Journal of Science and Medicine in Sport 22 (2019) 677–683

Contents lists available at ScienceDirect

Journal of Science and Medicine in Sport

journal homepage: www.elsevier.com/locate/jsams

Original research

Wrist-specific accelerometry methods for estimating free-living physical activity

Michael I.C. Kingsley[a,*], Rashmika Nawaratne[b], Paul D. O'Halloran[c], Alexander H.K. Montoye[d], Damminda Alahakoon[b], Daswin De Silva[b], Kiera Staley[e], Matthew Nicholson[e]

[a] Exercise Physiology, La Trobe Rural Health School, La Trobe University, Australia
[b] Research Centre for Data Analytics and Cognition, School of Business, La Trobe University, Australia
[c] School of Psychology and Public Health, La Trobe University, Australia
[d] Research in Applied Physiology Laboratory, Integrative Physiology and Health Science Department, Alma College, USA
[e] Centre of Sport and Social Impact, School of Business, La Trobe University, Australia

ARTICLE INFO

Article history:
Received 20 June 2018
Received in revised form 1 November 2018
Accepted 3 December 2018
Available online 11 December 2018

Keywords:
Accelerometer
Actigraph
Physical activity
Wrist
Hip
Artificial neural network

ABSTRACT

Objectives: To compare accelerometry-derived estimates of physical activity from 9 wrist-specific predictive models and a reference hip-specific method.
Design: Prospective cohort repeated measures study.
Methods: 110 participants wore an accelerometer at wrist and hip locations for 1 week of free-living. Accelerometer data from three axes were used to calculate physical activity estimates using existing wrist-specific models (3 linear and 6 artificial neural network models) and a reference hip-specific method. Estimates of physical activity were compared to reference values at both epoch (≤60-s) and weekly levels.
Results: 9044 h were analysed. Physical activity ranged from 7 to 96 min per day of moderate-to-vigorous physical activity (MVPA). Method of analysis influenced determination of sedentary behaviour (<1.5 METs), light physical activity (1.5 to <3 METs) and MVPA (>3 METs) ($p < 0.001$, respectively). All wrist-specific models produced total weekly MVPA values that were different to the reference method. At the epoch level, Hildebrand et al. (2014) produced the strongest correlation ($r = 0.69$, 95%CI: 0.67–0.71) with tightest ratio limits of agreement (95%CI: 0.53–1.30) for MVPA, and highest agreement to predict MVPA (94.1%, 95%CI: 94.0–94.1%) with sensitivity of 63.1% (95%CI: 62.6–63.7%) and specificity of 96.0% (95%CI: 95.9–96.0%).
Conclusions: Caution is required when comparing results from studies that use inconsistent analysis methods. Although a wrist-specific linear model produced results that were most similar to the hip-specific reference method when estimating total weekly MVPA, modest absolute and relative agreement at the epoch level suggest that additional analysis methods are required to improve estimates of physical activity derived from wrist-worn accelerometers.

1. Introduction

Physical activity (PA) guidelines advise that adults should accumulate at least 150 min of moderate-intensity activity or 75 min of vigorous-intensity activity every week to achieve major health benefits. Global concern about the high proportion of people who do not meet PA guidelines is clear and considerable resource is committed to support attempts that aim to reduce this burden on societies and health care systems.

Accurate measurement is essential for surveillance of population-based PA and evaluation of both research interventions and community-based PA promotion programmes.[1] Self-report PA measures tend to overestimate time spent undertaking moderate-intensity and vigorous-intensity PA[1] and wearable devices, such as accelerometers, have emerged as a popular method of measuring PA.[2] Traditionally, the magnitude of acceleration accumulated over an epoch (counts) has been used as an independent variable in linear regression equations to estimate intensity of activity. When worn at the waist/hip, accelerometers can provide valid and reliable measurement of PA duration and intensity in a range of

* Corresponding author at: Discipline of Exercise Physiology, La Trobe Rural Health School, La Trobe University, Victoria, Australia.
E-mail address: M.Kingsley@latrobe.edu.au (M.I.C. Kingsley).

https://doi.org/10.1016/j.jsams.2018.12.003
1440-2440/

populations.[3] However, compliance has been reported to be low (e.g., ~25% of participants provided the requested 7 days of data in the National Health and Nutrition Examination Study (NHANES) during the 2003–2004 and 2005–2006 cycles) due to participant discomfort and inconvenience,[4] which can result in selection bias and misclassification of activity. Additionally, hip-worn devices are typically worn only during waking hours, which does not allow for assessment of sleep, a variable that can impact health.[5]

The adoption of wrist-worn accelerometers, including large scale surveillance studies such as NHANES and UK Biobank, allows sleep assessment[6] and can improve participant wear-time. For example, the proportion of people in NHANES who achieved ≥10 hours/day of wear-time on ≥6 days increased from 40–70% (depending upon age) in 2003–2004 to 70–80% in 2011–2012 when accelerometers were wrist-worn.[4] Although strong relationships (r = 0.88) have been reported between counts derived from hip-worn accelerometers and wrist-worn accelerometers,[7] analyses that rely on linear regression equations originally developed for hip-worn accelerometers overestimate the intensity of PA.[8,9] In attempts to address this issue, two main strategies have been employed for use with wrist-worn accelerometers: (i) to apply correction factors to cut-points established for the hip location (e.g., Actilife v7; Actigraph, USA), or (ii) to develop specific algorithms for wrist-worn accelerometers. For example, wrist-specific regression models have been developed for youth[10,11] and adult populations[11,12] to estimate the intensity of activities in metabolic equivalents (METs). With enhanced computing capability, more complex statistical methods and machine learning techniques have become popular.[12,13] These techniques use the distribution of acceleration and temporal dynamics of the signal as inputs to classify intensity of activity.

