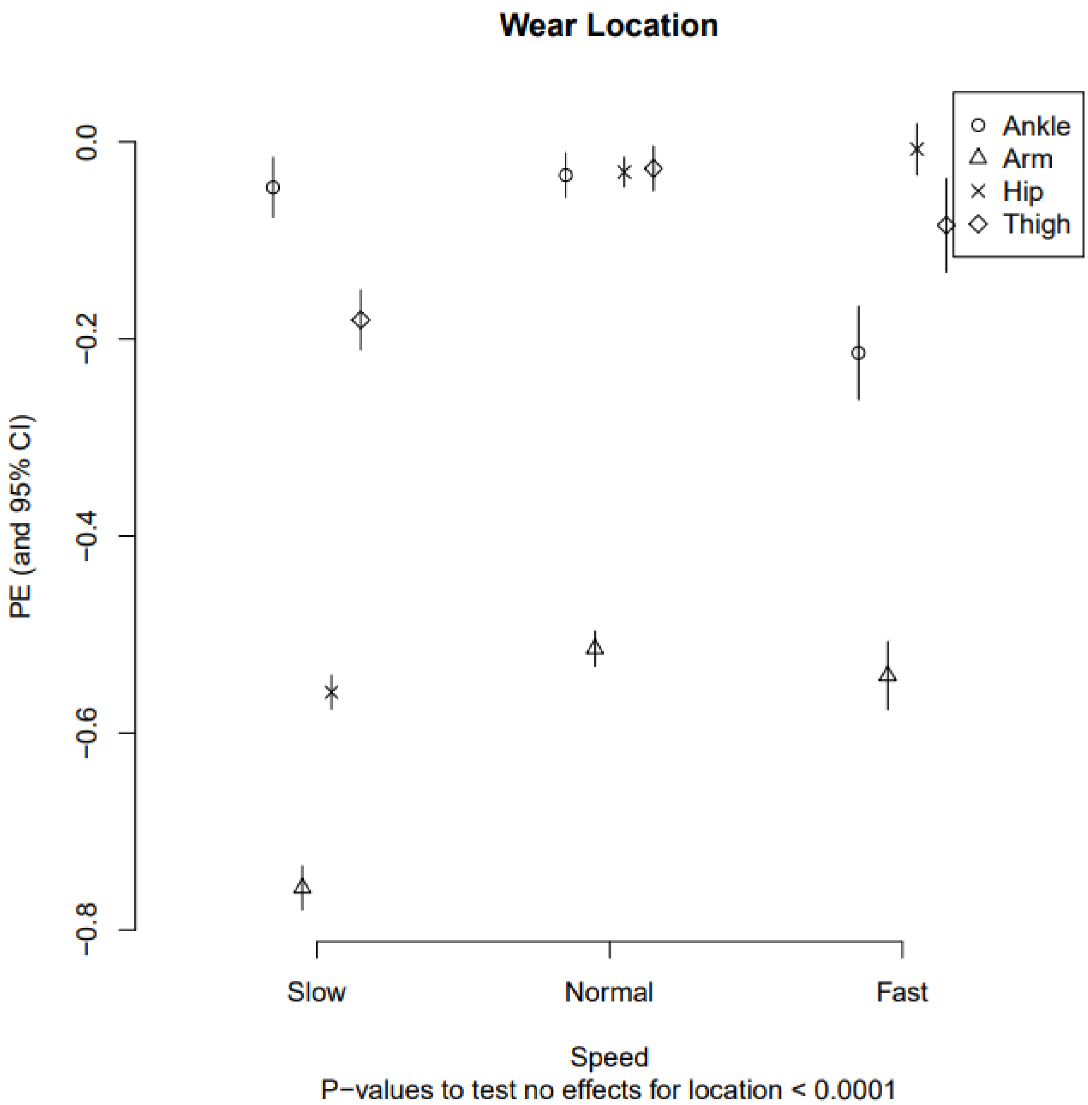
Notes: MAPE and corresponding 95% CIs of each wear location are presented at slow, normal, and fast walking speeds. MAPE values are averaged across technologies respective to each wear location for slow, normal, and fast walking speeds. The X axis refers to the speed group, and Y axis refers to MAPE values. MAPE values closer to 0% indicate greater accuracy. Wear location has a significant effect (p>0.0001) on technology accuracy. Further where 95% CIs do not overlap, there are significant differences between locations

