Gait & Posture 61 (2018) 98–110

Contents lists available at ScienceDirect

Gait & Posture

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

Review

Calibration of raw accelerometer data to measure physical activity: A systematic review

Márcio de Almeida Mendes[a,*], Inácio C.M. da Silva[a], Virgílio V. Ramires[a], Felipe F. Reichert[b], Rafaela C. Martins[a], Elaine Tomasi[a,*]

[a] Post-Graduate Program in Epidemiology, Federal University of Pelotas, Brazil
[b] School of Physical Education, Federal University of Pelotas, Brazil

ARTICLE INFO

Keywords:
Physical activity
Accelerometry
Calibration study
Raw accelerometer data

ABSTRACT

Most of calibration studies based on accelerometry were developed using count-based analyses. In contrast, calibration studies based on raw acceleration signals are relatively recent and their evidences are incipient. The aim of the current study was to systematically review the literature in order to summarize methodological characteristics and results from raw data calibration studies. The review was conducted up to May 2017 using four databases: PubMed, Scopus, SPORTDiscus and Web of Science. Methodological quality of the included studies was evaluated using the Landis and Koch's guidelines. Initially, 1669 titles were identified and, after assessing titles, abstracts and full-articles, 20 studies were included. All studies were conducted in high-income countries, most of them with relatively small samples and specific population groups. Physical activity protocols were different among studies and the indirect calorimetry was the criterion measure mostly used. High mean values of sensitivity, specificity and accuracy from the intensity thresholds of cut-point-based studies were observed (93.7%, 91.9% and 95.8%, respectively). The most frequent statistical approach applied was machine learning-based modelling, in which the mean coefficient of determination was 0.70 to predict physical activity energy expenditure. Regarding the recognition of physical activity types, the mean values of accuracy for sedentary, household and locomotive activities were 82.9%, 55.4% and 89.7%, respectively. In conclusion, considering the construct of physical activity that each approach assesses, linear regression, machine-learning and cut-point-based approaches presented promising validity parameters.

1. Introduction

Questionnaires have historically been the main physical activity measurement instrument in epidemiological studies. However, accelerometers are currently a feasible alternative to objectively measure physical activity. Accelerometers are portable devices, which measure the acceleration from body movements in one, two or three axes: vertical (Y), horizontal right-left (X) and horizontal front-back axis (Z) [1].

The use of accelerometers entails advantages and disadvantages, as any other method of measurements of physical activity. Regarding the disadvantages, accelerometers do not indicate the context and the purpose of the physical activities. Furthermore, accelerometers are not valid to measure specific physical activities such as isometric activities, physical activities against a resistance force (e.g. strength exercises) and cycling [2]. However, the major source of overall physical activity energy expenditure (PAEE) is derived form dynamic physical activities (e.g. walking, running), and these activities are accurately measured by accelerometers [2]. Data from accelerometers are also free of information bias introduced by interviewers or participants. The data are gathered by the devices at the exact moment in which the activities are taking place, providing a reliable physical activity measure in free-living conditions [2,3].

An important challenge regarding the use of accelerometers to measure physical activity lies in the interpretation of the signals provided by the devices, which need to be translated into measurements with biological and/or behavioral meaning. In this context, several calibration studies have been performed [4]. There are important methodological differences in calibration studies (e.g. sample sizes and characteristics, physical activity protocols and statistical approaches), which might influence the results of such studies. Accordingly, the accelerometer signal, which is one of the main variables analysed in these studies is not the same across studies. Some studies, notoriously the most recent ones, have analysed the signal as a gravitational equivalent (g, where $1g = 9.81\ \mathrm{m\,s^{-2}}$), whilst other analysed it as

* Correspondence to: Rua Marechal Deodoro, 1160, CEP: 96020-220, Pelotas, Brazil.
E-mail addresses: marciopenha.esef@yahoo.com.br (M. de Almeida Mendes), inacio_cms@yahoo.com.br (I.C.M. da Silva), virgilioramires@hotmail.com (V.V. Ramires), ffreichert@gmail.com (F.F. Reichert), rafamartins1@gmail.com (R.C. Martins), tomasiet@gmail.com (E. Tomasi).

https://doi.org/10.1016/j.gaitpost.2017.12.028
Received 27 July 2017; Received in revised form 17 December 2017; Accepted 29 December 2017
0966-6362/

counts. Direct comparison between count values from different accelerometer brands is limited, mainly because manufacturers use different undisclosed algorithms to define the acceleration outputs [5]. In contrast, analyses based on gravitational equivalent (raw data) are performed using open source packages and, therefore, provides more transparency and better comparability across studies.

More recently accelerometer data can be analysed using different acceleration signals, thus, a high number of calibration studies based on raw data have been carried out. In this context, it is crucial to understand how the raw signal from accelerometers has been translated into physical activity measures. Thus, the aim of this study was to systematically review the literature in order to summarize the methods and results from calibration studies based on raw accelerometer data to measure physical activity.

2. Methods

The systematic review was conducted up to May 30th 2017 using PubMed, Scopus, SPORTDiscus and Web of Science databases. The following terms were searched in abstracts and titles: [("motor activity" OR "physical activity" OR "physical fitness" OR "physical exercise") AND ("accelerometry" OR "accelerometer" OR "motion sensor") AND ("calibration" OR "cut-off" OR "cut-point" OR "threshold" OR "validity" OR "validation") AND ("raw acceleration" OR "raw data")].

Only articles assessing raw acceleration signal and including healthy people were included. Articles in which the calibration was a secondary purpose were also eligible. Studies identifying only sedentary behavior thresholds were not considered.

Initially, all identified titles and abstracts were read by the first and third authors. In case of disagreement between them, all eligibility criteria were discussed. Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines were followed to conduct and describe all methodological process and reported results [6]. The following data were extracted from each study included in the review: place of publication, sample, accelerometer placement, accelerometer model, sample frequency and epoch (interval in which acceleration signals are summarized), activity protocol used for calibration, criterion measure, statistical approach applied, physical activity intensity thresholds and prediction equations of PAEE. Finally, the methodological quality of the included studies was evaluated using the Landis and Koch's guidelines [7].

3. Results

In total, 1669 articles were identified (23 references in PubMed, 19 in Scopus, 1590 in SPORTDiscus and 37 in Web of Science). After checking duplicate studies, 45 references were excluded and 1624 titles were considered eligible for reading. From the 1624, 180 titles were kept. After evaluating the abstracts, 79 articles remained. All 79 articles were read and 17 studies were considered eligible. The reference list from selected studies was checked and three articles were added, resulting in a total of 20 studies [8–27] (Fig. 1).

A detailed description of all studies included is available in Table 1. Studies were published between 1994 and 2017 and most of them were carried out in the United States of America (eight) and the remaining in a European country. Both sexes were included in all studies and there was no information regarding body composition and physical fitness in the studied samples. Eight studies were performed with children/adolescents and adults, eight exclusively with adults and three with children/adolescents. Regarding the studies with adults, only four of them presented age range greater than four decades. Studies with adolescents included participants from six to eighteen years old (Table 1).

The number of physical activities included in the protocols ranged from two to 23 and included a broad spectrum of intensity. Walking and running, as main components of PAEE, were included in 16 studies. Only four studies included physical activities practiced outside the laboratory setting (see Supplementary Table S1 in the online version at DOI:10.1016/j.gaitpost.2017.12.028).

Table 1 indicates that most studies (15) placed the accelerometer on the waist of the participants, but it was also tested on other parts of the body (low back, wrist, foot, chest, waist, thigh and ankle).

Actigraph GT3X (nine studies) and GENEActiv (five studies) were the main accelerometers used. Other accelerometers used were: ICSensors 3031–010, 7164 Actigraph, IDEEA, GENEA, Tracmor, DynaPort, Hookie AM13, Hookie AM20, GulfCoast X6-1A, MotionLogs and MICA2DOT (Table 1). Except for the 7164 Actigraph (1 axis), MICA2DOT and IDEEA (two axes), all other accelerometers collected body movements into three axes. The sampling frequency (number of measurements in each axis per second) varied from 10 to 100 Hz, and epoch lengths were analysed as one, five, six, 30 and 60 s (Table 1).