Recent results suggest that although linear equations appear suitable for hip-worn accelerometers, more complex models might be necessary when accelerometers are wrist-worn.[12–14] Findings obtained from controlled laboratory conditions are promising, where strong relationships have been reported between wrist-worn accelerometry and energy expenditure derived from a variety of measurement techniques including portable indirect calorimetry.[15] However, the performance of wrist-worn accelerometry models to estimate energy expenditure and PA behaviours in free-living conditions appears variable.[16,17]

Kerr et al.[17] evaluated the influence of wear location (wrist *vs.* hip) on accelerometer counts and average daily PA estimates obtained from free-living middle aged and older women, concluding that caution is required when comparing results obtained from devices located at different sites. Although these authors applied accelerometer location-specific cut-points, their comparison of estimated PA intensity did not include more recently developed wrist-specific algorithms, focussed on average accumulated daily PA, and recruited from a relatively homogenous group of participants. With these limitations in-mind, the effect of wear location and methods of analyses on the agreement between PA estimates obtained in free-living populations are difficult to generalise; consequently, comparing results across a range of studies is challenging, particularly when different wear locations and analyses have been applied.[11,12] Therefore, this study aimed to compare the performance of available wrist-specific accelerometer models with an established waist-specific accelerometer method to estimate activity intensities for one week in free-living healthy adults.

2. Methods

110 participants gave written informed consent to participate in the study, which was approved by La Trobe University Human Ethics Committee (ref: E15-121). Participants were free of disorders affecting energy expenditure and conditions that influence their ability to perform free-living activities. Participants (84 female, 23 male, 3 other) were adults (45 ± 11 years of age) with body mass index (27.2 ± 5.0 kg/m^2; range: 19.7–43.8 kg/m^2). Participants were instructed to wear triaxial accelerometers (GT3X+; ActiGraph LLC, USA) on a belt with the device oriented above the right hip and on the non-dominant wrist for 1 week during waking hours. Accelerometers were calibrated and synchronised to record triaxial accelerations at 100 Hz.

Raw acceleration recordings were obtained using Actilife software (version 7.0; Actilife Corp., USA). Non-wear time was determined from hip-worn acceleration data for 60-s time intervals using the Choi wear time validation algorithm.[18] Wear time was classified into sleep or wake time using the wrist-worn accelerometer and Sadeh algorithm.[19] Periods classified as non-wear and/or sleep were removed from both wrist- and hip-worn accelerometer datasets. A valid day was determined when wake wear time was ≥300 min. The derived time-domain statistics included the magnitude of vertical axis (VA), vector magnitude (VM), mean of vector magnitude (mvm), standard deviation of vector magnitude (sdvm), euclidian norm minus one (ENMO), mean angle of acceleration (mangle), sum of vector magnitudes with gravity subtracted (SVMgs), 10th, 25th, 50th, 75th, 90th percentile of x, y and z axes (e.g., X10per) and covariance between adjacent windows for x, y and z axes (XCov) (Supplementary Table 1) for four epoch durations (5-s, 15-s, 30-s and 60-s). Relevant time-domain statistics were used to determine intensity of activity using previously derived linear[11,12,20] and machine learning models.[8,21]

Reference values for activity intensities (METs) were determined from the hip worn accelerometer using a modified Freedson VM3 Combination equation (Actilife Corp., USA). Activity intensity was calculated using a threshold for sedentary activity of 150 cpm in VA[22] (1.5 METs) and the work energy theorem when activity was low with a validated equation for determining activity intensity during physical activity[23] that has been applied extensively in studies of free-living activity.[24] The combination equation to predict activity intensity was METs = 0.000863(VM) + 0.668876, when VM cpm was >2453 and 0.001092(VA) + 1.336129, when VM cpm was ≤2453. Reference activity intensities were categorise with MET thresholds[25] as follows: Sedentary Behaviour (SB: ≤1.5 METs), Light PA (LPA: 1.5 to ≤3 METs), Moderate PA (MPA: 3 to <6 METs), Vigorous PA (VPA: ≥6 METs) and Moderate to Vigorous PA (MVPA ≥3 METs). This approach was chosen for the reference comparison for activity intensity because alternative methods of continually determining energy expenditure (e.g., portable metabolic analysers) are not feasible for measurement during free-living activities and the combination equation resulted in the following recommended waist cut points: 150 cpm VA (LPA), 2690 cpm VM (MPA) and 6167 cpm VM (VPA).[22]

Nine predictive models, specifically developed for wrist-worn accelerometers, were independently used to determine activity intensity. These models included three linear regression equations that predict activity intensity, two artificial neural network (ANN) models that predict activity level (categorical outcome) and four ANN models that predict activity intensity.

Staudenmayer et al.[12] developed a linear regression to estimate METs using non-overlapping 15-s windows of triaxial dominant wrist accelerometer measurements (St-LR), where METs = 1.89378 + 5.50821(sdvm) + 0.02705(mangle).

Sirichana et al.[20] developed a piecewise linear regression equation (Si-LR) to estimate the intensity of PA using two slope relationships between METs and the summation of vector magnitudes with gravity subtracted (SVMgs), where SVMgs = 1708.1 + 373.4(METs), when MET >6 and SVMgs = 32.5 + 83.3(METs), when MET ≤6.