Figure 22: The effect of wear location on the overall precision of wearable technology's step counting ability



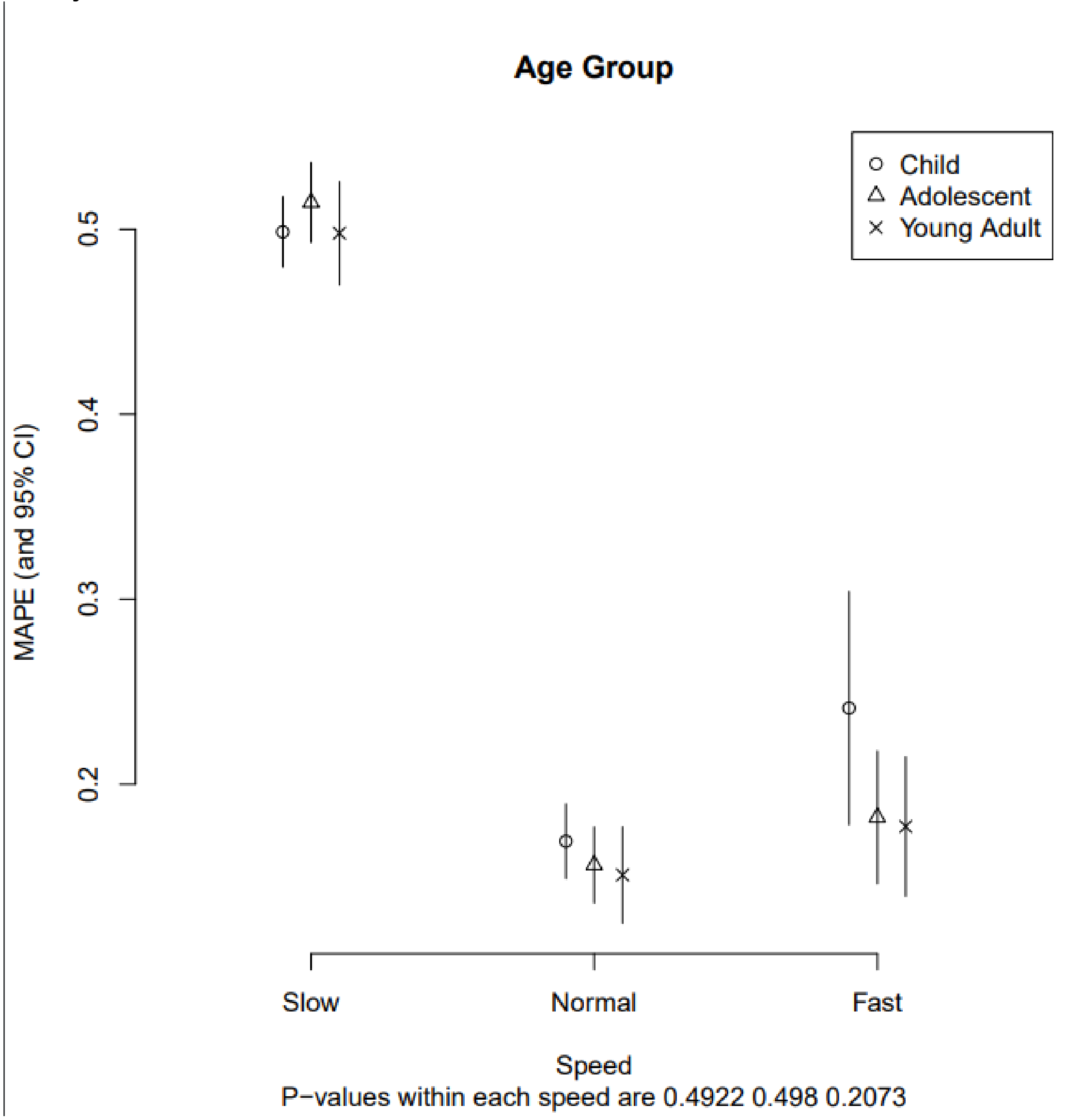
Notes: Correlation coefficients and corresponding 95% CIs of each wear location are presented. Correlation coefficients are averaged across technologies respective to each wear location for all walking bouts. The X axis refers to the wear location, and Y axis refers to correlation coefficients. Correlation coefficients closer to 1.0 indicate greater precision. Wear location does not have a significant effect (p=0.06) on technology precision. Further where 95% CIs do not overlap, there are significant differences between locations. The only significant difference occurred between waist, and arm/wrist wear locations.

