Assessing the validity and reliability and determining cut-points of the Actiwatch 2 in measuring physical activity

Kemp, C., Pienaar, P. R., Henst, R. H. P., Roden, L. C., Kolbe-Alexander, T. L. & Rae, D. E.

Original citation & hyperlink:
Kemp, C, Pienaar, PR, Henst, RHP, Roden, LC, Kolbe-Alexander, TL & Rae, DE 2020, 'Assessing the validity and reliability and determining cut-points of the Actiwatch 2 in measuring physical activity', Physiological Measurement, vol. 41, no. 8, 085001.
https://dx.doi.org/10.1088/1361-6579/aba80f

DOI 10.1088/1361-6579/aba80f
ISSN 0967-3334
ESSN 1361-6579

Publisher: IOP Publishing

ORIGINAL RESEARCH

Assessing the Validity and Reliability and Determining Cut-Points of the Actiwatch 2 in Measuring Physical Activity

Chadley Kemp BSc (Med)(Hons) [1], Paula R. Pienaar MSc (Med) [1,2], Rob H.P. Henst MSc (Med) [1], Laura C. Roden PhD [3,1], Tracy L. Kolbe-Alexander PhD [4,5,1], Dale E. Rae PhD [1]

Running head: Validation, Reliability & Calibration of the AW2

Affiliations:

[1] Health through Physical Activity, Lifestyle and Sport Research Centre & Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, South Africa.

[2] Department of Public and Occupational Health, Amsterdam Public Health Research Institute, VU University, Amsterdam UMC, Amsterdam, The Netherlands

[3] School of Life Sciences, Faculty of Health and Life Sciences, Coventry University, United Kingdom.

[4] School of Health and Wellbeing, Faculty of Health, Engineering & Sciences, University of Southern Queensland, Australia.

[5] School of Human Movement and Nutrition Sciences, The University of Queensland, Australia

Corresponding author:

Dr Dale Rae

Health through Physical Activity, Lifestyle and Sport Research Centre, Department of Human Biology, Faculty of Health Sciences, University of Cape Town

P.O. Box 115, Newlands, Cape Town, South Africa, 7725

Tel: + 27 21 650 4577

Email: Dale.Rae@uct.ac.za

Abstract

Objective: The Actiwatch 2 (AW2) is a wrist-worn accelerometer typically used to measure sleep. Although it can measure physical activity, there is limited evidence supporting its validity. We assessed the validity and reliability of the AW2 to measure sedentary behavior and physical activity (light, moderate, vigorous intensities), and reported their respective count cut-points.

Approach: 28 males and 22 females completed a task battery comprising three sedentary tasks and six randomized physical activity tasks at varying intensities, whilst wearing the AW2, a reference accelerometry device (Actigraph GT3X) and a cardiopulmonary gas analyzer on two separate occasions. Validity was assessed using correlations (AW2 counts versus GT3X counts and metabolic equivalent (MET) values), reliability using Bland-Altman analyses, and cut-points were determined using Receiver Operating Characteristic Area Under the Curve (AUC) analyses.

Main results: AW2 counts were positively correlated with GT3X counts (rho=0.902, p<0.001) and METs (rho=0.900, p<0.001). AW2-derived counts were comparable across independent assessment periods. Sedentary (AUC=0.99, cut-point: 256cpm) and vigorous activity (AUC=0.95, cut-point: 720cpm) were strongly characterized, and moderate activity (AUC=0.66, cut-point: 418cpm) was weakly characterized.

Significance: The use of the AW2 in physical activity monitoring looks promising for sedentary behavior, moderate and vigorous activity, however, further validation is needed.

Keywords: accelerometry; motion sensors; sedentary behavior; physical activity

Introduction

Wearable devices have gained increasing popularity in health research for their ability to return continuous objective measures of various health-related outcomes in free-living, habitual settings.[1] Among these outcomes are physical activity and sleep, both of which are important factors in the prevention and management of chronic diseases, including obesity, hypertension and type 2 diabetes mellitus.[2] At present, many devices exist to independently measure sedentary behavior, physical activity and sleep. Although it is possible to use multiple devices to independently measure parameters of sleep and physical activity, it is not ideal. Thus, a major challenge has been to identify a single valid and reliable monitoring device capable of measuring two or more of these variables concurrently.[3,4]

Accelerometers are favored as valid and reliable alternatives to traditional methods of objective physical activity and energy intake monitoring for their practical and non-invasive designs.[5,6] Among the most commonly used accelerometer devices for physical activity monitoring are the Actigraph GT1M and GT3X model devices (Actigraph, Pensacola, FL, USA).[6] Accelerometers are typically small wearable devices fixed to various points on the body (including the waist, wrist, hip, thigh or ankle) and can detect gross bodily movement in up to three orthogonal planes: anteroposterior, mediolateral and vertical.[5] Since the wearer typically receives no visual feedback relating to their measured physical activity, the risk of participants inadvertently altering or manipulating their habitual activity patterns is thought to be low.[7] The use of accelerometers is considered a gold-standard approach in the direct assessment of waking (total volume) movement behavior in free-living settings.[8] In addition to wake-time physical activity, accelerometers are now also used to indirectly monitor sleep patterns in free-living settings.[8] Despite differences in placement (hip-worn for measuring waking movement versus wrist-worn for measuring sleep),[9] the mechanics of sleep and physical activity monitoring by accelerometry are identical. The devices produce acceleration forces which, after being converted into voltage signals, are integrated as an average or peak acceleration according to a user-defined interval (or epoch) and are finally reported as arbitrary units called counts.[10,11] The conversion of raw accelerations into counts is done by manufacturer-specific algorithms, which may either be proprietary or open-source. Proprietary algorithms incur additional research challenges as raw data are often unavailable, meaning counts from one device may not be interpretable or comparable against those from other devices.[10] The premise of these algorithms in physical activity monitoring is to estimate physiological outcomes, including energy expenditure and the dose of physical activity exposure at various intensities (i.e., sedentary, light, moderate or vigorous).[12] For sleep monitoring, these algorithms are used to determine sleep-wake intervals by assessing whether gross motion is indicative of the wearer being awake, using the magnitude and duration of the acceleration signal.[13]

The Actiwatch 2 (AW2; Philips Respironics, Eindhoven, Netherlands) is widely used for research to directly measure parameters of sleep in free-living settings and has been validated against polysomnography.[8,13,14] While the AW2 is also capable of measuring physical activity using the native (albeit proprietary) algorithm to produce activity counts, there is limited evidence supporting its validity in physical activity. To the best of our knowledge, only a few studies have attempted to validate the AW2 for physical activity monitoring. For example, Neil-Sztramko et al[3] validated it's use in a convenience sample of mostly older, lean, active female shift workers; and Lee et al[15] validated it using a task menu comprising only treadmill running activities. The extrapolation of these findings to a broader demographic or the general population, or across a wide range of activities, however, is limited.

The purpose of this study was to test the validity of a single monitoring device (AW2), usually applied to measure sleep patterns, to quantify sedentary behavior and physical activity. This study expands upon the work by Neil-Sztramko et al[3] by including both males and females, encompassing a younger age range with wide variation in physical fitness. The aims were to (i) validate the AW2 as a tool for assessing sedentary behavior and physical activity by comparing its physical activity counts to both a reference physical activity monitor (Actigraph GT3X) and energy expenditure using indirect calorimetry, (ii) determine the AW2-derived count cut-points, maximizing sensitivity and specificity, for sedentary, light, moderate and vigorous physical activity using metabolic equivalent (MET) and count data, and Receiver Operating Characteristic (ROC) Curve analyses and (iii) assess the reliability of the AW2 to measure physical activity by comparing the physical activity counts across two independent assessment periods.