Hildebrand et al.[11] developed a linear regression method to estimate aspects of PA using triaxial dominant wrist accelerometer measurements using the time domain statistic ENMO, where

Table 1
Duration of weekly time (hours) of SB, LPA and MVPA during the evaluation period.

	PA category	Fr-VM3 (reference)	St-LR	Si-LR	Hb-LR	Mo-LW	Mo-RW	Mo-LW-V1V2	Mo-RW-V1V2	Mo-LW-V2	Mo-RW-V2
All (n = 110)	SB	56.9 ± 18.5	7.9 ± 11.1	79.3 ± 23.6	N/A	39.2 ± 14.4	25.8 ± 10.9	20.5 ± 9.6	9.7 ± 4.6	31.7 ± 14.2	21.4 ± 11.0
	LPA	21.0 ± 8.3	25.8 ± 10.3	2.5 ± 2.0	N/A	28.8 ± 12.1	39.2 ± 14.4	50.7 ± 15.1	59.4 ± 17.5	37.2 ± 12.1	54.1 ± 16.2
	SB + LPA	77.9 ± 23.7	33.7 ± 17.5	81.8 ± 24.2	76.6 ± 23.2	68.0 ± 20.6	65.0 ± 19.5	71.2 ± 22.1	69.0 ± 20.7	68.9 ± 21.8	75.5 ± 23.2
	MVPA	4.8 ± 2.5	49.0 ± 20.7	0.9 ± 0.8	6.2 ± 3.9	14.7 ± 5.9	17.7 ± 8.0	11.6 ± 5.8	13.7 ± 4.8	13.8 ± 6.2	7.2 ± 4.8
Male (n = 23)	SB	58.9 ± 17.3	7.9 ± 11.0	82.6 ± 22.3	N/A	37.9 ± 15.7	30.8 ± 10.4	24.2 ± 9.6	10.0 ± 5.0	37.6 ± 14.8	26.2 ± 12.1
	LPA	20.7 ± 8.1	27.7 ± 11.1	2.0 ± 1.2	N/A	32.1 ± 14.1	34.4 ± 13.8	51.9 ± 15.9	61.9 ± 15.7	37.0 ± 12.4	53.6 ± 16.1
	SB + LPA	79.6 ± 21.7	35.6 ± 17.8	84.7 ± 22.5	79.6 ± 21.8	69.9 ± 19.4	65.2 ± 16.9	76.1 ± 21.7	72.0 ± 18.9	74.5 ± 21.3	79.8 ± 22.0
	MVPA	5.8 ± 3.0	49.9 ± 23.7	0.9 ± 0.6	5.9 ± 3.0	15.5 ± 6.1	20.3 ± 9.0	9.4 ± 5.9 a	13.5 ± 4.9	11.0 ± 5.1	5.7 ± 3.3 a
Female (n = 84)	SB	55.9 ± 19.0	7.1 ± 9.7	77.7 ± 24.1	N/A	39.2 ± 14.3	23.6 ± 9.9	19.6 ± 9.4	9.6 ± 4.5	30.1 ± 13.5	20.2 ± 10.3
	LPA	20.8 ± 8.2	24.9 ± 9.8	2.5 ± 2.0	N/A	27.7 ± 11.6	40.4 ± 14.4	49.7 ± 15.0	57.9 ± 17.7	36.7 ± 11.7	53.6 ± 16.1
	SB + LPA	76.7 ± 24.4	32.0 ± 15.7	80.3 ± 24.7	75.1 ± 23.8	66.9 ± 21.1	64.0 ± 19.9	69.3 ± 22.3	67.5 ± 21.1	66.8 ± 21.8	73.8 ± 23.7
	MVPA	4.45 ± 2.3	49.2 ± 19.6	0.9 ± 0.8	6.0 ± 3.7	14.3 ± 5.7	17.1 ± 7.6	11.9 ± 5.1	13.7 ± 4.8	14.4 ± 6.1	7.4 ± 4.8
Other (n = 3)	SB	68.9 ± 5.1	30.2 ± 24.8	98.6 ± 4.2	N/A	48.7 ± 2.9	47.0 ± 8.6	18.7 ± 12.0	8.8 ± 4.0	33.7 ± 23.3	18.7 ± 17.8
	LPA	28.7 ± 8.2	37.4 ± 11.1	5.0 ± 4.4	N/A	34.7 ± 2.9	42.7 ± 11.1	67.2 ± 1.4	80.2 ± 7.7	52.2 ± 14.5	72.4 ± 11.6
	SB + LPA	97.5 ± 5.3	67.6 ± 33.7	103.6 ± 6.0	92.9 ± 5.5	83.4 ± 1.8	89.7 ± 13.9	85.9 ± 10.6	89.0 ± 3.9	85.9 ± 9.5	91.1 ± 6.6
	MVPA	7.6 ± 2.7	37.6 ± 29.5	1.6 ± 1.2	12.3 ± 10.1	21.8 ± 4.5	15.5 ± 9.5	19.3 ± 15.3	16.2 ± 2.4	19.2 ± 12.6	14.1 ± 10.4
Normal (n = 38) BMI 18.5 to <25	SB	56.1 ± 17.1	9.6 ± 11.0	78.8 ± 22.4	N/A	37.7 ± 15.5	22.0 ± 9.5	18.0 ± 8.8	9.0 ± 4.6	27.4 ± 12.8	18.4 ± 11.0
	LPA	21.3 ± 8.9	26.4 ± 11.3	2.7 ± 2.1	N/A	30.9 ± 13.7	43.3 ± 14.1	50.3 ± 14.0	59.4 ± 17.1	38.8 ± 11.3	55.2 ± 15.5
	SB + LPA	77.4 ± 22.7	36.0 ± 18.8	81.5 ± 23.1	75.4 ± 21.9	68.6 ± 19.8	65.4 ± 17.9	68.3 ± 20.6	68.4 ± 19.9	66.2 ± 20.6	73.7 ± 22.7
	MVPA	5.0 ± 2.4	46.4 ± 22.5	1.0 ± 0.8	7.1 ± 4.1	13.8 ± 5.7	17.0 ± 9.2	14.1 ± 6.6	14.1 ± 5.0	16.2 ± 6.7	8.8 ± 4.7
Overweight (n = 44) BMI 25 to <30	SB	53.7 ± 19.8	5.5 ± 9.1	74.7 ± 24.6	N/A	37.1 ± 13.6	25.5 ± 11.1	19.8 ± 10.5	9.9 ± 5.0	32.4 ± 15.6	21.4 ± 11.7
	LPA	19.7 ± 7.1	22.9 ± 9.4	2.3 ± 1.6	N/A	26.2 ± 10.0	35.4 ± 13.1	48.0 ± 15.1	55.5 ± 17.6	33.9 ± 10.9	50.8 ± 14.8
	SB + LPA	73.4 ± 24.4	28.4 ± 13.9	77.1 ± 24.9	72.5 ± 24.2	63.3 ± 20.5	60.9 ± 19.9	67.8 ± 23.9	65.4 ± 21.6	66.3 ± 23.4	72.1 ± 24.2
	MVPA	4.6 ± 2.5	49.6 ± 19.9	0.9 ± 0.6	5.5 ± 3.0	14.6 ± 6.3	17.1 ± 7.0	10.2 ± 4.5	12.6 ± 4.2	11.7 ± 4.7	5.8 ± 3.1
Obese (n = 24) BMI ≥30	SB	61.1 ± 16.4	6.2 ± 8.1	85.3 ± 20.8	N/A	42.0 ± 12.8	31.0 ± 9.7	25.2 ± 8.4	10.0 ± 3.9	37.7 ± 11.5	26.7 ± 10.1
	LPA	22.7 ± 8.9	28.1 ± 9.5	2.2 ± 1.6	N/A	29.6 ± 11.9	37.6 ± 13.3	53.5 ± 15.2	63.4 ± 15.0	38.0 ± 12.8	56.0 ± 16.6
	SB + LPA	83.8 ± 21.1	34.3 ± 14.0	87.6 ± 21.6	83.0 ± 20.2	71.5 ± 19.2	68.6 ± 16.8	78.7 ± 18.9	73.4 ± 17.7	75.8 ± 18.8	82.7 ± 20.4
	MVPA	4.6 ± 2.5	54.1 ± 18.6	0.9 ± 0.7	5.4 ± 3.1	16.8 ± 6.0	19.8 ± 7.4	9.7 ± 3.9	14.9 ± 5.1	12.6 ± 4.7	5.7 ± 2.6

