Journal for the Measurement of Physical Behaviour, 2025, 8, 1-10
https://doi.org/10.1123/jmpb.2024-0053

Human Kinetics
TECHNICAL NOTE

Identifying Days of Valid Wear for Studies Using Consumer-Grade Accelerometer-Based Activity Trackers: ActiWearCheck, a Python Library

Matthieu Gallou-Guyot,[1,2,3] Mor Diaw,[4,5] Nathanael Aubert-Kato,[3,6] Mame Saloum Coly,[7] Keyne Charlot,[8,9] Philippe Connes,[10] Tsukasa Yoshida,[11] Motohiko Miyachi,[11,12] and Julien Tripette[2,3,11]

[1]Japan Society for the Promotion of Science (JSPS), Chiyoda, Japan; [2]Department of Human Life and Environmental Sciences, Ochanomizu University, Bunkyo, Japan; [3]Center for Interdisciplinary AI and Data Science, Ochanomizu University, Bunkyo, Japan; [4]Laboratoire de Physiologie Humaine, Faculté de Médecine, de Pharmacie et d'Odontologie, Université Cheikh Anta Diop, Dakar, Sénégal; [5]IRL3189—CNRS Environnement, Santé, Sociétés, Dakar, Senegal; [6]Department of Information Sciences, Ochanomizu University, Bunkyo, Japan; [7]Laboratoire de Physiologie et Explorations Fonctionnelles, Université de Thiès, Thies, Sénégal; [8]Unité de Physiologie des Exercices et Activités en Conditions Extrêmes, Département Environnements Opérationnels, Institut de Recherche Biomédicale des Armées, Bretigny-Sur-Orge, France; [9]LBEPS, Univ Evry, IRBA, Université Paris-Saclay, Evry, France; [10]Laboratoire LIBM EA7424, Equipe "Biologie vasculaire et du globule rouge", Université Claude Bernard Lyon 1, Villeurbanne, France; [11]National Institute of Health and Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan; [12]Faculty of Sport Sciences, Waseda University, Tokorozawa, Japan

Obtaining reliable data from activity trackers is crucial for physical activity studies. Except for higher end models that offer heart rate (HR) monitoring, consumer-grade devices typically do not provide a method for estimating wearing time. This manuscript introduces a new Python software ("ActiWearCheck") developed to identify days of valid wear of Fitbit wrist-worn activity trackers, without relying on HR data. The software utilizes minute-per-minute time series for energy expenditure (EE) estimates extracted from accelerometer data and freely provided by the official Fitbit Web API. A case analysis was conducted on a data set of 3,339 days collected from 72 individuals who used the Alta HR or Inspire 2 devices to test the feasibility of estimating days of valid wear using ActiWearCheck with two different configurations, EE estimates ≥ 600 min over the basal metabolic rate and ≥10 hr with at least 1 min of EE over the basal metabolic rate. The valid wear classifications under both configurations showed accuracies of 72% and 86%, respectively, against the one performed using HR data (also available). The average daily estimates of the number of steps and EE were calculated for the 72 participants after identifying the days of valid wear, performed using ActiWearCheck. This procedure allowed exclusion of up to 25% of potentially invalid days, leading to 12% and 6% increases in the step count and EE estimates, respectively. These results demonstrate the feasibility and necessity of obtaining valid wear estimates for physical activity studies employing consumer-grade devices.

Keywords: physical activity, Fitbit, data processing, energy expenditure, step count

Key Points

- ActiWearCheck is an open-source Python software designed to determine days of valid wear for consumer-grade wrist-worn activity trackers, even when heart rate data is unavailable.
- By integrating ActiWearCheck, researchers using consumer-grade activity trackers can improve the reliability of their physical activity data, addressing known issues of underestimation in step count and energy expenditure.

Activity tracker devices are frequently used in physical activity (PA) studies. Although contemporary devices feature multisensing capabilities, the core of energy expenditure (EE) and step count predictions still relies on data using the accelerometer sensor chip (Chen & Bassett, 2005). Activity tracker devices are often categorized into two types. On the one hand, research-grade devices, such as Actigraph monitors, are designed to provide reliable estimations of daily PA volumes, which can be used to explore the relationship between lifestyle habits and a wide range of health-related endpoints. These devices are usually equipped with batteries and memory chips capable of powering the system and recording data for maximum periods of approximately 2–4 weeks depending on the integrated functionalities (e.g., liquid crystal display and high sampling rate modality) and selected time resolution (Actigraph, 2013; Activinsights, n.d.). They rarely offer >4 weeks of autonomy, except for some higher end devices, such as the Actigraph LEAP device (Actigraph, 2024). On the other hand, consumer-grade devices are made for the general public, enabling individuals to self-monitor their physical behaviors and enhance their long-term lifestyle. For this type of device, specific development requirements include miniaturization of the devices, which often limits their autonomy to a few days.

Gallou-Guyot https://orcid.org/0000-0002-2616-4850
Diaw https://orcid.org/0000-0003-0525-6235
Aubert-Kato https://orcid.org/0000-0002-9100-1855
Saloum Coly https://orcid.org/0000-0003-0893-9365
Charlot https://orcid.org/0000-0002-4366-4013
Connes https://orcid.org/0000-0002-9232-0268
Yoshida https://orcid.org/0000-0003-2304-6872
Miyachi https://orcid.org/0000-0002-1146-0905
Tripette (tripette.julien@ocha.ac.jp) is corresponding author, https://orcid.org/0000-0002-5669-5337

For instance, the wrist-worn Inspire 2 Fitbit features a lithium-polymer battery that lasts no more than 10 days on minimal settings (e.g., excluding heart rate [HR] measurements), as it stores minute-per-minute data for a maximum of 7 days before clearing the memory and switching to daily records (Fitbit, 2021). Although lithium-polymer batteries can be easily charged at home, the limited memory capacity of consumer-grade devices is compensated by the ability to be paired with smartphone handsets. This enables cloud data saving at the highest resolution when synchronized at the recommended time interval. Some researchers may find the standalone combination of a consumer-grade activity tracker and smartphone handset attractive because it facilitates conducting PA studies that may involve long observation periods. Additional benefits of consumer-grade devices in research include their lower cost compared with traditional research-grade devices and the validity of key parameters commonly used in PA studies, such as step count, EE, and derived outcomes (Diaz et al., 2015; Murakami et al., 2016, 2019; Nakagata et al., 2022; Sushames et al., 2016; Tully et al., 2014).

To ensure reliable daily estimates of PA parameters, the first step in processing research-level activity tracker data often involves estimating the wearing time. Activity tracker devices may continue to record data even when they are not worn, and the data recorded during nonwear periods can be similar to those recorded during sedentary periods (Randhawa et al., 2023). Several studies have highlighted this issue (Banda et al., 2016; Cain et al., 2013; Mannini & Sabatini, 2010; Randhawa et al., 2023; Skovgaard et al., 2023; Vanhelst et al., 2019); some have developed wearing-time algorithms based on the number of epochs with zero count values for uniaxial acceleration (Banda et al., 2016; Cain et al., 2013), whereas others have proposed more sophisticated triaxial statistical learning-based algorithms (Mannini & Sabatini, 2010; Randhawa et al., 2023; Skovgaard et al., 2023). These different approaches are based on processing the raw signal from the accelerometer and are generally included in the software of research-grade devices. However, consumer-grade devices do not typically include such algorithms or provide access to raw data. Recently, higher end consumer-grade device manufacturers have proposed the use of minute-per-minute HR data to estimate wearing time. To the best of our knowledge, only one validation study has been published for this method (Gorny et al., 2017), and only one research protocol has attempted to evaluate the wearing time using HR records (Claudel et al., 2020). Last, because the core of PA evaluation relies on the treatment of accelerometer data, some researchers may prefer disabling the HR functionality to preserve the memory and battery life of the devices.

