Using Wearable Devices to Measure Physical Activity in Manual Wheelchair Users with Spinal Cord Injuries

by

Yousif Shwetar

Bachelor of Science in Bioengineering, University of Pittsburgh, 2020

Submitted to the Graduate Faculty of the

University Honors College in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

UNIVERSITY HONORS COLLEGE

This thesis/dissertation was presented

by

Yousif Shwetar

It was defended on

April 10, 2020

and approved by

Shivayogi Hiremath, PhD, Department of Health and Rehabilitation Sciences, Temple University

Alicia Koontz, PhD, Department of Rehabilitation Science and Technology,
University of Pittsburgh

Theresa Crytzer, PhD, Department of Rehabilitation Science and Technology,
University of Pittsburgh

Thesis Advisor: Dan Ding, PhD, Department of Rehabilitation Science and Technology,
University of Pittsburgh

Using Wearable Devices to Measure Physical Activity in Manual Wheelchair Users with Spinal Cord Injury

Yousif Shwetar, BS

University of Pittsburgh, 2020

Abstract

Manual wheelchair users (MWUs) with spinal cord injury (SCI) generally exhibit low levels of physical activity (PA), placing them at a greater risk for many chronic diseases. Accurately measuring levels of PA in this population could potentially lead to better health management among these individuals. Recently, there has been a growth in the use of wearable devices to help individuals track free-living PA for self-management. This has been explored extensively in the ambulatory population, specifically with research grade activity monitors such as ActiGraph wearable devices. However, the literature lacks adequate investigation for energy expenditure (EE) assessment and PA estimation using wearable devices in the non-ambulatory population. The objective of this thesis is to assess the ability of wearable devices in estimating EE and PA in wheelchair users with SCI. In the first study, we conducted a literature search for existing EE predictive algorithms using an ActiGraph activity monitor for MWUs with SCI and evaluated their validity using an out-of-sample dataset collected from MWUs with chronic SCI. None of the five sets of predictive equations demonstrated equivalence within 20% of the criterion measure based on an equivalence test. The mean absolute error (MAE) for the five sets of predictive equations ranged from 0.87 – 6.41 kilocalories per minute (kcal·min^{-1}) when compared with the criterion measure, and the intraclass correlation (ICC) estimates ranged from 0.06 – 0.59. Given the unsatisfactory performance of the existing EE predictive models, in the second study,

we used machine learning techniques to develop a random forest model (RFM) for activity intensity estimation using data collected from MWUs with SCIs. Based on a 10-fold cross validation, the RFM had an average overall accuracy of 81.3% in distinguishing among sedentary, light-intensity PA, and MVPA with a precision of 0.82, 0.77, and 0.87, and a recall of 0.84, 0.79, and 0.82 for each intensity category, respectively. The results indicate that the RFM could classify sedentary and MVPA time reasonably well, but may lack the ability to classify light-intensity PA.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I first would like to give thanks to Dr. Dan Ding for her persistent support and guidance throughout the entirety of this project and research. Her knowledge and mentoring have greatly influenced my goal to pursue medical research. As only an undergraduate, Dr. Ding has entrusted me with an array of tasks and responsibilities that I could've only dreamed of prior to working under her and with that I'm forever grateful for the path that she has helped create for me. I hope to follow in her footsteps as a prominent scientist in the near future.

I would also like to give thanks to the Human Engineering Research Laboratories faculty, staff, and students. All members of HERL at some point have assisted me with my work and I am thankful for all their contributions. Specifically, I would like to extend a special thanks to Akhila Veerubhotla. When I began working at the HERL, it was one of my first real engineering positions and with that came several difficulties. Akhila was patient with me the whole way helping me become the best team member I could be.

I am thankful for Dr. Arash Mahboobin. Not only has Dr. Mahboobin provided me substantial knowledge in several of his courses, but he has also provided me endless interpersonal support. Several times in my academic career I needed advice or guidance yet no idea what to do. In each of these times, Dr. Mahboobin has been available, able to provided me assistance and support.

I would like to extend my gratitude to Dr. Shivayogi Hiremath for agreeing to be my external examiner and providing valuable feedback for my dissertation.

Lastly, I would like to thank my parents, brother and sister for their emotional support and providing me with the opportunities that have led me to where I am today.

1.0 Comparative Validity of Energy Expenditure Prediction Algorithms Using Wearable Devices For People With Spinal Cord Injury

1.1 Introduction

It's estimated that up to 500,000 people worldwide every year suffer from SCI, with a majority occurring due to traffic incidents, falls, or violence [1]. Individuals with SCIs often experience decreased levels of mobility, and many of them use a wheelchair as a primary means of mobility [2]. Despite this regained mobility, these individuals still experience lower levels of mobility in comparison to their ambulatory counterparts and hence are subject to lower levels of PA [3]. Decreased levels of PA in this population has been associated with high incidences of chronic conditions such as type 2 diabetes, cardiovascular diseases, fatigue, weight gain, pain, and depression [4].

With the exponential growth of wearable devices in recent years, these devices are increasingly used to help people track free-living PA for self-management [5,6] as well as for supporting a wide variety of research [7]. One of the common measures provided by these devices is EE which is often used to monitor energy balance (i.e., the difference between food intake and EE) for weight management. EE is composed of three components including the resting metabolic rate (RMR) or resting energy expenditure (REE) that contributes 60%-75% to the overall EE, thermic effect of food (TEF) that contributes approximately 10%, and physical activity energy expenditure (PAEE) that is the most flexible component contributing 15%-30% to the overall EE, and is easily modifiable through PA participation [8]. For people with SCI, the measured REE is 14%-27% lower than people without disabilities due to the reduced fat-free mass and sympathetic nervous system activity [9]. These individuals, especially those who use wheelchairs for mobility, also tended to have lower PAEE due to the primary use of small muscle groups in the upper body [10]. With the lower overall EE in this population, they are at a higher risk for weight gain and associated health problems. Thus, it is important for the wearable devices to provide accurate feedback of everyday EE in people with SCI.

Many wearable devices on the market today are designed to be wrist-worn for better compliance, however, they are calibrated using a protocol that involves predominantly lower-extremity movement such as walking and running, which excludes their usage in the non-ambulatory population. In recent years, there has been some work on developing custom EE predictive models using wearable devices for wheelchair users [10]. Of the few commercial wearable devices used to estimate EE in MWUs with SCI, ActiGraph activity monitors (ActiGraph, LLC., Pensacola, FL, USA) have been the primary device of choice by researchers [11]. The ActiGraph devices feature a primary accelerometer and an inertia measurement unit (IMU), and provides user access to proprietary variables such as accelerometer counts as well as raw sensor signals at various frequencies. Several studies developed custom EE predictive equations based on ActiGraph activity monitors for MWUs with SCI [11-15]. These studies often used a metabolic cart to obtain the criterion measure of EE while participants perform a variety of PA wearing an ActiGraph device on the wrist. They then developed custom EE predictive equations that relate the outputs of ActiGraph devices with the criterion measure. As the performance of these predictive equations can be affected by the activity protocol and evaluation method, it is difficult to determine the comparative validity of these predictive equations and know which one(s) could be potentially used by future work.

The aim of this study was to conduct a literature search for existing EE predictive equations using ActiGraph activity monitors for MWUs with SCI, and evaluate their validity using an out-of-sample dataset. Using the same dataset collected separately from these studies to evaluate these predictive equations will provide an unbiased result to help guide appropriate and informed use of these predictive equations.

1.2 Methods

1.2.1 Existing EE Prediction Equations

A literature search was conducted to collect studies that had developed EE predictive equations based on wearable devices for MWUs. The eligibility criteria are 1) the output of the predictive equation should be in a form related to EE (e.g., overall EE, PAEE, and VO_2); 2) the input of the predictive equation should include variables from a wrist-worn ActiGraph activity monitor; and 3) the data used to develop the predictive equations should be from people with physical disabilities leading to the use of a manual wheelchair for mobility, and at least 25% of the sample should be MWUs with SCI. Three databases – PubMed, Institute of Electrical and Electronics Engineers (IEEE) and Scopus were used for the search. A set of search terms was used for wheelchair users, EE, and activity monitors including different spellings and synonyms. The search terms were then logically joined by “OR” and “AND”. The end date for the search was March 5th, 2019. The search of the three databases yielded 76 results, and four of them met the eligibility criteria [12-15]. An additional fifth study was acquired from a university thesis catalog [11]. A flow diagram describing the selection process is shown in figure 1.