Regarding the studies in which the three-dimensional raw data were transformed into a single-dimensional signal vector magnitude (SVM) of acceleration, this conversion was performed using different metrics and the equation $SVM = \Sigma\sqrt{x^2+y^2 + z^2}$ is the most common metric adopted. Indirect calorimetry was the most widely used criterion measure (13 studies) (Table 1).

Among the studies using cut-point-based statistical approach, the mean values and standard deviation (± SD) of sensitivity, specificity and accuracy from the intensity thresholds were: 93.7% (± 7.0), 91.9% (± 9.6) and 95.8% (± 0.1), respectively. Values of sensitivity, specificity or accuracy were similar according to the different intensity thresholds and accelerometer placements. The values of accuracy ranged from 84% to 100% (Table 2).

Five studies used regression models to estimate PAEE. Four of these studies presented predictive equations, in which the mean value of coefficient of determination (R^2) was 0.79 (± 0.12) (Table 3).

Machine learning-based modelling to estimate PAEE and to recognize physical activity types was the calibration statistical approach most frequently applied (11 studies). Most of these studies used Artificial Neural Network technique to create their predictive models (seven studies). Among the predictive models for PAEE, the mean value of R^2 was 0.70 (± 0.11). Regarding predictive models for recognition of physical activity types, the mean values of accuracy for sedentary (e.g. lying down, sitting, standing), household (washing dishes, folding towels and stacking them nearby, vacuuming carpet) and locomotive (walking, cycling, running) activities were 82.9% (± 20.2), 55.4% (± 26.6) and 89.7% (± 11.2), respectively (Table 4).

4. Discussion

All articles found in this review were conducted in high-income countries and most of them had relatively small samples and specific population groups, with low variability in terms of individual characteristics. The sample composition from accelerometry calibration studies hinders the extrapolation of their results to other settings [4]. Greater heterogeneity is required regarding the characteristics such as age, body mass index and physical fitness. Future calibration studies using more representative samples of their target populations in terms of demographic and physiological characteristics would represent an important step forward.

Distinct physical activity protocols were applied in the included calibration studies and were summarized in the present review. Protocols including the whole spectrum of physical activity intensities (sedentary pursuits, low, moderate and vigorous activities) were identified and the number of physical activities performed varied across studies. It is important to highlight that the number of activities tested shall not affect the internal validity of the intensity thresholds or algorithms. Therefore, high accuracy in the prediction of PAEE might be found even in studies assessing few activities. In contrast, a low number of activities tested or the inclusion of activities that are rarely performed in free-living conditions by the target population could impair

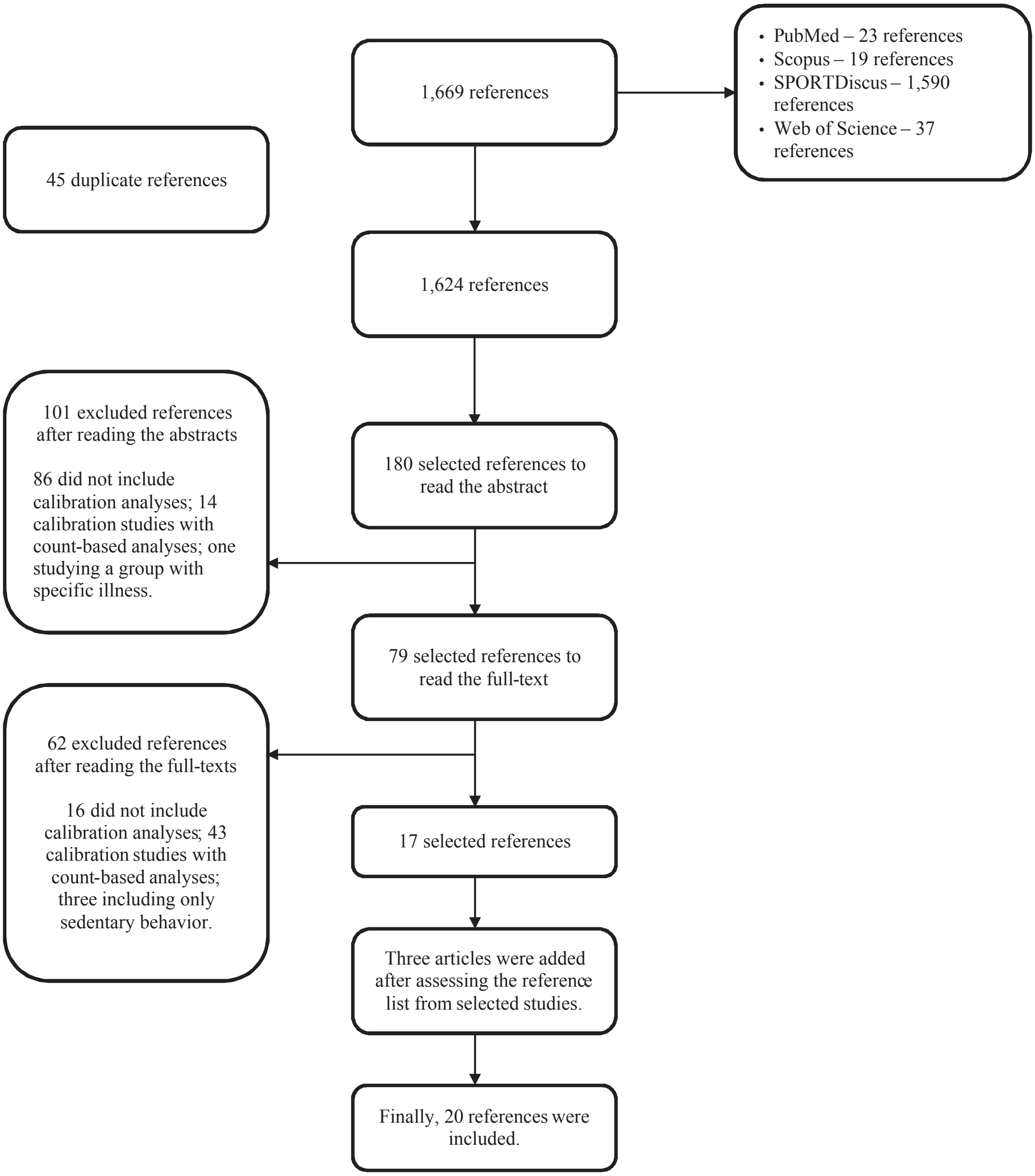

Fig. 1. Description of the search strategy process and results.

the generalization and application of such thresholds/algorithms to other settings [2,28].

Considering that accelerometer-based physical activity assessment is mostly implemented in epidemiological studies interested in habitual physical activity, protocols from accelerometry calibration studies should be as close as possible to free-living activities. However, physical activity protocols applied in controlled conditions (both indoor and outdoor) have been implementing a small increase in intensities and movement patterns [11,29]. In addition, in most studies, participants perform each activity for a reduced period of time, which probably does not accurately reflect daily physical activity patterns. As a result, the application of these intensity thresholds or algorithms developed under controlled conditions may lead to inaccurate estimates of physical activity in free-living [29]. Our results showed that some studies collected physical activity data outside the laboratory (semi-free-living) [10–12,20], however none of them in real-life conditions.

One of the main decisions to be made to design an accelerometer-based study is the accelerometer placement. A study conducted by Rosenberguer et al. [30], analyzing count-based acceleration signals, provided evidence that hip-worn accelerometer presents sedentary activity and physical activity measures more accurate than wrist-worn accelerometer. However, studies based on raw accelerometer data have shown different results. Wrist-worn accelerometers are able to provide accurate estimates regarding the activity type, PAEE, activity intensity similar to hip-worn devices [31]. In our review, which is restricted to raw data accelerometry, waist and wrist were the most frequently studied placements, and there was no evidence of any difference in the accuracy between them. Furthermore, other evidences have shown that raw acceleration signal based on placements such as thigh-worn accelerometers also provided accurate estimates of PAEE [21] and physical activity types [18,25].

