Development of Two-Regression Models to Predict Energy Expenditure in Youth Using a GENEActiv and Axivity AX3 Activity Monitor

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Andrew Scott Kaplan
August 2018

ABSTRACT

PURPOSE: The purpose was to develop two regression models (2RM) to estimate energy expenditure (EE) using wrist-worn GENEActiv (GENEA) and Axivity AX3 (AX3) activity monitors in youth. **METHODS:** Youth (N=100; mean ± [plus or minus] SD; age, 12.2±3.5 years) performed 16 activities ranging from sedentary behaviors (SB) to vigorous physical activities (VPA). Participants wore a GENEA and AX3 monitors on the opposite wrists. Monitors were randomized for which device was worn on which wrist. A Cosmed $K4b^2$ (K4b squared) was used as the criterion measure of EE. Raw 100 Hz acceleration data were expressed as Euclidean norm minus one (ENMO) and reduced to one-second epochs. 2RMs were developed for the GENEA and AX3 worn on the left and right wrists. Leave-one-participant-out cross-validation (LOOCV) was used to assess model performance. Using the entire activity bout, estimates of average EE from the four 2RMs and a previously developed single regression equation were calculated and estimates of time spent in different physical activity (PA) intensity levels were calculated using the four 2RMs and five single regression equations and ROC cut-points. **RESULTS:** Log-transformed ENMO was used for the development of the classifiers. Log-transformed ENMO and age were used as predictor variables in the regression equations. For the LOOCV, the four 2RMs had root mean square errors (RMSE) of 0.84-0.95 youth metabolic equivalents (MET_y [MET y]) and mean absolute percent errors (MAPE) of 19.21-20.71%. For the entire activity bout, RMSE for the 2RMs ranged from 0.40 MET_y to 0.60 MET_y and the Hildebrand single regression ranged from 0.97 MET_y to 1.25 MET_y. The four 2RMs were within ± 10.3 minutes of measured minutes of SB, light PA (LPA), moderate PA (MPA), and VPA. All other methods were within ± 61.5 minutes of measured minutes of SB, LPA, MPA, and VPA. **CONCLUSION:** Compared to indirect calorimetry, the newly developed

2RMs had lower RMSE and MAPE for estimates of MET_y and time spent in PA intensity levels than previously developed methods. Future studies should validate the 2RMs using an independent sample in a free-living environment.

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION *1*

Statement of Problem **4**

Statement of Purpose **4**

Research Questions **4**

Delimitations **5**

Limitations **6**

CHAPTER II: LITERATURE REVIEW *7*

GENEActiv **7**
Model Development 8
Comparisons Between GENEActiv and ActiGraph Activity Monitors 17

Axivity AX3 ..... **23**
Model Development 23
Comparisons Between Axivity AX3 and ActiGraph 23

Gaps in the Literature **23**

CHAPTER III: MANUSCRIPT *25*

Introduction **25**

Methods **27**
Participants 27
Procedures 28
Equipment 30
Data Reduction and Cleaning 32
Model Development 33
Additional Models 35
Statistical Analysis 37

Results **38**
Missing Data 38
Model Development 39
Estimation of METy for Structured Activities 40
Estimation of METy for Entire Activity Routine 42
Cross Validation of Models Predicting Time Spent in Different Physical Activity Intensities Levels 43

Discussion **46**
Improvements on Estimations of Energy Expenditure 47
Improvements on Estimations of Time Spent in Different Physcial Activity Intensity Levels 48
Applying a 2RM on the Opposite Wrist or Different Monitor than it was Developed For 50
Strengths and Limitations 52
Conclusion 53

LIST OF REFERENCES *54*

APPENDIX *60*

Tables **61**

Figures **76**

Informed Consent Form **82**

Informed Assent Form **85**
Health History Questionnaire **87**
VITA ***89***

LIST OF TABLES

Table 1. List and descriptions of the activities……..……..……...60

Table 2. Descriptive characteristics of the participants of the final analytical dataset.…….……61

Table 3. Mean (SD) acceleration (g) and coefficient of variation (CV) for the Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) for each activity used in the calibration data set..……...62

Table 4. Sedentary behavior (SB) and continuous walking and running (CWR) classifier sensitivities (SEN), specificities (SPEC), and area under the curve (AUC) for the Axivity left wright ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) 2-regression models...……..…...……64

Table 5. Root mean squared error (RMSE) and mean absolute percent error (MAPE) for Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) data from leave-one-participant-out cross validation when using combinations of linear, cubic, and log-transformed acceleration with and without age as a predictor variable.….……..……..……..……...65

Table 6. Sedentary behavior (SB) and continuous walking and running (CWR) thresholds and CWR and intermittent activity (INT) regression equations developed for the Axivity left wright ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$).…….…….…….……..……..…….67

Table 7. Axivity AX3 Left Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities.…….……...…...68

Table 8. Axivity AX3 Right Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities..……...…...……69

Table 9. GENEActiv Left Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities……...……70

Table 10. GENEActiv Right Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities..……...……...……71

Table 11. Mean (SD) measured youth metabolic equivalents (MET_y) from the Cosmed K4b2 and predicted MET_y from the Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$

2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model across the entire data collection period..……...72

Table 12. Root mean squared error for Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$ 2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model for the entire data collection period..……...73

Table 13. Mean bias and prediction intervals (PI) for Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$ 2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model for the entire data collection period. Units are youth metabolic equivalents (MET_y)..……..……...74

LIST OF FIGURES

Figure 1. Classification between sedentary behavior and non-sedentary behavior for Axivity left wrist (A), GENEActiv left wrist (B), Axivity right wrist (C), and GENEActiv right wrist (D) data using log(ENMO). Values to the left of the threshold are classified as sedentary while values to the right of the threshold are classified as non-sedentary……..75

Figure 2. Classification between continuous walking and running and intermittent activity for Axivity left wrist (A), GENEActiv left wrist (B), Axivity right wrist (C), and GENEActiv right wrist (D) data using the coefficient of variation (CV) of log(ENMO). Values below the threshold are classified as continuous walking and running while values above the threshold are classified as intermittent activity. ……..……..……..……76

Figure 3. Axivity AX3 Left Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed……...77

Figure 4. Axivity AX3 Right Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed…….……..……..78

Figure 5. GENEActiv Left Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed…….……..……..79

Figure 6. GENEActiv Right Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed…….……..80

LIST OF ABBREVIATIONS

2RM- Two-regression model
ANN- Artificial neural networks
AUC- Area under the curve
$AX3_{LW}$- Axivity left wrist
$AX3_{RW}$- Axivity right wrist
CV- Coefficient of variation
CWR- Continuous walking and running
ENMO- Euclidean norm minus one
$GENEA_{LW}$- GENEActiv left wrist
$GENEA_{RW}$- GENEActiv right wrist
g's- gravitational units
Hz- Hertz
ICC- Interclass correlation coefficients
LOOCV- Leave-one-participant-out cross-validation
LPA- Light physical activity
MAPE- Mean absolute percent error
MET_y- Youth metabolic equivalents
MPA- Moderate physical activity
MVPA- Moderate-to-vigorous physical activity
PA- Physical activity
RMR- Resting metabolic rate
RMSE- Root mean squared error
ROC- Receiver operating characteristics
SB- Sedentary behavior
TN- True negative
TP- True positive
VM- Vector magnitude
VPA- Vigorous physical activity

CHAPTER I: INTRODUCTION

Physical activity (PA) is important in youth due to a variety of health benefits, such as increased cognition (3), prevention of obesity (33), and decreased depression (34). Research determining dose-response relationships between PA and health outcomes is reliant on accurate and reliable methods to estimate energy expenditure (EE) and time spent in different PA intensity levels (71). Recall surveys are a common subjective measure that rely on an individual to recall their PA behavior, which includes context of activity, by asking participants to recall their PA for several months to a year (72). Recall surveys are commonly used because of the low cost of administering them to large groups of participants. However, they are limited because individuals may not accurately recall the intensity, frequency, or duration of PA they perform (58). To eliminate the subjective nature of recall surveys, researchers have used objective measures of PA such as accelerometer-based activity monitors to collect PA data (74).

Accelerometer-based activity monitors are lightweight, non-invasive wearable devices that contain accelerometers which measure acceleration and deceleration of the human body. Through the use of predictions models (e.g. regression equations (25, 32), artificial neural networks (ANN) (44, 45, 61), random forest (49), hidden Markov models (51), etc.), these activity monitors can be used to estimate EE, intensity, duration, and frequency of PA for several consecutive days to weeks, making them a popular tool amongst researchers, health care providers, and the general public (6, 21). Various accelerometer-based activity monitors have been uniaxial, measuring acceleration in the vertical axis (VA), while the majority of current accelerometer-based activity monitors are triaxial and measure movement in three axes: mediolateral axis (x-axis), VA (y-axis), and anteroposterior axis (z-axis). Accelerometer-based activity monitors can be worn at various locations on the body such as the wrist, hip, thigh, or

ankle. Traditionally, accelerometer-based activity monitors were worn on the hip. However, wearing accelerometer-based activity monitors on the wrist have become more popular due to increased compliance in wear time by the participant (23, 70) and the ability to estimate sleep duration and quality (65). Accelerometer-based activity monitors can also be combined with other sensors, such as heart rate monitors, gyroscopes, magnetometers, and altimeters to be used in prediction models for estimating EE and time spent in different PA intensity levels. Fifty-one percent of published studies use the ActiGraph, making it the most commonly used accelerometer-based activity monitor (74). However, the GENEActiv (GENEA) is currently being utilized in prospective cohort studies such as the Whitehall II study (N = 10,314) (43), Fenland study (N = 1,695) (73), and the Cork and Kerry Diabetes and Heart Disease Study (N = 464) (36) and the Axivity AX3 (AX3) is being used in the UK BioBank with a total sample of over 500,000 participants (18), making the two activity monitors widely used.

Prediction models that have been developed on adult populations cannot be applied to youth populations. Youth require separate calibrations than adults due to differences in movement economy (46) and increased resting metabolic rate (RMR) (60). Between the ages of 1.5 and 18 years old, running economy improves 2% per year (46). In addition, RMR in youth declines from ~10 ml $O_2 \cdot kg^{-1} \cdot min^{-1}$ at five years old to 3.5 ml $O_2 \cdot kg^{-1} \cdot min^{-1}$ at 18 years old (60). In adults, a metabolic equivalent (MET) is defined as $MET = \frac{Activity\ VO_2}{3.5\ ml \cdot kg^{-1} \cdot min^{-1}}$, however the use of this in youth will result in over-estimations of EE (60). Thus, researchers either use the predicted basal metabolic rate from the Schofield Equation (60) or measured RMR to convert oxygen consumption to youth metabolic equivalents (MET_y, $MET_y = \frac{Activity\ VO_2}{Resting\ VO_2}$). These differences between youth and adult physiology highlight the need for youth-specific regression equations to estimate EE and activity intensity.

Regression equations using acceleration output as the predictor variable are the most common technique to predict EE and time spent in sedentary behavior (SB; <1.5 METs), light physical activity (LPA; 1.5 – 2.99 METs), moderate physical activity (MPA; 3.0 – 5.99 METs), and vigorous physical activity (VPA; ≥6.0 METs) based on the assumption that EE and acceleration are linearly related (25). Although regression techniques are common, no single-regression equation predicts EE or different PA intensities across a wide range of activities (17, 66).

Pober et al. (51) showed that in an adult population, activities with the same MET value can have varying average activity count values, with intermittent activities often having a 2-3 times higher EE than continuous walking or running (CWR) at the same average activity count value. As a result, Crouter and colleagues have developed separate two-regression models (2RM) using the Actical (9) and the VA from a hip-worn ActiGraph 7164 for adults (10, 15) and the VA and vector magnitude (VM) from hip-worn ActiGraph GT3X (GT3X) in youth (13). These 2RMs first apply a threshold that discriminates between sedentary and non-sedentary behavior based on acceleration, and second, based on the variability in count values from one epoch to the next to discriminate between intermittent activity and continuous walk/run activity, a separate regression equation is applied that predicts EE based on the count values. Based off of a validation study using an independent sample in youth, the use of a 2RM improves estimates of EE compared to indirect calorimetry. The Crouter youth-specific VA and VM 2RM had a RMSE of 1.50 and 1.55 MET_y which were lower than commonly used single regression equations (1.56-1.65 MET_y) (14). More recently, Hibbing et al. (30) developed 2RMs for the hip, left wrist and right wrist in adult populations while no 2RMs have been developed for the wrists in youth populations. The left and right wrist 2RMs had RMSEs of 1.24 METs and 1.29 METs,

respectively, which were comparable to the RMSE for the hip which was 1.14 METs suggesting there may be utility for wrist-worn 2RMs.

Statement of Problem

The GENEA and AX3 are two accelerometer-based activity monitors that are used by researchers (74) and have been previously validated using a mechanical shaker for their ability to measure acceleration (20, 39). Single regression equations are still the most commonly used method of estimating PA in youth when using GENEA (32). However, no prediction equation has been developed for the AX3. Using the Actical (9) and ActiGraph series of activity monitors, previous 2RMs decreased the error of EE predictions (13, 14) and provided closer estimates of time spent in different PA intensity levels (14) compared to criterion values derived from indirect calorimetry. Currently, there is a gap in the scientific literature because there is no adequate method for assessing PA using the wrist-worn GENEA and AX3 activity monitors.

Statement of Purpose

The primary purpose of this study was to develop 2RMs to estimate EE using wrist-worn GENEActiv and AX3 activity monitors in youth.

Research Questions

Question 1: Compared to indirect calorimetry, did the wrist-specific 2RMs for the GENEA and AX3 activity monitors provide improved estimates of EE compared to the wrist-specific Hildebrand single regression equation.

Hypothesis 1a: It was hypothesized that using a wrist-specific 2RM for the GENEA in youth would improve estimates of EE to indirect calorimetry compared to the Hildebrand single regression equation.

Hypothesis 1b: It was hypothesized that using a wrist-specific 2RM for the AX3 in youth would improve estimates of EE and time spent in different PA intensity levels to indirect calorimetry compared to the Hildebrand single regression equation.

Question 2: Compared to indirect calorimetry, did the wrist-specific 2RMs for the GENEA and AX3 activity monitors provide improved estimates of time spent in different PA intensity levels, compared to the Hildebrand single regression equation, Phillips cut-points, and Schafer cut-points?

Hypothesis 2a: It was hypothesized that using a wrist-specific 2RM for the GENEA in youth would improve estimates of time spent in different PA intensity levels to indirect calorimetry compared to the Hildebrand single regression equation, Phillips cut-points, and Schaefer cut-points.

Hypothesis 2b: It was hypothesized that using a wrist-specific 2RM for the AX3 in youth would improve estimates of time spent in different PA intensity levels to indirect calorimetry compared to the Hildebrand single regression equation, Phillips cut-points, and Schaefer cut-points.

Delimitations

1. Participants were between 6 and 18 years old.
2. Participants reported no cardio-respiratory conditions, metabolic conditions or medications that affect metabolic processes, and no musculo-skeletal injury within the past six months via a health history questionnaire.
3. Activities were limited to the facilities within and around the Health, Physical Education, and Recreation Building on The University of Tennessee, Knoxville campus.

4. Participants were asked to refrain from PA for 24 hours prior to testing and refrain from eating and drinking, except for water, for 3 hours prior to testing.

Limitations

1. Participants were exposed to some risks inherent to vigorous intensity physical activity
2. Participants' parents were expected to answer the health history questionnaire truthfully.
3. Weather and campus events may have interfered with the ability to collect data.
4. It is assumed that participants followed directions and refrained from exercise, eating, and drinking prior to testing, though participants who did not follow instructions could have affected EE measurements.

CHAPTER II: LITERATURE REVIEW

Accelerometer-based activity monitors are increasingly popular devices that measure movement of the human body. Researchers use accelerometer-based activity monitors because they are minimally invasive. In addition, through the use of prediction models, they can estimate duration, frequency, and intensity of physical activity (PA) (16, 74) through data collected in either a single axis (y-axis) or in three axes (x-, y-, and z-axes). Two commercially available activity monitors that are used by researchers include GENEActiv (GENEA, Activinsights Ltd, Kimbolton, Cambridgeshire, UK) and Axivity AX3 (AX3, Axivity, Newcastle, UK). The GENEA is currently being utilized in prospective cohort studies such as the Whitehall II study (N = 10,314) (43), Fenland study (N = 1,695) (73), and the Cork and Kerry Diabetes and Heart Disease Study (N = 464) (36). The AX3 is being used in the UK BioBank with a total sample of over 500,000 participants (18).

The purpose of the literature review is to discuss previously developed energy expenditure (EE) and PA intensity prediction models for the GENEA and AX3 activity monitors in adults and youth. The review will be organized by activity monitor. Two subsections for each activity monitor will be discussed: 1) model development split into adult and youth and 2) comparisons between the activity monitor and ActiGraph, the most widely used accelerometer-based activity monitor (74).

GENEActiv

GENEA activity monitors are typically worn on the wrist or hip. The GENEA is a small (43 x 40 x 13 mm) lightweight (16g) device that contains a triaxial accelerometer. It records raw acceleration with a range up to ±8 gravitational units (g's) at a sampling frequency between 10-100 Hertz (Hz) in 10 Hz increments. The GENEA has the battery and memory capacity to collect

100 Hz data for up to seven days. GENEA raw acceleration output (g's) have been validated using a mechanical shaker (21) and calibration of prediction models to estimate EE and time spent in different PA intensity levels have been developed for youth (19, 31, 32, 50, 55, 56, 59) and adult (20, 31, 32, 35, 44, 45, 49, 52, 57, 69, 75) populations.

Model Development

Adult

Esliger et al. (20) investigated the technical reliability and validity of GENEA raw acceleration output, and calibrated activity intensity thresholds. Technical reliability and validity was assessed using a multi-axis shaking table which was calibrated to oscillate at 15 physiologically relevant accelerations with ranges that spanned light, moderate, and vigorous intensities. Across all 15 accelerations, intrainstrument and interinstrument coefficients of variation (CV) were 1.8% and 2.4%, respectively with a correlation $r = 0.97$ ($p < 0.001$). In addition to the testing using a multi-axis shaking table, activity intensity thresholds for sedentary behavior (SB; <1.5 METs), light physical activity (LPA; 1.5 – 2.99 METs), moderate physical activity (MPA; 3 – 5.99 METs), and vigorous physical activity (VPA; ≥ 6.00 METs) were developed using receiver operating characteristics (ROC) curve analyses. A sample of 60 adults (age range: 40-63 years). completed 10 to 12 activities for 4.5 minutes each, except for lying, which was completed for ten minutes. A Cosmed $K4b^2$ portable metabolic system was used as the criterion measure of EE. Reported correlations between acceleration and EE from the Cosmed $K4b^2$ were $r = 0.86$ (left wrist), $r = 0.83$ (right wrist), and $r = 0.87$ (hip). For the left wrist, right wrist, and hip, sensitivities ranged from 97-99% (SB), 95-100% (MPA), and 73-78% (VPA) and specificities ranged from 95-96% (SB), 56-80% (MPA), and 97-99% (VPA). LPA

was not assessed individually because the upper limit of SB and lower limit of MPA were used to create the thresholds for LPA.

Zhang et al. (75) developed algorithms for classification of SB, household, and locomotion activities. The sample included 60 adults (mean age [SD] = 49.4 [6.5] years). The protocol consisted of 10 to 12 activities and each activity was completed for 4.5 minutes, except for supine rest, which was performed for ten minutes. Participants wore GENEA monitors on the hip, left wrist, and right wrist. To determine activity classifications, decision tree models were developed for each attachment site. Precision ($\frac{True\ Positive\ (TP)}{TP + True\ Negative\ (TN)}$) for the left wrist, right wrist, and hip, ranged from 98-99% (SB), 91-97% (household), 96-100% (walking), and 99-100% (running). Accuracy ($\frac{TP + TN}{Estimated\ Positive + Estimated\ Negative}$) of all activities combined were 96% (left wrist), 97% (right wrist), and 99% (hip).

Hildebrand et al. (32) created one of the most commonly used regression equations to predict EE. A sample of thirty adults (mean age [SD] = 34.2 [10.7] years) completed a protocol consisting of eight structured activities split into two SB (lying down and sitting) and six activities ranging from LPA to VPA (circuit of activities of daily living, slow walking at 50 $m \cdot min^{-1}$, fast walking at 83.3 $m \cdot min^{-1}$, running at 133.3 $m \cdot min^{-1}$, stepping up stairs, and standing). The circuit of daily activities was coded as one activity and consisted of taking off shoes standing, moving eight things in a bookshelf, writing a sentence, putting a sheet of paper in an envelope, and sitting down. Each activity was completed for five minutes, except for lying down, which was completed for ten minutes. The participants wore GENEA activity monitors on their non-dominant wrist and right hip. A VMax Encore indirect calorimeter was used as the criterion measure of EE. Raw acceleration measures for all three axes were combined into one measure of body acceleration using Euclidean norm minus one (ENMO) which subtracts one gravitational

unit from vector magnitude (VM; $[x^2 + y^2 + z^2]^{1/2} – 1g$) (68). Leave-one-out cross validation was used to assess model performance. Two regression equations for predicting VO_2 were developed: hip, VO_2 (ml $O_2 \cdot kg^{-1} \cdot min^{-1}$) = 0.0530 ENMO + 6.86, and wrist, VO_2 (ml $O_2 \cdot kg^{-1} \cdot min^{-1}$) = 0.0323 ENMO + 7.49. The hip and wrist equations are unable to estimate time spent in SB, because when there is zero acceleration, the estimations of VO_2 are 6.86 and 7.49 ml $O_2 \cdot kg^{-1} \cdot min^{-1}$ for the hip and wrist, respectively, which is about two times higher than the average RMR of an adult, which is 3.5 ml $O_2 \cdot kg^{-1} \cdot min^{-1}$. Therefore, the authors combined SB and LPA estimations into a single category. For SB/LPA, intensity classification accuracies for the hip and wrist were 93%-100% for all activities except for slow walking (80%, hip, and 0%, wrist). For MPA, the hip and wrist classifications were 100% for fast walking, however, <47 and <30% for all other activities, respectively. For VPA, intensity classification accuracies for the hip and wrist for running were 100% and 97%, respectively, and for stepping were 0% for both locations.