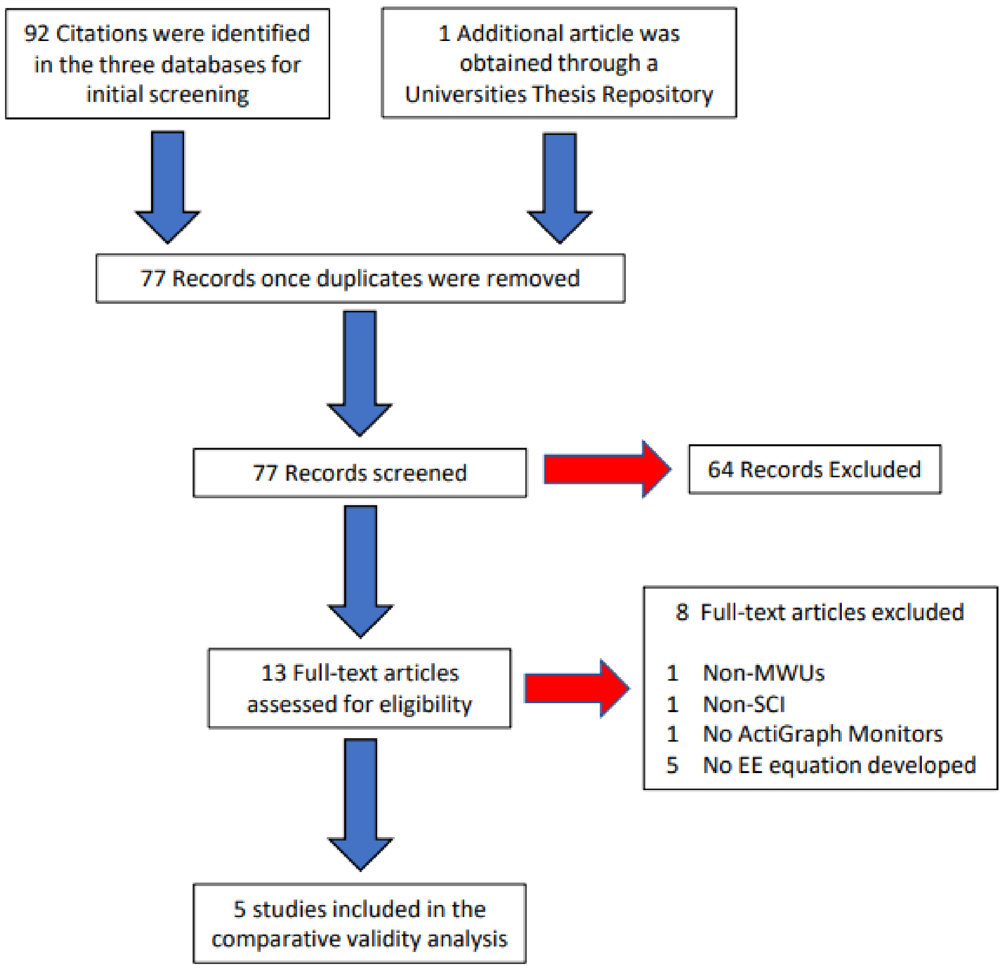


Figure 1: PRISMA flow diagram displaying selection process

The five sets of EE predictive equations and their related information are summarized in table 1. It is worth noting that the outputs of the predictive equations differ either by using different EE related variables or using different units. Eq. #1 [12] and #2 [13] predicted PAEE in unit of kcal·min^{-1} and kJ·min^{-1}, respectively, using per minute vector magnitude counts (VMC), a proprietary activity unit of ActiGraph devices, as the sole predictor variable. The criterion PAEE was obtained by subtracting the measured REE from the EE measured by a metabolic system during activities. Eq. #3 [14] predicted the VO_2 in units of ml·kg^{-1}·min^{-1}. It used variables that correspond to features extracted statistically and from the discrete wavelet transform (DWT) of the signals from each axis, i.e., x-axis counts (XC), y-axis counts (YC), z-axis counts (ZC), and VMC. The statistical features included the 50th, 75th, and 90th percentiles of each minute for each axis. The DWT features included the first level determination coefficient and the second level approximation coefficient. Eq. #4 [11] was from a non-peer-reviewed source (i.e., dissertation), which predicted EE in units of kcal·min^{-1} using features from the raw accelerometer signals as well as two basal metabolic rate (BMR) equations including the Mifflin/St. Jeor BMR equation (BMR1) and the World Health Organization (WHO) equation (BMR2) as predictors. Eq. #5 [15] includes two different equations, one for left-handed individuals, and one for right-handed individuals. They predict VO_2 in units of ml·kg^{-1}·min^{-1} using per minute VMC as predictors.

1.2.2 The Out-of-Sample Dataset

The out-of-sample dataset was collected from a study performed at two sites including the Human Engineering Research Laboratories (HERL), Pittsburgh, PA, and the James J Peters VA Medical Center, Bronx, NY. This study was approved by the US Department of Veterans Affairs

Author	Sample Size and Diagnoses	Activity Protocol	ActiGraph specifications	Equation (Units)	Reported Accuracy
Nightingale [12]	Spinal Cord Injury (Paraplegia): 10 Spina Bifida: 3 Cerebral Palsy: 1 Amputee: 1 Scoliosis: 1 Able bodied MWU athlete: 1	Desk Work WP at: 2 $km \cdot hr^{-1}$, 4 $km \cdot hr^{-1}$, 6 $km \cdot hr^{-1}$, 8 $km \cdot hr^{-1}$	Wrist: Right Model: GT3X+ f_s: 30 Hz Software: ActiLife 6	Eq. #1 $PAEE = 0.000245 \times VMC + 0.291708$ ($kcal \cdot min^{-1}$)	R = 0.93 Standard Error of Estimation = 3.34 $kJ \cdot hr^{-1}$
Nightingale [13]	Spinal Cord Injury (Paraplegia): 9 Fibromyalgia: 1 Spina Bifida: 2 Complex Regional Pain Syndrome: 1 Able-Bodied Individuals: 2	Resting Folding Clothes WP at 3 $km \cdot hr^{-1}$, 4 $km \cdot hr^{-1}$, 5 $km \cdot hr^{-1}$, 6 $km \cdot hr^{-1}$, 7 $km \cdot hr^{-1}$ WP with additional 8% body mass at 4 $km \cdot hr^{-1}$ WP on a 3% gradient at 4 $km \cdot hr^{-1}$	Wrist: Right Model: GT3X+ f_s: 30 Hz Software: ActiLife	Eq. #2 $PAEE = 0.000929 \times VMC - 0.284818$ ($kJ \cdot min^{-1}$)	MAE = 0.69 ± 0.63 $kcal \cdot min^{-1}$ MAPE = 33% ± 40%
Garcia-Massó [14]	Spinal Cord Injury (T2 – L5): 20	Lying Down; Body Transfers; Moving Items; Mopping; Watching TV; Working on a Computer; Arm-Ergometry exercise; Passive Propulsion Slow Propulsion; Fast Propulsion	Wrist: Dominant Model: GT3X f_s: 30 Hz Software: Not Specified	Eq. #3[a] $O_2 = 4.1355 + 0.0376 \times XC_{50} - 0.0155 \times XC_{90} - 0.0047 \times XC_{NA2} + 0.0062 \times XC_{ND1} + 0.02 \times ZC_{75} - 0.0363 \times ZC_{90} + 0.0161 \times VMC_{75} + 0.0253 \times VMC_{90}$ ($ml \cdot kg^{-1} \cdot min^{-1}$)	Mean Squared Error = 5.16 $ml^2 \cdot kg^{-2} \cdot min^{-2}$ MAE = 1.67 $ml \cdot kg^{-1} \cdot min^{-1}$ Root Mean Square Error = 3.32 $ml \cdot kg^{-1} \cdot min^{-1}$
Tsang [11]	Spinal Cord Injury (Paraplegia): 49 Spinal Cord Injury (Tetraplegia): 18 Spina Bifida: 8 Cerebral Palsy: 2 Amputation: 2 Other: 7 Did not report: 4	WP on a flat tile surface at, slow normal, and fast self-selected speed WP at self-selected normal speed on a track, low pile carpeted surface, sidewalk, up/down a ramp; Wheelchair basketball; TheraBand Exercising; Weight Lifting; Arm Ergometry at self-selected slow, normal, fast speed; Watching TV; Washing dishes; Folding clothes/bedsheets; Cleaning house Reading; Using Computer; Playing Games; Propelling around neighborhood; Stretching, chair aerobic, and strength exercises	Wrist: Dominant Model: GT9X Link f_s: 30 Hz Software: ActiLife (Version 4)	Eq. #4 $EE = -0.006197602220975 + 0.000000088463104 \times BMR2 \times VM + 0.000823693371782 \times BMR1 + 0.000577607827818 \times X$ ($kcal \cdot min^{-1}$)	MAPE = 31% ± 7% MSPE = -9 ± 16% ICC (2,1) = 0.84
Learmonth [15]	Spinal Cord Injury: 10 Spina Bifida: 5 Multiple Sclerosis: 4 Amputation: 2 Cerebral Palsy: 1 Congenital Bone Disorder: 2 Demyelinating disease: 1	Resting WP at 1.5 $miles \cdot hr^{-1}$, 3 $miles \cdot hr^{-1}$, and 4.5 $miles \cdot hr^{-1}$	Wrist: Both Model: GT3X f_s: 30 Hz Software: Not Specified	Eq. #5 – Right Handed $O_2 = 0.0022 \times VMC + 3.13$ ($ml \cdot kg^{-1} \cdot min^{-1}$) Eq. #5 – Left Handed $O_2 = 0.0021 \times VMC + 3.14$ ($ml \cdot kg^{-1} \cdot min^{-1}$)	R = 0.95 ± 0.37 R^2 = 0.90 ± 0.14 R = 0.93 ± 0.44 R^2 = 0.87 ± 0.19

[a]During data analysis, it was noted that Eq. #3 as represented in the original article had abnormal results due to the last coefficient ($0.253 \times VMC_{90}$). The corresponding author was contacted who confirmed that the last coefficient should be ($0.0253 \times VMC_{90}$) as shown in Eq. #3 above.

Table 1: EE Prediction Equations for MWUs

(VA) Central Institutional Review Board. The inclusion criteria are 1) between the ages of 18 and 65; 2) having a SCI at least one-year post injury and medically stable, and 3) using a manual wheelchair as their primary means of mobility for at least 40 hours/week.

Participants in the study were asked to refrain from engaging in moderate or vigorous intensity PA from the previous night. They were also asked to refrain from taking caffeine and eating on the day prior to their testing. Once participants gave consent, they completed a demographics questionnaire. Measurement of height was recorded to the nearest centimeter using a tape measure when participants laid supine, while weight was recorded to the nearest decimal using a wheelchair weight scale (Detecto, Webb City, MO, US). Participants rested in a seated position for almost half an hour before they were asked to rest in a supine position for the measurement of REE for 20 minutes. Participants were instructed not to talk and stay awake during the REE measurement. They then performed a number of activities of daily living (ADLs) and exercise in random order. Activities included: resting in a wheelchair; propulsion at self-selected slow, normal, and fast pace on a flat tile surface; propulsion up/down a 1:60 sloped tile surface; watching TV; working on a computer; playing basketball; sweeping/vacuuming the floor; loading and unloading a dishwasher; weight lifting; TheraBand exercises; arm ergometry at a self-selected slow and fast pace; folding laundry; and being pushed in their wheelchair. Each of these activities was performed for 10 minutes with at least a 3-minute break. Participants were equipped with a COSMED K4b2 portable metabolic cart (COSMED Inc, Rome, Italy) and an ActiGraph GT9X Link on their dominant wrist. The metabolic cart measures oxygen intake (VO_2) and carbon dioxide production (VCO_2), and uses the Weir Equation [16] to predict EE in units of $kcal{\cdot}min^{-1}$ based on VO_2 and VCO_2. The ActiGraph GT9X Link was configured to record raw acceleration signals at 30 Hz. The raw signals as well as activity counts for each axis and VMC in 1-second epoch size were obtained from the ActiGraph ActiLife software (v6.11.9).

1.2.3 Data Analysis

Prior to data analysis, steady-state data during each activity trial, defined as VO_2 and VCO_2 measured by the K4b2 having less than 10% changes within 5 consecutive minutes [17, 18], was extracted. When 5 consecutive steady-state minutes was not available, at least 3 consecutive minutes of data was attempted [19] or data from the activity was discarded [19]. Only steady-state data was used to evaluate the performance of the five sets of EE predictive equations shown in table 1. To obtain reliable REE measurement, the first 5 minutes of data was deleted before analysis [20].

To make sure the outputs from all equations are consistent for comparison, we performed a number of conversions to convert all outputs to EE in $kcal \cdot min^{-1}$. For Eq. #1 [12] and #2 [13], we added the measured REE from resting in a supine position for each participant to their predicted PAEE to obtain the predicted EE in $kcal \cdot min^{-1}$. For Eq. #3 [14] and #5 [15], we used the caloric equivalent based on the respiratory exchange ratio (RER) and the participant weight to convert oxygen consumption in $ml \cdot min^{-1} \cdot kg^{-1}$ to EE in $kcal \cdot min^{-1}$. Eq. #4 [11] predicted EE in $kcal \cdot min^{-1}$, and thus no conversion was required. Standardizing and processing of all equations was done using MATLAB 2018b (MathWorks Inc, Natick, MA, USA).