Methods

Participants

Apparently healthy males and females were eligible to participate in the study if they were between 18 and 60 years of age. Fifty participants with varying levels of physical activity were recruited through media posting. Participants first underwent health screening using the American College of Sports Medicine Exercise Pre-Participation Screening criteria[16] to ascertain participant safety during moderate to vigorous physical activity. Participants who answered 'Yes' to any of the questions during the screening process, as well as pregnant women, were excluded due to health risks and for integrity of cardiometabolic data. Ethical approval was obtained from the University of Cape Town's Human Research Ethics Committee (HREC No. 334/2017) and all participants provided written informed consent. This study was conducted in accordance with the ethical principles of the Declaration of Helsinki.[17]

Procedures

An overview of the study procedures is depicted in Figure 1. All testing was performed at the Division of Exercise Science and Sports Medicine, University of Cape Town with data collection sessions scheduled between 07h00 and 11h00 to control for diurnal intra-individual variability. Participants were asked not to eat or drink (except water) at least 2h prior to testing, not to consume caffeine at least 3h prior to testing and to avoid moderate to vigorous physical activity at least 6h prior to testing. Participants completed the Global Physical Activity Questionnaire[18] to characterize their habitual physical activity. Outcome variables were moderate-to-vigorous physical activity metabolic equivalent hours per week (MET h/wk). The investigator measured each participant's height (to the nearest 0.1cm) and weight using a stadiometer and digital scale, respectively. Waist circumference measurements (at the level of the umbilicus) were taken in triplicate using a standard tape measure and averaged. Waist circumference measurements were missing for four participants. Anthropometric outcome variables were height (m), weight (kg), waist circumference (cm) and body mass index (kg/m^2).

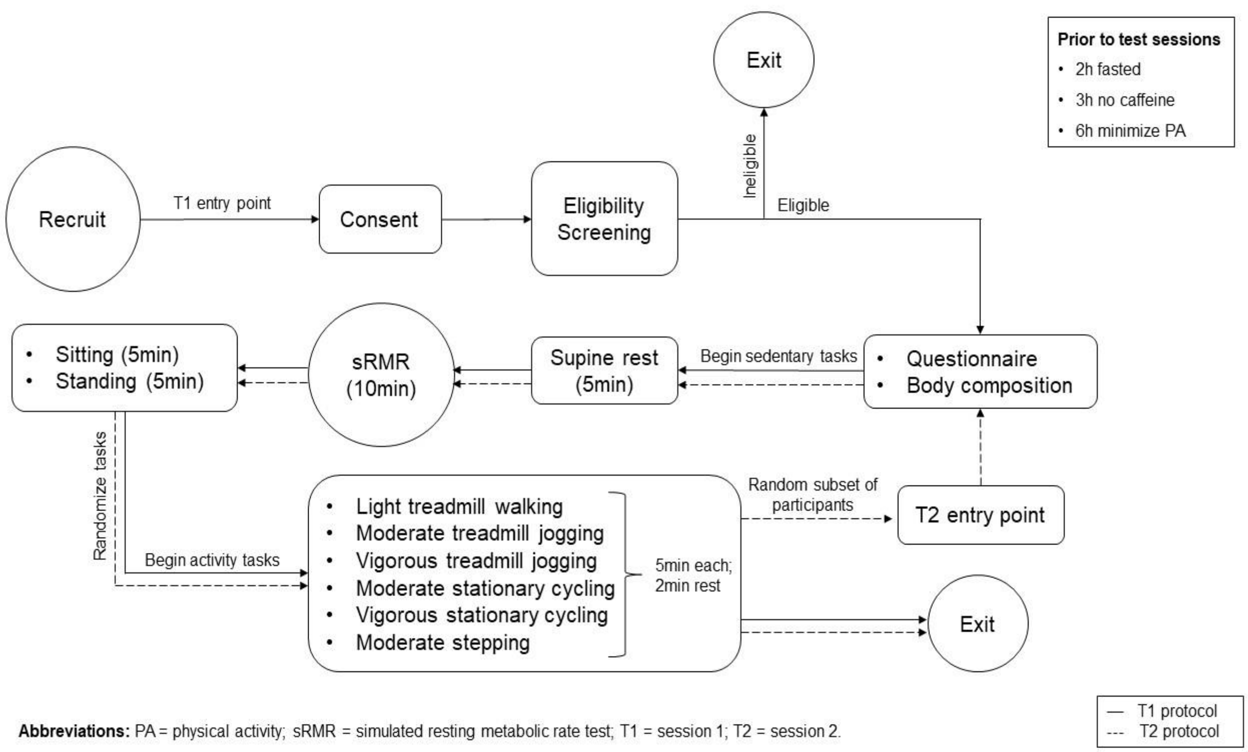

Figure 1. Study procedures overview

All participants were fitted with the AW2, worn on their non-dominant wrist, and GT3X, worn around the waist on an adjustable belt at hip level, in line with the midline of the thigh. They were also fitted with a mask connected to a cardiopulmonary gas analyzer (CPET; Cosmed CPET, Rome, Italy) for collection of metabolic data. The CPET was calibrated prior to each data collection session with a 3L calibration syringe and a standard gas mixture of 16% oxygen, 4% carbon dioxide and the balance nitrogen (BOC Special Gas, Afrox, Cape Town, South Africa).

Sedentary testing included 5min of supine rest, followed by 10min of a simulated resting metabolic rate (sRMR) test for normalization, and 5min each of sitting and standing. Participants were not permitted to speak, although they were permitted to listen to music and use their mobile phones during the supine rest, sitting and standing tasks. During the sRMR test, participants were required to remain awake, whilst lying still and listening to white noise. Respiratory exchange ratio data during supine rest were used to confirm fasting states.

Following the sedentary testing, a battery of six randomized physical activity tasks were performed. Participants exercised at self-selected paces eliciting light, moderate or vigorous intensities based on their rating of perceived exertion (RPE) scores during the task using the Borg 20-point scale.[19] RPE scores of 8±1, 12±1 and 15±1 were used for light, moderate and vigorous intensity exercise, respectively. Activity tasks comprised self-paced walking (light intensity) and jogging (moderate and vigorous intensities) on a treadmill (HP Cosmos treadmill, LE500CE, Nussdorf-Traunstein, Germany), stationary cycling (moderate and vigorous intensities) on a cycle ergometer (Wattbike Pro/Trainer, Wattbike Ltd., Nottingham, England) and stepping up and down a two-step 21cm stepping block (moderate intensity). Each activity was performed for 5min with at least 2min of rest preceding each task for a total data collection time of approximately 2h. A subset (n=18) of participants, chosen at random, were asked to return to the laboratory to repeat baseline and physical activity tasks to assess reliability of the AW2 to measure physical activity. These repeat sessions (T2) were scheduled 7±1 days after each participant's initial laboratory session. Participants were also required to comply with the same pre-test inclusion criteria as the initial session (T1).

