Original Research Article

Marco Giurgiu*, Eni Hysenllari, Jörg Ottenbacher, Ulrich Ebner-Priemer and Irina Timm

Calibration and evaluation of MET models for estimating energy expenditure using thigh and ankle-worn move 4 accelerometer

https://doi.org/10.1515/teb-2025-0027
Received August 18, 2025; accepted November 24, 2025;
published online December 4, 2025

Abstract

Objectives: Valid estimation of energy expenditure remains a challenge, particularly when using ankle- and thigh-worn devices. The Move 4 is a research-grade accelerometer previously tested for predicting metabolic equivalents (METs) when worn at the waist or wrist. This study aimed to calibrate and evaluate regression models to estimate METs from Move 4 data when worn at the ankle and thigh.
Methods: Participants completed walking and jogging tasks under laboratory conditions while wearing Move 4 sensors and with indirect calorimetry as a reference measure. Models were calibrated using study 1 (n=160) and evaluated in an independent dataset (study 2; n=15). Performance was assessed using mean absolute error (MAE), root mean square error (RMSE), and Bland-Altman analyses.
Results: The MET models demonstrated strong agreement across both locations and datasets. For the thigh position, the MAE ranged from 0.60 METs (walking) to 1.38 METs (jogging), with RMSE of 0.82 and 1.70 in the evaluation data. Calibration metrics were comparable (jogging: MAE=1.24, RMSE=1.63). The ankle models showed similar accuracy, with MAEs of 0.66 (walking) and 1.39 (jogging), and RMSEs of 0.85 and 1.67, respectively. Systematic bias remained low (mean differences between −0.34 and −0.01 METs).
Conclusions: This study provides the first calibration and evaluation for estimating METs from ankle- and thigh-worn Move 4 accelerometers. The model indicated accurate, high-resolution MET estimation for walking and jogging. Future work should expand independent performance evaluations, including diverse activities such as static activities, and diverse samples under free-living conditions.

Keywords: validation; evaluation; accelerometer; sensor; energy expenditure; model performance

Introduction

Movement and non-movement behaviors - collectively referred to as physical behavior (i.e., sleep, physical activity, and sedentary behavior [1]) - are crucial for preventing and regulating various health conditions, including chronic diseases such as physical and mental health issues [2–5]. In recent years, the 24-h physical behavior perspective has gained prominence in conceptual models and approaches. Tremblay and colleagues argue that the 24-h approach has the potential to serve as a foundation for individualized precision 24-h movement behavior guidelines – tailored to individual characteristics and life contexts – thus promoting more precise and equitable health recommendations [6]. One central requirement to provide health precision based on 24-h physical behavior is a valid measurement of all components of the 24-h physical behavior approach. However, based on a systematic review of 967 evaluation studies [7], only very few devices evaluated outcomes for all three dimensions of 24-h physical behavior [8], thus revealing limited validity in capturing all facets of the 24-h physical behavior construct [9]. Moreover, Doherty et al. (2024) found in their systematic reviews that only about 11 % of commercially available wearable devices have published evaluation data for any physiological outcome [10]. They highlight those inconsistencies in validation methods and reporting standards across studies, making it difficult to

***Corresponding author: Dr. Marco Giurgiu**, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Hertzstr. 16, 76187 Karlsruhe, Germany, E-mail: Marco.Giurgiu@kit.edu. https://orcid.org/0000-0001-6684-3463
Eni Hysenllari and Jörg Ottenbacher, movisens GmbH, Karlsruhe, Germany
Ulrich Ebner-Priemer, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany; and Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany. https://orcid.org/0000-0002-2769-5944
Irina Timm, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany. https://orcid.org/0000-0002-2957-242X

assess the true measurement quality of these devices. As a result, they advocate for more standardized and ongoing validation efforts to ensure up-to-date and comparable accuracy assessments [10, 11]. Another concern is transparency: only about 31 % of evaluation articles even report the algorithm (e.g., equations or cut-points) used to compute outcomes, making it hard or unfeasible to reproduce or compare studies [7]. According to Bassett et al. (2012), calibration and evaluation are essential prerequisites for obtaining accurate and objective data from wearable monitors. However, valid calibration of accelerometer raw signals into metabolic equivalents (METs) remains challenging, as accuracy depends on device type, placement, and population [12, 13]. Without proper calibration, errors in energy expenditure estimation propagate into downstream health recommendations, limiting the utility of wearable-based assessments.

The Move 4 sensor by movisens GmbH is widely used in scientific studies for high-resolution, sensor-based assessment of physical behavior [14]. The Move 4 records tri-axial acceleration (64 Hz), angular rotation (gyroscope at 64 Hz), and barometric altitude (8 Hz). Its companion software can estimate, based on the triaxial raw acceleration signals, various metrics such as movement acceleration intensity, step count, activity classes, body posture, as well as energy expenditure. Moreover, the Move 4 includes live analysis (i.e., processing raw data on the sensor), which can then be used for real-time approaches in combination with ecological momentary assessment (EMA) [15–18]. The Move 4 can be worn on different body sites (ankle, wrist, hip, chest, thigh) and already has validated cut-points for some positions [19]. However, different wearing locations enable different abilities to capture acceleration signals and body postures. For example, chest or wrist-worn devices are more sensitive to measure upper body movement compared to hip or thigh-worn location, whereas a thigh position is favored to differentiate between sitting/lying and standing postures. Thus, the development of algorithms is specific to the respective wearing location and cannot be readily transferred to other wearing positions. Previous studies evaluated the performance of MET models derived from move sensors at the hip. For example, Anastasopoulou et al. (2014) found that the Movisens Move II (worn at the hip) produced very small errors for walking and moderate errors for jogging, while the Move II overestimated walking by +0.21 kcal/min (6 % bias) and underestimated jogging by −0.93 kcal/min (−7.5 % bias) relative to indirect calorimetry [20]. Yet, to date, no validated MET models exist for ankle- or thigh-worn Move 4 sensors, despite their increasing use in ambulatory research. This study aimed to calibrate and evaluate MET models for ankle and thigh-worn Move 4 accelerometers against indirect calorimetry, while using two independent studies. The summary of this article is presented in Figure 1.