In addition to validity parameters of physical activity measures, other aspects such as adherence to protocol (wear compliance) and research purpose (e.g. comparability between studies or previous data

Table 1
- Description of the methods and results from calibration studies based on raw accelerometer data.

Author(s); year; country	Sample	Brand; axes; sampling frequency; epoch; placement	Definition of signal vector magnitude of acceleration (SVM)	Criterion measure	Statistical approach
Bouten et al.; 1994; Netherlands [8].	11 subjects aged 21–27 years.	ICSensors 3031–010; 3; 100 Hz; 30 s; low back.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$	Indirect calorimetry	Regression model
Rothney et al.; 2007; United States [9].	102 subjects aged 18–70 years.	7164 Actigraph and IDEEA; 1 and 2; 30 Hz; 60 s; waist, soles of the feet, thighs and upper sternum.	NA	Indirect calorimetry	Machine-learning algorithm
Esliger et al.; 2011; England [10].	60 subjects aged 40–65 years.	GENEA; 3; 80 Hz; 1 s; left wrist, right wrist and waist.	$SVM = \Sigma\left\|\sqrt{x^2+y^2+z^2} - g\right\|$	Indirect calorimetry	ROC curve
Gyllensten et al.; 2011; Netherlands [11]	20 subjects aged 22–51 years.	Tracmor and IDEEA; 3 and 2; 20 and 30 Hz; waist, soles of the feet, thighs and upper sternum.	NA	Acceleration signal pattern during each activity performed	Machine-learning algorithm
Zhang et al.; 2011; England [12].	60 subjects aged 40–65 years.	GENEA; 3; 80 Hz; left wrist, right wrist and waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$	Indirect calorimetry	Machine-learning algorithm
Brandes et al.; 2012; Germany [13].	186 subjects aged 6 or more years.	DynaPort; 3; 100 Hz; low back.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$	Indirect calorimetry	Regression model
Phillips et al.; 2013; England [14].	44 subjects aged 8–14 years.	GENEA; 3; 80 Hz; 1 s; left wrist, right wrist and waist.	$SVM = \Sigma\left\|\sqrt{x^2+y^2+z^2} - g\right\|$	Indirect calorimetry	ROC curve
Bing He et al.; 2014; United States [15].	60 subjects (Mean age ± standard deviate = 80.6 ± 4.8)	Actigraph GT3X; 3; 80 Hz; left wrist, right wrist and waist.	NA	Acceleration signal pattern during each activity performed	Machine-learning algorithm
Hildebrand et al.; 2014; Norway [16].	30 subjects aged 7–65 years.	- GENEActiv and Actigraph GT3X; 3; 60 Hz; 1 s; non-domi-nant wrist and waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2} - g$	Indirect calorimetry	Regression model
Schaefer et al.; 2014; United States [17].	24 subjects aged 6–11 years.	GENEActiv; 3; 75 Hz; 1 s; non-dominant wrist.	$SVM = \sum_{i=1}^{f}\left\|\sqrt{x^2+y^2+z^2}\right\|/(f)$	Indirect calorimetry	ROC curve
Skotte et al.; 2014; Denmark [18].	17 subjects (Mean age ± standard deviate = 34 ± 11)	Actigraph GT3X; 3; 30 Hz; right thigh and waist.	NA	Acceleration signal pattern during each activity performed	Machine-learning algorithm
Aittasalo et al.; 2015; Finland [19].	20 subjects aged 13–15 years.	Actigraph GT3X and Hookie AM13; 3; 30 and 100 Hz; 5 s; waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$ $SVMmean = \frac{1}{N}\sum_{i=j}^{j+N-1} SVM$ $MAD = \frac{1}{N}\sum_{i=j}^{j+N-1}\left\|SVM - SVMmean\right\|$	Arbitrary category of physical activity intensity	Regression model
Bastian et al.; 2015; France [20].	59 subjects aged 18–55 years.	MotionLogs; 3; 100 Hz; waist.	NA	Acceleration signal pattern during each activity performed	Machine-learning algorithm
Montoye et al.; 2015; United States [21].	44 subjects aged 18–44 years.	GENEActiv and Actigraph GT3X; 3; 20 and 40 Hz; waist, left wrist, right wrist and right thigh.	NA	Indirect calorimetry	Machine-learning algorithm
Vähä-Ypyä et al.; 2015; Finland [22].	29 subjects (Mean age ± standard deviate = 35 ± 11).	Hookie AM20; 3; 100 Hz; 6 s; waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$ $SVMmean = \frac{1}{N}\sum_{i=j}^{j+N-1} SVM$ $MAD = \frac{1}{N}\sum_{i=j}^{j+N-1}\left\|SVM - SVMmean\right\|$	Indirect calorimetry	Regression model and ROC curve
Vähä-Ypyä et al.; 2015; Finland [23].	21 subjects (Mean age ± standard deviate = 42 ± 11).	Actigraph GT3X, GulfCoast X6-1A and Hookie AM13; 3; 30, 20 and100 Hz; 5 s; waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$ $SVMmean = \frac{1}{N}\sum_{i=j}^{j+N-1} SVM$ $MAD = \frac{1}{N}\sum_{i=j}^{j+N-1}\left\|SVM - SVMmean\right\|$	Arbitrary category of physical activity intensity	ROC curve
Montoye et al.; 2016; United States [24].	34 subjects aged 18–30 years.	Actigraph GT3X and MICA2DOT; 3 and 2; 30 and 10 Hz; waist, right wrist, right thigh and right ankle.	NA	Indirect calorimetry	Machine-learning algorithm

(continued on next page)

Table 1 (*continued*)

Author(s); year; country	Sample	Brand; axes; sampling frequency; epoch; placement	Definition of signal vector magnitude of acceleration (SVM)	Criterion measure	Statistical approach
Montoye et al.; 2016; United States [25].	44 subjects aged 18–44 years.	GENEActiv and Actigraph GT3X; 3; 20 and 40 Hz; left wrist, right wrist and right thigh.	NA	Indirect calorimetry	Machine-learning algorithm
Montoye et al.; 2016; United States [26].	44 subjects aged 18–44 years.	GENEActiv; 3; 20 Hz; left and right wrists.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$	Indirect calorimetry	Machine-learning algorithm
Rosenberg et al.; 2017; United States [27].	39 subjects aged 55–96 years.	Actigraph GT3X; 3; 40 Hz; 60 s; waist.	$SVM = \Sigma\sqrt{x^2+y^2+z^2}$	Global positioning systems (GPS)	Machine-learning algorithm

y, x and z – Correspond to axes: vertical, horizontal right-left and horizontal front-back, respectively.
g – Gravitational equivalent unit, where 1 g = 9.81 m s^{-2}.
f – Sampling frequency.
SVMmean – Mean resultant for each analysed epoch.
MAD – Mean amplitude deviation.
i – Corresponds to each measurement point with acceleration signals in x, y and z axes.
N – Number of samples in the epoch.
j – Starting point of the epoch.
NA – Not applicable. It was used for studies in which the three-dimensional raw data (from x, y, and z axes) were not transformed into a single-dimensional SVM.

collections) should be considered to choose the accelerometer placement. NHANES (National Health and Nutrition Examination Survey), a large-scale study in the United States of America, collected data from hip-attached accelerometers in the first wave and changed to wrist placement thereafter. The decision was basically attributed to a greater compliance found to wrist-worn accelerometer in real life situations compared to waist-worn [31].