Oxygen consumption (VO_2, mL/kg/min) was measured continuously using the CPET and values for each task, for each participant, were subsequently converted to MET values and normalized using their resting metabolic rate derived from the sRMR test. Both the AW2 and GT3X count data were collected in 15s epochs and reported as counts per minute (cpm) for physical activity measurements after processing using Philips Actiware (version 6.0.2) and ActiLife (version 6.10.4) software packages, respectively. Data from the AW2 were synchronized with the GT3X and Cosmed using event markers which created timestamps in the data.

Data and statistical analyses

Descriptive statistics are presented as mean ± standard deviation, or median and interquartile range. Normality was assessed using the Shapiro-Wilk test. Differences between gender groups were analyzed using a Mann-Whitney U test. Data during the activity tasks were collected in 15s epochs and only data from the 4th and 5th minute of each task were used to ensure steady state of metabolic data.[20] Correlations were performed using Spearman's rho tests. Receiver operating characteristic curve (ROC) analysis was used to calculate area under the ROC curve (AUC) so that count cut-points for the AW2, which maximized sensitivity and specificity, could be determined for sedentary, light, moderate and vigorous activities, defined as ≤ 1.5 METs, 1.5 - 3.0 METs, ≥ 3.0 METs and ≥ 6.0 METs, respectively. Count cut-points were subsequently confirmed with Youden's J statistic.[21] Quantification of predictive accuracy was determined using effect size equivalencies for AUC, Cohen's d and r.[22] Pairwise comparisons to determine differences between the AUC under independent ROC curves at varying intensities were also calculated. Differences in accelerometry and metabolic data between sessions T1 and T2 were analyzed using a paired-sample t-test or Wilcoxon matched-pair sign-ranked test. Repeatability of the AW2 to measure physical activity was performed using Bland-Altman analyses. Significance was accepted at $p < 0.05$. All data were analyzed using SPSS (IBM Corp., IBM SPSS Statistics, Version 20.0. Armonk, NY, USA).

Results

Descriptive characteristics of the 50 participants are presented in Table 1. Their ages ranged from 19 to 59 years, males were taller ($p < 0.01$), heavier ($p < 0.01$) and had greater waist circumferences ($p < 0.05$) relative to females. The self-reported levels of moderate-vigorous physical activity during the week were similar among males and females (female range: 0 - 74 MET h/wk, male range: 0 – 44 MET h/wk).

Table 1. Descriptive characteristics of the participants.

	All (n=50)	Males (n=28)	Females (n=22)	p value
Age (y)	29.5 (18.0)	29.5 (18.0)	29.0 (20.0)	0.822
Weight (kg)	69.9 (15.2)	75.1 (18.2)	67.8 (16.4)	0.004
Height (m)	1.71 ± 0.09	1.77 ± 0.07	1.63 ± 0.05	<0.001
BMI (kg/m^2)	24.1 (5.5)	24.1 (5.6)	23.9 (5.7)	0.922
Waist circumference (cm)	78.5 (12.2) [a]	82.5 (13.8) [b]	74.5 (12.6) [c]	0.027
MVPA (MET h/wk)	10.0 (8.7)	10.0 (9.4)	10.0 (8.5)	0.645

Data are presented as mean ± standard deviation or median (interquartile range). BMI: body mass index; MVPA: moderate-vigorous physical activity; MET: metabolic equivalent. The p value represents the gender comparison tested using an independent t-test or Mann-Whitney U test. Waist circumference available for a subset: [a] n=46; [b] n=26; [c] n=20.

MET and count data for each task, as well as correlations between the device counts and METs, and between the devices themselves are presented in Table 2. While the purpose of the study was to validate the AW2 device, the correlations between the GT3X counts and MET are presented as a comparator for the AW2 v MET correlation. AW2 activity counts were positively correlated to MET values for the sitting ($p = 0.007$), standing ($p = 0.007$), light treadmill walking

($p = 0.010$), moderate treadmill jogging ($p < 0.001$), vigorous treadmill jogging ($p = 0.009$), and vigorous stationary cycling ($p = 0.028$) tasks. GT3X activity counts were positively correlated with MET values for the sitting ($p = 0.020$), light treadmill walking ($p < 0.001$), moderate treadmill jogging ($p < 0.001$), vigorous treadmill jogging ($p = 0.001$), and moderate stepping ($p = 0.009$) tasks. AW2 and GT3X counts were positively correlated for the moderate treadmill jogging ($p = 0.002$) and moderate stepping ($p = 0.011$) tasks. In most cases, the AW2 correlations were weak but significant, except for correlations for moderate treadmill jogging, which were all moderate and significant.

Table 2. Metabolic and count data measured during each task as well as their correlations.

Task	Predicted MET	Measured MET	AW2 (cpm)	GT3X (cpm)	AW2 vs. MET	GT3X vs. MET	AW2 vs. GT3X
					Spearman's rho (p value)		
Supine rest	1.3	1.1 ± 0.1	8.5 (23.1)	0.0 (0.0)	-0.024 (0.869)	0.146 (0.311)	0.014 (0.925)
Sitting	1.5	1.2 ± 0.1	34.8 (68.1)	0.0 (4.1)	0.377 (0.007)	0.328 (0.020)	0.220 (0.126)
Standing	1.8	1.2 ± 0.2	37.8 (79.3)	0.0 (1.0)	0.377 (0.007)	0.177 (0.219)	0.106 (0.465)
Light treadmill walking	3.5	3.3 ± 0.6	484.3 (248.8)	2131.0 (1413.0)	0.359 (0.010)	0.590 (<0.001)	0.063 (0.665)
Moderate treadmill jogging	>6.0	6.8 ± 1.4	1749.0 (1529.3)	7601.8 (3497.5)	0.570 (<0.001)	0.574 (<0.001)	0.428 (0.002)
Vigorous treadmill jogging	>6.0	8.3 ± 1.9	2610.1 ± 1140.9	7509.1 ± 2106.6	0.364 (0.009)	0.440 (0.001)	0.120 (0.407)
Moderate stationary cycling	6.8	4.9 (1.8)	80.0 (119.0)	74.8 (685.8)	0.058 (0.687)	-0.081 (0.576)	0.033 (0.820)
Vigorous stationary cycling	8.8	6.8 (3.2)	219.5 (193.6)	891.5 (1509.1)	0.312 (0.028)	-0.045 (0.755)	0.100 (0.490)
Moderate stepping	7.5	6.0 ± 1.2	712.2 ± 332.2	3206.1 ± 720.1	0.114 (0.431)	0.368 (0.009)	0.355 (0.011)

Data are presented as mean ± standard deviation, median (interquartile range), or Spearman's rho. MET: metabolic equivalent; AW2: Actiwatch 2; GT3X: Actigraph GT3X; cpm: counts per minute. Correlations were determined using Spearman's rho test. Significance was accepted at $p < 0.05$.