To examine the validity of the five sets of EE predictive equations, an equivalence test between each predictive equation and the criterion measure was performed based on a confidence interval (CI) method [21]. We first obtained both the mean criterion and estimated EE for each activity across all participants. A regression model was then fitted to the pairs of mean criterion

EE (X-axis) and mean estimated EE (Y-axis) for each activity. If the predictive equation is equivalent to the criterion across all activities, the intercept of this regression should be 0 and the slope should be 1. To make sure the intercept describes an average activity, the X and Y values of the regression were further adjusted by subtracting the overall criterion mean (averaged over all activities) [21]. Research suggested regression-based equivalence regions as ±10% of the criterion mean for the intercept and (0.9, 1.1) for the slope (i.e., ±10% of the slope of 1 that would be expected for equal means on the two measures) [21, 22]. We also tested the equivalence regions of ±15% and ±20% of the criterion mean for the intercept and of the slope of 1, respectively. To claim equivalence across the array of activities tested at α=0.05, two 90% CIs, one for the intercept and one for the slope, should fall inside their respective equivalence regions.

In addition, the mean absolute error (MAE), mean absolute percent error (MAPE), mean signed error (MSE), and mean signed percent error (MSPE) were calculated for each participant by comparing the per minute estimated and criterion EE (kcal·min^{-1}). The ICC using a two-way mixed-effects model with absolute agreement ICC (3,1) was obtained along with the 95% CIs between the estimated and criterion EE across all participants. As proposed by Koo and Li [23], ICC values higher than 0.9 are considered as excellent, between 0.75 and 0.9 as good, between 0.5 and 0.75 as moderate, and lower than 0.5 as poor reliability. Also, the BA plot and analysis was used to compare the per minute estimated EE against the criterion across all activity trials of all participants. It plotted the differences between the estimated and criterion EE against their average for each trial of each participant, and also calculated the mean difference between the per minute estimated and criterion EE (the 'bias'), and 95% limits of agreement (LoA) as the mean difference plus and minus 1.96 times the standard deviation (SD) of the differences [24].

We further looked into how the predictive equations perform with different intensities of PA. We first averaged the VO_2 values ($ml \cdot kg^{-1} \cdot min^{-1}$) from the metabolic cart for each minute and then divided it by the resting metabolic equivalent of 2.7 $ml \cdot kg^{-1} \cdot min^{-1}$ for individuals with SCI [25] to obtain the metabolic equivalent task (MET) for the minute. MET values ≤ 1.5 are classified as sedentary behavior, 1.5-3 METs are classified as light intensity, and ≥ 3 METs are classified as MVPA [25]. The MAE, MAPE, MSE, MSPE, and ICC (3,1) were then calculated for each subject by comparing the per minute estimated and criterion EE for each intensity group. Finally, the same set of measures was calculated for each subject by comparing the per minute estimated and criterion EE for each type of activity. All statistical analysis was performed using IBM SPSS Statistics v. 25 (IBM, Armonk NY, USA).

1.3 Results

A total of 30 participants were recruited and tested in this study. One participant did not have steady-state REE after the first 5 minutes of resting data was removed, and thus was not included in the data analysis. Demographic information for 29 participants is shown in table 2. The total steady-state activity minutes for all participants ranged from 19 to 104 minutes with a mean (SD) of 73 (21) minutes. Based on the criterion VO_2, 21% of the time was sedentary, 44% of time was in light intensity PA, and 35% of time was in MVPA. The performance of the EE predictive equations in terms of MAE, MAPE, MSE, MSPE, and ICC (3,1) between the per minute estimated and criterion EE across all participants and intensity groups can be found in table 3. Same measures for each type of activity across all participants can be found in appendix A. Bland-Altman (BA) plots in figure 2 show the differences between the per minute estimated and criterion EE against their average for each trial of each participant. The range between the 95% lower LoA and upper LoA for Eq. #1-#5 [11-15] are 5.01 kcal·min^{-1}, 4.70 kcal·min^{-1}, 5.37 kcal·min^{-1}, 4.74 kcal·min^{-1}, and 25.09 kcal·min^{-1}, respectively.

In terms of the equivalence testing, the overall criterion mean of all 17 activities was 2.89 kcal·min^{-1} so the intercept equivalence region is (-0.29, 0.29) at 10%, (-0.43, 0.43) at 15%, and (-0.58, 0.58) at 20%. The slope equivalence region is (0.9, 1.1) at 10%, (0.85, 1.15) at 15%, and (0.8, 1.2) at 20%. The results from regression predicting the criterion mean from the estimated mean by each predictive equation for 17 activities is shown in table 4. For all five sets of EE predictive equations, the 90% CI for both the intercept and the slope table 4 was outside their respective equivalence regions at all levels including 10%, 15%, and 20%, and thus none of the equations demonstrated statistical equivalence against the criterion measure.

Variables	Mean (SD) or Number of Participants, % of total
Age (Years)	39.4 (12.8)
Weight (Kg)	82.9 (22.1)
Height (in)	68.6 (4.1)
Gender	
Male	23, 80%
Female	6, 20%
Handedness	
Right	25, 86%
Left	4, 14%
Body Mass Index	
BMI ≤ 25	12, 41%
25 < BMI < 30	11, 38%
30 ≤ BMI	6, 21%
Time using Wheelchair (Years)	8.9 (7.6)
Neurological Level of Lesion	
T1 – T12	25, 86%
L2	2, 7%
Not Reported	2, 7%
Lesion Type	
Complete	21, 72%
Incomplete	6, 21%
Not Reported	2, 7%

Abbreviations: BMI, Body Mass Index

Table 2: Demographic Data

Category	Equations	MAE – kcal·min^{-1}	MAPE – %	MSE – kcal·min^{-1}	MSPE – %	ICC (3,1) [95% CI]
All Activities	Eq. #1 – Nightingale [12]	1.03 (0.32)	37 (15)	0.73	29	0.40 [-0.11, 0.74]
	Eq. #2 – Nightingale [13]	0.87 (0.27)	31 (10)	0.45	17	0.59 [0.28, 0.89]
	Eq. #3 – Garcia-Massó [14]	1.10 (0.67)	44 (25)	0.86	37	0.40 [-0.07, 0.70]
	Eq. #4 – Tsang [11]	0.95 (0.39)	37 (11)	0.54	15	0.28 [-0.07, 0.58]
	Eq. #5 – Learmonth [15]	6.41 (2.42)	206 (66)	6.35	203	0.06 [-0.05. 0.23]
Sedentary (METs ≤ 1.5)	Eq. #1 – Nightingale [12]	0.34 (0.43)	20 (15)	0.31	18	0.64 [0.02, 0.87]
	Eq. #2 – Nightingale [13]	0.33 (0.42)	19 (12)	0.27	14	0.83 [0.65, 0.92]
	Eq. #3 – Garcia-Massó [14]	0.64 (0.57)	45 (30)	0.64	45	0.33 [-0.10, 0.66]
	Eq. #4 – Tsang [11]	0.83 (0.38)	68 (35)	0.83	68	0.16 [-0.05, 0.51]
	Eq. #5 – Learmonth [15]	0.79 (1.14)	48 (54)	0.76	45	0.21 [-0.11, 0.52]
Light (1.5 < METs < 3.0)	Eq. #1 – Nightingale [12]	0.98 (0.37)	44 (21)	0.77	36	0.46 [-0.10, 0.78]
	Eq. #2 – Nightingale [13]	0.79 (0.27)	35 (15)	0.46	22	0.65 [0.16, 0.85]
	Eq. #3 – Garcia-Massó [14]	1.25 (0.87)	51 (27)	1.11	45	0.43 [-0.07, 0.73]
	Eq. #4 – Tsang [11]	0.43 (0.16)	18 (7)	0.28	12	0.73 [0.49, 0.87]
	Eq. #5 – Learmonth [15]	6.15 (3.89)	243 (114)	6.02	236	0.08 [-0.08, 0.31]
MVPA (3.0 ≤ METs)	Eq. #1 – Nightingale [12]	1.28 (0.68)	32 (19)	0.91	24	0.44 [0.08, 0.69]
	Eq. #2 – Nightingale [13]	1.17 (0.55)	29 (29)	0.70	17	0.53 [0.20, 0.75]
	Eq. #3 – Garcia-Massó [14]	1.28 (1.53)	29 (27)	0.94	22	0.46 [0.12, 0.70]
	Eq. #4 – Tsang [11]	1.61 (0.66)	36 (8)	1.61	36	0.14 [-0.06, 0.45]
	Eq. #5 – Learmonth [15]	10.51 (7.31)	244 (122)	10.50	244	0.06 [-0.08, 0.26]

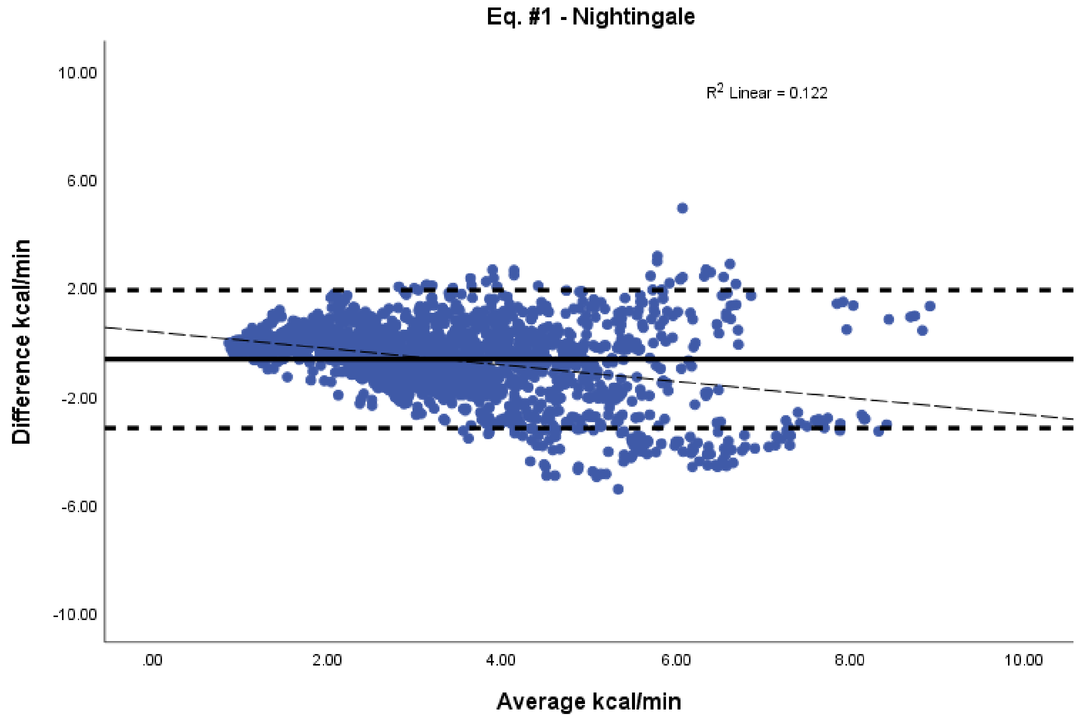
Abbreviations: MAE, Mean Absolute Error, MAPE, Mean Absolute Percent Error, MSE, Mean Signed Error, MSPE, Mean Signed Percent Error, ICC. Intraclass Correlation Coefficient, METs, Metabolic Equivalent of Task