Figure 23: The effect of wear location on the overall bias of wearable technology's step counting abilities.



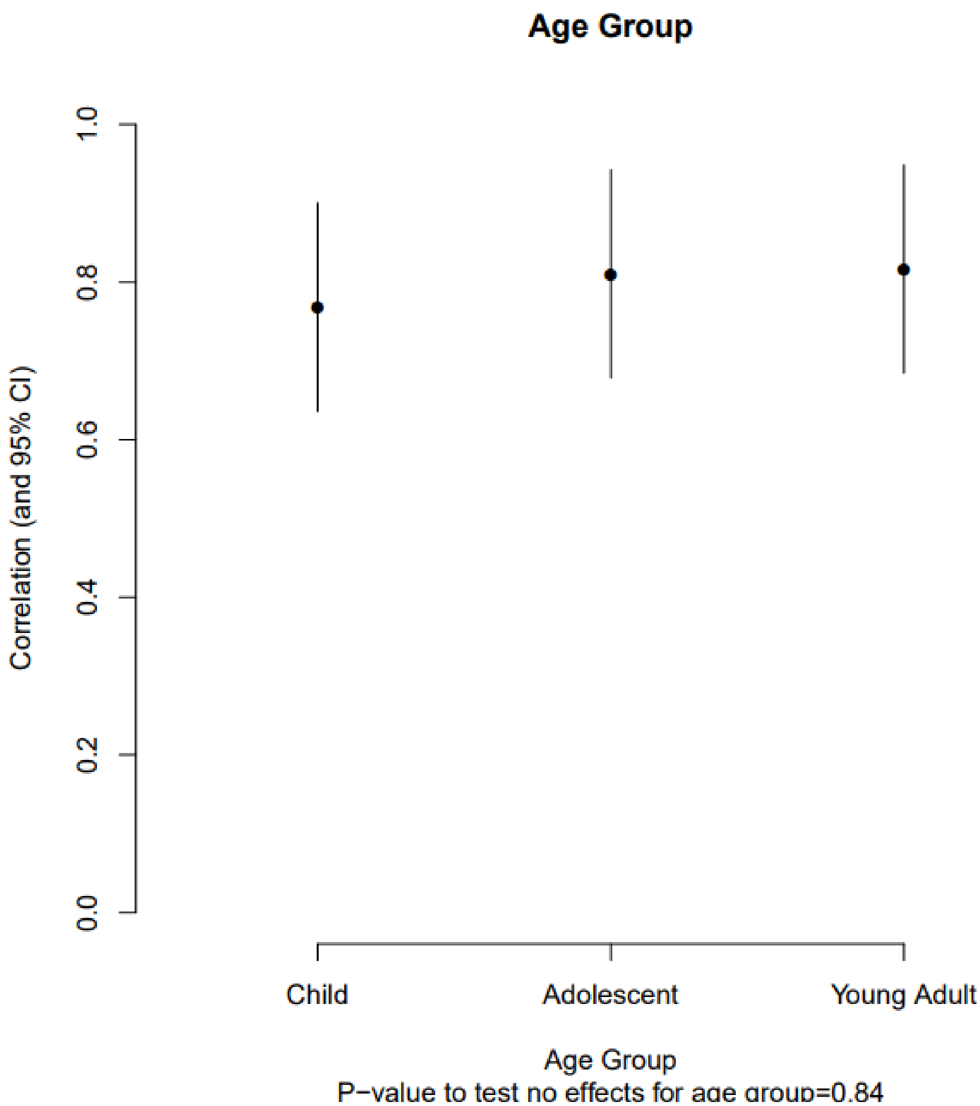
Notes: PE and corresponding 95% CIs of each wear location are presented at slow, normal, and fast walking speeds. PE values are averaged across technologies respective to each wear location for slow, normal, and fast walking speeds. The X axis refers to the speed group, and Y axis refers to PE values. PE values closer to 0% indicate improved bias. Wear location has a significant effect (p>0.0001) on technology bias. Further where 95% CIs do not overlap, there are significant differences between locations.