Materials and methods

Participants – study 1

In study 1, 222 employees from a university setting were recruited between October 2022 and December 2023. Interested individuals underwent a brief screening interview (via email or phone) to verify eligibility. Participants were included if they were capable of performing typical daily activities without notable limitations and reported no current physical injuries or mental health conditions that could affect adherence to the study protocol (see Table 1). Three participants were subsequently excluded due to their inability to complete the study protocol. Additional participants were excluded (n=62) due to missing or faulty data, synchronization failures, measurement interruptions, device detachment, and artefacts in accelerometer or indirect calorimetry recordings (see Supplementary Material, Figure S1). After data cleaning, 160 participants from study 1 were included, comprising 75 males and 85 females. Male participants had a mean age of 36.4±11.2 years and a mean BMI of 24.0±2.8 kg/m^2. Female participants had a mean age of 36.4±10.7 years and a mean BMI of 24.2±3.6 kg/m^2. The average VO_2max across all participants was 40 mL/kg/min, indicating a generally moderate level of cardiorespiratory fitness. The study was approved by the Ethics Committee of the responsible university. All participants received detailed written and verbal information regarding the study procedures and provided written informed consent before participation. Participation was voluntary, and individuals could withdraw from the study at any time without any disadvantages.

Study procedure – study 1

Participants completed a structured series of consecutive tasks, divided into two sections: i) a semi-standardized protocol (conditions 1–2) and ii) an incremental treadmill test to volitional exhaustion following a 6-2-1 protocol (starting at 6 km/h and increasing by 1 km/h every 2 min; condition 3). The rationale for choosing this incremental protocol was to ensure volitional exhaustion, as the present study is a secondary analysis of a within-person encouragement design (see reference [21] for details). A minimum 20-min break was scheduled between the two sections to allow for the

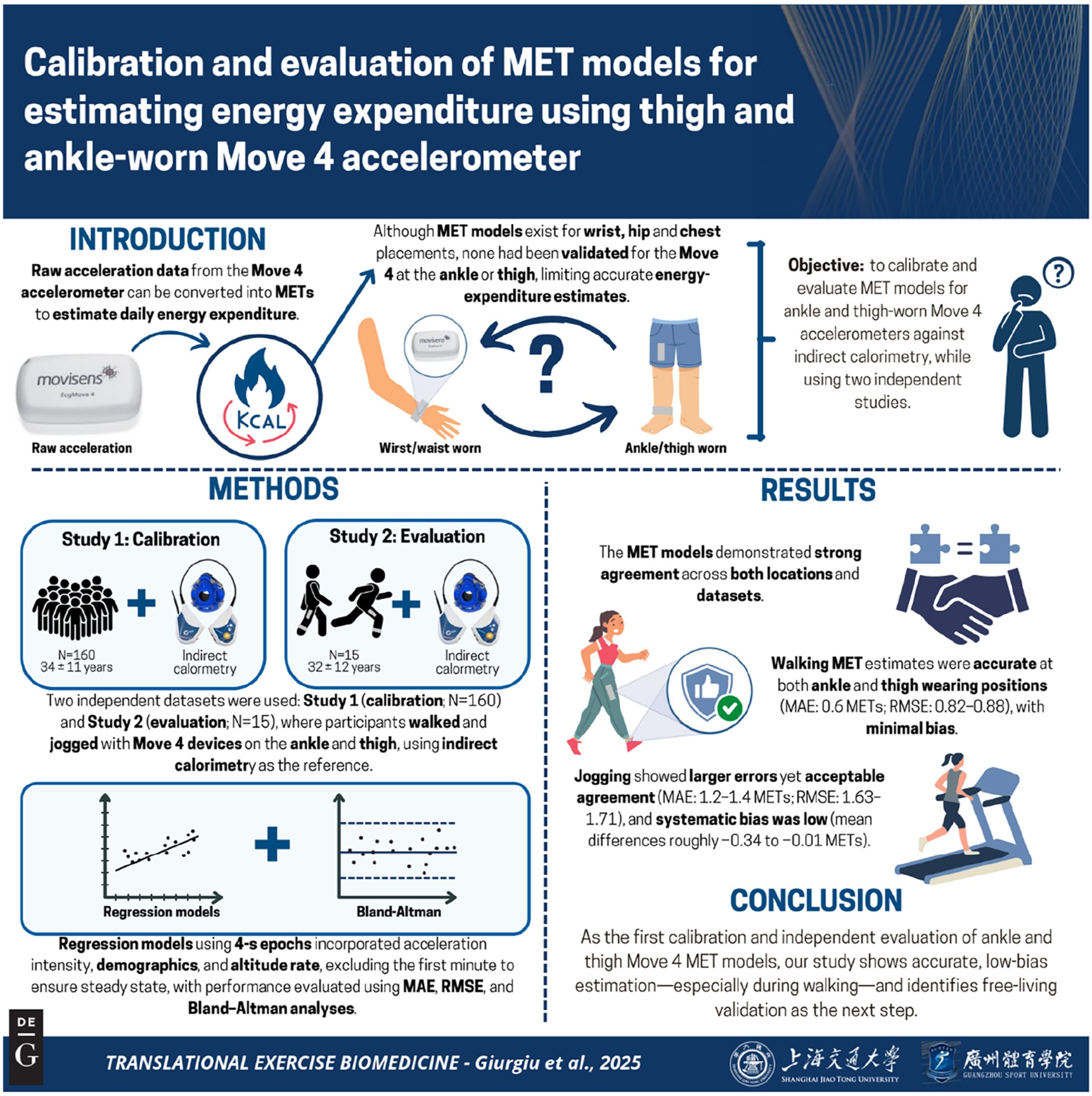

Figure 1: Graphical representation of this study. Key points: (1) this study advances translational measurement methodology by linking laboratory-based indirect calorimetry with free-living 24-h physical behavior monitoring. Accelerometer-derived movement acceleration was calibrated to metabolic equivalents (MET) to support precise assessment of everyday physical activity. (2) MET prediction models for thigh- and ankle-worn Move 4 accelerometers were developed in 160 adults and validated in an independent sample performing standardized walking and jogging tasks. (3) The models showed good accuracy for walking, and moderate accuracy for jogging, with small systematic bias and proportional error at higher intensities. These models enhance the reliability of activity assessment for research and applied health monitoring. Figure created with BioRender.

transition from portable to stationary measurement systems. In conditions 1 and 2, walking speed was self-selected. Participants received verbal instructions such as “Please walk at a relaxed pace for 3 min” and “Please walk at a brisk pace for 3 min”, with observed mean accelerometer magnitudes of 669 mg (ankle) and 267 mg (thigh) during relaxed walking, and 966 mg (ankle) and 369 mg (thigh) during brisk walking. Throughout the testing session, participants were continuously observed by the study supervisor to ensure correct task execution.