Hildebrand et al. (31) calibrated SB cut-points from the same data set discussed previously (32). These cut-points were validated in a free-living setting with the ActivPal activity monitor used as the criterion measure of time spent in SB. Cut-points for lie/sit and stand/step were created using ROC curve analyses. Sensitivity for SB thresholds were 93% and 98% for the hip and wrist, respectively, although specificity for SB thresholds were 73% and 78% for the hip and wrist, respectively. When these cut-points were applied in a free-living setting, estimates of SB time were significantly different by at least 30 minutes at both attachment sites compared to the ActivPal ($p < 0.001$).

One of the few studies to use machine learning techniques with the GENEA activity monitor was conducted by Montoye et al. (44). Artificial neural networks (ANN) were developed to predict EE from wrist-worn monitors while identifying simple feature sets to

maximize the accuracy of the predictions. Forty-four adults (mean age [SD] = 22.1 [4.3] years) were recruited with an equal percentage of male and female participants. Participants wore an Oxycon Mobile metabolic analyzer which was used as the criterion measure of EE. GENEA monitors were worn on both wrists while the participants completed 14 activities ranging from SB to VPA which included ambulatory, lifestyle, and exercise movements. Using a semi-structured format, participants could choose the order of activities and when to transition between activities as long the duration was between 3-10 minutes. Four feature sets were used to develop the neural networks: 1) 36-time domain features and three participant features, 2) only mean and variance of acceleration, 3) mean, variance, and minimum and maximum acceleration signal, and 4) 10th, 25th, 75th, and 90th percentiles and the covariance of acceleration. The average measured MET value was 3.3 across all activities. Left wrist root mean square error (RMSE) for feature sets 1-4 were 1.18, 1.26, 1.26, and 1.15 METs, respectively, and right wrist RMSE were 1.18, 1.25, 1.27, and 1.21 METs, respectively.

Montoye et al. (45) developed ANNs for wrist-worn monitors using the same sample set as Montoye et al (44). Separate ANNs were used to predict EE and whether the activity monitor was worn on the left or right wrist. For EE predictions, three feature sets were tested for each wrist: 1) mean and variance of the VM, 2) the absolute values of the mean and variance of the VM, and 3) mean raw acceleration. For prediction of whether the GENEA was located on the left or right wrist, only the third feature set was tested. During evaluation, both models were applied to data from the wrist they were developed for (same wrist prediction) and the opposite wrist (opposite wrist prediction). Same wrist RMSE for feature sets 1-3 were 1.47, 1.33, and 1.25 METs, respectively. Opposite wrist RMSE for feature sets 1-3 were 1.48, 1.35, and 1.97 METs,

respectively. The ANN developed to predict whether the GENEA was located on left or right wrist was correct 100% of the time.

Pavey et al. (49) used random forest models to predict activity intensity in a free-living environment. A sample of 21 adults (mean age [SD] = 27.6 [6.2] years) wore GENEA activity monitors while completing seven activities in a laboratory setting. The first activity was always lying down while the following six (sitting still, standing still, sitting while using a computer or sorting papers, washing dishes or cleaning windows, walking at a self-selected pace, and running at a self-selected pace) were completed in a random order for three minutes each. Sixteen of the 21 participants also completed a 24-hour free-living trial the day following the laboratory protocol. ActivPals were worn on the thigh as the criterion measure of posture and stepping vs. non-stepping behavior. A random forest classifier model was developed using frequency domain features to classify SB, stationary+ (sitting active or standing still/active), walking, and running. Another random forest classifier model was developed to determine stepping vs. non-stepping behavior. Model performance was assessed using balanced accuracy, which is the average of sensitivity and specificity. Balanced accuracy for SB, stationary+, walking, and running were 89%, 93%, 95%, and 97%, respectively. The sensitivity and specificity for the random forest model that distinguishes between stepping and non-stepping behavior were 54% and 96%, respectively.

In summary, regression equations, ROC cut-points and machine learning algorithms have been developed to predict EE and time spent in different PA intensity levels using a GENEA activity monitor in adults. However, more complex machine learning approaches have shown to improve EE and time spent in different intensity estimates compared to single regression

approaches. In addition, an ANN was developed to determine whether the GENEA activity monitor is being worn on the left or right wrist with perfect accuracy.

Youth

Phillips et al. (50) developed cut-points using a GENEA activity monitor in youth. A sample of 44 youth (mean age [SD] = 10.9 [1.9] years) completed eight activities while wearing GENEA activity monitors on both wrists and the right hip and a Cosmed K4b^2 as a criterion measure of EE. Cut-points were developed using ROC analyses to split activities into four intensities; SB (<1.5 youth metabolic equivalents [MET_y]), LPA (1.5-2.99 MET_y), MPA (3-5.99 MET_y), and VPA (>6.00 MET_y). For the left wrist, right wrist, and hip, sensitivity ranged from 95-96% (SB), 82-89% (MPA), and 89-92% (VPA) and specificity ranged from 96-98% (SB), 83-88% (MPA), and 86-89% (VPA). LPA was not assessed individually because the upper limit of SB and lower limit of MPA were used to create the thresholds for LPA.

Hildebrand et al. (32) developed one of the most commonly used youth specific regression equations to predict EE using a GENEA. Thirty youth (mean age [SD] = 8.9 [0.9] years) completed an activity protocol similar to the adults previously discussed (32). Differences in the youth and adult activity protocol included allowing youth to watch television during lying down and drawing on a white board during standing in lieu of using a mobile phone. Two regression equations for predicting VO_2 were developed: hip, VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) = 0.0497 ENMO + 10.39 (r^2 = 0.75), and wrist, VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) = 0.0357 ENMO + 11.16 (r^2 = 0.72). Similar to the adult wrist equation, the youth hip and wrist equations are unable to estimate EE and time spent in SB, because when there is zero acceleration, the estimations of VO_2 are 10.39 and 11.16 $ml \cdot kg^{-1} \cdot min^{-1}$ for the hip and wrist, respectively. Those values are almost twice the average RMR of youth in the study which was 6 $ml \cdot kg^{-1} \cdot min^{-1}$. Therefore, the authors combined

SB and LPA estimations into a single category. For SB/LPA, intensity classification accuracies for the hip and wrist were 93%-100% for all activities except for fast walking (0%). For MPA, intensity classification accuracies for the hip and wrist were <64% for all activities except for fast walk (85%, hip) and running (100%, wrist). For VPA, intensity classification accuracies for the hip and wrist for stepping were 67% and 0%, respectively and for running were 89% and 72%, respectively.

Hildebrand et al. (31) created SB cut-points in youth using the same data set mentioned previously (32). Cut-points for lie/sit and stand/step were developed using ROC curve analyses. For the hip and wrist, sensitivity for SB thresholds was 100% and 97%, respectively, while specificity was 68% and 75%, respectively. When applied to free-living data, the hip and wrist GENEA predictions overestimated time spent in SB by 26% and 15%, respectively, compared to the ActivPal.

Schaefer et al. (59) developed GENEA cut-points for the wrist in a youth population. The study included 24 youth (mean age [SD] = 9.4 [1.2] years) and each participant wore a GENEA activity monitor on the non-dominant wrist. The participants performed a resting trial for six minutes before completing seven activities in order of increasing intensity for six minutes each, that included: coloring, Lego® building, Wii sports tennis, Wii sports boxing, treadmill walking at 45 and 75 $m \cdot min^{-1}$, jogging at 105 $m \cdot min^{-1}$, and running at 135 $m \cdot min^{-1}$. Average gravity-subtracted signal VM was calculated for each one second epoch using the formula: signal VM = $\sum ([x^2 + y^2 + z^2]^{1/2} / (f)$ where f is the sampling frequency. ROC analyses were used to establish cut-points for SB, LPA, MPA, and VPA intensities. The cut-points created for SB, MPA, and VPA were 0.190, 0.314, and 0.998 signal VM. Sensitivity was 97% (SB), 91% (MPA), and 95% (VPA), and specificity was 88% (SB), 87% (MPA), 85% (VPA). No cut point was created for

LPA, so the upper threshold of SB and lower threshold of MPA activity created the thresholds for LPA.

Duncan et al. (19) validated the Phillips et al. (50) cut-points in a 5-8 year old population (mean age [SD] = 6.8 [1.4] years). The Phillips cut-points were developed in youth between the ages of 8 and 14 years old. Fifteen youth wore GENEA activity monitors on their non-dominant wrist while performing a series of six semi-structured activities for five minutes each. The activities included lying, sitting and playing with Legos, slow-paced walking (3 $km \cdot hour^{-1}$), medium-paced walking (4.5 $km \cdot hour^{-1}$), fast-paced walking (6 $km \cdot hour^{-1}$), and a medium-paced run (8 $km \cdot hour^{-1}$). All walking and running activities were completed on a treadmill. Participants wore a MetaMax 3B portable gas analyzer as the criterion measure of EE. Classification accuracy of the cut-points were evaluated using sensitivity and specificity. Sensitivity was 92% (SB), 81% (LPA), 97% (MPA), and 96% (VPA), and sensitivity was 90% (SB), 56% (LPA), 83% (MPA), and 84% (VPA).

Recently, Roscoe et al. (55) developed wrist cut-points for PA intensity in preschool children four to five years old. Twenty-one participants wore GENEA activity monitors on both wrists and simultaneously wore a MetaMax 3B gas analyzer as the criterion measure of EE. The participants performed two SB activities (lying and playing with Lego® blocks) for five minutes each and four treadmill walking and running activities (41.7 $m \cdot min^{-1}$, 56.7 $m \cdot min^{-1}$, 71.7 $m \cdot min^{-1}$, and 90.0 $m \cdot min^{-1}$) for four minutes each. Using ROC curve analyses, activity intensity cut-points were developed for SB, LPA, and MPA for the dominant and non-dominant wrists. VPA cut-points were not developed because the youth who participated in the study could not run at a speed for four continuous minutes that would be classified as VPA intensity. For the non-dominant wrist, sensitivities were 90% (SB) and 86% (MPA) and specificities were 90% (SB)

and 40% (MPA). For the dominant wrist, sensitivities were 100% (SB) and 76% (MPA) and specificities were 10% (SB) and 40% (MPA). The authors also evaluated LPA based off the SB and MPA cut-points. For the non-dominant wrist, sensitivity and specificity were 40% and 20%, respectively. For the dominant wrist, sensitivity and specificity were 10% and 85%, respectively.

Okely et al. (47) validated previous developed wrist specific cut-points for estimating time spent in MPA, VPA, and moderate-to-vigorous PA (MVPA) in youth. The study included 57 participants (mean age [SD] = 9.2 [2.3] years) who wore a MetaMax 3B portable metabolic system and GENEA activity monitors on their non-dominant wrist. Participants completed 15 semi-structured activities ranging from SB to VPA in increasing intensity for five minutes each, with the exception of lying down which was completed for 10 minutes. Metabolic VO_2 data from a MetaMax 3B portable respiratory gas analysis system were averaged over 10-second epochs and converted into MET_y using measured RMR. Activities were categorized into MPA ($\geq$ 3 - 5.99 MET_y), VPA ($\geq$ 6 MET_y) and MVPA ($\geq$ 3 MET_y) based off indirect calorimetry. Acceleration data reduction were performed in three ways that were specific to the calibration studies of Hildebrand et al. (32), Phillips, et al. (50), and Schaefer et al. (59). Measured and predicted minutes spent in MPA, VPA, and MVPA cut-points were reported in contingency tables and 95% equivalence testing were conducted to determine group-level agreement between the cut-points using indirect calorimetry as a criterion. The Hildebrand, Phillips, and Schaefer cut-points correctly classified MPA 47%, 45%, and 52% and VPA 70%, 80%, and 94% of the time, respectively. The Schaefer cut-points were equivalent to the criterion for estimating minutes of MPA, but no cut-points were equivalent to the criterion for estimating minutes of VPA. The Hildebrand and Phillips cut-points were equivalent to the criterion for estimating

minutes of MVPA. When wearing a GENEA activity monitor on the wrist, all three prediction methods exhibited large group-level error with high misclassification of MVPA as non-MVPA.

In summary, only regression equations and ROC cut-points have been developed to predict EE and time spent in different PA intensity levels using GENEA activity monitors in youth. No machine learning algorithms have been developed using the GENEA in youth. The Hildebrand regression equation is unable to predict sedentary behavior while the Phillips and Schaefer cut-points have a high misclassification rate of MVPA.

Comparisons Between GENEActiv and ActiGraph Activity Monitors

One of the clear advantages to using raw acceleration is the ability to compare output between activity monitors and potentially develop prediction models that can be used by any brand of accelerometer-based activity monitor (24). Comparing the GENEA to the ActiGraph series of activity monitors is important because ActiGraph activity monitors are used most frequently amongst researchers (74). ActiGraph activity monitors are used in 51% of published studies (74) and the ActiGraph 7164 and ActiGraph GT3X+ (GT3X+) have been used in the National Health and Nutrition Examination Survey.

Acceleration Output

Comparisons of raw acceleration output between GT3X+ and GENEA activity monitors have been previously investigated (32, 35, 56, 57). John et al. (35) conducted a study comparing raw acceleration (g's) values between the GENEA and GT3X+. Comparisons using a mechanical shaker were conducted by attaching each activity monitor to a mechanical shaker that oscillated at frequencies between 0.7 - 4.0 Hz and a fixed radius of 5.08 cm for a duration of ten minutes. These frequencies correlate to locomotion speeds ranging between 1.5 mph to 16 mph. Using a linear mixed model with likelihood ratio tests, mean VM raw acceleration between the GT3X+

and GENEA at each oscillation frequency were compared. Raw triaxial acceleration output between brands of activity monitors were statistically different at each oscillation frequency ($p < 0.05$). The GENEA acceleration were consistently higher by 3.5-6.5% with larger differences seen as oscillation frequency increased. Significant differences were found between GENEA and GT3X+ VM raw acceleration output will affect estimates of EE and time spent in different PA intensity levels when using models developed on the opposite activity monitor.

As a part of the same study discussed previously, John et al. (35) compared raw acceleration values from the GT3X+ and GENEA activity monitors in a laboratory protocol designed to simulate free-living. Eight adults (mean age [SD] = 23.8 [5.4] years) completed eight activities which included walking at 2.0 and 3.5 mph on a treadmill, running at 5.5 and 7.5 mph on a treadmill, seated computer work, vacuuming, cleaning a room, and throwing a ball for two minutes each. Two separate random forest models were developed for each monitor to predict activity type, one using frequency domain features and one using time domain features. The prediction accuracy was compared when the models were applied to the activity monitor they were trained for (e.g. GT3X+ model on GT3X+ data) and applied to the other activity monitor (e.g. GT3X+ model on GENEA data). When using frequency domain features, accuracy was 94.3-95.8% when the model was applied to the same activity monitor as it was calibrated for and 93.8% when the model was applied to the opposite monitor than the one it was calibrated on. When using time domain features, accuracy was 91.7-94.3% when the model was applied to the same activity monitor it was calibrated on and 86.5-94.5% when the model was applied to the opposite monitor than the one it was calibrated on. Significant differences were found in VM raw acceleration output between the GT3X+ and GENEA activity monitors which will affect time domain features but not frequency domain features. When developing models that are intended

to be used between the GENEA and GT3X+, using frequency domain features are recommended.

Rowlands et al. (57) compared raw acceleration outputs between wrist-worn GT3X+ and GENEA activity monitors in adults. The outputs investigated in this study included acceleration and time spent in MVPA. Thirty-four participants (mean age [SD] = 28.2 [5.8] years) participated in a two-day free-living trial. The participants were instructed to wear both monitors on their non-dominant wrist during all waking hours with the GENEA always being proximal to the GT3X+. For each activity monitor, ENMO was calculated over 5-second epochs. Time spent in MVPA was classified using a 100 mg per five second cut-point. Mean ENMO for the GENEA and were 22.9 ± 20.7 mg and 27.8 ± 21.4 mg, respectively, and were significantly different ($p < 0.05$). Estimates of time spent in MVPA from the GENEA and GT3X+ were 91.8 ± 46.0 minutes and 89.3 and 46.0 minutes, respectively, and were not significantly different from one another ($p > 0.05$). Intraclass correlation coefficients (ICC) and 95% confidence intervals were used to determine agreement between activity monitor brand outputs (ENMO and time spent in MVPA). ENMO output between the two monitor brands were highly correlated (ICC = 0.987, 95% CI = 0.707-0.997). Time spent in MVPA between the two monitor brands were highly correlated (ICC = 0.982, 95% CI = 0.943-0.993). In conclusion, the ENMO outputs (mg) were significantly different. However, no statistical differences were found for estimates of average time spent in MVPA.

Hildebrand et al. (32) compared raw acceleration between GT3X+ and GENEA activity monitors using the same participants and the same protocol discussed previously. Agreement between placement and brands of activity monitors were evaluated using a two-way mixed model ANOVA, ICC, and mean bias with limits of agreement. The raw acceleration outputs

were not significantly different between brands in adults (p = 0.12) or youth (p = 0.73). ICC and 95% CI between the raw acceleration for each brand was 0.979 (0.979-0.980) for the hip and 0.987 (0.986-0.987) for the wrist in adults and 0.964 (0.929-0.932) for the hip and 0.976 (0.976-0.977) for the wrist in youth. Mean bias and limits of agreement for the hip and wrist in youth and adults ranged from -6.9 mg to 10.3 mg with all limits of agreement ranging from ±45mg to ±55mg. Raw acceleration output between brands at the same wear location has a high correlation although individual variability is large as shown by wide limits of agreement.

In summary, inconsistent findings were found when comparing raw acceleration between the GENEA and GT3X+ activity monitors. When the two brands of activity monitors were attached to a mechanical shaker, significant differences in VM raw acceleration were seen with the differences becoming larger as oscillation frequency increased (35). In a free-living environment, Rowlands et al. (57) found significant differences in average ENMO (mg) output but no statistical differences in average time spent in MVPA.

Comparisons of Estimations from Prediction Equations between the GENEActiv and ActiGraph Series of Activity Monitors

Van Loo et al. (69) validated previously developed cut-points for estimating SB using wrist accelerometry in 57 youth (mean age [SD] = 9.2 [2.3] years), split into groups of 5-8 years old and 9-12 years old for analyses. Participants wore a GT3X+ and GENEA on each wrist, and an GT3X+ while completing 15 activities over two testing days. The order of activities for each day of testing were in increasing intensity. The 15 activities included: lying down, TV watching, handheld electronic gaming, writing/coloring, computer gaming, getting ready for school, standing class activity [e.g. writing on a whiteboard], slow walking at a self-selected pace, picking up toys/clothes, brisk walk at a self-selected pace, soccer, basketball, running at a self-

selected pace, and an obstacle course. The criterion measure for activity classification was direct observation. Nine ActiGraph and two GENEA wrist cut-points were validated. The authors used 95% equivalence testing to assess group level equivalence between measured and predicted SB time. For 5-8 year olds, no equation was statistically equivalent to direct observation. For 9-12 year olds, the Crouter et al. VA ROC cut-points (12) and Kim et al. (38) single regression equation developed for the ActiGraph were statistically equivalent to direct observation ($p < 0.001$). The Schaefer et al. (59) and Phillips et al. (50) cut-points developed for the GENEA overestimated SB time by 9.6%-17.8% for both age groups.

Rowlands et al. (56) compared estimates of time spent in different PA intensity levels between GENEA cut-points and ActiGraph cut-points for the hip and wrist in youth. The GENEA and ActiGraph cut-points were calibrated using raw acceleration and activity counts, respectively. A sample of 51 youth (mean age [SD] = 10.7 [0.8] years) wore a GT3X+ and a GENEA activity monitor on the right hip and GENEA activity monitors on their non-dominant wrist during a seven-day free-living trial. Estimates for time spent in SB, LPA, and MVPA for GT3X+ data were calculated using the Evenson et al. (22) cut-points, which uses the VA, and the Hanggi et al. (28) cut-points, which uses VM. Estimates for time spent in SB, LPA, and MVPA using GENEA data were calculated using the Phillips et al. (50) wrist and hip cut-points, which uses ENMO. Phillips developed cut-points for the left and right wrist so the cut-points that corresponded with the wrist in which the GENEA was worn was used for analysis. Mean output between GT3X+ activity count output and GENEActiv raw acceleration ($g \cdot sec$-1) were examined using four correlations: 1) mean daily VA GT3X+ ($counts \cdot sec^{-1}$) with mean daily GENEActiv hip output ($g \cdot sec^{-1}$), 2) mean daily VA GT3X+ ($counts \cdot sec^{-1}$) with mean daily GENEActiv wrist output ($g \cdot sec^{-1}$), 3) mean daily VM GT3X+ ($counts \cdot sec^{-1}$) with mean daily GENEActiv hip output

($g \cdot sec^{-1}$), and 4) mean daily VM GT3X+ ($counts \cdot sec^{-1}$) with mean daily GENEActiv wrist output ($g \cdot sec^{-1}$). Repeated measures ANOVAs were used to investigate statistical differences between the four cut-point methods for time spent in SB, LPA, MPA, VPA, and MVPA. For mean daily output, all correlations (r^2) were significant ranging from 0.86-0.90 ($p < 0.001$). For time spent in SB, the Evenson equation was significantly different from both GENEA equations ($p < 0.01$) while the Hanggi equation was significantly different from the Schafer cut-points ($p < 0.01$) but not the Philips cut-points ($p > 0.01$). For time spent in SB and LPA, the Evenson cut-points was significantly different from both GENEA cut-points ($p < 0.01$) while the Hanggi cut-points was significantly different from the Schafer cut-points ($p < 0.01$) but not the Philips cut-points ($p > 0.01$). For time spent in MPA and VPA, the Evenson cut-points was significantly different from the Schafer cut-points and the Philips cut-points ($p < 0.01$). No comparisons were made using the Hanggi equation for time spent in MPA and VPA because those intensities were combined into one MVPA cut-point. For time spent in MVPA, all equations were significantly different ($p < 0.01$).