Further discussion on accelerometer axes, sampling frequency, epoch and the SVM calculation is necessary. Different axes and sampling frequency criteria were adopted by the studies included in our review. However, according to Zhang et al. [32], these differences are not determinant for the accuracy of the classification of physical activities. This result is important, since lower sampling rate combined with fewer measurement axes allow for longer periods of use of the monitor with a single battery charge and also reduces the processing time required to analyse the accelerometer data. Furthermore, by reducing the data size, it is also possible to reduce the accelerometer chips responsible for storing the measurements, allowing the monitors to be smaller and with more space to add other functions (e.g. watch face on accelerometer). These extra functions may improve the compliance and the quality of the physical activity measurements [32].

In contrast, potential implications in terms of epoch and SVM choices must be highlighted. Calibration studies based on cut-point approaches usually address the total volume of acceleration signals converted to SVM (under count or raw acceleration signals). As the SVM of acceleration is not changed according to the choice of epoch length, the identified thresholds and their accuracy results are also not influenced by epoch criteria. However, if the objective is to apply these intensity thresholds to obtain the amount of time spent in different levels of intensity (from data collected in free-living), the epoch length may be determinant. For instance, if the intensity varies between moderate/vigorous and sedentary activities within an epoch length of 60 s, it will only reflect the average activity level during this period and, eventually, short bursts of moderate to vigorous physical activity will not be identified. Therefore, shorter epochs (e.g. one or five seconds) are recommended for sensitive assessment of intermittent patterns of physical activity (especially vigorous physical activity, in all age groups) [33]. Among the studies based on machine learning techniques, the discussion about epoch length is less relevant due to specific time windows usually adopted and due to the fact that the main information to be analysed is the movement pattern, instead of the magnitude of acceleration signals summarized within each epoch [31].

Another important topic is the potential advantage that has been attributed to raw data accelerometry, which is expected to generate better comparability between the results from different accelerometer brands. Although small differences might still occur between data collection from different brands, higher comparability will only be achieved if data reduction is performed in a standardized manner. There is evidence presenting high comparability between Actigraph GT3X and GENEActiv, when the same analytical approach (same metric) is performed [34,35]. However, without a standardized data filtering process among studies, as observed in the current literature, the main limitation from count-based analyses remains for raw acceleration signal-based studies.

Due to the mentioned variability in the data reduction and different SVM calculations among the included studies, it was not possible to perform direct comparisons between the cut-points presented. This limitation in terms of comparability has also been highlighted elsewhere [36] and these issues have been consistently described as relevant definitions influencing calibration studies and their threshold proposals [34,35].

Another limitation to be highlighted is the choice of the criterion measure among studies using cut-point-based approaches. Most studies – epidemiological or laboratory-based – addressing physical activity and its health benefits, classify the activities performed using absolute intensity categories [37]. Similarly, most of the included studies used

Table 2
Intensity thresholds and their validity parameters.

Study	Intensity	Thresholds	Sensitivity (%)	Specificity (%)	Accuracy ($CI_{95\%}$)
Esliger et al. [10]	Left wrist (adults)				
	Sedentary (≤1.5 METs)	< 217 g min^{-1}	97	95	0.98 (0.98–0.99)
	Light (1.6–3.9 METs)	217–644 g min^{-1}	NA	NA	NA
	Moderate (4.0–6.9 METs)	645–1810 g min^{-1}	95	72	0.91 (0.88–0.93)
	Vigorous (≥7 METs)	> 1810 g min^{-1}	78	98	0.91 (0.86–0.95)
	Right wrist (adults)				
	Sedentary	< 386 g min^{-1}	99	96	0.98 (0.97–0.99)
	Light	386–439 g min^{-1}	NA	NA	NA
	Moderate	440–2098 g min^{-1}	100	56	0.84 (0.81–0.87)
	Vigorous	> 2098 g min^{-1}	78	97	0.89 (0.84–0.94)
	Waist (adults)				
	Sedentary	< 77 g min^{-1}	99	96	0.97 (0.96–0.98)
	Light	77–219 g min^{-1}	NA	NA	NA
	Moderate	220–2056 g min^{-1}	96	80	0.93 (0.91–0.95)
	Vigorous	> 2056 g min^{-1}	73	99	0.92 (0.88–0.96)
Phillips et al. [14]	Left wrist (children)				
	Sedentary (≤1.5 METs)	< 6 g	94.7	96.7	0.97 (0.96–0.99)
	Light (1.6–2.9 METs)	6–21 g	NA	NA	NA
	Moderate (3–5.9 METs)	22–56 g	88.1	84	0.93 (0.90–0.95)
	Vigorous (≥6)	> 56 g	91.3	89.2	0.94 (0.92–0.97)
	Right wrist (children)				
	Sedentary	< 7 g	94.9	97.7	0.97 (0.95–0.99)
	Light	7–19 g	NA	NA	NA
	Moderate	20–60 g	82.4	83.3	0.92 (0.89–0.94)
	Vigorous	> 60 g	89.4	85.5	0.93 (0.91–0.96)
	Waist (children)				
	Sedentary	< 3 g	96	96.1	0.99 (0.97–0.99)
	Light	3–16 g	NA	NA	NA
	Moderate	17–51 g	88.5	88	0.95 (0.93–0.97)
	Vigorous	> 51 g	92	88.9	0.94 (0.91–0.96)
Hildebrand et al. [16]	AG – waist (adults)				
	Sedentary (NP)	NP	NP	NP	NP
	Light (< 3 METs)	< 69.1 mg	NP	NP	NP
	Moderate (3–5.9 METs)	69.1–258.7 mg	NP	NP	NP
	Vigorous (≥6)	> 258.7 mg	NP	NP	NP
	GA – waist (adults)				
	Sedentary	NP	NP	NP	NP
	Light	< 68.7 mg	NP	NP	NP
	Moderate	68.7–266.8 mg	NP	NP	NP
	Vigorous	> 266.8 mg	NP	NP	NP
	AG – non-dominant wrist (adults)				
	Sedentary	NP	NP	NP	NP
	Light	100.6 mg	NP	NP	NP
	Moderate	100.6–428.8 mg	NP	NP	NP
	Vigorous	> 428.8 mg	NP	NP	NP
	GA – non-dominant wrist (adults)				
	Sedentary	NP	NP	NP	NP
	Light	< 93.2 mg	NP	NP	NP
	Moderate	93.2–418.3 mg;	NP	NP	NP
	Vigorous	> 418,3 mg	NP	NP	NP
Hildebrand et al. [16]	AG – waist (children)				
	Sedentary	NP	NP	NP	NP
	Light	< 142.6 mg	NP	NP	NP
	Moderate	142.6–464.6 mg	NP	NP	NP
	Vigorous	> 464.6 mg	NP	NP	NP
	GA – waist (children)				
	Sedentary	NP	NP	NP	NP
	Light	< 152.8 mg	NP	NP	NP
	Moderate	152.8–514.3 mg	NP	NP	NP
	Vigorous	> 514.3 mg	NP	NP	NP
	AG – non-dominant wrist (children)				
	Sedentary	NP	NP	NP	NP
	Light	< 201.4 mg	NP	NP	NP
	Moderate	201.4–707.0 mg	NP	NP	NP
	Vigorous	> 707.0 mg	NP	NP	NP

(continued on next page)

Table 2 (*continued*)