All procedures took place in a controlled laboratory setting. Participants were equipped with multiple devices: Move 4 sensors (movisens GmbH, Karlsruhe, Germany) were worn on the right hip, wrist, thigh, and ankle. In addition, participants wore an EcgMove 4 (movisens GmbH, Karlsruhe, Germany) sensor on the chest to record electrocardiographic data. Participants also wore a portable indirect calorimetry system throughout the protocol. Before starting the protocol, all activity monitors were initialized simultaneously using the same computer to

Table 1: Specification of conditions during the study 1 protocol (activities).

Activities	Description	Duration, min	Compendium of physical activities (MET)[a]
Slow walking	Slow walking in a habitual way and self-defined pace	3	2.8–3.0
Fast walking	Fast walking habitually and self-defined pace	3	4.3–7.0
Treadmill protocol (6-1-2)	Starting with 6 km/h and changing the speed by 1 km/h after every 2 min	Until exhaustion	4.8–14.8

MET, metabolic equivalents. [a]MET values from the compendium of physical activities [22].

Table 2: Specification of conditions during the study 2 protocol (activities).

Activities	Description	Duration, min	Compendium of physical activities (MET)[a]
Slow walking	Slow walking in a habitual way and self-defined pace	3	2.3–3.0
Normal walking	Normal walking freely at an individual's pace	3	3.8–4.0
Fast walking	Fast walking habitually and self-defined pace	3	4.3–7.0
Jogging	Free jogging at a self-defined pace	3	7.5
Treadmill: slow walking	Slow walking (~4 km/h) on a treadmill	3	3.5
Treadmill: normal walking	Normal walking (~5 km/h) on a treadmill	3	3.8
Treadmill: fast walking	Fast walking (~6 km/h) on a treadmill	3	4.8
Treadmill: normal jogging	Normal jogging (~7 km/h) on a treadmill	3	6.5
Treadmill: fast jogging	Fast jogging (~8 km/h) on a treadmill	3	8.5

MET, metabolic equivalents. [a]MET values from the compendium of physical activities [22].

ensure synchronized time-stamping across devices. The accelerometer and indirect calorimetry datasets were merged by visual inspection, aligning the time-stamped signals at second-by-second epoch length. Participants' body weight and height were measured without shoes using an electronic column scale (Seca GmbH & Co. KG, Hamburg, Germany). Throughout the entire protocol, participants wore all activity monitors concurrently.

Participants – study 2

In study 2, a sample of 15 volunteers was recruited with the same eligibility criteria as in study 1. The sample included nine male and six female participants. Male participants had a mean age of 31.7±11.7 years and a mean BMI of 23.2±2.1 kg/m^2. Female participants had a mean age of 41.1±14.9 years and a mean BMI of 23.7±2.6 kg/m^2. The average VO_2max across all participants was 42 mL/kg/min. All participants received detailed written information about the study procedures, including a data privacy statement and a consent form. Participation in the study was conditional upon providing written informed consent, and individuals could withdraw from the study at any time without any disadvantages.

Study procedure – study 2

Data collection was conducted on the premises of the Karlsruhe Institute of Technology (KIT). During the protocol, participants were equipped with multiple measurement devices: one Move 4 sensor was worn at each of the following locations: on the right hip, left upper arm, wrist, thigh (secured with a strap over clothing), and ankle. In addition, participants wore an EcgMove 4 sensor on the chest to record electrocardiographic data. Body weight and height were assessed without shoes using a calibrated electronic scale. Before beginning the activities (see Table 2), the study supervisor verified the correct placement and fit of all sensors. All devices were started immediately before the study protocol on the same computer. To ensure precise synchronization across all devices, participants performed a series of standardized jumps immediately after sensor attachment and again at the end of the protocol, following recommended procedures for defining clear temporal reference events [23]. Participants also wore a portable indirect calorimetry system throughout the protocol. During the testing session, participants were closely guided by the study supervisor to ensure the correct execution of all prescribed tasks.

Criterion measure

Indirect calorimetry assessed with a gas analyzer (MetaMax3B Cortex; Metalyser3B; CORTEX Biophysik GmbH, Leipzig, Germany) served as the criterion measure in both studies, and has been shown to provide valid and reliable estimates of energy expenditure [24]. In study 1, conditions 1 and 2 were assessed with the portable MetaMax3B system, while the treadmill protocol was assessed with the stationary Metalyser3B. In study 2, the portable system was used for all conditions. The MetaMax3B is a portable metabolic measurement system composed of a measurement module, a battery module, and a face mask, which covers the mouth and nose of the participant. It measures gas volume with a bidirectional digital turbine. The O_2 and CO_2 concentrations are measured by using an electrochemical cell and an infrared analyzer. The Metalyser 3B works the same way, without being carried by the participant. The volume of oxygen consumed (VO_2) and the volume of carbon dioxide produced (VO_2) were calculated by standard metabolic algorithms based on the Haldane transformation [25]. The system was paired to the Metasoft 3 software, v3.02.36, for analysis. According to established recommendations, the data were smoothed over a moving average of 20 s [26]. For subsequent processing of the data, the Meta Soft Studio export was used as an Excel file with a one-second resolution. As a final step, the Metabolic Equivalent of Task values (MET values) were calculated from the reference based on gender and oxygen uptake per kilogram of body weight so that they correspond to the following formula, which was also used for the Move sensors: Men: 1 MET=3.5 mL O_2/(kg min); Women: 1 MET=3.15 mL O_2/(kg min). Before each test, the MetaMax was calibrated according to the manufacturer's guidelines.

Move accelerometer

The Move 4 and EcgMove 4 (movisens GmbH, https://www.movisens.com/en/, Karlsruhe, Germany) are triaxial activity monitors with a size of 62 × 39 × 11 mm and a mass of 25 g. The sensor records acceleration at a range of ±16 g and a sampling rate of 64 Hz. The Move 4 accelerometers were attached to the right upper thigh with a strip of tape (study 1) or with a strap over clothing (study 2), on the right hip using an elastic belt, at the non-dominant wrist using a bracelet, and at the ankle using a strip. The EcgMove 4 was attached to the left side of the chest by using two disposable adhesive electrodes. The manufacturer's software SensorManager (v1.11.19) was used to initialize and download the data, and the software DataAnalyzer (v1.16.8) was applied to calculate time-stamped data with a 1-s resolution. The outcome parameters were calculated for each sensor location separately. The calculation of MET values is performed in two steps. First, the activity type - such as walking, sitting/lying, or cycling - is identified based on acceleration and barometric pressure signals. Once the activity type is determined, the corresponding activity-specific model is selected to estimate MET values. Each activity type has its own dedicated model that has been specifically fitted for that activity (for details, see Ref. [27]).