Data are mean ± SD (hours). N/A: Hb-LR model was developed considering both SB and LPA together as a single category (SB + LPA). PA: physical activity. SB: sedentary behaviour. LPA: light physical activity. MVPA: moderate-to-vigorous physical activity.

METs = [0.0320(ENMO) + 7.28]/3.5. Data were calculated in 1-s epochs and averaged in minute intervals (Hb-LR).

Montoye et al.[8] developed ANN models to measure activity intensity in 3-class categories (SB, LPA, MVPA) for accelerometers worn on left and right wrists (Mo-LW and Mo-RW, respectively). The 10th, 25th, 50th, 75th and 90th percentile of raw accelerometer readings from three axes were derived in 30-s epochs.

Montoye et al.[21] developed ANN models to predict energy expenditure in METs from wrist-worn accelerometer data. ANN models were trained using structured 5-min activities (V1) and simulated free-living activities (V2). Data were integrated in 30-s time windows to derive 10th, 25th, 50th, 75th and 90th percentiles, and covariance between adjacent windows for all accelerometer axes. Four models were used, two ANNs trained with V1V2 for left and right wrists (Mo-LW-V1V2 and Mo-RW-V1V2, respectively), and two ANNs trained with V2 for left and right wrists (Mo-LW-V2 and Mo-RW-V2, respectively). Trained ANN models were downloaded as described in the original publications.[8,21]

Group values are expressed as mean and standard deviation. Repeated measures analyses of covariance (ANCOVA) were performed using IBM SPSS Statistics for Windows (Version 24; IBM Corporation, USA) to compare models for weekly SB, LPA, SB + LPA and MVPA using wake time as a covariate. Significant main effects were explored by pairwise-comparisons between each wrist-worn model and the reference method with Bonferroni correction. Agreement between the predictive models from wrist-worn accelerometers (three linear regression models and six ANN models) and reference values from accelerometers worn at the hip were determined for four epoch durations (5-s, 15-s, 30-s and 60-s). Relative agreement was determined using Pearson correlation coefficient for ratio data and Spearman's rank correlation for ordinal data. Absolute agreement was assessed with log-transformed data replicated in pairs to determine 95% ratio limits of agreement (RLOA).[26] Categorical agreement, sensitivity and specificity of models to determine SB, LPA and MVPA were calculated using confusion matrices and Clopper–Pearson confidence intervals.