Study	Intensity	Thresholds	Sensitivity (%)	Specificity (%)	Accuracy ($CI_{95\%}$)
	GA – non-dominant wrist (children)				
	Sedentary	NP	NP	NP	NP
	Light	< 191.6 mg	NP	NP	NP
	Moderate	191.6–695.8 mg	NP	NP	NP
	Vigorous	> 695.8 mg	NP	NP	NP
Schaefer et al. [17]	Non-dominant wrist (children)				
	Sedentary (≤1.5 METs)	< 0.190 g	97	87.6	0.96 (NP)
	Light (1.6–2.9 METs)	0.190–0.313 g	NA	NA	NA
	Moderate (3–5.9 METs)	0.314–0.997 g	91	87.3	0.95 (NP)
	Vigorous (≥6)	> 0.997 g	94.9	85.3	0.94 (NP)
Aittasalo et al. [19]	AG – Waist (adolescents)				
	Sedentary	< 26.9 mg	95	100	0.98 (NP)
	Light	26.9–331.9 mg	NA	NA	NA
	Moderate	332.0–558.2 mg	96.7	98.6	0.98 (NP)
	Vigorous	> 558.2 mg	100	100	1.00 (NP)
	Hookie – Waist (adolescents)				
	Sedentary	< 28.7 mg	96	99	0.98 (NP)
	Light	28.7–337.9 mg	NA	NA	NA
	Moderate	338.0–603.7 mg	98.3	99.3	0.99 (NP)
	Vigorous	> 603.7 mg	100	100	1.00 (NP)
Vähä-Ypyä et al. [22]	Waist (adults)				
	Sedentary (NP)	NP	NP	NP	NP
	Light (< 3 METs)	91 mg	NA	NA	NA
	Moderate (3–5.9 METs)	91,1–414 mg	100	96	0.97 (0.95–0.99)
	Vigorous (≥6)	> 414 mg	96	95	1.00 (0.99–1.00)
Vähä-Ypyä et al. [23]	Waist (adults)				
	Threshold 1	16.7 mg	98.7	99.7	1.00 (NP)
	Threshold 2	157.4 mg	100	100	1.00 (NP)
	Threshold 3	331.2 mg	98.9	96.9	1.00 (NP)
	Threshold 4	599.3 mg	98.3	98.8	1.00 (NP)

GA – GENE
GA – GENEActiv; AG – Actigraph GT3X;
NP – corresponds to information not provided by the study;
NA – not applicable because the sedentary and moderate thresholds provide the limits for the light-intensity category;
In Aittasalo et al. [19] study, the intensities are defined as sedentary (lying supine on a bed, sitting on a chair, sitting while working on a computer, standing, standing while moving light, 1 kg, weights on a table surface), light (slow and normal walking), moderate (brisk walking) e vigorous (jogging and running);
In Vähä-Ypyä et al. [23] study, threshold 1 separates the intensity category 0 (lying supine on a bed, sitting on a chair, sitting while working on a computer, standing, standing while moving light, 1 kg, weights on a table surface) of the category 1 (slow walking), threshold 2 separates the category 1 of the category 2 (normal walking), threshold 3 separates the category 2 of the category 3 (brisk walking) and threshold 4 separates the category 4 (jogging and running) of the category 3.

absolute intensity thresholds as criterion measure. The use of absolute intensity thresholds as reference measure might cause significant misclassification due to individual characteristics [37]. Another point to be highlighted is that even though calibration studies have showed high values of sensitivity, specificity and accuracy, if the criterion measure is inaccurate, an important misclassification will always be present.

Different statistical approaches were applied in the studies included in this review. Linear regression is one of the most widely used method found in the literature, allowing to identify intensity thresholds and predict PAEE adjusted for variables such as age and body weight, among other characteristics. Another approach to identify physical activity intensity thresholds is the Receive Operating Characteristics (ROC) curve analysis. This statistical procedure allows minimizing false positives and false negatives [4], which might be considered an advantage, enabling the researcher to select the intensity thresholds with the highest values of sensitivity and/or specificity [10].

Both statistical approaches to calibration also present limitations. The development of cut-point-based approaches for accelerometer data is applicable in specific research contexts, limiting direct comparisons among physical activity results obtained from different cut-points. Furthermore, considering the diversity of intensity thresholds proposed in the literature, this approach allows the researcher to obtain different results according to the choice of a particular cut-point [31].

Despite these limitations, the cut-points-based approach is still often used to estimate time spent on different physical activity intensities. In this context, it is important that the researcher takes into account the methodological procedures when choosing the cut-points (e.g. data filtering process, accelerometer placement and sample characteristics).

Machine learning-based modelling is a more recent and sophisticated statistical approach, which has been adopted in calibration studies to assess PAEE and physical activity patterns. Artificial Neural Network was the most frequent machine leaning technique applied among the included studies. This technique is a flexible and robust approach, which allows estimating both continuous (PAEE) and categorical variables (physical activity types) [38]. Other techniques such as Decision Tree, Support Vector Machine, Naive Bayes, among others, were also observed in this review, but there is no consensus in the literature on which machine learning is the most feasible technique. For example, Zhang et al. [12], tested different machine learning techniques and the highest accuracy was shown by the Support Vector Machine and Decision Tree. However, considering both efficiency (time spent for classification) and accuracy, the Decision Tree approach was considered the most advantageous compared to other approaches. In the set of results assessed in this review, the accuracy of machine learning-based algorithms for pattern recognition of physical activities was dependent on the type of activity; higher accuracy was observed for locomotive and sedentary activities compared to other household activities.

Overall, promising validity parameters were identified in most studies, considering distinct statistical method applied. For example,

Table 3
Regression equations developed for the prediction of physical activity energy expenditure (PAEE).

Study	Equation	Coefficient of determination
Bouten et al. [8]	Using the integral of the absolute value accelerometer output in antero-posterior direction (IAAx)	$R^2 = 0.94$
	EEact = −0.176 + 0.085 · IAAx	
	Using the sum of the integrals of absolute accelerometer output from three orthogonal measurement directions (IAAtot)	$R^2 = 0.90$
	EEact = 0.104 + 0.023 · IAAtot	
Hildebrand et al. [16]	Adults	
	AG – hip	$R^2 = 0.81$
	VO2 = 0.0554 mg + 6.67	
	GA – hip	$R^2 = 0.79$
	VO2 = 0.0530 mg + 6.86	
	AG – wrist	$R^2 = 0.75$
	VO2 = 0.0320 mg + 7.28	
	GA – wrist	$R^2 = 0.76$
	VO2 = 0.0323 mg + 7.49	
	Children	
	AG – hip	$R^2 = 0.78$
	VO2 = 0.0559 mg + 10.03	
	GA – hip	$R^2 = 0.75$
	VO2 = 0.0498 mg + 10.39	
	AG – wrist	$R^2 = 0.71$
	VO2 = 0.0356 mg + 10.83	
	GA – wrist	$R^2 = 0.72$
	VO2 = 0.0357 mg + 11.16	
Brandes et al. [13]	AEE = Acc + gender	$R^2 = 0.59$
	AEE = Acc + age	$R^2 = 0.57$
Vähä-Ypyä et al. [22]	Direct relationship between the incident mean amplitude deviation (MAD) and VO2	$R^2 = 0.92$
	$VO2 = 10.015\ e^{0.0017 MAD(mg)}$	
	Walking only	$R^2 = 0.88$
	VO2 = 7.920 + 0.0331 MAD (mg)	
	All measured values	$R^2 = 0.96$
	$VO2 = 2.351^{e(0.00177\ MAD\ (mg) - 0.282\ Vmax\ (ms^{-1}) + 0.0183\ VO2peak\ (ml\ kg^{-1}\ min^{-1}) + 0.0117\ height\ (cm) - 0.0142\ weight\ (kg) + 0.00693\ waist\ circumference\ (cm) - 0.00211\ age\ (years))}$	

VO_2 – Oxygen uptake expressed under milliliters per kilogram per minute ($ml\,kg^{-1}\,min^{-1}$).
Vmax – Maximum oxygen uptake.
VO_2peak – Peak oxygen uptake.
EEact – Energy expenditure due to physical activity expressed in watts per kilogram ($W\,kg^{-1}$).
AEE – Energy expenditure due to physical activity expressed in joules per minute per kilogram ($J\,min^{-1}\,kg^{-1}$).
Acc – Acceleration.

Table 4
Quality indicators from machine learning-based predict models.