To summarize, although the distinction between nonwear and sedentary times is necessary to ensure the reliability of daily PA estimates, no previous study using Fitbit devices has ever mentioned a "valid wear" evaluation protocol. The present study introduced and tested a new configurable Python software that allows the evaluation of days of valid wear by exploiting minute-per-minute time-series data collected from the official Fitbit Web API, compared with using HR.

Software Description

The software is named "ActiWearCheck" and was developed in Python primarily using the NumPy (Harris et al., 2020) and Pandas (McKinney, 2010) libraries. The full code is open source and accessible elsewhere under the MIT License (https://github.com/OchaUni-Physical-Activity-Measurement/ActiWearCheck). This software was developed to identify the days of valid wear in the Fitbit activity tracker time series collected in the drePAnon clinical trial (UMIN000042826).

Concept Used for Evaluating Valid Wear

The main inputs used by the software are minute-per-minute EE data (in kilocalories per minute). Although minute-per-minute data cannot be downloaded through the official Fitbit application, they are accessible through the Fitbit Web API, which is freely available to third-party developers. Because the decimal rounding used to report hourly data may not reveal a slight EE increase associated with small and rare movements executed during sedentary behaviors, accessing minute-per-minute data is critical for the evaluation of valid wear. Indeed, the slightest decimal deviation between a 1-min EE estimate and the EE corresponding to basal metabolic rate (BMR) data indicates that a movement was performed and that the activity tracker was worn during this minute. Depending on the wearing criteria decided by the researcher, valid wear can be assessed for each monitoring day based on the number of minutes that present EE estimates higher than those corresponding to the BMR (see Figure S1 in the Supplementary Material [available online] for an illustrative example).

Software Configuration

To test ActiWearCheck, two configurations, based on minute counts above the estimated BMR and their periodicity, were proposed to evaluate the days of valid and invalid wear (Figure 1).

Method Cal-worn: Days of valid wear required a minimum of 600 min above the estimated BMR, as previously proposed (Gorny et al., 2017; Migueles et al., 2017).

Method Cal-worn (per hour): Days of valid wear required a minimum of 10 hr, containing at least 1 min above the BMR, adapted from a previous study on the treatment of Actigraph data (Choi et al., 2012).

Methods Cal-worn and *Cal-worn (per hour)* corresponded to two different software configurations. The valid wear estimates of these two configurations were compared with the outcome of the HR-based method, which was described as follows:

Method HR-worn: Days of valid wear required a minimum of 600 min with HR data.

Example Case

The analysis was performed using Python (Jupyter Notebook, version 6.4.12) and the SciPy library (Virtanen et al., 2020).

Population and Protocol

The data used for the analyses were acquired during the drePAnon clinical trial, which received ethical approval from Cheikh Anta Diop University, Dakar, Senegal (number: 0388/2019/CER/UCAD). Seventy-two out of 124 young Senegalese men (28.6 ± 7.9 years of age, including 52 patients with sickle cell anemia and 20 participants without hemoglobinopathy) were enrolled in the clinical trial and were equipped with a wrist-worn Fitbit activity tracker device on the nondominant arm to record both accelerometer-derived PA parameters and HR data. The study protocol consisted of a monitoring period of 5–15 weeks. The participants were instructed to wear their activity tracker device continuously throughout the day and to remove it only during showering, bathing, or performing water-related activities. They were advised to charge the device while showering/bathing or at bedtime if the battery was almost empty. Participants who reported sleep quality impairment due to wearing the activity tracker device were allowed to wear it during waking hours only. The participants had access to the Fitbit application, which was configured to show only the time and

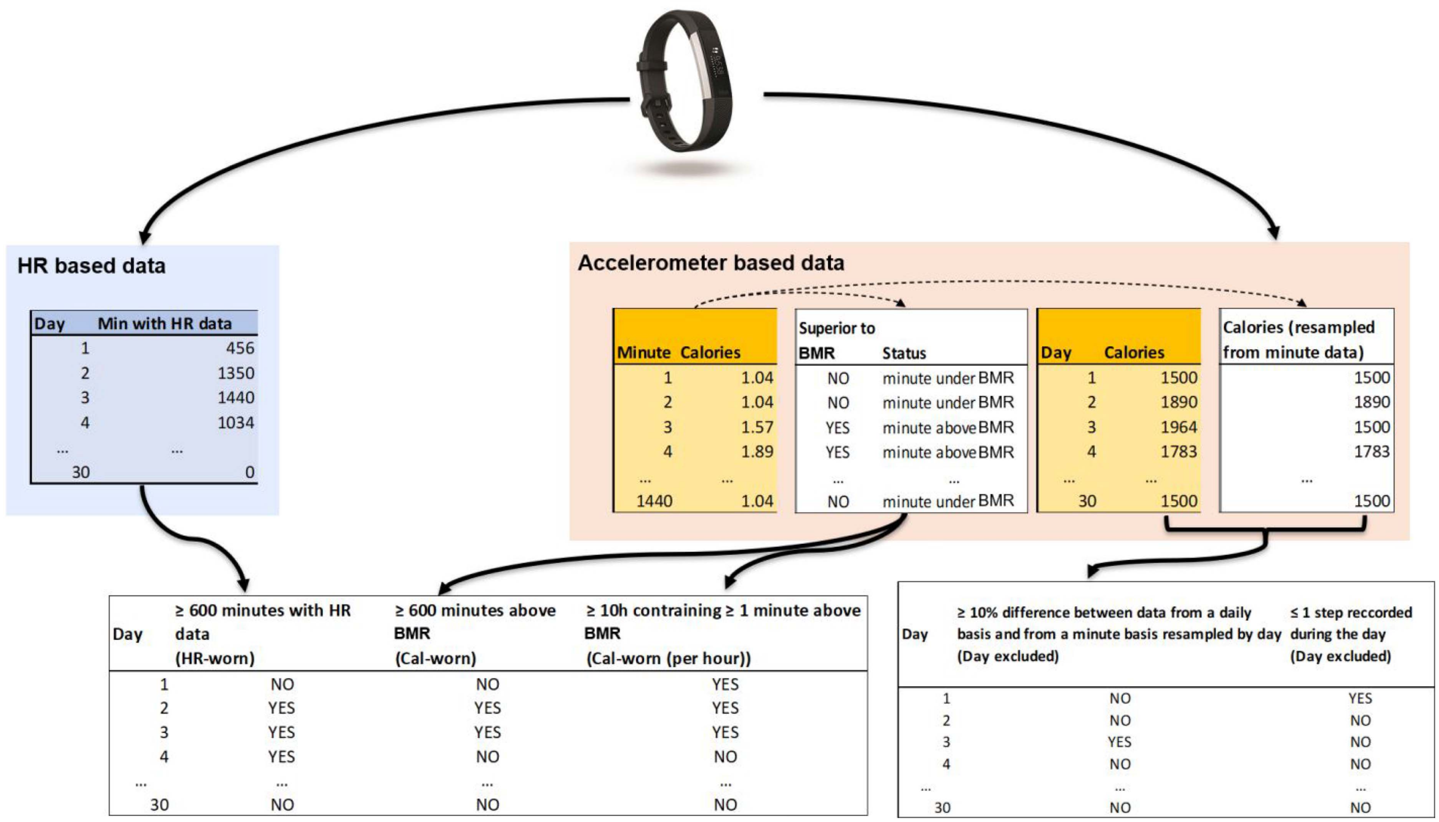

Figure 1 — Data processing flow used to estimate days of valid wear using three different methods. The Fitbit Web API provides the number of minutes and HR data for each day (called "WearTime" in the Fitabase third party application) and the EE estimates for each minute and day. The presence of one step count and absence of substantial differences between the data sets from a daily basis and a minute basis resampled by day were systematically checked (see Figure 2). *Method Cal-worn* evaluates whether there are at least 600 min above the BMR throughout the day. *Method Cal-worn (per hour)* evaluates whether there are at least 10 hr containing at least 1 min above the BMR throughout the day. *Method HR-worn* evaluates whether there are at least 600 min of HR data throughout the day. BMR = basal metabolic rate; Cal = calories; EE = energy expenditure; HR = heart rate.

battery levels. All notifications were disabled. Data presented in the example case were collected between February 2021 and August 2023. Each participant was monitored for an average of 72 ± 41 days.