Figure 2 displays the correlations for all tasks combined between AW2 and GT3X counts (A and B), AW2 counts and METs (C and D) and GT3X counts and METs (E and F, for comparison purposes). The left panel (A, C, E) includes all activities while the right panel excludes the cycling tasks (B, D, F), since the intensity of the cycling tasks was poorly estimated by both devices (Table 2). The counts measured by the AW2 were positively correlated with counts measured by the GT3X regardless of whether cycling was included (Figure 2A, $p < 0.001$) or excluded (Figure 2B, $p < 0.001$) from the analyses. Counts measured by the AW2 were positively correlated with task METs for analyses including (Figure 2C, $p < 0.001$) and excluding (Figure 2D, $p < 0.001$) the cycling tasks. Similarly, counts measured using the GT3X were positively correlated with task METs for analyses including (Figure 2E, $p < 0.001$) and excluding (Figure 2F, $p < 0.001$) the cycling tasks. In all cases, the strengths of the correlations were improved through removal of the cycling tasks.

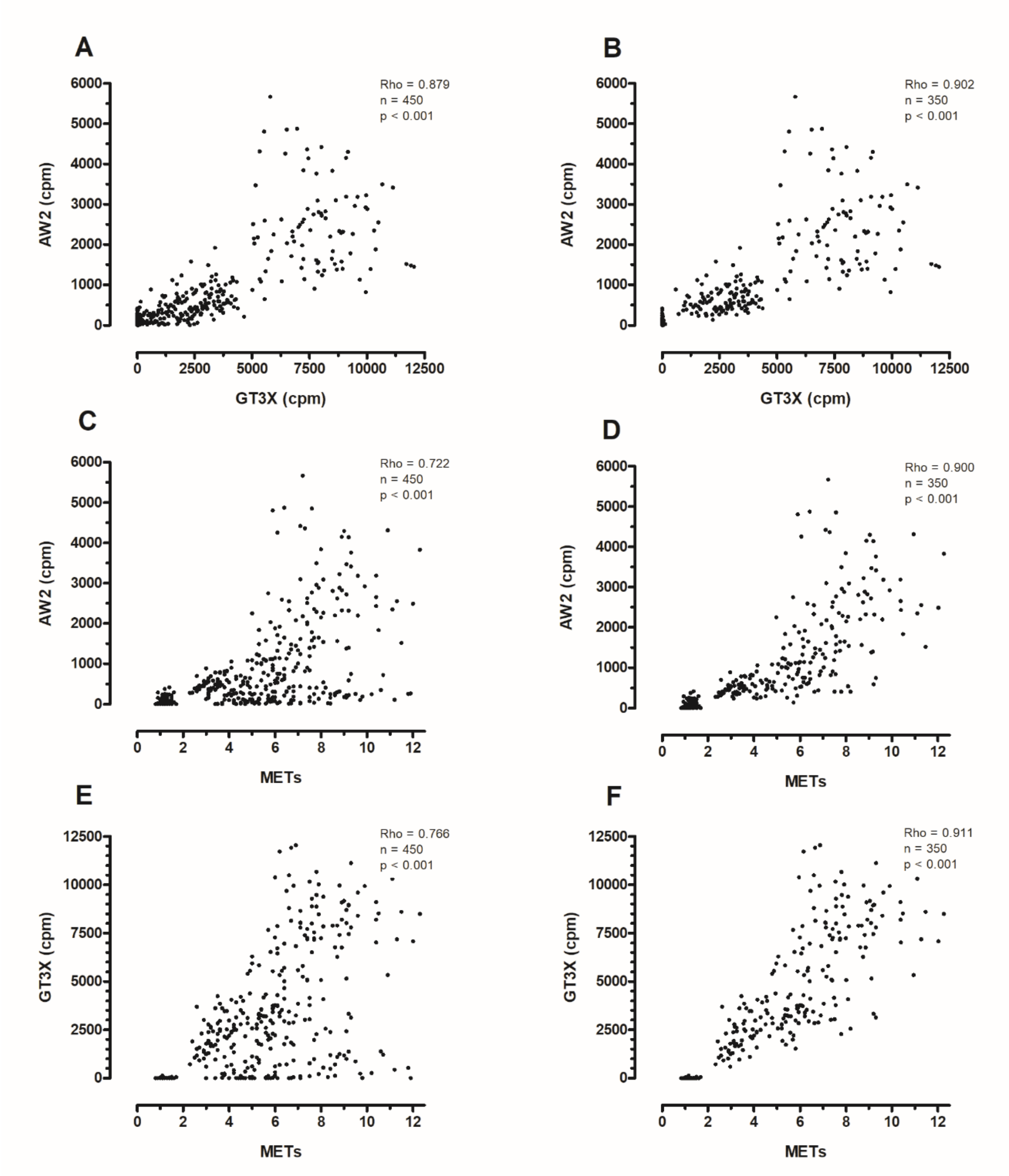

Figure 2. Correlations between counts measured by the AW2 and GT3X accelerometers **(A, B)**, AW2 counts and task MET **(C, D)** and GT3X counts and task MET **(E, F)**. The left panel includes all tasks **(A, C, E)** while the cycling tasks are omitted from the right panel graphs **(B, D, F)**. cpm: counts per minute; MET: metabolic equivalents. Correlations were determined using Spearman's rho test.

ROC analysis revealed that the AW2's ability to characterize sedentary activity was strong (AUC = 0.93, 95% CI: 0.90 to 0.95, $p < 0.001$) and that using a count cut-point of 99 cpm produced 85.5% sensitivity and 86.6% specificity (Figure 3A). The ability of the AW2 to characterize light activity was weak (AUC = 0.47, 95% CI: 0.38 to 0.55, $p = 0.600$), and a cut-point of 578 cpm elicited a 90.9% sensitivity but 33.9% specificity (Figure 3C). AW2 characterization of moderate activity was also weak (AUC = 0.58, 95% CI: 0.53 to 0.63, $p = 0.007$) and a count cut-point of 259 cpm gave 63.0% sensitivity and 56.5% specificity, indicative of high false-positive rates (Figure 3B). While vigorous activity characterization by the AW2 was acceptable (AUC = 0.84, 95% CI: 0.80 to 0.88, $p < 0.001$), a count cut-point of 400 cpm yielded 72.3% sensitivity and 73.5% specificity (Figure 3D). Based on the poor sensitivity and specificity for the characterization of light activity, the ability of the AW2 to reliably determine the cut-points for light activity intensity was not justified; and are therefore not reported.

The ability of the GT3X to characterize sedentary activity was almost perfect (AUC = 0.97, 95% CI: 0.95 to 0.98, $p < 0.001$) using a cut-point of 42 cpm, yielding 99% sensitivity and 89.5% specificity (Figure 3A). In contrast, the ability of the GT3X to characterize light activity was weak (AUC = 0.52, 95% CI: 0.45 to 0.60, $p = 0.699$) using a cut-point of 2328 cpm, produced 90.9% sensitivity and 39.7% specificity (Figure 3C). Moderate activity was weakly characterized by the GT3X (AUC = 0.62, 95% CI: 0.56 to 0.67, $p < 0.001$) using a cut-point of 1442, producing 66.7% sensitivity and 62.5% specificity. The ability of the GT3X to characterize vigorous activity (AUC = 0.86, 95% CI: 0.82 to 0.89, $p < 0.001$) was acceptable, and using a cut-point of 2836 cpm, resulted in 68.2% sensitivity and 84.8% specificity.

A pairwise comparison of AW2 and GT3X ROC curves indicated that there were no differences between the AUC for light ($p = 0.522$), moderate ($p = 0.419$) or vigorous ($p = 0.531$) cut-points, except for the AUC for sedentary ($p = 0.036$) cut-points which were significantly different.