Table 3: Performance of EE Predictive Equations for All, Sedentary, Light, and MVPA Activities

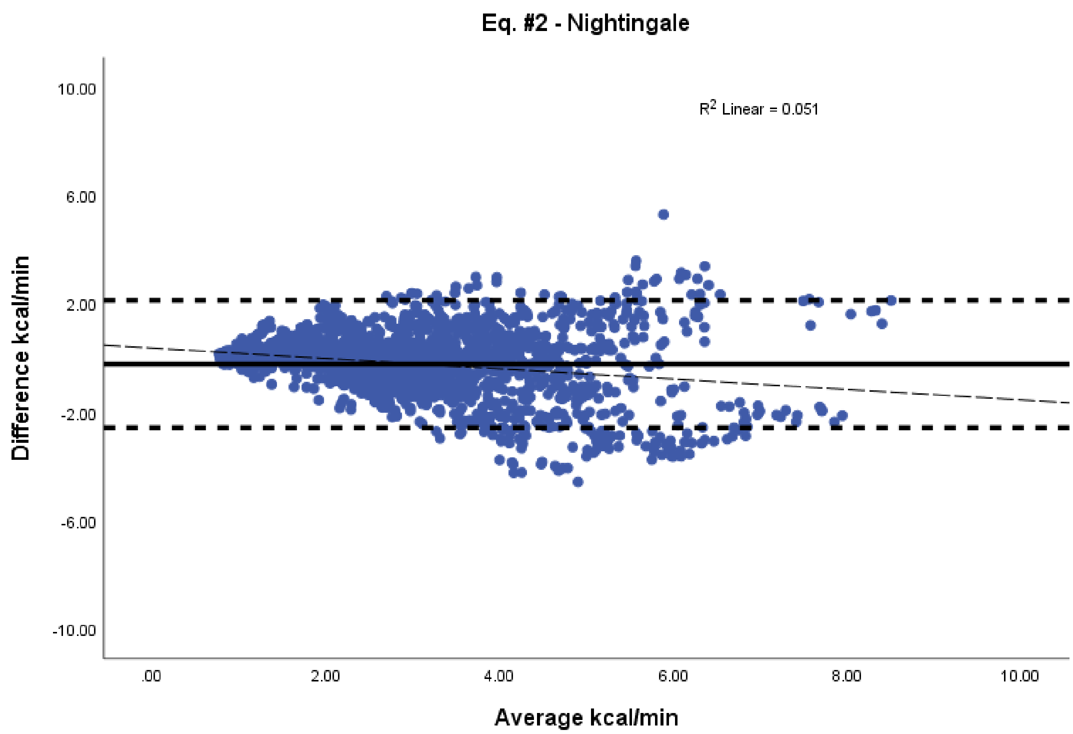
Equations	Intercept			Slope		
	Estimate	SE	90% CI	Estimate	SE	90% CI
Eq. #1 – Nightingale [12]	0.65	0.25	(0.21, 1.08)	1.23	0.24	(0.81, 1.66)
Eq. #2 – Nightingale [13]	0.24	0.22	(-0.16, 0.63)	1.12	0.22	(0.73, 1.50)
Eq. #3 – Garcia-Massó [14]	0.92	0.18	(0.61, 1.24)	1.19	0.17	(0.90, 1.49)
Eq. #4 – Tsang [11]	-0.57	0.03	(-0.62, -0.53)	0.20	0.02	(0.16, 0.25)
Eq. #5 – Learmonth [15]	6.89	0.91	(5.30, 8.47)	4.73	0.83	(3.27, 6.19)

Abbreviations: SE, Standard Error, CI, Confidence Interval

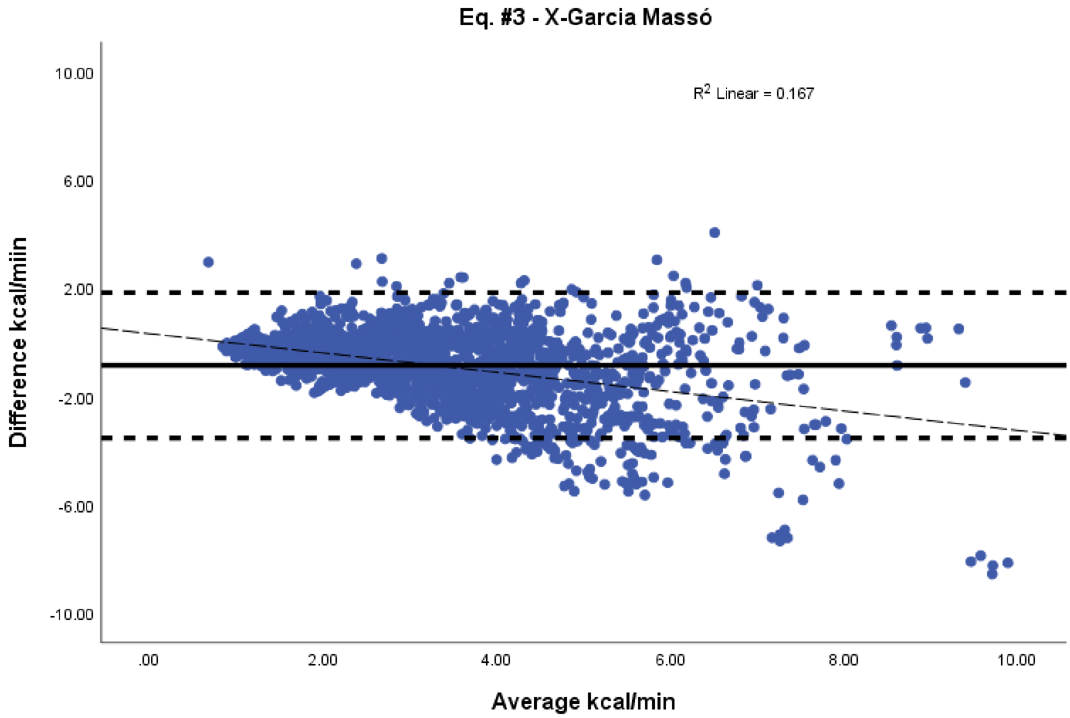
Table 4: Equivalence Testing



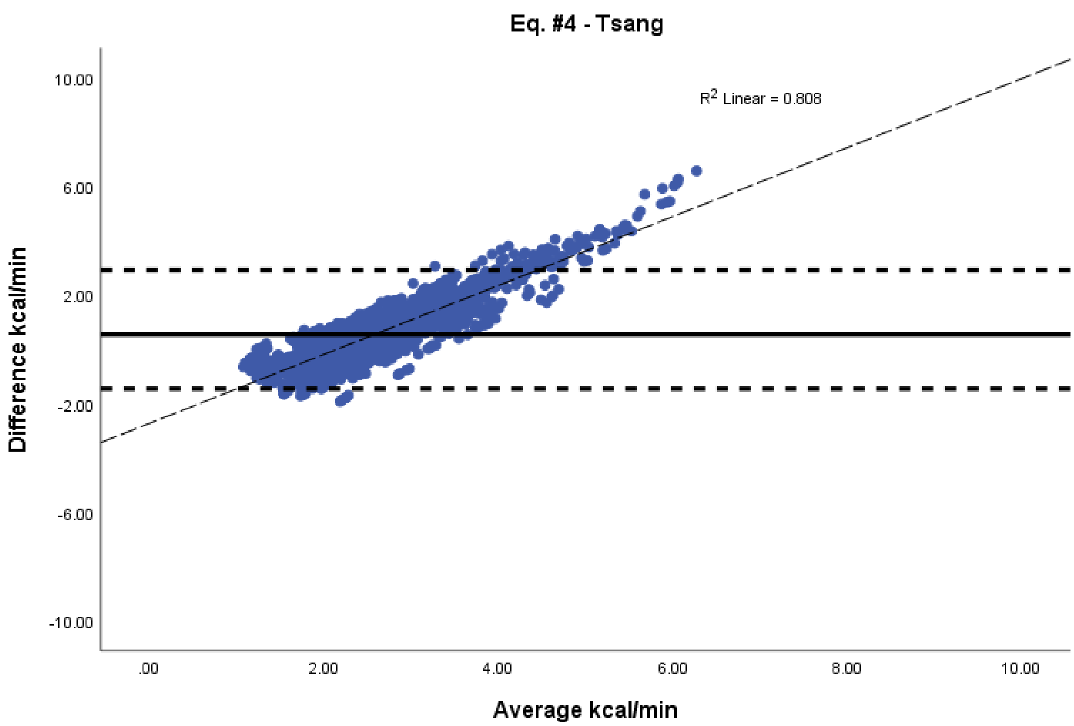
Mean = -0.65 kcal·min^{-1} [-1.11 kcal·min^{-1}, -0.18 kcal·min^{-1}]
Lower LoA = -3.20 kcal·min^{-1} [-4.00 kcal·min^{-1}, -2.39 kcal·min^{-1}]
Upper LoA = 1.90 kcal·min^{-1} [1.09 kcal·min^{-1}, 2.70 kcal·min^{-1}]



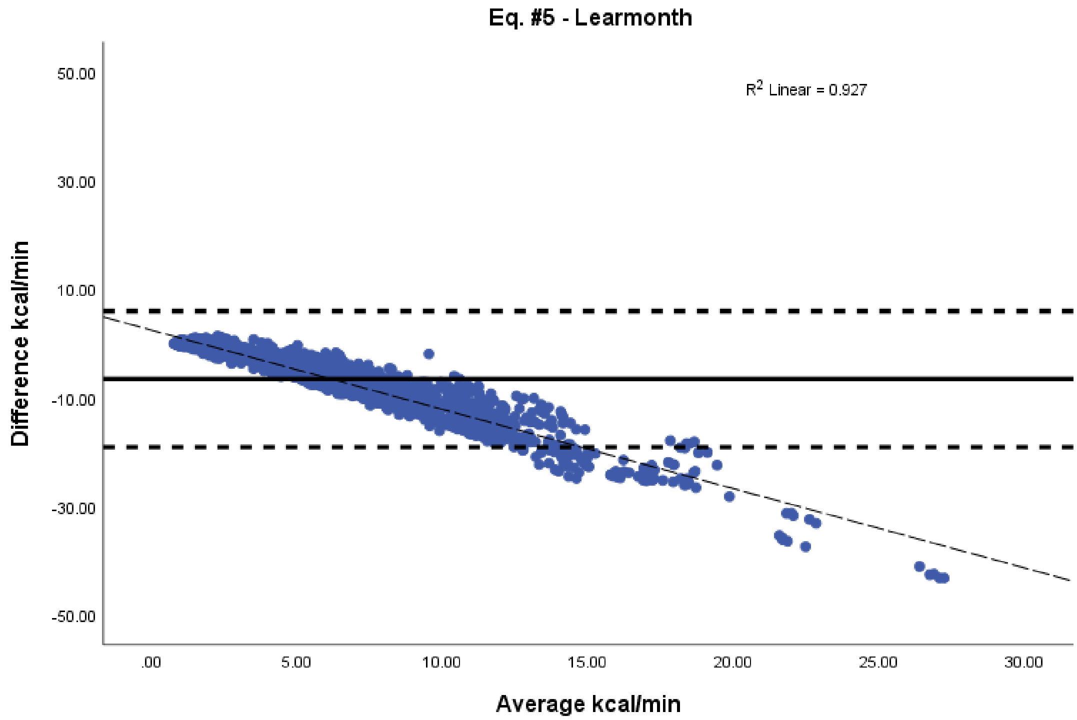
Mean = -0.23 kcal·min^{-1} [-0.66 kcal·min^{-1}, 0.20 kcal·min^{-1}]
Lower LoA = -2.58 kcal·min^{-1} [-3.33 kcal·min^{-1}, -1.84 kcal·min^{-1}]
Upper LoA = 2.12 kcal·min^{-1} [1.38 kcal·min^{-1}, 2.87 kcal·min^{-1}]