Study	Machine learning technique adopted	Quality indicators
Prediction of PAEE		
Rothney et al. [9]	Artificial Neural Network – waist	R^2 = 0.85
Montoye et al. [21]	Artificial Neural Network – right wrist	R^2 = 0.74
	Artificial Neural Network – left wrist	R^2 = 0.72
	Artificial Neural Network – thigh	R^2 = 0.81
	Artificial Neural Network – waist	R^2 = 0.74
Montoye et al. [24]	Artificial Neural Network – thigh, right and left wrists combined	R^2 = 0.62
	Artificial Neural Network – waist	R^2 = 0.52
Montoye et al. [25]	Artificial Neural Network – right wrist (using accelerometer data from right wrist)	R^2 = 0.66
	Artificial Neural Network – right wrist (using accelerometer data from left wrist)	R^2 = 0.58
Prediction of physical activity types		
Gyllensten et al. [11]	Decision Tree – waist, soles of the feet, thighs and upper sternum combined	
	Overall activities	Accuracy (%) = 92.9
	Artificial Neural Network – waist, soles of the feet, thighs and upper sternum combined	
	Overall activities	Accuracy (%) = 94.2
	Support Vector Machines – waist, soles of the feet, thighs and upper sternum combined	
	Overall activities	Accuracy (%) = 95.4
	Majority voting – waist, soles of the feet, thighs and upper sternum combined	
	Overall activities	Accuracy (%) = 95.9
Zhang et al. [12]	Decision Tree – right wrist	
	Sedentary activities	Accuracy (%) = 98.1
	Household	Accuracy (%) = 92.9
	Walking	Accuracy (%) = 97.8
	Running	Accuracy (%) = 98.7
	Overall activities	Accuracy (%) = 97.0
	Decision Tree – left wrist	
	Sedentary activities	Accuracy (%) = 98.5
	Household	Accuracy (%) = 91.0
	Walking	Accuracy (%) = 95.9
	Running	Accuracy (%) = 100
	Overall activities	Accuracy (%) = 95.9
	Decision Tree – waist	
	Sedentary activities	Accuracy (%) = 99.4
	Household	Accuracy (%) = 96.6
	Walking	Accuracy (%) = 100
	Running	Accuracy (%) = 100
	Overall activities	Accuracy (%) = 99.1
Zhang et al. [12]	Naive Bayes – right wrist	
	Overall activities	Accuracy (%) = 95.3
	Naive Bayes – left wrist	
	Overall activities	Accuracy (%) = 95.3
	Naive Bayes – waist	
	Overall activities	Accuracy (%) = 98.2
	Support Vector Machine – right wrist	
	Overall activities	Accuracy (%) = 96.8
	Support Vector Machine – left wrist	
	Overall activities	Accuracy (%) = 96.4
	Support Vector Machine – waist	
	Overall activities	Accuracy (%) = 99.3
	Artificial Neural Network – right wrist	
	Overall activities	Accuracy (%) = 96.8
	Artificial Neural Network – left wrist	
	Overall activities	Accuracy (%) = 95.9
	Artificial Neural Network – waist	
	Overall activities	Accuracy (%) = 99.6
Bing He et al. [15]	Single-accelerometer Movelet – right wrist	
	Lying down	Accuracy (%) = 97.1
	Standing	Accuracy (%) = 95.3
	Washing dishes	Accuracy (%) = 28.7
	Kneading a ball of dough	Accuracy (%) = 55.2
	Putting jacket on	Accuracy (%) = 6.0
	Folding towels and stacking them nearby	Accuracy (%) = 13.7
	Vacuuming carpet	Accuracy (%) = 67.0
	Simulated shopping	Accuracy (%) = 38.0
	Writing	Accuracy (%) = 98.9
	Dealing cards	Accuracy (%) = 62.2
	Standing up from a chair and sitting back down	Accuracy (%) = 94.3
	Normal walking with arm swing	Accuracy (%) = 98.9

(continued on next page)

Table 4 (*continued*)

Study	Machine learning technique adopted	Quality indicators
	Normal walking without arm swing	Accuracy (%) = 99.2
	Brisk walking with arm swing	Accuracy (%) = 95.9
	Brisk walking without arm swing	Accuracy (%) = 97.9
	Single-accelerometer Movelet – left wrist	
	Lying down	Accuracy (%) = 98.1
	Standing	Accuracy (%) = 96.3
	Washing dishes	Accuracy (%) = 15.2
	Kneading a ball of dough	Accuracy (%) = 51.1
	Putting jacket on	Accuracy (%) = 11.0
	Folding towels and stacking them nearby	Accuracy (%) = 11.4
	Vacuuming carpet	Accuracy (%) = 84.1
	Simulated shopping	Accuracy (%) = 24.4
	Writing	Accuracy (%) = 97.8
	Dealing cards	Accuracy (%) = 95.9
Bing He et al. [15]	Standing up from a chair and sitting back down	Accuracy (%) = 98.8
	Normal walking with arm swing	Accuracy (%) = 98.9
	Normal walking without arm swing	Accuracy (%) = 98.5
	Brisk walking with arm swing	Accuracy (%) = 97.7
	Brisk walking without arm swing	Accuracy (%) = 98.7
	Single-accelerometer Movelet – waist	
	Lying down	Accuracy (%) = 98.9
	Standing	Accuracy (%) = 96.1
	Washing dishes	Accuracy (%) = 63.4
	Kneading a ball of dough	Accuracy (%) = 64.3
	Putting jacket on	Accuracy (%) = 41.1
	Folding towels and stacking them nearby	Accuracy (%) = 34.9
	Vacuuming carpet	Accuracy (%) = 52.3
	Simulated shopping	Accuracy (%) = 24.7
	Writing	Accuracy (%) = 98.9
	Dealing cards	Accuracy (%) = 52.8
	Standing up from a chair and sitting back down	Accuracy (%) = 88.1
	Normal walking with arm swing	Accuracy (%) = 93.5
	Normal walking without arm swing	Accuracy (%) = 83.5
	Brisk walking with arm swing	Accuracy (%) = 85.0
	Brisk walking without arm swing	Accuracy (%) = 81.6
	Expanded Movelets – waist, right and left wrists combined	
	Lying down	Accuracy (%) = 100
	Standing	Accuracy (%) = 100
	Washing dishes	Accuracy (%) = 39.7
	Kneading a ball of dough	Accuracy (%) = 64.1
	Putting jacket on	Accuracy (%) = 0.3
	Folding towels and stacking them nearby	Accuracy (%) = 13.7
	Vacuuming carpet	Accuracy (%) = 84.1
	Simulated shopping	Accuracy (%) = 32.8
	Writing	Accuracy (%) = 100
	Dealing cards	Accuracy (%) = 100
	Standing up from a chair and sitting back down	Accuracy (%) = 98.4
	Normal walking with arm swing	Accuracy (%) = 99.2
	Normal walking without arm swing	Accuracy (%) = 99.1
	Brisk walking with arm swing	Accuracy (%) = 96.7
	Brisk walking without arm swing	Accuracy (%) = 98.9
	Movelets Voting – waist, right and left wrists combined	
	Lying down	Accuracy (%) = 99.3
	Standing	Accuracy (%) = 97.8
	Washing dishes	Accuracy (%) = 64.5
	Kneading a ball of dough	Accuracy (%) = 77.4
	Putting jacket on	Accuracy (%) = 33.8
	Folding towels and stacking them nearby	Accuracy (%) = 29.9
	Vacuuming carpet	Accuracy (%) = 87.2
	Simulated shopping	Accuracy (%) = 40.5
	Writing	Accuracy (%) = 97.5
Bing He et al. [15]	Dealing cards	Accuracy (%) = 80.3
	Standing up from a chair and sitting back down	Accuracy (%) = 98.0
	Normal walking with arm swing	Accuracy (%) = 98.5
	Normal walking without arm swing	Accuracy (%) = 98.2
	Brisk walking with arm swing	Accuracy (%) = 98.2
	Brisk walking without arm swing	Accuracy (%) = 97.3
	Movelets Decision Tree – waist, right and left wrists combined	
	Lying down	Accuracy (%) = 100
	Standing	Accuracy (%) = 98.1
	Washing dishes	Accuracy (%) = 53.6

(*continued on next page*)

Table 4 (*continued*)