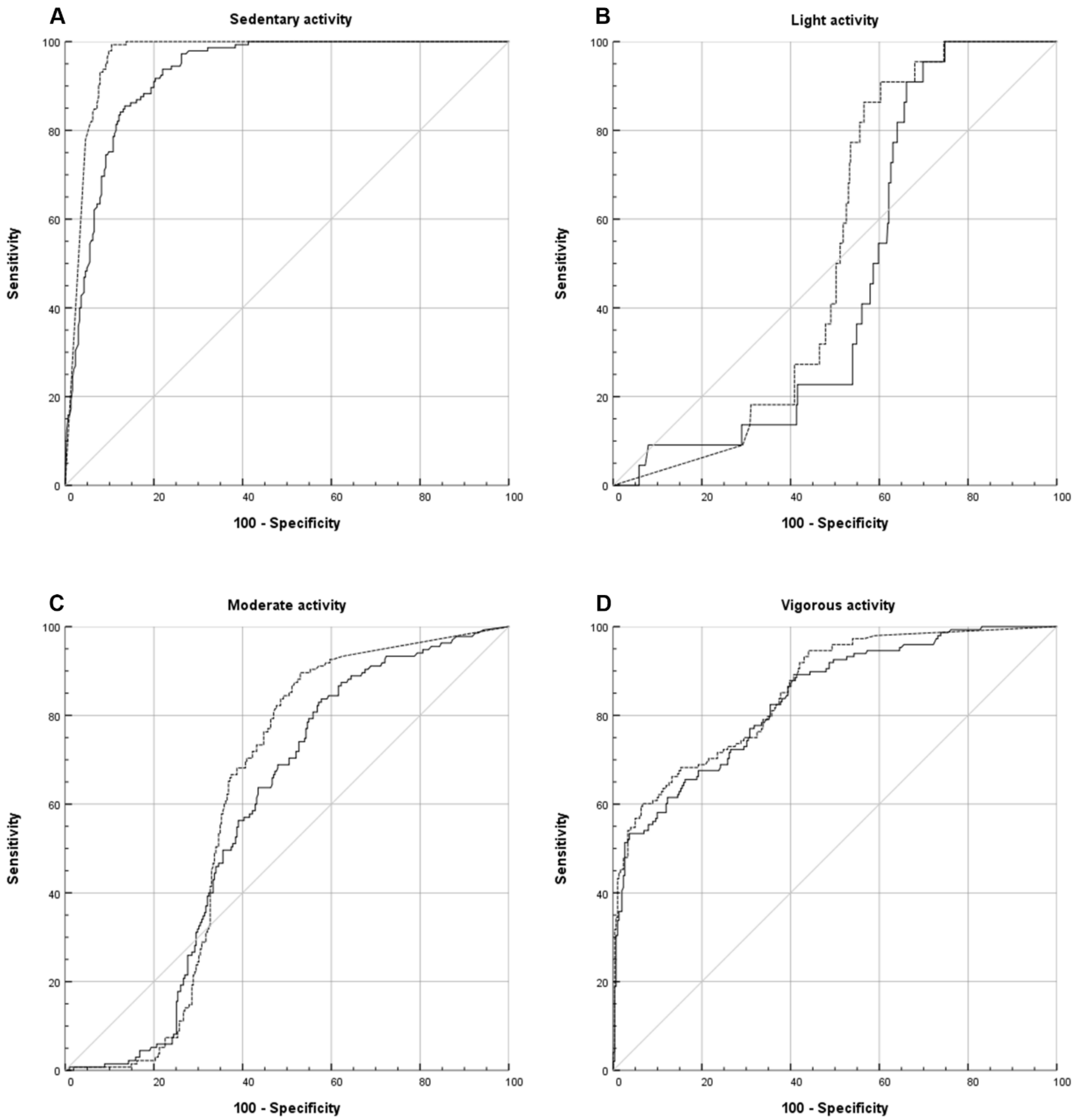

Figure 3. Receiver operating characteristic (ROC) analyses for sedentary, light, moderate and vigorous activities for the Actiwatch 2 (solid line) and Actigraph GT3X (dashed line) devices.

Removing cycling tasks from the sample pool improved the ability of the AW2 to characterize sedentary activity to nearly perfect (AUC = 0.99, 95% CI: 0.98 to 1.00, $p < 0.001$) with a count cut-point of 256 cpm, giving 97.9% sensitivity and 96.6% specificity (Figure 4A). Characterization of light activity using the AW2 remained weak (AUC = 0.46, 95% CI: 0.38 to 0.54, $p = 0.548$) with a cut-point of 273 cpm, giving 81.0% sensitivity and 45.0% specificity (Figure 4C). AW2 characterization of moderate activity improved but remained weak (AUC = 0.66, 95% CI: 0.60 to 0.71, $p < 0.001$), with a count cut-point of 418 cpm producing 80.5% sensitivity and 59.7% specificity (Figure 4B). Characterization of vigorous activity using the AW2 was almost perfect (AUC = 0.95, 95% CI: 0.93 to 0.97, $p < 0.001$) with a count cut-point of 720 cpm, yielding 88.2% sensitivity and 85.1% specificity (Figure 4D).

Similarly, with the cycling tasks removed, the ability of the GT3X to characterize sedentary activity remained nearly perfect (AUC = 0.99, 95% CI: 0.98 to 1.0, $p < 0.001$) with a cut-point of 46 cpm, giving 99.3% sensitivity and 97.8% specificity (Figure 4A). GT3X characterization of light activity remained weak (AUC = 0.44, 95% CI: 0.37 to 0.50, $p = 0.341$) with a cut-point of 655 cpm, producing 71.4% sensitivity and 44.1% specificity (Figure 4C). Characterization of moderate activity with the GT3X improved, but remained weak (AUC = 0.65, 95% CI: 0.60 to 0.71, $p < 0.001$) using a count cut-point of 1585 cpm, producing 93.9% sensitivity and 60.1% specificity (Figure 4B). Vigorous activity (AUC = 0.96, 95% CI: 0.95 to 0.98, $p < 0.001$) was characterized almost perfectly using the GT3X using a cut-point of 3707 cpm, giving 86.3% sensitivity and 91.9% specificity (Figure 4D).

A pairwise comparison of AW2 and GT3X ROC curves with the cycling tasks excluded yielded no differences between the AUC for sedentary ($p = 0.471$), light ($p = 0.800$), moderate ($p = 0.922$) or vigorous ($p = 0.384$) cut-points.