Mean = -0.86 kcal·min^{-1} [-1.35 kcal·min^{-1}, -0.37 kcal·min^{-1}]
Lower LoA = -3.55 kcal·min^{-1} [-4.39 kcal·min^{-1}, -2.70 kcal·min^{-1}]
Upper LoA = 1.83 kcal·min^{-1} [0.98 kcal·min^{-1}, 2.67 kcal·min^{-1}]



Mean = 0.55 kcal·min^{-1} [0.12 kcal·min^{-1}, 0.98 kcal·min^{-1}]
Lower LoA = -1.82 kcal·min^{-1} [-2.57 kcal·min^{-1}, -1.07 kcal·min^{-1}]
Upper LoA = 2.92 kcal·min^{-1} [2.17 kcal·min^{-1}, 3.67 kcal·min^{-1}]



Mean = -6.61 kcal·min^{-1} [-8.90 kcal·min^{-1}, -4.32 kcal·min^{-1}]
Lower LoA = -19.15 kcal·min^{-1} [-23.12 kcal·min^{-1}, -15.19 kcal·min^{-1}]
Upper LoA = 5.93 kcal·min^{-1} [1.97 kcal·min^{-1}, 9.90 kcal·min^{-1}]

Figure 2: Bland-Altman plots showing the differences between the criterion and estimated EE using the predictive equations including Nightingale (Eq. #1) [12], Nightingale (Eq. #2) [13], Garcia-Massó (Eq. #3) [14], Tsang (Eq. #4) [11], Learmonth (Eq. #5) [15]

1.4 Discussion

This study examined the performance of five sets of published EE predictive equations for MWUs using an independent dataset from 29 MWUs with SCI. The out-of-sample validation showed that these predictive equations did not demonstrate statistical equivalence against the criterion measure based on 20% equivalence regions. They also had varied performance when compared with the criterion measure. The MAE (MAPE) for the five sets of predictive equations ranged from 0.87-6.41 kcal·min^{-1} (31%-206%) with the ICC estimates ranging from 0.06-0.59. From the BA plots in figure 2, Eq. #1-#3 [12-14] all demonstrated considerable heteroskedasticity (i.e., increasing error as the intensity of activity increases). Eq. #4 [11] showed a tendency (R^2=0.808) to over-predict with higher intensity activities whereas for Eq. #5 [15], the negative correlation (R^2=0.927) implies there was a tendency to under-predict with higher intensity activities.

Though none of the equations demonstrated statistical equivalence against the criterion measure, the regression table 4 based on Nightingale's Eq. #2 [13] yielded a slope of 1.118 (closest to 1) and an intercept of 0.237 (closest to 0). From table 3, this equation also showed the lowest MAE and highest ICC. However, it should be noted that when standardizing this equation along with Nightingale's Eq. #1 [12], we added the measured REE from the metabolic cart to the estimated PAEE to obtain the estimated overall EE. Therefore, it is expected that these equations may yield better accuracies in estimating the overall EE, as the REE needed in such estimations is from direct measurement instead of by the predictive equations. Thus, the accuracies of the Nightingale's Eq. #1 [12] and Eq. #2 [13] equations should be used with caution when the overall EE prediction is of primary interest.

Similar to both Nightingale's equations [12, 13], Learmonth's Eq. #5 [15] also utilized the VMC as their only predictor variable, however, the equations yielded far higher errors and lower ICC values in comparison to the rest of the equations. One difference between the Learmonth's study [15] and others was that the former used a protocol including only three propulsion activities at 1.5 mph, 3.0 mph, and 4.5 mph, respectively, while other studies included a mix of propulsion activities and other ADLs [11-14]. In addition, there could be other unidentified systematic errors that caused the large estimation errors, as evidenced by the BA plots in figure 2 and large deviations from the ideal slope of 1 and intercept of 0 in the equivalence testing table 4.

Tsang's study [11] included the largest and widest array of activity recordings among the five studies with a total of 24 activities, 13 of which came from a lab session and 11 of which came from a home session. It also had the largest sample size among the five studies. As light-intensity PA accounted for a large portion of the activities, Tsang's Eq. #4 [11] showed the lowest MAE (MAPE) of 0.43 $kcal \cdot min^{-1}$ (18%) and highest ICC of 0.73 for light-intensity activities. However, the performance of Tsang's Eq. #4 [11] fell short for sedentary behavior and MVPA as shown in table 3. The lowest MSPE of 15% but a poor ICC value of Tsang's equation [11] for all activities indicate that Tsang's Eq. #4 [11] consistently over and underestimated different activities, causing these errors to weigh each other out. In addition, the BA plots in figure 2 and large deviations from the ideal slope of 1 in the equivalence testing from table 4 also indicate there might be systematic errors in the modeling process. One of the issues could be related to its use of the able-bodied BMR prediction equation, which have previously been shown to demonstrate considerable error when used in individuals with SCI [26].

Garcia-Massó's Eq. #3 [14] is the only one that utilizes statistical methods as well as signal processing techniques in the modeling process. Compared with the other equations, it showed reasonable performance especially considering that it estimates the overall EE directly and thus does not need to use measured REE in the total EE estimation as the Nightingale's Eq. #1 [12] and Eq. #2 [13]. The more complex modeling technique used in Garcia-Massó's study [14] may have contributed to the reasonable performance achieved by Garcia-Massó's Eq. #3 [14] for predicting the overall EE as compared with the other equations.

While these predictive equations seemed to yield moderate to large estimation errors for varying reasons, we want to contrast their performance with what was available among the general ambulatory population. We found some literature that evaluated the validity of off-the-shelf wearable devices including accelerometer-only devices and multi-sensor devices that incorporate heart rate in estimating EE among the general ambulatory population. A recent study [27] assessed the accuracies of several consumer-grade activity monitors such as the Fitbit Surge (Fitbit Inc., San Francisco CA, United States), Jawbone Up3 (Jawbone Inc., San Francisco, CA, United States), and Apple Watch 2 (Apple Inc., Cupertino, CA, United States) in 44 ambulatory participants. The study found that the Jawbone Up3 gave the best performance with a MAPE (SD) of 28% (27%), whereas the Fitbit Surge had the worst performance with a MAPE (SD) of 67% (80%). The other devices including the Apple Watch 2 performed similarly, with an MAPE (SD) of 49% (47%). Another study assessed the laboratory and daily EE estimates from four consumer-grade devices including the three aforementioned devices in [27] and two research-grade devices. While the MAPE values reported for the three consumer devices differed from those provided by [27], the MAPE for both consumer- and research-grade devices ranged from 20%-40% for laboratory assessment and 15%-34% for 24-hour free-living assessment [28]. A 2018 systematic review and meta-analysis of the validity of activity monitors in estimating EE in the general population found large and significant heterogeneity for many devices and concluded that EE estimates from wrist and arm-worn device differ in accuracy depending on activity type [29]. Additionally, a pilot study from 2018 assessed the feasibility of using Fitbit Charge 2 to monitor daily physical activity and hand-bike training in 6 wheelchair users with SCI. The study provided descriptive graphs to show the possibility of using total daily step counts to detect training days and interpersonal/intrapersonal variations in the daily PA level, however, it did not incorporate any criterion measure and provide meaningful statistics [30]. These findings are similar to what was found in this paper, with the predictive algorithms varying in accuracy across different activity types and intensities.

Compared with the EE prediction performance for the general ambulatory population, the existing EE prediction algorithms for MWUs with SCI demonstrated greater inaccuracy with MAE (MAPE) ranging from 0.87-6.41 kcal·min^{-1} (31%-206%). Extrapolating this error over a 24-hour period would lead to an EE estimation error of 1,253-9,230 kcal/day. However, as accuracies of these predictive EE equations vary across different types and intensities of activities, the EE estimation error yielded in the study based on the specific activity protocol may not accurately reflect the daily EE estimation errors in free-living conditions. Nonetheless, future work is needed to develop more accurate EE algorithms for MWUs with SCI. With the growing prevalence of multi-sensor consumer devices, new EE predictive algorithms could consider incorporating physiological signals such as heart rate to help improve EE prediction accuracy for MWUs with SCI, especially for MVPA and with more practical calibration procedures [31]. In addition, EE estimation could use signal features and patterns extracted from high resolution raw acceleration signals with machine-learning techniques to better classify the activity types and derive more sophisticated EE prediction models. Finally, as REE accounts for a large percent of daily EE, more research is needed to improve REE estimation accuracy for this population based on readily available demographic and anthropometric information.

Although this is the first study that offers a comparative evaluation of all predictive ActiGraph EE algorithms in MWUs with SCI, the study has a few limitations. The study attempted to follow a systematic review process, however, the search covered only three databases and limited effort has been made to locate unpublished work. Thus, the study may not include all relevant EE predictive algorithms. In terms of the out-of-sample data collection, all 29 participants in the study had paraplegia. A larger sample size with various levels of diagnosis including tetraplegia could further improve our understanding of predictive equation performance. This study only equipped participants with one ActiGraph monitor on their dominant wrist. Garcia-Massó's study [14] also developed a non-dominant wrist equation, which could not be evaluated in this study. Finally, although the activity protocol in our study included a relatively large range of typical daily activities, they were not performed in natural settings. The proportion of different types of activities in the protocol does not necessarily reflect the typical activity profile in everyday living. As the estimation accuracy of the EE predictive equations may be dependent on the types of activities, the results of this evaluation should be interpreted with caution.

1.5 Conclusion

EE estimates from all five sets of predictive equations based on ActiGraph monitors for MWUs with SCI failed to fall into the ±10%, ±15%, and ±20% equivalence regions set by the criterion. These equations yielded a MAE of 0.87-6.41 kcal·min^{-1} and a MAPE of 31%-206%. Future work is needed to develop more accurate EE predictive algorithms for MWUs with SCI.

2.0 Predicting Physical Activity Intensity Using Raw Accelerometer Signals In Manual Wheelchair Users With Spinal Cord Injury

2.1 Introduction

Life Expectancy of individuals with SCI has increased with 43% of those who experience their injuries from ages 25 to 34 living for another 40 years [32]. Yet, SCIs result in lowered levels of mobility, primarily requiring these individuals to rely on a wheelchair for daily mobility [2]. Despite the use of a manual wheelchair for PA, this population of MWUs has lesser accessibility and fewer opportunities to engage in PA in comparison to the general population [33]. These issues lead to an increased prevalence of many chronic diseases associated with physical inactivity, including cardiovascular disease, diabetes, cancer, hypertension, obesity, depression, and osteoporosis [4].