Data preprocessing, calibration, and model evaluation

Prior to the statistical analyses, several participants were excluded through the data cleaning procedure. In particular, in study 1 data were excluded due to i) complete or partial missing data (n=2); failed synchronization between accelerometer and criterion measures (n=8); measurement interruption of the accelerometer during the protocol (n=1); accelerometer detached from the body during the protocol (n=2); faulty air pressure signal from the accelerometer (n=2); missing indirect calorimetry measurements (n=23); artifacts in indirect calorimetry measurement such as incorrect breathing technique, technical errors or (n=31) (Supplementary Material, Figure S1). Resulting in a total of 160 participants with complete data from at least one sensor location (thigh or ankle). Besides the cleaning procedure, the Excel files of the calorimetry system and the accelerometer data were timely synchronized by manually merging all files into a final data set via SPSS (version 28). Accelerometer data that could not be assigned a label during synchronization was classified as unknown.

To generate MET prediction models for each activity, linear regression analysis was employed. The input variables included demographics (i.e., age, height, and weight) of the participants, as well as a physical activity metric derived from movement acceleration intensity (for details see Ref. [28]). This metric was calculated based on raw triaxial acceleration signals collected from the move accelerometer. Each axis signal underwent bandpass filtering using a fourth-order Butterworth filter with a frequency range of 0.25–11 Hz, effectively eliminating the gravitational offset and high-frequency noise unrelated to voluntary movements. Subsequently, the vector magnitude of the filtered triaxial signals was computed for each sample. Within each output interval, the vector magnitude was aggregated by calculating the mean value across all samples. The resulting parameter, movement acceleration intensity, is expressed in units of gravitational acceleration (g, where 1 g=9.81 m/s^2). In addition, altitude differences were included as an input feature. Altitude above sea level was determined from barometric pressure measurements, and the altitude signal was derived according to the method described in the manufacturer's documentation (for details see: Ref. [29]). The first derivative of the altitude signal was computed to obtain the rate of altitude change, enabling the detection of relative height variations with a resolution of approximately 10 cm. To ensure the reliability of steady-state activity data, measurements obtained during the first minute of each condition (walking, jogging) were excluded from analysis, accounting for the physiological transition phase. MET values were computed in 4-s epochs, providing a fine-grained temporal resolution suitable for detailed activity-energy expenditure modeling. Data from study 1 were used to calibrate the models, whereas data from study 2 were used to evaluate the MET models. To compare the predicted MET with the reference MET values from the indirect calorimetry, various statistical indices and visualizations were used. In particular, root mean square error, mean absolute error, mean relative error, lower and upper limit of agreement, as well as Bland-Altman analysis focusing on overall agreement without testing proportional bias. Bland–Altman plots (Figures 2 and 3) were created to visually represent inter-method agreement on a 4-s epoch-by-epoch level between predicted MET values from sensor data compared to the MET values from the indirect calorimetry. In these figures, the brown line represents the mean, whereas the black lines represent the 95 % limit-of-agreement interval. From these plots, it is possible to detect trends in inter-method discrepancy based on the mean difference value of the variable being considered. All analyses were conducted using MATLAB 2015b.

Results

Calibration

Across all four Bland-Altman analyses (Figure 2a–d), there is a balance between over- and underestimated numbers of data points when comparing sensor-predicted MET values and the MET values derived from the indirect calorimetry. However, for both thigh and ankle wearing locations, the differences in walking conditions are smaller. The mean MET difference from thigh-sensor predicted MET values to the MET values derived from the indirect calorimetry was −0.01 (SD=1.6) METs, 95 % confidence interval (CI): - 3.2 to 3.2, in the jogging conditions based on 13,824 epochs of 4 s segments, and 0.01 (SD=0.8), 95 % CI: −1.6 to 1.6, in the walking conditions based on 7,957 epochs. The mean MET difference from ankle-sensor predicted MET values to the MET values derived from the indirect calorimetry was −0.001 (SD=1.2) METs, 95 % CI: −2.4 to 2.4, in the jogging conditions based on 15,680 epochs, and 0.01 (SD=0.9), 95 % CI: −1.7 to 1.7, in the walking conditions based on 8,249 epochs. Notably, periods prior to the achievement of a steady state were excluded from the analysis (i.e., the first minute of each activity). Although mean differences were small across locations and activities, the Bland–Altman plots showed a downward trend consistent with proportional bias - overestimation at lower METs and underestimation at higher METs.

MAE was calculated to assess the accuracy of sensor-predicted MET values compared to MET values derived from the indirect calorimetry. MAE was calculated separately for both the walking and jogging conditions. As seen in Table 3, across both sensor positions, the walking condition had a MAE of 0.6 METs, whereas the MAEs of the jogging conditions were 1 MET for the ankle and 1.2 METs for the thigh position. These values correspond to approximately 10–20 % of the expected energy cost of walking and jogging, respectively. In summary, the largest prediction error was observed during the jogging activity, likely due to the model being trained on data that did not reach a steady state during jogging.

Evaluation

Across all four Bland-Altman analyses (Figure 3a–d), there is a balance between over- and underestimated numbers of data points when comparing sensor-predicted MET values and the MET values derived from the indirect calorimetry. However, for both thigh and ankle wearing locations, the differences in walking conditions are smaller. Overall, the analyses indicate that the results for the evaluation data are comparable to those of the calibration phase, suggesting consistent model performance. In particular, the mean MET difference between thigh-sensor predicted MET values to the MET values derived from indirect calorimetry was 0.05 (SD=1.6) METs, 95 % confidence interval (CI): −2.8 to 3.7, in the jogging conditions based on 1,355 epochs of 4-s segments, and 0.01 (SD=0.8) METs, 95 % CI: −1.6 to 1.6, in the walking condition based on 3,325 epochs. The mean MET difference from ankle-sensor predicted MET values to the MET values derived from the indirect calorimetry was −0.09 (SD=1.7) METs, 95 % CI: −3.4 to 3.2, in the jogging conditions based on 1,355 epochs, and −0.1 (SD=0.8), 95 % CI: −1.8 to 1.5, in the walking conditions based on 3,325 epochs. Similar to the calibration models, Bland-Altman plots showed a downward trend consistent with proportional bias - overestimation at lower METs and underestimation at higher METs.