Material

The participants received a Fitbit Alta HR or Inspire 2 activity tracker device (Fitbit), which was paired with the smartphone Fitbit application. Fifty-six participants wore Alta HR devices, while 16 wore Inspire 2 devices. The protocol was modified to use the latter in January 2022, due to market unavailability of the former. Although the official Fitbit application provides access to daily summaries only, the Fitbit Web API allows access to minute-per-minute data. In the drePAnon study, the Fitabase platform (Small Steps Labs LLC) was used to download minute-per-minute data and daily summaries of step counts and EE parameters. The HR data were also accessed through the Fitabase server.

Both the Fitbit Alta HR and Fitbit Inspire 2 have memory limitations that prevent them from recording minute-by-minute data after a few days (5 days for the Fitbit Alta HR and 7 days for the Fitbit Inspire with optimal configuration). In case of synchronization failure between the activity tracker device and smartphone handset, the minute-per-minute data cannot be transferred to the cloud, and the record of minute-by-minute data is erased to make room for the storage of daily summaries. This characteristic allows the identification of days with unreliable minute-per-minute records, as described in the "Times Series Cleaning Procedure" section and in Figure 2.

Times Series Cleaning Procedure

Before applying *Cal-worn, Cal-worn (per hour),* and *HR-worn*, the time series were preprocessed to remove days 1) with no step count record, which indicates that the device was not worn or that the battery was discharged during that day, or 2) a discrepancy occurred between the daily EE estimates provided by the original Fitbit application and the EE estimates computed by resampling the minute-per-minute EE data, which in turn indicates that the minute-per-minute data record was corrupted (possibly due to memory or sync issues). The preprocessing procedure is shown in Figure 2.

Of the initial 3,339 days of observation, 3,282 days contained at least one recorded step. Among them, 3,253 days did not show any differences (±10%) between daily summaries and daily estimates calculated using the resampled minute-per-minute data. This represents 97% of the initial data set that underwent a valid wear estimation procedure. The participants were rather active (9,037 ± 5,052 steps and 2,668 ± 628 kcal per day).

Test 1: Accuracies of the Valid Wear Estimation Methods

Each of the 3,253 days that were selected after the above cleaning procedure was classified as "valid wear" or "invalid wear" using *Cal-worn* and *Cal-worn (per hour)* in the developed software. In

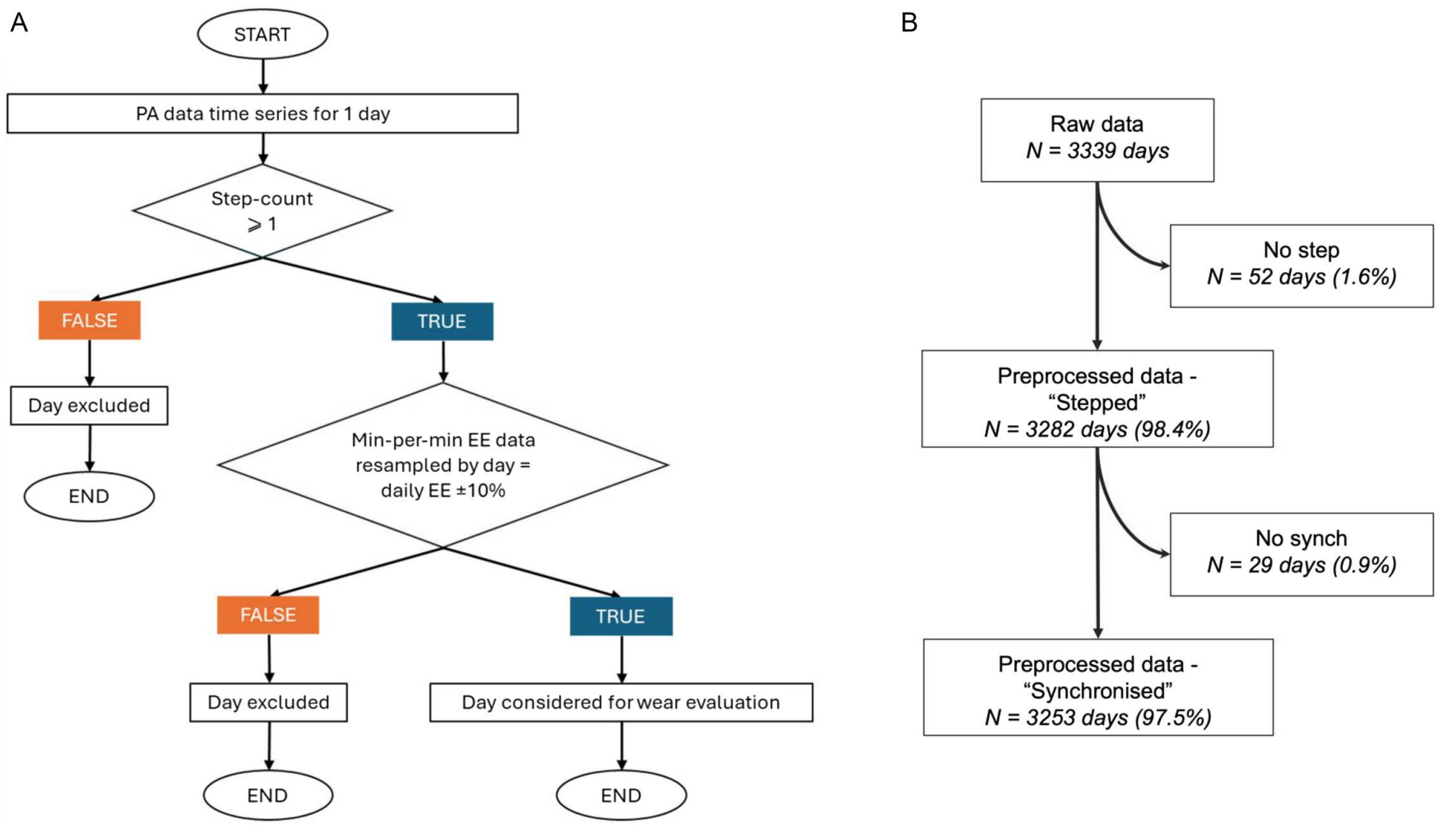

Figure 2 — Flowchart of data preprocessing. (A) Algorithmic logic of the data ActiWearCheck preprocessing flow. (B) Outcomes of data preprocessing for the example case. EE = energy expenditure; PA = physical activity.

the absence of a gold standard, the classifications of these two methods were compared with that of *HR-worn*, which was unlikely to produce false positives. The results are presented as confusion matrices in Figure 3. The accuracies were calculated by summing all true positive and true negative outcomes divided by the total number of predictions (here, 3,253 days).