Activity monitors have been widely utilized to track and promote changes in PA, with ActiGraph activity monitors being the most extensively studied devices used to track PA in research settings [34]. The ActiGraph devices are capable of collecting raw acceleration signals at a set frequency, as well as producing a proprietary variable called 'count' for each accelerometer axis. These 'count' recordings from all 3 axes are used to obtain a vector magnitude count, being VMC, which is often utilized in predictive algorithms for PA intensity and EE in many research manuscripts. For example, several research groups including Learmonth [15], McCracken [35], Veerubhotla [36] and Holmlund [37] developed VMC-based thresholds for classifying different PA intensities performed by MWUs. Yet, as VMC can only be obtained through ActiGraph devices and its associated software license that costs $1700 [38], the applicability of these algorithms is restricted. Meanwhile, there are many other commercial wearables devices that are just as capable of recording raw acceleration signals but are offered at an affordable price without requiring specialized software. Therefore, it is important to develop predictive algorithms for PA intensity based on raw accelerometer signals instead of proprietary 'counts'.

The purpose of this study is to develop a PA intensity classification model based on raw accelerometer signals for MWUs with SCI. A study protocol that encapsulates a series of ADLs of varying intensities was used to develop and cross-validate the model.

2.2 Methods

2.2.1 Study Logistics

This study was conducted at two sites including HERL in Pittsburgh, PA and the James J. Peters VA Medical Center in Bronx, NY. Approval for this study was granted by the US Department of Veterans Affairs' Central Institutional Review Board. The inclusion criteria are 1) between the ages of 18 and 65; 2) having an SCI at least one-year post injury and medically stable, and 3) using a manual wheelchair as their primary means of mobility for at least 40 hours/week.

2.2.2 Study Protocol

Participants were asked to avoid taking part in any MVPA the night before testing, along with ingesting any caffeine or food on the day of testing. Individuals first gave informed consent and completed a demographics questionnaire. Individuals were then instructed to lay in supine position while their height was measured using a tape measure. Weight was measured while individuals were in their wheelchair, on a wheelchair weight scale. This weight was then subtracted by the weight of the wheelchair alone. For the activity protocol, individuals were asked to first rest in a seated position for 30 minutes, and then rest in supine position for 20 minutes. This was then followed by a randomly selected array of ADLs including: resting in a wheelchair; propulsion at self-selected slow, normal, and fast pace on flat tiled surface; propulsion up/down a 1:60 sloped tile surface, watching TV; working on a computer; playing basketball; sweeping/vacuuming the floor; loading and unloading a dishwasher; weight lifting; TheraBand exercises; arm ergometry at a self-selected slow and fast pace; folding laundry; and being pushed in their wheelchair. Each activity was performed for 10 minutes with a minimum break of 3 minutes between each activity.

2.2.3 Instrumentation

Individuals were equipped with a COSMED $K4b^2$ portable metabolic cart, that measures VO_2 intake and VCO_2 output. Individuals were also equipped with an ActiGraph GT9X Link on the dominant wrist, recording raw acceleration data at 30 Hz. Raw-signal data was obtained from the ActiGraph ActiLife software (v6.11.9).

2.2.4 Data Preprocessing

All data that did not constitute as activity data was removed. If either data for the $K4b^2$ or the ActiGraph was not available for a minute due to device malfunctioning, data from both devices was removed. Only steady-state data for each activity trial was retained in the final dataset. Steady-state is defined as VO_2 and VCO_2 measured by the $K4b^2$ having changed less than 10% for 5 continuous minutes [18]. If this wasn't available for an activity, a minimum of 3 minutes was attempted [19], or the data was removed [19].

All data was organized into different PA intensity categories, being the metabolic equivalent of task (METs), defined as the average VO_2, in units of ml kg^{-1} min^{-1}, divided by 2.7 ml kg^{-1} min^{-1} [25]. This served as the criterion for PA intensity, with values below 1.5 as resting, those in-between 1.5 and 3.0 as light-intensity, and those above 3.0 as MVPA.

2.2.5 Data Modeling & Validation

2.2.5.1 Feature Development and Selection

A total of 130 features were extracted in 1-min non-overlapping windows based on raw accelerometer signals. Sixty-three features were derived directly from raw signals, and sixty-seven features were based on activity count derived from raw signals based on an algorithm by Brønd et al. [39]. The algorithm first removes noise and artifacts from raw-signals using anti-aliasing and a band-pass filter. The signals are then truncated with 2.13g constant amplitude and rectified. Lastly, the signals are applied with a dead-band threshold of 0.068g, converted to 8-bit resolution and integrated into 1-second epoch.

A variety of statistical and time-domain signal measures were utilized as the basis to generate features based on raw signals, and activity counts converted from raw signals for all three axes and the vector magnitude. These features include mean, standard deviation, minimum, maximum, summation, 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles, interquartile range, correlation between each axis, coefficient of variation, skewness, kurtosis, and signal power. Weka (Machine Learning Group, University of Waikato), an open-source machine learning software was used for feature selection. A correlation-based feature selection method along with a greedy stepwise search through the feature set was used to find a feature subset that contains features highly correlated with the activity intensity, yet uncorrelated with each other. Based on a 10-fold cross validation, the features that were selected by at least 7 of the 10 folds were selected to build the predictive model.

2.2.5.2 Model Development

A RFM was developed to classify sedentary activity, light-intensity PA, and MVPA using a feature subset selected from the aforementioned feature selection step. A 10-fold cross validation process was used to evaluate the model performance, where 90% of the data was used to develop the model and the remaining 10% of data was used to evaluate the model in each fold. The data was stratified based on intensity levels to ensure the same intensity distribution in the model development and evaluation datasets. The cross-validation was also used to tune the number of trees used in the RFM (starting at 10 trees with a 10-tree interval until 50 trees). The number of trees that yielded the best accuracy was used by the final model. Given the imbalance time spent in sedentary behavior, light-intensity PA, and MVPA (shown by the confusion matrix in the results section), the accuracy measures including recall, precision and area under the precision-recall curve (PRC) for each intensity were obtained. A confusion matrix summing the values of all 10 models on the 10 validation datasets was obtained along with an overall accuracy and kappa statistic.

2.3 Results

A total of 32 participants were recruited and tested in this study. There was a total of 2,152 steady-state minutes of activity data. Of this data, 478 minutes (22%) were classified as sedentary behavior, 888 minutes (41%) being light-intensity PA, and 786 minutes (37%) being MVPA based on the criterion measure. Across these participants, the total steady-state activity minutes obtained ranged from 13 to 94 minutes with an average of 67 ± 20 minutes from each participant. Additional demographic information can be found in Table 5.

Variables	Mean (SD) or Number of Participants, % of total
Age (Years)	39.7 (12.6)
Weight (Kg)	83.0 (21.2)
Height (in)	68.6 (3.9)
Gender	
Male	26, 81%
Female	6, 19%
Handedness	
Right	28, 88%
Left	4, 12%
Time using Wheelchair (Years)	9.0 (7.8)
Neurological Level of Lesion	
Cervical	2, 6%
Thoracic	26, 82%
Lumbar	2, 6%
Not Reported	2, 6%
Lesion Type	
Complete	18, 72%
Incomplete	8 21%
Not Reported	6, 7%

Table 5: Demographic Data

A subset of twenty-one features was selected which includes fifteen features based on raw-signals (i.e., the mean, minimum, sum, 10th, 50th, 90th percentile, interquartile range, and coefficient of variation of VM, the maximum, 25th percentile, and signal power of x-axis, the minimum and 25th percentile of y-axis, the 50th percentile of z-axis, and correlation between y-axis and z-axis), and six features based on activity counts (i.e., the 10th percentile of VMC, the 50th percentile of x-axis counts, the 50th percentile of z-axis counts, and the interquartile range of x-axis, y-axis, and z-axis counts).

Based on the 10-fold cross validation, the RFM with 20 trees yielded the best average accuracy of 81.3% with the accuracy of each fold ranging from 78.6% to 87.4%. Table 6 presents the precision, recall, and area under PRC for each PA intensity. Table 7 shows the confusion matrix detailing the correctly and incorrectly classified PA intensity minutes across all ten folds, with a kappa statistic of 0.71.

	Precision	**Recall**	**Area under PRC**
Sedentary	0.82	0.84	0.87
Light	0.77	0.79	0.83
MVPA	0.87	0.82	0.92

Table 6: Overall model performance

	Estimated Sedentary	**Estimated Light**	**Estimated MVPA**
Criterion Sedentary	400	76	2
Criterion Light	86	705	97
Criterion MVPA	5	136	645

Table 7: Confusion matrix showing minutes spent at each intensity based on criterion and model estimation

2.4 Discussion

In this study, we built an activity intensity classification model based on raw acceleration signals from a wrist-worn wearable device for MWUs with SCI. As no proprietary information is used in the classification model, the model could be potentially used by any wrist-worn device that records raw acceleration signals.

We used a machine learning approach (i.e., RFM) to build a classification model based on a number of measures that are computationally easy to calculate. We chose the RFM, as it is a powerful ensemble learning algorithm that combines multiple decision trees, trains each one on a slightly different set of the training data, and splits nodes in each tree considering a subset of the features. The aggregation of many decision trees in this manner helps limit overfitting as well as error due to bias. The study results showed the RFM yielded a reasonable performance in classifying sedentary behavior (with a precision of 0.82 and a recall of 0.84) and MVPA (with a precision of 0.87 and recall of 0.82), indicating that the predicted sedentary and MVPA instances are likely true and most sedentary and MVPA minutes could be correctly detected by the RFM. The area under PRC for sedentary behavior (0.87) and MVPA (0.92) were also considered excellent or outstanding [40]. However, the RFM lacked the ability to classify light-intensity PA with a precision of 0.77 and a recall of 0.79. From the confusion matrix, it can also be seen that light-intensity PA could be wrongly classified into either sedentary or MVPA category. We noticed that some light-intensity activity such as sweeping or folding laundry that involves consistent and large ranges of upper limb movements, may give higher raw acceleration values, leading to wrong classification into MVPA. While other light-intensity activities such as weight-lifting for some individuals was light-intensity based on criterion METs, but were wrongly classified into sedentary category due to the infrequent upper limb movements. From the confusion matrix, there are also a portion of MVPA minutes that were wrongly classified into light-intensity. A trend we observed was that the model tended to incorrectly predict resistance-based activities. The lack of changes in raw acceleration signals during these activities for some participants may have resulted in features recording values indicative of light-intensity PA during MVPA, leading to an increase in misclassification. Nonetheless, based on the kappa statistic, the RFM was able to improve the classification performance from the expected accuracy of 35% (random chance based on the criterion activity intensity composition) to the observed accuracy of 81.3%.