MAE was calculated to assess the accuracy of sensor-predicted MET values compared to MET values derived from the indirect calorimetry. MAE was calculated separately for

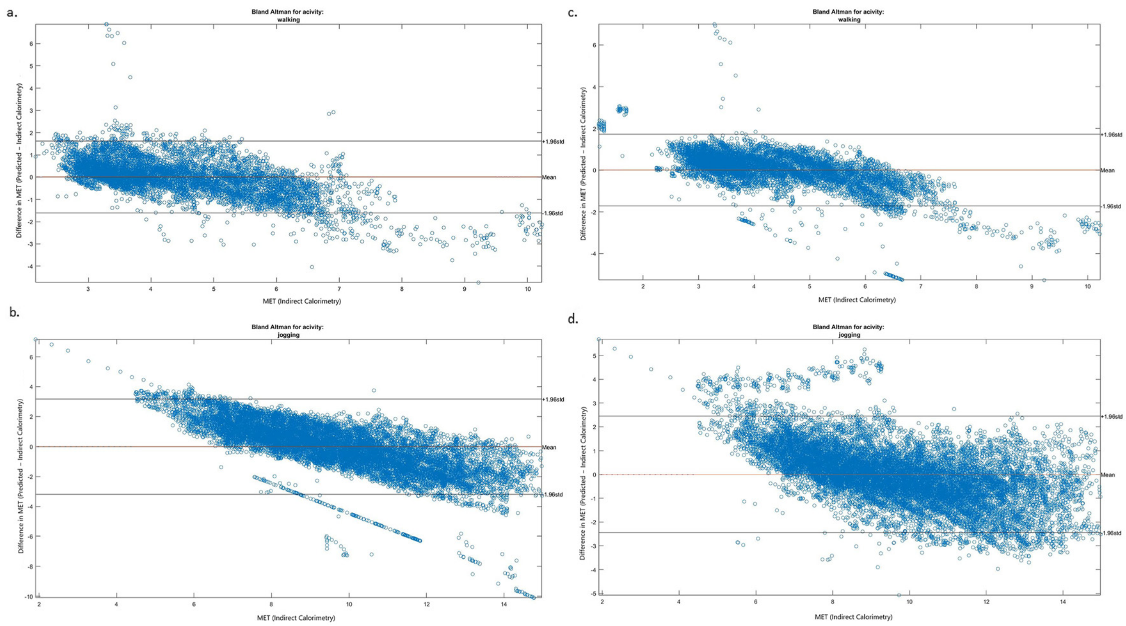

Figure 2: A. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (thigh placement) and MET values derived from indirect calorimetry during walking activity. B. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (thigh placement) and MET values derived from indirect calorimetry during jogging activity. C. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (ankle placement) and MET values derived from indirect calorimetry during walking activity. D. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (ankle placement) and MET values derived from indirect calorimetry during jogging activity.

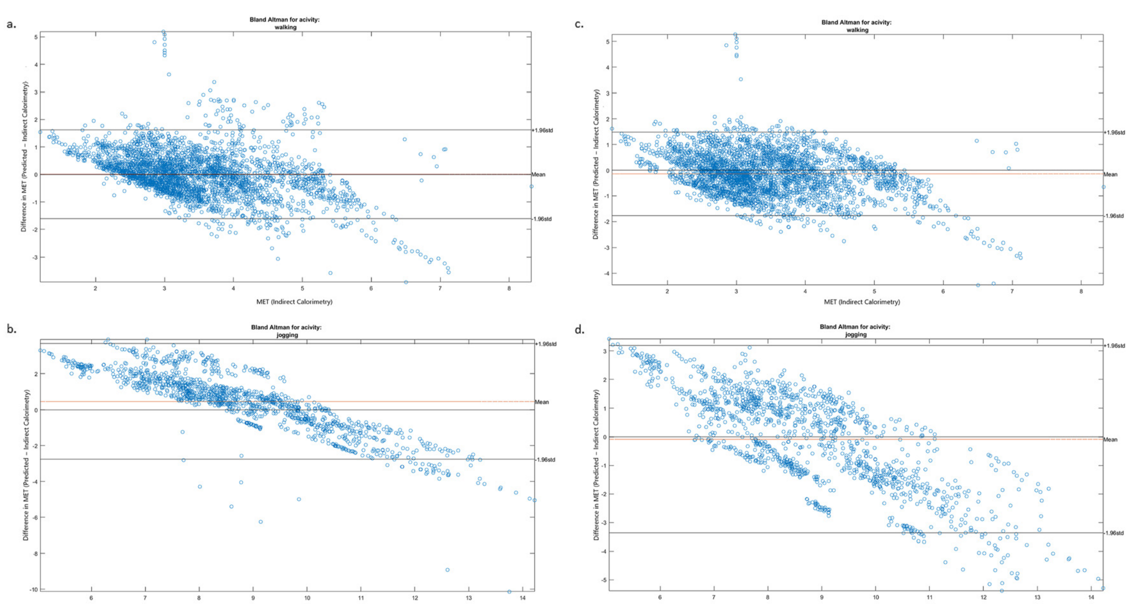

Figure 3: A. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (thigh placement) and MET values derived from indirect calorimetry during walking activity. B. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (thigh placement) and MET values derived from indirect calorimetry during jogging activity. C. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (ankle placement) and MET values derived from indirect calorimetry during walking activity. D. Bland-Altman plot showing agreement between predicted metabolic equivalent (MET) values from the Move 4 sensor (ankle placement) and MET values derived from indirect calorimetry during jogging activity.

Table 3: Evaluation parameters for both calibration and evaluation data sets (thigh).