The accuracy of *Cal-worn* (at least 600 min above BMR throughout the day) against *HR-worn* (at least 600 min of HR data throughout the day) was 72.0%. As shown in Figure 3, 83.6% of the days evaluated as valid wear using *HR-worn* were similarly evaluated using *Cal-worn*. Conversely, 16.4% of the days evaluated as valid wear using the *HR-worn* were evaluated as invalid wear using *Cal-worn*. In addition, 76.5% of the days evaluated as invalid wear using *HR-worn* were similarly evaluated using *Cal-worn*, and 23.5% of the days evaluated as invalid wear using the *HR-worn* were evaluated as valid wear using *Cal-worn.*

The accuracy of *Cal-worn (per hour)* (at least 10 hr containing 1 min above the BMR throughout the day) against *HR-worn* (at least 600 min of HR data throughout the day) was 86.0%. As shown in Figure 3, 100% of the days evaluated as valid wear using *HR-worn* were also similarly evaluated using *Cal-worn (per hour)*. In addition, 39.3% of the days evaluated as invalid wear using *HR-worn* were similarly evaluated using *Cal-worn (per hour)* and 60.7% of the days evaluated as invalid wear using the *HR-worn* were evaluated as valid wear using *Cal-worn (per hour)*. Further analyses are provided as Supplementary Material (available online), and Figure S2 (see Supplementary Material [available online]) shows that the more active the participant was in the day, the higher the agreement between the methods. Furthermore, Figure S3 (see Supplementary Material [available online]) provides comparative metrics between methods regarding PA values for each subject, including the results of the mean average error, mean average percentage error, and root mean square error tests. The figure shows overall a good concordance between methods (mean average percentage error between ~1.5% and 6%). Yet, larger discrepancies for some subjects show why relying on a single method could incorrectly consider that the tracker was not worn, justifying the proposed approach.

Test 2: Impact of Using the Valid Wear Estimation Software on General Statistics

The estimation of days of valid wear performed using *HR-worn* resulted in the exclusion of 14% of the days. *Cal-worn* and *Cal-worn (per hour)* excluded 25% and 5% of the days, respectively. Sixty-eight of the 72 participants had more than 1 day of valid wear application of *Cal-worn* or *Cal-worn (per hour)* and were included in the statistical analysis. Considering the three tested methods, it is noteworthy that the higher the number of excluded days, the higher the average step count and EE. The results of the data selection process for the entire population along with the mean daily step counts and EE estimates are presented in Table 1. The average daily step count and EE predictions were calculated for each participant and method, and a Friedman test (nonnormal distribution of EE values, Shapiro–Wilk test, $p < .05$) was employed to assess the effect of different valid wear methods (*Cal-worn, Cal-worn [per hour], HR-worn,* and none) on the predictions. The method had an effect on both the step count and EE predictions ($p < .001$). Conover post hoc tests indicated a statistically significant difference

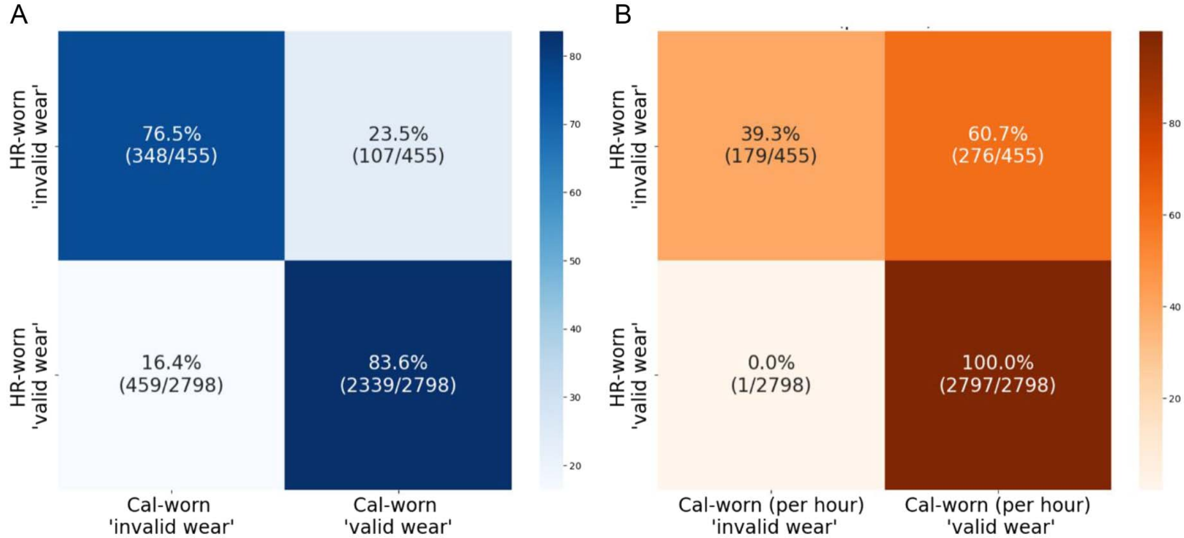

Figure 3 — Accuracy of accelerometer data-based evaluation of days of valid wear relative to Method HR. (A) Accuracy of wear classification for *Cal-worn* versus *HR-worn*. (B) Accuracy of wear classification for *Cal-worn (per hour)* versus *HR-worn*, expressed in percentage. Cal = calories; HR = heart rate.

Table 1 Day and Step Counts and EE After Application of the Preprocessing Procedure Using *Cal-Worn, Cal-Worn (per Hour)*, and *HR-Worn*

	Day count	Steps (step/day)	EE (kcal/day)
Before cleaning	3,339	8,896 ± 5,144	2,644 ± 641
After cleaning ("No Method")	3,253	9,037 ± 5,052	2,668 ± 628
Valid wear evaluation			
Cal-worn	2,446 (75%)	10,137 ± 4,563	2,837 ± 592
Cal-worn (per hour)	3,073 (95%)	9,292 ± 4,861	2,708 ± 612
HR-worn	2,798 (84%)	9,327 ± 4,752	2,742 ± 617

Note. The cleaning procedure consisted of removing days without any step record or with a difference greater than 10% between the daily total of minute-per-minute data and daily summaries. Percentages are calculated for *Cal-worn, Cal-worn (per hour)*, and *HR-worn*, relative to the number of days included after the preprocessing procedure (*No Method*). Cal = calories; EE = energy expenditure; HR = heart rate.

between the step count and EE estimates calculated after the application of the *Cal-worn* and the ones obtained without any valid wear evaluation procedure ($p = .001$ and .001, respectively). These differences are illustrated in Figure 4. The outcome of the valid wear evaluation software on the step count outcome for a selected participant is illustrated in Figure 5.

Finally, in the absence of a gold standard method, the average step count and EE values representing days of valid wear for each participant, estimated using *Cal-worn* or *Cal-worn (per hour),* were compared with those obtained for days of valid wear, estimated using *HR-worn,* using Bland–Altman plots. These plots are presented in Figure 6. The results revealed two key differences between methods. First, the *Cal-worn* method showed a slight positive bias (+738 steps/day, +88 kcal/day), suggesting a tendency to overestimate step count and EE compared with the *HR-worn* method. In contrast, the *Cal-worn (per hour)* method produced estimates closer to *HR-worn* (−227 steps/day, +60 kcal/day). Second, the limits of agreement indicated that step count estimates were more dispersed with the *Cal-worn* method (±15.5%) than with the *Cal-worn (per hour)* method (±8.7%), suggesting better interchangeability with *HR-worn* in the latter case. For EE, the limits of agreement were lower and similar between methods, that is, ±7.0% for *Cal-worn* and ±6.7% for *Cal-worn (per hour)*.