There were a few studies mentioned in the introduction section that had defined ActiGraph count cut-points for different PA intensities [15,35 – 37]. There is no doubt that using cut-off points is the simplest method to classify PA intensities. However, as these cut-off points were based on proprietary counts from ActiGraph devices, they cannot be used by other wearable devices. Moreover, these studies reported a wide range of cut-points for MVPA, i.e., 3,644 counts min^{-1} by Learmonth et al. [15], 11,652 counts min^{-1} by McCracken et al. [35], 12,467 counts min^{-1} by Veerubhotla et al. [36], and 9,515 count min^{-1} for motor-complete paraplegics and 4,887 count min^{-1} for motor-complete tetraplegics by Holmlund et al. [37]. The large discrepancy in MVPA cut-points could be due to the diversity within wheelchair users. For example, Learmonth et al. tested 24 wheelchair users with a wide array of neurological disorders, which may adversely affect the generation of an accurate MVPA threshold [15]. McCracken et al. showed that the individual MVPA cut-off points ranged from 7,395 – 15,909 counts min^{-1}, indicating that a generalized threshold of 11,652 counts min^{-1} may not be effective for everyone [35]. Holmlund et al. further showed that the MVPA cut-points varied greatly between people with motor-complete paraplegics and tetraplegics. It is worth noting that the activity count used in these studies is an aggregate measure that summarizes 3-axis accelerometer signals sampled at 30Hz into the vector magnitude activity counts in one-minute epochs. In contrast, our approach relied on raw signals at 30Hz and derived multiple features of individual axis as well as vector magnitude signals to better capture movement patterns within each minute (e.g., variability, speed, jerkiness, and regularity). These features could potentially help detect the differences in movement patterns among people with different injury levels and completeness. Finally, most existing work derived the cut-off points based on a limited variety of activities. For example, Learmonth et al. included only wheelchair propulsion trials at different intensities (over a treadmill) in the testing protocol. Holmlunds et al. included six types of activities and specifically mentioned that some MVPA such as weight training and arm crank were not included in the data analysis. Our study developed the RFM based on a wide range of activities including resistance-based activities such as TheraBand exercises, weight lifting, and arm crank, and the cross-validation results reflected the model performance across the wide range of PA. Thus, the model presented in this study may be better for predicting PA intensity in real world settings.

There are a few limitations in the study. First, although the study used the cross-validation approach, the study lacked a separate testing dataset to validate the final classification model. Second, the number of subjects is relatively small and the activity minutes for each participant varied to a large degree. Therefore, we split the datasets into training and testing datasets based on minutes instead of participants during cross-validation. This approach has also prevented us from incorporating demographic variables such as injury level and completeness into the model for addressing the diverse movement patterns in wheelchair users with SCI. Future work will involve gathering more data from more participants to allow for developing machine learning based classification models at the participant level. Furthermore, different modeling techniques and additional features could be attempted for different applications. For example, classification models that will be used during post processing for research/clinical applications could use more complex models and features such as frequency-domain features for best performance, while classification models that aim to provide real-time feedback could adopt simpler models as well as a small subset of computationally basic features.

2.5 Conclusion

A classification model to predict time in sedentary, light-intensity PA, and MVPA for MWUs with SCI was developed based on raw accelerometer signals and assessed using 10-fold cross-validation. Results from this study show that the model can potentially be used to predict sedentary and MVPA with moderate accuracy, however it should be used with caution when trying to measure time in light-intensity PA.

Appendix A

Activity	Equations	MAE – kcal·min^{-1}	MAPE – %	MSE – kcal·min^{-1}	MSPE – %	ICC (3,1) [95% CI]
Resting	**Eq. #1 – Nightingale [12]**	-	-	-	-	-
	Eq. #2 – Nightingale [13]	-	-	-	-	-
	Eq. #3 – Garcia-Massó [14]	0.57 (0.37)	52 (34)	0.57	52	0.21 [-0.10, 0.53]
	Eq. #4 – Tsang [11]	1.01 (0.41)	95 (49)	1.00	95	0.05 [-0.05, 0.23]
	Eq. #5 – Learmonth [15]	0.37 (0.36)	32 (29)	0.36	31	0.36 [0.00, 0.64]
Propelling at Self-Pace	**Eq. #1 – Nightingale [12]**	0.64 (0.60)	23 (26)	0.61	22	0.85 [0.59, 0.94]
	Eq. #2 – Nightingale [13]	0.48 (0.50)	16 (20)	0.46	15	0.89 [0.78, 0.95]
	Eq. #3 – Garcia-Massó [14]	0.91 (1.01)	31 (34)	0.89	31	0.74 [0.30, 0.89]
	Eq. #4 – Tsang [11]	0.76 (0.76)	21 (17)	0.75	21	0.67 [0.35, 0.84]
	Eq. #5 – Learmonth [15]	6.40 (3.82)	199 (109)	6.40	199	0.20 [-0.09, 0.53]
Propelling at Slow-Pace	**Eq. #1 – Nightingale [12]**	0.07 (0.22)	3 (11)	0.06	3	0.90 [0.31, 0.99]
	Eq. #2 – Nightingale [13]	0.07 (0.19)	3 (9)	0.06	3	0.92 [0.21, 0.99]
	Eq. #3 – Garcia-Massó [14]	0.07 (0.21)	2 (8)	0.06	2	0.91 [0.15, 0.99]
	Eq. #4 – Tsang [11]	0.14 (0.46)	4 (12)	0.15	4	0.12 [-0.86, 0.90]
	Eq. #5 – Learmonth [15]	0.75 (2.20)	23 (62)	0.75	23	0.25 [-0.14, 0.88]
Propelling at Fast-Pace	**Eq. #1 – Nightingale [12]**	0.78 (0.65)	19 (20)	0.70	21	0.82 [0.65, 0.91]
	Eq. #2 – Nightingale [13]	0.79 (0.68)	17 (16)	0.73	16	0.80 [0.60, 0.90]
	Eq. #3 – Garcia-Massó [14]	0.97 (0.94)	23 (25)	0.91	25	0.75 [0.51, 0.88]
	Eq. #4 – Tsang [11]	1.84 (1.32)	35 (20)	1.81	20	0.29 [-0.11, 0.64]
	Eq. #5 – Learmonth [15]	9.10 (4.62)	199 (110)	9.07	110	0.13 [-0.06, 0.42]
Propelling on Ramp	**Eq. #1 – Nightingale [12]**	0.80 (0.67)	16 (12)	0.75	15	0.79 [0.20, 0.93]
	Eq. #2 – Nightingale [13]	1.09 (0.80)	23 (14)	1.07	22	0.66 [-0.07, 0.89]
	Eq. #3 – Garcia-Massó [14]	0.64 (0.51)	14 (11)	0.55	12	0.89 [0.74, 0.95]
	Eq. #4 – Tsang [11]	1.60 (1.23)	33 (19)	1.60	33	0.33 [-0.11, 0.67]
	Eq. #5 – Learmonth [15]	4.75 (3.39)	104 (63)	4.75	104	0.29 [-0.10, 0.64]
Pushed by Investigator	**Eq. #1 – Nightingale [12]**	0.21 (0.27)	12 (14)	0.18	14	0.91 [0.82, 0.96]
	Eq. #2 – Nightingale [13]	0.28 (0.32)	16 (16)	0.27	15	0.84 [0.50, 0.94]
	Eq. #3 – Garcia-Massó [14]	0.41 (0.40)	27 (26)	0.37	25	0.80 [0.61, 0.90]
	Eq. #4 – Tsang [11]	0.38 (0.37)	28 (31)	0.36	27	0.83 [0.45, 0.94]
	Eq. #5 – Learmonth [15]	0.44 (0.43)	27 (25)	0.40	25	0.82 [0.63, 0.92]
Working on Computer	**Eq. #1 – Nightingale [12]**	0.18 (0.16)	13 (11)	0.16	12	0.84 [0.69, 0.93]
	Eq. #2 – Nightingale [13]	0.24 (0.20)	16 (11)	0.22	15	0.75 [0.29, 0.90]
	Eq. #3 – Garcia-Massó [14]	0.43 (0.40)	34 (36)	0.42	33	0.43 [0.06, 0.69]
	Eq. #4 – Tsang [11]	0.53 (0.34)	43 (32)	0.53	43	0.40 [-0.10, 0.73]
	Eq. #5 – Learmonth [15]	0.45 (0.42)	33 (30)	0.40	29	0.47 [0.14, 0.72]
Watching TV	**Eq. #1 – Nightingale [12]**	0.22 (0.18)	16 (14)	0.20	15	0.88 [0.75, 0.94]
	Eq. #2 – Nightingale [13]	0.23 (0.18)	16 (11)	0.20	15	0.86 [0.71, 0.93]
	Eq. #3 – Garcia-Massó [14]	0.46 (0.42)	36 (35)	0.42	34	0.61 [0.10, 0.83]
	Eq. #4 – Tsang [11]	0.56 (0.36)	46 (33)	0.54	45	0.57 [-0.09, 0.85]
	Eq. #5 – Learmonth [15]	0.54 (0.55)	39 (34)	0.48	35	0.59 [0.19, 0.81]

(Continued)