In summary, there is limited evidence that estimates of EE and time spent in different PA intensity levels from prediction equations developed for the GENEA or ActiGraph when applied across-device are different, thus making them comparable. Within the same device, the estimates of EE and time spent in different PA intensity levels are highly dependent on the prediction method being used. More research should be done to investigate the comparability of using a regression equation developed using data from one monitor but applied to data from a different monitor. For example, how do estimates of time spent in different PA intensity levels differ when applying a model developed using GENEA data compare when applied to GENEA and ActiGraph data using an independent sample.

Axivity AX3

The AX3 is a small (23.0 x 32.5 x 7.6 mm), lightweight (9g) activity monitor that contains a triaxial accelerometer that can be worn at multiple body locations. The monitor can measure acceleration with a sampling rate ranging between 12.5 Hz and 3200 Hz and a seismic acceleration range of ± 2, 4, 8, or 16 g's. The AX3 has a battery life of 14 days when a sampling frequency of 100 Hz is used. Other features of the activity monitor include a real-time clock, 512 MB of memory, temperature sensor, light sensor, and it is dust and water resistance.

Model Development

No models have been developed for the AX3 to estimate EE or activity type in adults or youth. However, the manufacturer of the AX3 recommends using models developed for the GENEA; specifically the wrist-specific Hildebrand regression equations for estimating EE in adults and youth and the wrist-specific Phillips cut-points for time spent in different PA intensity levels for youth (2). No hip-specific regression equations were recommended, although Axivity Ltd. guides users to a variety of review articles that overview methods to analyze accelerometer-based activity monitor data at various wear locations (1, 37, 62).

Comparisons Between Axivity AX3 and ActiGraph

Currently there are no studies that have investigated the raw acceleration output between the AX3 and ActiGraph series of activity monitors. No models have been developed for the AX3 to predict EE or time spent in different PA intensity levels; thus, no comparisons can be made between AX3 and ActiGraph prediction models.

Gaps in the Literature

Prediction models for estimates of EE and time spent in different PA intensity levels for the GENEA activity monitors have been developed. However, no prediction models have been

developed for the AX3. Single regression equations and cut-points methods are variable in their estimates of EE and time spent in different activities intensities when predicting across a wide range of PA intensity levels and activity domains (i.e. household chores, CWR, sporting activities, etc.). The use of a two-regression model (2RM) that distinguishes between continuous walking and running and intermittent lifestyle activities has improved estimations of EE and time spent in different PA intensity levels in adults (9-11, 15) and youth (12, 13) using the Actical and ActiGraph series of activity monitors for the hip and wrist but no 2RM has been developed for wrist-worn GENEA and AX3 activity monitors. The development of a 2RM will advance future research by providing a more valid prediction method for estimating EE, which in turn can be used for more valid estimates of time spent in different PA intensity levels. The valid estimations of time spent in different PA intensity levels can be used in large scale studies to investigate associations between frequency, duration, and intensity of PA and various health outcomes.

CHAPTER III: MANUSCRIPT

Introduction

Valid methods for estimating physical activity (PA) energy expenditure (EE) are important to determine associations between PA and various health outcomes. A common method of estimating PA outcomes is the use of wearable sensors like heart rate monitors, pedometers, and accelerometers (7, 65). Accelerometer-based activity monitors are common wearable devices used by researchers and healthcare professionals (7, 65). Three common accelerometer-based activity monitors are the ActiGraph (ActiGraph, LLC, Pensacola, FL), GENEActiv (GENEA; Activinsights Ltd, Kimbolton, Cambridgeshire, UK), and Axivity AX3 (AX3, Axivity, Newcastle, UK). The ActiGraph is the most widely used activity monitor and the majority of prediction methods for estimating EE and time spent in different PA intensity levels have been developed for the ActiGraph. However, the GENEA has been utilized in prospective cohort studies such as the Whitehall II study (N = 10,314) (43), Fenland study (N = 1,695) (73), and the Cork and Kerry Diabetes and Heart Disease Study (N = 464) (36) and the AX3 is being used in the UK biobank study, a large-scale study with over 500,000 participants (18). Prediction methods have been developed for the GENEA, however no prediction method has been developed for the AX3.

Through the use of prediction equations and machine learning algorithms, researchers have been able to estimate EE and time spent in sedentary behavior (SB; <1.5 metabolic equivalents [METs]), light physical activity (LPA; 1.5 – 2.99 METs), moderate physical activity (MPA; 3.0 – 5.99 METs), and vigorous physical activity (VPA; ≥6.0 METs) using accelerometer-based activity monitors. Regression equations are the most common method for estimating EE. Initially, researchers constructed regression equations relating accelerometer

counts to EE based on treadmill walking and running (25). However, Pober et al. (51) showed that, at the same activity counts, adults performing intermittent activities can have 2-3 times higher EE than continuous walking and running (CWR). Thus, Crouter and colleagues developed two-regression models (2RM) that first apply a threshold that discriminates between sedentary and non-sedentary behavior based on count values, and second, based on the variability in count values to discriminate between CWR and intermittent activity, a separate regression equation is applied that predicts EE based on the count values (10, 13, 15).

Previously developed 2RMs using hip-worn ActiGraph in youth or wrist-worn ActiGraph in adults have improved estimates of EE and time spent in different PA intensity levels compared to single regression equations. Crouter et al. (14) conducted an independent validation in a free-living environment in youth, which compared estimated EE from youth-specific ActiGraph Crouter vertical axis (VA) and vector magnitude (VM) 2RMs (13) and the Freedson (25), Trueth (64), Trost (67), and Puyau (53) single regression equations to indirect calorimetry. The VA 2RM and VM 2RM had the lowest root mean squared error (RMSE) of 1.55 and 1.50 MET_y (youth metabolic equivalent, $MET_y = \frac{Activity\ VO_2}{Resting\ VO_2}$), respectively, which was 0.6% – 9.1% lower than the single regression equations. More recently, Hibbing et al. (30) developed 2RMs for the ActiGraph GT9X worn on the hip, left wrist, right wrist, left ankle and right ankle in adults. The 2RMs for the left wrist and right wrist had RMSEs of 1.24 METs and 1.29 METs, respectively. These were similar to the 2RMs developed for the hip, left ankle, and right ankle using the same dataset, which had RMSEs of 1.14 METs, 1.16 METs and 1.18 METs, respectively. This shows potential that the wrist location can be used for development of a 2RM.

No prediction equation has been developed for the AX3 activity monitor and only one single regression equation has been developed to predict EE in youth using a GENEA activity

monitor (32). The manufacturer of the AX3 recommends using the Hildebrand wrist specific model for estimations of EE when using the AX3 (2). In addition to the Hildebrand regression equation, Philips et al. (50) and Schaefer et al. (59) developed youth specific cut-points using ROC curve analyses to measure time spent in PA intensity levels using the GENEA in youth. To date, no 2RM has been developed for the GENEA or AX3 in youth. Thus, the purposes of this study were to: 1) develop left and right wrist 2RMs to predict youth metabolic equivalents (MET_y) for the AX3 and GENEA activity monitors in youth, 2) compare MET_y estimates from the left and right wrist GENEA and AX3 2RMs and the Hildebrand single regression equation to the Cosmed $K4b^2$, and 3) compare time spent in different PA intensity levels from the left and right wrist GENEA and AX3 2RMs, Hildebrand single regression equation, Phillips left and right wrist cut-points, and Schaefer cut-points to the $K4b^2$. Secondary purposes of this study were to 1) compare the estimates of MET_y and time spent in different PA intensity levels of the 2RMs when applied to data from a different wrist than it was developed and 2) compare the estimates of MET_y and time spent in different PA intensity levels of the 2RMs when applied to data from a different activity monitor than it was developed.

Methods

Participants

Youth (N = 100) between the ages of 6 and 18 years old were recruited from the greater Knoxville area via schools, after-school sports camps, word of mouth, and flyers. A parent or legal guardian of each participant signed a written informed consent and completed a health history questionnaire and each participant signed a written informed assent before participating in the study. Approximately 25 participants with a 50% split of males and females were recruited from each of the following four age groups: 1) 6 - 9 years old, 2) 10 - 12 years old, 3) 13 - 15 and

4) 16 - 18 years. Participants were excluded if their parents reported musculo-skeletal injuries within the past six months, any metabolic condition or medication that may alter metabolic processes, and cardio-respiratory conditions reported via a health history questionnaire. Participants were instructed to report to the laboratory having fasted for at least three hours and having refrained from exercise for the previous 24 hours. The University of Tennessee Knoxville Institutional Review Board approved the study before recruitment began.

Procedures

Data were collected at The University of Tennessee Knoxville Applied Physiology Laboratory in the Health, Physical Education, and Recreation building. Testing was performed on two visits on separate days. On the first day of testing, participants completed the informed consent and assent process, and anthropometric measurements were taken. All participants had their standing and seated height (cm) measured using a wall-mounted stadiometer (Seca Co., Hamburg, Germany). Body mass (kg) and body fat percentage were measured using a Tanita Body Composition Analyzer BC-418 segmented bioelectrical impedance analyzer (Tanita Co., Tokyo, Japan). Participants wore light weight athletic clothing and removed their shoes and socks prior to the anthropometric measurements. Participants then completed 30 minutes of supine rest to allow for the measurement resting metabolic rate (RMR), along with eight of the 16 activities which are summarized in Table 1. On day two of testing, the participants completed the remaining eight activities.

Participants completed each activity twice, once for 60 to 90 seconds and once for four to five minutes. Before the first visit, eight activities were randomly selected, and the order of the short and long bouts were randomized. Before testing, each participant was able to choose the order of activities, however the same activity was not allowed to be performed consecutively.

This study was a part of a larger study in which participants wore 14 activity monitors (five ActiGraph GT9X [left wrist, right wrist, right hip, left ankle, right ankle], Apple Watch 2, [left wrist], either a Fitbit Charge 2 or Samsung Gearfit 2 [right wrist], AX3 [left or right wrist], GENEA [left or right wrist], Mymo activity tracker [right hip], two misfit shine 2 [right hip and right shoe], and two ActivPals [right and left thigh]). The GENEA and AX3 are the only activity monitors that are used in the analyses for the current study. Each participant wore a GENEA and AX3 on opposite wrists. The wrist location (i.e. left or right wrist) of the monitors was switched every 25 participants until all 100 participants were collected. For example, the first 25 participants wore the GENEA on the left wrist and the AX3 on the right wrist, while the next 25 participants reversed the locations of those two activity monitors. Multiple monitors were worn on each wrist simultaneously with the most distal monitor (always ActiGraph GT9X) positioned at the level of the ulnar process on the posterior aspect of the wrist. The GENEA and AX3 were located proximal to the ActiGraph GT9X without touching. A Cosmed K4b^2 (K4b^2, Cosmed, Rome, Italy) indirect calorimeter was used as the criterion measure of EE.

Noldus Observer XT (Noldus International Technology, Wageningen, Netherlands) software was used to code activity behaviors in real time using a Samsung Galaxy Tab 4 tablet (Samsung, Seoul, South Korea). A two-class coding scheme was used to code the activity being performed and the posture of the participant during the activity. Duration of each activity was calculated from the time stamps corresponding with the start and end of an activity. Posture was recorded as lying, sitting, standing, or stepping. The tablet system clock for the Noldus Observer XT program was different than the GENEA, AX3, and K4b^2. To synchronize the live coding data with the accelerometer and metabolic data, a comment was inserted into the Noldus data containing the time of the PC system clock. This permitted alignment of the time series of

acceleration and Noldus data. If the participants and their parents consented, data collection was video recorded. Video recordings were used during the data cleaning process when atypical metabolic data was observed or to correct live coding errors (e.g. due to the device becoming dislodged, activity being interrupted, or an errant keystroke).

Equipment

Activity Monitors

GENEActiv: The GENEA is a small (43 x 40 x 13 mm) light weight (16g) activity monitor that is recommended to be worn on the wrist or hip. It has a range of up to ± 8 gravitational units (g's) in three planes of motion and can be initialized to collect data at sampling frequencies of 10 - 100 Hz in 10 Hz increments. Battery life and memory capacity of GENEA activity monitors can measure data for up to seven days at 100 Hz. For the present study, GENEA activity monitors were initialized to collect data at 100 Hz.

Axivity: The AX3 is a small (23 x 32.5 x 7.6) lightweight (9g) activity monitor that can be worn on the wrist, hip, ankle, or upper arm. It can collect data in the range of ± 2, 4, 8, or 16 g's in three planes of motion and can collect data at a sampling rates of 12.5 - 3200 Hz. Battery and memory capacity of the AX3 can measure data for up to 14 days at 100 Hz. In addition to the accelerometer, the AX3 activity monitor houses a real-time clock, 512 MB of memory, a temperature sensor, a light sensor, and is dust and water resistant. For the present study, AX3 activity monitors were initialized to record at a sample rate of 100 Hz with a seismic acceleration range of ± 8 g's.

Indirect Calorimetry

Cosmed K4b^2: The K4b^2 weighs 1.5 kg, which includes the battery and harness to hold the equipment. It measures ventilation rate, oxygen consumption (VO_2), and carbon dioxide

(CO_2) production on a breath by breath basis. The K4b[2] has been shown to be valid for measurement of VO_2 and CO_2 compared to Douglas bag measurements during light to vigorous intensity cycling on a stationary ergometer (42). Before the start of each test, a four step calibration was done: 1) room air calibration, 2) the gas analyzers were calibrated using a reference gas tank containing 15.93% O_2 and 4.92% CO_2, 3) the flow meter was calibrated using a Hans-Rudolph 3.00-liter syringe, and 4) a delay calibration was performed to adjust for the time lag between the expiratory flow and the expired gas fractions measurements. All calibration procedures were done according to manufactures instructions (8).

Live Coding and Video Recording

Noldus Observer XT: The Noldus Observer XT is a program designed for coding activities, either from videos or in real-time (i.e. live coding, for example in a laboratory setting). Using the Noldus Pocket Observer application (version 3.2), live coding is performed with Android tablets. Data collected from the Noldus Pocket Observer application can be downloaded to a PC using the Noldus Observer XT 12.5 software. Before the start of data collection, each researcher was trained on the Noldus Observer XT and pocket observer software.

A Canon Vixia HFR700 camcorder (Canon Inc., Melville, NY) with a Vivitar HD[4] MC AF High Definition 0.43X Wide Angle Converter with Macro Japan Optics attachment (Vivitar, Santa Monica, CA) was used to record the entirety of the data collection period. At the start and end of data collection, a camera shot of the system clock of the PC was taken to allow for synchronization of timestamps. The recording settings were set at a resolution of 1080p and 29 frames per second. A SanDisk Ultra 64GB SD card was used to store the video recordings.

Data Reduction and Cleaning

RMR was calculated from data collected during the 30-minute supine rest. A sliding window approach was used that examined breath-by-breath K4b^2 data in five-minute windows from minute 10 to the end of the RMR test (~30 minutes). Beginning at minute 15, each breath was averaged together with all the breaths from the preceding five minutes. This process was repeated for each succeeding breath until the end of the RMR test. The lowest 5-minute average VO_2 value was used as the measured RMR.

For each activity, the metabolic data were reduced to obtain a single MET_y value for each participant. The long bout of activity was used to compute steady state MET_y. Steady state MET_y was calculated by excluding the last the last ten seconds of the activity bout and using the previous sixty seconds. Using breath-by-breath data, all breaths that occurred within the 60-second window were averaged to obtain a steady-state absolute VO_2 ($ml \cdot min^{-1}$) and converted to relative VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$). Two kg were added to the participant's body mass for weightbearing activities to account for the weight of equipment. Relative VO_2 values were divided by measured RMR to convert to MET_y.

For acceleration data reduction, raw triaxial accelerometer data were collected at 100Hz, downloaded, and reduced down to one second averages in g's. Euclidean norm minus one (ENMO) was calculated on the raw 100 Hz data and averaged into one-second epochs. ENMO is a process that combines the three axes into VM for a single-orientation-independent value and subtracts one gravitational unit ($ENMO = \sqrt{x^2 + y^2 + z^2} - 1$). Any negative acceleration values after subtracting one gravitational unit were rounded up to zero. The average ENMO was calculated for each second of data and the 60 one-second values that matched what was used for

the metabolic steady state data for each activity were averaged and used for the calibration of the 2RM.

The metabolic data (VO_2 and MET_y) were cleaned by removing potential outliers. This was accomplished by removing any measured average VO_2 during RMR that were ≥ 2 SD away from the mean. Activity VO_2 data were removed if one or more of the following criteria were met: 1) the activity bout was under 220 seconds, 2) MET_y values ≤ 0.2 MET_y, which were physiologically unreasonable, or 3) MET_y values were ≥ 2 SD from the mean.

Model Development

2RMs were developed using techniques similar to those used by Crouter and colleagues (10, 13, 15, 30). The overall analytic dataset included 100 participants. Fifty participants wore AX3 on the right wrist and the GENEA on the left wrist. The other 50 participants wore the AX3 on the left wrist and GENEA on the right wrist. There were four subsets of data that correspond to the activity monitor being used and the wrist location, which were: 1) Axivity left wrist ($AX3_{LW}$), 2) Axivity right wrist ($AX3_{RW}$), 3) GENEActiv left wrist ($GENEA_{LW}$), and 4) GENEActiv right wrist ($GENEA_{RW}$). A 2RM was developed for each of the four subsets of data. Cycling was removed during the entire model development process.

All four 2RMs were developed using the same procedures as described below. 2RM development used all activities except cycling. Using the pROC package for R (54), a classifier that distinguishes between SB and non-SB was developed with receiver operating characteristic (ROC) curve analyses. During the development process, ENMO and log-transformed ENMO were investigated as predictor variables. The threshold for the classifier was selected by choosing the point closest to the top-left corner of the ROC curve using the 'closest.topleft' function which maximized sensitivity and specificity (26). SB was defined as any activity that had a measured

EE of <1.5 MET_y and was in a seated or lying position (63) while all other activities were defined as non-SB.

After the development of the SB classifier, a second classifier was developed that distinguishes between CWR and intermittent activity. Classification was made based on variability in acceleration because CWR activities are very consistent and rhythmic activities with low variations in acceleration while intermittent activities are often stop-and-go and have high variation in acceleration. The subset used for the development of the CWR classifier excluded all activities that were classified as SB from the first classifier. The activities used to classify CWR were slow walking, brisk walking, and running while all other activities were considered intermittent activity. The classifier was developed with ROC analysis, and the predictor variable was a coefficient of variation (CV) for ENMO or log-transformed ENMO. This was calculated based on the one-second ENMO values and the ENMO values of the preceding and succeeding nine seconds. Specifically, for each one-second ENMO value, ten CV's were calculated in the following manner: 1) one-second ENMO value and preceding nine one-second ENMO values, 2) one-second ENMO value, preceding eight one-second ENMO values, and succeeding one-second ENMO value, 3) one-second ENMO value, preceding seven one-second ENMO values, and succeeding two one-second ENMO values, and so on, up to 10) one-second ENMO value and succeeding nine one-second ENMO values. The CV used in analyses was the minimum of the ten CVs calculated.

Following classifier development, regression models were developed using a three-step process. First, the SB classifier was applied to distinguish between SB and non-SB. Second, the CWR classifier was applied to divide the non-SB subset into a CWR subset and intermittent activity subset. Third, using the CWR subset and intermittent activity subset from the second

step, regression equations were developed. Different models (i.e. linear [CWR and intermittent activity] and cubic [intermittent activity]) and expressions of ENMO data (i.e. ENMO and log-transformed ENMO) were examined for each regression equation. Further testing was conducted to determine if adding age as a predictor variable improved the fit of the models.

Any activity that was classified as SB was applied a MET_y value of 1.25, which differs from the norm of 1.00 MET_y for SB. The use of assigning 1.25 METs to SB has been used by Hibbing et al. (30) in the development of a 2RM in adults. The Sedentary Behavior Research Network defines SB as any activity that is under 1.5 METs and where the person is in a seated, reclining, or lying posture (63). The clear majority of SBs in the youth compendium have a MET_y value around 1.3 MET_y while only one activity (watching TV/movies while lying down) has an average MET_y value of 1.00 MET_y (5). Time use surveys suggest that screen-time constitutes for about half of all SB in youth living in the United States and Scotland (4, 41). Therefore, the use of 1.25 MET_y is a reasonable value to use for SB EE estimations.

Additional Models

Previously developed prediction models were applied for estimations of EE and time spent in different PA intensity levels using each of the following methods:

- The Hildebrand equation and cut-points were originally calibrated for the non-dominant wrist. The regression equation was: VO_2 [ml $O_2 \cdot kg^{-1} \cdot min^{-1}$] = 0.0357*mg* + 11.16. The cut-points for SB/LPA, MPA and VPA were: SB/LPA, <192 mg; MPA, 192-695 mg; VPA, ≥696 mg. Classification of SB and LPA were combined because when there is zero acceleration the estimation of VO_2 is 11.16 $ml \cdot kg^{-1} \cdot min^{-1}$, which is almost twice the average RMR of youth in the Hildebrand study (6 $ml \cdot kg^{-1} \cdot min^{-1}$). In addition to the cut-points, time spent in different SB, LPA, MPA, and VPA were estimated by converting

VO_2 to MET_y by dividing by measured RMR for each activity. Each MET_y value was classified as: SB, <1.5 MET_y; LPA, 1.5 MET_y -2.99 MET_y; MPA, 3.00 MET_y – 5.99 MET_y; and VPA, >6.00 MET_y;

- The Phillips cut-points for the right wrist are SB, ≤87.5 mg; LPA >87.5 to 275 mg; MPA, >275 to 700 mg; and VPA, >700 mg and left wrist are SB, ≤ 75 mg; LPA >75 to 250 mg; MPA, >250 to 750 mg; and VPA, >750 mg. Application of the Phillips cut-points is sampling frequency dependent because the calibration of the cut-points was developed using the sum of gravity-based accelerations. The cut-points were developed using 80 Hz data but the activity monitors in the present study were set to initialize collect data at 100 Hz. Therefore, the original Phillips cut-points were divided by 80 to adjust them to match match the collection frequency in the present study.