Figure 24: The effect of age on the accuracy of wearable technology's step counting ability



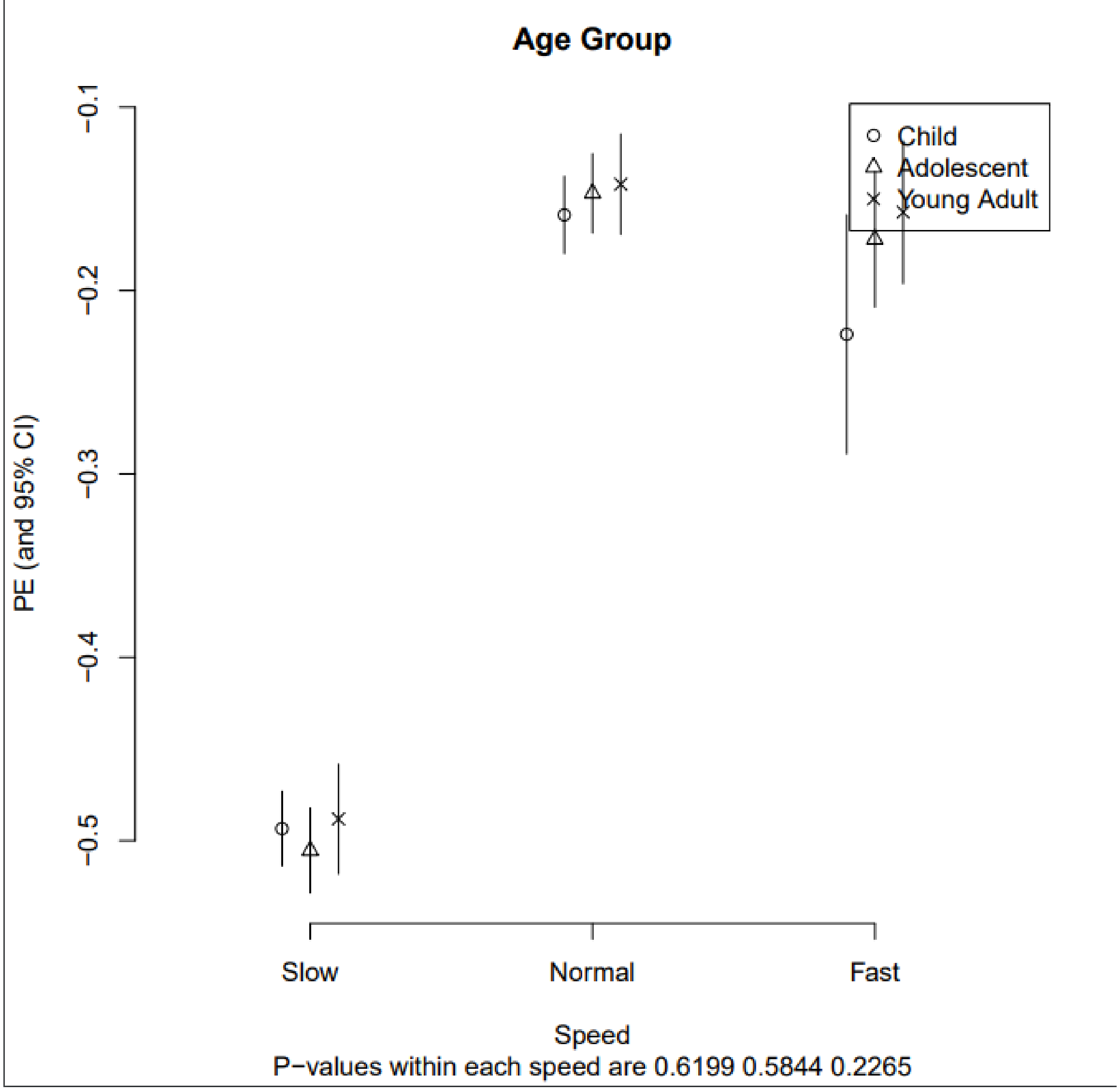
Notes: MAPE and corresponding 95% CIs of each age group are presented at slow, normal, and fast walking speeds. MAPE values are averaged across technologies respective to each age group for slow, normal, and fast walking speeds. The X axis refers to the speed group, and Y axis refers to MAPE values. MAPE values closer to 0% indicate greater accuracy. Age group did not have a significant effect (p=0.49, 0.50, 0.21) on technology accuracy at slow, normal or fast walking, respectively. Further where 95% CIs do not overlap, there are significant differences between locations

Figure 25: The effect of age on overall precision of wearable technology's step counting ability



Notes: Correlation coefficients and corresponding 95% CIs of each age group are presented. Correlation coefficients are averaged across technologies respective to each age group for all walking bouts. The X axis refers to the age group, and Y axis refers to correlation coefficients. Correlation coefficients closer to 1.0 indicate greater precision. Wear location does not have a significant effect (p=0.84) on technology precision. Further where 95% CIs overlap, there are no significant differences between age groups.