Study	Machine learning technique adopted	Quality indicators
	Kneading a ball of dough	Accuracy (%) = 63.8
	Putting jacket on	Accuracy (%) = 28.8
	Folding towels and stacking them nearby	Accuracy (%) = 30.8
	Vacuuming carpet	Accuracy (%) = 58.0
	Simulated shopping	Accuracy (%) = 28.4
	Writing	Accuracy (%) = 99.6
	Dealing cards	Accuracy (%) = 33.8
	Standing up from a chair and sitting back down	Accuracy (%) = 55.8
	Normal walking with arm swing	Accuracy (%) = 97.6
	Normal walking without arm swing	Accuracy (%) = 93.2
	Brisk walking with arm swing	Accuracy (%) = 84.8
	Brisk walking without arm swing	Accuracy (%) = 96.3
Skotte et al. [18]	Decision Tree – waist	
	Standing	Accuracy (%) = 100
	Sitting	Accuracy (%) = 100
	Walking	Accuracy (%) = 99.6
	Stair climbing and descending	Accuracy (%) = 97.7
	Cycling	Accuracy (%) = 100
	Running	Accuracy (%) = 99.3
Bastian et al. [20]	Bayesian – waist	
	Lying	Accuracy (%) = 93.4
	Slouching	Accuracy (%) = 85.2
	Sitting	Accuracy (%) = 52.7
	Standing	Accuracy (%) = 81.3
	Small utilitarian movements	Accuracy (%) = 79.4
	Walking	Accuracy (%) = 95.4
	Cycling	Accuracy (%) = 66.3
	Running	Accuracy (%) = 93.5
Montoye et al. [26]	Artificial Neural Network – right wrist	
	Lying down	Accuracy (%) = 81.1
	Reading	Accuracy (%) = 59.9
	Using computer	Accuracy (%) = 85.9
	Standing	Accuracy (%) = 90.0
	Folding towels and putting them in a laundry basket	Accuracy (%) = 79.3
	Sweeping	Accuracy (%) = 77.5
	Flexing and extending the biceps with a light load	Accuracy (%) = 90.8
Montoye et al. [26]	Slow walking	Accuracy (%) = 72.2
	Squats	Accuracy (%) = 87.9
	Brisk walking	Accuracy (%) = 69.4
	Stair climbing and descending	Accuracy (%) = 74.4
	Cycling	Accuracy (%) = 84.1
	Jogging	Accuracy (%) = 96.1
	Overall activities	Accuracy (%) = 81.1
	Artificial Neural Network – left wrist	
	Lying down	Accuracy (%) = 76.2
	Reading	Accuracy (%) = 66.9
	Using computer	Accuracy (%) = 78.7
	Standing	Accuracy (%) = 90.2
	Folding towels and putting them in a laundry basket	Accuracy (%) = 83.8
	Sweeping	Accuracy (%) = 79.5
	Flexing and extending the biceps with a light load	Accuracy (%) = 87.5
	Slow walking	Accuracy (%) = 71.5
	Squats	Accuracy (%) = 89.8
	Brisk walking	Accuracy (%) = 69.9
	Stair climbing and descending	Accuracy (%) = 73.3
	Cycling	Accuracy (%) = 87.2
	Jogging	Accuracy (%) = 95.3
	Overall activities	Accuracy (%) = 80.9
	Artificial Neural Network – thigh	
	Lying down	Accuracy (%) = 53.9
	Reading	Accuracy (%) = 38.4
	Using computer	Accuracy (%) = 54.6
	Standing	Accuracy (%) = 69.6
	Folding towels and putting them in a laundry basket	Accuracy (%) = 59.3
	Sweeping	Accuracy (%) = 72.7
	Flexing and extending the biceps with a light load	Accuracy (%) = 53.4
	Slow walking	Accuracy (%) = 82.2
	Squats	Accuracy (%) = 88.9
	Brisk walking	Accuracy (%) = 71.9
	Stair climbing and descending	Accuracy (%) = 89.1
	Cycling	Accuracy (%) = 66.0
	Jogging	Accuracy (%) = 93.1

(*continued on next page*)

Table 4 (*continued*)

Study	Machine learning technique adopted	Quality indicators
	Overall activities	Accuracy (%) = 71.4
	Artificial Neural Network – waist	
	Lying down	Accuracy (%) = 90.8
	Reading	Accuracy (%) = 36.5
	Using computer	Accuracy (%) = 39.2
	Standing	Accuracy (%) = 56.9
	Folding towels and putting them in a laundry basket	Accuracy (%) = 50.6
	Sweeping	Accuracy (%) = 60.0
	Flexing and extending the biceps with a light load	Accuracy (%) = 44.8
	Slow walking	Accuracy (%) = 66.9
	Squats	Accuracy (%) = 78.9
Montoye et al. [26]	Brisk walking	Accuracy (%) = 77.1
	Stair climbing and descending	Accuracy (%) = 87.3
	Cycling	Accuracy (%) = 66.0
	Jogging	Accuracy (%) = 92.5
	Overall activities	Accuracy (%) = 66.2
Rosemberg et al. [27]	Random forest – waist	
	Sitting	Accuracy (%) = 90.0
	Riding in a vehicle	Accuracy (%) = 91.0
	Standing still	Accuracy (%) = 67.0
	Standing moving	Accuracy (%) = 79.0
	Walking	Accuracy (%) = 84.0

PAEE – Physical activity energy expenditure;
R^2 – Coefficient of determination.

high mean values of intensity thresholds accuracy (84%), high mean coefficient of determination to estimate PAEE (80% and 70%, approximately for regression-based and machine learn-based models, respectively), as well as high accuracy for the recognition of physical activity patterns (except household activities) were found in the studies.

Comparisons between accuracy of machine learning technique and other traditional methods must be carefully performed. Usually the physical activity construct being assessed is not the same. Accuracy of physical activity patterns recognition, for example, cannot be directly compared to accuracy of intensities thresholds. Otherwise, PAEE is one of the few constructs in which both traditional methods (linear regression) and machine learning technique might be compared. In this context, an important characteristic is the use of more complete acceleration records using machine-learning technique, unlike approaches based on linear regression models in which only the mean acceleration for each interval of activity is considered. Therefore, these algorithms are theoretically able to better adjust the relationship between movement acceleration and PAEE by reducing the misclassification of activities with similar energy expenditures and different accelerations levels [31,39]. However, no relevant differences regarding accuracy of predict PAEE was identified comparing such methods in this review.

5. Conclusions

The use of accelerometers to measure physical activity has been increasingly frequent [31], however it is still necessary to advance in the interpretation of these measures. In this scenario, the approximation among researchers from different areas (e.g. program designers, engineers, and statisticians) is a relevant alternative to expand the developing and the use of more sophisticated analytical techniques. A closer relationship is essential to obtain even more accurate and detailed physical activity measures in the future.

This systematic review has summarized the methodological characteristics and main results of calibration studies based on raw-accelerometry. In conclusion, despite the relatively small number of studies identified, important differences in methodological decisions have influenced the results and comparability. Higher sample heterogeneity is required for better generalization of results from calibration studies. Furthermore, standardization of SVM calculations would dramatically contribute for comparability between physical activity estimates. Finally, the different statistical approaches used in the studies presented promising validity parameters.

Conflict of interest

None.

Acknowledgements

I extend my acknowledgements to all Brazilian citizens who, through the payment of their taxes, allow so many researchers to improve their scientific knowledge in public graduate programs.