  The Schaefer cut-points were originally calibrated for the non-dominant wrist. The cut-points were calibrated using signal vector magnitude using the following equation: $SVM = \sum_{i=1}^{f} |\sqrt{x^2 + y^2 + z^2}| / Sampling\ Frequency.$ The cut-points are SB, < 190 mg; LPA, 190 to <314 mg; MPA, 314 to <998 mg; and VPA, ≥998 mg.

For the primary analysis, the newly developed 2RMs were applied to data from the same wrist and same device as it was developed on. For the secondary analysis, each 2RM was applied to data from: 1) the opposite wrist but same activity monitor, 2) the same wrist but opposite activity monitor, and 3) the opposite wrist and activity monitor it was developed for. For each activity monitor and wrist location, the Hildebrand single regression equation, Hildebrand cut-points, Phillips left and right wrist cut-points, and Schaefer cut-points were applied.

Statistical Analysis

Descriptive statistics on the participants' age, height, weight, and BMI were calculated. For each activity, average ENMO and CV for the 60 seconds of steady state activity were calculated for each of the four subsets. Model performance for each activity was assessed using a leave-one-participant-out cross-validation (LOOCV) for the $AX3_{LW}$ 2RM, $AX3_{RW}$ 2RM, $GENEA_{LW}$ 2RM, and $GENEA_{RW}$ 2RM. LOOCV is a process where model development is repeated on subsets of the data where each subset has one participant held out and every participant is held out from one subset. Within each subset, the model development procedure described above (i.e. SB and CWR classifier development and CWR and intermittent activity regression equation development) was completed. For the participant who was held out of the model development procedure, activity EE predictions were obtained from the newly developed 2RM. When all subsets had been processed, there were measured and predicted values for all activities performed by each participant. These values were then used to calculate root mean squared error (RMSE) and mean absolute percent error (MAPE).

In addition to cross-validation of EE of specific activities, cross-validation of EE and time spent in different activity intensity estimates was performed using the entire data collection period. The entire data collection period in the present study included the long bouts of activity, short bout of activity, and transitions between activities. In order to keep the data collection duration approximately the same duration, only participants with two complete days of data collection were included in this analysis.

Statistical analyses were completed using SPSS (version 25.0; SPSS Inc., Chicago, IL). Using the $AX3_{LW}$ data, two one-way repeated measures analysis of variance (ANOVA) were used to compare measured ($K4b^2$) and predicted MET_y for each activity and measured ($K4b^2$) and

predicted METy for the entire activity bout from the four 2RMs and the Hildebrand single regression equation. A third one-way repeated measures ANOVA was used to compare measured ($K4b^2$) and predicted time spent in different PA intensity levels from the four 2RMs, Hildebrand single regression equation, Hildebrand cut-points, Phillips left and right wrist cut-points, and Schaefer cut-points. Using one-way repeated measures ANOVA, the same comparisons were made using $AX3_{RW}$, $GENEA_{LW}$, and $GENEA_{RW}$ data. Planned contrasts with Bonferroni adjustments were performed to locate differences between the criterion measure and prediction methods for MET_y and time spent in different PA intensity levels. The significance level was set at $p < 0.05$ for all comparisons.

Results

Missing Data

One participant withdrew from the study after enrollment, before completing any testing. Six other participants withdrew from the study after completing one visit, and available data were included from those cases. Additional data loss occurred for participants who did not complete the RMR protocol (n = 3), stopped wearing the $K4b^2$ following the RMR protocol (n = 3), RMR values were > 2 SD (n = 5) from the mean, and removed after manual inspection of the values (n = 2). Seven AX3 and 11 GENEA download errors occurred.

Metabolic data cleaning resulted in a loss of data for individual activities for one or more of the following reasons: $K4b^2$ malfunctions (21 activities), activity duration was under the minimum duration of 220 seconds (74 activities), activity MET_y values under 0.2 (6 activities), and activity MET_y values ≥2 SD away from the mean (47 activities). In total, 800 observations (50 participants times 16 activities) were expected for each of the monitors and attachment sites. The four calibration subsets included 582 observations ($AX3_{LW}$), 577 observations ($AX3_{RW}$),

543 observations ($GENEA_{LW}$), and 541 observations ($GENEA_{RW}$). Participant characteristics are summarized in Table 2. Average ENMO across one minute of steady state activity and average of the one-second CV's across one minute of steady state activity for each activity are summarized in Table 3.

Model Development

SB and CWR classifier performance for the $AX3_{LW}$ 2RM, $AX3_{RW}$ 2RM, $GENEA_{LW}$ 2RM, and $GENEA_{RW}$ 2RM development is shown Table 4. Figure 1 shows the SB classifier threshold separating SB and non-SB for GENEA and AX3 data. Activities to the left of the threshold line are classified as SB while activities to the right of the line are classified as non-SB. The SB classifier sensitivities, specificities, and AUC for all devices and wrist locations were ≥91.6%, ≥97.1%, and ≥97.6%, respectively. Figure 2 shows the CWR classifier threshold separating CWR and intermittent activity for GENEA and AX3 data. Activities below the threshold line are classified as CWR, while activities above the threshold line are classified as intermittent activity. The CWR classifier sensitivities, specificities, and AUC for all devices and wrist locations were ≥85.3%, ≥87.5%, and ≥92.3%, respectively. Sensitivity, specificity, and AUC of the classifiers were the same regardless of whether ENMO or log-transformed ENMO were used as predictor variables.

MAPE and RMSE for different combinations of linear, logarithmic, and cubic transformations of ENMO with and without age added as a predictor variable when predicting MET_y are shown in Table 5. Log-transforming versus not log-transforming ENMO reduced RMSE by 0.06 ($GENEA_{RW}$ 2RM) – 0.10 ($GENEA_{LW}$ 2RM) MET_y and MAPE by 1.18% ($AX3_{RW}$ 2RM) – 2.67% ($GENEA_{LW}$ 2RM), respectively. Adding age as a predictor versus only using log(ENMO) as a predictor reduced RMSE by 0.03 ($GENEA_{RW}$ 2RM) – 0.06 ($AX3_{RW}$ 2RM)

MET_y and MAPE by 0.02% ($AX3_{LW}$ 2RM) – 1.12% ($AX3_{RW}$ 2RM). Therefore, log-transformed ENMO and age were chosen as predictor variables in the regression equations. To remain consistent with the regression equations, the classifiers also used log-transformed ENMO. Final classifier thresholds and regression equations for $AX3_{LW}$, $AX3_{RW}$, $GENEA_{LW}$, and $GENEA_{RW}$ data are shown in Table 6.

Estimation of METy for Structured Activities

Table 7 shows measured ($K4b^2$) and predicted MET_y from the four 2RMs and the Hildebrand single regression model when applied to $AX3_{LW}$ data. When applying the 2RM that was developed on the same activity monitor and wrist ($AX3_{LW}$ 2RM), five activities were significantly different from the $K4b^2$ by 0.08 MET_y (games) to 2.39 MET_y (cycling, $p < 0.05$). When the Hildebrand single regression equation was applied to $AX3_{LW}$ data, eight activities were significantly different from the $K4b^2$ by 1.03 MET_y (reclining) to 2.81 MET_y (jumping jacks, $p < 0.05$). When applying the 2RM developed on the same activity monitor but opposite wrist to $AX3_{LW}$ data, seven activities were significantly different from the $K4b^2$ by 0.08 MET_y (games) to 2.51 MET_y (cycling, $p < 0.05$). When applying the 2RMs that were developed on the opposite activity monitor to $AX3_{LW}$ data, ≥4 activities were significantly different from the $K4b^2$ by 0.08 MET_y (games) to 2.43 MET_y (cycling, $p < 0.05$).

Table 8 shows measured ($K4b^2$) and predicted MET_y from the four 2RMs and the Hildebrand single regression model when applied to $AX3_{RW}$ data. When applying the 2RM that was developed on the same activity monitor and wrist ($AX3_{RW}$ 2RM), six activities were significantly different from the $K4b^2$ by 0.45 MET_y (brisk walk) to 2.14 MET_y (cycling, $p < 0.05$). When the Hildebrand single regression equation was applied to $AX3_{RW}$ data, 13 activities were significantly different from the $K4b^2$ by 0.67 MET_y (dust) to 4.32 MET_y (jumping jacks, p <

0.05). When applying the 2RM developed on the same activity monitor but opposite wrist to $AX3_{RW}$ data, five activities were significantly different from the $K4b^2$ by 0.08 MET_y (games) to 2.09 MET_y (cycling, $p < 0.05$). When applying the 2RMs that were developed on the opposite activity monitor to $AX3_{RW}$ data, ≥4 activities were significantly different from the $K4b^2$ by 0.09 MET_y (games) to 2.10 MET_y (cycling, $p < 0.05$).

Table 9 shows measured ($K4b^2$) and predicted MET_y from the four 2RMs and the Hildebrand single regression model when applied to $GENEA_{LW}$ data. When applying the 2RM that was developed on the same activity monitor and wrist ($GENEA_{LW}$ 2RM),, four activities were significantly different from the $K4b^2$ by 0.10 MET_y (games) to 2.08 MET_y (cycling, $p < 0.05$). When the Hildebrand single regression equation was applied to $GENEA_{LW}$ data, 12 activities were significantly different from the $K4b^2$ by 0.47 MET_y (dust) to 4.39 MET_y (jumping jacks, $p < 0.05$). When applying the 2RM developed on the same activity monitor but opposite wrist to $GENEA_{LW}$ data, four activities were significantly different from the $K4b^2$ by 0.10 MET_y (games) to 2.10 MET_y (cycling, $p < 0.05$). When applying the 2RMs that were developed on the opposite activity monitor to $GENEA_{LW}$ data, ≥4 activities were significantly different from the $K4b^2$ by 0.10 MET_y (games) to 2.07 MET_y (cycling, $p < 0.05$).

Table 10 shows measured ($K4b^2$) and predicted MET_y from the four 2RMs and the Hildebrand single regression model when applied to $GENEA_{RW}$ data. When applying the 2RM that was developed on the same activity monitor and wrist ($GENEA_{RW}$ 2RM),, four activities were significantly different from the $K4b^2$ by 0.10 MET_y (games) to 2.21 MET_y (cycling, $p < 0.05$). When the Hildebrand single regression equation was applied to $GENEA_{RW}$ data, 12 activities were significantly different from the $K4b^2$ by 0.47 MET_y (dust) to 4.39 MET_y (jumping jacks, $p < 0.05$). When applying the 2RM developed on the same activity monitor but opposite

wrist to $GENEA_{RW}$ data, six activities were significantly different from the $K4b^2$ by 0.08 MET_y (lying) to 2.19 MET_y (cycling, $p < 0.05$). When applying the 2RMs that were developed on the opposite activity monitor to $GENEA_{RW}$ data, ≥7 activities were significantly different from the $K4b^2$ by 0.08 MET_y (lying) to 2.21 MET_y (cycling, $p < 0.05$).

Estimation of MET_y for Entire Activity Routine

Descriptive statistics, RMSE, and mean biases when applying the four 2RMs and the Hildebrand single regression model to the four subsets of data for the entire activity routine are summarized in tables 11, 12, and 13, respectively. When the four 2RMs were applied to the same dataset in which they were developed, the $AX3_{RW}$ 2RM was the only model not statistically different from the $K4b^2$, while the other three 2RMS were statistically different from the $K4b^2$ by 0.26 MET_y ($GENEA_{LW}$ 2RM) to 0.31 MET_y ($GENEA_{RW}$ 2RM, $p < 0.05$). When the four 2RMs were applied to the same dataset in which they were developed, RMSE ranged from 0.46 MET_y ($GENEA_{LW}$ 2RM) to 0.58 MET_y ($GENEA_{RW}$ 2RM) and mean bias ranged from -0.07 MET_y ($AX3_{RW}$ 2RM) to 0.31 MET_y ($GENEA_{RW}$ 2RM). When applied to each dataset, the Hildebrand single regression equation was statistically different from the $K4b^2$ by 0.47 MET_y to 0.73 MET_y ($p < 0.05$), had a RMSE ranging from 0.93 – 1.09 MET_y, and a mean bias of -0.47 MET_y to -0.73 MET_y. When the four 2RMs were applied to data collected from the same activity monitor but opposite wrist in which they were developed, the $AX3_{RW}$ 2RM was the only model not significantly different, while the other three 2RMs were statistically different from the $K4b^2$ by 0.20 MET_y to 0.36 MET_y ($p < 0.05$). When the four 2RMs were applied to data collected from the same activity monitor but opposite wrist in which they were developed, RMSE ranged from 0.50 MET_y to 0.55 MET_y and mean bias ranged from -0.02 MET_y to 0.36 MET_y. When the four 2RMs were applied to data collected from the opposite activity monitor in which they were

developed, the $AX3_{RW}$ 2RM was the only model not significantly different while the other three 2RMs were statistically different to the $K4b^2$ by 0.06 MET_y to 0.37 MET_y ($p < 0.05$). When the four 2RMs were applied to data collected from the opposite activity monitor in which they were developed, RMSE ranged from 0.40 MET_y to 0.60 MET_y and mean bias ranged from -0.06 MET_y to 0.37 MET_y.

Cross Validation of Models Predicting Time Spent in Different Physical Activity Intensities Levels

Figure 3 shows average minutes spent in SB, LPA, MPA, and VPA, measured from the $K4b^2$ and predicted from the four 2RMs, Hildebrand single regression equation, and four cut-point methods when applied to $AX3_{LW}$ data. Average data collection duration was 147.4 minutes for participants included in the $AX3_{LW}$ whole trial cross-validation. When applying the 2RM that was developed on the same activity monitor and wrist ($AX3_{LW}$ 2RM), minutes spent in SB and VPA were significantly different from the $K4b^2$ by 7.5 minutes and 6.1 minutes, respectively ($p < 0.05$). When the 2RM developed on the same monitor but opposite wrist was applied to $AX3_{LW}$ data, minutes spent in SB and LPA were significantly different from the $K4b^2$ by 8.3 minutes and 5.7 minutes, respectively ($p < 0.05$). When the 2RM developed on a different activity monitor but same wrist was applied to $AX3_{LW}$ data, minutes spent in VPA were significantly different from the $K4b^2$ by 5.7 minutes ($p < 0.05$). When the 2RM developed on a different activity monitor and different wrist was applied to $AX3_{LW}$ data, minutes spent in SB and VPA were significantly different ($p < 0.05$) from the $K4b^2$ by 9.1 minutes and 7.0 minutes, respectively. The five previously developed models vary widely in their estimates of time spent in different PA intensity levels. All models were significantly different from the $K4b^2$ for ≥2 PA intensity levels. The greatest differences for minutes spent in SB, LPA, MPA and VPA were

61.5 minutes (Schaefer, $p < 0.001$), 50.2 minutes (Hildebrand cut-points, $p < 0.001$), 27.0 minutes (Phillips right wrist cut-points, $p < 0.001$), and 6.4 minutes (Schaefer, $p < 0.001$), respectively.

Figure 4 shows average minutes of SB, LPA, MPA, and VPA, measured from the $K4b^2$ and predicted from the four 2RMs, Hildebrand single regression equation, and four cut-point methods when applied to $AX3_{RW}$ data. Average data collection duration was 141.3 minutes for participants included in the $AX3_{RW}$ whole trial cross-validation. When applying the 2RM that was developed on the same activity monitor and wrist ($AX3_{RW}$ 2RM), minutes spent in SB and LPA were significantly different from the $K4b^2$ by 8.1 minutes and 10.3 minutes, respectively ($p < 0.05$). When the 2RM developed on the same monitor but opposite wrist was applied to $AX3_{RW}$ data, minutes spent in SB and VPA were significantly different from the $K4b^2$ by 8.1 minutes and 6.6 minutes, respectively ($p < 0.05$). When the 2RM developed on a different activity monitor was applied to $AX3_{RW}$ data, minutes spent in SB was significantly different from the $K4b^2$ by 4.6 minutes to 9.0 minutes from the $K4b^2$, and minutes spent in VPA were significantly different by 5.7 minutes to 7.0 minutes from the $K4b^2$($p < 0.05$). The five previously developed models vary widely in their estimates of time spent in different PA intensity levels. All models were significantly different from the $K4b^2$ for ≥2 PA intensity levels. The greatest differences for minutes spent in SB, LPA, MPA and VPA were 55.9 minutes (Schaefer, $p < 0.001$), 45.9 minutes (Hildebrand cut-points, $p < 0.001$), 23.7 minutes (Phillips right wrist cut-points, $p < 0.001$), and 6.6 minutes (Schaefer, $p < 0.001$), respectively.

Figure 5 shows average minutes of SB, LPA, MPA, and VPA, measured from the $K4b^2$ and predicted from the four 2RMs, Hildebrand single regression equation, and four cut-point methods when applied to $GENEA_{LW}$ data. Average data collection duration was 134.5 minutes

for participants included in the $GENEA_{LW}$ whole trial cross-validation. When applying the 2RM that was developed on the same activity monitor and wrist ($GENEA_{LW}$ 2RM), minutes spent in SB and VPA were significantly different from the $K4b^2$ by 6.1 minutes and 6.9 minutes, respectively ($p < 0.05$). When the 2RM developed on the same monitor but opposite wrist was applied to $GENEA_{LW}$ data, minutes spent in SB and VPA were significantly different from the $K4b^2$ by 6.2 minutes and 8.1 minutes, respectively ($p < 0.05$). When the 2RM developed on a different activity monitor was applied to $GENEA_{LW}$ data, minutes spent in SB was significantly different from the $K4b^2$ by 6.7 minutes to 6.8 minutes, and minutes spent in VPA were significantly different from the $K4b^2$ by 6.9 minutes to 8.1 minutes ($p < 0.05$). The five previously developed models vary widely in their estimates of time spent in different PA intensity levels. All models were significantly different from the $K4b^2$ for ≥2 PA intensity levels. The greatest differences for SB, LPA, MPA and VPA were 53.4 minutes (Schaefer, $p < 0.001$), 45.1 minutes (Hildebrand cut-points, $p < 0.001$), 24.0 minutes (Phillips right wrist cut-points, $p < 0.001$), and 7.6 minutes (Schaefer, $p < 0.001$), respectively.

Figure 6 shows average minutes of SB, LPA, MPA, and VPA, measured from the $K4b^2$ and predicted from the four 2RMs, Hildebrand single regression equation, and four cut-point methods when applied to $GENEA_{RW}$ data. Average data collection duration was 140.4 minutes for participants included in the $GENEA_{RW}$ whole trial cross-validation. When applying the 2RM that was developed on the same activity monitor and wrist ($GENEA_{RW}$ 2RM), minutes spent in SB and VPA were significantly different from the $K4b^2$ by 6.8 minutes and 5.8 minutes, respectively ($p < 0.05$). When the 2RM developed on the same monitor but opposite wrist was applied to $GENEA_{RW}$ data, minutes spent in SB and VPA were significantly different from the $K4b^2$ by 4.1 minutes and 4.5 minutes, respectively ($p < 0.05$). When the 2RM developed on a

different activity monitor but same wrist was applied to $GENEA_{RW}$ data, minutes spent in VPA were significantly different from the $K4b^2$ by 4.5 minutes ($p < 0.05$). When the 2RM developed on a different activity monitor and different wrist was applied to $GENEA_{RW}$ data, minutes spent in SB and VPA were significantly different from the $K4b^2$ by 6.8 minutes and 5.9 minutes, respectively ($p < 0.05$). The five previously developed models vary widely in their estimates of time spent in different PA intensity levels. All models were significantly different from the $K4b^2$ for ≥2 PA intensity levels. The greatest differences for SB, LPA, MPA and VPA were 56.7 minutes (Schaefer, $p < 0.001$), 45.1 minutes (Hildebrand cut-points, $p < 0.001$), 23.5 minutes (Phillips right wrist cut-points, $p < 0.001$), and 5.1 minutes (Schaefer, $p < 0.001$), respectively.

Discussion

The primary aim of this study was to develop 2RMs for the left and right wrist for the GENEA and AX3 in youth. The three primary findings of this study were: 1) estimates of EE from the 2RMs when developed and validated using data from the same wrist and activity monitor had lower error compared to the $K4b^2$ than the Hildebrand single regression equation. 2) estimates of time spent in different PA intensity levels from the 2RMs when developed and validated using data from the same wrist and activity monitor had lower error compared to the $K4b^2$ than previously developed regression equations and cut-points, and 3) applying 2RMs to data from different activity monitors and wrist locations had minimal impact on their error compared to the $K4b^2$ for estimates of EE and time spent in different PA intensity levels.

The wrist-specific 2RM in youth appeared to improve estimates of EE and time spent in different PA intensity levels compared to previous single regression and ROC cut-points. However, it is important to note that the 2RMs were developed and cross-validated using the same activities which could have contributed to the lower RMSE for the 2RMs. In contrast, the

Hildebrand single regression equation was developed on different activities and participants than the current study. Although using the same activities biased the results in favor of the 2RMs, it is important to note that not all of the 2RMs were developed using the same group of participants. The $AX3_{LW}$ and $GENEA_{RW}$ subsets were drawn from the same group of 50 participants and the $AX3_{RW}$ and $GENEA_{LW}$ subsets were drawn from a different group of 50 participants.