Figure 26: The effect of age on the bias of wearable technology's step counting ability



Notes: PE and corresponding 95% CIs of each age group are presented at slow, normal, and fast walking speeds. PE values are averaged across technologies respective to each age group for slow, normal, and fast walking speeds. The X axis refers to the speed group, and Y axis refers to PE values. PE values closer to 0% indicate improved bias. Age group did not have a significant effect (p=0.62, 0.58, 0.23) on technology bias at slow, normal or fast walking, respectively. Further where 95% CIs do not overlap, there are significant differences between locations

REFERENCES

1. Schmitz K, Jacobs D, Leon A, Schreiner P, Sternfeld B. Physical activity and body weight: Associations over ten years in the CARDIA study. *International Journal of Obesity*. 2000;24(11):1475-1487. https://search.proquest.com/docview/219284768. doi: 10.1038/sj.ijo.0801415.

2. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for americans. *Journal of the American Medical Association*. 2018;320(19):2020-2028. https://www.ncbi.nlm.nih.gov/pubmed/30418471. doi: 10.1001/jama.2018.14854.

3. Ainsworth, Barbara E., Haskell, William L, Herrmann, Stephen D., Meckes, Nathanael1, Bassett, David R. Jr., Tudor-Locke, Catrine, Greer, Jennifer L., Vezina, Jesse1, Whitt-Glover, Melicia C., Leon, Arthur S.7, 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values, *Medicine and Science in Sports and Exercise*: August 2011 - Volume 43 - Issue 8 - p 1575-1581 doi: 10.1249/MSS.0b013e31821ece12

4. Strong WB, Malina RM, Blimkie CJR, et al. Evidence based physical activity for school-age youth. *The Journal of Pediatrics*. 2005;146(6):732-737. https://www.sciencedirect.com/science/article/pii/S0022347605001009. doi: 10.1016/j.jpeds.2005.01.055.

5. Janz KF, Letuchy EM, Burns TL, Eichenberger Gilmore JM, Torner JC, Levy SM. Objectively measured physical activity trajectories predict adolescent bone strength: Iowa bone development study. *British Journal of Sports Medicine*. 2014;48(13):1032-1036. http://dx.doi.org/10.1136/bjsports-2014-093574. doi: 10.1136/bjsports-2014-093574.

6. Andersen LB, Harro M, Sardinha LB, et al. Physical activity and clustered cardiovascular risk in children: A cross-sectional study (the European youth heart study). *The Lancet*. 2006;368(9532):299-304. https://www.sciencedirect.com/science/article/pii/S0140673606690752. doi: 10.1016/S0140-6736(06)69075-2.

7. Janz KF, EdD, Kwon S, MS, Letuchy EM, MS, et al. Sustained effect of early physical activity on body fat mass in older children. *American Journal of Preventive Medicine*. 2009;37(1):35-40. https://www.clinicalkey.es/playcontent/1-s2.0-S0749379709002074. doi: 10.1016/j.amepre.2009.03.012.

8. Knowles G, Pallan M, Thomas G, et al. Physical activity and blood pressure in primary school children: A longitudinal study. *Hypertension*. 2013;61(1):70-75. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=fulltext

&D=ovft&AN=00004268-201301000-00015. doi: 10.1161/HYPERTENSIONAHA.112.201277.

9. Haugen HA, Chan L, Li F. Indirect calorimetry: A practical guide for clinicians. *Nutrition in Clinical Practice*. 2007;22(4):377-388. https://journals.sagepub.com/doi/full/10.1177/0115426507022004377. doi: 10.1177/0115426507022004377.

10. Kenny G, Notley S, Gagnon D. Direct calorimetry: A brief historical review of its use in the study of human metabolism and thermoregulation. *European Journal of Applied Physiology*. 2017;117(9):1765-1785. https://www.ncbi.nlm.nih.gov/pubmed/28689303. doi: 10.1007/s00421-017-3670-5.

11. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sports Medicine* 2001;31(6):439-454. http://www.ingentaconnect.com/content/adis/smd/2001/00000031/00000006/art00004. doi: 10.2165/00007256-200131060-00004.

12. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jequier E. Energy expenditure by doubly labeled water: Validation in humans and proposed calculation. *American Journal of Physiology;* (United States). 1986. https://www.osti.gov/biblio/5425867.

13. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Medicine and Science in Sports and Exercise*. 2017;49(8):1697-1703. https://www.ncbi.nlm.nih.gov/pubmed/28709155. doi: 10.1249/MSS.0000000000001284.

14. Eston RG, Rowlands AV, Ingledew DK. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology*. 1998;84(1):362. http://jap.physiology.org/cgi/content/abstract/84/1/362.

15. Rowlands AV, Eston RG. The measurement and interpretation of children's physical activity. *Journal of Sports Science and Medicine*. 2007;6(3):270-276. https://www.ncbi.nlm.nih.gov/pubmed/24149412.

16. Silsbury Z, Goldsmith R, Rushton A. Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations. *BMJ Open*. 2015;5(9):e008430. http://dx.doi.org/10.1136/bmjopen-2015-008430. doi: 10.1136/bmjopen-2015-008430.

17. Dunton GF. Ecological momentary assessment in physical activity research. *Exercise and Sport Sciences Reviews*. 2017;45(1):48-54.

https://www.ncbi.nlm.nih.gov/pubmed/27741022. doi: 10.1249/JES.0000000000000092.

18. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. *Progress in Cardiovascular Diseases*. 2014;57(4):387-395. https://www.clinicalkey.es/playcontent/1-s2.0-S0033062014001674. doi: 10.1016/j.pcad.2014.10.005.

19. Wallace JP, McKenzie TL, Nader PR. Observed vs. recalled exercise behavior: A validation of a seven day exercise recall for boys 11 to 13 years old. *Research Quarterly for Exercise and Sport*. 1985;56(2):161-165. https://doi.org/10.1080/02701367.1985.10608451. doi: 10.1080/02701367.1985.10608451.

20. Olds TS, Olds ST, Gomersall SR, Ridley K. A source of systematic bias in self-reported physical activity: The cutpoint bias hypothesis. *Journal of Science and Medicine in Sport*. 2019;22(8):924-928. https://www.sciencedirect.com/science/article/pii/S1440244018311447. doi: 10.1016/j.jsams.2019.03.006.

21. Strath S, Kaminsky L, Ainsworth B, 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. https://www.ncbi.nlm.nih.gov/pubmed/24126387. doi: 10.1161/01.cir.0000435708.67487.da.

22. Bassett J, David R, Toth LP, LaMunion SR, Crouter SE. Step counting: A review of measurement considerations and health-related applications. *Sports Medicine (Auckland, N.Z.)*. 2017;47(7):1303-1315. https://www.ncbi.nlm.nih.gov/pubmed/28005190. doi: 10.1007/s40279-016-0663-1.

23. Tudor-Locke C, Bassett DR, Shipe MF, McClain JJ. Pedometry methods for assessing free-living adults. *Journal of Physical Activity and Health*. 2011;8(3):445-453. http://silk.library.umass.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=60408399&site=ehost-live&scope=site.

24. Chen K, Bassett D. The technology of accelerometry-based activity monitors: Current and future. *Medicine and Science in Sports and Exercise*. 2005;37(11 Suppl):S490-S500. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=fulltext&D=ovft&AN=00005768-200511001-00002. doi: 10.1249/01.mss.0000185571.49104.82.

25. 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. https://www.ncbi.nlm.nih.gov/pubmed/9588623. doi: 10.1097/00005768-199805000-00021.

26. Matthew CE, Department of Medicine, Division of General Internal Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN.. 2005;37. doi: 10.1249/01.mss.0000185659.11982.3d.

27. Singer E. The measured life. *Utne Reader*. 2011(168):71. https://search.proquest.com/docview/902189921.

28. Gibbs-Smith CH. *The inventions of Leonardo da Vinci.* ; 1978.

29. Dumbauld E. *Thomas Jefferson, American Tourist.* Vol 9. Norman: Univ. of Oklahoma Press; 1946.

30. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: Convergent validity. *Sports Medicine*. 2002;32(12):795-808. http://www.ingentaconnect.com/content/adis/smd/2002/00000032/00000012/art00004. doi: 10.2165/00007256-200232120-00004.

31. Hatano Y. Use of the pedometer for promoting daily walking exercise. *Journal of the International Committee on Health, Physical Education and Recreation*. 1993;29:4-8.

32. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for americans. *Journal of the American Medical Association*. 2018;320(19):2020-2028. https://www.ncbi.nlm.nih.gov/pubmed/30418471. doi: 10.1001/jama.2018.14854.

33. Bravata DM, Smith-Spangler C, Sundaram V, et al. Using pedometers to increase physical activity and improve health: A systematic review. *Journal of the American Medical Association*. 2007;298(19):2296-2304. http://dx.doi.org/10.1001/jama.298.19.2296. doi: 10.1001/jama.298.19.2296.

34. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. *Annals of Family Medicine*. 2008;6(1):69-77. http://www.annfammed.org/cgi/content/abstract/6/1/69. doi: 10.1370/afm.761.

35. Kang M, Marshall SJ, Barreira TV, Lee J. Effect of pedometer-based physical activity interventions. *Research Quarterly for Exercise and Sport*. 2009;80(3):648-655. http://www.tandfonline.com/doi/abs/10.1080/02701367.2009.10599604. doi: 10.1080/02701367.2009.10599604.

36. Lubans DR, Morgan PJ, Tudor-Locke C. A systematic review of studies using pedometers to promote physical activity among youth. *Preventive Medicine*. 2009;48(4):307-315. https://www.clinicalkey.es/playcontent/1-s2.0-S009174350900111X. doi: 10.1016/j.ypmed.2009.02.014.

37. Tudor-Locke C, Rowe D. Using cadence to study free-living ambulatory behaviour. *Sports Medicine*. 2012;42(5):381-398. https://www.ncbi.nlm.nih.gov/pubmed/22462794. doi: 10.2165/11599170-000000000-00000.

38. Tudor-Locke C, Han H, Aguiar EJ, et al. How fast is fast enough? walking cadence (steps/min) as a practical estimate of intensity in adults: A narrative review. *British Journal of Sports Medicine*. 2018;52(12):776-788. http://dx.doi.org/10.1136/bjsports-2017-097628. doi: 10.1136/bjsports-2017-097628.

39. Tudor-Locke C, Schuna J, John M, Han H, et al. Cadence (steps/min) and intensity during ambulation in 6–20 year olds: The CADENCE-kids study. *The International Journal of Behavioral Nutrition and Physical Activity*. 2018;15(1):20. https://www.ncbi.nlm.nih.gov/pubmed/29482554. doi: 10.1186/s12966-018-0651-y.

40. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Frontiers in Physiology*. 2018;9:743. https://www.ncbi.nlm.nih.gov/pubmed/30002629. doi: 10.3389/fphys.2018.00743.

41. Straiton N, Alharbi M, Bauman A, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review. *Maturitas*. 2018;112:85-93. https://www.sciencedirect.com/science/article/pii/S0378512218301828. doi: 10.1016/j.maturitas.2018.03.016.

42. Aloba A., Luc A., Woodward J. Quantifying Differences Between Child and Adult Motion Based on Gait Features. *In: Antona M., Stephanidis C. eds) Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments.* HCII 2019. Lecture Notes in Computer Science, vol 11573. Springer, Cham

43. Beets MW, Patton MM, Edwards S. The accuracy of pedometer steps and time during walking in children. *Medicine and Science in Sports and Exercise*. 2005;37(3):513-520. https://www.ncbi.nlm.nih.gov/pubmed/15741852. doi: 10.1249/01.MSS.0000155395.49960.31.

44. Physical Activity Monitoring for Fitness, Wearables: Step Counting. ANSI/CTA standard.

45. Bruno A. Walther, Joslin L. Moore, Carsten Rahbek. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*. 2005;28(6):815-829. https://www.jstor.org/stable/3683546. doi: 10.1111/j.2005.0906-7590.04112.x.

46. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. *Medicine and Science in Sports and Exercise*. 2012;44(4):733-741. https://www.ncbi.nlm.nih.gov/pubmed/21904249. doi: 10.1249/MSS.0b013e3182351913.

47. Hatano Y. Prevalence and use of pedometers. *Research Journal of Walking*. 1997;1:45-54.

48. Hazra A, Gogtay N. Biostatistics series module 6: Correlation and linear regression. *Indian Journal of Dermatology*. 2016;61(6):593-601. https://www.ncbi.nlm.nih.gov/pubmed/27904175. doi: 10.4103/0019-5154.193662.

49. Dueker D, Gauderman WJ, McConnell R. Accuracy of a new time-resolved step counter in children. *Pediatric exercise science*. 2012;24(4):622-633. https://www.ncbi.nlm.nih.gov/pubmed/23196770. doi: 10.1123/pes.24.4.622.