Parameter	Study 1 (calibration)		Study 2 (evaluation)	
	Jogging	Walking	Jogging	Walking
Root mean square error	1.625	0.826	1.698	0.816
Mean absolute error	1.243	0.598	1.376	0.596
Mean relative error	0.138	0.137	0.166	0.179
Value number	32,073	32,073	6,437	6,437
Degrees of freedom	32,072	32,072	6,436	6,436
Difference mean	−0.013	−0.013	−0.104	−0.104
Standard derivation	1.165	1.165	1.277	1.277
Lower limit of agreement	−2.296	−2.296	−2.606	−2.606
Upper limit of agreement	2.271	2.271	2.398	2.398

both the walking and jogging conditions. As seen in Table 4, across both sensor positions, the walking condition had a MAE of 0.6 METs, whereas the MAE of the jogging conditions was 1.4 METs for both the ankle and thigh positions. The same analyses for energy expenditure were conducted using both ankle- and thigh-worn accelerometer data; the corresponding plots are provided in the Supplementary Material (Figures S2–9).

Discussion

This study aimed to calibrate and evaluate regression-based models to estimate energy expenditure (METs) from Move 4 data when worn at the thigh and ankle, using two independent laboratory datasets. The study findings demonstrate that both thigh- and ankle-based models yielded accurate MET predictions under steady-state conditions, particularly during level walking with MAE≈0.6 METs for walking and ≈1.2–1.4 METs for jogging. Although average bias was close to zero, the Bland–Altman distributions revealed a tendency of proportional bias. In particular, the models tended to overestimate at low METs and underestimate at high METs. This within-category trend aligns with the larger errors observed during jogging. In the calibration phase, models trained on indirect calorimetry-based MET values showed high agreement for walking at various speeds, with only small mean biases for both sensor positions. While the thigh placement generally yielded slightly more precise estimates, the ankle placement also demonstrated acceptable performance. Higher errors during jogging are plausible, which might relate to increased variability in gait mechanics and VO_2 kinetics at higher running speeds (e.g., [30–32]). In the evaluation phase, model performance was impressively consistent with the calibration results, confirming that both models generalize well to independent samples. Again, predictions during walking were particularly robust across participants and conditions, with only minor deviations from the reference. Jogging performance remained more variable, with slightly larger RMSE and wider limits of agreement, especially at the ankle, but overall agreement was still within acceptable bounds for practical use. Overall, the models showed robust performance, especially for level walking, in line with the objectives of creating an accurate thigh-based energy expenditure estimator.

Table 4: Evaluation parameters for both calibration and evaluation data sets (ankle).

Parameter	Study 1 (calibration)		Study 2 (evaluation)	
	Jogging	Walking	Jogging	Walking
Root mean square error	1.247	0.877	1.671	0.847
Mean absolute error	0.950	0.606	1.386	0.655
Mean relative error	0.107	0.146	0.161	0.198
Value number	34,640	34,640	6,437	6,437
Degrees of freedom	34,639	34,639	6,436	6,436
Difference mean	−0.058	−0.058	−0.336	−0.336
Standard derivation	0.960	0.960	1.236	1.236
Lower limit of agreement	−1.938	−1.938	−2.757	−2.757
Upper limit of agreement	1.823	1.823	2.086	2.086

The study findings are in line with previous validation efforts using movisens devices [20] and extend this work by providing a calibrated and evaluated MET model for ankle and thigh locations. Studies employing different devices and wear locations further support the utility of accelerometer-based energy expenditure estimates: Nagayoshi et al. (2019) validated a waist-worn accelerometer against indirect calorimetry, reporting high correlations (r˜0.85–0.88), but systematic underestimation at higher intensities (mean errors −0.6 to −0.1 METs) [33], mirroring the observation that model performance deteriorates during vigorous activity (e.g., jogging), likely due to limited steady-state representation in training data. Research-grade devices tend to underestimate energy expenditure, whereas consumer devices like Fitbit often overestimate, highlighting the variability across platforms [34]. In free-living adults, White et al. (2019) demonstrated that thigh acceleration predicts activity energy expenditure moderately well (r>0.6 vs. doubly labeled water) and yields high agreement for total energy expenditure (r≈0.9) [35]. Together, these results support thigh accelerometry as a valid surrogate for energy expenditure.

In contrast, simple posture-based monitors are less accurate for energy expenditure: activPAL's proprietary algorithm achieved r≈0.76 and RMSE≈1.74 METs for energy expenditure in a semi-structured protocol, whereas a custom neural network on raw thigh data improved to r≈0.89 and RMSE≈1.07 METs [36]. Collectively, these comparisons suggest that the Move 4 thigh-based regression models achieve comparable or better performance than existing approaches, particularly in walking, though accuracy declines at vigorous intensities. Notably, it is difficult to compare evaluation studies due to the heterogeneity of study protocols, samples, and settings [7].

Ho et al. (2020) showed that combining ankle-mounted accelerometer data with heart rate can feasibly predict energy expenditure in both athletes and non-athletes [37]. In older adults, ankle-worn accelerometers outperformed other placements for physical activity intensity classification against indirect calorimetry, with excellent discrimination of sedentary behavior and MVPA and matching or exceeding waist placement [38]. Building on this, ankle-worn accelerometers showed high validity (r=0.81) and reliability (ICC=0.92) in adolescent girls with low misclassification error vs. indirect calorimetry [39, 40]. Collectively, these studies indicate that ankle-worn sensors are promising proxies for metabolic energy expenditure across youth and adults under controlled conditions.

This study represents an important step toward a comprehensive MET prediction model applicable across different activities and populations. Following the phase-based validation framework by Keadle et al. (2019), which progresses from mechanical testing (Phase 0) to calibration (Phase I), structured laboratory validation (Phase II), free-living validation (Phase III), and application in health research (Phase IV) [41], this work can be positioned within the phase I and II of the model. Phase I was completed with a large calibration sample (n=160), allowing robust model development across multiple sensor placements. Following initial calibration work, phase II was addressed through structured walking and jogging tasks in a separate evaluation cohort (n=15). However, while the phase I calibration benefited from high participant numbers, the subsequent phase II evaluation involved fewer individuals. Importantly, the results showed strong agreement with indirect calorimetry and low systematic error across both thigh and ankle placements. Future work will need to move to Phase III, which is inherently difficult due to the nuances of real-life physical behaviors and limited ability to capture energy expenditure through portable indirect calorimetry.