Taken together, these results indicate that *Cal-worn (per hour)* is more conservative than *Cal-worn* in the estimation of days of valid wear, resulting in step count and EE predictions that are closer to those obtained after the application of *HR-worn*. All data and scripts used for the analysis are available at https://github.com/OchaUni-Physical-Activity-Measurement/ActiWearCheck.

Discussion

In this study, a new open-access Python software ("ActiWear Check") was presented, enabling the selection of days of valid wear in PA studies using consumer-grade accelerometer-based activity trackers. The software excluded days of "invalid wear" from the data set, thus enhancing the validity of PA values considered for further analyses. Days of valid wear can be selected using either HR or accelerometer data. If accelerometer data are used, the evaluation criteria included in the software can be configured in two ways: (a) the total number of minutes above the BMR that the activity tracker should reach or (b) the minimum number of minutes above the BMR per hour for a specified number of hours. In the example case, the latter configuration set to identify days of valid wear as having ≥10 hr with at least 1 min of EE over

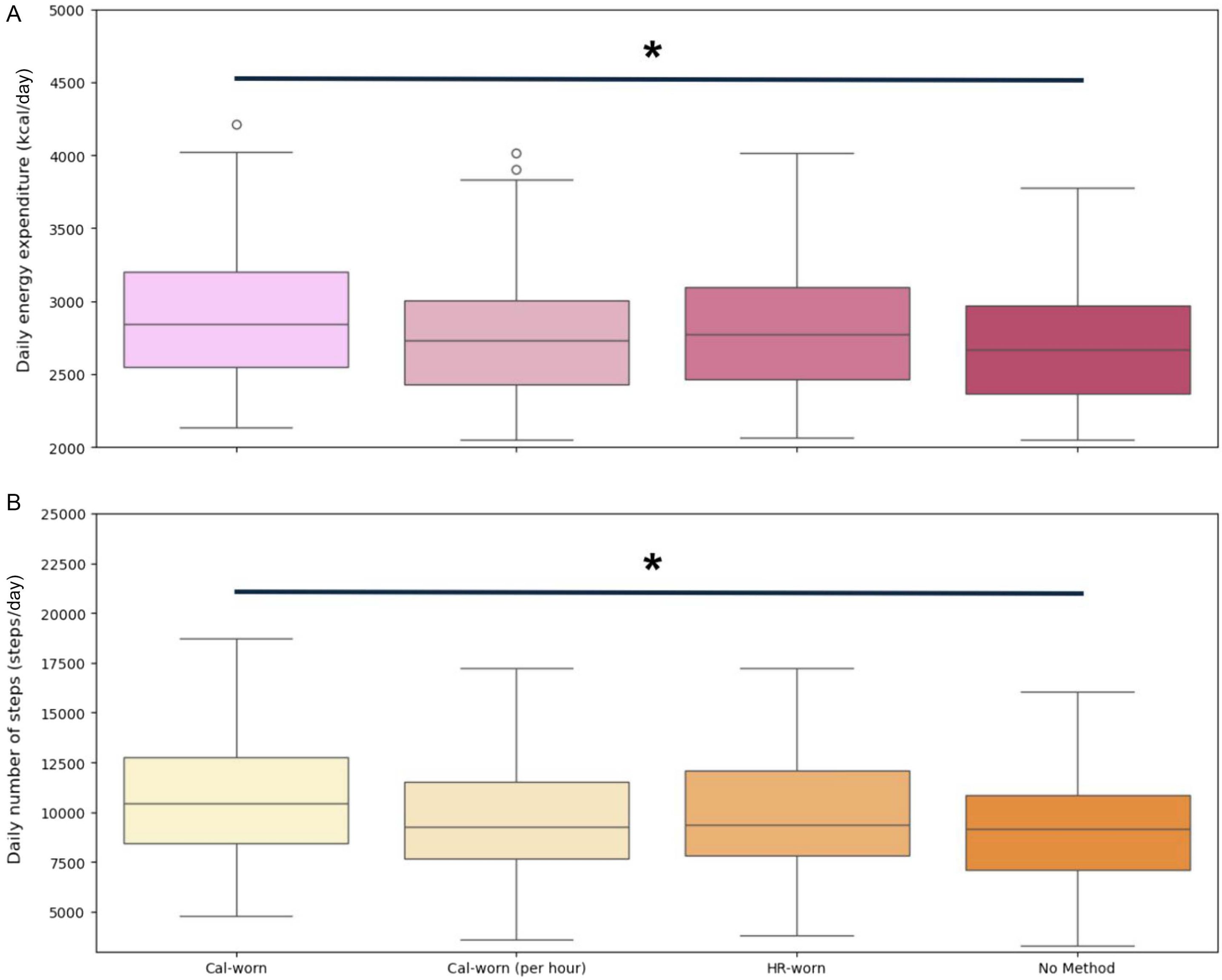

Figure 4 — Comparison of methods in relation to daily estimates of step count and energy expenditure. (A) Daily energy expenditure estimates. (B) Daily number of step estimates. Sixty-eight of the 72 participants had more than 1 day of valid wear after the application of *Cal-worn* and *Cal-worn (per hour)* and were therefore included in the statistical analysis. *No Method* corresponds to data only treated with the *times series cleaning procedure*. ○: outlier data. $^{*}p < .01$. Cal = calories; HR = heart rate.

the BMR (*Cal-worn [per hour]*) presented satisfying accuracies using the selection method based on HR data (*HR-worn*). The average step count and EE predictions calculated for the participant also exhibited low dispersion scores in relation to the HR data-based selection method.

Applying ActiWearCheck to a Set of Fitbit Data

ActiWearCheck was configured to use *Cal-worn* and *Cal-worn (per hour)*, respectively, with the minute count and periodicity criteria described above to evaluate the days of valid wear for the wrist-worn Fitbit activity trackers used in the drePAnon clinical study ($n = 72$), compared with the HR-based method. *Cal-worn* and *Cal-worn (per hour)* showed accuracies of 72% and 86%, respectively, compared with the *HR-worn* for the classification of valid/invalid wear days. In the example case, the valid wear evaluation method can exclude up to 25% of daily records. Excluding the days during which the activity tracker device was only partially worn increased the average daily step count and EE values. In some cases, the overall step count or EE estimates were reduced. For example, some individuals could have walked continuously from 8 a.m. to 11 a.m. and remained completely inactive for the rest of the day. In this case, this particular day might be classified as "nonvalid" and the activity performed during the morning will not be accounted for, potentially reducing the average step count and EE estimates. However, this scenario did not reflect the observed general case (Figure S2 in the Supplementary Material [available online]). In a recent review of the literature assessing the accuracy of consumer-grade activity tracker devices, Germini et al. (2022) reported that Fitbit devices correctly estimate steps but are not accurate in measuring EE. In another review, Chevance et al. (2022) reported that Fitbit devices are likely to underestimate the