Improvements for Estimation of Energy Expenditure

A reason that the 2RM had improved performance compared to the Hildebrand equation is in part due to the classification accuracy of SBs. All four of the SB classifiers in the current study had sensitivities over 91.6% and specificities over 97.1%, demonstrating a high level of classification accuracy between sedentary and non-sedentary behaviors. The high level of classifications accuracy leads to estimates of EE that are closer to the $K4b^2$ compared to the Hildebrand single regression equation. Previous research has shown the Hildebrand equation to significantly over-predict EE of SB (29). The inability to estimate the EE of SB using the Hildebrand single regression equation for the wrist in youth is due to the intercept of the equation being 11.16 $ml \cdot kg^{-1} \cdot min^{-1}$. This means that when there is zero acceleration, the minimum VO_2 that can be estimated is 11.16 $ml \cdot kg^{-1} \cdot min^{-1}$, which is almost twice as much as the average resting VO_2 of youth from the Hildebrand study (6 $ml \cdot kg^{-1} \cdot min^{-1}$). The results from the current study showed the Hildebrand equation significantly over-estimated the five sedentary activities by ≥1 MET_y whereas the four 2RMs were within 0.10 MET_y to the $K4b^2$, which is consistent with previous research.

A limitation of the 2RM for predicting MET_y arises in part from being able to correctly differentiate between CWR and intermittent activity. During the development of the CWR classifiers (with CV as the predictor variable), jumping jacks was misclassified as CWR 48.5%

to 59.0% of the time making it the most commonly misclassified activity. The high misclassification rate was because the variation in acceleration for jumping jacks is low, similar to CWR activities although the relationship between EE and acceleration for jumping jacks and CWR activities are different. Therefore, future researchers should be aware that activities with low-variability in acceleration will be classified as CWR and thus, impact EE predictions from the 2RMs (10, 13).

Previous research has investigated whether adding age as a predictor variable in youth regression equations improved estimations. Trost et al. (66) conducted a validation study examining multiple youth regression equations with participants between 5-15 years old. Trost and colleagues concluded that estimations of time spent in different intensities from single regression equations which include age as a predictor variable performed similarly to those without. The current study decided to add age as a predictor due to a decrease in MAPE and RMSE across all developed 2RMs compared to when age was not included which supports the use of age as a predictor variable. Preliminary results from the current study showed stratifying the sample into two age groups (6 - 12 years and 13 - 18 years) reduces error for the younger age group while increasing error in the older age group compared to indirect calorimetry. Future studies should investigate age-stratification within regression equations to predict EE and time spent in different PA intensity levels. Differences in metabolic rate and movement patterns between children and adolescents may support the implementation of age-stratification in regression equations.

Improvements for Estimation of Time Spent in Different Physical Activity Intensity Levels

In recent years, differences between regression equations and ROC cut-points for estimating time spent in different physical activity intensity levels has been investigated using the Actical and ActiGraph series of devices. Schaefer et al. developed regression equations and

ROC cut-points for a wrist-worn Actical in youth and using a free-living independent sample. The ROC cut-points gave higher estimates of time spent in MPA and VPA compared to regression analyses by 45.2 minutes and 12.3 minutes, respectively. However, the study had no criterion measure of time spent in MPA and VPA, so it is uncertain to know whether the ROC cut-points or regression equation provided closer estimates. Additionally, Crouter et al. (12) developed regression equations and ROC cut-points for the dominant wrist to estimate time spent in different physical activity intensity levels in youth using ActiGraph GT3X and GT3X+ accelerometers. Compared to indirect calorimetry, the ROC cut-points had a mean bias of 22% to 69%, which was higher than the mean bias of the regression equations (2% to 8%) for estimates of time spent in SB, LPA, MPA, and VPA. The results of the current study are consistent with the findings of Schaefer and Crouter, which showed ROC cut-points have greater error for estimating time spent in SB, LPA, MPA, and VPA compared to indirect calorimetry, supporting the use of a 2RM.

There has been increasing interest by researchers in developing prediction models that can estimate SB because increased time in SB independent of time spent in moderate-to-vigorous PA (MVPA) has been shown to be associated with negative health outcomes (27, 48). When estimating minutes spent in SB, the four 2RMs developed in this study provide closer estimates of minutes spent in SB to the $K4b^2$ than previously developed prediction methods. The Hildebrand single regression equation is the most commonly used prediction method for estimating EE and minutes spent in different PA intensity levels using the GENEA. The Hildebrand single regression equation has been shown to have error when estimating time spent in SB (31, 32). Therefore, Hildebrand developed cut-points that distinguish between SB and non-SB using the sample used to develop the single regression equations (31). However, it is

unknown how using the Hildebrand SB cut-point and the Hildebrand single regression equation together will impact estimates of LPA. In addition, previous research has shown that the Phillips and Schaefer cut-points over-estimate minutes spent in SB to direct observation (69). The results from the current study show the four 2RMs over-estimated time spent in SB to indirect calorimetry regardless of the activity monitor or wrist it was applied to by <9.1 minutes while all other methods were different by >27.2 minutes. Therefore, the 2RMs offered improvements in time spent in different PA intensity levels compared to ROC cut-points.

The ability to estimate time spent in MPA and VPA is important for youth interventions that focus on the association between MVPA and various health outcomes (47). Previous research has shown that the Hildebrand, Phillips, and Schaefer prediction models do not provide equivalent estimates of VPA to indirect calorimetry while the Hildebrand and Schaefer prediction models do not provide equivalent estimates of MPA to indirect calorimetry (47). The four 2RMs developed in the present study improved estimates of time spent in MPA compared to the Hildebrand, Phillips and Schaefer prediction models to indirect calorimetry. However, the $AX3_{LW}$ 2RM, $GENEA_{LW}$ 2RM and $GENEA_{RW}$ 2RM all significantly underestimated time spent in VPA from the $K4b^2$ while the $AX3_{RW}$ 2RM was not significantly different in estimating minutes of VPA from the $K4b^2$. It is important to note that the models are being validated using the same participants and the same activities that were used in the development process, but they also include short bouts of activity and transitions that were not used in the development process, which is how the 2RMs are applied by researchers.

Applying a 2RM on the Opposite Wrist or Different Monitor than it was Developed For

Accelerometers worn on the wrist have become popular with researchers because of the increased wear-time compliance (23) and the ability to estimate sleep duration and quality (65).

Researchers have been interested on the effect that attachment site at the wrist has on EE predictions (30, 32, 40, 44, 45, 50, 75). Hand dominance has been a common distinction made when developing prediction models using the GENEA because many daily activities (e.g. writing, eating, etc.) are typically performed with the dominant hand. Previous research has shown that applying site-specific models to the opposite wrist (e.g. applying a dominant wrist model to non-dominant wrist data) has little effect on estimations of EE and minutes spent in different PA intensity levels in adults (45) and youth (40). The results from the current study are consistent with previous studies. Mean bias and RMSE of the 2RMs were approximately the same regardless of whether it was being applied to data from the same or opposite wrist. In addition, the participants in the development of the $GENEA_{LW}$ and $AX3_{RW}$ 2RM were different than the $GENEA_{RW}$ and $AX3_{LW}$ 2RM, so cross-validating the 2RM on data using the opposite wrist but the same activity monitor used an independent sample of participants, which makes the comparisons in the current study less biased. However, future research should investigate the application of 2RMs on data using a different activity protocol, such as free-living data.

Researchers have also been interested in whether prediction models that are developed for one activity monitor can be applied to data from a different activity monitor. Harmonization between activity monitors will allow for more direct comparisons between studies to be made. Currently, the AX3 is being used in the UK BioBank study (18) while the GENEA is being used in prospective cohort studies like the Whitehall II (43) and Fenland (73) studies. In order for comparisons to be made between the studies, comparisons between the GENEA and AX3 have to be conducted. Previous research has shown that applying models that were developed using a GENEActiv on ActiGraph data or vice versa provide similar estimations of EE and minutes spent in different PA intensity levels (32, 35, 69), but no prior comparisons have been made

between the GENEA and AX3. In the current study, mean bias and RMSE for estimations of MET_y for the entire activity bout were approximately the same when applying 2RMs across activity monitor brands, indicating that any of the four 2RMs can be used when analyzing GENEA or AX3 raw acceleration data.

Strengths and Limitations

The present study has strengths and limitations. A strength of the study is development of both right and left wrist 2RMs. Previous regression equations developed for the GENEA in youth have usually been developed for one wrist, typically the non-dominant wrist. The exception is the Phillips wrist cut-points which were developed for left and right wrists separately. Another strength of the study is the 2RMs were developed using activities with varying intensities and multiple activity domains (e.g. household chores, sport and gaming, continuous locomotion). Using a wide range of intensities and types of activities allows for a more generalizable model. A limitation of this study is the 2RMs were cross-validated on the same activities in which they were developed. Using the same activities to develop and cross-validate the 2RM biased the results and made estimations of EE and time spent in different PA intensity levels appear more accurate than other single regression models and ROC cut-points that were developed in other studies and cross-validation in the present study. However, the participants in the development of the $GENEA_{LW}$ and $AX3_{RW}$ 2RM were different than the $GENEA_{RW}$ and $AX3_{LW}$ 2RM, providing cross-validation using independent participants. Future research should validate the 2RMs and the Hildebrand model using an independent sample in a free-living environment.

Conclusion

In conclusion, the current study has developed 2RMs for the left and right wrist using GENEA and AX3 activity monitors in youth. To our knowledge, these are the first prediction models that estimates EE and time spent in different PA intensity levels using the AX3. Compared to indirect calorimetry, when estimating MET_y, the 2RMs have lower RMSE and MAPE than the Hildebrand single regression equation. In addition, for estimating PA intensity levels, the 2RMs had lower mean bias than the Hildebrand single regression equation and cut-point methods. In addition, the findings from the present study suggest that applying 2RMs to the opposite wrist (i.e. left versus right wrist) or across activity monitor brand (AX3 versus GENEA) do not change estimates of MET_y or minutes spent in different PA intensity levels. Future work should validate the newly developed 2RMs in an independent free-living sample.

LIST OF REFERENCES

1. Atkin AJ, Gorely T, Clemes SA et al. Methods of measurement in epidemiology: sedentary Behaviour. *Int J Epidemiol*. 2012;41(5):1460-71.
2. Axivity Web site [Internet]. United Kingdom: Axivity Ltd. Available from: http://axivity.com/.
3. Becker DR, McClelland MM, Loprinzi P, Trost SG. Physical activity, self-regulation, and early academic achievement in preschool children. *Early Educ Dev*. 2014;25(1):56-70.
4. Biddle SJ, Gorely T, Marshall SJ, Cameron N. The prevalence of sedentary behavior and physical activity in leisure time: A study of Scottish adolescents using ecological momentary assessment. *Prev Med*. 2009;48(2):151-5.
5. Butte NF, Watson KB, Ridley K et al. A youth compendium of physical activities: activity codes and metabolic intensities. *Med Sci Sports Exerc*. 2018;50(2):246-56.
6. Chen KY, Bassett DR, Jr. The technology of accelerometry-based activity monitors: current and future. *Med Sci Sports Exerc*. 2005;37(11S):S490-500.
7. Chen KY, Janz KF, Zhu W, Brychta RJ. Redefining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc*. 2012;44(1S):S13-23.
8. Cosmed S.r.l. *K4b2 user manual*. Rome, Italy: Cosmed S.r.l.; 1998.
9. Crouter SE, Bassett DR, Jr. A new 2-regression model for the Actical accelerometer. *Br J Sports Med*. 2008;42(3):217-24.
10. Crouter SE, Clowers KG, Bassett DR, Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*. 2006;100(4):1324-31.
11. Crouter SE, Dellavalle DM, Horton M, Haas JD, Frongillo EA, Bassett DR, Jr. Validity of the Actical for estimating free-living physical activity. *Eur J Appl Physiol*. 2011;111(7):1381-9.
12. Crouter SE, Flynn JI, Bassett DR, Jr. Estimating physical activity in youth using a wrist accelerometer. *Med Sci Sports Exerc*. 2015;47(5):944-51.
13. Crouter SE, Horton M, Bassett DR, Jr. Use of a two-regression model for estimating energy expenditure in children. *Med Sci Sports Exerc*. 2012;44(6):1177-85.
14. Crouter SE, Horton M, Bassett DR, Jr. Validity of ActiGraph child-specific equations during various physical activities. *Med Sci Sports Exerc*. 2013;45(7):1403-9.
15. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR, Jr. Refined two-regression model for the ActiGraph accelerometer. *Med Sci Sports Exerc*. 2010;42(5):1029-37.
16. Dannecker KL, Sazonova NA, Melanson EL, Sazonov ES, Browning RC. A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*. 2013;45(11):2105-12.
17. de Graauw SM, de Groot JF, van Brussel M, Streur MF, Takken T. Review of prediction models to estimate activity-related energy expenditure in children and adolescents. *Int J Pediatr*. 2010;2010:489304.
18. Doherty A, Jackson D, Hammerla N et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One*. 2017;12(2):e0169649.
19. Duncan MJ, Wilson S, Tallis J, Eyre E. Validation of the Phillips et al. GENEActiv accelerometer wrist cut-points in children aged 5-8 years old. *Eur J Pediatr*. 2016;175(12):2019-21.
20. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA accelerometer. *Med Sci Sports Exerc*. 2011;43(6):1085-93.

21. Esliger DW, Tremblay MS. Physical activity and inactivity profiling: the next generation. *Can J Public Health*. 2007;98(2S):S195-207.
22. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. *J Sports Sci*. 2008;26(14):1557-65.
23. Fairclough SJ, Noonan R, Rowlands AV, van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. *Med Sci Sports Exerc*. 2016;48(2):245-53.
24. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc*. 2012;44(1S):S1-4.
25. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777-81.
26. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. *PLoS One*. 2014;9(12):e114468.
27. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too Little Exercise and Too Much Sitting: Inactivity Physiology and the Need for New Recommendations on Sedentary Behavior. *Curr Cardiovasc Risk Rep*. 2008;2(4):292-8.
28. Hanggi JM, Phillips LR, Rowlands AV. Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. *J Sci Med Sport*. 2013;16(1):40-4.
29. Hibbing PR, Ellingson LD, Dixon PM, Welk GJ. Adapted Sojourn Models to Estimate Activity Intensity in Youth: A Suite of Tools. *Med Sci Sports Exerc*. 2018;50(4):846-54.
30. Hibbing PR, LaMunion SR, Kaplan AS, Crouter SE. Estimating Energy Expenditure with ActiGraph GT9X Inertial Measurement Unit. *Med Sci Sports Exerc*. 2018;50(5):1093-102.
31. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. *Scand J Med Sci Sports*. 2017;27(12):1814-23.
32. Hildebrand M, van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc*. 2014;46(9):1816-24.
33. Hinkley T, Teychenne M, Downing KL, Ball K, Salmon J, Hesketh KD. Early childhood physical activity, sedentary behaviors and psychosocial well-being: a systematic review. *Prev Med*. 2014;62:182-92.
34. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int J Behav Nutr Phys Act*. 2010;7:40.
35. John D, Sasaki J, Staudenmayer J, Mavilia M, Freedson PS. Comparison of raw acceleration from the GENEA and ActiGraph GT3X+ activity monitors. *Sensors (Basel)*. 2013;13(11):14754-63.
36. Kearney PM, Harrington JM, Mc Carthy VJ, Fitzgerald AP, Perry IJ. Cohort profile: The Cork and Kerry Diabetes and Heart Disease Study. *Int J Epidemiol*. 2013;42(5):1253-62.
37. Kim Y, Beets MW, Welk GJ. Everything you wanted to know about selecting the "right" ActiGraph accelerometer cut-points for youth, but...: a systematic review. *J Sci Med Sport*. 2012;15(4):311-21.

38. Kim Y, Lee JM, Peters BP, Gaesser GA, Welk GJ. Examination of different accelerometer cut-points for assessing sedentary behaviors in children. *PLoS One*. 2014;9(4):e90630.
39. Ladha C, Ladha K, Jackson D, Olivier P. Shaker table validation of Openmovement Ax3 accelerometer. In: *Proceedings of the 3rd International Conference on Ambulatory Monitoring of Physical Activity and Movement*. 2013, Jun 17-19: Amherst (United States of America). p. 69-70.
40. Mackintosh KA, Montoye AH, Pfeiffer KA, McNarry MA. Investigating optimal accelerometer placement for energy expenditure prediction in children using a machine learning approach. *Physiol Meas*. 2016;37(10):1728-40.
41. Matthews CE, Chen KY, Freedson PS et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. *Am J Epidemiol*. 2008;167(7):875-81.
42. McLaughlin JE, King GA, Howley ET, Bassett DR, Jr., Ainsworth BE. Validation of the Cosmed K4b2 portable metabolic system. *Int J Sports Med*. 2001;22(4):280-4.
43. Menai M, van Hees VT, Elbaz A, Kivimaki M, Singh-Manoux A, Sabia S. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. *Sci Rep*. 2017;8:45772.
44. Montoye AH, Mudd LM, Biswas S, Pfeiffer KA. Energy expenditure prediction using raw accelerometer data in simulated free living. *Med Sci Sports Exerc*. 2015;47(8):1735-46.
45. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiol Meas*. 2016;37(10):1770-84.
46. Morgan DW, Martin PE, Krahenbuhl GS. Factors affecting running economy. *Sports Med*. 1989;7(5):310-30.
47. Okely AD, Batterham MJ, Hinkley T et al. Wrist acceleration cut points for moderate-to-vigorous physical activity in youth. *Med Sci Sports Exerc*. 2018;50(3):609-16.
48. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. *Exerc Sport Sci Rev*. 2010;38(3):105-13.
49. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport*. 2017;20(1):75-80.
50. Phillips LR, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. *J Sci Med Sport*. 2013;16(2):124-8.
51. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc*. 2006;38(9):1626-34.
52. Powell C, Carson BP, Dowd KP, Donnelly AE. Simultaneous validation of five activity monitors for use in adult populations. *Scand J Med Sci Sports*. 2017;27(12):1881-92.
53. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res*. 2002;10(3):150-7.
54. Robin X, Turck N, Hainard A et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *Bioinformatics*. 2011;12:77.
55. Roscoe CMP, James RS, Duncan MJ. Calibration of GENEActiv accelerometer wrist cut-points for the assessment of physical activity intensity of preschool aged children. *Eur J Pediatr*. 2017;176(8):1093-8.

56. Rowlands AV, Rennie K, Kozarski R et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc*. 2014;46(12):2308-16.
57. Rowlands AV, Yates T, Davies M, Khunti K, Edwardson CL. Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Med Sci Sports Exerc*. 2016;48(10):1935-41.
58. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport*. 2000;71(2S):S1-14.
59. Schaefer CA, Nigg CR, Hill JO, Brink LA, Browning RC. Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth. *Med Sci Sports Exerc*. 2014;46(4):826-33.
60. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr*. 1985;39(1S):5-41.
61. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300-7.
62. Taraldsen K, Chastin SF, Riphagen, II, Vereijken B, Helbostad JL. Physical activity monitoring by use of accelerometer-based body-worn sensors in older adults: a systematic literature review of current knowledge and applications. *Maturitas*. 2012;71(1):13-9.
63. Tremblay MS, Aubert S, Barnes JD et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. *Int J Behav Nutr Phys Act*. 2017;14(1):75.
64. Treuth MS, Schmitz K, Catellier DJ et al. Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc*. 2004;36(7):1259-66.
65. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med*. 2014;48(13):1019-23.
66. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc*. 2011;43(7):1360-8.
67. Trost SG, Ward DS, Moorehead SM, Watson PD, Riner W, Burke JR. Validity of the computer science and applications (CSA) activity monitor in children. *Med Sci Sports Exerc*. 1998;30(4):629-33.
68. van Hees VT, Gorzelniak L, Dean Leon EC et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One*. 2013;8(4):e61691.
69. Van Loo CM, Okely AD, Batterham MJ et al. Wrist Accelerometer Cut Points for Classifying Sedentary Behavior in Children. *Med Sci Sports Exerc*. 2017;49(4):813-22.
70. Ward V, Schaefer CA, Lampe S. Achieving excellent compliance in multi-day free-living activity monitoring in elementary school students: the IPLAY study. *Med Sci Sports Exerc*. 2011;45(5):S486.
71. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Int J Obes Relat Metab Disord*. 1998;22(S2):S30-8.
72. Warren JM, Ekelund U, Besson H et al. Assessment of physical activity - a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17(2):127-39.

73. White T, Westgate K, Wareham NJ, Brage S. Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS One*. 2016;11(12):e0167472.
74. Wijndaele K, Westgate K, Stephens SK et al. Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. *Med Sci Sports Exerc*. 2015;47(10):2129-39.
75. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. *Med Sci Sports Exerc*. 2012;44(4):742-8.

APPENDIX

Tables

Table 1. List and descriptions of the activities.

Activity	Description of Activity
Supine Rest	Lying supine with arms by their side in a quiet room.
Internet Games	Sitting at a desk playing a self-selected internet computer game.
Surf Internet	Sitting at a desk browsing self-selected internet websites on a computer
Reclining	Sitting in a desk chair. If desired, participants were able to lean back in the chair and put feet up on a desk.
Book Reading	Sitting in a desk chair reading a book.
Dusting	Wiping down tables and similar surfaces with a paper towel and spray bottle.
*Slow Walking	Walking at a self-selected pace around an indoor basketball court or outdoor tennis court. Participants were instructed to walk at a leisurely pace.
Sweeping	Sweeping a pile of paper shreds in a hallway with a straw broom.
*Brisk Walking	Walking at a self-selected pace around an indoor basketball court or outdoor tennis court.. Participants were instructed to walk as if they were late for class.
Playing Catch	Passing a football with a partner at a comfortable distance.
Cycling	Riding a stationary ergometer at a self-selected pace and resistance.
Soccer	Soccer gameplay (e.g., dribbling, passing, shooting, etc.) with a partner.
Basketball	Basketball gameplay (e.g., playing one-on-one, dribbling, shooting, etc.) with a partner.
Stair Walking	Walking up and down a staircase in an indoor or outdoor location.
Jumping Jacks	Continuous jumping jacks.
Running	Running at a self-selected pace around an indoor basketball court or outdoor tennis court.. Participants were instructed to choose as pace they could sustain for 4-5 minutes.