Limitations

Several limitations should be noted. First, the calibrated and evaluated model was confined to steady-state, dynamic locomotor activities (walking, jogging). While these behaviors represent common components of daily movement, the models were not calibrated or tested on static or quasi-static activities such as cycling, resistance exercises, slow stepping, or sedentary tasks involving small lower-limb movements. Therefore, errors may be larger in activities with weak acceleration-energy relationships. Moreover, the relationship between thigh acceleration and true energy cost can differ: for instance, cycling produces low acceleration but moderate METs, and small leg movements (fidgeting) may raise MET only slightly, and thus the regression models may not generalize well to these types of behavior. Future validation should include a wider range of tasks (e.g., cycling at various intensities, stair-climbing) to refine the model coefficients. Second, although the sample size was relatively large for a laboratory-based calibration study and included a fully independent test dataset, the sample was limited to healthy adults under controlled conditions, and additional participants had to be excluded due to several technical reasons. This restricts the generalizability of the findings to broader populations. According to recent expert reviews, many wearable validation studies suffer from methodological weaknesses (small n, restricted tasks, inconsistent protocols [9, 10]). These issues were addressed by implementing: (1) independent training and validation cohorts; (2) careful removal of non-steady-state periods; (3) alignment with indirect calorimetry as gold-standard reference. However, the sample lacked diversity in terms of age, BMI range, functional capacity, and contextual behavior. Third, evaluation results should be interpreted cautiously given the small sample size (n=15), which restricts statistical power, limits the generalizability of the findings, and reduces the precision of error estimates. The evaluation was confined to a controlled setting. Differences in attachment methods between studies (tape in study 1 vs. strap in study 2) may have introduced minor variability in signal quality. However, both fixation types ensured stable sensor placement and reflected the development of improved accessories over time. Fourth, due to the primary purpose of the data collection, the treadmill protocol used was designed to test participants' volitional exhaustion, which does not represent the recommended gold standard for performance evaluation, as proposed by the INTERLIVE network (i.e., a steady-state treadmill test with a consistent and self-selected walking speed [42]. However, to account for steady-state activity data, the first minute of data was removed in each case. Finally, while energy expenditure metric offers an interpretable unit for comparing activities, its regression-based estimation from thigh acceleration should be treated with caution outside the trained activity spectrum. Nonlinear approaches (e.g., deep learning) may better capture complex motion–energy relationships, particularly when paired with multimodal sensor data and contextual inputs. However, these approaches will also require large, diverse, and ecologically valid datasets for training and evaluation. These study findings lay important groundwork, but further calibration and validation efforts are essential before the model can be applied in translational or public health contexts.

Conclusions

In conclusion, this study provides new evidence that an ankle- and thigh-worn accelerometer (Move 4) can yield accurate MET estimates during walking and jogging in a laboratory setting. These regression models extend the capabilities of posture sensors by adding energy expenditure information. They performed robustly, particularly for walking, and comparably to existing devices that use multi-sensor or activity-dependent algorithms [20, 35]. In the future, combining posture and energy expenditure outputs could improve physical behavior classification as defined by current guidelines and enable smarter real-time interventions [16, 43]. Notably, both calibration and evaluation samples were employed, a practice still uncommon in many validation studies, and high-resolution indirect calorimetry reference data and detailed quality control throughout the processing pipeline. Future research may expand existing analytical approaches by leveraging modern machine learning techniques, in keeping with recommended validation frameworks [41].

Acknowledgments: We thank all participants and Manuel Leitner for research support.
Research ethics: The Ethics Committee of the Karlsruhe Institute of Technology (KIT) approved this study (date: 10/19/2020). The standard institutional documentation of ethics can see Supplementary Material.
Informed consent: All eligible participants received written and oral information regarding the study procedures before written informed consent was obtained. Participants were free to withdraw from the study at any time.
Author contributions: M.G. and I.T. collected the data. M.G., E.H. and J.O. analyzed the data. I.T. and M.G. wrote the manuscript. I.T., M. G., E.H., J.O. and U.E-P. edited the manuscript. All authors read and approved the final manuscript.
Use of Large Language Models, AI and Machine Learning Tools: Parts of the manuscript were linguistically revised using an AI-based large language model (ChatGPT, OpenAI, DeepL).
Conflict of interest: Author E.H. is an employee of movisens GmbH, which develops, manufactures, and distributes the sensors and analysis software used in this study. Author J.O. is a founder and shareholder of movisens GmbH. These affiliations may be considered potential conflicts of interest. I.T., M. G., and U.E-P. state no conflict of interest.
Research funding: This research was funded by the German Research Foundation [Grant Number: 496846758].
Data availability: Correspondence and requests for materials should be addressed to Marco Giurgiu.

References

1. Falck RS, Davis JC, Li L, Stamatakis E, Liu-Ambrose T. Preventing the “24-hour babel”: the need for a consensus on a consistent terminology scheme for physical activity, sedentary behaviour and sleep. Br J Sports Med 2022;56:367–8.
2. Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. J Psychiatr Res 2016;77:42–51.
3. Pinto AJ, Bergouignan A, Dempsey PC, Roschel H, Owen N, Gualano B, et al. Physiology of sedentary behavior. Physiol Rev 2023;103: 2561–622.
4. Scott AJ, Webb TL, Martyn-St James M, Rowse G, Weich S. Improving sleep quality leads to better mental health: a meta-analysis of randomised controlled trials. Sleep Med Rev 2021;60:101556.
5. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017;32:541–56.
6. Tremblay MS, Duncan MJ, Kuzik N, Silva DAS, Carson V. Towards precision 24-hour movement behavior recommendations-the next new paradigm? J Sport Health Sci 2024;13:743–8.
7. Giurgiu M, Von Haaren-Mack B, Fiedler J, Woll S, Burchartz A, Kolb S, et al. The wearable landscape: issues pertaining to the validation of the measurement of 24-h physical activity, sedentary, and sleep behavior assessment. J Sport Health Sci 2025;14:101006.
8. Stevens ML, Gupta N, Inan Eroglu E, Crowley PJ, Eroglu B, Bauman A, et al. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. BMJ Open Sport Exerc Med 2020;6:000874.
9. Giurgiu M, Ketelhut S, Kubica C, Nissen R, Doster AK, Thron M, et al. Assessment of 24-hour physical behaviour in adults via wearables: a systematic review of validation studies under laboratory conditions. Int J Behav Nutr Phys Act 2023;20:68.
10. Doherty C, Baldwin M, Keogh A, Caulfield B, Argent R. Keeping pace with wearables: a living umbrella review of systematic reviews evaluating the accuracy of consumer wearable technologies in health measurement. Sports Med 2024;54:2907–26.
11. Giurgiu M, Timm I, Becker M, Schmidt S, Wunsch K, Nissen R, et al. Quality evaluation of free-living validation studies for the assessment of 24-hour physical behavior in adults via wearables: systematic review. JMIR Mhealth Uhealth 2022;10:e36377.
12. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc 2012;44:S32–38.
13. Clanchy K, Stanfield M, Smits E, Liimatainen J, Ritchie C. Calibration and validation of physical behaviour cut-points using wrist-worn