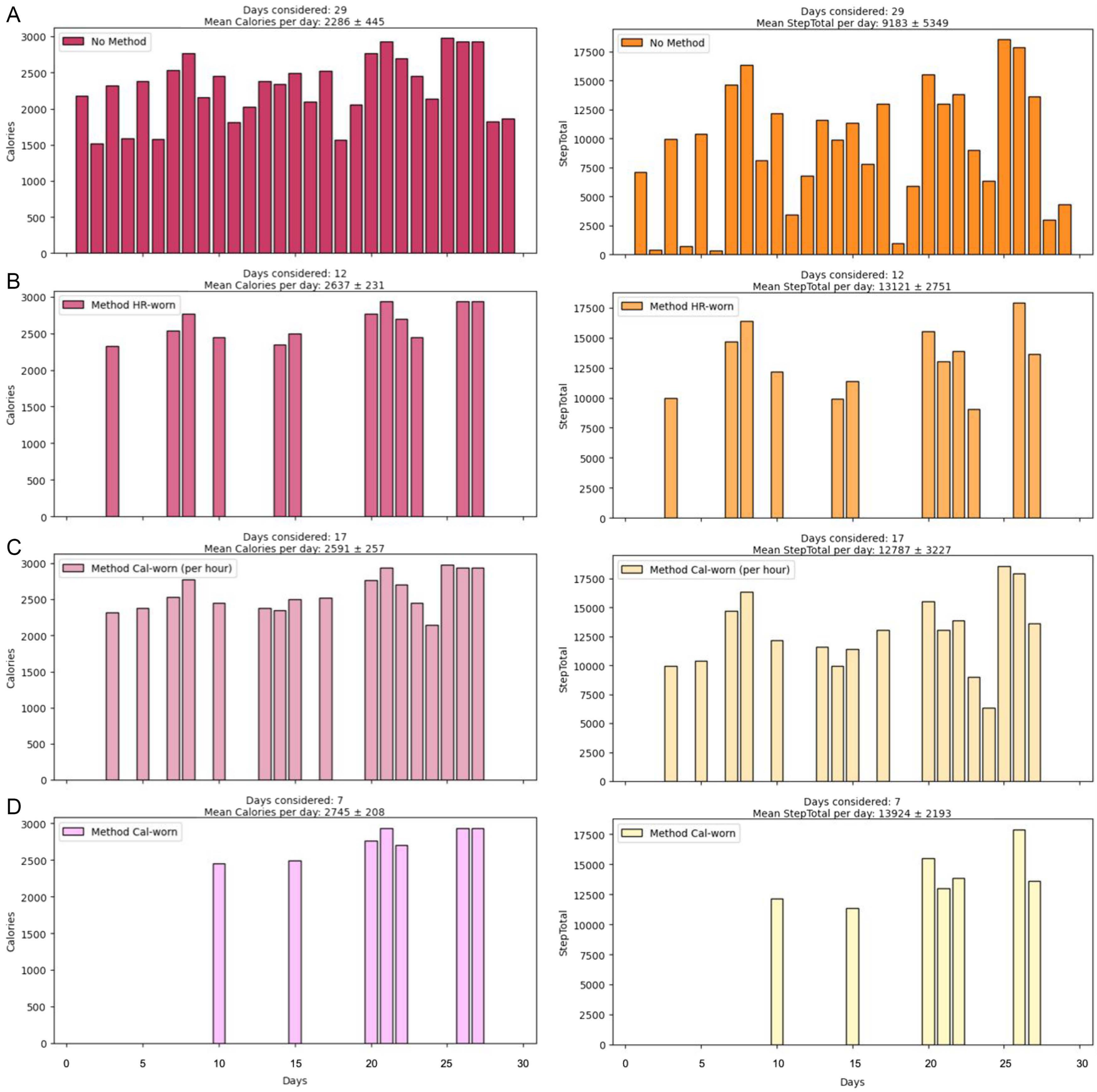

Figure 5 — Effect of *HR-worn, Cal-worn (per hour),* and *Cal-worn* on day selection and calculation of energy expenditure (left) and step count (right) estimates for one selected participant during a 30-day period. (A) Data were treated with the *times series cleaning procedure*, but no assessment of valid wear was performed. (B) Valid days assessed using *HR-worn* (at least 600 min of HR data throughout the day). (C) Days assessed using *Cal-worn (per hour)* (at least 10 hr containing at least 1 min above the BMR throughout the day). (D) Days assessed using *Cal-worn* (at least 600 min above the BMR throughout the day). HR = heart rate; Cal = calories; BMR = basal metabolic rate.

step count and EE. Therefore, the application of valid wear criteria using a simple Python software has the potential to offer substantial benefits for the accuracy of daily estimates of PA metrics in studies employing consumer-grade accelerometer-based activity tracker devices. Specifically, regarding the underestimation issue noted in a previous study (Chevance et al., 2022), implementing ActiWearCheck can assist in identifying and excluding days with partial wear thereby preventing them from contributing to artificially low estimates.

Data Treatment and Configuration

In the context of consumer-grade activity trackers, manufacturers usually do not provide access to raw accelerometer data or wearing time processing algorithms. The proposed Python software relies on the treatment of minute-per-minute EE, a physiological estimate that Fitbit measures with increments above the BMR at the first decimal resolution. Therefore, the minimum change observable at this resolution was assumed to capture the smallest changes above the resting EE, which are typically movements executed during sedentary behavior. To evaluate wearing time, the resolution offered by Fitbit devices, that is, minute summaries, may be sufficient to counterbalance the lack of access to raw data. While future studies are needed to establish the true sensitivity of Fitbit devices, Figure S1 (see Supplementary Material [available online]) provides an example suggesting that EE may be responsive to fine or upper body movements, such as those occurring during sedentary or low-intensity activities. In addition, the configuration used in *Cal-worn*