50. Giavarina D. Understanding bland altman analysis. *Biochemia medica*. 2015;25(2):141-151. https://www.ncbi.nlm.nih.gov/pubmed/26110027. doi: 10.11613/BM.2015.015.

51. Catrine Tudor-Locke, Susan B. Sisson, Sarah M. Lee, Cora L. Craig, Ronald C. Plotnikoff, Adrian Bauman. Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique*. 2006;97:S10-S15. https://www.jstor.org/stable/41994626.

52. Crouter Se, P. L. Schneider, M. Karabulut, D. R. Bassett J. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise*. 2003;8(35):1455-1460.

53. Sandroff BM, Motl RW, Pilutti LA, et al. Accuracy of StepWatch™ and ActiGraph accelerometers for measuring steps taken among persons with multiple sclerosis. *PloS one*. 2014;9(4):e93511. https://www.ncbi.nlm.nih.gov/pubmed/24714028. doi: 10.1371/journal.pone.0093511.

54. Kanoun N. Validation of the ActivPAL activity monitor as a measure of walking at pre-determined slow walking speeds in a healthy population in a controlled

setting. *Reinvention: a Journal of Undergraduate Research*. 2009;2(2). http://www.warwick.ac.uk/go/reinventionjournal/archive/volume2issue2/kanoun.

55. Tudor-Locke, C., Barreira, T.V., Schuna, J.M. *et al.* Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). *International Journal of Behavioral Nutrition and Physical Activity* **12,** 11 (2015). https://doi.org/10.1186/s12966-015-0172-x

56. Zhang M, Zhang Y, Luo M, Nie R. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. *International Journal of Medical Informatics*. 2017;108:97-109. https://www.sciencedirect.com/science/article/pii/S1386505617303568. doi: 10.1016/j.ijmedinf.2017.09.016.

57. Tudor-Locke C, Barreira T, Schuna J. Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine and Science in Sports and Exercise*. 2015;47(4):839-842. https://www.ncbi.nlm.nih.gov/pubmed/25121517. doi: 10.1249/MSS.0000000000000476.

58. Trapp GSA, Giles-Corti B, Bulsara M, et al. Measurement of children's physical activity using a pedometer with a built-in memory. *Journal of Science and Medicine in Sport*. 2012;16(3):222-226. https://www.clinicalkey.es/playcontent/1-s2.0-S1440244012001296. doi: 10.1016/j.jsams.2012.06.011.

59. Ramírez-Marrero FA, Smith BA, Kirby TE, Leenders N, Sherman WM. Evaluation of a step-counter during treadmill walking in 7-12 year old African-American children. *Journal of National Black Nurses' Association : JNBNA*. 2002;13(1):1. https://www.ncbi.nlm.nih.gov/pubmed/12242745.

60. Mitre N, Lanningham-Foster L, Foster R, Levine JA. Pedometer accuracy for children: Can we recommend them for our obese population? *Pediatrics*. 2009;123(1):e127-e131. http://pediatrics.aappublications.org/cgi/content/abstract/123/1/e127. doi: 10.1542/peds.2008-1908.

61. Aminian S, Hinckson EA. Examining the validity of the ActivPAL monitor in measuring posture and ambulatory movement in children. *The International Journal of Behavioral Nutrition and Physical Activity*. 2012;9(1):119. https://www.ncbi.nlm.nih.gov/pubmed/23031188. doi: 10.1186/1479-5868-9-119.

62. Rosenkranz RR, Rosenkranz SK, Weber C. Validity of the actical accelerometer step-count function in children. *Pediatric Exercise Science*. 2011;23(3):355-365. https://www.ncbi.nlm.nih.gov/pubmed/21881156. doi: 10.1123/pes.23.3.355.

63. Assaiante C. Development of locomotor balance control in healthy children. *Neuroscience and Biobehavioral Reviews*. 1998;22(4):527-532. http://dx.doi.org/10.1016/S0149-7634(97)00040-7. doi: 10.1016/S0149-7634(97)00040-7.

64. Toth LP, Bassett DR, Crouter SE, et al. StepWatch accuracy during walking, running, and intermittent activities. *Gait & Posture*. 2016;52:165-170. https://www.clinicalkey.es/playcontent/1-s2.0-S0966636216306737. doi: 10.1016/j.gaitpost.2016.11.035.