Activity	Equations	MAE – kcal·min^{-1}	MAPE – %	MSE – kcal·min^{-1}	MSPE – %	ICC (3,1) [95% CI]
Vacuuming	**Eq. #1 – Nightingale [12]**	0.75 (0.56)	37 (38)	0.72	36	0.54 [0.05, 0.79]
	Eq. #2 – Nightingale [13]	0.60 (0.43)	28 (29)	0.57	27	0.63 [0.33, 0.82]
	Eq. #3 – Garcia-Massó [14]	0.96 (0.84)	44 (40)	0.94	43	0.47 [-0.09, 0.78]
	Eq. #4 – Tsang [11]	0.48 (0.41)	21 (26)	0.48	21	0.60 [0.27, 0.80]
	Eq. #5 – Learmonth [15]	4.95 (3.19)	208 (132)	4.95	208	0.09 [-0.06, 0.34]
Folding Laundry	**Eq. #1 – Nightingale [12]**	0.99 (0.51)	44 (25)	0.99	44	0.59 [-0.08, 0.87]
	Eq. #2 – Nightingale [13]	0.65 (0.40)	29 (20)	0.65	29	0.74 [-0.05, 0.92]
	Eq. #3 – Garcia-Massó [14]	1.10 (0.89)	47 (39)	1.09	47	0.49 [-0.07, 0.79]
	Eq. #4 – Tsang [11]	0.43 (0.45)	16 (12)	0.42	16	0.70 [0.42, 0.85]
	Eq. #5 – Learmonth [15]	6.57 (3.41)	272 (129)	6.56	272	0.10 [-0.07, 0.35]
Eating a Meal	**Eq. #1 – Nightingale [12]**	0.59 (0.35)	26 (18)	0.53	24	0.79 [0.27, 0.92]
	Eq. #2 – Nightingale [13]	0.42 (0.26)	18 (12)	0.36	16	0.87 [0.75, 0.94]
	Eq. #3 – Garcia-Massó [14]	0.90 (0.80)	39 (33)	0.83	36	0.58 [0.10, 0.81]
	Eq. #4 – Tsang [11]	0.33 (0.26)	14 (10)	0.31	13	0.87 [0.74, 0.94]
	Eq. #5 – Learmonth [15]	4.55 (3.22)	191 (121)	4.55	191	0.14 [-0.09, 0.41]
Arm Ergometer at Self-Selected Pace	**Eq. #1 – Nightingale [12]**	1.83 (1.40)	55 (45)	1.83	55	0.63 [-0.06, 0.87]
	Eq. #2 – Nightingale [13]	1.36 (1.14)	41 (37)	1.36	41	0.72 [0.02, 0.91]
	Eq. #3 – Garcia-Massó [14]	1.71 (1.60)	54 (56)	1.70	54	0.59 [0.00, 0.84]
	Eq. #4 – Tsang [11]	0.84 (0.80)	22 (18)	0.84	22	0.71 [0.07, 0.89]
	Eq. #5 – Learmonth [15]	11.43 (7.84)	339 (239)	11.43	339	0.12 [-0.09, 0.40]
Arm Ergometer at Slow-Pace	**Eq. #1 – Nightingale [12]**	0.91 (1.03)	34 (41)	0.91	33	0.75 [0.19, 0.91]
	Eq. #2 – Nightingale [13]	0.65 (0.82)	24 (32)	0.64	24	0.82 [0.49, 0.93]
	Eq. #3 – Garcia-Massó [14]	1.14 (1.19)	41 (46)	1.14	41	0.69 [0.09, 0.89]
	Eq. #4 – Tsang [11]	0.56 (0.64)	17 (17)	0.56	17	0.80 [0.38, 0.92]
	Eq. #5 – Learmonth [15]	6.80 (5.63)	234 (203)	6.81	234	0.18 [-0.10, 0.49]
Arm Ergometer at Fast-Pace	**Eq. #1 – Nightingale [12]**	2.39 (1.70)	63 (55)	2.39	63	0.56 [-0.10, 0.85]
	Eq. #2 – Nightingale [13]	1.81 (1.39)	48 (46)	1.81	48	0.65 [-0.07, 0.87]
	Eq. #3 – Garcia-Massó [14]	2.05 (2.15)	53 (59)	2.04	53	0.54 [-0.12, 0.81]
	Eq. #4 – Tsang [11]	1.26 (1.03)	28 (19)	1.26	28	0.57 [-0.07, 0.84]
	Eq. #5 – Learmonth [15]	15.30 (11.65)	378 (293)	15.30	378	0.09 [-0.08, 0.34]
Basketball	**Eq. #1 – Nightingale [12]**	0.95 (0.92)	27 (30)	0.92	27	0.56 [-0.10, 0.85]
	Eq. #2 – Nightingale [13]	0.78 (0.77)	21 (22)	0.77	20	0.65 [-0.07, 0.89]
	Eq. #3 – Garcia-Massó [14]	1.15 (1.20)	32 (35)	1.07	30	0.54 [-0.01, 0.81]
	Eq. #4 – Tsang [11]	1.10 (1.28)	23 (20)	1.09	23	0.57 [-0.07, 0.84]
	Eq. #5 – Learmonth [15]	8.59 (6.90)	220 (177)	8.59	220	0.09 [-0.08, 0.34]
TheraBand Exercises	**Eq. #1 – Nightingale [12]**	0.67 (0.56)	25 (22)	0.57	21	0.85 [0.69, 0.93]
	Eq. #2 – Nightingale [13]	0.59 (0.49)	21 (17)	0.48	17	0.87 [0.73, 0.94]
	Eq. #3 – Garcia-Massó [14]	0.95 (0.90)	35 (35)	0.72	28	0.75 [0.44, 0.89]
	Eq. #4 – Tsang [11]	0.54 (0.60)	17 (15)	0.54	17	0.78 [0.30, 0.92]
	Eq. #5 – Learmonth [15]	4.28 (3.26)	157 (127)	4.24	156	0.24 [-0.10, 0.57]
Weight Lifting	**Eq. #1 – Nightingale [12]**	0.80 (0.62)	30 (27)	0.70	26	0.83 [0.66, 0.92]
	Eq. #2 – Nightingale [13]	0.82 (0.59)	29 (22)	0.71	24	0.81 [0.63, 0.90]
	Eq. #3 – Garcia-Massó [14]	1.00 (0.76)	38 (34)	0.81	31	0.83 [0.57, 0.93]
	Eq. #4 – Tsang [11]	0.64 (0.75)	19 (17)	0.63	18	0.75 [0.43, 0.89]
	Eq. #5 – Learmonth [15]	3.77 (2.88)	136 (114)	3.71	134	0.33 [-0.10, 0.66]

Bibliography

1. Spinal cord injury Key facts WHO, 2013

2. Shields RK, Dudley-Javoroski S. Monitoring standing wheelchair use after spinal cord injury: a case report. Disabil Rehabil. 2005;27(3):142-6.

3. Buchholz AC, McGillivray CF, Pencharz PB. Physical activity levels are low in free-living adults with chronic paraplegia. Obes Res. 2003;11(4):563-70.

4. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ. 2006;174(6):801-9.

5. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med. 2015;13:77-.

6. Hiremath SV, Amiri AM, Thapa-Chhetry B, Snethen G, Schmidt-Read M, Ramos-Lamboy M, et al. Mobile health-based physical activity intervention for individuals with spinal cord injury in the community: A pilot study. PloS one. 2019;14(10):e0223762.

7. Scott J, Grierson A, Gehue L, Kallestad H, Macmillan I, Hickie I. Can consumer grade activity devices replace research grade actiwatches in youth mental health settings? Sleep and Biol Rhythms. 2019;17.

8. Poehlman ET. A review: exercise and its influence on resting energy metabolism in man. Med Sci Sports Exerc. 1989;21(5):515-25.

9. Buchholz AC, Pencharz PB. Energy expenditure in chronic spinal cord injury. Curr Opin Clin Nutr Metab Care. 2004;7(6):635-9.

10. Nightingale TE, Rouse PC, Thompson D, Bilzon JLJ. Measurement of Physical Activity and Energy Expenditure in Wheelchair Users: Methods, Considerations and Future Directions. Sports Med Open. 2017;3(1):10.

11. Tsang K. Using Wearable Sensors for Physical Activity Measurement and Promotion in Manual Wheelchair Users [Dissertation]. Pittsburgh (PA): University of Pittsburgh; 2018.

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

13. Nightingale TE, Walhin JP, Thompson D, Bilzon JL. Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users. PLoS One. 2015;10(5):e0126086.

14. Garcia-Masso X, Serra-Ano P, Garcia-Raffi LM, Sanchez-Perez EA, Lopez-Pascual J, Gonzalez LM. Validation of the use of Actigraph GT3X accelerometers to estimate energy expenditure in full time manual wheelchair users with spinal cord injury. Spinal Cord. 2013;51(12):898-903.

15. Learmonth YC, Kinnett-Hopkins D, Rice IM, Dysterheft JL, Motl RW. Accelerometer output and its association with energy expenditure during manual wheelchair propulsion. Spinal Cord. 2016;54(2):110-4.

16. Weir JBDB. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol. 1949;109(1-2):1-9.

17. McClave SA, Spain DA, Skolnick JL, Lowen CC, Kieber MJ, Wickerham PS, et al. Achievement of steady state optimizes results when performing indirect calorimetry. JPEN J Parenter Enteral Nutr. 2003;27(1):16-20.

18. Schlein KM, Coulter SP. Best practices for determining resting energy expenditure in critically ill adults. Nutr Clin Pract. 2014;29(1):44-55.

19. Reeves MM, Davies PS, Bauer J, Battistutta D. Reducing the time period of steady state does not affect the accuracy of energy expenditure measurements by indirect calorimetry. J Appl Physiol (1985). 2004;97(1):130-4.

20. Compher C, Frankenfield D, Keim N, Roth-Yousey L. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. J Acad Nutr Diet. 2006;106(6):881-903.

21. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement. Med Sci Sports Exerc. 2018;50(4):837-45.

22. Robinson AP, Duursma RA, Marshall JD. A regression-based equivalence test for model validation: shifting the burden of proof. Tree Physiol. 2005;25(7):903-13.

23. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155-63.

24. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb). 2015;25(2):141-51.

25. Collins EG, Gater D, Kiratli J, Butler J, Hanson K, Langbein WE. Energy cost of physical activities in persons with spinal cord injury. Med Sci Sports Exerc. 2010;42(4):691-700.

26. Nightingale TE, Gorgey AS. Predicting Basal Metabolic Rate in Men with Motor Complete Spinal Cord Injury. Med Sci Sports Exerc. 2018;50(6):1305-12.

27. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR mHealth and uHealth. 2018;6(4):e94-e.

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

29. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med. 2018

30. Maijers MC, Verschuren O, Stolwijk-Swüste JM, van Koppenhagen CF, de Groot S, Post MWM. Is Fitbit Charge 2 a feasible instrument to monitor daily physical activity and handbike training in persons with spinal cord injury? A pilot study. Spinal Cord Ser Cases. 2018;4:84-.

31. Nightingale TE, Walhin JP, Thompson D, Bilzon JLJ. Predicting physical activity energy expenditure in wheelchair users with a multisensor device. BMJ Open Sport Exerc Med. 2015;1(1:bmjsem-2015-000008.

32. Wyndaele M, Wyndaele JJ. Incidence, prevalence and epidemiology of spinal cord injury: what learns a worldwide literature survey? Spinal cord. 2006;44(9):523-9.

33. Myers J, Lee M, Kiratli J. Cardiovascular disease in spinal cord injury: an overview of prevalence, risk, evaluation, and management. Am J Phys Med Rehabil. 2007;86(2):142-52.

34. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and Reliability of Fitbit Flex for Step Count, Moderate to Vigorous Physical Activity and Activity Energy Expenditure. PloS one. 2016;11(9):e0161224-e.

35. McCracken LA, Ma JK, Voss C, Chan FH, Martin Ginis KA, West CR. Wrist Accelerometry for Physical Activity Measurement in Individuals With Spinal Cord Injury-A Need for Individually Calibrated Cut-Points. Archives of physical medicine and rehabilitation. 2018;99(4):684-9.

36. Veerubhotla A, MS, Tsang K, BS, James K, Ding D. Classifying Intensity of Activity Using ActiGraph Monitor in Wheelchair Users. Proceedings of the RESNA Annual Conference. Arlington: RESNA Press; 2018.

37. Holmlund T, Ekblom-Bak E, Franzén E, Hultling C, Wahman K. Defining accelerometer cut-points for different intensity levels in motor-complete spinal cord injury. Spinal Cord. 2020;58(1):116-24.

38. Kang S, Kim Y, Byun W, Suk J, Lee J-M. Comparison of a Wearable Tracker with Actigraph for Classifying Physical Activity Intensity and Heart Rate in Children. Int J Environ Res Public Health. 2019;16(15):2663.

39. Brønd, J. C., Andersen, L. B., & Arvidsson, D. (2017). Generating ActiGraph counts from raw acceleration recorded by an alternative monitor.

40. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology. 2010;5(9):1315-6.