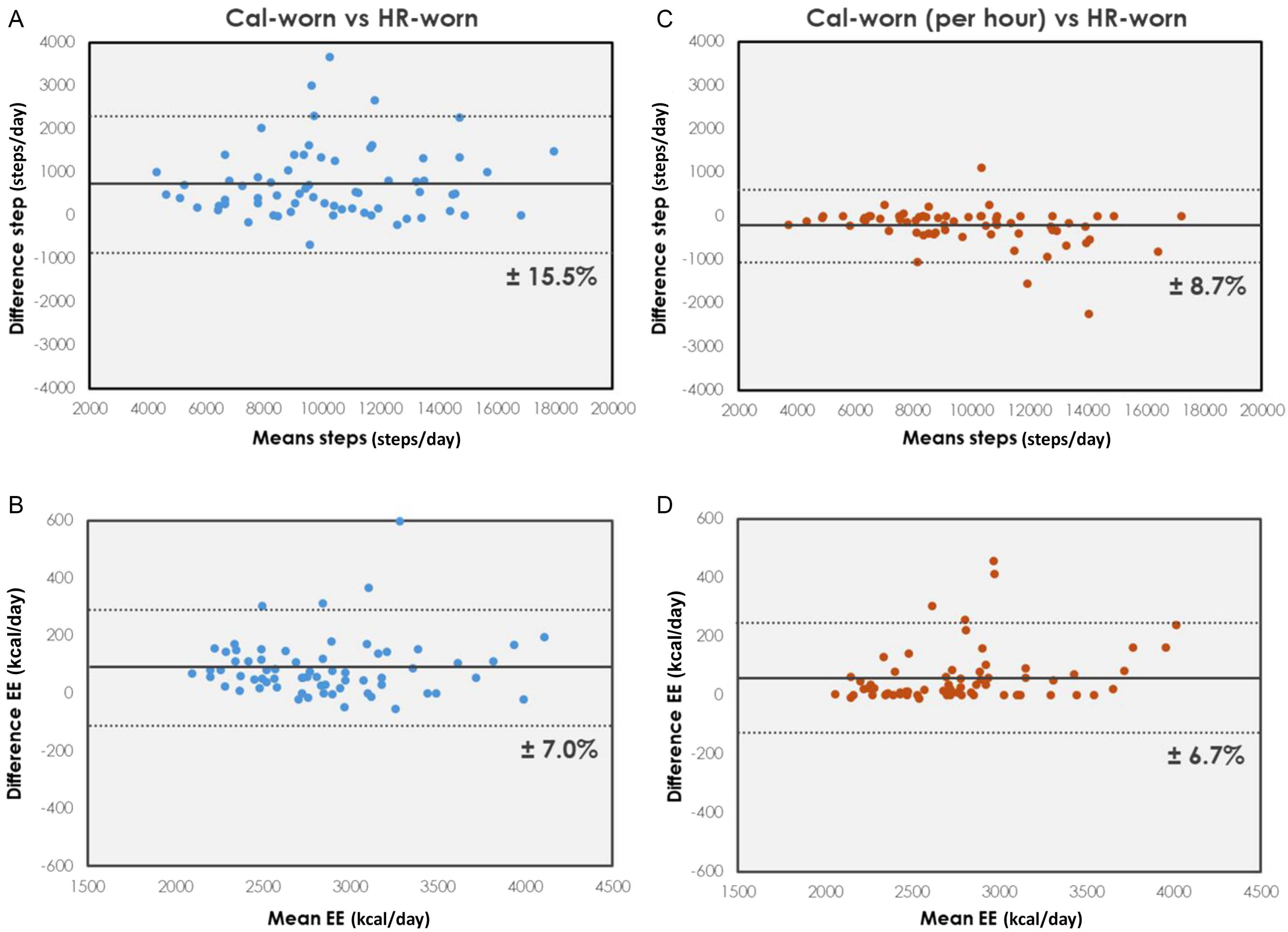

Figure 6 — Difference plot (Bland–Altman) between *Cal-worn* and *HR-worn*, and between *Cal-worn (per hour)* and *HR-worn*. (A) *Cal-worn* versus *HR-worn* regarding differences in average step count estimates. (B) *Cal-worn* versus *HR-worn* regarding differences in average EE estimates. (C) *Cal-worn (per hour)* versus *HR-worn* regarding differences in average step count estimates. (D) *Cal-worn (per hour)* versus *HR-worn* regarding differences in average EE estimates. One data point corresponds to the average step count or EE estimates calculated for days of valid wear for one participant. The dispersion rate around the mean is indicated in the plot. Cal = calories; HR = heart rate; EE = energy expenditure.

(per hour) can be employed to address the issue related to the possible inability of minute-per-minute EE estimates to capture all movements of sedentary behaviors. Indeed, some physical behaviors executed during sedentary hours (a few steps to the bathroom, changing posture, etc.) may allow the activation of the Fitbit algorithm in a way that will increase the EE estimates by one minute. Thus, only 10 min above the BMR equally distributed over 10 of the 24 hr were needed to include a day in the analysis. In addition, the software can be configured to more or less conservative criteria (e.g., at least 6 hr with at least 2 min above the BMR or 16 hr with at least 1 min above the BMR) depending on the characteristics of the studied population. The additional configurations proposed in the software were not tested in the example case. The *valid wear (steps)* method checks that there are enough steps in a given day, and the *valid wear (steps, hours)* method checks that enough hours in a given day included at least a few steps (parameterizable). Restricting valid wear analysis to the usual waking hours (e.g., from 5:00 to 22:59) or any other period of the day is also possible. Finally, the software is open source, and researchers can add criteria to fulfill the needs of their protocol. For instance, one may be interested in validating a day as worn after a certain number of realized steps or considering bouts of continuous PA (Hardcastle et al., 2020).

Limitations and Strengths

First, the lack of a gold standard measure of the actual wearing time is a limitation of the present analytical design. In this study, the HR-based parameter was compared with the evaluations performed using *Cal-worn* and *Cal-worn (per hour)*, which were solely based on accelerometer-derived PA parameters. However, the validity of HR measurements using activity trackers is still being investigated (Claudel et al., 2020; Gorny et al., 2017), and studies evaluating the use of this parameter to assess the valid wear of the device are yet to be published. Moreover, the HR lens can be manually disabled by the participant through the Fitbit application; at least this was the case for the Alta HR and Inspire 2 devices at the time of the drePAnon study. Future studies should aim to assess the validity of valid wear evaluations using consumer-grade accelerometer-based devices, considering control procedures, such as wearing time diaries or the concomitant use of activity trackers compatible with open research that allows the extraction of raw accelerometer data. Nevertheless, without considering a gold standard, a method based on HR data is unlikely to produce false-positive outcomes (i.e., detect heartbeat when not worn), allowing for some level of interpretation of the example case results. The handling of false negatives (no HR but wear detected through step or EE estimates) is left to the user of ActiWearCheck. In the absence of an appropriate gold standard, HR measures proposed by some manufacturers to evaluate valid wear provide a reference for evaluating the proposed software.

Second, the example case illustrated the functionality of the software user data from a study conducted in Senegal. The association between darker skin and issues in HR measurement accuracy has been discussed (Bent et al., 2020; Shcherbina et al., 2017), questioning the reliability of such methods in non-White populations. However, errors reported by Bent et al. (2020) for darker skin tones measured 3–4 beats/min compared with lighter tone skins. Regardless of the skin tone, errors typically measure approximately 5 beats/min at rest and 15 beats/min during walking. Although the differences noted by Bent et al. (2020) and other authors may be of primary importance for research aimed at improving the accuracy of HR measurements, it is less critical for binary wear/nonwear evaluation relevant to ActiWearCheck.

Finally, the software is currently designed to work with data exported from Fitabase. To address this limitation, a program that converts JSON data obtained from the Fitbit Web API is also provided to facilitate the use of the software. The software will also benefit from contributions aimed at generalizing its use to data obtained from a wider range of activity tracker devices.

Conclusions

ActiWearCheck is a new Python software that enables the selection of days of valid wear in PA studies. This software is open access (https://github.com/OchaUni-Physical-Activity-Measurement/ActiWearCheck). Although consumer-grade activity tracker devices have been linked to frequent underestimations of daily step counts and EE (Chevance et al., 2022), researchers using this type of device, especially accelerometer-based monitors, may benefit from the use of ActiWearCheck to improve the accuracy of their PA outcomes, as suggested in the example case presented in this manuscript.

Acknowledgments

The authors would like to thank all participants of the drePAnon clinical trial (UMIN000042826) who consented to participate in the study and whose data were used in the example case shown in the present manuscript. The authors also thank the reviewers for their contributions in improving the software functionalities. **Funding**: This study was supported by the Japanese Society for the Promotion of Science (19KK0248). **Author Contributions**: *Study design, software development, data collection, data cleaning, data processing, statistical analysis, data interpretation, manuscript draft, and data and code repository maintenance:* Gallou-Guyot. *Fund acquisition, data collection, manuscript review:* Diaw. *Software development, data and code repository maintenance, and software improvement:* Aubert-Kato. *Data collection and manuscript review:* Saloum Coly. *Data collection and manuscript review*: Charlot. *Fund acquisition, data collection, and manuscript review:* Connes. *Fund acquisition and manuscript review:* Yoshida. *Fund acquisition, data interpretation, and manuscript review:* Miyachi. *Study design, fund acquisition, software development, data collection, data cleaning, data interpretation, manuscript draft, and data and code repository maintenance*: Tripette.