Table 2. Descriptive characteristics of the participants of the final analytical dataset.

	Male (n = 41)	Female (n = 45)	Total (N = 86)
Age (years; Mean [SD])	12. (3.5)	12.2 (3.6)	12.2 (3.5)
6 – 9 years (n)	11	14	25
10 – 12 years (n)	12	10	22
13 – 15 years (n)	11	10	21
16 – 18 years (n)	7	11	18
Height (cm; Mean [SD])	152.1 (20.4)	148.4 (17.9)	150.2 (19.1)
Body Mass (kg; Mean [SD])	46.2 (20.1)	43.8 (17.2)	45.0 (18.7)
BMI Classification (%)			
Underweight (<5th Percentile)	4.9%	4.4%	4.7%
Normal Weight (5th - <85th Percentile)	78.0%	75.6%	76.7%
Overweight (85th - <95th Percentile)	7.3%	15.6%	11.6%
Obese (≥95th Percentile)	9.8%	4.4%	7.0%

cm = centimeter. kg = kilogram. BMI = body mass index.

Table 3. Mean (SD) acceleration (g) and coefficient of variation (CV) for 60 seconds of steady state activity for the Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) for each activity used in the calibration data set.

Activity	$AX3_{LW}$			$AX3_{RW}$			$GENEA_{LW}$			$GENEA_{RW}$		
	n	Acceleration (g)	CV	n	Acceleration (g)	CV	n	Acceleration (g)	CV	n	Acceleration (g)	CV
Supine Rest	37	0.013 (0.015)	18.39 (25.51)	36	0.013 (0.015)	23.85 (39.07)	33	0.009 (0.007)	21.98 (30.40)	37	0.017 (0.012)	11.41 (23.43)
Internet Games	34	0.011 (0.014)	41.98 (39.81)	36	0.009 (0.015)	64.75 (52.10)	32	0.015 (0.011)	41.98 (39.81)	34	0.024 (0.013)	16.09 (18.86)
Surf Internet	41	0.013 (0.014)	57.86 (51.00)	40	0.011 (0.018)	61.62 (47.41)	39	0.015 (0.011)	57.86 (51.00)	36	0.023 (0.011)	16.58 (15.19)
Reclining	39	0.016 (0.016)	37.18 (38.22)	39	0.011 (0.015)	49.53 (42.74)	36	0.017 (0.011)	37.18 (38.22)	33	0.016 (0.011)	28.62 (38.58)
Book Reading	36	0.024 (0.015)	41.19 (46.61)	38	0.011 (0.012)	52.80 (46.33)	37	0.010 (0.011)	41.19 (46.61)	35	0.015 (0.011)	45.18 (32.75)
Dusting	38	0.060 (0.035)	50.21 (16.59)	37	0.104 (0.055)	44.48 (12.58)	34	0.067 (0.035)	50.21 (16.59)	36	0.100 (0.065)	45.04 (12.36)
Slow Walking	40	0.155 (0.083)	20.62 (5.07)	36	0.146 (0.078)	20.53 (6.82)	34	0.239 (0.103)	20.62 (5.07)	34	0.141 (0.042)	21.51 (4.65)
Sweeping	41	0.089 (0.034)	37.13 (8.34)	34	0.097 (0.042)	37.98 (12.17)	34	0.103 (0.040)	37.13 (8.34)	35	0.093 (0.035)	41.96 (6.30)
Brisk Walking	33	0.239 (0.117)	18.83 (7.79)	37	0.234 (0.113)	17.88 (6.31)	34	0.239 (0.103)	18.83 (7.79)	32	0.225 (0.097)	17.90 (6.45)
Playing Catch	36	0.272 (0.096)	71.77 (15.79)	37	0.370 (0.109)	81.64 (14.06)	35	0.261 (0.091)	71.77 (15.79)	31	0.384 (0.126)	79.80 (18.39)
Cycling	35	0.033 (0.017)	32.50 (26.58)	3	0.023 (0.023)	52.56 (49.75)	34	0.042 (0.020)	32.50 (26.58)	34	0.039 (0.025)	21.00 (10.38)
Soccer	34	0.305 (0.134)	63.60 (17.12)	35	0.343 (0.194)	53.32 (13.89)	32	0.388 (0.202)	63.60 (17.12)	35	0.281 (0.114)	65.94 (18.09)
Basketball	34	0.371 (0.116)	60.34 (11.39)	28	0.451 (0.137)	58.58 (14.80)	26	0.435 (0.156)	60.34 (11.39)	31	0.414 (0.128)	63.54 (9.94)

Table 3. Continued.

Activity	$AX3_{LW}$			$AX3_{RW}$			$GENEA_{LW}$			$GENEA_{RW}$		
	n	Acceleration (g)	CV	n	Acceleration (g)	CV	n	Acceleration (g)	CV	n	Acceleration (g)	CV
Stair Walking	35	0.184 (0.052)	32.20 (5.39)	39	0.187 (0.047)	33.28 (7.33)	37	0.192 (0.051)	32.20 (5.39)	33	0.177 (0.056)	33.84 (7.20)
Jumping Jacks	33	0.897 (0.368)	28.90 (13.75)	39	1.051 (0.451)	26.01 (15.52)	36	1.045 (0.444)	28.90 (13.75)	34	0.968 (0.399)	27.93 (13.50)
Running	36	0.865 (0.191)	16.27 (7.24)	31	0.825 (0.228)	14.71 (4.81)	30	0.813 (0.202)	16.27 (7.24)	31	0.888 (0.205)	15.17 (5.80)

Table 4. Sedentary behavior (SB) and continuous walking and running (CWR) classifier sensitivities (SEN), specificities (SPEC), and area under the curve (AUC) for the Axivity left wright ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) 2-regression models

Activity Monitor and Location	SB			CWR		
	SEN	SPEC	AUC	SEN	SPEC	AUC
$AX3_{LW}$ (n = 43)	93.8%	97.1%	97.6%	87.8%	90.9%	92.9%
$AX3_{RW}$ (n = 43)	91.6%	100 %	96.7%	86.0%	87.5%	92.3%
$GENEA_{LW}$ (n = 42)	91.9%	98.7%	97.7%	85.3%	93.9%	93.2%
$GENEA_{RW}$ (n = 42)	95.0%	98.8%	98.5%	89.1%	93.8%	93.7%

Table 5. Root mean squared error (RMSE) and mean absolute percent error (MAPE) for Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$) data from leave-one-participant-out cross validation when using combinations of linear, cubic, and log-transformed acceleration with and without age as a predictor variable

Calibration Data	CWR	INT	ENMO and Age		Only ENMO	
			RMSE (MET_y)	MAPE (%)	RMSE (MET_y)	MAPE (%)
$AX3_{LW}$	Linear	Linear	0.92	20.85	0.96	21.20
	Linear	Cubic	0.89	19.96	0.93	20.21
	log(ENMO)	log(ENMO)	**0.85**	**19.21**	0.89	19.23
	log(ENMO)	Linear	0.90	20.30	0.93	20.28
	log(ENMO)	Cubic	0.86	19.40	0.90	19.28
	Linear	log(ENMO)	0.88	19.76	0.92	19.96
$AX3_{RW}$	Linear	Linear	1.02	21.89	1.06	22.73
	Linear	Cubic	1.00	21.48	1.05	22.42
	log(ENMO)	log(ENMO)	**0.95**	**20.71**	1.02	21.83
	log(ENMO)	Linear	0.97	21.81	1.03	22.11
	log(ENMO)	Cubic	0.96	20.77	1.02	21.80
	Linear	log(ENMO)	1.00	21.42	1.05	22.45
$GENEA_{LW}$	Linear	Linear	1.00	22.28	1.05	23.15
	Linear	Cubic	0.96	20.88	1.02	22.08
	log(ENMO)	log(ENMO)	**0.90**	**19.61**	0.96	20.65
	log(ENMO)	Linear	0.95	21.34	1.01	22.15
	log(ENMO)	Cubic	0.91	19.94	0.97	21.07
	Linear	log(ENMO)	0.95	20.56	1.02	21.65
$GENEA_{RW}$	Linear	Linear	0.90	20.65	0.92	20.92
	Linear	Cubic	0.88	20.08	0.90	20.44
	log(ENMO)	log(ENMO)	**0.84**	**19.44**	0.87	19.75
	log(ENMO)	Linear	0.87	20.02	0.90	20.22

Table 5. Continued

Calibration Data	CWR	INT	ENMO and Age		Only ENMO	
			RMSE (MET_y)	MAPE (%)	RMSE (MET_y)	MAPE (%)
$GENEA_{RW}$	log(ENMO)	Cubic	0.85	19.45	0.88	19.76
	Linear	log(ENMO)	0.87	20.07	0.90	20.43

ENMO = Euclidean norm minus one, CWR = continuous walking and running, INT = intermittent activity. Bold indicates the final model.

Table 6. Sedentary behavior (SB) and continuous walking and running (CWR) thresholds and CWR and intermittent activity regression equations developed for the Axivity left wright ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$), and GENEActiv right wrist ($GENEA_{RW}$)

Activity Monitor and Location	SB Threshold (1-s epoch, g)	CWR Threshold (CV)	Regression Equations
$AX3_{LW}$	log(ENMO) ≤ -3.33	log(ENMO) ≤ 25.8%	CWR MET_y = 4.500 + 1.568(log(ENMO [g's])) + 0.134(Age [yrs]) [0.63, 1.14] Intermittent Activity MET_y = 5.044 + 1.184(log(ENMO [g's])) + 0.075(Age [yrs]) [0.44, 1.06]
$AX3_{RW}$	log(ENMO) ≤ -3.08	log(ENMO) ≤ 27.9%	CWR MET_y = 5.992 + 1.447(log(ENMO [g's])) + 0.194(Age [yrs]) [0.51, 1.37] Intermittent Activity MET_y = 4.541 + 1.183(log(ENMO [g's])) + 0.102(Age [yrs]) [0.63, 1.14]
$GENEA_{LW}$	log(ENMO) ≤ -3.11	log(ENMO) ≤ 27%	CWR MET_y = 4.351 + 1.490(log(ENMO [g's])) + 0.151(Age [yrs]) [0.56, 1.30] Intermittent Activity MET_y = 4.489 + 1.274(log(ENMO [g's])) + 0.129(Age [yrs]) [0.45, 1.09]
$GENEA_{RW}$	log(ENMO) ≤ -2.93	log(ENMO) ≤ 25.6%	CWR MET_y = 4.779 + 1.487(log(ENMO [g's])) + 0.098(Age [yrs]) [0.68, 0.97] Intermittent Activity MET_y = 4.840 + 1.148(log(ENMO [g's])) + 0.071(Age [yrs]) [0.38, 1.13]

ENMO = Euclidean norm minus one. Values in brackets are [R^2, Standard error of the estimate (SEE)].

Table 7. Axivity AX3 Left Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities.

Activity	n	$K4b^2$	Axivity		*GENEActiv*		Hildebrand
			Left Wrist	*Right Wrist*	*Left Wrist*	*Right Wrist*	
Supine Rest	37	1.16 (0.02)	1.25 (0.00)*	*1.25 (0.00)**	*1.25 (0.00)**	*1.25 (0.00)**	2.22 (0.10)*
Games	34	1.17 (0.03)	1.25 (0.00)*	*1.25 (0.00)**	*1.25 (0.00)*	*1.25 (0.00)**	2.24 (0.10)*
Internet	41	1.17 (0.03)	1.25 (0.00)	*1.25 (0.00)*	*1.25 (0.00)*	*1.25 (0.00)*	2.25 (0.11)*
Reclining	39	1.18 (0.04)	1.25 (0.00)	*1.25 (0.00)*	*1.27 (0.03)*	*1.25 (0.00)*	2.21 (0.09)*
Book	36	1.22 (0.03)	1.30 (0.03)	*1.25 (0.00)*	*1.30 (0.03)*	*1.27 (0.01)*	2.29 (0.10)*
Dust	38	2.32 (0.07)	2.26 (0.13)	*2.02 (0.14)*	*2.37 (0.11)*	*2.09 (0.12)*	2.51 (0.11)
Slow Walk (mean speed = 74.5 $m \cdot min^{-1}$)	40	2.95 (0.09)	3.22 (0.11)	*3.47 (0.09)**	*3.25 (0.11)*	*3.15 (0.10)*	3.15 (0.15)
Sweep	41	3.02 (0.11)	2.97 (0.07)	*2.85 (0.09)*	*2.98 (0.07)*	*2.82 (0.06)*	2.73 (0.12)
Brisk Walk (mean speed = 96.6 $m \cdot min^{-1}$)	33	3.65 (0.15)	3.85 (0.13)	*4.06 (0.11)**	*3.85 (0.13)*	*3.76 (0.12)*	3.73 (0.20)
Catch	36	3.72 (0.13)	4.33 (0.07)*	*4.18 (0.07)**	*4.33 (0.07)**	*4.14 (0.07)**	3.90 (0.17)
Cycling	35	3.82 (0.15)	1.43 (0.06)*	*1.31 (0.05)**	*1.46 (0.07)**	*1.39 (0.05)**	2.37 (0.11)
Soccer	34	4.70 (0.23)	4.46 (0.09)	*4.30 (0.09)*	*4.46 (0.09)*	*4.27 (0.09)*	4.31 (0.27)
Basketball	34	4.72 (0.17)	4.71 (0.07)	*4.54 (0.07)*	*4.71 (0.07)*	*4.51 (0.07)*	4.62 (0.25)
Stairs	35	5.16 (0.21)	3.85 (0.07)*	*3.78 (0.06)**	*3.88 (0.06)**	*3.68 (0.06)**	3.42 (0.16)*
Jumping Jacks	33	5.73 (0.23)	5.88 (0.11)	*5.80 (0.12)*	*5.87 (0.11)*	*5.67 (0.11)*	8.54 (0.66)*
Running (mean speed = 155.0 $m \cdot min^{-1}$)	36	6.30 (0.27)	5.96 (0.06)	*6.06 (0.06)*	*5.96 (0.06)*	*5.75 (0.06)*	7.72 (0.38)*

Values are Mean (SD). *Significantly different from Cosmed $K4b^2$, $p < 0.05$. Italicized indicates models developed on the opposite wrist or activity monitor than the calibration subset.

Table 8. Axivity AX3 Right Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities.

Activity	n	$K4b^2$	Axivity		*GENEActiv*		Hildebrand
			Left Wrist	Right Wrist	*Left Wrist*	*Right Wrist*	
Supine Rest	36	1.20 (0.03)	*1.25 (0.00)*	1.25 (0.00)	*1.25 (0.00)*	*1.25 (0.00)*	2.33 (0.11)*
Games	36	1.18 (0.03)	*1.26 (0.01)**	1.25 (0.00)	*1.26 (0.01)*	*1.27 (0.01)**	2.31 (0.10)*
Internet	40	1.28 (0.04)	*1.28 (0.01)*	1.25 (0.00)	*1.28 (0.01)*	*1.28 (0.02)*	2.37 (0.11)*
Reclining	39	1.24 (0.04)	*1.28 (0.03)*	1.25 (0.00)	*1.27 (0.04)*	*1.28 (0.03)*	2.40 (0.11)*
Book	38	1.19 (0.03)	*1.25 (0.00)*	1.25 (0.00)	*1.27 (0.02)*	*1.25 (0.00)*	2.32 (0.11)*
Dust	37	2.39 (0.09)	*3.03 (0.15)**	2.94 (0.15)*	*3.11 (0.13)**	*2.90 (0.14)**	3.06 (0.16)*
Slow Walk (mean speed = 73.7 $m \cdot min^{-1}$)	36	2.85 (0.11)	*3.17 (0.12)*	3.40 (0.11)*	*3.18 (0.12)*	*3.08 (0.11)*	3.40 (0.16)*
Sweep	34	3.08 (0.12)	*3.07 (0.07)*	3.03 (0.06)	*3.07 (0.07)*	*2.88 (0.07)*	3.08 (0.15)
Brisk Walk (mean speed = 97.0 $m \cdot min^{-1}$)	37	3.60 (0.16)	*3.83 (0.12)*	4.05 (0.11)*	*3.83 (0.12)*	*3.72 (0.11)*	4.03 (0.23)
Catch	37	3.96 (0.13)	*4.74 (0.06)**	4.57 (0.05)*	*4.74 (0.06)**	*4.54 (0.06)**	4.91 (0.25)*
Cycling	35	3.52 (0.22)	*1.43 (0.08)**	1.38 (0.07)*	*1.49 (0.08)**	*1.42 (0.07)**	2.49 (0.12)*
Soccer	35	4.88 (0.28)	*4.53 (0.12)*	4.38 (0.11)	*4.53 (0.12)*	*4.34 (0.11)*	4.81 (0.37)
Basketball	28	4.54 (0.22)	*4.97 (0.07)*	4.80 (0.07)	*4.97 (0.07)*	*4.76 (0.07)*	5.32 (0.32)*
Stairs	39	5.09 (0.18)	*3.84 (0.07)**	3.80 (0.05)*	*3.88 (0.06)**	*3.68 (0.06)**	3.62 (0.15)*
Jumping Jacks	39	5.73 (0.22)	*6.03 (0.14)*	6.00 (0.14)	*6.03 (0.14)*	*5.81 (0.13)*	10.05 (0.75)*
Run (mean speed = 154.4 $m \cdot min^{-1}$)	31	6.67 (0.33)	*5.85 (0.08)*	5.95 (0.08)	*5.85 (0.08)*	*5.66 (0.08)**	8.31 (0.54)*

Values are Mean (SD). *Significantly different from Cosmed $K4b^2$, $p < 0.05$. Italicized indicates models developed on the opposite wrist or activity monitor than the calibration subset.

Table 9. GENEActiv Left Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities.

Activity	n	$K4b^2$	*Axivity*		GENEActiv		Hildebrand
			Left Wrist	*Right Wrist*	Left Wrist	*Right Wrist*	
Supine Rest	33	1.18 (0.03)	*1.25 (0.00)*	*1.25 (0.00)*	1.25 (0.00)	*1.25 (0.00)*	2.38 (0.12)*
Games	32	1.15 (0.03)	*1.25 (0.00)**	*1.25 (0.00)**	1.25 (0.00)*	*1.25 (0.00)**	2.42 (0.11)*
Internet	39	1.29 (0.04)	*1.25 (0.00)*	*1.25 (0.00)*	1.25 (0.00)	*1.25 (0.00)*	2.47 (0.11)*
Reclining	36	1.24 (0.05)	*1.25 (0.00)*	*1.25 (0.00)*	1.28 (0.03)	*1.25 (0.00)*	2.52 (0.11)*
Book	37	1.19 (0.03)	*1.25 (0.00)*	*1.25 (0.00)*	1.25 (0.00)	*1.25 (0.00)*	2.37 (0.11)*
Dust	34	2.39 (0.10)	*2.42 (0.16)*	*2.23 (0.16)*	2.51 (0.14)	*2.27 (0.15)*	2.86 (0.15)*
Slow Walk (mean speed = 73.7 $m \cdot min^{-1}$)	34	2.94 (0.11)	*3.27 (0.11)*	*3.52 (0.09)**	3.28 (0.11)	*3.20 (0.10)*	3.55 (0.15)*
Sweep	34	3.15 (0.13)	*3.08 (0.10)*	*3.01 (0.11)*	3.11 (0.08)	*2.94 (0.09)*	3.18 (0.14)
Brisk Walk (mean speed = 97.0 $m \cdot min^{-1}$)	34	3.57 (0.16)	*3.82 (0.11)*	*4.08 (0.11)**	3.82 (0.11)	*3.74 (0.11)*	4.11 (0.25)*
Catch	35	3.93 (0.14)	*4.28 (0.08)*	*4.13 (0.07)*	4.28 (0.08)	*4.10 (0.07)*	4.15 (0.18)
Cycling	34	3.58 (0.23)	*1.51 (0.09)**	*1.51 (0.10)**	1.50 (0.09)*	*1.49 (0.07)**	2.71 (0.13)*
Soccer	32	5.12 (0.28)	*4.67 (0.13)*	*4.52 (0.12)*	4.67 (0.13)	*4.48 (0.12)*	5.22 (0.38)
Basketball	26	4.66 (0.24)	*4.90 (0.09)*	*4.72 (0.09)*	4.90 (0.09)	*4.70 (0.09)*	5.35 (0.34)
Stairs	37	5.20 (0.18)	*3.94 (0.06)**	*3.86 (0.05)**	3.94 (0.06)*	*3.76 (0.06)**	3.78 (0.16)*
Jumping Jacks	36	5.87 (0.21)	*6.06 (0.14)*	*6.01 (0.15)*	6.04 (0.14)	*5.84 (0.14)*	10.26 (0.76)*
Run (mean speed = 154.4 $m \cdot min^{-1}$)	30	6.71 (0.33)	*5.83 (0.08)**	*5.95 (0.07)*	5.83 (0.08)*	*5.63 (0.07)**	8.28 (0.51)*

Values are Mean (SD). *Significantly different from Cosmed $K4b^2$, $p < 0.05$. Italicized indicates models developed on the opposite wrist or activity monitor than the calibration subset.

Table 10. GENEActiv Right Wrist Dataset: Measured (Cosmed K4b2) and predicted (Axivity and GENEActiv left and right wrist two-regression models and Hildebrand equation) youth metabolic equivalents (MET_y) for 16 structured activities.