ActiGraphs for children and adolescents: a systematic review. J Sci Med Sport 2024;27:92–104.
14. movisens GmbH. Move 4 | movisens Docs [online]. Available from: https://docs.movisens.com/Sensors/Move4/ [Accessed 13 Nov 2025].
15. Giurgiu M, Koch ED, Plotnikoff RC, Ebner-Priemer UW, Reichert M. Breaking up sedentary behavior optimally to enhance mood. Med Sci Sports Exerc 2020;52:457–65.
16. Giurgiu M, Niermann C, Ebner-Priemer U, Kanning M. Accuracy of sedentary behavior-triggered ecological momentary assessment for collecting contextual information: development and feasibility study. JMIR Mhealth Uhealth 2020;8:e17852.
17. Ebner-Priemer UW, Koudela S, Mutz G, Kanning M. Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. Front Psychol 2013;3:596.
18. Timm I, Reichert M, Ebner-Priemer UW, Giurgiu M. Momentary within-subject associations of affective states and physical behavior are moderated by weather conditions in real life: an ambulatory assessment study. Int J Behav Nutr Phys Act 2023;20:117.
19. Beck F, Marzi I, Eisenreich A, Seemüller S, Tristram C, Reimers AK. Determination of cut-off points for the Move4 accelerometer in children aged 8–13 years. BMC Sports Sci Med Rehabil 2023;15:163.
20. Anastasopoulou P, Tubic M, Schmidt S, Neumann R, Woll A, Härtel S. Validation and comparison of two methods to assess human energy expenditure during free-living activities. PLoS One 2014;9:e90606.
21. Giurgiu M, Timm I, Ebner-Priemer UW, Schmiedek F, Neubauer AB. Causal effects of sedentary breaks on affective and cognitive parameters in daily life: a within-person encouragement design. NPJ Mental Health Res 2024 21:1–11.
22. Herrmann SD, Willis EA, Ainsworth BE, Barreira TV, Hastert M, Kracht CL, et al. 2024 adult compendium of physical activities: a third update of the energy costs of human activities. J Sport Health Sci 2024;13:6–12.
23. Argent R, Hetherington-Rauth M, Stang J, Tarp J, Ortega FB, Molina-Garcia P, et al. Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure: expert statement and checklist of the INTERLIVE network. Sports Med 2022;52:1817–32.
24. Vogler AJ, Rice AJ, Gore CJ. Validity and reliability of the cortex MetaMax3B portable metabolic system. J Sports Sci 2010;28:733–42.
25. Haugen HA, Chan LN, Li F. Indirect calorimetry: a practical guide for clinicians. Nutr Clin Pract 2007;22:377–88.
26. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing from expired gas analysis indirect calorimetry. Sports Med 2010;40:95–111.
27. movisens GmbH. Welcome to the movisens documentation. Energy expenditure 2024. Available from: https://docs.movisens.com/Algorithms/energy_expenditure/#energy-expenditure.
28. movisens GmbH. Movement acceleration | movisens Docs [Online]. Available from: https://docs.movisens.com/Algorithms/physical_activity/#movement-acceleration-movementacceleration [Accessed 13 Nov 2025].
29. movisens GmbH. Physical activity | movisens Docs [Online]. Available from: https://docs.movisens.com/Algorithms/energy_expenditure/#energy-expenditure [Accessed 13 Nov 2025].
30. Jordan K, Newell KM. The structure of variability in human walking and running is speed-dependent. Exerc Sport Sci Rev 2008;36:200.
31. Carter H, Jones AM, Barstow TJ, Burnley M, Williams CA, Doust JH. Oxygen uptake kinetics in treadmill running and cycle ergometry: a comparison. J Appl Physiol 2000;89:899–907.
32. Padulo J, Borrelli M, Antiglio A, Esposito F. Gait variability and fatigability during a simulated 10-km running race in trained runners. Eur J Appl Physiol 2025;125:2529–35.
33. Nagayoshi S, Oshima Y, Ando T, Aoyama T, Nakae S, Usui C, et al. Validity of estimating physical activity intensity using a triaxial accelerometer in healthy adults and older adults. BMJ Open Sport Exerc Med 2019;5:e000592.
34. Rieckmann A, Jordan B, Burczik F, Meixner J, Thiel C. Validation of activity trackers to estimate energy expenditure in older adults with cardiovascular risk factors. PLoS One 2024;19:e0309481.
35. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. Int J Obes 2019;43:2333–42.
36. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Evaluation of the activPAL accelerometer for physical activity and energy expenditure estimation in a semi-structured setting. J Sci Med Sport 2017;20:1003–7.
37. Ho CS, Chang CH, Hsu YJ, Tu YT, Li F, Jhang WL, et al. Feasibility of the energy expenditure prediction for athletes and non-athletes from ankle-mounted accelerometer and heart rate monitor. Sci Rep 2020;10:8816.
38. Duncan MJ, Rowlands A, Lawson C, Leddington Wright S, Hill M, Morris M, et al. Using accelerometry to classify physical activity intensity in older adults: what is the optimal wear-site? Eur J Sport Sci 2020;20:1131–9.
39. Duncan MJ, Dobell A, Noon M, Clark CCT, Roscoe CMP, Faghy MA, et al. Calibration and cross-validation of accelerometery for estimating movement skills in children aged 8–12 years. Sensors 2020;20:2776.
40. Hager ER, Treuth MS, Gormely C, Epps L, Snitker S, Black MM. Ankle accelerometry for assessing physical activity among adolescent girls: threshold determination, validity, reliability, and feasibility. Res Q Exerc Sport 2015;86:397–405.
41. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev 2019;47:206–14.
42. Johnston W, Judice PB, García PM, Mühlen JM, Skovgaard EL, Stang J, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. Br J Sports Med 2021;55:780–93.
43. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018;6:288.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/teb-2025-0027).