3. Results

A total of 9044 h of wake time was accumulated from valid wear days (~10 billion data points) with an average of 10.5 ± 2.8 h/day analysed. Reference activity during the week was 56.9 ± 18.5 h (69 ± 8%) of SB, 20.0 ± 8.3 h (25 ± 6%) of LPA and 4.9 ± 2.5 h (6 ± 3%) of MVPA.

Table 1 presents weekly time in SB, LPA, SB + LPA and MVPA. Significant main effects existed for all categories of PA ($F_{(9,972)} \geq 49.2$, $p < 0.001$). Hb-LR was closest to the reference method for MVPA (MD: 1.3 h, 95%CI: 0.4–2.2 h) and SB + LPA (MD: −1.3 h 95%CI: −0.4 to −2.2 h). Mo-LW was closest to the reference method for SB (MD: −17.7 h 95%CI: −15.4 to −20.0 h). Neither gender nor BMI influenced patterns in these comparisons (Table 1).

Relative and absolute agreement between wrist-models and reference hip method were not influenced by epoch duration (5-s to 60-s; data not presented). Agreement between wrist-models and reference hip method are presented in Table 2, where results reflect the epoch duration used to develop each wrist-model. Overall correlations between wrist-models and the hip method for the epoch that resulted in the strongest association produced medium to large correlations ($r = 0.30–0.69$, $p < 0.001$), with Hb-LR producing the largest correlation.

Hb-LR displayed highest categorical agreement to predict MVPA at the epoch level (94.1%, 95%CI: 94.0–94.1%), which was associated with sensitivity of 63.2% (95%CI: 62.6–63.7%) and specificity of 96.0% (95%CI: 95.9–96.0%). None of the other models resulted in categorical agreement, sensitivity and specificity values that were all above 50% (Table 2).

Table 2
Relative agreement (bivariate correlations), absolute agreement (ratio limits of agreement), sensitivity, specificity and categorical agreement (agreement) of predictive models compared to reference at the epoch level.

	St-LR (15-s)	Si-LR (60-s)	Hb-LR (60-s)	Mo-LW (30-s)+	Mo-RW (30-s)+	Mo-LW-V1V2 (30-s)	Mo-RW-V1V2 (30-s)	Mo-LW-V2 (30-s)	Mo-RW-V2 (30-s)
SB									
Sensitivity % (95%CI)	11.7–11.8	99.2–99.2	N/A	50.4–50.6	33.4–33.6	26.5–26.7	12.5–12.6	40.9–41.1	27.8–27.9
Specificity % (95%CI)	95.2–95.3	11.3–11.6	N/A	88.7–89.1	95.5–95.7	95.5–95.8	98.2–98.3	91.2–91.6	96.5–96.7
LPA									
Correlation r (95%CI)	0.14–0.16	0.17–0.22	0.52–0.55**	N/A	N/A	0.09–0.11	0.17–0.19	0.13–0.16	0.08–0.10
RLOA (95%CI)	0.79–3.62	0.34–0.95	1.20–2.19	N/A	N/A	0.6–2.57	0.68–2.68	0.5–2.96	0.57–2.26
Sensitivity % (95%CI)	21.6–21.8	59.1–61.6	95.9–96.0	68.1–69.6	54.8–56.4	63.2–64.8	57.0–58.6	51.5–53.1	73.7–75.2
Specificity % (95%CI)	65.6–65.7	98.0–98.1	62.6–63.7	65.6–65.8	54.6–52.8	38.7–38.9	28.0–28.2	55.0–55.2	34.6–34.8
MVPA									
Correlation r (95%CI)	0.17–0.26	0.00–0.07	0.45–0.53*	N/A	N/A	0.06–0.15	0.07–0.14	0.03–0.12	0.07–0.16
RLOA (95%CI)	0.51–1.77	0.12–0.93	0.53–1.30	N/A	N/A	0.31–1.38	0.36–1.61	0.27–1.65	0.28–1.27
Sensitivity % (95%CI)	85.3–85.7	11.2–11.9	62.6–63.7	31.7–32.4	54.7–55.4	32.4–33.1	46.6–47.4	39.4–40.1	23.7–24.3
Specificity % (95%CI)	42.3–42.4	99.5–99.6	95.9–96.0	83.1–83.3	80.9–81.0	87.3–87.4	85.5–85.7	84.9–85.1	92.3–92.4
Overall									
Correlation r (95%CI)	0.32–0.36**	0.22–0.28*	0.67–0.71**	0.18–0.21*	0.22–0.25**	0.29–0.33**	0.36–0.39**	0.33–0.36**	0.27–0.31**
RLOA (95%CI)	0.78–4.28	0.37–1.17	0.99–2.42	N/A	N/A	0.57–2.73	0.69–2.78	0.47–2.79	0.55–2.52
Agreement % (95%CI)	18.5–18.6	70.3–70.5	94.0–94.1	49.5–49.6	35.1–35.3	27.4–27.6	15.4–15.5	40.9–41.1	28.1–28.3

SB: sedentary behaviour. LPA: light physical activity. MVPA: moderate-to-vigorous physical activity. RLOA: ratio limits of agreement. 95%CI: 95% confidence interval. N/A: results not available; Hb-LR model was developed considering both SB and LPA as one category; +data are ordinal therefore correlations, where possible, represent Spearman's and RLOA calculations are not possible.