Activity	n	$K4b^2$	*Axivity*		GENEActiv		Hildebrand
			Left Wrist	*Right Wrist*	*Left Wrist*	Right Wrist	
Supine Rest	37	1.17 (0.02)	*1.25 (0.00)**	*1.25 (0.00)**	*1.25 (0.00)**	1.25 (0.00)*	2.26 (0.10)*
Games	34	1.16 (0.03)	*1.27 (0.01)**	*1.25 (0.00)**	*1.26 (0.01)**	1.27 (0.01)*	2.30 (0.09)*
Internet	36	1.18 (0.03)	*1.25 (0.00)*	*1.25 (0.00)*	*1.23 (0.01)*	1.25 (0.00)	2.31 (0.10)*
Reclining	33	1.18 (0.04)	*1.25 (0.00)*	*1.25 (0.00)*	*1.24 (0.01)*	1.25 (0.00)	2.18 (0.09)*
Book	35	1.21 (0.03)	*1.25 (0.00)*	*1.25 (0.00)*	*1.25 (0.00)*	1.25 (0.00)	2.23 (0.09)*
Dust	36	2.27 (0.08)	*2.82 (0.17)**	*2.74 (0.16)**	*2.94 (0.14)**	2.69 (0.16)	2.72 (0.10)*
Slow Walk (mean speed = 73.6 $m \cdot min^{-1}$)	34	2.87 (0.09)	*3.11 (0.08)**	*3.42 (0.07)**	*3.16 (0.08)*	3.07 (0.07)	3.10 (0.13)
Sweep	35	2.96 (0.11)	*3.08 (0.07)*	*2.95 (0.08)*	*3.08 (0.07)*	2.92 (0.07)	2.71 (0.11)
Brisk Walk (mean speed = 96.6 $m \cdot min^{-1}$)	32	3.53 (0.15)	*3.78 (0.12)*	*4.01 (0.10)**	*3.79 (0.12)*	3.68 (0.11)	3.61 (0.18)
Catch	31	3.61 (0.14)	*4.77 (0.07)**	*4.60 (0.07)**	*4.77 (0.07)**	4.57 (0.07)*	4.63 (0.25)*
Cycling	34	3.78 (0.13)	*1.58 (0.09)**	*1.57 (0.10)**	*1.59 (0.09)**	1.57 (0.08)*	2.39 (0.11)*
Soccer	35	4.55 (0.24)	*4.36 (0.08)*	*4.21 (0.08)*	*4.36 (0.08)*	4.17 (0.08)	4.15 (0.23)
Basketball	31	4.72 (0.17)	*4.86 (0.07)*	*4.69 (0.07)*	*4.86 (0.07)*	4.66 (0.07)	5.06 (0.27)
Stairs	33	4.91 (0.19)	*3.81 (0.07)**	*3.73 (0.06)**	*3.81 (0.07)**	3.64 (0.07)*	3.28 (0.16)*
Jumping Jacks	34	5.58 (0.22)	*5.95 (0.11)*	*5.88 (0.12)*	*5.95 (0.11)*	5.73 (0.11)	8.88 (0.70)*
Run (mean speed = 155.2 $m \cdot min^{-1}$)	31	6.12 (0.24)	*5.98 (0.06)*	*6.07 (0.07)*	*5.98 (0.06)*	5.77 (0.06)	7.79 (0.40)*

Values are Mean (SD). *Significantly different from Cosmed $K4b^2$, $p < 0.05$. Italicized indicates models developed on the opposite wrist or activity monitor than the calibration subset.

Table 11. Mean (SD) measured youth metabolic equivalents (MET_y) from the Cosmed K4b2 and predicted MET_y from the Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$ 2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model across the entire data collection period.

Subset	Prediction Model					
	$K4b^2$	$AX3_{LW}$ 2RM	$AX3_{RW}$ 2RM	$GENEA_{LW}$ 2RM	$GENEA_{RW}$ 2RM	Hildebrand
$AX3_{LW}$ EE (MET_y) (n = 42)	2.73 (0.53)	**2.44 (0.23)***	2.75 (0.33)	2.50 (0.33)*	2.37 (0.21)*	3.20 (0.91)*
$AX3_{RW}$ EE (MET_y) (n = 41)	2.83 (0.49)	2.52 (0.24)*	**2.89 (0.33)**	2.58 (0.31)*	2.45 (0.21)*	3.56 (0.79)*
$GENEA_{LW}$ EE (MET_y) (n = 39)	2.76 (0.57)	2.47 (0.34)*	2.82 (0.45)	**2.50 (0.38)***	2.40 (0.32)*	3.50 (0.67)*
$GENEA_{RW}$ EE (MET_y) (n = 40)	2.70 (0.56)	2.46 (0.24)*	2.78 (0.33)	2.50 (0.31)*	**2.39 (0.22)***	3.34 (0.98)*

Bold indicates that the model was developed and applied on the same wrist and activity monitor

Table 12. Root mean squared error for Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$ 2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model for the entire data collection period

Subset	Prediction Model				
	$AX3_{LW}$ 2RM	$AX3_{RW}$ 2RM	$GENEA_{LW}$ 2RM	$GENEA_{RW}$ 2RM	Hildebrand
$AX3_{LW}$ EE (MET_y)	**0.56**	0.50	0.55	0.60	0.93
$AX3_{RW}$ EE (MET_y)	0.55	**0.50**	0.55	0.58	1.06
$GENEA_{LW}$ EE (MET_y)	0.50	0.40	**0.46**	0.54	1.09
$GENEA_{RW}$ EE (MET_y)	0.55	0.50	0.55	**0.58**	1.06

Bold indicates that the model was developed and applied on the same wrist and activity monitor

Table 13. Mean bias and prediction intervals (PI) for Axivity left wrist 2-regression model ($AX3_{LW}$ 2RM), Axivity right wrist 2-regression model ($AX3_{RW}$ 2RM), GENEActiv left wrist 2-regression model ($GENEA_{LW}$ 2RM), GENEActiv right wrist 2-regression model ($GENEA_{RW}$ 2RM), and Hildebrand single regression model for the entire data collection period. Units are youth metabolic equivalents (MET_y)

Subset	Prediction Method				
	$AX3_{LW}$ 2RM	$AX3_{RW}$ 2RM	$GENEA_{LW}$ 2RM	$GENEA_{RW}$ 2RM	Hildebrand
	Mean Bias (Upper PI, Lower PI)	Mean Bias (Upper PI, Lower PI)	Mean Bias (Upper PI, Lower PI)	Mean Bias (Upper PI, Lower PI)	Mean Bias (Upper PI, Lower PI)
$AX3_{LW}$ MET_y	**0.29 (-0.66, 1.24)**	-0.02 (-1.01, 0.96)	0.24 (-0.75, 1.22)	0.36 (-0.59, 1.31)	-0.47 (-2.06, 1.13)
$AX3_{RW}$ MET_y	0.30 (-0.52, 1.13)	**-0.07 (-0.91, 0.78)**	0.25 (-0.59, 1.09)	0.37 (-0.46, 1.21)	-0.73 (-2.46, 1.00)
$GENEA_{LW}$ MET_y	0.29 (-0.50, 1.09)	-0.06 (-0.84, 0.72)	**0.26 (-0.51, 1.03)**	0.36 (-0.44, 1.17)	-0.71 (-2.34, 0.92)
$GENEA_{RW}$ MET_y	0.24 (-0.74, 1.21)	-0.09 (1.07, 0.90)	0.20 (-0.81, 1.21)	**0.31 (-0.67, 1.28)**	-0.65 (-2.31, 1.02)

Bold indicates that the model was developed and applied on the same wrist and activity monitor

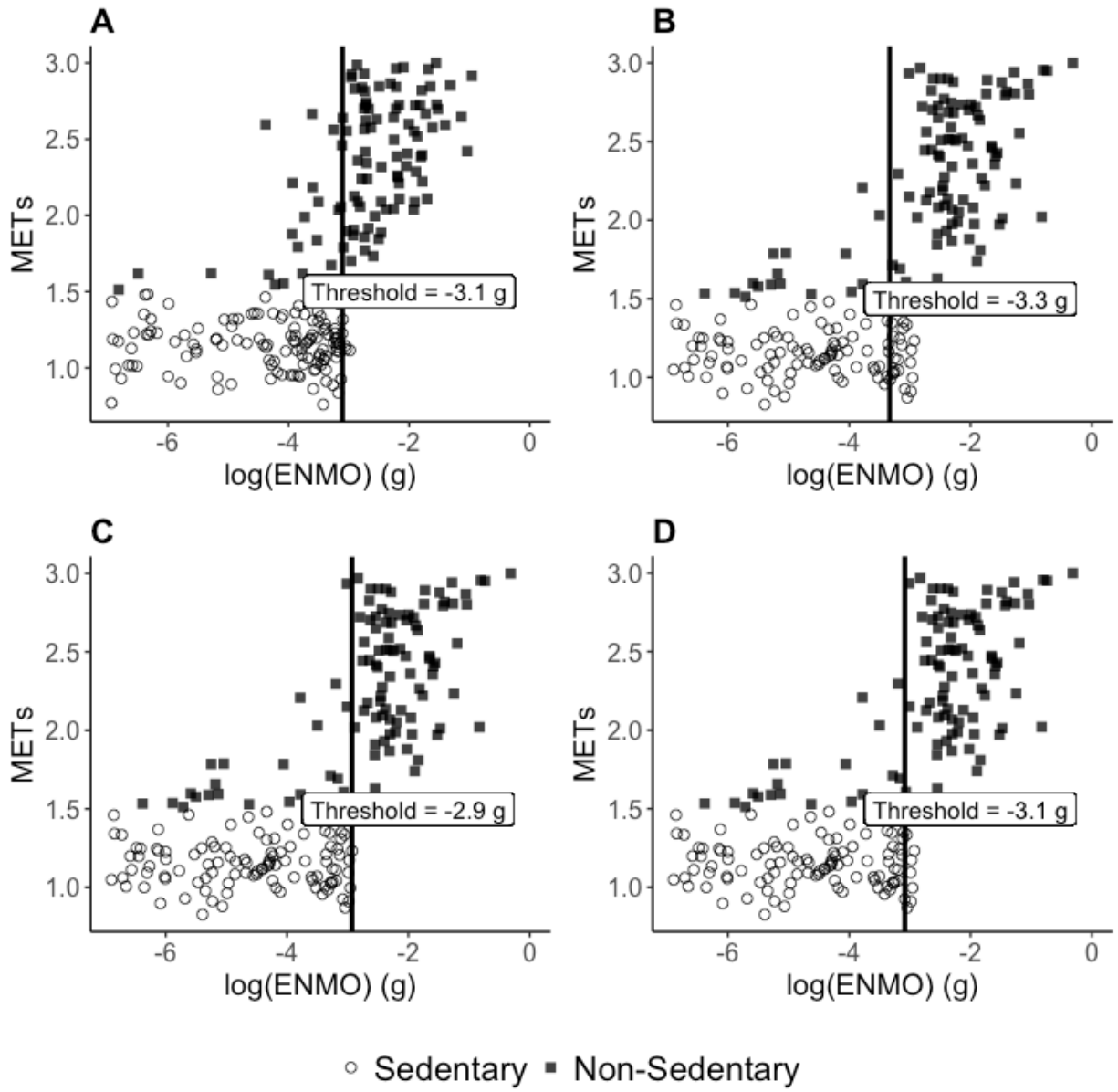

Figure 1. Classification between sedentary behavior and non-sedentary behavior for A) Axivity left wrist, B) GENEActiv left wrist, C) Axivity right wrist, and D) GENEActiv right wrist data using log(ENMO). Values to the left of the threshold are classified as sedentary while values to the right of the threshold are classified as non-sedentary.

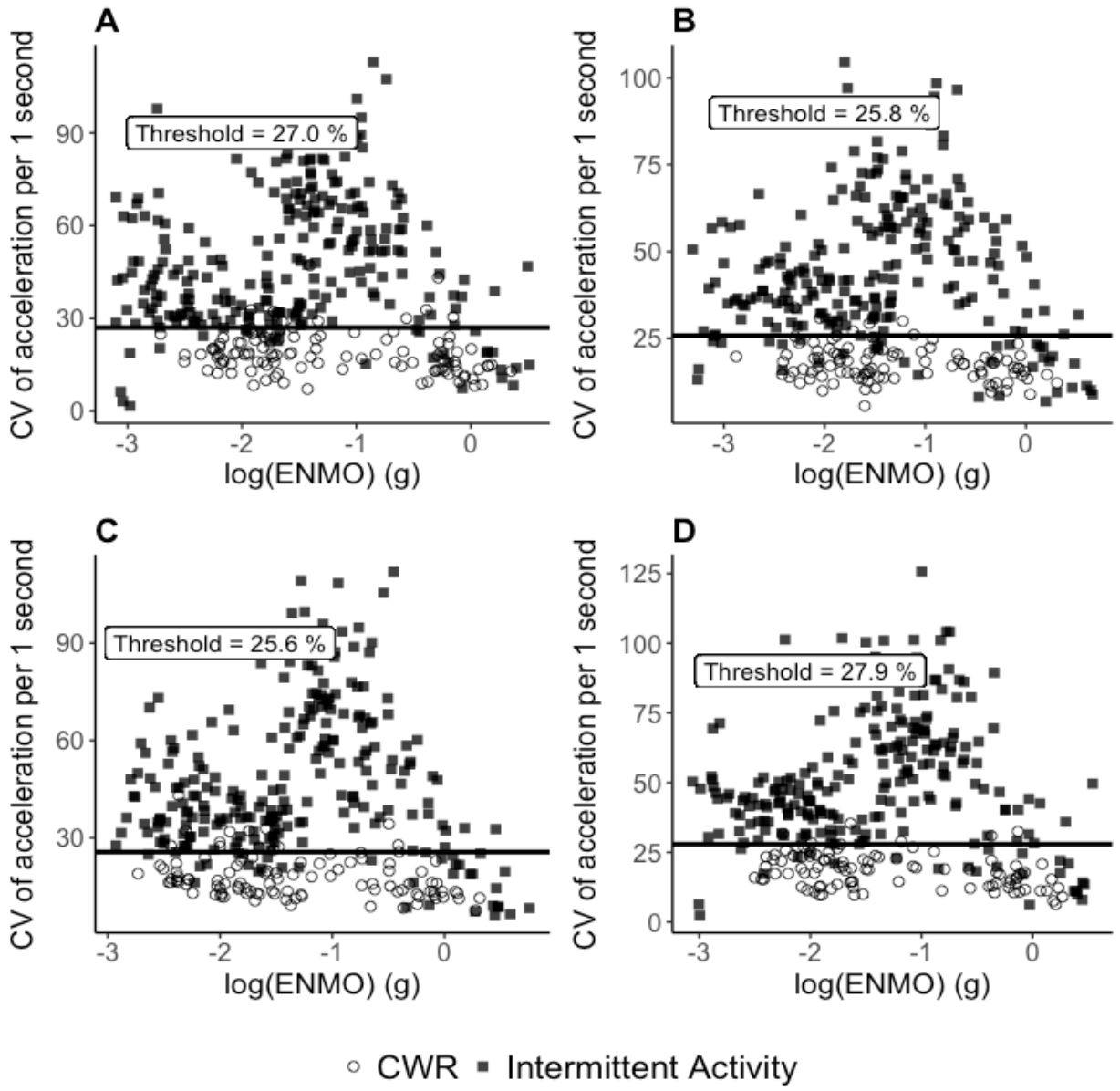

Figure 2. Classification between continuous walking and running and intermittent activity for A) Axivity left wrist, B) GENEActiv left wrist, C) Axivity right wrist, and D) GENEActiv right wrist data using the coefficient of variation (CV) of log(ENMO). Values below the threshold are classified as continuous walking and running while values above the threshold are classified as intermittent activity.

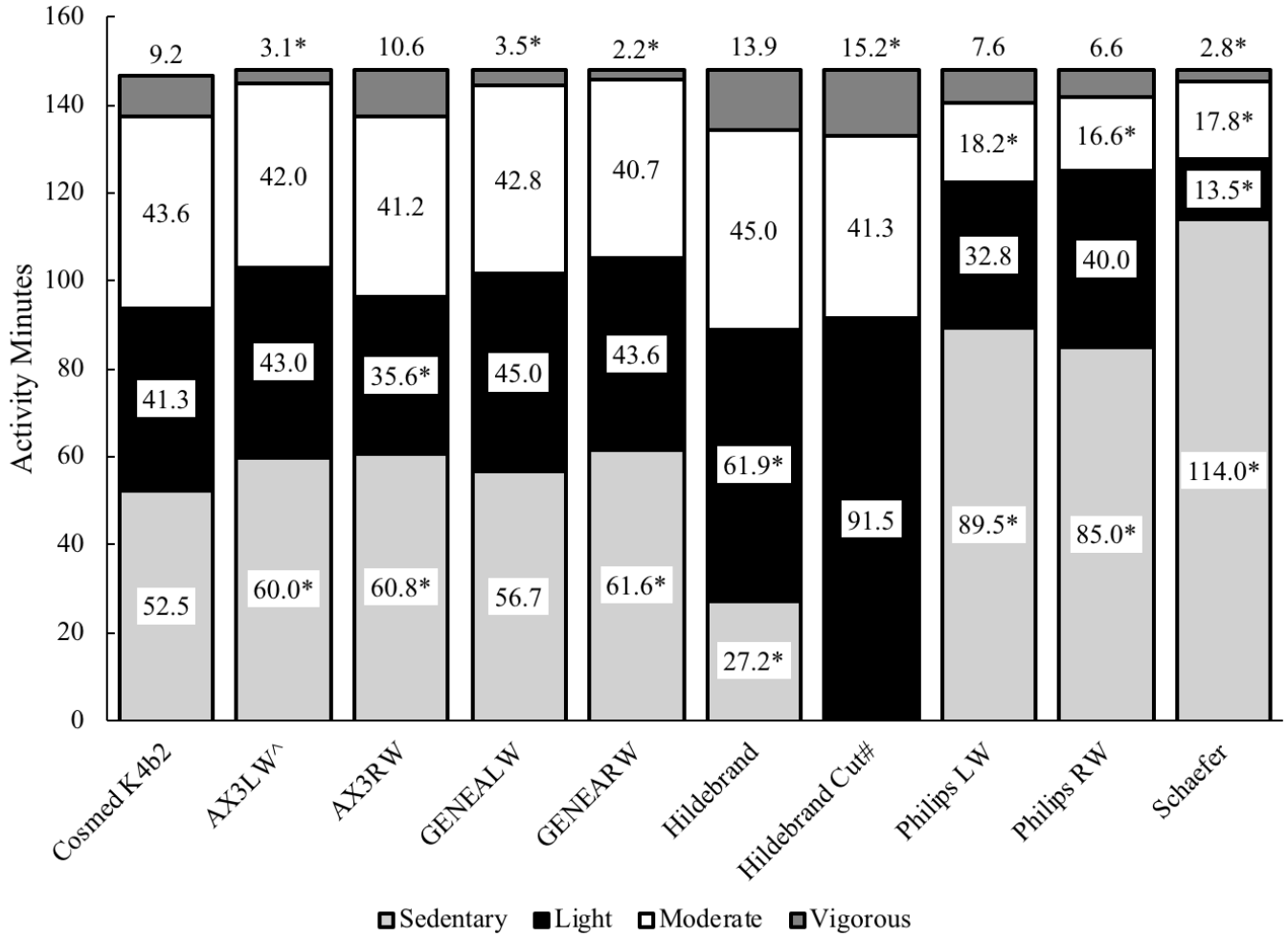

Figure 3. Axivity AX3 Left Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist (AX3LW), Axivity right wrist (AX3RW), GENEActiv left wrist (GENEALW) and GENEActiv right wrist (GENEARW) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed.

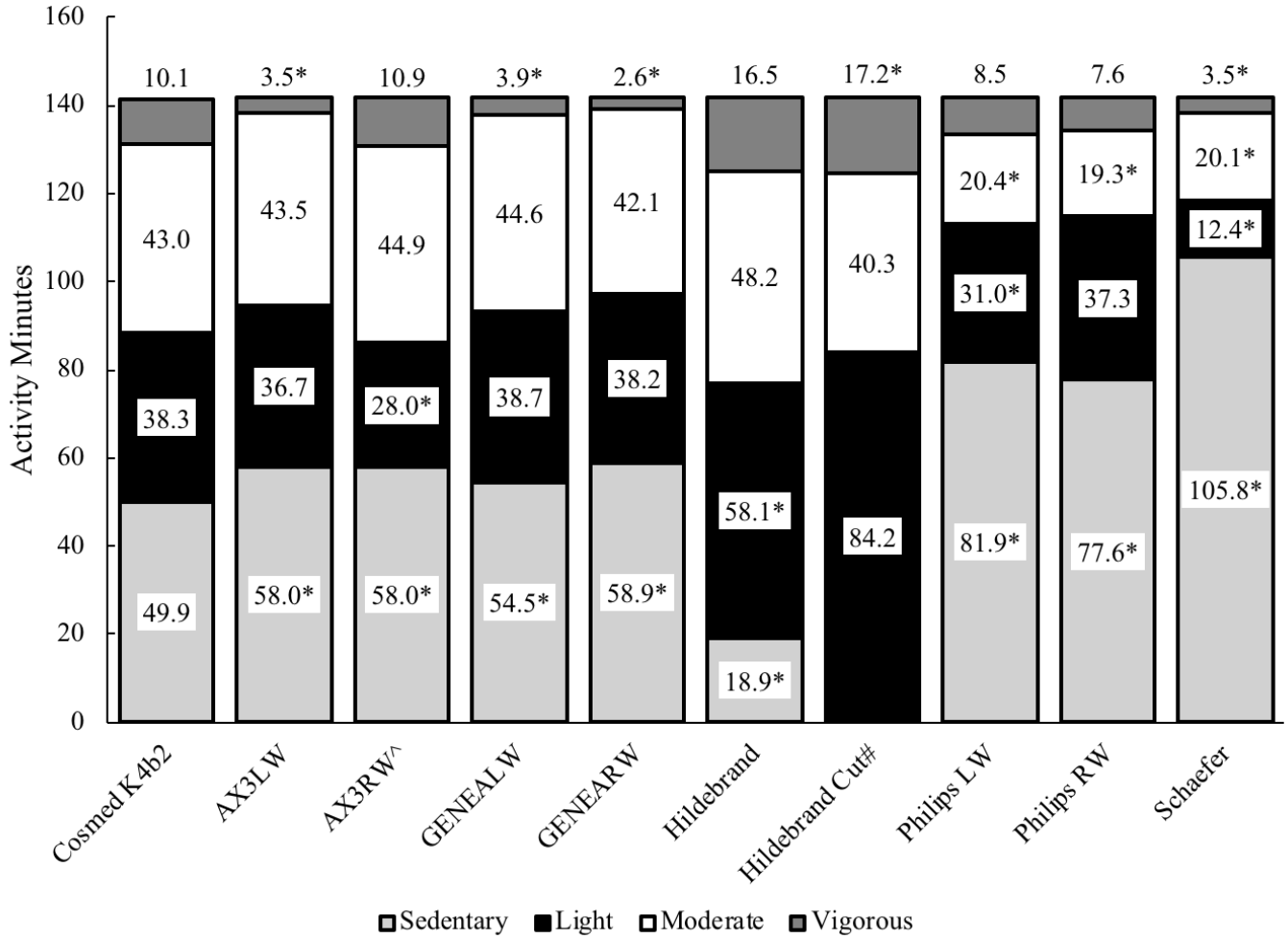

Figure 4. Axivity AX3 Right Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed.