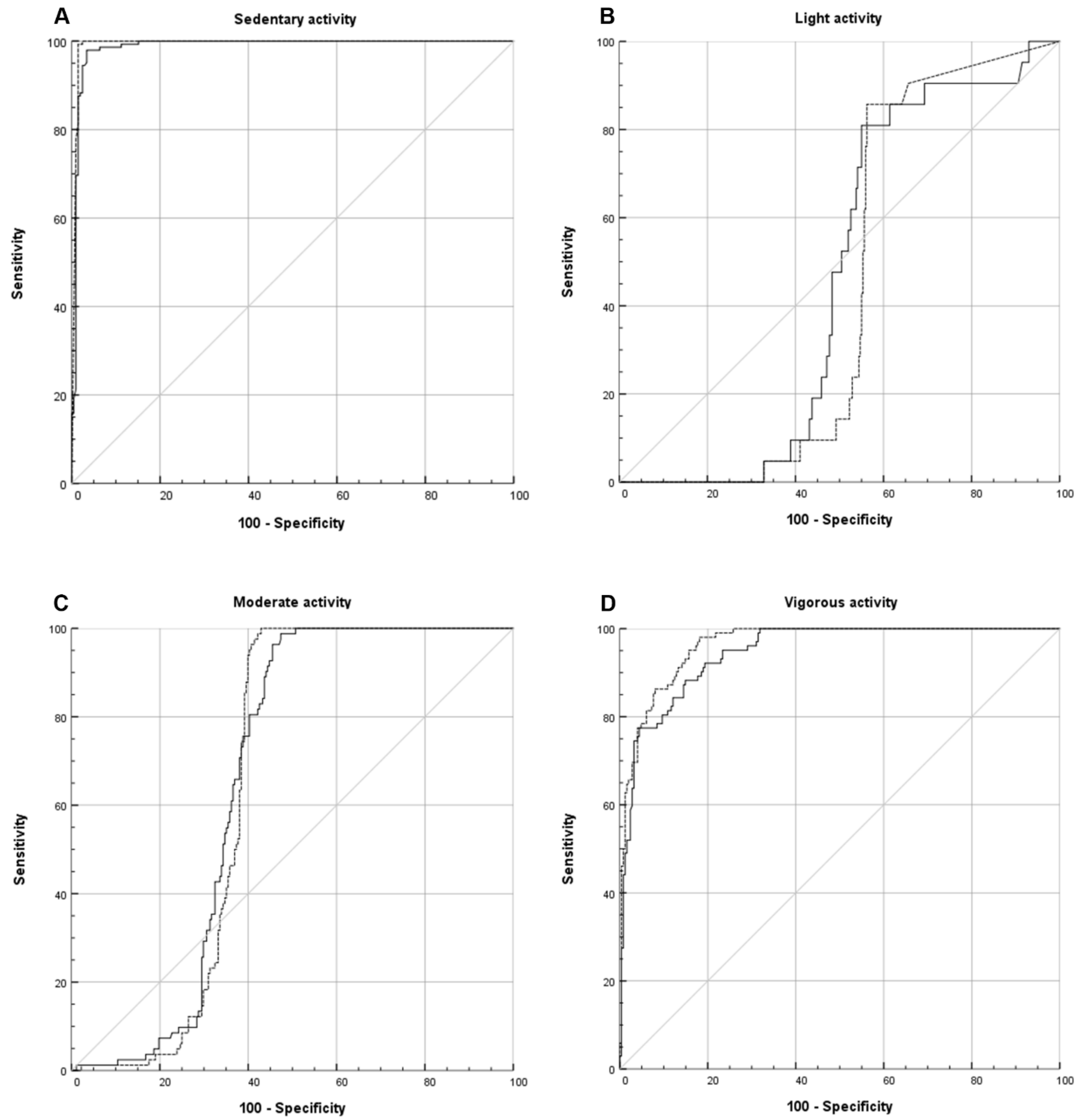

Figure 4. Receiver operating characteristic (ROC) analyses for sedentary, light, moderate and vigorous activities for the Actiwatch 2 (solid line) and Actigraph GT3X (dashed line) devices with cycling activity tasks omitted.

A comparison of metabolic and AW2 count data measured on two separate occasions (T1 and T2) are presented in Table 3. There were no differences in MET values or AW2 counts measured between T1 and T2. Bland-Altman analyses (Figure 5A - 5I) show the agreement of AW2 counts in the activity tasks graphically.

Table 3. Comparison of metabolic and Actiwatch 2 count data measured on two separate occasions (T1 and T2) in a subset of participants (n=18).

Tasks	MET			AW2 (cpm)		
	T1	T2	p value	T1	T2	p value
Supine rest	1.1 ± 0.1	1.1 ± 0.1	0.668	15.0 (30.5)	20.8 (43.0)	0.647
Sitting	1.2 (0.2)	1.1 (0.2)	0.557	35.3 (99.9)	48.8 (63.9)	0.472
Standing	1.2 ± 0.2	1.2 ± 0.2	0.750	35.3 (41.4)	38.5 (73.5)	0.246
Light treadmill walking	3.5 ± 0.5	3.7 ± 0.7	0.205	518.4 ± 179.0	518.5 ± 132.7	0.998
Moderate treadmill jogging	6.9 ± 1.2	7.3 ± 1.6	0.232	2115.0 (1567.5)	2001.5 (1021.1)	0.983
Vigorous treadmill jogging	8.2 ± 1.4	8.5 ± 1.8	0.408	2609.0 (1349.8)	2367.8 (1152.6)	0.845
Moderate stationary cycling	5.1 ± 1.1	5.3 ± 1.3	0.621	64.0 (211.1)	104.0 (214.6)	0.420
Vigorous stationary cycling	6.9 ± 1.8	6.7 ± 1.9	0.658	204.0 (247.0)	178.5 (284.3)	0.396
Moderate stepping	6.1 ± 1.0	6.1 ± 1.0	0.810	720.8 (383.8)	581.3 (307.1)	0.828

Data are presented as mean ± standard deviation or median (interquartile range). MET = metabolic equivalent; AW2 = Actiwatch 2 (presented in counts per minute). Significance was determined using either a paired t-test or Wilcoxon matched-pair signed-rank test. Significance was accepted at $p < 0.05$.