References

[1] K.Y. Chen, D.R. Bassett, The technology of accelerometry-based activity monitors: current and future, Med. Sci. Sports Exerc. 37 (2005) 490–500.
[2] C.E. Matthews, Calibration of accelerometer output for adults, Med. Sci. Sports Exerc. 37 (2005) S512–522.
[3] C.C. Yang, Y.L. Hsu, A review of accelerometry-based wearable motion detectors for physical activity monitoring, Sensors (Basel) 10 (2010) 7772–7788.
[4] D.R. Bassett Jr., A. Rowlands, S.G. Trost, Calibration and validation of wearable monitors, Med. Sci. Sports Exerc. 44 (2012) S32–S38.
[5] J.J. Kavanagh, H.B. Menz, Accelerometry: a technique for quantifying movement patterns during walking, Gait Posture 28 (2008) 1–15.
[6] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, et al., Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement, Syst. Rev. 4 (2015) 1.
[7] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.
[8] C.V. Bouten, K.R. Westerterp, M. Verduin, J.D. Janssen, Assessment of energy expenditure for physical activity using a triaxial accelerometer, Med. Sci. Sports Exerc. 26 (1994) 1516–1523.
[9] M.P. Rothney, M. Neumann, A. Beziat, K.Y. Chen, An artificial neural network model of energy expenditure using nonintegrated acceleration signals, J. Appl. Physiol. (1985) 103 (2007) 1419–1427.
[10] D.W. Esliger, A.V. Rowlands, T.L. Hurst, M. Catt, P. Murray, R.G. Eston, Validation of the GENEA accelerometer, Med. Sci. Sports Exerc. 43 (2011) 1085–1093.
[11] I.C. Gyllensten, A.G. Bonomi, Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life, IEEE Trans. Biomed. Eng. 58 (2011) 2656–2663.
[12] S. Zhang, A.V. Rowlands, P. Murray, T.L. Hurst, Physical activity classification using the GENEA wrist-worn accelerometer, Med. Sci. Sports Exerc. 44 (2012) 742–748.

[13] M. Brandes, V.T. Van Hees, V. Hannover, S. Brage, Estimating energy expenditure from raw accelerometry in three types of locomotion, Med. Sci. Sports Exerc. 44 (2012) 2235–2242.
[14] L.R. Phillips, G. Parfitt, A.V. Rowlands, Calibration of the GENEA accelerometer for assessment of physical activity intensity in children, J. Sci. Med. Sport 16 (2013) 124–128.
[15] H.E. Bing, B.A.I. Jiawei, V.V. Zipunnikov, A. Koster, P. Caserotti, B. Lange-Maia, et al., Predicting human movement with multiple accelerometers using movelets, Med. Sci. Sports Exerc. 46 (2014) 1859–1866.
[16] M. Hildebrand, V.T. Van Hees, B.H. Hansen, U. Ekelund, Age-Group comparability of raw accelerometer output from wrist- and hip-worn monitors, Med. Sci. Sports Exerc. 46 (2014) 1816–1824.
[17] C.A. Schaefer, C.R. Nigg, J.O. Hill, L.A. Brink, R.C. Browning, Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth, Med. Sci. Sports Exerc. 46 (2014) 826–833.
[18] J. Skotte, M. Korshøj, J. Kristiansen, C. Hanisch, A. Holtermann, Detection of physical activity types using triaxial accelerometers, J. Phys. Act. Health 11 (2014) 76–84.
[19] M. Aittasalo, H. Vaha-Ypya, T. Vasankari, P. Husu, A.M. Jussila, H. Sievanen, Mean amplitude deviation calculated from raw acceleration data: a novel method for classifying the intensity of adolescents' physical activity irrespective of accelerometer brand, BMC Sports Sci. Med. Rehabil. 7 (2015) 18.
[20] T. Bastian, A. Maire, J. Dugas, A. Ataya, C. Villars, F. Gris, et al., Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough, J. Appl. Physiol. 118 (2015) 716–722.
[21] A.H.K. Montoye, L.M. Mudd, S. Biswas, K.A. Pfeiffer, Energy expenditure prediction using raw accelerometer data in simulated free living, Med. Sci. Sports Exerc. 47 (2015) 1735–1746.
[22] H. Vähä-Ypyä, T. Vasankari, P. Husu, A. Manttari, T. Vuorimaa, J. Suni, et al., Validation of cut-Points for evaluating the intensity of physical activity with accelerometry-based mean amplitude deviation (MAD), PLoS One 10 (2015) e0134813.
[23] H. Vähä-Ypyä, T. Vasankari, P. Husu, J. Suni, H. Sievänen, A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer, Clin. Physiol. Funct. Imaging 35 (2015) 64–70.
[24] A.H.K. Montoye, D. Bo, S. Biswas, K.A. Pfeiffer, Validation of a wireless accelerometer network for energy expenditure measurement, J. Sports Sci. 34 (2016) 2130–2139.
[25] A.H.K. Montoye, J.M. Pivarnik, L.M. Mudd, S. Biswas, K.A. Pfeiffer, Comparison of activity type classification accuracy from accelerometers worn on the hip wrists, and thigh in young, apparently healthy adults, Meas. Phys. Educ. Exerc. Sci. 20 (2016) 173–183.
[26] A.H.K. Montoye, J.M. Pivarnik, L.M. Mudd, S. Biswas, K.A. Pfeiffer, Wrist-independent energy expenditure prediction models from raw accelerometer data, Physiol. Meas. 37 (2016) 1770–1784.
[27] D. Rosenberg, S. Godbole, K. Ellis, D.I. Chongzhi, A. Lacroix, L. Natarajan, et al., Classifiers for accelerometer-measured behaviors in older women, Med. Sci. Sports Exerc. 49 (2017) 610–617.
[28] P.S. Freedson, D. John, Comment on estimating activity and sedentary behavior from an accelerometer on the hip and wrist, Med. Sci. Sports Exerc. 45 (2013) 962–963.
[29] V.T. van Hees, R. Golubic, U. Ekelund, S. Brage, Impact of study design on development and evaluation of an activity-type classifier, J. Appl. Physiol. 114 (2013) 1042–1051.
[30] M.E. Rosenberger, W.L. Haskell, F. Albinali, S. Mota, J. Nawyn, S. Intille, Estimating activity and sedentary behavior from an accelerometer on the hip or wrist, Med. Sci. Sports Exerc. 45 (2013) 964–975.
[31] R.P. Troiano, J.J. McClain, R.J. Brychta, K.Y. Chen, Evolution of accelerometer methods for physical activity research, Br. J. Sports Med. 48 (2014) 1019–1023.
[32] S. Zhang, P. Murray, R. Zillmer, R.G. Eston, M. Catt, A.V. Rowlands, Activity classification using the GENEA: optimum sampling frequency and number of axes, Med. Sci. Sports Exerc. 44 (2012) 2228–2234.
[33] S.G. Trost, K.L. McIver, R.R. Pate, Conducting accelerometer-based activity assessments in field-based research, Med. Sci. Sports Exerc. 37 (2005) S531–543.
[34] A.V. Rowlands, T. Yates, M. Davies, K. Khunti, C.L. Edwardson, Raw accelerometer data analysis with GGIR R-package: does accelerometer brand matter? Med. Sci. Sports Exerc. 48 (2016) 1935–1941.
[35] A.V. Rowlands, E.M. Mirkes, T. Yates, S. Clemes, M. Davies, K. Khunti, et al., Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? Med. Sci. Sports Exerc. (October (2)) (2017), http://dx.doi.org/10.1249/MSS.0000000000001435 [Epub ahead of print].
[36] M. Hildebrand, B.H. Hansen, V.T. van Hees, U. Ekelund, Evaluation of raw acceleration sedentary thresholds in children and adults, Scand. J. Med. Sci. Sports. 27 (December 12) (2017) 1814–1823, http://dx.doi.org/10.1111/sms.12795 Epub 2016 Nov 22.Source: Pubmed.
[37] C.E. Garber, B. Blissmer, M.R. Deschenes, B.A. Franklin, M.J. Lamonte, I.M. Lee, et al., American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise, Med. Sci. Sports Exerc. 43 (2011) 1334–1359.
[38] S.J. Preece, J.Y. Goulermas, L.P. Kenney, D. Howard, K. Meijer, R. Crompton, Activity identification using body-mounted sensors–a review of classification techniques, Physiol. Meas. 30 (2009) R1–33.
[39] N.F. Butte, U. Ekelund, K.R. Westerterp, Assessing physical activity using wearable monitors: measures of physical activity, Med. Sci. Sports Exerc. 44 (2012) S5–12.