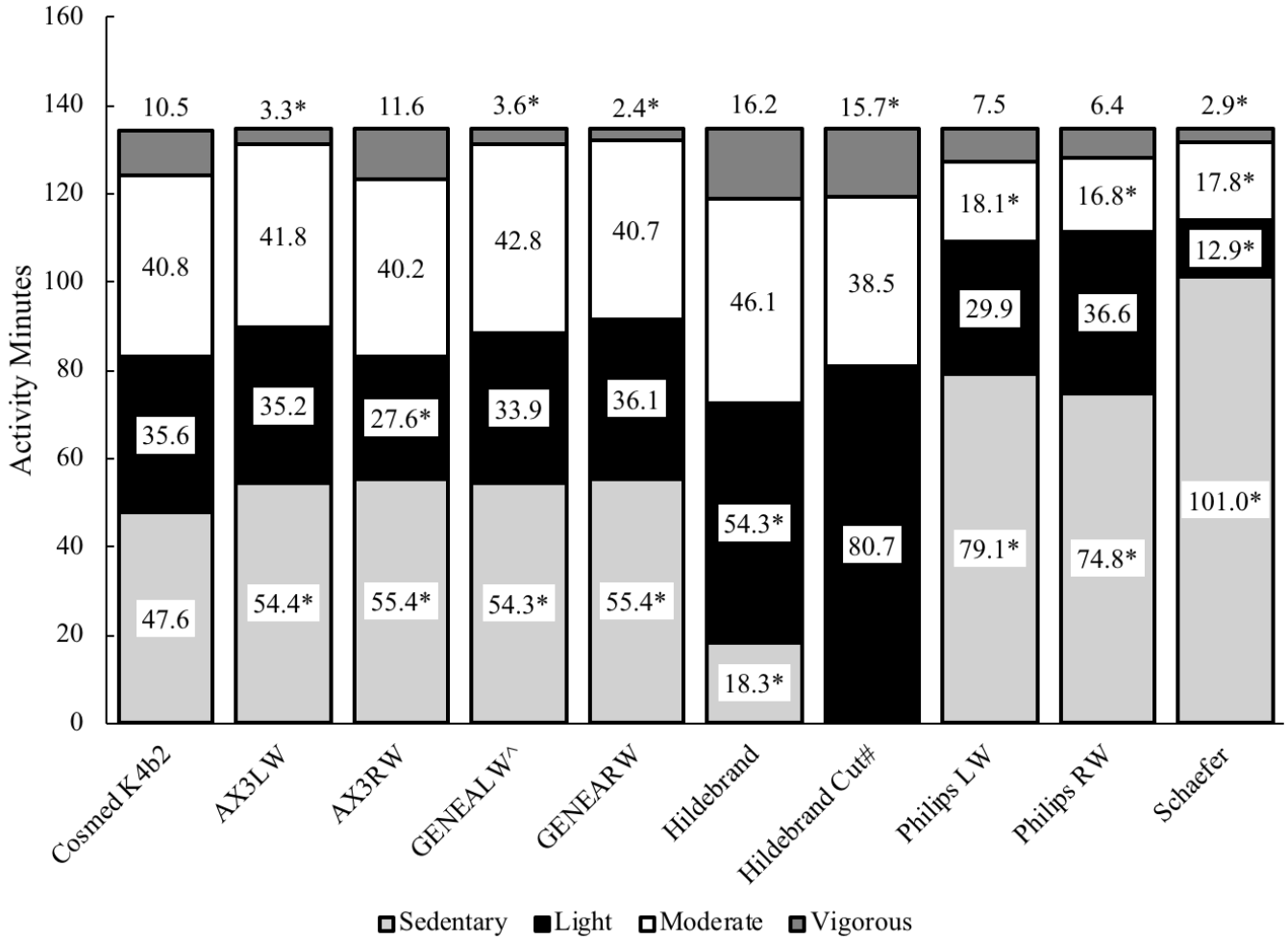

Figure 5. GENEActiv Left Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed.

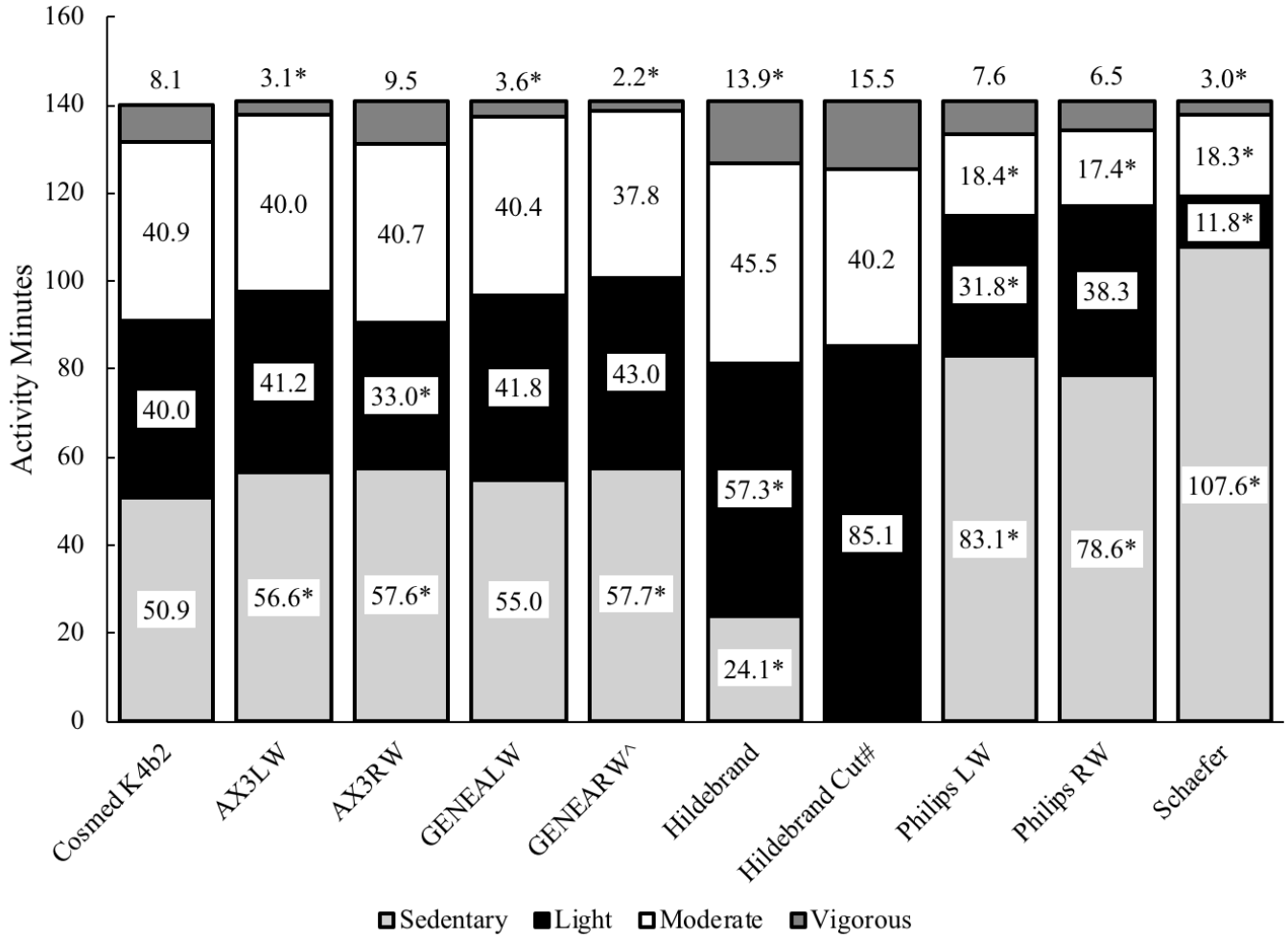

Figure 6. GENEActiv Right Wrist Dataset: Measured (Cosmed K4b^2) and predicted (Axivity left wrist ($AX3_{LW}$), Axivity right wrist ($AX3_{RW}$), GENEActiv left wrist ($GENEA_{LW}$) and GENEA right wrist ($GENEA_{RW}$) two-regression models, Hildebrand equation, and four cut-points) time spent in different PA intensity levels. *Statistically different from Cosmed K4b^2, $p < 0.05$. ^ The model was developed and applied on the same wrist and activity monitor. #Sedentary behavior and light intensity activity were combined because no cut-point was developed.

Informed Consent Form

Informed Consent Statement for Parent or Guardian

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 1 – Development of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to give permission for your child to participate in a research project that will examine new methods to measure the amount of physical activity youth obtain using small devices called accelerometers. These devices are attached to a belt which is worn around the waist, wrist, or ankle similar to a pedometer. Your child will be eligible if he or she is: 1) 6 to 18 years of age, and 2) you and your child can read and speak English. Your child will not be eligible if: 1) he or she has a serious health problem or injury that would prevent them from participating safely in regular physical activity, 2) he or she has a Pacemaker or other internal medical device.

The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

Here is what the study involves:

1) You will be asked to fill out a health history questionnaire pertaining to your child's health status to determine eligibility for the study.
2) Your child will be asked to visit the Applied Physiology Laboratory in the HPER building on 2 separate days for approximately 100 minutes on visit 1 and 60 minutes on visit 2, for a total time commitment of 160 minutes (2 hrs 40 minutes).
    a. During visit 1 your child will have his/her height, weight, percent body fat measured, resting metabolic rate measured, and will perform 8 activities randomly selected from a list of 16 activities.
    b. During visit 2 your child will perform the remaining 8 activities, from the list of 16 that they did not perform during visit 1. The 2 testing days will occur within 2 weeks of each other.
3) The percent body fat will be measured using bio-electrical impedance (BIA). BIA is a non-invasive technique that involves standing on a scale while also holding on to a handlebar. During this time an undetectable and harmless electrical current is passed through the body.
4) The resting metabolic rate test will take 35 minutes. This test will involve having your child rest in a lying position for 25 minutes and it will take 10 minutes to connect the equipment and prepare for the test. During this time he/she will have 3 accelerometers placed on their waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will also be attached to each thigh using a sticky pad. Your child will also be attached to a Cosmed portable system, which will measure the air that they breathe. This device is light weight and straps onto the body. It also requires the use of a facemask, over the nose and mouth. While wearing the facemask your child will be able to breathe normally and talk, and the mask does not block or restrict airflow.

_____ **Participant's Initials**

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/30/2017
IRB EXPIRATION DATE: 09/14/2018

5) Your child will perform 16 different activities. During each visit 8 activities will be performed which include:

   1) self-paced slow walk, 2) self-paced brisk walk, 3) self-paced running, 4) self-paced stair climbing, 5) shooting a basketball, 6) playing catch with a ball, 7) playing soccer, 8) sitting in a reclined position, 9) sitting upright while reading, 10) sitting upright while searching the internet, 11) sitting upright while playing computer games, 12) lying, 13) cycling at a moderate pace, 14) performing jumping jacks, 15) sweeping, and 16) dusting

   a. While performing the activities your child will be wearing the elastic belts with the accelerometers and the Cosmed metabolic cart as described above. Each visit your child will receive a list of 8 activities from which to choose from and they will be instructed to perform each activity for two times; once for between 60-90 seconds and once for between 4-5 minutes. A trained research assistant will work with your child to ensure they perform the activities correctly.
   b. While performing the activities a trained research assistant will also video record your child so that exact transition times between activities can be obtained. This is an essential part of this project for development of better tracking methods in real-world situations.

RISKS

There are minimal risks associated with participation in the research study. Since the study involves children engaging in physical activities, there is the possibility of minor injuries such as cuts, bruises, and muscle soreness, and a slight risk of sprains. The risk is similar to that involved in a physical education class or a community sport program. There is also a risk of skin irritation from wearing the sticky pad on the thigh, which will be minimized by cleaning the area after it is removed. If irritation does occur during testing it will be removed immediately. Additionally, there is also a potential for social and psychological risks associated with wearing the Cosmed $K4b^2$ and being followed during the activities. Lastly, there is a risk of the identity of your child being known through the use of the video recording.

BENEFITS

The benefits of the proposed research include improved methods for measuring free-living physical activity. This has implications for intervention researchers and those looking at the health of children and adolescents. Participants will learn about their general health characteristics (e.g. BMI) and strategies to become more active.

CONFIDENTIALITY

Information and records included in this study will be kept confidential. Data will be stored in a secure location and will only be made available to the people conducting the study, unless you specifically give permission in writing to do otherwise. The video recordings will be kept on a secure password protected server and only investigators involved with this study will have access to these files. The results of the study will be published, but no reference will be made in oral or written reports that could link participants to the study.

The data files will also be sent to study investigators at the University of Massachusetts Boston and the University of California Riverside. Only non-identifiable data will be sent to these investigators and they will not receive the videos of your child, thus they would not be able to identify who your child is. The Office of Human Research Protections (OHRP) may also review the study records at any time.

COMPENSATION

For participation in the study, your child will receive a $30 gift card for completion of visit 1 and a $45 gift card for completion of visit 2, for a total of $75 in gift cards if both days are completed.

_____ **Participant's Initials**

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/30/2017
IRB EXPIRATION DATE: 09/14/2018

EMERGENCY MEDICAL TREATMENT
The University of Tennessee does not "automatically" reimburse subjects for medical claims or other compensation. If physical injury is suffered in the course of research, or for more information, please notify the investigator in charge, Dr. Scott Crouter, at (865) 974-1272.

CONTACT INFORMATION
If you have questions at any time about the study or the procedures (or you experience adverse effects as a result of participating in this study), you should immediately contact the principal investigator, Dr. Scott Crouter, 334 HPER Building, The University of Tennessee, Knoxville, TN 37996, (865) 974-1272. If you have questions concerning your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-7697.

PARTICIPATION
Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

STATEMENT OF CONSENT
I have read the above information, and I have received a copy of this form. I consent for my child to participate in this study. I also certify that I am 18 years of age or older.

Printed name of your child _____

Parent or Legal Guardian signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT
I consent for my child to be videotaped while completing the activity described above. Not signing, does not exclude your child from the study and they can still participate, however they will not be videotaped.

Parent or Legal Guardian signature _____ Date _____

I consent for my child's videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, your child can still be videotaped for use in the current study, but there file will not be used for future training or educational purposes.

Parent or Legal Guardian signature _____ Date _____

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/30/2017
IRB EXPIRATION DATE: 09/14/2018

Informed Assent Form

Children's Assent form

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 1 – Development of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to participate in a research project that will use small devices called accelerometers to measure how much you move. These devices are attached to a belt and can be worn around the waist, hip, or ankle. The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter or a research assistant will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

If you decide that you want to be in this study, this is what you will be asked to do:

1) Visit the Applied Physiology Laboratory in the HPER building on 2 different days for approximately 1.5 hours on visit 1 and 1 hour on visit 2.
2) Have your height, weight, and percent body fat measured, which requires you to stand barefoot on a scale.
    a. During your 2 visits you will be asked to do the following tests:
    b. Resting metabolic rate (visit 1): During this test you will be asked to rest quietly while lying on a table. During the test you will have 3 small devices placed on your waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will also be attached to each thigh using a sticky pad. You will also have devices placed on your shoes and You will also be asked to wear a mask over your nose and mouth which will be connected to a device that is worn on your chest. The mask over your face will not affect how you breathe and you will be able to talk. This test will take 35 minutes to complete.
3) Perform 16 different activities (visits 1 and 2): During each visit 8 activities from the list below will be performed:

    1) walk slowly, 2) walk quickly, 3) run as fast or as slow as you want, 4) walk up and down stairs, 5) play basketball, 6) play catch with a ball, 7) playing soccer, 8) sit in a reclined position, 9) sit in a chair and read, 10) sit in a chair and play on the internet, 11) sit in a chair and play computer games, 12) lay on your back, 13) indoor cycling, 14) jumping jacks, 15) sweeping a floor, and 16) dusting a table

    a. While performing the activities you will be asked to wear the elastic belts facemask as you will do for the resting metabolic measurement test. You will do each activity two for either 60-90 seconds or 4-5 minutes. A research assistant will work with you to perform each activity.
    b. While doing the activities a research assistant will also video record what you are doing.

CAN ANYTHING BAD HAPPEN TO ME?

Although there are very few risks to taking part in this study, you could get hurt just like with any type of physical activity such as what you do in PE class. Dr. Crouter and the other research team members will help you to avoid being hurt or upset. There is also the risk that your skin will become itchy while wearing the sticky pad on your thigh.

_____ **Participant's Initials**

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/30/2017
IRB EXPIRATION DATE: 09/14/2018

CAN ANYTHING GOOD HAPPEN TO ME?
By participating in this study, you will learn about other physical activities that will be good for your health while also having fun.

DO I HAVE OTHER CHOICES?
You can choose not to be in this study if you don't want to.

WILL ANYONE KNOW I AM IN THE STUDY?
We won't tell anyone you took part in this study. Other people who are in the Applied Physiology Laboratory will know that you are being tested but we will not tell them any of your personal information. When we are done with the study, we will write a report about what we found out. We won't use your name in the report.

WHAT WILL YOU RECEIVE?
You will receive a $30 gift card for completing visit 1 and a $45 gift card for completing visit 2. If you complete both days of testing you will get a total of $75 in gift cards.

WHAT IF I DO NOT WANT TO DO THIS?
You don't have to be in this study. It's up to you. If you say yes now, but you change your mind later, that's okay too. All you have to do is tell us.

Before you tell us your decision about being a part of this study; be sure to ask Dr. Crouter or one of his research assistants to tell you more about anything you don't understand.

STATEMENT OF CONSENT
I have read the above information, and I have received a copy of this form. By signing this form I give assent to participate in this study.

Participant's signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT
I give assent to be videotaped while completing the activity described above. Not signing, does not exclude you from the study and you can still participate, however you will not be videotaped.

Participant's signature _____ Date _____

I give assent for my videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, you can still be videotaped for use in the current study, but there file will not be used for future training or educational purposes.

Participant's signature _____ Date _____

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/30/2017
IRB EXPIRATION DATE: 09/14/2018

Health History Questionnaire

HEALTH HISTORY QUESTIONNAIRE

Today's Date: _____

Name of parent/legal guardian completing the form:_____

ABOUT THE PARTICIPANT
Please answer the following questions about your child.

Name: _____

Address: _____

City: _____ **Zip Code:** _____

Parent's Phone: _____ **Date of Birth (month/day/year):** _____

Age: **Gender:** ___ M ___ F

Current School (if summer, School entering in Fall): _____

Current Grade (if summer, Grade entering in the Fall): _____

How do you identify your child?
_____Asian, Non-Hispanic
_____Asian, Hispanic
_____Black/African American, Non-Hispanic
_____Black/African American, Hispanic
_____Native Hawaiian/Pacific Islander, Non-Hispanic
_____Native Hawaiian/Pacific Islander, Hispanic
_____Native American/Alaskan, Non-Hispanic
_____Native American/Alaskan, Hispanic
_____White, Non-Hispanic
_____White, Hispanic

Is this child Multiracial? _____Y _____N

Who resides in the household with your child?
_____Both parents _____Single Mother _____Single Father _____Grandparent(s) _____Other

Please Turn Over

Health History
Has your child ever been diagnosed with any of the follow conditions? If yes, please explain.

	NO	YES	Current Condition	Explanation
Heart Disease				
Diabetes/High Blood Sugar				
High Blood Pressure				
Seizure				
Asthma or Other Lung Condition				
Cancer				
Musculoskeletal Condition or Injury				
Other Serious Illness				
Allergy				

Please describe any additional medical conditions that may affect your child's participation in physical activity.

_____

_____

_____

_____

_____

Is your child taking any medication (including prescription and non-prescription? If yes, please state below.

YES _____ NO _____

Name of Medication Reason for Taking For How Long?

_____

_____

_____

_____

_____

Emergency Contact

Name: _____

Relationship: _____ Phone: Work: _____

Home: _____

VITA

Andrew Scott Kaplan, born July 5, 1993 to Mark G. Kaplan and Jodell A. Keppler. Andrew completed his Bachelor's of Science in Kinesiology with a specialization in coaching at Michigan State University in May of 2015. In the Fall of 2016, Andrew enrolled in the Master's of Science in Kinesiology program with a concentration in Exercise Physiology at The University of Tennessee Knoxville. Andrew graduated from the Master's program in the August of 2016 at which time he enrolled at The University of Wisconsin Milwaukee for Fall 2018 to begin work as a doctoral candidate in Kinesiology with a concentration in Exercise Physiology. Andrew intends to pursue a teaching position at the University level upon completion of his doctoral studies.