References

Actigraph. (2013, May 13). *GT3X+ and wGT3X+—Device Manual*. https://dl.theactigraph.com/gt3xp_wgt3xp_device_manual.pdf

Actigraph. (2024, December 4). *ActiGraph LEAP—Instruction for use*. ™https://6407355.fs1.hubspotusercontent-na1.net/hubfs/6407355/User%20Manuals/ActiGraph%20LEAP™%20User%20Manual.pdf

Activinsights. (n.d.). *FAQs—GENEActiv*. https://activinsights.com/resources/faqs/

Banda, J.A., Haydel, K.F., Davila, T., Desai, M., Bryson, S., Haskell, W.L., Matheson, D., & Robinson, T.N. (2016). Effects of varying epoch lengths, wear time algorithms, and activity cut-points on estimates of child sedentary behavior and physical activity from accelerometer data. *PLoS One, 11*(3), Article e0150534. https://doi.org/10.1371/journal.pone.0150534

Bent, B., Goldstein, B.A., Kibbe, W.A., & Dunn, J.P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digital Medicine, 3*(1), Article 18. https://doi.org/10.1038/s41746-020-0226-6

Cain, K.L., Sallis, J.F., Conway, T.L., Dyck, D.V., & Calhoon, L. (2013). Using accelerometers in youth physical activity studies: A review of methods. *Journal of Physical Activity and Health, 10*(3), 437–450. https://doi.org/10.1123/jpah.10.3.437

Chen, K.Y., & Bassett, D.R. (2005). The technology of accelerometry-based activity monitors: Current and future. *Medicine & Science in Sports & Exercise, 37*(11), S490–S500. https://doi.org/10.1249/01.mss.0000185571.49104.82

Chevance, G., Golaszewski, N.M., Tipton, E., Hekler, E.B., Buman, M., Welk, G.J., Patrick, K., & Godino, J.G. (2022). Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing fitbits against reference measures: Systematic review and meta-analysis. *JMIR mHealth and uHealth, 10*(4), Article e35626. https://doi.org/10.2196/35626

Choi, L., Ward, S.C., Schnelle, J.F., & Buchowski, M.S. (2012). Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Medicine and Science in Sports and Exercise, 44*(10), 2009–2016. https://doi.org/10.1249/MSS.0b013e318258cb36

Claudel, S.E., Tamura, K., Troendle, J., Andrews, M.R., Ceasar, J.N., Mitchell, V.M., Vijayakumar, N., & Powell-Wiley, T.M. (2020). Comparing methods to identify wear-time intervals for physical activity with the fitbit charge 2. *Journal of Aging and Physical Activity, 29*(3), 529–535. https://doi.org/10.1123/japa.2020-0059

Diaz, K.M., Krupka, D.J., Chang, M.J., Peacock, J., Ma, Y., Goldsmith, J., Schwartz, J.E., & Davidson, K.W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology, 185*, 138–140. https://doi.org/10.1016/j.ijcard.2015.03.038

Fitbit. (2021). *Fitbit Inspire 2 Manual*. https://help.fitbit.com/manuals/manual_inspire_2_fr.pdf

Germini, F., Noronha, N., Borg Debono, V., Abraham Philip, B., Pete, D., Navarro, T., Keepanasseril, A., Parpia, S., de Wit, K., & Iorio, A. (2022). Accuracy and acceptability of wrist-wearable activity-tracking devices: Systematic review of the literature. *Journal of Medical Internet Research, 24*(1), Article e30791. https://doi.org/10.2196/30791

Gorny, A.W., Liew, S.J., Tan, C.S., & Müller-Riemenschneider, F. (2017). Fitbit charge HR wireless heart rate monitor: Validation study conducted under free-living conditions. *JMIR mHealth and uHealth, 5*(10), Article e157. https://doi.org/10.2196/mhealth.8233

Hardcastle, S.J., Jiménez-Castuera, R., Maxwell-Smith, C., Bulsara, M.K., & Hince, D. (2020). Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: Exploratory analysis from a randomised controlled trial. *PLoS One, 15*(10), Article e0240967.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., … Oliphant, T.E. (2020). Array programming with NumPy. *Nature, 585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Mannini, A., & Sabatini, A.M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors, 10*(2), 1154–1175. https://doi.org/10.3390/s100201154

McKinney, W. (2010). *Data structures for statistical computing in python*. Proceedings of the 9th Python in Science Conference, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Migueles, J.H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J.R., & Ortega, F.B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine, 47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Murakami, H., Kawakami, R., Nakae, S., Nakata, Y., Ishikawa-Takata, K., Tanaka, S., & Miyachi, M. (2016). Accuracy of wearable devices for estimating total energy expenditure: Comparison with metabolic chamber and doubly labeled water method. *JAMA Internal Medicine, 176*(5), 702–703. https://doi.org/10.1001/jamainternmed.2016.0152

Murakami, H., Kawakami, R., Nakae, S., Yamada, Y., Nakata, Y., Ohkawara, K., Sasai, H., Ishikawa-Takata, K., Tanaka, S., & Miyachi, M. (2019). Accuracy of 12 wearable devices for estimating physical activity energy expenditure using a metabolic chamber and the doubly labeled water method: Validation study. *JMIR mHealth and uHealth, 7*(8), Article e13938. https://doi.org/10.2196/13938

Nakagata, T., Murakami, H., Kawakami, R., Tripette, J., Nakae, S., Yamada, Y., Ishikawa-Takata, K., Tanaka, S., & Miyachi, M. (2022). Step-count outcomes of 13 different activity trackers: Results from laboratory and free-living experiments. *Gait & Posture, 98*, 24–33. https://doi.org/10.1016/j.gaitpost.2022.08.004

Randhawa, S., Sharma, M., Fiterau, M., Banda, J.A., Haydel, F., Kapphahn, K., Matheson, D., Moore, H., Ball, R.L., Kushida, C., Delp, S., Wall, D.P., Robinson, T., & Desai, M. (2023). Statistical learning methods to identify nonwear periods from accelerometer data. *Journal for the Measurement of Physical Behaviour, 6*(2), 124–133. https://doi.org/10.1123/jmpb.2022-0034

Shcherbina, A., Mattsson, C.M., Waggott, D., Salisbury, H., Christle, J.W., Hastie, T., Wheeler, M.T., & Ashley, E.A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine, 7*(2), Article 3. https://doi.org/10.3390/jpm7020003

Skovgaard, E.L., Roswall, M.A., Pedersen, N.H., Larsen, K.T., Grøntved, A., & Brønd, J.C. (2023). Generalizability and performance of methods to detect non-wear with free-living accelerometer recordings. *Scientific Reports, 13*(1), Article 2496. https://doi.org/10.1038/s41598-023-29666-x

Sushames, A., Edwards, A., Thompson, F., McDermott, R., & Gebel, K. (2016). Validity and reliability of fitbit flex for step count, moderate to vigorous physical activity and activity energy expenditure. *PLoS One, 11*(9), Article e0161224. https://doi.org/10.1371/journal.pone.0161224

Tully, M.A., McBride, C., Heron, L., & Hunter, R.F. (2014). The validation of Fitbit Zip™ physical activity monitor as a measure of free-living physical activity. *BMC Research Notes, 7*(1), Article 952. https://doi.org/10.1186/1756-0500-7-952

Vanhelst, J., Vidal, F., Drumez, E., Béghin, L., Baudelet, J.-B., Coopman, S., & Gottrand, F. (2019). Comparison and validation of accelerometer wear time and non-wear time algorithms for assessing physical activity levels in children and adolescents. *BMC Medical Research Methodology, 19*(1), Article 72. https://doi.org/10.1186/s12874-019-0712-1

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2