* Indicates that lower limit of 95%CI is $p < 0.05$.

** Indicates that lower limit of 95%CI is $p < 0.001$.

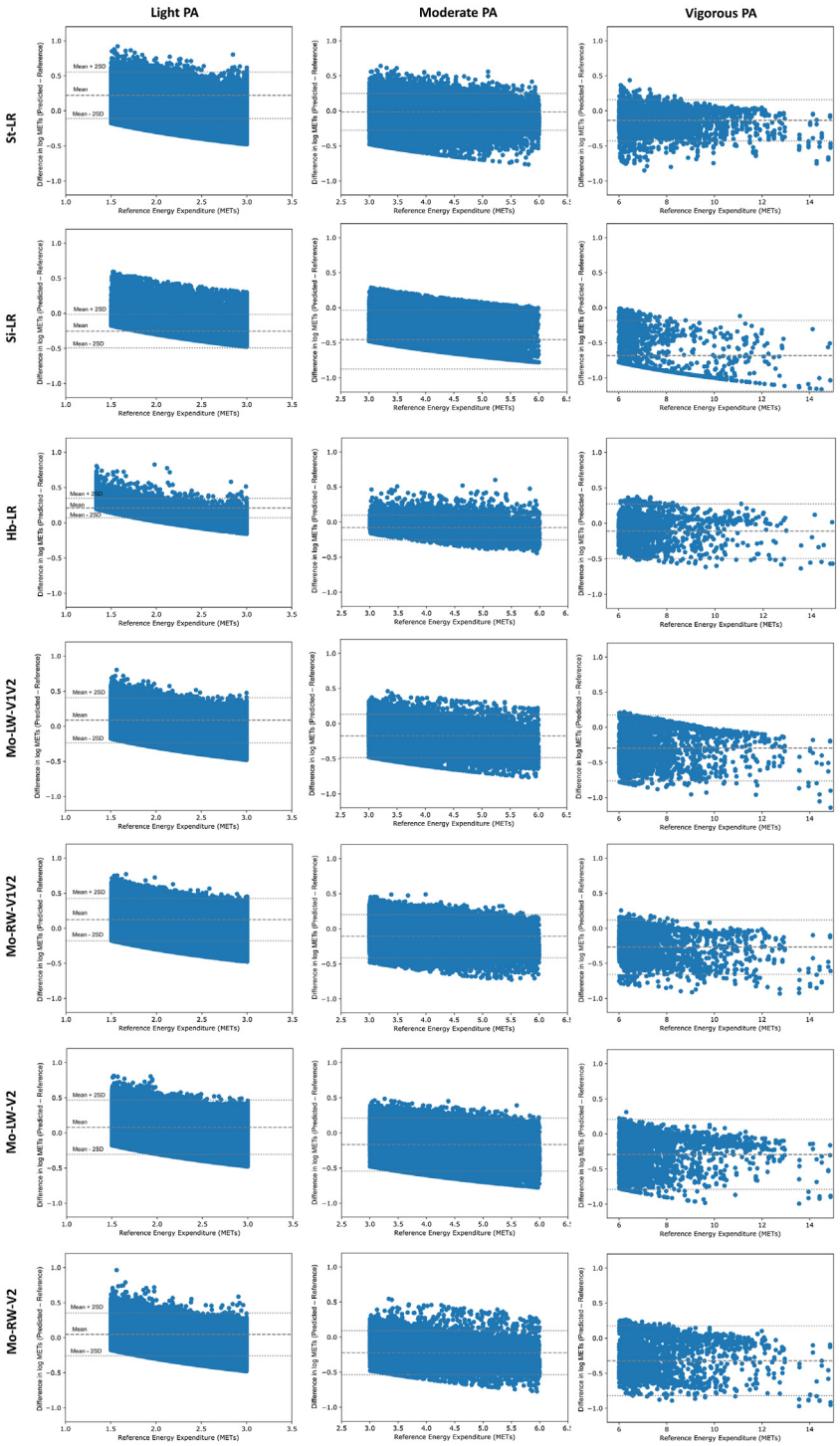

Fig. 1. Bland–Altman plots demonstrating agreement between predictive models and reference for activity intensity.

Fig. 1 illustrates differences in log transformed activity intensity between wrist models and reference hip method by PA category. Absolute agreement for weekly MVPA was tightest for Hb-LR with 95%RLOA being 0.53–1.30 (Table 2). Absolute agreement for weekly LPA was tightest for Si-LR with 95%RLOA being 0.34–0.95 (Table 2).

4. Discussion

We compared the agreement between nine wrist-specific accelerometry predictive models with a reference hip-specific method to determine activity intensities and categories of activity intensity. When estimating weekly PA, Hb-LR produced the most similar results for total MVPA and SB + LPA when compared to the reference method. At the epoch level (5-s to 60-s), Hb-LR produced MET values with the strongest relative agreement, tightest RLOA, and with categorical agreement, sensitivity and specificity values greater than 60% for intensities classified by the reference method to be above 3 METs.

Substantial differences existed in weekly MVPA, LPA and SB determined by the models under investigation. These findings support the conclusions of previous researchers that caution is required when comparing PA estimates from studies that

employ different accelerometer wear location and processing procedures.[17] In the free-living environment, all of wrist-specific models, including more complex machine learning models, produced higher estimated total weekly time of MVPA when compared to the reference hip-specific method. Hb-LR produced total weekly MVPA that were most similar to the reference values. Importantly these findings refer to participants across a wide range of PA patterns, including individuals who achieved from 7 min/day to 96 min/day of MVPA according to the reference method.