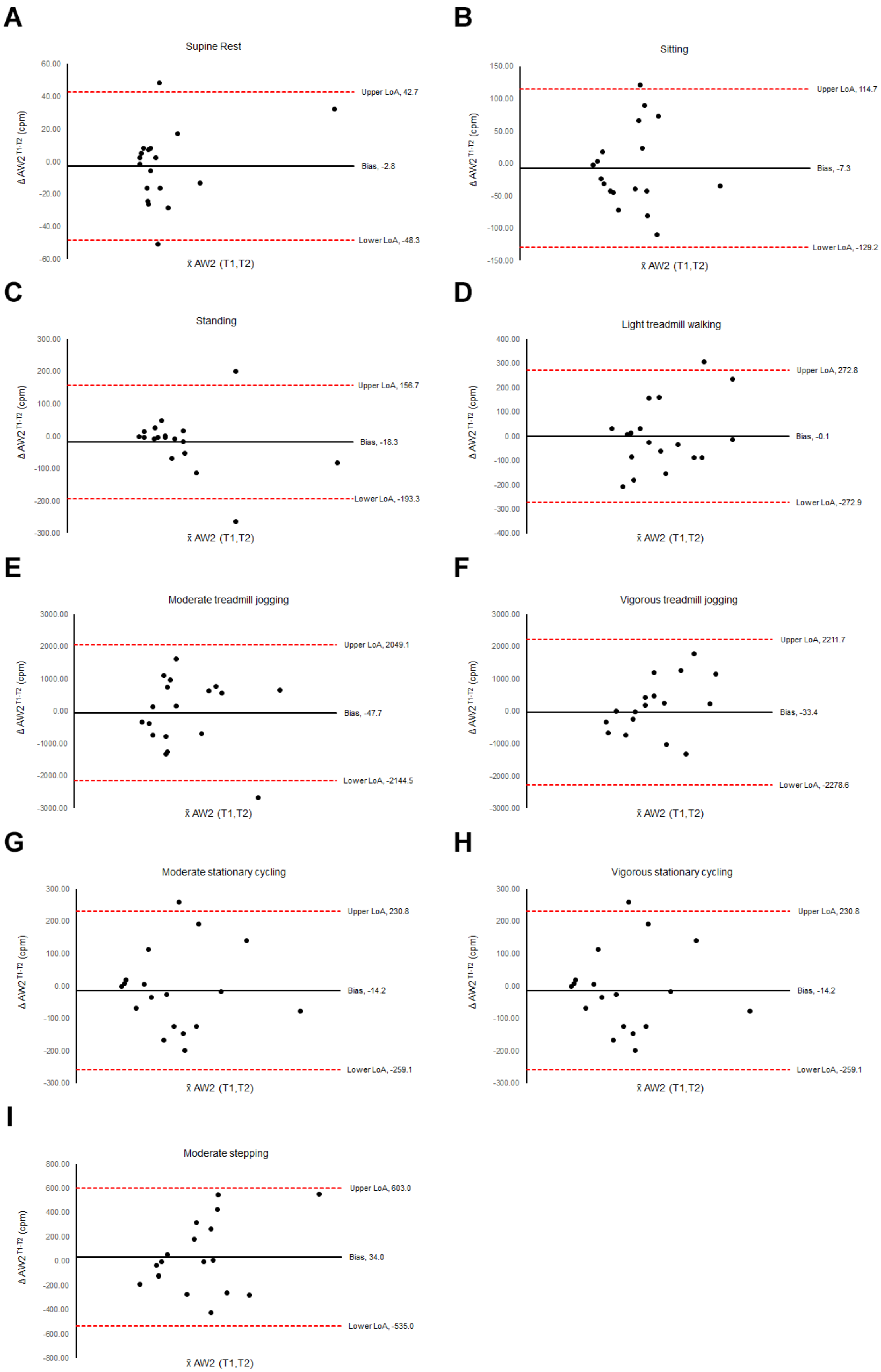
A
Supine Rest
Upper LoA, 42.7
Bias, -2.8
Lower LoA, -48.3
B
Sitting
Upper LoA, 114.7
Bias, -7.3
Lower LoA, -129.2
C
Standing
Upper LoA, 156.7
Bias, -18.3
Lower LoA, -193.3
D
Light treadmill walking
Upper LoA, 272.8
Bias, -0.1
Lower LoA, -272.9
E
Moderate treadmill jogging
Upper LoA, 2049.1
Bias, -47.7
Lower LoA, -2144.5
F
Vigorous treadmill jogging
Upper LoA, 2211.7
Bias, -33.4
Lower LoA, -2278.6
G
Moderate stationary cycling
Upper LoA, 230.8
Bias, -14.2
Lower LoA, -259.1
H
Vigorous stationary cycling
Upper LoA, 230.8
Bias, -14.2
Lower LoA, -259.1
I
Moderate stepping
Upper LoA, 603.0
Bias, 34.0
Lower LoA, -535.0
Δ AW2 T1-T2 (cpm)
x̄ AW2 (T1,T2)

Figure 5. Bland-Altman plots demonstrating agreement in AW2 count data (counts per minute; cpm) across two independent assessment periods (T1, session 1; T2, session 2) during supine rest **(A)**, sitting **(B)**, standing **(C)**, light treadmill walking **(D)**, moderate treadmill jogging **(E)**, vigorous treadmill jogging **(F)**, moderate stationary cycling **(G)**, vigorous stationary cycling **(H)** and moderate stepping **(I)** activity tasks. The solid line represents the mean (bias) while the dotted lines represent the upper and lower limits of agreement (LoA, ± 1.96 SD). The difference in AW2 activity counts are plotted on the y-axis, and the mean of the AW2 activity counts across are plotted on the x-axis.

Discussion

The present study performed a calibration and validation of the AW2 by comparing AW2-derived physical activity outcomes (counts per minute) against objectively measured oxygen consumption (MET), and physical activity outcomes of a reference device (GT3X). The calibration component of the study was performed using ROC curve analyses, and AUC to determine count cut-points using the native algorithm typically used in the analysis of sleep parameters. Additionally, this study assessed the reliability of the AW2 to measure physical activity outcomes in an array of tasks over two independent assessment periods, using Bland-Altman analyses.

By using the native sleep algorithm to predict physical activity outcomes at varying intensities, the AW2 may be an effective tool to concurrently report on both sleep and physical activity behavior in a research setting. This may provide physical activity researchers with a broader understanding of the relationship between sleep and activity during wakeful periods, including: the combined effect of time spent in sleep, sedentary activity, light and moderate-to-vigorous physical activity intensities, and dose of the activity exposure. Moreover, the ability of the AW2 to concurrently deliver meaningful information that is comparable against criterion approaches of sleep, physical activity and energy expenditure measurements is important to minimize participant burden and measurement bias.

Currently there are a limited number of studies assessing physical activity using the AW2 device.[3,15] While one study has reported AW2 count cut-points for the measurement of physical activity at sedentary, light, moderate and vigorous intensities, the findings are limited to predominantly older (40.0 ± 14.9 years), lean (22.4 ± 3.1 kg/m^2), active (37.5 ± 24.5 MET h/wk) female shift-workers[3]. The present study sought to expand upon the work of Neil-Sztramko et al[3] by including both male and female participants of varying degrees of cardiorespiratory fitness from the general population. Moreover, the present study employed an alternative menu of physical activity tasks (including sedentary and stationary cycle tasks) and used stronger methodological steps, namely: inclusion of an objective parameter of activity task intensity using the Borg 20-point scale;[19] using a sRMR to determine baseline oxygen consumption in normalizing metabolic data during ROC analysis to determine MET values (versus using the Harris-Benedict predicted resting metabolic rate); and confirming participants' fasted states in real-time using the CPET to minimize the thermic effect of feeding for metabolic integrity.

The data reported in the present study suggest that the AW2 count cut-points were acceptable for characterizing sedentary activity (256 cpm) and vigorous (720 cpm) intensity activity. However, the ability of the AW2 to characterize moderate (418 cpm) intensity activity was weak and it was unable to characterize light activity. Overall, count data between both AW2 and GT3X monitoring devices were acceptably correlated with each other, and objective energy expenditure measurements (MET) suggesting that the AW2 is comparable to a valid and reliable reference activity monitor, despite differences in the anatomical placement (i.e. wrist- versus waist-worn) of the respective monitoring devices. Moreover, the AW2 illustrates a high level of reproducibility in its ability to predict physical activity outcomes during sedentary and active tasks, in a laboratory environment.

Neil-Sztramko et al[3] reported AW2 physical activity count cut-points for sedentary activity and moderate and vigorous activity intensity of 145 cpm, 274 cpm and 597 cpm, respectively. Light intensity count cut-points were reported to be 145 – 274 cpm. These cut-points are lower than the cut-points reported in the present study and have a stronger ability to discriminate between light and moderate activity intensities. However, the present study reports cut-points with a stronger ability to discriminate sedentary activity and vigorous intensity activity. Additionally, correlations reported in the present study were stronger between AW2 and Actigraph (GT3X/+) counts, and MET values than those reported by Neil-Sztramko et al[3]. While the present study did report positive correlations between individual activity tasks, it is noted that these correlations were mostly weak, as were similarly reported by Neil-Sztramko et al[3].