Although gender and level of obesity might influence PA patterns,[27] results from the current study demonstrate that neither of these factors influenced the relative or absolute agreement between wrist-specific models and the reference hip-specific method to determine MVPA, LPA or SB. Previous findings that PA behaviours in middle-age and older women vary greatly according to data processing techniques and device placement,[17] can be extended to include PA intensities determined for males as well as individuals classified as normal, overweight or obese.

This is the first study to compare the performance of a range of wrist-specific activity intensity prediction models to a hip-specific method with recommended accelerometer cut-off values[22] at the epoch level in free-living participants (≤60-s). These findings show that, of the current freely available analyses models, Hb-LR resulted in the strongest agreement to the hip-specific reference method with categorical agreement, sensitivity and specificity above 60% for activity intensities greater than 3 METs. This finding is important for a number of reasons. Firstly, most current PA guidelines refer to the requirement to accumulate PA in sustained bouts of greater than or equal to 10 min. Although it is possible to argue that this requirement lacks consistent support,[27] our findings suggest that Hb-LR provides the most comparable method to the hip-specific reference method when evaluating the accumulation of sustained MVPA bouts in the free-living environment. Also, as identified by many authors[28] and acknowledged in most PA guidelines, additional health benefits are afforded by undertaking activity above 3 METs. In this respect, Hb-LR was the analysis method that provided the strongest relative and absolute agreement with the reference hip-specific method for free-living MVPA.

Although Hb-LR was the most comparable wrist-specific model to the reference hip-specific method when quantifying SB + LPA and MVPA, with a floor effect of 2.08 METs this model cannot be used to differentiate between SB and LPA. Therefore, Hb-LR model can only be used for assessment of MVPA and not for SB, which can impact health outcomes independently of MVPA.[25] The potential to use wrist-worn accelerometry to accurately determine LPA and SB in free-living people is important because evidence supports the independent association between excessive SB with mortality[29] and adverse health outcomes,[30] and higher LPA is associated with lower mortality risk in women with low MVPA.[31] Although practical advantages of wrist-worn accelerometry has meant that this wear site has gained popularity in large scale surveillance studies, our findings suggest that wrist-specific predictive models can lead to large variations in the classification of PA compared to the reference hip-specific method. In order to allow comparability between studies, future work should explore new analysis models to improve the agreement between results obtained from wrist-worn and hip-worn accelerometers so that more consistent analysis methods can be adopted with confidence.

In accordance with the aim of this study, identical wrist-worn raw accelerometer data collected from the ActiGraph GT3X+ accelerometer were used to directly compare the performance of predictive models even though some of the models were originally developed using different accelerometer devices (e.g., GENEActiv, Axivity) and sampling rates. While raw data should theoretically be comparable across device manufacturers, small differences exist between data derived from ActiGraph and GENEActiv accelerometers,[32] while Axivity and GENEActiv produce near perfect agreement in raw acceleration data.[32] Nevertheless, good agreement (>85% agreement when classifying activity intensity) has been reported when raw data from different accelerometer brands are used in a predictive model.[32] Therefore, while some of the predictive models in the present study were not developed for ActiGraph data, accelerometer choice is likely to have had minimal impact on performance of these models. Given the wide RLOA for all models tested, further advances in modelling approaches are warranted for wrist-worn accelerometers to improve PA intensity prediction.

Participants in this study provide representation across gender groups, adult-aged groupings (young adult, middle aged and older adults) and BMI (normal BMI, overweight and obese). These volunteers were recruited from one regional city location and the individuals were not representative of the population as a whole because the sample included greater representation from female and other genders. Even though neither gender nor BMI groupings influenced the comparisons between wrist-specific models with the reference method, their activity values might not be generalizable. In support of this statement, the average total MVPA was 38 ± 19 min/day, which demonstrates that the group were more active than people of similar age in the NHANES study.[27]

The current study did not include a gold standard method for determining activity intensity at the epoch level (i.e., indirect calorimetry) in order to establish accuracy of the wrist-specific models. While this is a limitation of the study, current indirect calorimetry techniques are not feasible for extended surveillance in free-living environments. Furthermore, camera-based monitoring systems are only suitable for classifying activity behaviours[16] rather than activity intensity. Therefore, we chose to compare available wrist-specific models to a reference hip-specific accelerometry method that has been validated in a range of populations. It should, however, be acknowledged that our choice of reference method might influence conclusions drawn from this study.

5. Conclusions

Large differences can exist in estimates of activity intensity and total weekly MVPA, LPA and SB derived from accelerometers placed at different locations and using different analysis techniques. Of the nine wrist-specific prediction models investigated, a linear regression model (Hb-LR) produced estimates of PA that were most comparable with the reference hip-specific method for activities at or above moderate intensity (3 METs). As more studies are published with wrist-worn accelerometers, it is important to understand that differences in findings might reflect variations in analysis methods rather than differences in activity patterns between populations and interventions.

Practical implications

- Choice of accelerometer wear location and model of analysis can cause large variations in estimates of total weekly MVPA, LPA and SB.
- All wrist-specific models produced higher estimated total weekly MVPA when compared to the reference hip-specific method.
- Hb-LR produced the closest results to the reference hip-specific method for total weekly MVPA.
- At the epoch level (60-s), Hb-LR was able to classify MVPA with categorical agreement, sensitivity and specificity greater than 60%.

Acknowledgements

The authors would like to thank Mr Ian Bailie and Mr Craig Staunton for their assistance with data collection. The work was supported by a VicHealth grant and a La Trobe University Rehabilitation Sport, Exercise and Rehabilitation Research Focus Area Small Innovation Projects grant scheme.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jsams.2018.12.003.

References

1. Hallal PC, Andersen LB, Bull FC et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* 2012; 380(9838):247–257.
2. Montoye AH, Moore RW, Bowles HR et al. Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. *Br J Sports Med* 2016; 52(23):1507–1516.
3. Butte NF, Ekelund U, Westerterp KR. Assessing physical activity using wearable monitors: measures of physical activity. *Med Sci Sports Exerc* 2012; 44(1 Suppl 1):S5–12.
4. Troiano RP, McClain JJ, Brychta RJ et al. Evolution of accelerometer methods for physical activity research. *Br J Sports Med* 2014; 48(13):1019–1023.
5. Kanagasabai T, Chaput JP. Sleep duration and the associated cardiometabolic risk scores in adults. *Sleep Health* 2017; 3(3):195–203.
6. Slater JA, Botsis T, Walsh J et al. Assessing sleep using hip and wrist actigraphy. *Sleep Biol Rhythms* 2015; 13(2):172–180.
7. Dieu O, Mikulovic J, Fardy PS et al. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. *Clin Physiol Funct Imaging* 2017; 37(5):525–529.
8. Montoye AHK, Pivarnik JM, Mudd LM et al. Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. *AIMS Public Health* 2016; 3(2):298–312.
9. Rosenberger ME, Haskell WL, Albinali F et al. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc* 2013; 45(5):964–975.
10. Crouter SE, Kuffel E, Haas JD et al. Refined two-regression model for the actigraph accelerometer. *Med Sci Sports Exerc* 2010; 42(5):1029–1037.
11. Hildebrand M, Van Hees VT, Hansen BH et al. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc* 2014; 46(9):1816–1824.
12. Staudenmayer J, He S, Hickey A et al. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol (1985)* 2015; 119(4):396–403.
13. Montoye AH, Pivarnik JM, Mudd LM et al. Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiol Meas* 2016; 37(10):1770–1784.
14. White T, Westgate K, Wareham NJ et al. Estimation of physical activity energy expenditure during free-living from wrist accelerometry in uk adults. *PLoS One* 2016; 11(12):e0167472.
15. Ellingson LD, Hibbing PR, Kim Y et al. Lab-based validation of different data processing methods for wrist-worn actigraph accelerometers in young adults. *Physiol Meas* 2017; 38(6):1045–1060.
16. Ellis K, Kerr J, Godbole S et al. Hip and wrist accelerometer algorithms for free-living behavior classification. *Med Sci Sports Exerc* 2016; 48(5):933–940.
17. Kerr J, Marinac CR, Ellis K et al. Comparison of accelerometry methods for estimating physical activity. *Med Sci Sports Exerc* 2017; 49(3):617–624.
18. Choi L, Liu Z, Matthews CE et al. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc* 2011; 43(2):357–364.
19. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep* 1994; 17(3):201–207.
20. Sirichana W, Dolezal BA, Neufeld EV et al. Wrist-worn triaxial accelerometry predicts the energy expenditure of non-vigorous daily physical activities. *J Sci Med Sport* 2017; 20(8):761–765.
21. Montoye AHK, Conger SA, Connolly CP et al. Validation of accelerometer-based energy expenditure prediction models in structured and simulated free-living settings. *Meas Phys Educ Exerc Sci* 2017; 21(4):223–234.
22. Migueles JH, Cadenas-Sanchez C, Ekelund U et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports Med* 2017; 47(9):1821–1845.
23. Sasaki JE, John D, Freedson PS. Validation and comparison of actigraph activity monitors. *J Sci Med Sport* 2011; 14(5):411–416.
24. Brown MA, Howatson G, Quin E et al. Energy intake and energy expenditure of pre-professional female contemporary dancers. *PLoS One* 2017; 12(2):e0171998.
25. Tremblay MS, Aubert S, Barnes JD et al. Sedentary behavior research network (sbrn) – terminology consensus project process and outcome. *Int J Behav Nutr Phys Act* 2017; 14(1):75.
26. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8(2):135–160.
27. Tucker JM, Welk GJ, Beyler NK et al. Associations between physical activity and metabolic syndrome: comparison between self-report and accelerometry. *Am J Health Promot* 2016; 30(3):155–162.
28. Atienza AA, Moser RP, Perna F et al. Self-reported and objectively measured activity related to biomarkers using nhanes. *Med Sci Sports Exerc* 2011; 43(5):815–821.
29. Thorp AA, Owen N, Neuhaus M et al. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996–2011. *Am J Prev Med* 2011; 41(2):207–215.
30. Powell C, Herring MP, Dowd KP et al. The cross-sectional associations between objectively measured sedentary time and cardiometabolic health markers in adults – a systematic review with meta-analysis component. *Obes Rev* 2018; 19(3):381–395.
31. Borgundvaag E, Janssen I. Objectively measured physical activity and mortality risk among american adults. *Am J Prev Med* 2017; 52(1):e25–e31.
32. Rowlands AV, Mirkes EM, Yates T et al. Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? *Med Sci Sports Exerc* 2018; 50(2):257–265.