In addition to performing a calibration of the AW2, this study also reported count cut-points for the Actigraph GT3X for a qualitative assessment of count data. The Freedson VM3 (2011)[23] count cut-points are among the most frequently used cut-points in physical activity research. The corresponding count cut-points for physical activity at either light (0 - 2690 cpm), moderate (2691 – 6166 cpm) or vigorous (>6167 cpm) intensity are 655 cpm, 1585 cpm and 3707 cpm, respectively. Whilst Freedson-reported count cut-points are higher than those measured, it is thought that discrepancies may be attributed to a small sample size, population bias or procedural differences. For instance, cycling tasks were found to distort physical activity count data (and consequently, physical activity intensity and energy expenditure) in both AW2 and GT3X monitoring devices owing possibly to the static nature of the cycling task, and variability of reported activity by each device. This is substantiated by an improvement in the strength of correlations between AW2 and GT3X activity counts with METs respectively (Figure 2) after cycling tasks were omitted from the analysis. In hindsight, substituting cycling tasks with alternative habitual lifestyle activity tasks (such as lifting or carrying tasks) may have yielded superior findings.

The methodological steps in the present study followed best practice recommendations,[24] which included using criterion approaches for energy expenditure (via indirect calorimetry); using a broad age range of participants; including both men and women with a range of body mass index and cardiorespiratory fitness; and a menu of activity tasks spanning a MET range of 1.1 to 10 METs to discriminate sedentary activity, light, moderate and vigorous physical activity. A key methodological step was having a personalized and objectively measured independent variable to compare AW2 and GT3X count data, which was afforded via indirect calorimetry.

Another methodological strength was the comparison of AW2 physical activity count data with a waist-worn (GT3X) monitor, which is the preferred approach for detection of moderate to vigorous physical activity intensities.[25] The AW2 demonstrated the ability to discriminate vigorous activity tasks acceptably, despite being wrist-worn. Both AW2 and GT3X devices poorly discriminated light and moderate intensity activity tasks. It is thought that the inclusion of additional lower intensity activity tasks may have provided stronger compliance by both devices within these lower MET ranges (i.e. 1.5 - 6.0 METs).

A major limitation to the data presented is due to a small sample size, which resulted in a loss of statistical power. While efforts were made to maintain homogeneity between male and female participants, the cohort was not normally distributed. Moreover, it was not possible to verify participants' cardiorespiratory fitness using the subjective physical activity questionnaire, meaning the broad range of MVPA (MET h/wk) may be the result of misrepresentation of habitual physical activity. Future work using a larger cohort of participants, comprising a broader distribution of age and habitual physical activity (or cardiorespiratory fitness) is necessary. Another major limitation includes using cut-point methods (ROC-AUC) to define intensity categories. Future work should embrace alternative analytical techniques (such as pattern recognition analysis) for predicting energy expenditure, which utilizes components of raw acceleration signals and minimizes the over- or under-representation of energy expenditure.[24]

Given the variability of tasks performed over a 24h period, or over multiple days, it is also recommended that future research determine the ability of the AW2 to measure waking movement behavior in free-living settings. Given the stringency of controlled laboratory conditions, natural human movement patterns may have been restricted, and thus not properly reflected in this study as in a free-living setting.

Conclusion

In conclusion, the count cut-points reported in the present study provides promising evidence for the use of the AW2 to discriminate between sedentary activity, as well as moderate and vigorous physical activity intensities in apparently healthy adult males and females. Further work is required to confirm these findings and to refine best practice recommendations for concurrent sleep and physical activity data collection in the general population, and in niche cohorts. Further cross-validation of the AW2 to concurrently measure physical activity of varying intensities, and parameters of sleep would also aid to broaden the understanding of the combined effect of sleep and dose exposure to physical activity intensities during wake periods.

Acknowledgements

The authors have no conflicts of interest to declare, financial or otherwise. We would like to thank Dr Jacolene Kroff from the University of Cape Town for assisting us with the use of the Cosmed cardiopulmonary gas analyzer, and interpretation of its data. This study was supported by a grant from the AstraZeneca Research Trust. Chadley Kemp was supported through scholarships from the National Research Foundation and the University of Cape Town.

References

1. Bassett Jr DR, Rowlands AV, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S32.
2. Ndahimana D, Kim E. Measurement methods for physical activity and energy expenditure: A review. Clinical nutrition research. 2017;6(2):68-80.
3. Neil-Sztramko SE, Rafn BS, Gotay CC, Campbell KL. Determining activity count cut-points for measurement of physical activity using the Actiwatch2 accelerometer. Physiol Behav. 2017;173:95-100.
4. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. Sleep health. 2015;1(4):275-284.
5. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical guide to measuring physical activity. Journal of the Academy of Nutrition and Dietetics. 2014;114(2):199-208.
6. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: An overview of objective measures. Frontiers in nutrition. 2014;1:5.
7. Ainslie PN, Reilly T, Westerterp KR. Estimating human energy expenditure. Sports medicine. 2003;33(9):683-698.
8. Lambiase MJ, Gabriel KP, Chang Y, Kuller LH, Matthews KA. Utility of actiwatch sleep monitor to assess waking movement behavior in older women. Med Sci Sports Exerc. 2014;46(12):2301.
9. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep. 2003;26(3):342-392.
10. Sasaki JE, da Silva KS, da Costa, Bruno Gonçalves Galdino, John D. Measurement of physical activity using accelerometers. In: Computer-assisted and web-based innovations in psychology, special education, and health. Elsevier; 2016:33-60.

11. Rowlands AV, Stiles VH. Accelerometer counts and raw acceleration output in relation to mechanical loading. J Biomech. 2012;45(3):448-454.

12. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the american heart association. Circulation. 2013;128(20):2259-2279.

13. Cheung J, Zeitzer JM, Lu H, Mignot E. Validation of minute-to-minute scoring for sleep and wake periods in a consumer wearable device compared to an actigraphy device. Sleep Science and Practice. 2018;2(1):11.

14. Chow CM, Wong SN, Shin M, et al. Defining the rest interval associated with the main sleep period in actigraph scoring. Nature and science of sleep. 2016;8:321.

15. Lee P, Tse CY. Calibration of wrist-worn ActiWatch 2 and ActiGraph wGT3X for assessment of physical activity in young adults. Gait & posture. 2019;68:141-9.

16. Riebe D, Franklin BA, Thompson PD, et al. Updating ACSM's recommendations for exercise preparticipation health screening. Medicine & Science in Sports & Exercise. 2015;47(11):2473-9.

17. General Assembly of the World Medical Association. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. J Am Coll Dent. 2014;81(3):14.

18. Armstrong T, Bull F. Development of the world health organization global physical activity questionnaire (GPAQ). Journal of Public Health. 2006;14(2):66-70.

19. Borg G. Borg's perceived exertion and pain scales. Human kinetics; 1998.

20. Kaminsky LA, Ozemek C. A comparison of the actigraph GT1M and GT3X accelerometers under standardized and free-living conditions. Physiol Meas. 2012;33(11):1869.

21. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-35.

22. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, cohen's d, and r. Law Hum Behav. 2005;29(5):615-620.

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

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(1 Suppl 1):S1.

25. Trost SG, Mciver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc. 2005;37(11 Suppl):531.